Geometry and physics of black holes

Lecture notes

IAP, Paris, March-April 2016
CP3, UCL, Louvain-la-Neuve, November 2016
DIAS-TH, BLTP, Dubna, May 2017
Les Houches, July 2018

Éric Gourgoulhon
Laboratoire Univers et Théories
CNRS, Observatoire de Paris, Université de Paris
Université Paris Sciences et Lettres
eric.gourgoulhon@obspm.fr

https://luth.obspm.fr/~luthier/gourgoulhon/bh16

— DRAFT —

version of 16 May 2021

Corrections and comments are welcome

Preface

These notes correspond to lectures given

- at *Institut d'Astrophysique de Paris* (France) in March-April 2016, within the framework of the *IAP Advanced Lectures*:
  http://www.iap.fr/vie_scientifique/cours/cours.php?nom=cours_iap&annee=2016
- at the *Centre for Cosmology, Particle Physics and Phenomenology* in Louvain-la-Neuve (Belgium) in November 2016, within the framework of the *Chaire Georges Lemaître*:
  https://uclouvain.be/fr/instituts-recherche/irmp/chaire-georges-lemaitre-2016.html
- at the *Bogoliubov Laboratory of Theoretical Physics*, in Dubna (Russia) in May 2017, within the framework of the *Dubna International Advanced School of Theoretical Physics*:
  http://www.jinr.ru/posts/lecture-course-geometry-and-physics-of-black-holes/
- at the Summer School *Gravitational Waves 2018*, taking place at Les Houches (France) in July 2018:
  http://www.lkb.upmc.fr/gravitationalwaves2018/

In complement to these notes, one may recommend various monographs devoted to black holes: O'Neill (1995) [200], Poisson (2004) [214], Frolov & Novikov (1998) [110], Frolov & Zelnikov (2011) [111], Bambi (2017) [14] and Chruściel (2020) [57], as well as the review articles by Carter (1987) [47], Chruściel (2002, 2005) [55, 56], Chruściel, Lopes Costa & Heusler (2012) [60] and Wald (2001) [257]. In addition, let us point other lecture notes on black holes: Andersson, Bäckdahl & Blue (2016) [10] Compère (2006, 2019) [67, 68], Dafermos and Rodnianski (2008) [75], Deruelle (2009) [84] Hawking (1994) [147, 148], Reall (2020) [217] and Townsend (1997) [249].

The history of black holes in theoretical physics and astrophysics is very rich and fascinating. It is however not discussed here, except in some small historical notes. The interested reader is referred to Nathalie Deruelle's lectures [84], to Kip Thorne's textbook [246] and to Jean Eisenstaedt's articles [99, 101].

The web pages associated to these notes are

https://luth.obspm.fr/~luthier/gourgoulhon/bh16/

They contain supplementary material, such as the SageMath notebooks presented in Appendix D.

I warmly thank Cyril Pitrou for the organization of the Paris lectures, Fabio Maltoni and Christophe Ringeval for the organization of the Louvain-la-Neuve ones, Anastasia Golubtsova and Irina Pirozhenko for the organization of the Dubna ones, as well as Bruce Allen, Marie-Anne Bizouard, Nelson Christensen and Pierre-François Cohadon for the organization of Les Houches ones.

Contents

Chapter 1

General framework

Contents

1.1 Introduction

This chapter presents succinctly the spacetime framework used in these lectures (Sec. 1.2) and recalls useful basic concepts, such as worldlines of particles and observers (Sec. 1.3 and 1.4). In most of these lectures, we shall assume that the theory of gravitation is general relativity; this means that the spacetime metric obeys Einstein equation, which is recalled in Sec. 1.5.

This chapter is by no means an introduction to general relativity. We recommend the textbooks [38, 52, 86, 142, 190, 237, 256] in this respect, as well as [85, 124, 172] for the French-speaking reader. The reader might also find useful to start the reading by Appendix A, which recaps the concepts from differential geometry employed in the main text.

1.2 Spacetime

1.2.1 General settings

In these lectures we consider a n-dimensional ***spacetime***, i.e. a pair $(\mathcal{M}, \boldsymbol{g})$, where $\mathcal{M}$ is a n-dimensional smooth manifold, with $n \geq 2$, and $\boldsymbol{g}$ is a Lorentzian metric on $\mathcal{M}$. In many parts, n will be set to 4 — the standard spacetime dimension — but we shall also consider spacetimes with $n > 4$, especially in Chap. 19.

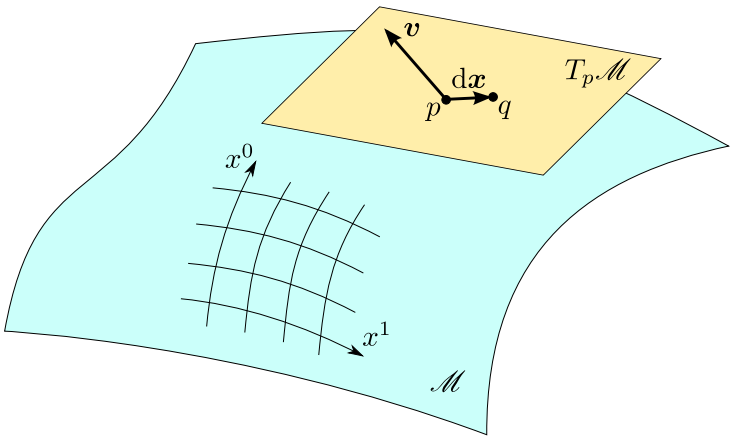


Figure 1.1: A smooth manifold $\mathscr{M}$: the infinitesimal vector $\mathrm{d}\boldsymbol{x}$ connects the nearby points p and q and thus can thought as a displacement within the manifold, while the finite vector $\boldsymbol{v}$ does not correspond to any displacement in the manifold and "lives" in the tangent space $T_p\mathscr{M}$.

The precise definition and basic properties of a *smooth manifold* are recalled in Appendix A. Here let us simply say that, in loose terms, a ***manifold*** $\mathscr{M}$ of dimension n is a "space" that *locally* resembles $\mathbb{R}^n$, i.e. can be described by a n-tuple of coordinates $(x^1, \dots, x^n)$. However, globally, $\mathscr{M}$ can be very different from $\mathbb{R}^n$, in particular regarding its topology.

The smooth structure endows the manifold with the concept of ***infinitesimal displacement vectors*** $\mathrm{d}\boldsymbol{x}$, which connect infinitely close points of $\mathscr{M}$ (cf. Fig. 1.1 and Sec. A.2.3 of Appendix A). However, for finitely separated points, there is no longer the concept of connecting vector (contrary for instance to points in $\mathbb{R}^n$). In other words, vectors on $\mathscr{M}$ do not live in the manifold but in the ***tangent spaces*** $T_p\mathscr{M}$, which are defined at each point $p \in \mathscr{M}$. Each $T_p\mathscr{M}$ is a n-dimensional vector space, which is generated for instance by the infinitesimal displacement vectors along the n coordinate lines of some coordinate system. Unless explicitly specified, we assume that $\mathscr{M}$ is an orientable manifold (cf. Sec. A.3.4).

The full definition of the ***metric tensor*** $\boldsymbol{g}$ is given in Sec. A.3 of Appendix A. At each point $p \in \mathscr{M}$, $\boldsymbol{g}$ induces a (non positive definite) scalar product on $T_p\mathscr{M}$, which we shall denote by a dot:

$$\forall (\boldsymbol{u}, \boldsymbol{v}) \in T_p\mathscr{M} \times T_p\mathscr{M}, \quad \boldsymbol{u} \cdot \boldsymbol{v} := \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}). \tag{1.1}$$

The fact that its signature is Lorentzian, i.e.

$$\operatorname{sign} \boldsymbol{g} = (-, \underbrace{+, \dots, +}_{n-1 \text{ times}}), \tag{1.2}$$

implies that from each point $p \in \mathscr{M}$, there are privileged directions, which form the so-called ***null cones*** or ***light cones*** (cf. Fig. 1.2). The null cones constitute an absolute structure of spacetime, independent from any observer. A vector at a point $p \in \mathscr{M}$ that is either timelike or null is said to be ***causal***. It lies necessarily inside the null cone at p (timelike vector) or along it (null vector).

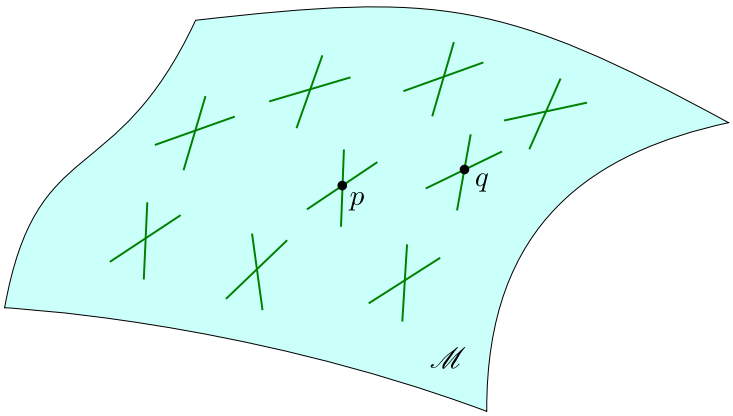


Figure 1.2: A Lorentzian manifold $(\mathscr{M}, \boldsymbol{g})$: at each point, the metric tensor $\boldsymbol{g}$ defines privileged directions: those lying in the null cone at p.

1.2.2 Time orientation

When dealing with black hole spacetimes, it is very important to have clear concepts of "past" and "future". Therefore, we assume that the spacetime $(\mathscr{M}, \boldsymbol{g})$ is *time-orientable*, i.e. that it is possible to divide *continuously* all causal vectors (i.e. timelike or null) into two classes, the *future-directed* ones and the *past-directed* ones. More precisely, at each tangent space $T_p\mathscr{M}$, we may split the causal vectors in two classes by declaring that two causal vectors belong to the same class iff they are located inside or onto the same sheet of the null cone at p. This defines an equivalence relation on causal vectors at p, with two equivalence classes. The spacetime $(\mathscr{M}, \boldsymbol{g})$ is then called ***time-orientable*** iff some choice of an equivalence class can be performed continuously over the entire manifold $\mathscr{M}$. The vectors belonging to the chosen equivalence class are called ***future-directed*** and the other ones ***past-directed***.

As a characterization of future-oriented causal vectors, we shall use quite often the following lemmas:

Lemma 1 (scalar product of a timelike vector with a causal vector)

Let $(\mathscr{M}, \boldsymbol{g})$ be a time-orientable spacetime and $\boldsymbol{u}$ a future-directed timelike vector. For any null or timelike vector $\boldsymbol{v}$, we have

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) < 0 \iff \boldsymbol{v} \text{ is future-directed} \tag{1.3a}$$

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) > 0 \iff \boldsymbol{v} \text{ is past-directed.} \tag{1.3b}$$

In particular, one cannot have $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = 0$.

Proof. Without any loss of generality, we may assume that $\boldsymbol{u}$ is a unit vector: $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = -1$. Let then $(\boldsymbol{e}_i)_{1 \leq i \leq n-1}$ be a family of $n-1$ unit spacelike vectors such that $(\boldsymbol{u}, \boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n-1})$ is an orthonormal basis of $T_p\mathscr{M}$. We may expend $\boldsymbol{v}$ on this basis: $\boldsymbol{v} = v^0\boldsymbol{u} + v^i\boldsymbol{e}_i$. We have necessarily $v^0 \neq 0$, otherwise $\boldsymbol{v} = v^i\boldsymbol{e}_i$ would be a spacelike vector, which is excluded by hypothesis. Moreover, the time-orientation of $\boldsymbol{v}$ is the same as that of $\boldsymbol{u}$ iff $v^0 > 0$. Since $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = -v^0$, this establishes (1.3a) and (1.3b). □

Lemma 2 (scalar product of a null vector with a causal vector)

Let $(\mathscr{M}, \boldsymbol{g})$ be a time-orientable spacetime and $\boldsymbol{u}$ a future-directed null vector. For any null or timelike vector $\boldsymbol{v}$, we have

$$\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) < 0 \iff \boldsymbol{v} \text{ is not collinear with } \boldsymbol{u} \text{ and is future-directed} \tag{1.4a}$$
$$\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) = 0 \iff \boldsymbol{v} \text{ is collinear with } \boldsymbol{u} \text{ (and thus null)} \tag{1.4b}$$
$$\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) > 0 \iff \boldsymbol{v} \text{ is not collinear with } \boldsymbol{u} \text{ and is past-directed.} \tag{1.4c}$$

Proof. Without any loss of generality, we may find an orthonormal basis $(\boldsymbol{e}_\alpha)_{0\leq\alpha\leq n-1}$ of $T_p\mathscr{M}$ such that $\boldsymbol{u} = \boldsymbol{e}_0 + \boldsymbol{e}_1$, where the timelike unit vector $\boldsymbol{e}_0$ is future-directed since $\boldsymbol{u}$ is. Let us expand $\boldsymbol{v}$ on this basis: $\boldsymbol{v} = v^0\boldsymbol{e}_0 + v^i\boldsymbol{e}_i$, with $v^0 \neq 0$ since $\boldsymbol{v}$ is not spacelike. We have then $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) = -v^0 + v^1$. Now, since $\boldsymbol{v}$ is null or timelike, $\boldsymbol{g}(\boldsymbol{v},\boldsymbol{v}) \leq 0$, which is equivalent to

$$(v^0)^2 \geq \sum_{i=1}^{n-1} (v^i)^2. \tag{1.5}$$

This implies $|v^0| \geq |v^1|$. If $|v^0| > |v^1|$, then $\boldsymbol{v}$ cannot be collinear with $\boldsymbol{u}$ (since this would imply $v^0 = v^1$) and $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) = -v^0 + v^1 \neq 0$, with a sign identical to that of $-v^0$. If $|v^0| = |v^1|$, Eq. (1.5) implies $v^2 = v^3 = \cdots = v^{n-1} = 0$. We have then either $\boldsymbol{v} = v^0(\boldsymbol{e}_0 + \boldsymbol{e}_1) = v^0\boldsymbol{u}$ or $\boldsymbol{v} = v^0(\boldsymbol{e}_0 - \boldsymbol{e}_1)$. In the first case, $\boldsymbol{v}$ is collinear with $\boldsymbol{u}$ and $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) = 0$. In the second case, $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) = -2v^0 \neq 0$. To summarize, the only case where $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v}) = 0$ is when $\boldsymbol{v}$ is collinear with $\boldsymbol{u}$. This establishes (1.4b). In all the other cases, $\boldsymbol{v}$ is not collinear with $\boldsymbol{u}$ and the sign of $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{v})$ is that of $-v^0$. Since $\boldsymbol{v}$ is future-directed if $v^0 > 0$ and past-directed if $v^0 < 0$, this establishes (1.4a) and (1.4c). □

Two useful properties are immediate consequences of the above lemmas. From Lemma 1, we get

Corollary 1

A timelike vector can be neither orthogonal to a timelike vector nor orthogonal to a null vector.

From the part (1.4b) of Lemma 2, we get

Corollary 2

Two null vectors are orthogonal if, and only if, they are collinear.

1.3 Worldlines

1.3.1 Definitions

In relativity, a particle is described by its spacetime extent, which is a smooth curve, $\mathscr{L}$ say, and not a point. This curve is called the particle's ***worldline*** and might be thought

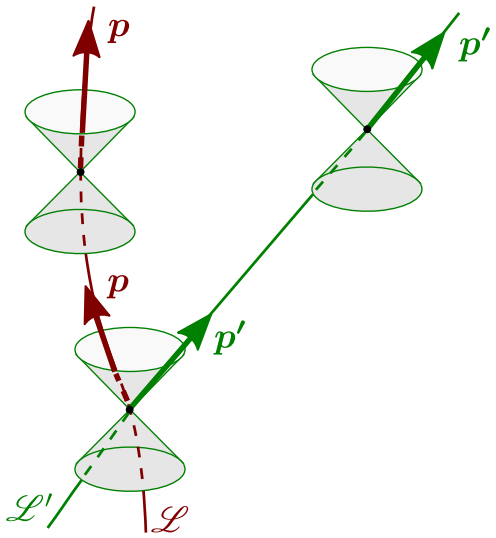


Figure 1.3: Worldlines of a massive particle ($\mathscr{S}$) and of a massless one ($\mathscr{S}'$).

of as the set of the "successive positions" occupied by the particle as "time evolves". Except for pathological cases (tachyons), the worldline has to be a ***causal curve***, i.e. at any point, a tangent vector to $\mathscr{S}$ is either timelike or null. This reflects the impossibility for the particle to travel faster than light with respect to any local inertial frame. The dynamics of a simple particle (i.e. a particle without any internal structure nor spin) is entirely described by its ***4-momentum*** or ***energy-momentum vector***[1], which is a vector field $\boldsymbol{p}$ defined along $\mathscr{S}$, tangent to $\mathscr{S}$ at each point and future-directed (cf. Fig. 1.3).

One distinguishes two types of particles:

- the ***massive particles***, for which $\mathscr{S}$ is a timelike curve, or equivalently, for which $\boldsymbol{p}$ is a timelike vector:
$$\boldsymbol{g}(\boldsymbol{p},\boldsymbol{p}) = \boldsymbol{p}\cdot\boldsymbol{p} < 0; \tag{1.6}$$
- the ***massless particles***, such as the photon, for which $\mathscr{S}$ is a null curve, or equivalently, for which $\boldsymbol{p}$ is a null vector:
$$\boldsymbol{g}(\boldsymbol{p},\boldsymbol{p}) = \boldsymbol{p}\cdot\boldsymbol{p} = 0. \tag{1.7}$$

In both cases, the ***mass*** of the particle is defined by[2]
$$m = \sqrt{-\boldsymbol{p}\cdot\boldsymbol{p}}. \tag{1.8}$$

Of course, for a massless particle, we get $m = 0$.

[1] When $n \neq 4$, *energy-momentum vector* is definitely a better name than *4-momentum*!
[2] Unless specified, we use geometrized units, for which $G = 1$ and $c = 1$.

1.3.2 Geodesic motion

If the particle feels only gravitation, i.e. if no non-gravitational force is exerted on it, the energy-momentum vector must be a ***geodesic vector***, i.e. it obeys

$$\boxed{\boldsymbol{\nabla}_{\boldsymbol{p}}\, \boldsymbol{p} = 0}, \tag{1.9}$$

or, in index notation,

$$p^\mu \nabla_\mu p^\alpha = 0. \tag{1.10}$$

This implies that the worldline $\mathscr{L}$ must be a ***geodesic*** of the spacetime $(\mathscr{M}, \boldsymbol{g})$ (cf. Appendix B).

Remark 1: The reverse is not true, i.e. having $\mathscr{L}$ geodesic and $\boldsymbol{p}$ tangent to $\mathscr{L}$ does not imply (1.9), but the weaker condition $\boldsymbol{\nabla}_{\boldsymbol{p}}\, \boldsymbol{p} = \alpha\, \boldsymbol{p}$, with α a scalar field along $\mathscr{L}$. In this case, one says that $\boldsymbol{p}$ is a ***pregeodesic vector*** (cf. Sec. B.2.2 in Appendix B).

For massive particles, Eq. (1.9) can be derived from a variational principle, the action being simply the worldline's length as given by the metric tensor:

$$S = \int_A^B \mathrm{d}s = \int_{\lambda_A}^{\lambda_B} \sqrt{-\boldsymbol{g}\left(\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}, \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}\right)}\, \mathrm{d}\lambda \tag{1.11}$$

(cf. Sec. B.4.2 for details). For photons, Eq. (1.9) can be derived from Maxwell equations within the geometrical optics approximation (see e.g. Box 5.6 of Ref. [216]), with the assumption that the photon energy-momentum vector is related to the wave 4-vector $\boldsymbol{k}$ by

$$\boldsymbol{p} = \hbar \boldsymbol{k}. \tag{1.12}$$

1.3.3 Massive particles

For a massive particle, the constraint of having the worldline $\mathscr{L}$ timelike has a simple geometrical meaning: $\mathscr{L}$ must always lie inside the light cones of events along $\mathscr{L}$ (cf. Fig. 1.3). The fundamental link between physics and geometry is that the ***proper time*** τ of the particle is nothing but the metric length along the worldline, increasing towards the future:

$$\mathrm{d}\tau = \sqrt{-\boldsymbol{g}(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{x})} = \sqrt{-g_{\mu\nu} \mathrm{d}x^\mu\, \mathrm{d}x^\nu}, \tag{1.13}$$

where $\mathrm{d}\boldsymbol{x}$ is an infinitesimal future-directed[3] displacement along $\mathscr{L}$.

The particle's ***4-velocity*** is defined as the derivative vector $\boldsymbol{u}$ of the parametrization of $\mathscr{L}$ by the proper time:

$$\boxed{\boldsymbol{u} := \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\tau}}. \tag{1.14}$$

By construction, $\boldsymbol{u}$ is tangent to $\mathscr{L}$ and is a unit timelike vector:

$$\boldsymbol{u} \cdot \boldsymbol{u} = -1. \tag{1.15}$$

[3] Cf. Sec. 1.2.2.

For a simple particle (no internal structure), the 4-momentum $\boldsymbol{p}$ is tangent to $\mathscr{L}$; it is then necessarily collinear to $\boldsymbol{u}$. Since both vectors are future-directed, Eqs. (1.8) and (1.15) lead to

$$\boxed{\boldsymbol{p} = m\,\boldsymbol{u}}. \tag{1.16}$$

1.3.4 Massless particles (photons)

For a massless particle, Eq. (1.13) would lead to $\mathrm{d}\tau = 0$ since the displacement $\mathrm{d}\boldsymbol{x}$ would be a null vector. There is then no natural parameter along a null geodesic. However, one can single out a whole family of them, called *affine parameters*. As recalled in Appendix B, an ***affine parameter*** along a null geodesic $\mathscr{L}$ is a parameter λ such that the associated tangent vector,

$$\boldsymbol{v} := \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}, \tag{1.17}$$

is a geodesic vector field: $\boldsymbol{\nabla}_{\boldsymbol{v}}\,\boldsymbol{v} = 0$. In general, the tangent vector associated to a given parameter fulfils only $\boldsymbol{\nabla}_{\boldsymbol{v}}\,\boldsymbol{v} = \alpha\,\boldsymbol{v}$, with α a scalar field along $\mathscr{L}$ (cf. Remark 1 above).

The qualifier *affine* arises from the fact any two affine parameters λ and λ' are necessarily related by an affine transformation:

$$\lambda' = a\lambda + b, \tag{1.18}$$

with a and b two constants. Given that the photon energy-momentum vector $\boldsymbol{p}$ is a geodesic vector [Eq. (1.9)], a natural choice of the affine parameter λ is that associated with $\boldsymbol{p}$:

$$\boldsymbol{p} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}. \tag{1.19}$$

This sets $a = 1$ in the transformation (1.18).

1.4 Quantities measured by an observer

In the simplest modelization, an ***observer*** $\mathscr{O}$ in the spacetime $(\mathscr{M}, \boldsymbol{g})$ is described by a timelike worldline $\mathscr{L}_{\mathscr{O}}$ that is equipped with an orthonormal basis $(\boldsymbol{e}_\alpha)$ at each point, such that $\boldsymbol{e}_0$ is future-directed and tangent to $\mathscr{L}_{\mathscr{O}}$ and $(\boldsymbol{e}_\alpha)$ varies smoothly along $\mathscr{L}_{\mathscr{O}}$ (see e.g. Sec. 13.6 of Ref. [190] or Chap. 3 of Ref. [123] for an extended discussion). The vector $\boldsymbol{e}_0$ is then the 4-velocity of $\mathscr{O}$ and the vectors $(\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3)$ form an orthonormal basis of the 3-dimensional local rest space of $\mathscr{O}$. $(\boldsymbol{e}_\alpha)$ is called the ***observer's frame***.

Let us suppose that the observer $\mathscr{O}$ encounters a particle at some event A. Geometrically, this means that the worldline $\mathscr{L}$ of the particle intersects $\mathscr{L}_{\mathscr{O}}$ at A. Then, the ***energy*** E and the ***linear momentum*** $\boldsymbol{P}$ of the particle, both measured by $\mathscr{O}$, are given by the orthogonal decomposition of the particle's energy-momentum vector $\boldsymbol{p}$ with respect to $\mathscr{L}_{\mathscr{O}}$ (cf. Fig. 1.4):

$$\boxed{\boldsymbol{p} = E\boldsymbol{u}_{\mathscr{O}} + \boldsymbol{P}}, \quad \text{with} \quad \boldsymbol{u}_{\mathscr{O}} \cdot \boldsymbol{P} = 0, \tag{1.20}$$

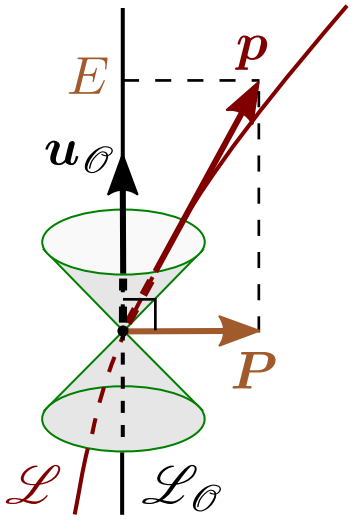


Figure 1.4: Orthogonal decomposition of the energy-momentum vector $\boldsymbol{p}$ of a particle with respect to the 4-velocity $\boldsymbol{u}_{\mathscr{O}}$ of an observer $\mathscr{O}$, giving birth to the energy E and linear momentum $\boldsymbol{P}$ as measured by $\mathscr{O}$.

where $\boldsymbol{u}_{\mathscr{O}} = \boldsymbol{e}_0$ is the 4-velocity of observer $\mathscr{O}$. By taking the scalar product of Eq. (1.20) with $\boldsymbol{u}_{\mathscr{O}}$, we obtain the following expressions for E and $\boldsymbol{P}$:

$$\boxed{E = -\boldsymbol{u}_{\mathscr{O}} \cdot \boldsymbol{p}} \tag{1.21}$$

$$\boxed{\boldsymbol{P} = \boldsymbol{p} + (\boldsymbol{u}_{\mathscr{O}} \cdot \boldsymbol{p})\, \boldsymbol{u}_{\mathscr{O}}}. \tag{1.22}$$

The scalar square of Eq. (1.20) leads to

$$\underbrace{\boldsymbol{p} \cdot \boldsymbol{p}}_{-m^2} = E^2 \underbrace{\boldsymbol{u}_{\mathscr{O}} \cdot \boldsymbol{u}_{\mathscr{O}}}_{-1} + 2E \underbrace{\boldsymbol{u}_{\mathscr{O}} \cdot \boldsymbol{P}}_{0} + \boldsymbol{P} \cdot \boldsymbol{P}, \tag{1.23}$$

where we have used Eq. (1.8) to let appear the particle's mass m. Hence we recover Einstein's relation:

$$\boxed{E^2 = m^2 + \boldsymbol{P} \cdot \boldsymbol{P}}. \tag{1.24}$$

An infinitesimal displacement $\mathrm{d}\boldsymbol{x}$ of the particle along its worldline is related to the energy-momentum vector $\boldsymbol{p}$ by

$$\mathrm{d}\boldsymbol{x} = \boldsymbol{p}\, \mathrm{d}\lambda, \tag{1.25}$$

where λ is the affine parameter along the particle's worldline whose tangent vector is $\boldsymbol{p}$ [cf. Eq. (1.19) for a massless particle and Eqs. (1.14) and (1.16) with $\lambda := \tau/m$ for a massive particle]. Substituting (1.20) for $\boldsymbol{p}$ in (1.25), we get the orthogonal decomposition of $\mathrm{d}\boldsymbol{x}$ with respect to $\mathscr{L}_{\mathscr{O}}$:

$$\mathrm{d}\boldsymbol{x} = E \mathrm{d}\lambda\, \boldsymbol{u}_{\mathscr{O}} + \mathrm{d}\lambda\, \boldsymbol{P}. \tag{1.26}$$

$\mathscr{O}$'s proper time elapsed during the particle displacement is the coefficient in front of $\boldsymbol{u}_{\mathscr{O}}$: $\mathrm{d}\tau_{\mathscr{O}} = E\mathrm{d}\lambda$ and the the particle's displacement in $\mathscr{O}$'s rest frame is the part orthogonal

to $\boldsymbol{u}_{\mathscr{O}}$: $\mathrm{d}\boldsymbol{X} = \mathrm{d}\lambda\, \boldsymbol{P}$. By definition, the particle's velocity with respect to $\mathscr{O}$ is

$$\boldsymbol{V} := \frac{\mathrm{d}\boldsymbol{X}}{\mathrm{d}\tau_{\mathscr{O}}} = \frac{\mathrm{d}\lambda\, \boldsymbol{P}}{E \mathrm{d}\lambda}. \tag{1.27}$$

Hence the relation

$$\boxed{\boldsymbol{P} = E\, \boldsymbol{V}}. \tag{1.28}$$

By combining with (1.20), we get the following orthogonal decomposition of the particle's 4-momentum:

$$\boxed{\boldsymbol{p} = E\, (\boldsymbol{u}_{\mathscr{O}} + \boldsymbol{V})}. \tag{1.29}$$

Relations (1.24), (1.29) and (1.28) are valid for any kind of particle, massive or not. For a massive particle, the energy-momentum vector $\boldsymbol{p}$ is related to the particle's 4-velocity $\boldsymbol{u}$ via (1.16). Inserting this relation into (1.21), we obtain

$$\boxed{E = \Gamma\, m}, \tag{1.30}$$

where

$$\Gamma := -\boldsymbol{u}_{\mathscr{O}} \cdot \boldsymbol{u} \tag{1.31}$$

is the ***Lorentz factor*** of the particle with respect to the observer. If we depart from units with $c = 1$, Eq. (1.30) becomes the famous relation $E = \Gamma m c^2$. Furthermore, Combining (1.28) and (1.30) yields the familiar relation between the linear momentum and the velocity:

$$\boldsymbol{P} = \Gamma m\, \boldsymbol{V}. \tag{1.32}$$

Finally, inserting (1.30) and (1.32) into (1.24) leads to the well-known expression of the Lorentz factor in terms of the velocity:

$$\Gamma = (1 - \boldsymbol{V} \cdot \boldsymbol{V})^{-1/2}. \tag{1.33}$$

If we divide Eq. (1.20) by m and use Eqs. (1.16), (1.30) and (1.32) to express respectively $\boldsymbol{p}/m$, E/m and $\boldsymbol{P}/m$, we get the following orthogonal split of the particle's 4-velocity $\boldsymbol{u}$ with respect to observer $\mathscr{O}$:

$$\boxed{\boldsymbol{u} = \Gamma\, (\boldsymbol{u}_{\mathscr{O}} + \boldsymbol{V})}. \tag{1.34}$$

For a massless particle (photon), inserting (1.28) into the Einstein relation (1.24) with $m = 0$ yields

$$\boldsymbol{V} \cdot \boldsymbol{V} = 1. \tag{1.35}$$

This means that the norm of the velocity of the massless particle with respect to $\mathscr{O}$ equals the speed of light c ($= 1$ in our units). For a photon associated with a monochromatic radiation, the wave 4-vector $\boldsymbol{k}$ admits the following orthogonal decomposition:

$$\boxed{\boldsymbol{k} = \omega\, (\boldsymbol{u}_{\mathscr{O}} + \boldsymbol{V})}, \tag{1.36}$$

where $\omega = 2\pi\nu$ and ν is the radiation frequency as measured by observer $\mathscr{O}$. In view of Eq. (1.29) with $\boldsymbol{p} = \hbar \boldsymbol{k}$ [Eq. (1.12)], we get the Planck-Einstein relation:

$$\boxed{E = h\nu}. \tag{1.37}$$

1.5 Einstein equation

Saying that gravitation in the spacetime $(\mathscr{M}, \boldsymbol{g})$ is ruled by ***general relativity*** amounts to demanding that the spacetime dimension fulfils $n \geq 3$ and the metric $\boldsymbol{g}$ obeys ***Einstein equation***:

$$\boxed{\boldsymbol{R} - \frac{1}{2} R\, \boldsymbol{g} + \Lambda\, \boldsymbol{g} = 8\pi \boldsymbol{T}}, \tag{1.38}$$

where $\boldsymbol{R}$ is the Ricci tensor of $\boldsymbol{g}$, R is the Ricci scalar of $\boldsymbol{g}$ (cf. Sec. A.5.3 in Appendix A), Λ is some constant, called the ***cosmological constant***, and $\boldsymbol{T}$ is the energy-momentum tensor of matter and non-gravitational fields.

Remark 1: The case $n = 2$ has been excluded since the Einstein equation would no longer involve the spacetime curvature, given that the Einstein tensor $\boldsymbol{R} - R/2\, \boldsymbol{g}$ is identically zero for any metric tensor $\boldsymbol{g}$ if $n = 2$ (consider the trace of Eq. (A.112) in Appendix A). The exclusion of $n = 2$ follows also by noticing that the Einstein-Hilbert action, which gives birth to Eq. (1.38) for $n \geq 3$, reduces (via the Gauss-Bonnet formula) to the Euler characteristic of $\mathscr{M}$ — a topological invariant independent of $\boldsymbol{g}$.

By taking the trace of (1.38) with respect to $\boldsymbol{g}$, it is easy to show that the Einstein equation (1.38) is equivalent to

$$\boxed{\boldsymbol{R} = \frac{2}{n-2} \Lambda\, \boldsymbol{g} + 8\pi \left(\boldsymbol{T} - \frac{1}{n-2} T\, \boldsymbol{g} \right)}, \tag{1.39}$$

where $T := g^{\mu\nu} T_{\mu\nu}$ is the trace of $\boldsymbol{T}$ with respect to $\boldsymbol{g}$.

Remark 2: The spacetime dimension n does not appear in the Einstein equation (1.38); on the contrary, the variant (1.39) depends on n. Notice as well that (1.39) would have no mathematical meaning for $n = 2$ (cf. Remark 1).

Taking the covariant divergence of the Einstein equation (1.38) and invoking the contracted Bianchi identity (A.110) leads to

$$\boxed{\boldsymbol{\nabla} \cdot \overrightarrow{\boldsymbol{T}} = 0}, \tag{1.40}$$

where $\overrightarrow{\boldsymbol{T}}$ in the type-$(1,1)$ tensor associated by metric duality to $\boldsymbol{T}$ [cf. (A.44)]. In index notation, the above equation writes

$$\nabla_\mu T^\mu{}_\alpha = 0.$$

Equation (1.40) is often referred to as the ***equation of energy-momentum conservation***.

Chapter 2

The concept of black hole 1: Horizons as null hypersurfaces

Contents

2.1 Introduction

In this chapter, we shall start from a naive "definition" of a black hole, as a region of spacetime from which no particle can escape, and we shall convince ourselves that the black hole boundary — the so-called *event horizon* — must be a null hypersurface (Sec. 2.2). We shall then study the properties of these hypersurfaces (Secs. 2.3 and 2.4). The precise mathematical definition of a black hole will be given in Chap. 4.

2.2 Black holes and null hypersurfaces

2.2.1 A first definition of black holes

Given a n-dimensional spacetime $(\mathcal{M}, \boldsymbol{g})$ as presented in Chap. 1 (with $n \geq 2$), a naive definition of a black hole, involving only words, could be

> A ***black hole*** is a localized region of spacetime from which neither massive particles nor massless ones (photons) can escape.

There are essentially two features in this definition: *localization* and *inescapability*. Let us for a moment focus on the latter. It implies the existence of a *boundary*, which no particle

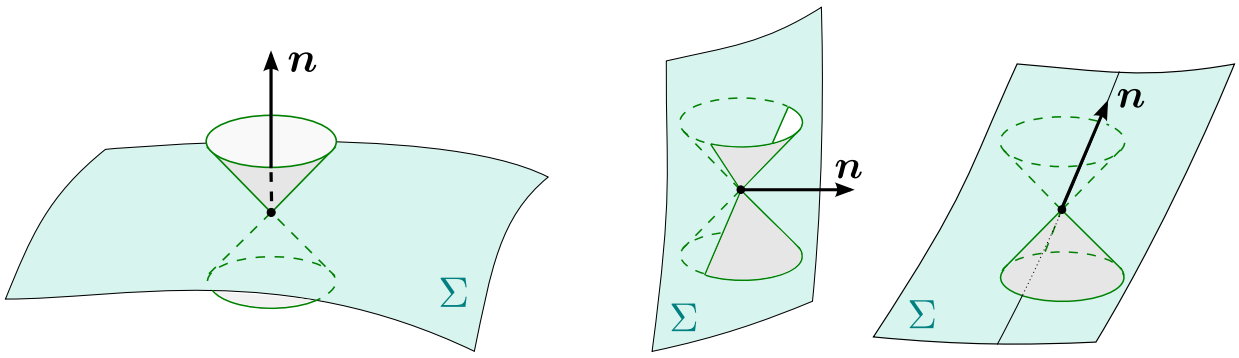


Figure 2.1: The three types of hypersurfaces: spacelike (left), timelike (middle) and null (right).

emitted in the black hole region can cross. This boundary is called the ***event horizon*** and is quite often referred to simply as the ***horizon***. It is a ***one-way membrane***, in the sense that it can be crossed from the black hole "exterior" towards the black hole "interior", but not in the reverse way. The one-way membrane must be a hypersurface of the spacetime manifold $\mathscr{M}$, for it has to divide $\mathscr{M}$ in two regions: the interior (the black hole itself) and the exterior region. Let us recall that a ***hypersurface*** is an embedded submanifold of $\mathscr{M}$ of codimension 1 (cf. Sec. A.2.7 in Appendix A).

2.2.2 The event horizon as a null hypersurface

To discuss further which hypersurface could act as a black hole boundary, one should recall that, on a Lorentzian manifold $(\mathscr{M}, \boldsymbol{g})$, a hypersurface Σ can locally be classified in three categories. The classification depends on the type of metric induced by $\boldsymbol{g}$ on Σ, the ***induced metric*** being nothing but the restriction $\boldsymbol{g}|_\Sigma$ of $\boldsymbol{g}$ to vector fields tangent to Σ. The hypersurface Σ is said to be

- ***spacelike*** iff $\boldsymbol{g}|_\Sigma$ is positive definite, i.e. iff sign $\boldsymbol{g}|_\Sigma = (+,+,+)$, i.e. iff $(\Sigma, \boldsymbol{g}|_\Sigma)$ is a Riemannian manifold;
- ***timelike*** iff $\boldsymbol{g}|_\Sigma$ is a Lorentzian metric, i.e. iff sign $\boldsymbol{g}|_\Sigma = (-,+,+)$, i.e. iff $(\Sigma, \boldsymbol{g}|_\Sigma)$ is a Lorentzian manifold;
- ***null*** iff $\boldsymbol{g}|_\Sigma$ is degenerate[1] i.e. iff sign $\boldsymbol{g}|_\Sigma = (0,+,+)$.

All these definitions are local, i.e. apply to a point $p \in \Sigma$. Of course, it may happen that a hypersurface has not the same type among all its points.

The hypersurface type can also be deduced from any normal vector[2] $\boldsymbol{n}$ to it (cf. Fig. 2.1):

- Σ spacelike $\iff$ $\boldsymbol{n}$ timelike;

[1] Cf. Sec. A.3.1 in Appendix A for the definition of a degenerate bilinear form; the degeneracy implies that the bilinear form $\boldsymbol{g}|_\Sigma$ is not, strictly speaking, a metric on Σ.

[2] The definition of a vector normal to a hypersurface is recalled in Sec. A.3.5 of Appendix A.

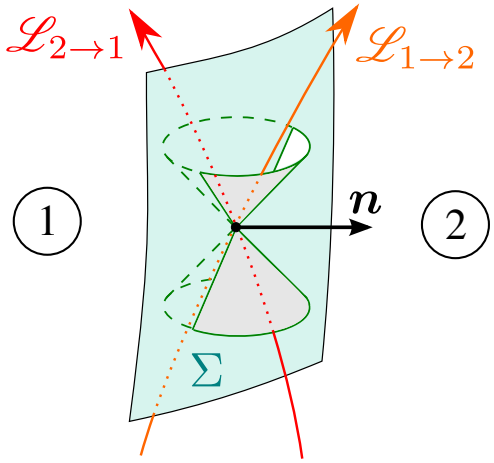


Figure 2.2: A timelike hypersurface is a two-way membrane: $\mathscr{L}_{1\to2}$ is a timelike worldline from Region 1 to Region 2, while $\mathscr{L}_{2\to1}$ is a timelike worldline from Region 2 to Region 1.

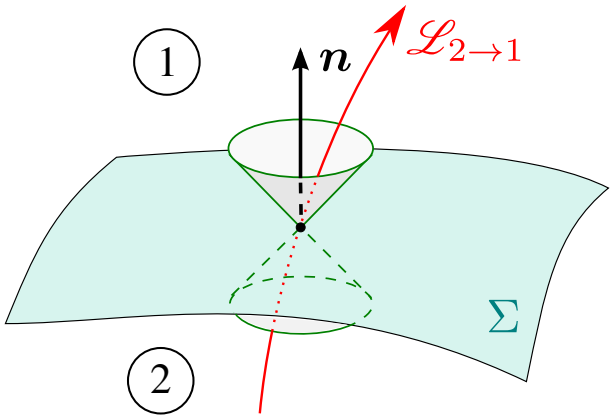


Figure 2.3: A spacelike hypersurface is a one-way membrane: $\mathscr{L}_{2\to1}$ is a timelike worldline from Region 2 to Region 1, while there is no timelike or null worldline from Region 1 to Region 2.

- Σ timelike $\iff$ $\boldsymbol{n}$ spacelike;
- Σ null $\iff$ $\boldsymbol{n}$ null.

These equivalences are easily proved by considering a $\boldsymbol{g}$-orthogonal basis adapted to Σ.

Remark 1: Null hypersurfaces have the distinctive feature that their normals are also tangent to them. Indeed, by definition, the normal $\boldsymbol{n}$ is null iff $\boldsymbol{n}\cdot\boldsymbol{n} = 0$, which is nothing but the condition for $\boldsymbol{n}$ to be tangent to Σ.

A timelike hypersurface is a two-way membrane: if it divides (locally) spacetime in two regions, 1 and 2 say, and a future-directed timelike or null worldline can cross it from Region 1 to Region 2, or from Region 2 to Region 1 (see Fig. 2.2). On the contrary, a spacelike hypersurface is a one-way membrane: a future-directed timelike or null worldline, which is constrained to move inside the light cones, can cross it only from Region 2 to Region 1, say (see Fig. 2.3). A null hypersurface is also a one-way membrane (see Fig. 2.4). At most, a null worldline that is not going from Region 2 to Region 1 must stay on the

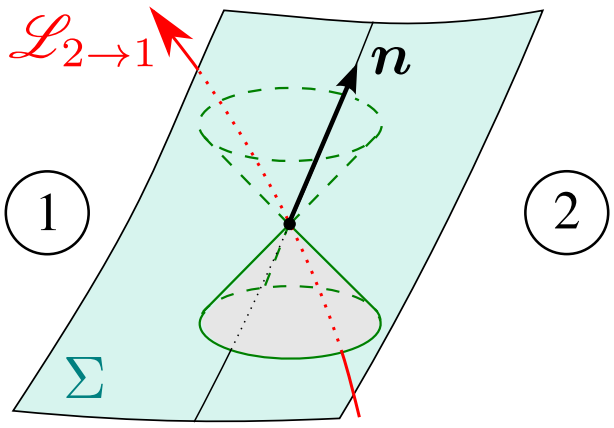


Figure 2.4: A null hypersurface is a one-way membrane: $\mathscr{L}_{2\to1}$ is a timelike worldline from Region 2 to Region 1, while there is no timelike or null worldline from Region 1 to Region 2.

hypersurface; an example of such null worldline is the one depicted in Fig. 2.4 as the thin black line tangent to the normal $\boldsymbol{n}$.

The limit case between two-way membranes (timelike hypersurfaces) and one-way ones being null hypersurfaces, it is quite natural to select the latter ones for the black hole boundary, rather than spacelike hypersurfaces. This choice will be fully justified in Chap. 4, where we shall see that the precise definition of a black hole implies that its boundary (the event horizon) is a null hypersurface as soon as it is smooth (Property 4 in Sec. 4.4.3). Note however that in Chap. 18, we shall see that spacelike hypersurfaces, called *dynamical horizons*, are involved in quasi-local approaches to black holes.

2.3 Geometry of null hypersurfaces

Having decided that the black hole event horizon must be a null hypersurface, let us examine the geometrical properties of such hypersurfaces. We shall denote the hypersurface under study by $\mathscr{H}$, for *horizon*, but the results of this section will be valid for any null hypersurface.

2.3.1 Hypersurfaces as level sets

As any hypersurface, $\mathscr{H}$ can be locally considered as a level set: around any point of $\mathscr{H}$, there exists an open subset $\mathscr{U}$ of $\mathscr{M}$ (possibly $\mathscr{U} = \mathscr{M}$) and a smooth scalar field $u : \mathscr{U} \to \mathbb{R}$ such that

$$\forall p \in \mathscr{U}, \quad p \in \mathscr{H} \iff u(p) = 0. \tag{2.1}$$

and

$$\boldsymbol{\nabla} u \neq 0 \quad \text{on } \mathscr{H}. \tag{2.2}$$

Condition (2.2) ensures that $\mathscr{H}$ is a regular hypersurface (an *embedded* submanifold, in mathematical terms); without it, $\mathscr{H}$ may be self-intersecting.

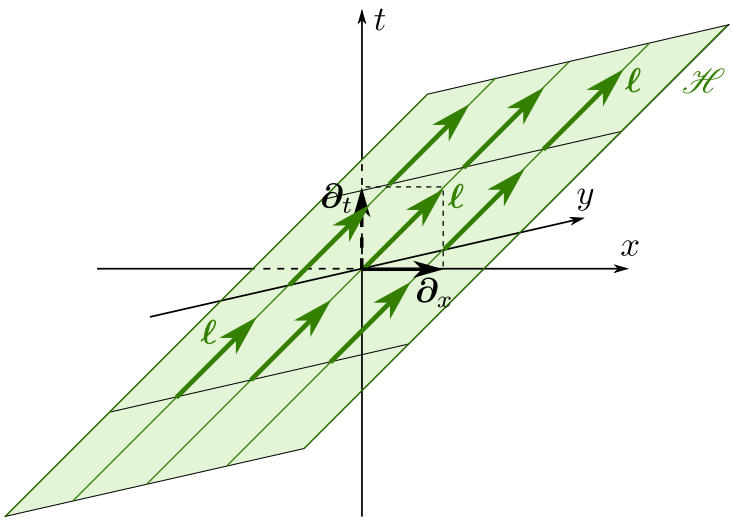


Figure 2.5: Null hyperplane $\mathscr{H}$ of equation $t - x = 0$ in Minkowski spacetime. The dimension along z has been suppressed, so that $\mathscr{H}$ is pictured as a 2-plane.

Example 1 (null hyperplane): A very simple example of null hypersurface is a null hyperplane of the 4-dimensional Minkowski spacetime[3]. If (t, x, y, z) are Minkowskian coordinates, the choice of the scalar field

$$u(t, x, y, z) = t - x \tag{2.3}$$

defines a null hyperplane $\mathscr{H}$ by $u = 0$ (cf. Fig. 2.5).

Example 2 (light cone): Another simple example of null hypersurface, still in the 4-dimensional Minkowski spacetime, is the future sheet $\mathscr{H}$ of a light cone, also called ***future light cone***. Note that we have to take out the cone apex from $\mathscr{H}$, in order to have a regular hypersurface. In the Minkowskian coordinates (t, x, y, z), the choice of the "retarded time"

$$u(t, x, y, z) = t - \sqrt{x^2 + y^2 + z^2} \tag{2.4}$$

defines a future light cone $\mathscr{H}$ by $u = 0$ and $t > 0$ (cf. Fig. 2.6).

Example 3 (Schwarzschild horizon): Let us consider the 4-dimensional spacetime $(\mathscr{M}, \boldsymbol{g})$ with $\mathscr{M}$ diffeomorphic to $\mathbb{R}^4$ and equipped with a coordinate system $(x^\alpha) = (t, r, \theta, \varphi)$ ($t \in \mathbb{R}$, $r \in (0, +\infty)$, $\theta \in (0, \pi)$ and $\varphi \in (0, 2\pi)$) such that $\boldsymbol{g}$ takes the form

$$g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = -\left(1 - \frac{2m}{r}\right) \mathrm{d}t^2 + \frac{4m}{r}\, \mathrm{d}t\, \mathrm{d}r + \left(1 + \frac{2m}{r}\right) \mathrm{d}r^2 + r^2 \mathrm{d}\theta^2 + r^2 \sin^2\theta\, \mathrm{d}\varphi^2, \tag{2.5}$$

where m is a positive constant. We shall see in Chap. 6 that $(\mathscr{M}, \boldsymbol{g})$ is actually a part of Schwarzschild spacetime, described in coordinates different from the standard Schwarzschild-Droste ones, $(\bar{t}, r, \theta, \varphi)$ say, by the choice of the time coordinate: $t = \bar{t} + 2m \ln|r/(2m) - 1|$. The present coordinates are called the ***ingoing Eddington-Finkelstein coordinates*** and have the

[3]The ***Minkowski spacetime*** is defined as $\mathscr{M} = \mathbb{R}^4$ with $\boldsymbol{g}$ a flat Lorentzian metric; ***Minkowskian coordinates*** are then coordinates with respect to which the metric components are $g_{\alpha\beta} = \mathrm{diag}(-1, 1, 1, 1)$.

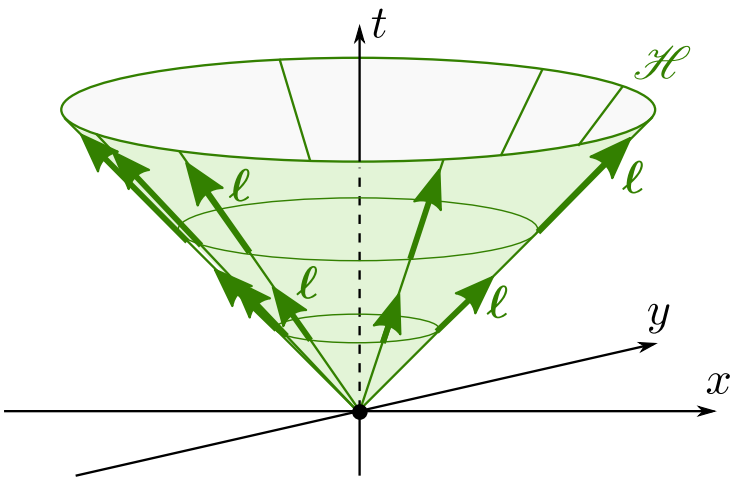


Figure 2.6: Future sheet $\mathscr{H}$ of the light cone of equation $t - \sqrt{x^2+y^2+z^2} = 0$ in Minkowski spacetime. The dimension along z has been suppressed, so that $\mathscr{H}$ looks 2-dimensional, whereas it is actually 3-dimensional.

advantage over the standard ones to be regular on the event horizon, which is located at $r = 2m$. Indeed, the metric components (2.5) remain finite when $r \to 2m$, as those of the inverse metric, which are

$$g^{\alpha\beta} = \begin{pmatrix} -\left(1+\frac{2m}{r}\right) & \frac{2m}{r} & 0 & 0 \\ \frac{2m}{r} & 1-\frac{2m}{r} & 0 & 0 \\ 0 & 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & 0 & \frac{1}{r^2\sin^2\theta} \end{pmatrix}. \tag{2.6}$$

Let us consider the scalar field defined on $\mathscr{M}$ by

$$u(t,r,\theta,\varphi) = \left(1-\frac{r}{2m}\right)\exp\left(\frac{r-t}{4m}\right). \tag{2.7}$$

It is then clear that the hypersurface $u = 0$ is the 3-dimensional "cylinder" $\mathscr{H}$ of equation $r = 2m$ (cf. Fig. 2.7). We shall see below[4] that $\mathscr{H}$ is indeed a null hypersurface.

2.3.2 Null normals

Let $\boldsymbol{\ell}$ be a vector field normal to $\mathscr{H}$. Since $\mathscr{H}$ is a null hypersurface, $\boldsymbol{\ell}$ is a null vector:

$$\boldsymbol{\ell}\cdot\boldsymbol{\ell} = 0. \tag{2.8}$$

Moreover, we choose $\boldsymbol{\ell}$ to be future-directed (cf. Sec. 1.2.2).

Remark 1: As a consequence of (2.8), there is no natural normalization of $\boldsymbol{\ell}$, contrary to the case of timelike or spacelike hypersurfaces, where one can always choose the normal to be a unit vector (scalar square equal to 1 or -1). It follows that there is no unique choice of $\boldsymbol{\ell}$. At this

[4] This should be obvious to the experienced reader, since a normal 1-form to $\mathscr{H}$ is $\mathbf{d}r$ and from Eq. (2.6), $g^{\mu\nu}\partial_\mu r\,\partial_\nu r = g^{rr} = 0$ on $\mathscr{H}$.

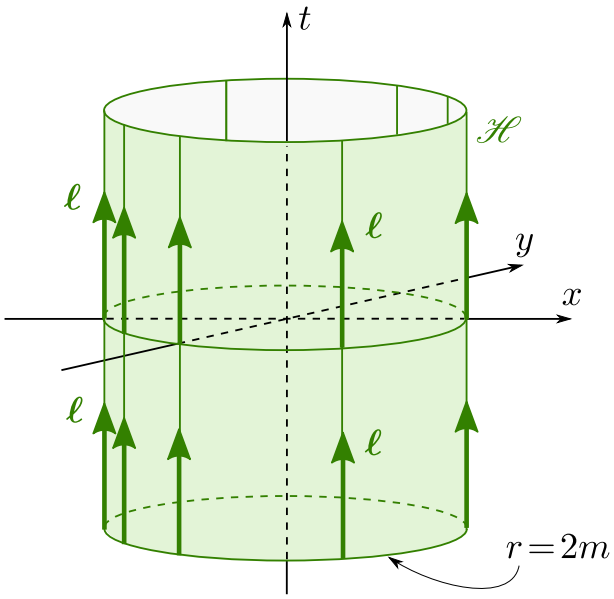


Figure 2.7: Schwarzschild horizon $\mathscr{H}$ introduced in Example 3; The figure is drawn for $\theta = \pi/2$ and is based on coordinates (t, x, y) related to the ingoing Eddington-Finkelstein coordinates (t, r, θ, φ) by $x = r\cos\varphi$ and $y = r\sin\varphi$.

stage, any rescaling $\boldsymbol{\ell} \mapsto \boldsymbol{\ell}' = \alpha\boldsymbol{\ell}$, with α some strictly positive (to preserve the future orientation of $\boldsymbol{\ell}$) scalar field on $\mathscr{H}$, yields a normal vector field $\boldsymbol{\ell}'$ as valid as $\boldsymbol{\ell}$.

The null normal vector field $\boldsymbol{\ell}$ is a priori defined on $\mathscr{H}$ only and not at points $p \notin \mathscr{H}$. However, it is worth to consider $\boldsymbol{\ell}$ as a vector field not confined to $\mathscr{H}$ but defined in some open subset of $\mathscr{M}$ around $\mathscr{H}$. In particular this would permit to define the spacetime covariant derivative $\boldsymbol{\nabla}\boldsymbol{\ell}$, which is not possible if the support of $\boldsymbol{\ell}$ is restricted to $\mathscr{H}$. Following Carter [48], a simple way to achieve this is to consider not only a single null hypersurface $\mathscr{H}$, but a foliation of $\mathscr{M}$ (in the vicinity of $\mathscr{H}$) by a family of null hypersurfaces, such that $\mathscr{H}$ is an element of this family. Without any loss of generality, we may select the value of the scalar field u defining $\mathscr{H}$ to label these hypersurfaces and denote the family by $(\mathscr{H}_u)$. The null hypersurface $\mathscr{H}$ is then nothing but the element $\mathscr{H} = \mathscr{H}_{u=0}$ of this family [Eq. (2.1)]. The vector field $\boldsymbol{\ell}$ can then be viewed as being defined in the part of $\mathscr{M}$ foliated by $(\mathscr{H}_u)$, such that at each point in this region, $\boldsymbol{\ell}$ is null and normal to $\mathscr{H}_u$ for some value of u.

Example 4: The scalar field u introduced in Example 1 (null hyperplane) does define a family of null hypersurfaces $(\mathscr{H}_u)$. A counter-example would be $u(t, x, y, z) = (t - x)(1 + x^2)$, since $u = a$ does not define a null hypersurface except for $a = 0$. Similarly, the scalar fields u of Example 2 (light cone) and Example 3 (Schwarzschild horizon) do define a family of null hypersurfaces $(\mathscr{H}_u)$. In the latter example, this would not have been the case for the simpler choice $u(t, r, \theta, \varphi) = r - 2m$. Some of these null hypersurfaces are represented in Fig. 2.8

Obviously the family $(\mathscr{H}_u)$ is non-unique but all geometrical quantities that we shall introduce hereafter do not depend upon the choice of the foliation $\mathscr{H}_u$ once they are evaluated at $\mathscr{H}$.

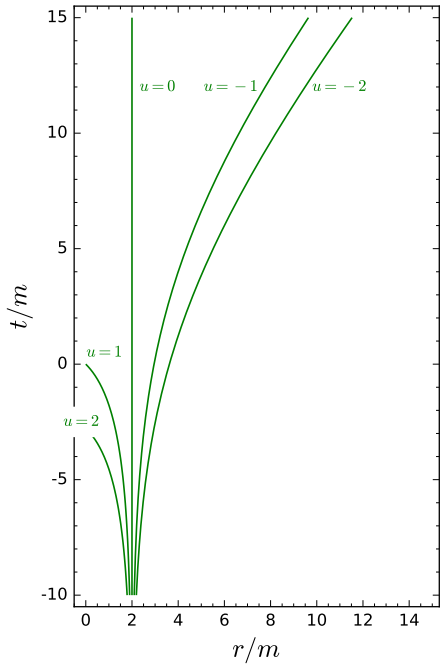


Figure 2.8: Hypersurfaces $\mathscr{H}_u$ defined by $u = \text{const}$ for the example of the Schwarzschild horizon (Example 3).

Since $\mathscr{H}$ is a hypersurface where u is constant [Eq. (2.1)], we have, by definition,

$$\begin{aligned} \forall \boldsymbol{v} \in T_p\mathscr{M}, \quad \boldsymbol{v} \text{ tangent to } \mathscr{H} &\iff \boldsymbol{\nabla}_{\boldsymbol{v}} u = 0 \\ &\iff \langle \boldsymbol{\nabla} u, \boldsymbol{v} \rangle = 0 \\ &\iff \vec{\boldsymbol{\nabla}} u \cdot \boldsymbol{v} = 0, \end{aligned} \tag{2.9}$$

where $\vec{\boldsymbol{\nabla}} u$ is the gradient vector field of the scalar field u, i.e. the vector field given in index-notation by (cf. Sec. A.3.3)

$$\nabla^\alpha u = g^{\alpha\mu} \nabla_\mu u = g^{\alpha\mu} \frac{\partial u}{\partial x^\mu}. \tag{2.10}$$

Property (2.9) means that $\vec{\boldsymbol{\nabla}} u$ is a normal vector field to $\mathscr{H}$. By uniqueness of the normal direction to a hypersurface, it must then be collinear to $\boldsymbol{\ell}$. Therefore, there must exist some scalar field ρ such that

$$\boxed{\boldsymbol{\ell} = -\mathrm{e}^{\rho}\, \vec{\boldsymbol{\nabla}} u}. \tag{2.11}$$

We have chosen the coefficient linking $\boldsymbol{\ell}$ and $\vec{\boldsymbol{\nabla}} u$ to be strictly negative, i.e. under the form of minus an exponential. This is always possible by a suitable choice of the scalar field u. The minus sign ensures that in the case of u increasing toward the future, $\boldsymbol{\ell}$ is future-directed, as the following example shows:

Example 5 (null hyperplane): We deduce from the expression (2.3) chosen for u in Example 1 that

$$\boldsymbol{\nabla} u = \mathbf{d}t - \mathbf{d}x.$$

The gradient vector field obtained by metric duality is $\overrightarrow{\boldsymbol{\nabla}} u = -\boldsymbol{\partial}_t - \boldsymbol{\partial}_x$. Choosing for simplicity $\rho = 0$, we get from formula (2.11)

$$\boldsymbol{\ell} = \boldsymbol{\partial}_t + \boldsymbol{\partial}_x. \tag{2.12}$$

The vector field $\boldsymbol{\ell}$ is depicted in Fig. 2.5.

Example 6 (light cone): Regarding Example 2, we have, given expression (2.4) for u,

$$\boldsymbol{\nabla} u = \mathbf{d}t - \frac{x}{r}\mathbf{d}x - \frac{y}{r}\mathbf{d}y - \frac{z}{r}\mathbf{d}z, \quad \text{with} \quad r := \sqrt{x^2+y^2+z^2}.$$

Choosing for simplicity $\rho = 0$ in (2.11), we get the normal

$$\boldsymbol{\ell} = \boldsymbol{\partial}_t + \frac{x}{r}\boldsymbol{\partial}_x + \frac{y}{r}\boldsymbol{\partial}_y + \frac{z}{r}\boldsymbol{\partial}_z. \tag{2.13}$$

The vector field $\boldsymbol{\ell}$ is depicted in Fig. 2.6.

Example 7 (Schwarzschild horizon): We deduce from the expression (2.7) chosen for u in Example 3 that

$$\boldsymbol{\nabla} u = \frac{1}{4m}\mathrm{e}^{(r-t)/(4m)}\left[-\left(1-\frac{r}{2m}\right)\mathbf{d}t - \left(1+\frac{r}{2m}\right)\mathbf{d}r\right].$$

The corresponding gradient vector field, is computed from (2.10) via expression (2.6) for $g^{\alpha\mu}$:

$$\overrightarrow{\boldsymbol{\nabla}} u = \frac{1}{4m}\mathrm{e}^{(r-t)/(4m)}\left[-\left(1+\frac{r}{2m}\right)\boldsymbol{\partial}_t + \left(1-\frac{r}{2m}\right)\boldsymbol{\partial}_r\right].$$

This time, we do not chose $\rho = 0$ but rather select ρ so that $\ell^t = 1$:

$$\mathrm{e}^{\rho} = -\frac{1}{\nabla^t u} \iff \rho = \frac{t-r}{4m} - \ln\left(1+\frac{r}{2m}\right) + \ln(4m). \tag{2.14}$$

Equation (2.11) leads then to

$$\boldsymbol{\ell} = \boldsymbol{\partial}_t + \frac{r-2m}{r+2m}\boldsymbol{\partial}_r. \tag{2.15}$$

Given the metric (2.5), we check that $\boldsymbol{g}(\boldsymbol{\ell},\boldsymbol{\ell}) = 0$. Since $\boldsymbol{\ell} \neq 0$, this proves that all hypersurfaces $\mathscr{H}_u$, and in particular $\mathscr{H}$, are null. The vector field $\boldsymbol{\ell}$ is depicted on $\mathscr{H}$ in Fig. 2.7 and in all space in Fig. 2.11.

2.3.3 Null geodesic generators

Frobenius identity

Let us take the metric dual of relation (2.11): it writes $\underline{\boldsymbol{\ell}} = -\mathrm{e}^{\rho}\,\boldsymbol{\nabla} u$, or, in index notation,

$$\ell_\alpha = -\mathrm{e}^{\rho}\,\nabla_\alpha u. \tag{2.16}$$

Taking the covariant derivative, we get

$$\nabla_\alpha \ell_\beta = -\mathrm{e}^{\rho}\nabla_\alpha\rho\nabla_\beta u - \mathrm{e}^{\rho}\nabla_\alpha\nabla_\beta u = \nabla_\alpha\rho\,\ell_\beta - \mathrm{e}^{\rho}\nabla_\alpha\nabla_\beta u$$

Antisymmetrizing and using the torsion-free property of $\boldsymbol{\nabla}$ (i.e. $\nabla_\alpha\nabla_\beta u - \nabla_\beta\nabla_\alpha u = 0$, cf. Eq. (A.68) in Appendix A), we get

$$\nabla_\alpha \ell_\beta - \nabla_\beta \ell_\alpha = \nabla_\alpha \rho\, \ell_\beta - \nabla_\beta \rho\, \ell_\alpha. \tag{2.17}$$

In the left-hand side there appears the exterior derivative of the 1-form $\underline{\boldsymbol{\ell}}$ (cf. Sec. A.4.3 in Appendix A), while one recognize in the right-hand side the exterior product of the two 1-forms $\mathbf{d}\rho$ and $\underline{\boldsymbol{\ell}}$. Hence we may rewrite (2.17) as

$$\boxed{\mathbf{d}\underline{\boldsymbol{\ell}} = \mathbf{d}\rho \wedge \underline{\boldsymbol{\ell}}}. \tag{2.18}$$

This reflects the ***Frobenius theorem*** in its dual formulation (see e.g. Theorem B.3.2 in Wald's textbook [256] or Theorem C.2 in Straumann's textbook [237]): the exterior derivative of the 1-form $\underline{\boldsymbol{\ell}}$ is the exterior product of $\underline{\boldsymbol{\ell}}$ itself with some 1-form ($\mathbf{d}\rho$ in the present case) if, and only if, $\underline{\boldsymbol{\ell}}$ defines hyperplanes that are integrable in some hypersurface ($\mathscr{H}$ in the present case).

Geodesic generators

Let us contract the Frobenius identity (2.17) with $\boldsymbol{\ell}$:

$$\ell^\mu \nabla_\mu \ell_\alpha - \ell^\mu \nabla_\alpha \ell_\mu = \ell^\mu \nabla_\mu \rho\, \ell_\alpha - \underbrace{\ell^\mu \ell_\mu}_{0} \nabla_\alpha \rho. \tag{2.19}$$

Now, since $\boldsymbol{\ell}$ is a null vector,

$$\ell^\mu \nabla_\alpha \ell_\mu = \nabla_\alpha \underbrace{(\ell^\mu \ell_\mu)}_{0} - \ell_\mu \nabla_\alpha \ell^\mu,$$

from which we get

$$\ell^\mu \nabla_\alpha \ell_\mu = 0. \tag{2.20}$$

Hence (2.19) reduces to

$$\ell^\mu \nabla_\mu \ell_\alpha = \kappa\, \ell_\alpha, \tag{2.21}$$

with

$$\kappa := \ell^\mu \nabla_\mu \rho = \boldsymbol{\nabla}_{\boldsymbol{\ell}} \rho. \tag{2.22}$$

The metric dual of (2.21) is

$$\boxed{\boldsymbol{\nabla}_{\boldsymbol{\ell}}\, \boldsymbol{\ell} = \kappa\, \boldsymbol{\ell}}. \tag{2.23}$$

This equation implies that the field lines of $\boldsymbol{\ell}$ are geodesics (cf. Appendix B). To demonstrate this, we note that a rescaling

$$\boldsymbol{\ell} \mapsto \boldsymbol{\ell}' = \alpha \boldsymbol{\ell} \tag{2.24}$$

with α a positive scalar field can be performed to yield a ***geodesic vector field*** $\boldsymbol{\ell}'$, i.e. a vector field that obeys[5] Eq. (B.1):

$$\boldsymbol{\nabla}_{\boldsymbol{\ell}'}\, \boldsymbol{\ell}' = 0. \tag{2.25}$$

[5] A vector field that obeys the weaker condition (2.23), with κ possibly different from zero, is called a ***pregeodesic vector field***, cf. Sec. B.2.2 in Appendix B.

Proof. Equations (2.24) and (2.23) imply

$$\boldsymbol{\nabla}_{\boldsymbol{\ell}'} \boldsymbol{\ell}' = \alpha \left(\boldsymbol{\nabla}_{\boldsymbol{\ell}} \alpha + \kappa \alpha \right) \boldsymbol{\ell}. \tag{2.26}$$

Hence, since $\alpha > 0$,

$$\boldsymbol{\nabla}_{\boldsymbol{\ell}'} \boldsymbol{\ell}' = 0 \iff \boldsymbol{\nabla}_{\boldsymbol{\ell}} \ln \alpha = -\kappa.$$

Therefore, it suffices to solve $\boldsymbol{\nabla}_{\boldsymbol{\ell}} \ln \alpha = -\kappa$, which is a first-order ordinary differential equation along each field line of $\boldsymbol{\ell}$, to ensure that $\boldsymbol{\ell}'$ is a geodesic vector field. □

Because of (2.25), the field lines of $\boldsymbol{\ell}'$ are null geodesics and $\boldsymbol{\ell}'$ is the tangent vector to them associated with some affine parameter λ. On the other side, if $\kappa \neq 0$, $\boldsymbol{\ell}$ is not a geodesic vector field and therefore cannot be associated with some affine parameter. For this reason the quantity κ is called the ***non-affinity coefficient*** of the null normal $\boldsymbol{\ell}$ (cf. Sec. B.2.2 in Appendix B).

Since $\boldsymbol{\ell}$ is collinear to $\boldsymbol{\ell}'$, it obviously shares the same field lines, which have just been shown to be null geodesics. These field lines are called the ***null geodesic generators*** of the hypersurface $\mathscr{H}$.

Hence, we have shown that

> Any null hypersurface $\mathscr{H}$ is ruled by a family of null geodesics, called the *generators of* $\mathscr{H}$, and each vector field $\boldsymbol{\ell}$ normal to $\mathscr{H}$ is tangent to these null geodesics.

Remark 2: The above result is not trivial: while it is obvious that the field lines of the normal vector field $\boldsymbol{\ell}$ are null curves that are tangent to $\mathscr{H}$, the reader must keep in mind that not all null curves are null geodesics. For instance, in Minkowski spacetime, the helix defined in terms of some Minkowskian coordinates $(x^\alpha) = (t, x, y, z)$ by the parametric equation $x^\alpha(\lambda) = (\lambda, \cos\lambda, \sin\lambda, 0)$ is a null curve, i.e. it has a null tangent vector at each point, but it is not a null geodesic: in Minkowski spacetime, all null geodesics are straight lines.

As a by-product of (2.26), we get the behaviour of the non-affinity coefficient under a rescaling of the null normal:

$$\boldsymbol{\ell}' = \alpha \boldsymbol{\ell} \implies \kappa' = \alpha\kappa + \boldsymbol{\nabla}_{\boldsymbol{\ell}} \alpha. \tag{2.27}$$

Example 8 (null hyperplane): It is clear on expression (2.12) for $\boldsymbol{\ell}$ that the covariant derivative $\boldsymbol{\nabla}\boldsymbol{\ell}$ vanishes identically. In particular $\boldsymbol{\nabla}_{\boldsymbol{\ell}}\boldsymbol{\ell} = 0$. Equation (2.23) then implies

$$\kappa = 0, \tag{2.28}$$

which is in agreement with Eq. (2.22) and the choice $\rho = 0$ performed in Example 5. The null geodesic generators of $\mathscr{H}$ are the straight lines defined by $t = x$, $y = y_0$ and $z = z_0$ for some constants $(y_0, z_0) \in \mathbb{R}^2$. They are depicted as green lines in Fig. 2.5. Either t or x can be chosen as affine parameters of these generators.

Example 9 (light cone): From expression (2.13) for $\boldsymbol{\ell}$ and the fact that $\nabla_\beta \ell^\alpha = \partial_\beta \ell^\alpha$ in the Minkowskian coordinates (t, x, y, z), we get

$$\nabla_\beta \ell^\alpha = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \frac{y^2+z^2}{r^3} & -\frac{xy}{r^3} & -\frac{xz}{r^3} \\ 0 & -\frac{xy}{r^3} & \frac{x^2+z^2}{r^3} & -\frac{yz}{r^3} \\ 0 & -\frac{xz}{r^3} & -\frac{yz}{r^3} & \frac{x^2+y^2}{r^3} \end{pmatrix} \qquad \begin{array}{l}(\alpha = \text{row index;} \\ \beta = \text{column index}).\end{array} \tag{2.29}$$

We obtain then $\ell^\mu \nabla_\mu \ell^\alpha = 0$. From Eq. (2.23), we conclude that

$$\kappa = 0,$$

which is in agreement with Eq. (2.22) and the choice $\rho = 0$ performed in Example 6. The null geodesic generators of $\mathscr{H}$ are the half-lines defined by $x = at$, $y = bt$, $z = \sqrt{1 - a^2 - b^2}t$, with $t > 0$ and $(a, b) \in \mathbb{R}^2$ such that $a^2 + b^2 \leq 1$. They are depicted as green lines in Fig. 2.6. Since from (2.13) $\boldsymbol{\nabla}_{\boldsymbol{\ell}} t = 1$ and $\kappa = 0$, $\lambda = t$ is an affine parameter along these null geodesic generators.

Example 10 (Schwarzschild horizon): The covariant derivative of the vector field $\boldsymbol{\ell}$ as given by (2.15) is (cf. Sec. D.3.1 for the computation)

$$\nabla_\beta \ell^\alpha = \begin{pmatrix} \frac{m}{r^2} & \frac{m}{r^2}\frac{3r+2m}{r+2m} & 0 & 0 \\ \frac{m}{r^2}\frac{r-2m}{r+2m} & \frac{m}{r^2}\frac{3r^2-4m(r+m)}{(r+2m)^2} & 0 & 0 \\ 0 & 0 & \frac{r-2m}{r(r+2m)} & 0 \\ 0 & 0 & 0 & \frac{r-2m}{r(r+2m)} \end{pmatrix} \qquad \begin{array}{l}(\alpha = \text{row index;} \\ \beta = \text{column index}).\end{array} \tag{2.30}$$

Contracting with $\boldsymbol{\ell}^\beta$, we obtain

$$\boldsymbol{\nabla}_{\boldsymbol{\ell}} \boldsymbol{\ell} = \frac{4m}{(r+2m)^2} \boldsymbol{\partial}_t + \frac{4m(r-2m)}{(r+2m)^3} \boldsymbol{\partial}_r = \frac{4m}{(r+2m)^2} \boldsymbol{\ell}.$$

Hence, for any $\mathscr{H}_u$, $\kappa = 4m/(r+2m)^2$. On $\mathscr{H}$ $(r = 2m)$, we get

$$\kappa = \frac{1}{4m}. \tag{2.31}$$

This value agrees with $\kappa = \boldsymbol{\nabla}_{\boldsymbol{\ell}} \rho$ [Eq. (2.22)] and the choice (2.14) made for ρ. Contrary to Examples 8 and 9, κ does not vanish; hence t, which is a parameter of the null geodesic generators associated with $\boldsymbol{\ell}$ (since $\boldsymbol{\nabla}_{\boldsymbol{\ell}} t = 1$ by virtue of (2.15)), is *not* an affine parameter. The null geodesic generators are depicted as vertical green lines in Fig. 2.7.

2.3.4 Cross-sections

Let us now focus on the first aspect of the black hole definition given in Sec. 2.2.1: *localization*. This feature is crucial to distinguish a black hole boundary from other types of null hypersurfaces. For instance the interior of a future null cone in Minkowski spacetime

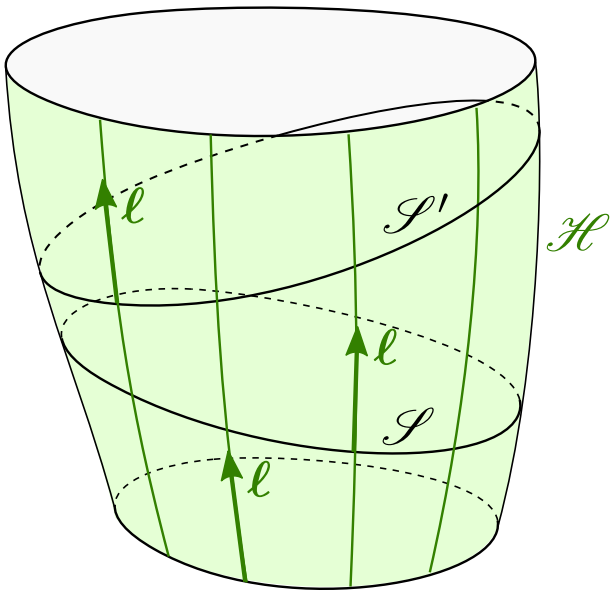


Figure 2.9: The null hypersurface $\mathscr{H}$ and two cross-sections $\mathscr{S}$ and $\mathscr{S}'$. The green curves represent some null geodesic generators, with the null normal $\boldsymbol{\ell}$ tangent to them.

is a region from which no particle may escape, but since the null cone is expanding, particles can travel arbitrary far from the center. Therefore, a null cone does not define a black hole. A key parameter is hence the *expansion* of null hypersurfaces, which we shall discuss in the next section, after having introduced cross-sections.

In the remaining of this chapter, we assume that the spacetime dimension n obeys $n \geq 3$. We define then a ***cross-section*** of the null hypersurface $\mathscr{H}$ as a submanifold $\mathscr{S}$ of $\mathscr{H}$ of codimension 2 (i.e. $\dim \mathscr{S} = n-2$), such that (i) the null normal $\boldsymbol{\ell}$ is nowhere tangent to $\mathscr{S}$ and (ii) each null geodesic generator of $\mathscr{H}$ intersects $\mathscr{S}$ once, and only once.

Notation: Indices relative to a cross-section will range from 2 to $n-1$ and will be denoted by a Latin letter from the beginning of the alphabet: a, b, etc.

To encompass the idea that an event horizon delimitates some region of spacetime, we shall assume that the cross-sections are ***closed manifolds***, i.e. are compact without boundary. The simplest example is the sphere, more precisely the $(n-2)$-dimensional sphere $\mathbb{S}^{n-2}$, where n is the spacetime dimension. It is the one relevant for standard 4-dimensional black holes. But at this stage, we shall allow for other closed-manifold topologies, like that of a torus.

Given the definition of a cross-section $\mathscr{S}$, the topology of $\mathscr{H}$ is then that of a "tube" or "cylinder" (cf. Fig. 2.9):

$$\mathscr{H} \simeq \mathbb{R} \times \mathscr{S}. \tag{2.32}$$

For the standard 4-dimensional black holes, this is $\mathscr{H} \simeq \mathbb{R} \times \mathbb{S}^2$.

A first important property is

Any cross-section $\mathcal{S}$ is spacelike, i.e. all vectors tangent to $\mathcal{S}$ are spacelike.

Proof. The spacelike character of $\mathcal{S}$ follows from

Lemma (vectors tangent to a null hypersurface)

Every nonzero vector tangent to a null hypersurface is either spacelike or null. Moreover, in the latter case, it is tangent to a null geodesic generator (i.e. it is normal to the hypersurface).

Proof. Tangent vectors to a null hypersurface $\mathcal{H}$ are by definition vectors $\boldsymbol{v}$ such that $\boldsymbol{g}(\boldsymbol{\ell}, \boldsymbol{v}) = 0$, where $\boldsymbol{\ell}$ is the normal to $\mathcal{H}$. Since $\boldsymbol{\ell}$ is null, it follows then from Corollary 1 of Sec. 1.2.2 that $\boldsymbol{v}$ cannot be timelike. Besides, if $\boldsymbol{v}$ is null, Corollary 2 of Sec. 1.2.2 implies that it must be collinear to $\boldsymbol{\ell}$. □

Let $p \in \mathcal{S}$ and $\boldsymbol{v} \in T_p\mathcal{M}$ be a nonzero vector tangent to $\mathcal{S}$. The above lemma implies that $\boldsymbol{v}$ is either spacelike or tangent to the null geodesic generator $\mathcal{L}$ going through p, but then $\mathcal{L}$ would be tangent to $\mathcal{S}$, which is not allowed, given the definition of a cross-section. We conclude that $\boldsymbol{v}$ is necessarily spacelike, which proves that $\mathcal{S}$ is a spacelike submanifold. □

Example 11 (light cone): From now on, we abandon the null hyperplane considered in Examples 1, 5 and 8, since its topology is $\mathbb{R}^3$, and therefore not of the type (2.32) with $\mathcal{S}$ compact. On the other side, the future sheet $\mathcal{H}$ of the Minkowski-spacetime light cone considered in Examples 2, 6 and 9 does obey (2.32), since we have excluded the cone apex from $\mathcal{H}$. A natural choice of cross-section is a sphere defined by $t = t_0$ for some positive constant t_0:

$$\mathcal{S} = \{p \in \mathcal{H},\ t(p) = t_0\}\,.$$

That $\mathcal{S}$ is a 2-dimensional sphere in the hyperplane $t = t_0$ is clear on its equation in terms of the Minkowskian coordinates (t, x, y, z):

$$\mathcal{S}: \quad t = t_0 \quad \text{and} \quad x^2 + y^2 + z^2 = t_0^2,$$

which follows immediately from $u = 0$ [cf. Eq. (2.4)]. Moreover, this equation shows that the radius of the sphere is t_0.

Example 12 (Schwarzschild horizon): The 3-dimensional cylinder $\mathcal{H}$ introduced in Example 3 has the topology (2.32), with $\mathcal{S} \simeq \mathbb{S}^2$ (cf. Fig. 2.7). Since it is defined by $r = 2m$ in terms of the ingoing Eddington-Finkelstein coordinates (t, r, θ, φ), a natural coordinate system on $\mathcal{H}$ is $x^A = (t, \theta, \varphi)$. Moreover, we have seen that the coordinate t is the (non-affine) parameter of the null geodesics generating $\mathcal{H}$ associated with the null normal $\boldsymbol{\ell}$. As in Example 11, a natural choice of cross-section is a sphere defined by $t = t_0$ for some constant t_0:

$$\mathcal{S} = \{p \in \mathcal{H},\ t(p) = t_0\}\,.$$

The equation of $\mathcal{S}$ in terms of the coordinates (t, r, θ, φ) is then

$$\mathcal{S}: \quad t = t_0 \quad \text{and} \quad r = 2m.$$

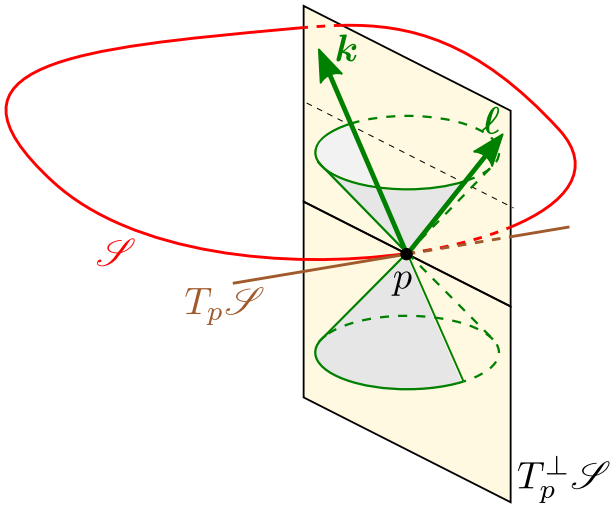


Figure 2.10: The tangent space $T_p\mathscr{S}$ to the cross-section $\mathscr{S}$ and its 2-dimensional orthogonal complement $T_p^\perp\mathscr{S}$. Only the dimensionality of the latter is respected in the figure: $\mathscr{S}$ and $T_p\mathscr{S}$ are depicted as 1-dimensional objects, while they are truly $(n-2)$-dimensional ones.

Note that $x^a = (\theta, \varphi)$ constitutes a coordinate system on $\mathscr{S}$.

Example 13 (binary black hole): Some cross-sections of the event horizon $\mathscr{H}$ in numerically generated binary black hole spacetimes are displayed in Figs. 4.20 and 4.21 of Chap. 4.

Let us denote by $\boldsymbol{q}$ the ***metric induced on*** $\mathscr{S}$ ***by*** $\boldsymbol{g}$, i.e. the bilinear form defined at any point $p \in \mathscr{S}$ by

$$\forall (\boldsymbol{u}, \boldsymbol{v}) \in T_p\mathscr{S} \times T_p\mathscr{S}, \quad \boldsymbol{q}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}). \tag{2.33}$$

Saying that $\mathscr{S}$ is spacelike is equivalent to saying that $\boldsymbol{q}$ is positive definite, i.e.

$$\forall \boldsymbol{v} \in T_p\mathscr{S}, \quad \boldsymbol{q}(\boldsymbol{v}, \boldsymbol{v}) \geq 0 \quad \text{and} \quad \boldsymbol{q}(\boldsymbol{v}, \boldsymbol{v}) = 0 \iff \boldsymbol{v} = 0. \tag{2.34}$$

In other words, $(\mathscr{S}, \boldsymbol{q})$ is a ***Riemannian manifold*** (cf Sec. A.3.2 in Appendix A).

Example 14 (Schwarzschild horizon): The metric induced by $\boldsymbol{g}$ on the cross-section $\mathscr{S}$ of the Schwarzschild horizon defined in Example 12 is readily obtained by setting $t = \text{const} = t_0$ and $r = \text{const} = 2m$ in Eq. (2.5), since $x^a = (\theta, \varphi)$ is a coordinate system on $\mathscr{S}$:

$$q_{ab} \mathrm{d}x^a \mathrm{d}x^b = 4m^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}^2\varphi \right). \tag{2.35}$$

An important consequence of $\mathscr{S}$ being spacelike is that, at each point $p \in \mathscr{S}$, the tangent space $T_p\mathscr{S}$ has an orthogonal complement $T_p^\perp\mathscr{S}$, which is a timelike plane such that $T_p\mathscr{M}$ is the direct sum of $T_p\mathscr{S}$ and $T_p^\perp\mathscr{S}$:

$$\boxed{\forall p \in \mathscr{S}, \quad T_p\mathscr{M} = T_p\mathscr{S} \oplus T_p^\perp\mathscr{S}}. \tag{2.36}$$

That $T_p^\perp\mathscr{S}$ is timelike is necessary for the signature of $\boldsymbol{g}$ to be $(-, +, \ldots, +)$. This can be seen by constructing an $\boldsymbol{g}$-orthogonal basis of $T_p\mathscr{M}$ by the Gram-Schmidt process, starting

form a $\boldsymbol{q}$-orthogonal basis of $T_p\mathscr{S}$. Since $\dim T_p\mathscr{S} = n-2$, we have $\dim T_p^\perp\mathscr{S} = 2$, i.e. $T_p^\perp\mathscr{S}$ is a timelike 2-plane. In other words, the metric induced by $\boldsymbol{g}$ on $T_p^\perp\mathscr{S}$ is Lorentzian:

$$\text{sign } \boldsymbol{g}|_{T_p^\perp\mathscr{S}} = (-,+). \tag{2.37}$$

Since $\mathscr{S} \subset \mathscr{H}$, the null normal $\boldsymbol{\ell}$ to $\mathscr{H}$ is orthogonal to any vector tangent to $\mathscr{S}$, i.e. $\boldsymbol{\ell} \in T_p^\perp\mathscr{S}$. Now, as a timelike plane, $T_p^\perp\mathscr{S}$ has two independent null directions, which can be seen as the two intersections of the null cone at p with the 2-plane $T_p^\perp\mathscr{S}$ (cf. Fig. 2.10). Let us denote by $\boldsymbol{k}$ a future-directed null vector in the null direction of $T_p^\perp\mathscr{S}$ that is not along $\boldsymbol{\ell}$. By a proper rescaling $\boldsymbol{k} \mapsto \alpha\boldsymbol{k}$, we may choose $\boldsymbol{k}$ so that

$$\boldsymbol{k}\cdot\boldsymbol{\ell} = -1. \tag{2.38}$$

Given $\boldsymbol{\ell}$ and $\mathscr{S}$, the condition (2.38) determines the null vector $\boldsymbol{k}$ uniquely. Since $\boldsymbol{\ell}$ and $\boldsymbol{k}$ are non-collinear vectors of $T_p^\perp\mathscr{S}$ and $\dim T_p^\perp\mathscr{S} = 2$, they constitute a basis of $T_p^\perp\mathscr{S}$:

$$T_p^\perp\mathscr{S} = \text{Span}\,(\boldsymbol{\ell},\boldsymbol{k})\,. \tag{2.39}$$

A priori, the bilinear form $\boldsymbol{q}$ is defined only on $T_p\mathscr{S}$, via (2.33). However, thanks to the orthogonal decomposition (2.36), we can extend it to all vectors of $T_p\mathscr{M}$ by requiring

$$\forall \boldsymbol{u} \in T_p^\perp\mathscr{S},\ \forall \boldsymbol{v} \in T_p\mathscr{M}, \quad \boldsymbol{q}(\boldsymbol{u},\boldsymbol{v}) = 0. \tag{2.40}$$

Indeed, given a pair $(\boldsymbol{u},\boldsymbol{v})$ of vectors in $T_p\mathscr{M}$, the direct sum (2.36) implies that there are unique decompositions

$$\boldsymbol{u} = \boldsymbol{u}^\parallel + \boldsymbol{u}^\perp \quad \text{and} \quad \boldsymbol{v} = \boldsymbol{v}^\parallel + \boldsymbol{v}^\perp, \quad \text{with} \quad \boldsymbol{u}^\parallel, \boldsymbol{v}^\parallel \in T_p\mathscr{S}, \quad \boldsymbol{u}^\perp, \boldsymbol{v}^\perp \in T_p^\perp\mathscr{S}. \tag{2.41}$$

Then, using the bilinearity of $\boldsymbol{q}$ and property (2.40), we obtain

$$\forall (\boldsymbol{u},\boldsymbol{v}) \in T_p\mathscr{M} \times T_p\mathscr{M}, \quad \boldsymbol{q}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{q}(\boldsymbol{u}^\parallel,\boldsymbol{v}^\parallel). \tag{2.42}$$

Equation (2.42), along with (2.33), can be considered as the definition of $\boldsymbol{q}$. An equivalent definition, which provides an explicit expression of $\boldsymbol{q}$, is

$$\boxed{\boldsymbol{q} = \boldsymbol{g} + \underline{\boldsymbol{\ell}} \otimes \underline{\boldsymbol{k}} + \underline{\boldsymbol{k}} \otimes \underline{\boldsymbol{\ell}}}, \tag{2.43}$$

or, in index notation,

$$\boxed{q_{\alpha\beta} = g_{\alpha\beta} + \ell_\alpha k_\beta + k_\alpha \ell_\beta}. \tag{2.44}$$

Proof. Let us show that (2.43) implies (2.42)-(2.33). Starting from (2.43), we have for any pair of vectors $(\boldsymbol{u},\boldsymbol{v})$ in $T_p\mathscr{M}$,

$$\boldsymbol{q}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{u}\cdot\boldsymbol{v} + (\boldsymbol{\ell}\cdot\boldsymbol{u})(\boldsymbol{k}\cdot\boldsymbol{v}) + (\boldsymbol{k}\cdot\boldsymbol{u})(\boldsymbol{\ell}\cdot\boldsymbol{v}). \tag{2.45}$$

Now, thanks to (2.39), we may write the orthogonal decompositions (2.41) as

$$\boldsymbol{u} = \boldsymbol{u}^\parallel + u^0\boldsymbol{\ell} + u^1\boldsymbol{k} \quad \text{and} \quad \boldsymbol{v} = \boldsymbol{v}^\parallel + v^0\boldsymbol{\ell} + v^1\boldsymbol{k}.$$

Using $\boldsymbol{\ell}\cdot\boldsymbol{\ell}=0$, $\boldsymbol{k}\cdot\boldsymbol{k}=0$ and $\boldsymbol{\ell}\cdot\boldsymbol{k}=-1$ [Eq. (2.38)], we have then

$$\begin{aligned}&\boldsymbol{u}\cdot\boldsymbol{v}=\boldsymbol{u}^{\parallel}\cdot\boldsymbol{v}^{\parallel}-u^0v^1-u^1v^0\\&\boldsymbol{\ell}\cdot\boldsymbol{u}=-u^1,\quad \boldsymbol{k}\cdot\boldsymbol{u}=-u^0,\quad \boldsymbol{\ell}\cdot\boldsymbol{v}=-v^1,\quad \boldsymbol{k}\cdot\boldsymbol{v}=-v^0.\end{aligned}$$

Hence (2.45) results in

$$\boldsymbol{q}(\boldsymbol{u},\boldsymbol{v})=\boldsymbol{u}^{\parallel}\cdot\boldsymbol{v}^{\parallel}-u^0v^1-u^1v^0+u^1v^0+u^0v^1=\boldsymbol{u}^{\parallel}\cdot\boldsymbol{v}^{\parallel},$$

which is nothing but (2.42). □

Example 15 (light cone): In continuation with Example 11, the null vector $\boldsymbol{k}$ orthogonal to the sphere $\mathscr{S}$ and obeying $\boldsymbol{k}\cdot\boldsymbol{\ell}=-1$ is

$$\boldsymbol{k}=\frac{1}{2}\boldsymbol{\partial}_t-\frac{x}{2r}\boldsymbol{\partial}_x-\frac{y}{2r}\boldsymbol{\partial}_y-\frac{z}{2r}\boldsymbol{\partial}_z.$$

Evaluating $\boldsymbol{q}$ via (2.43), given expression (2.13) for $\boldsymbol{\ell}$, we get the following components of $\boldsymbol{q}$ with respect to the Minkowskian coordinates $x^\alpha=(t,x,y,z)$:

$$q_{\alpha\beta}=\begin{pmatrix}0&0&0&0\\0&\frac{y^2+z^2}{r^2}&-\frac{xy}{r^2}&-\frac{xz}{r^2}\\0&-\frac{xy}{r^2}&\frac{x^2+z^2}{r^2}&-\frac{yz}{r^2}\\0&-\frac{xz}{r^2}&-\frac{yz}{r^2}&\frac{x^2+y^2}{r^2}\end{pmatrix}.$$

If we consider the spherical coordinates $x'^\alpha=(t,r,\theta,\varphi)$ deduced from the Minkowskian ones via the standard formulas:

$$\begin{cases}x=r\sin\theta\cos\varphi\\y=r\sin\theta\sin\varphi\\z=r\cos\theta\end{cases},$$

the components of $\boldsymbol{q}$ become instead

$$q'_{\alpha\beta}=\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&r^2&0\\0&0&0&r^2\sin^2\theta\end{pmatrix}. \tag{2.46}$$

and we recognize in $q_{ab}=\mathrm{diag}(r^2,r^2\sin^2\theta)$ the standard metric on the 2-sphere of radius r.

Example 16 (Schwarzschild horizon): For the Schwarzschild horizon case, we deduce from the metric (2.5) and the expression (2.15) for $\boldsymbol{\ell}$ that the null vector $\boldsymbol{k}$ orthogonal to the sphere $\mathscr{S}$ introduced in Example 12 and obeying $\boldsymbol{k}\cdot\boldsymbol{\ell}=-1$ is

$$\boldsymbol{k}=\left(\frac{1}{2}+\frac{m}{r}\right)\boldsymbol{\partial}_t-\left(\frac{1}{2}+\frac{m}{r}\right)\boldsymbol{\partial}_r. \tag{2.47}$$

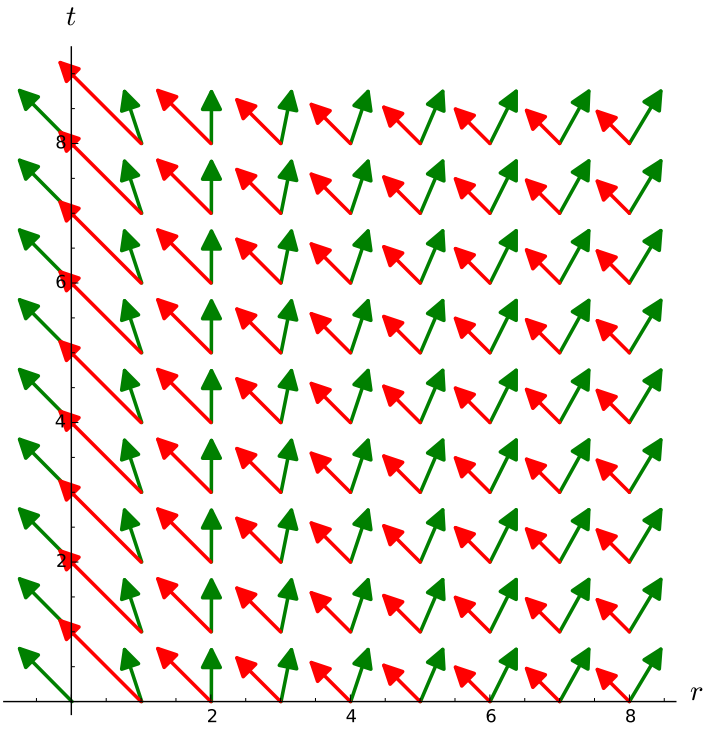


Figure 2.11: Null vector fields $\boldsymbol{\ell}$ (green) and $\boldsymbol{k}$ (red) corresponding to Example 16 (Schwarzschild horizon). The plot is a 2-dimensional slice $\theta = \text{const}$ and $\varphi = \text{const}$ of the spacetime $\mathscr{M}$, with t and r labelled in units of m. Note that since $\boldsymbol{k}$ diverges at $r = 0$ [cf. Eq. (2.47)], it is not represented there.

The vector field $\boldsymbol{k}$ is depicted in Fig. 2.11. We have (cf. Appendix D)

$$\underline{\boldsymbol{\ell}} = \frac{2m - r}{2m + r}\, \mathbf{d}t + \mathbf{d}r \qquad \text{and} \qquad \underline{\boldsymbol{k}} = -\left(\frac{1}{2} + \frac{m}{r}\right) \mathbf{d}t - \left(\frac{1}{2} + \frac{m}{r}\right) \mathbf{d}r, \tag{2.48}$$

so that Eq. (2.43) leads to the following components of $\boldsymbol{q}$ in terms of the ingoing Eddington-Finkelstein coordinates $x^\alpha = (t, r, \theta, \varphi)$:

$$q_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix}. \tag{2.49}$$

Having extended the definition of $\boldsymbol{q}$ via (2.43), we notice that the metric dual[6] of $\boldsymbol{q}$, i.e. the tensor of type $(1,1)$ defined by by

$$\boxed{\overrightarrow{\boldsymbol{q}} := \mathrm{Id} + \boldsymbol{\ell} \otimes \underline{\boldsymbol{k}} + \boldsymbol{k} \otimes \underline{\boldsymbol{\ell}}}, \tag{2.50}$$

or, in index notation,

$$\boxed{q^\alpha{}_\beta = \delta^\alpha{}_\beta + \ell^\alpha\, k_\beta + k^\alpha\, \ell_\beta}, \tag{2.51}$$

[6] See Eq. (A.44) of Appendix A for the explanation of the arrow notation.

is nothing but the ***orthogonal projector*** onto the cross-section $\mathscr{S}$:

$$\forall \boldsymbol{v} \in T_p\mathscr{M}, \quad \overrightarrow{\boldsymbol{q}}(\boldsymbol{v}) = \boldsymbol{v}^{\parallel}. \tag{2.52}$$

The demonstration follows from the decomposition $\boldsymbol{v} = \boldsymbol{v}^{\parallel} + v^0\boldsymbol{\ell} + v^1\boldsymbol{k}$ used above. In particular, we have

$$\overrightarrow{\boldsymbol{q}}(\boldsymbol{\ell}) = 0 \quad \text{and} \quad \overrightarrow{\boldsymbol{q}}(\boldsymbol{k}) = 0. \tag{2.53}$$

As stressed by (2.39), $(\boldsymbol{\ell}, \boldsymbol{k})$ forms a null basis of $T_p^\perp\mathscr{S}$. One can construct from it an *orthonormal* basis $(\boldsymbol{n}, \boldsymbol{s})$ as follows:

$$\begin{cases} \boldsymbol{n} &= \frac{1}{2}\boldsymbol{\ell} + \boldsymbol{k} \\ \boldsymbol{s} &= \frac{1}{2}\boldsymbol{\ell} - \boldsymbol{k}. \end{cases} \tag{2.54}$$

This system is easily inverted:

$$\begin{cases} \boldsymbol{\ell} &= \boldsymbol{n} + \boldsymbol{s} \\ \boldsymbol{k} &= \frac{1}{2}(\boldsymbol{n} - \boldsymbol{s}). \end{cases} \tag{2.55}$$

Since $\boldsymbol{\ell}\cdot\boldsymbol{\ell} = 0$, $\boldsymbol{k}\cdot\boldsymbol{k} = 0$ and $\boldsymbol{\ell}\cdot\boldsymbol{k} = -1$, it is easy to check that:

$$\boldsymbol{n}\cdot\boldsymbol{n} = -1, \quad \boldsymbol{s}\cdot\boldsymbol{s} = 1 \quad \text{and} \quad \boldsymbol{n}\cdot\boldsymbol{s} = 0. \tag{2.56}$$

In other words, $(\boldsymbol{n}, \boldsymbol{s})$ is an orthonormal basis of the Lorentzian plane $(T_p^\perp\mathscr{S}, \boldsymbol{g})$; in particular:

$$T_p^\perp\mathscr{S} = \mathrm{Span}\,(\boldsymbol{n}, \boldsymbol{s}). \tag{2.57}$$

If we substitute (2.55) for $\boldsymbol{\ell}$ and $\boldsymbol{k}$ in (2.43), we get

$$\boldsymbol{q} = \boldsymbol{g} + \frac{1}{2}(\underline{\boldsymbol{n}} + \underline{\boldsymbol{s}}) \otimes (\underline{\boldsymbol{n}} - \underline{\boldsymbol{s}}) + \frac{1}{2}(\underline{\boldsymbol{n}} - \underline{\boldsymbol{s}}) \otimes (\underline{\boldsymbol{n}} + \underline{\boldsymbol{s}}).$$

Expanding and simplifying results in

$$\boldsymbol{q} = \boldsymbol{g} + \underline{\boldsymbol{n}} \otimes \underline{\boldsymbol{n}} - \underline{\boldsymbol{s}} \otimes \underline{\boldsymbol{s}}. \tag{2.58}$$

2.3.5 Expansion along the null normal

Let us define the expansion of the cross-section $\mathscr{S}$ along the vector field $\boldsymbol{\ell}$ as follows. Given an infinitesimal parameter $\varepsilon \geq 0$, take a point $p \in \mathscr{S}$ and displace it by the (infinitesimal) vector $\varepsilon\boldsymbol{\ell}$, thereby getting a nearby point p_ε (cf. Fig. 2.12). Since $\boldsymbol{\ell}$ is tangent to $\mathscr{H}$ and $p \in \mathscr{H}$, we have $p_\varepsilon \in \mathscr{H}$. By repeating this for each point in $\mathscr{S}$, keeping the value of ε fixed, we define a new codimension-2 surface, $\mathscr{S}_\varepsilon$ say (cf. Fig. 2.12). One says that $\mathscr{S}_\varepsilon$ is obtained from $\mathscr{S}$ by ***Lie dragging along $\boldsymbol{\ell}$ by the parameter*** ε. Note that $\mathscr{S}_0 = \mathscr{S}$. Since $p_\varepsilon \in \mathscr{H}$ for every $p \in \mathscr{S}$, we have $\mathscr{S}_\varepsilon \subset \mathscr{H}$. Because the null

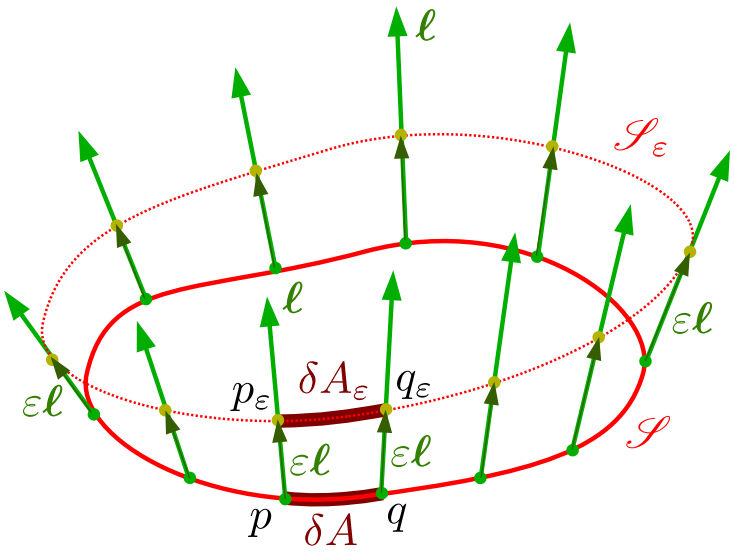


Figure 2.12: Lie dragging of the surface $\mathcal{S}$ along $\boldsymbol{\ell}$ by the small parameter ε. $\mathcal{S}$ is drawn as a 1-dimensional submanifold, while it is actually a $(n-2)$-dimensional one, n being the spacetime dimension.

direction $\boldsymbol{\ell}$ is transverse to $\mathcal{S}_\varepsilon$ by construction, it follows that $\mathcal{S}_\varepsilon$ is spacelike (cf. the lemma in Sec. 2.3.4).

At each point $p \in \mathcal{S}$, the ***expansion of*** $\mathcal{S}$ ***along*** $\boldsymbol{\ell}$ is defined from the rate of change $\theta_{(\boldsymbol{\ell})}$ of the area[7] δA of an element of surface δS of $\mathcal{S}$ around p:

$$\boxed{\theta_{(\boldsymbol{\ell})} := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \frac{\delta A_\varepsilon - \delta A}{\delta A}}. \tag{2.59}$$

In the above formula, δA_ε stands for the area of the surface element $\delta S_\varepsilon \subset \mathcal{S}_\varepsilon$ that is obtained from δS by Lie dragging along $\boldsymbol{\ell}$ by the parameter ε (cf. Fig. 2.12).

Remark 3: The reader may wonder why the expansion is not denoted by something like $\theta_{(\boldsymbol{\ell})}(\mathcal{S})$, since its definition depends explicitly on $\mathcal{S}$. We shall show below that, because $\mathcal{H}$ is a null hypersurface, $\theta_{(\boldsymbol{\ell})}$ is actually independent of the choice of the cross-section $\mathcal{S}$.

For concreteness, let us assume that the element of surface $\delta S \subset \mathcal{S}$ is a $(n-2)$-dimensional parallelogram delimited by some infinitesimal displacement vectors $\mathrm{d}\boldsymbol{x}_{(2)}$, ..., $\mathrm{d}\boldsymbol{x}_{(n-1)}$. The area of δS is then

$$\delta A = {}^{\mathcal{S}}\boldsymbol{\epsilon}(\mathrm{d}\boldsymbol{x}_{(2)}, \ldots, \mathrm{d}\boldsymbol{x}_{(n-1)}), \tag{2.60}$$

where ${}^{\mathcal{S}}\boldsymbol{\epsilon}$ is the Levi-Civita tensor associated with the metric $\boldsymbol{q}$ in $\mathcal{S}$ (cf. Sec. A.3.4 in Appendix A). Since $\boldsymbol{q}$ is the metric induced by $\boldsymbol{g}$ in $\mathcal{S}$ and $(\boldsymbol{n}, \boldsymbol{s})$ is an orthonormal basis of $T_p^\perp\mathcal{S}$, ${}^{\mathcal{S}}\boldsymbol{\epsilon}$ is actually the alternating form induced on $\mathcal{S}$ by the spacetime Levi-Civita tensor $\boldsymbol{\epsilon}$:

$$\boxed{{}^{\mathcal{S}}\boldsymbol{\epsilon} = \boldsymbol{\epsilon}(\boldsymbol{n}, \boldsymbol{s}, \ldots)}, \tag{2.61}$$

[7] We are using the words *area* and *surface* even if $n-2 \neq 2$, i.e. even if $n \neq 4$, being aware that for $n = 3$ the words *length* and *line* would be more appropriate, as well as *volume* for $n = 5$.

or, in index notation,

$$ {}^{\mathscr{S}}\epsilon_{\alpha_1\cdots\alpha_{n-2}} = \epsilon_{\mu\nu\alpha_1\cdots\alpha_{n-2}} n^\mu s^\nu . $$

Proof. To demonstrate (2.61), it suffices to note that its right-hand side defines a fully antisymmetric $(n-2)$-linear form on $T_p\mathscr{S}$. Since the space of such forms is 1-dimensional (for $\dim T_p\mathscr{S} = n-2$), we have then necessarily $\boldsymbol{\epsilon}(\boldsymbol{n}, \boldsymbol{s}, \ldots) = a\, {}^{\mathscr{S}}\boldsymbol{\epsilon}$ for some proportionality factor a. Since $\boldsymbol{\epsilon}(\boldsymbol{n}, \boldsymbol{s}, \mathrm{d}\boldsymbol{x}_{(2)}, \ldots, \mathrm{d}\boldsymbol{x}_{(n-1)})$ is the volume of the n-parallelepiped constructed on the vectors $\boldsymbol{n}, \boldsymbol{s}, \mathrm{d}\boldsymbol{x}_{(2)}, \ldots, \mathrm{d}\boldsymbol{x}_{(n-1)}$ and $\boldsymbol{n}$ and $\boldsymbol{s}$ are unit-length vectors for the metric $\boldsymbol{g}$, we have

$$ \boldsymbol{\epsilon}(\boldsymbol{n}, \boldsymbol{s}, \mathrm{d}\boldsymbol{x}_{(2)}, \ldots, \mathrm{d}\boldsymbol{x}_{(n-1)}) = \delta A. $$

This implies that $a = 1$, thereby establishing (2.61). □

An alternative expression of ${}^{\mathscr{S}}\boldsymbol{\epsilon}$ is obtained by substituting (2.54) for $\boldsymbol{n}$ and $\boldsymbol{s}$ in (2.61). Thanks to the multilinearity and antisymmetry of $\boldsymbol{\epsilon}$, we get

$$ \boxed{{}^{\mathscr{S}}\boldsymbol{\epsilon} = \boldsymbol{\epsilon}(\boldsymbol{k}, \boldsymbol{\ell}, \ldots)} . \tag{2.62} $$

Let us consider in some vicinity of $\mathscr{S}$ a coordinate system

$$ x^\alpha = \left(\varepsilon, u, x^2, \ldots, x^{n-1}\right) $$

that is adapted to $\mathscr{S}$ and $\boldsymbol{\ell}$ in the sense that

$$ \boldsymbol{\ell} = \frac{\partial}{\partial \varepsilon} \tag{2.63} $$

and the points of $\mathscr{S}$ are defined by $(\varepsilon, u) = (0, 0)$. Then, from the very definition of the Lie dragging of $\mathscr{S}$ along $\boldsymbol{\ell}$, we have

$$ \mathscr{S}_\varepsilon = \left\{ p \in \mathscr{M}, \quad (x^0(p), x^1(p)) = (\varepsilon, 0) \right\} \tag{2.64} $$

and $x^a = (x^2, \ldots, x^{n-1})$ can be viewed as a coordinate system[8] on each surface $\mathscr{S}_\varepsilon$. Let us choose the $n-2$ infinitesimal displacement vectors in (2.60) along the coordinate lines of this system:

$$ \mathrm{d}x^a_{(i)} = (\underbrace{0, \ldots, 0}_{i-2}, \mathrm{d}x^i, \underbrace{0, \ldots, 0}_{n-1-i}), \qquad 2 \leq i \leq n-1. \tag{2.65} $$

Then expression (2.60) for the area of δS becomes

$$ \begin{aligned} \delta A &= {}^{\mathscr{S}}\epsilon_{a_1\cdots a_{n-2}}\, \mathrm{d}x^{a_1}_{(2)} \cdots \mathrm{d}x^{a_{n-2}}_{(n-1)} \\ &= {}^{\mathscr{S}}\epsilon_{2\cdots(n-1)}\, \mathrm{d}x^2 \cdots \mathrm{d}x^{n-1} \\ \delta A &= \sqrt{q}\, \mathrm{d}x^2 \cdots \mathrm{d}x^{n-1}, \end{aligned} \tag{2.66} $$

where we have used (A.49) for the components of the Levi-Civita tensor ${}^{\mathscr{S}}\boldsymbol{\epsilon}$, q standing for the determinant of the metric $\boldsymbol{q}$ with respect to the coordinates $(x^2, \ldots, x^{n-1})$. By

[8] Let us recall that according to the convention stated in Sec. 2.3.4, Latin indices from the beginning of the alphabet, a, b, etc. range from 2 to $n-1$.

the very definition of the Lie dragging, the surface element δS_ε on $\mathcal{S}_\varepsilon$ is defined by the same values of the coordinates $(x^2, \ldots, x^{n-1})$ as δS. In particular, the small coordinate increments $\mathrm{d}x^2$, ..., $\mathrm{d}x^{n-1}$ take the same values as on $\mathcal{S}$. Therefore, the area of δS_ε is

$$\delta A_\varepsilon = \sqrt{q(\varepsilon)}\, \mathrm{d}x^2 \cdots \mathrm{d}x^{n-1}, \tag{2.67}$$

where $q(\varepsilon)$ stands for the determinant of the components of the metric $\boldsymbol{q}(\varepsilon)$ induced by $\boldsymbol{g}$ on $\mathcal{S}_\varepsilon$. Since $\mathcal{S}_\varepsilon$ is spacelike (cf. above), $\boldsymbol{q}(\varepsilon)$ is positive definite, so that $q(\varepsilon) \geq 0$.

In view of (2.66)-(2.67), the definition (2.59) of the expansion of $\mathcal{S}$ along $\boldsymbol{\ell}$ can be rewritten as

$$\theta_{(\boldsymbol{\ell})} = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \frac{\sqrt{q(\varepsilon)} - \sqrt{q(0)}}{\sqrt{q(0)}}.$$

We recognize the derivative of the function $\varepsilon \mapsto \ln \sqrt{q(\varepsilon)} = 1/2 \ln q(\varepsilon)$ at $\varepsilon = 0$:

$$\theta_{(\boldsymbol{\ell})} = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}\varepsilon} \ln q. \tag{2.68}$$

Given that $\mathcal{S}_\varepsilon$ is deduced from $\mathcal{S}$ by a Lie dragging along $\boldsymbol{\ell}$ and ε is the parameter associated with $\boldsymbol{\ell}$ [cf. Eq. (2.63)], we may rewrite this formula as the Lie derivative of $\ln q$ along $\boldsymbol{\ell}$:

$$\boxed{\theta_{(\boldsymbol{\ell})} = \frac{1}{2} \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \ln q}. \tag{2.69}$$

Example 17 (light cone): For the light cone in Minkowski spacetime, it is easy to evaluate $\theta_{(\boldsymbol{\ell})}$ by means of the spherical coordinates introduced in Example 15, since these coordinates are adapted to the surface $\mathcal{S}$, the metric of $\mathcal{S}$ being $q_{ab}\mathrm{d}x^a\mathrm{d}x^b = r^2\mathrm{d}\theta^2 + r^2 \sin^2\theta \mathrm{d}\varphi^2$ [cf. Eq. (2.46)]. We have then $q = \det(q_{ab}) = r^4 \sin^2\theta$. Moreover, the parameter ε can be chosen as $\varepsilon = t - t_0$ since t is an (affine) parameter associated with $\boldsymbol{\ell}$ (cf. Example 9). Given that $t = r$ on $\mathscr{H}$, we have $\varepsilon = r - t_0$, so that (2.68) yields

$$\theta_{(\boldsymbol{\ell})} = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}r} \ln q = \frac{1}{2} \frac{\mathrm{d}}{\mathrm{d}r} \left(4 \ln r + 2 \ln \sin\theta\right),$$

i.e.

$$\theta_{(\boldsymbol{\ell})} = \frac{2}{r}. \tag{2.70}$$

Example 18 (Schwarzschild horizon): As above, we have $q = r^4 \sin^2\theta$ [cf. Eq. (2.35)], so that (2.69) yields

$$\begin{aligned} \theta_{(\boldsymbol{\ell})} &= \frac{1}{2} \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \ln q = \frac{1}{2} \ell^\mu \frac{\partial}{\partial x^\mu} \ln q = \underbrace{\frac{\partial}{\partial t} \ln(r^2 \sin\theta)}_{0} + \frac{r - 2m}{r + 2m} \frac{\partial}{\partial r} \ln(r^2 \sin\theta) \\ &= \frac{2}{r} \frac{r - 2m}{r + 2m}. \end{aligned} \tag{2.71}$$

where we have used (2.15) for the components ℓ^μ. The above expression is valid for any null hypersurface $\mathscr{H}_u$. For the specific case of the Schwarzschild horizon, $r = 2m$ and (2.71) yields a vanishing expansion:

$$\theta_{(\boldsymbol{\ell})} = 0. \tag{2.72}$$

Note that for large r, Eq. (2.71) yields $\theta_{(\boldsymbol{\ell})} \sim 2/r$, i.e. we recover the flat spacetime result (2.70), which is consistent with the fact that for large r, $\mathscr{H}_u$ is closed to a Minkowskian light cone (cf. Fig. 2.8). Note also that Eq. (2.71) yields $\theta_{(\boldsymbol{\ell})} < 0$ for $r < 2m$ and $\theta_{(\boldsymbol{\ell})} > 0$ for $r > 2m$. These expansion values are in agreement with what can be inferred from Fig. 2.11, since r is directly related to the area of the cross-sections of $\mathscr{H}$: $A = 4\pi r^2$ from Eq. (2.49) and $\boldsymbol{\ell}$ points towards decreasing (resp. increasing) values of r for $r < 2m$ (resp. $r > 2m$).

Using the general law of variation of a determinant, as given by Eq. (A.73) in Appendix A, Eq. (2.69) can be rewritten as

$$\theta_{(\boldsymbol{\ell})} = \frac{1}{2} \operatorname{tr}\left(Q^{-1} \times \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} Q\right),$$

when Q is the matrix representing the components of $\boldsymbol{q}$ with respect to the coordinates $(x^a) = (x^2, \ldots, x^{n-1})$. In index notation, we have $Q = (q_{ab})$ and $Q^{-1} = (q^{ab})$. Hence

$$\boxed{\theta_{(\boldsymbol{\ell})} = \frac{1}{2} q^{ab} \mathcal{L}_{\boldsymbol{\ell}} q_{ab}}. \tag{2.73}$$

The Lie derivative along $\boldsymbol{\ell}$ of the metric $\boldsymbol{q}$ of the cross-section $\mathscr{S}$ that appears in this formula is defined as follows. As in Sec. A.4.2 of Appendix A, let us denote by Φ_ε the smooth map $\mathscr{S} \to \mathscr{H}$ that corresponds to the displacement of points of $\mathscr{S}$ by some infinitesimal quantity ε along $\boldsymbol{\ell}$. Using the notations of Fig. 2.12, we have then $p_\varepsilon = \Phi_\varepsilon(p)$, $q_\varepsilon = \Phi_\varepsilon(q)$ and $\mathscr{S}_\varepsilon = \Phi_\varepsilon(\mathscr{S})$. The ***Lie derivative along $\boldsymbol{\ell}$ of $\boldsymbol{q}$*** is then the field $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \boldsymbol{q}$ of bilinear forms on $\mathscr{S}$ defined by the following action on any pair of vectors $(\boldsymbol{u}, \boldsymbol{v})$ tangent to $\mathscr{S}$ at the same point p:

$$\boxed{\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \boldsymbol{q}\,(\boldsymbol{u}, \boldsymbol{v}) := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[\boldsymbol{q}(\varepsilon)(\Phi_\varepsilon^* \boldsymbol{u}, \Phi_\varepsilon^* \boldsymbol{v}) - \boldsymbol{q}(\boldsymbol{u}, \boldsymbol{v})\right]}, \tag{2.74}$$

where, as above, $\boldsymbol{q}(\varepsilon)$ is the metric induced by $\boldsymbol{g}$ on the $(n-2)$-surface $\mathscr{S}_\varepsilon$ deduced from $\mathscr{S}$ by Lie dragging along $\boldsymbol{\ell}$ by the quantity ε and $\Phi_\varepsilon^* \boldsymbol{u}$ (resp. $\Phi_\varepsilon^* \boldsymbol{v}$) is the vector tangent to $\mathscr{S}_\varepsilon$ at $\Phi_\varepsilon(p)$ that is the pushforward of $\boldsymbol{u}$ (resp. $\boldsymbol{v}$) by the map Φ_ε (cf. Sec. A.4.2).

Remark 4: Since $\boldsymbol{q}$ is nothing but the metric induced by the spacetime metric $\boldsymbol{g}$ on cross-sections of $\mathscr{H}$, we may rewrite the above formula as

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \boldsymbol{q}\,(\boldsymbol{u}, \boldsymbol{v}) := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[\boldsymbol{g}|_{\Phi_\varepsilon(p)}(\Phi_\varepsilon^* \boldsymbol{u}, \Phi_\varepsilon^* \boldsymbol{v}) - \boldsymbol{g}|_p\,(\boldsymbol{u}, \boldsymbol{v})\right]. \tag{2.75}$$

One may wonder about the link between the Lie derivative $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \boldsymbol{q}$ defined by Eq. (2.74), which is a tensor field on $\mathscr{S}$, and the Lie derivative along $\boldsymbol{\ell}$ of the spacetime extension $\boldsymbol{q}$ introduced by Eq. (2.43). For the sake of clarity, let us denote here the latter by $\bar{\boldsymbol{q}}$. More precisely, we may consider that $\bar{\boldsymbol{q}}$ is a field defined in some neighbourhood of the portion of $\mathscr{H}$ sliced by $\bigcup_\varepsilon \mathscr{S}_\varepsilon$ via Eq. (2.43), with $\boldsymbol{k}$ defined at each point $p \in \mathscr{S}_\varepsilon$ as the unique null vector of $T_p^\perp \mathscr{S}_\varepsilon$ obeying $\boldsymbol{\ell} \cdot \boldsymbol{k} = -1$. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be vector fields on $\mathscr{H}$ that are tangent to the cross-sections $\mathscr{S}_\varepsilon$. Applying the bilinear form $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \bar{\boldsymbol{q}}$ to them and using the Leibniz rule to expand $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,[\bar{\boldsymbol{q}}(\boldsymbol{u}, \boldsymbol{v})]$ yields

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \bar{\boldsymbol{q}}\,(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,[\bar{\boldsymbol{q}}(\boldsymbol{u}, \boldsymbol{v})] - \bar{\boldsymbol{q}}\,(\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \boldsymbol{u}, \boldsymbol{v}) - \bar{\boldsymbol{q}}\,(\boldsymbol{u}, \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \boldsymbol{v}). \tag{2.76}$$

Now, since $\boldsymbol{u}$ and $\boldsymbol{v}$ are tangent to $\mathscr{S}_\varepsilon$, we may write $\bar{\boldsymbol{q}}(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{q}(\boldsymbol{u},\boldsymbol{v})$. Moreover, by the very definition of the Lie derivative of a vector field (cf. Sec. A.4.2) and the fact that the cross-sections $\mathscr{S}_\varepsilon$ are Lie-dragged along $\boldsymbol{\ell}$, the vectors $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{u}$ and $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{v}$ are also tangent to $\mathscr{S}_\varepsilon$. Therefore, we have

$$\bar{\boldsymbol{q}}\,(\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{q}\,(\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{u},\boldsymbol{v}) \quad \text{and} \quad \bar{\boldsymbol{q}}\,(\boldsymbol{u},\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{v}) = \boldsymbol{q}\,(\boldsymbol{u},\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{v})$$

as well. Thus, we may rewrite (2.76) as

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\bar{\boldsymbol{q}}\,(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,[\boldsymbol{q}(\boldsymbol{u},\boldsymbol{v})] - \boldsymbol{q}\,(\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{u},\boldsymbol{v}) - \boldsymbol{q}\,(\boldsymbol{u},\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{v})\,.$$

The right-hand side is identical to what would be obtained by expressing $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{q}\,(\boldsymbol{u},\boldsymbol{v})$ via the Leibniz rule. Hence we conclude that

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\bar{\boldsymbol{q}}\,(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{q}\,(\boldsymbol{u},\boldsymbol{v}).$$

Since this identity holds for a pair $(\boldsymbol{u},\boldsymbol{v})$ of vectors tangent to $\mathscr{S}_\varepsilon$, we may express it for any pair of vectors, i.e. not necessarily tangent to $\mathscr{S}_\varepsilon$ by introducing the orthogonal projector $\overrightarrow{\boldsymbol{q}}$ onto $\mathscr{S}_\varepsilon$ [cf. Eq. (2.50)]:

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\bar{\boldsymbol{q}}\,(\overrightarrow{\boldsymbol{q}}(\boldsymbol{u}),\overrightarrow{\boldsymbol{q}}(\boldsymbol{v})) = \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{q}\,(\overrightarrow{\boldsymbol{q}}(\boldsymbol{u}),\overrightarrow{\boldsymbol{q}}(\boldsymbol{v})). \tag{2.77}$$

Using index notation, this is equivalent to

$$\mathcal{L}_{\boldsymbol{\ell}}\,\bar{q}_{\mu\nu}\,\bar{q}^{\mu}{}_{\alpha}\bar{q}^{\nu}{}_{\beta} = \mathcal{L}_{\boldsymbol{\ell}}\,q_{ab}\,\bar{q}^{a}{}_{\alpha}\bar{q}^{b}{}_{\beta}.$$

Taking the trace with respect to $\boldsymbol{g}$, we get

$$\mathcal{L}_{\boldsymbol{\ell}}\,\bar{q}_{\mu\nu}\,\bar{q}^{\mu}{}_{\sigma}\bar{q}^{\nu\sigma} = \mathcal{L}_{\boldsymbol{\ell}}\,q_{ab}\,\bar{q}^{a}{}_{\sigma}\bar{q}^{b\sigma}.$$

Now, since $\bar{\boldsymbol{q}}$ is symmetric and $\overrightarrow{\boldsymbol{q}}$ is a projector, $\bar{q}^{\mu}{}_{\sigma}\bar{q}^{\nu\sigma} = \bar{q}^{\mu}{}_{\sigma}\bar{q}^{\sigma\nu} = \bar{q}^{\mu\nu}$. Similarly, $\bar{q}^{a}{}_{\sigma}\bar{q}^{b\sigma} = \bar{q}^{ab}$. Hence

$$\bar{q}^{\mu\nu}\mathcal{L}_{\boldsymbol{\ell}}\,\bar{q}_{\mu\nu} = \bar{q}^{ab}\mathcal{L}_{\boldsymbol{\ell}}\,q_{ab} = q^{ab}\mathcal{L}_{\boldsymbol{\ell}}\,q_{ab},$$

where the second equality follows from $\bar{q}^{ab} = q^{ab}$. Hence we may rewrite (2.73) as

$$\boxed{\theta_{(\boldsymbol{\ell})} = \frac{1}{2}\,q^{\mu\nu}\mathcal{L}_{\boldsymbol{\ell}}\,q_{\mu\nu}}. \tag{2.78}$$

Note that we have dropped the bar over q, i.e. we revert to the previous notation.

Substituting (2.43) for $q_{\mu\nu}$, and using the Leibniz rule, we get

$$\theta_{(\boldsymbol{\ell})} = \frac{1}{2}\,q^{\mu\nu}\left(\mathcal{L}_{\boldsymbol{\ell}}\,g_{\mu\nu} + \mathcal{L}_{\boldsymbol{\ell}}\,\ell_\mu\;k_\nu + \ell_\mu\,\mathcal{L}_{\boldsymbol{\ell}}\,k_\nu + \mathcal{L}_{\boldsymbol{\ell}}\,k_\mu\;\ell_\nu + k_\mu\,\mathcal{L}_{\boldsymbol{\ell}}\,\ell_\nu\right).$$

If we express the Lie derivative $\mathcal{L}_{\boldsymbol{\ell}}\,g_{\mu\nu}$ in terms of the covariant derivative $\boldsymbol{\nabla}$ via Eq. (A.87) of Appendix A, we get

$$\mathcal{L}_{\boldsymbol{\ell}}\,g_{\mu\nu} = \ell^\sigma\underbrace{\nabla_\sigma g_{\mu\nu}}_{0} + \underbrace{g_{\sigma\nu}\nabla_\mu\ell^\sigma}_{\nabla_\mu\ell_\nu} + \underbrace{g_{\mu\sigma}\nabla_\nu\ell^\sigma}_{\nabla_\nu\ell_\mu} = \nabla_\mu\ell_\nu + \nabla_\nu\ell_\mu.$$

Moreover, since $\boldsymbol{\ell}$ and $\boldsymbol{k}$ are orthogonal to $\mathscr{S}$, we have $q^{\mu\nu}\ell_\nu = 0$ and $q^{\mu\nu}k_\nu = 0$. Hence we end up with

$$\theta_{(\boldsymbol{\ell})} = \frac{1}{2}\, q^{\mu\nu}\left(\nabla_\mu \ell_\nu + \nabla_\nu \ell_\mu\right),$$

i.e. since $q^{\mu\nu}$ is symmetric,

$$\boxed{\theta_{(\boldsymbol{\ell})} = q^{\mu\nu}\nabla_\mu \ell_\nu}. \tag{2.79}$$

We can transform further this relation by expressing $q^{\mu\nu}$ via (2.43):

$$\begin{aligned}
\theta_{(\boldsymbol{\ell})} &= \left(g^{\mu\nu} + \ell^\mu k^\nu + k^\mu \ell^\nu\right)\nabla_\mu \ell_\nu \\
&= \nabla_\mu \ell^\mu + k^\nu \underbrace{\ell^\mu \nabla_\mu \ell_\nu}_{\kappa \ell_\nu} + k^\mu \ell^\nu \nabla_\mu \ell_\nu \\
&= \nabla_\mu \ell^\mu + \kappa \underbrace{k^\nu \ell_\nu}_{-1} + \frac{1}{2}\underbrace{k^\mu \nabla_\mu(\ell_\nu \ell^\nu)}_{0} \\
&= \nabla_\mu \ell^\mu - \kappa,
\end{aligned} \tag{2.80}$$

where we have used respectively the properties (2.23), (2.20) and (2.38). Denoting the divergence of $\boldsymbol{\ell}$ by $\boldsymbol{\nabla}\cdot\boldsymbol{\ell} = \nabla_\mu \ell^\mu$, we have then

$$\boxed{\theta_{(\boldsymbol{\ell})} = \boldsymbol{\nabla}\cdot\boldsymbol{\ell} - \kappa}. \tag{2.81}$$

Remark 5: Contrary to $\theta_{(\boldsymbol{\ell})}$ or κ, the quantity $\boldsymbol{\nabla}\cdot\boldsymbol{\ell}$ depends a priori on the extension of $\boldsymbol{\ell}$ outside $\mathscr{H}$ (cf. the discussion in Sec. 2.3.2). For Eq. (2.81) to hold, we have supposed that $\boldsymbol{\ell}$ remains null outside $\mathscr{H}$, so that $k^\mu \nabla_\mu(\ell_\nu \ell^\nu)$, which is a derivative in a direction transverse to $\mathscr{H}$, could be set to zero in the computation leading to (2.80).

Example 19 (light cone): $\boldsymbol{\nabla}\cdot\boldsymbol{\ell}$ is easily computed by taking the trace of (2.29) and we have $\kappa = 0$ (cf. Example 9), so that (2.81) yields

$$\theta_{(\boldsymbol{\ell})} = \frac{2(x^2+y^2+z^2)}{r^3} = \frac{2}{r}.$$

Hence we recover the result obtained in Example 17.

Example 20 (Schwarzschild horizon): Here also, $\boldsymbol{\nabla}\cdot\boldsymbol{\ell}$ is easily computed by taking the trace of (2.30):

$$\boldsymbol{\nabla}\cdot\boldsymbol{\ell} = \frac{m}{r^2} + \frac{m}{r^2}\frac{3r^2 - 4m(r+m)}{(r+2m)^2} + 2\frac{r-2m}{r(r+2m)} = \frac{2(r^2+2mr-4m^2)}{r(r+2m)^2}.$$

Given the value $\kappa = 4m/(r+2m)^2$ found in Example 10, formula (2.81) leads to

$$\theta_{(\boldsymbol{\ell})} = \frac{2(r^2+2mr-4m^2) - 4mr}{r(r+2m)^2} = \frac{2(r^2-4m^2)}{r(r+2m)^2} = \frac{2}{r}\frac{r-2m}{r+2m}.$$

Hence we recover the result (2.71).

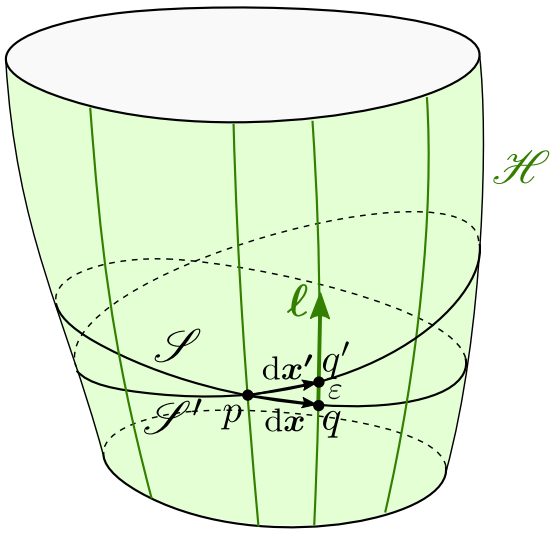


Figure 2.13: Two cross-sections $\mathscr{S}$ and $\mathscr{S}'$ through the same point p of $\mathscr{H}$.

We notice that the right-hand side of (2.81) is independent of the explicit choice of the cross-section $\mathscr{S}$: clearly both $\boldsymbol{\nabla}\cdot\boldsymbol{\ell}$ and κ depends only on the null normal $\boldsymbol{\ell}$ of $\mathscr{H}$. This justifies the notation $\theta_{(\boldsymbol{\ell})}$, which does not refer to $\mathscr{S}$ (cf. Remark 3 in page 44). This can be understood geometrically as follows. Let $p \in \mathscr{H}$ be a point where one would like to evaluate $\theta_{(\boldsymbol{\ell})}$. Let $\mathscr{S}$ and $\mathscr{S}'$ be two distinct cross-sections of $\mathscr{H}$ going through p (cf. Fig. 2.13). Let q be a point of $\mathscr{S}$ infinitely close to p and let q' be the point of $\mathscr{S}'$ located on the same null geodesic generator as q, i.e. $\overrightarrow{qq'} = \varepsilon\boldsymbol{\ell}$, with ε infinitely small. Let $\mathrm{d}\boldsymbol{x}$ (resp. $\mathrm{d}\boldsymbol{x}'$) be the infinitesimal vector connecting p to q (resp. p to q'). We have then

$$\mathrm{d}\boldsymbol{x}' = \mathrm{d}\boldsymbol{x} + \varepsilon\boldsymbol{\ell},$$

the scalar square of which is

$$\mathrm{d}\boldsymbol{x}'\cdot\mathrm{d}\boldsymbol{x}' = \mathrm{d}\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x} + 2\varepsilon\underbrace{\mathrm{d}\boldsymbol{x}\cdot\boldsymbol{\ell}}_{0} + \varepsilon^2\underbrace{\boldsymbol{\ell}\cdot\boldsymbol{\ell}}_{0},$$

where we have used the fact that $\boldsymbol{\ell}$ is normal to any vector tangent to $\mathscr{H}$, such as $\mathrm{d}\boldsymbol{x}$ and $\boldsymbol{\ell}$ itself. Hence

$$\mathrm{d}\boldsymbol{x}'\cdot\mathrm{d}\boldsymbol{x}' = \mathrm{d}\boldsymbol{x}\cdot\mathrm{d}\boldsymbol{x}.$$

In other words, the lengths of all segments from p do not depend on the cross-section in which they are taken, provided their second end lies on the same null geodesic generator of $\mathscr{H}$. It follows that all infinitesimal surfaces δS that (i) contain p and (ii) are enclosed in a tube made of null geodesic generators have the same area δA. Hence the expansion $\theta_{(\boldsymbol{\ell})}$ at p does not depend on the choice of δS, i.e. of the cross-section $\mathscr{S}$ through p. We conclude that

The expansion $\theta_{(\boldsymbol{\ell})}$ depends only on the choice of the null normal $\boldsymbol{\ell}$ on the null hypersurface $\mathscr{H}$.

For this reason, from now on, we shall call $\theta_{(\boldsymbol{\ell})}$ the ***expansion of the null hypersurface $\mathscr{H}$ along $\boldsymbol{\ell}$***.

The dependency of the expansion on $\boldsymbol{\ell}$ is given by the following behaviour under a rescaling of $\boldsymbol{\ell}$:

$$\boldsymbol{\ell}' = \alpha\boldsymbol{\ell} \implies \theta_{(\boldsymbol{\ell}')} = \alpha\theta_{(\boldsymbol{\ell})}, \tag{2.82}$$

where α is any positive scalar field on $\mathscr{H}$. This follows immediately from the expression (2.69) of $\theta_{(\boldsymbol{\ell})}$, given that the metric $\boldsymbol{q}$ is independent of $\boldsymbol{\ell}$ and $\boldsymbol{\mathcal{L}}_{\alpha\boldsymbol{\ell}} \ln q = \alpha \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \ln q$.

Remark 6: The reader may check that the rescaling laws (2.27) and (2.82) for respectively κ and $\theta_{(\boldsymbol{\ell})}$ are compatible with the expression (2.81) of $\theta_{(\boldsymbol{\ell})}$, given that $\boldsymbol{\nabla}\cdot\boldsymbol{\ell}' = \alpha\boldsymbol{\nabla}\cdot\boldsymbol{\ell} + \boldsymbol{\nabla}_{\boldsymbol{\ell}}\alpha$.

Let us gather all the expressions of the expansion $\theta_{(\boldsymbol{\ell})}$ obtained so far:

$$\boxed{\theta_{(\boldsymbol{\ell})} = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\frac{\delta A_\varepsilon - \delta A}{\delta A} = \frac{1}{2}\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}} \ln q = \frac{1}{2} q^{\mu\nu}\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, q_{\mu\nu} = q^{\mu\nu}\nabla_\mu \ell_\nu = \boldsymbol{\nabla}\cdot\boldsymbol{\ell} - \kappa}, \tag{2.83}$$

with the reminder that the last equality is valid insofar as the vector field $\boldsymbol{\ell}$ is null in some entire open neighbourhood of $\mathscr{H}$ (and not only on $\mathscr{H}$), as stressed in Remark 5.

2.3.6 Deformation rate and shear tensor

Let us consider a cross-section $\mathscr{S}$ of the null hypersurface $\mathscr{H}$. The ***deformation rate $\boldsymbol{\Theta}$ of $\mathscr{S}$ along $\boldsymbol{\ell}$*** is defined from the Lie derivative along $\boldsymbol{\ell}$ of the induced metric $\boldsymbol{q}$ on $\mathscr{S}$ [cf. Eq. (2.74)] as

$$\boxed{\boldsymbol{\Theta} := \frac{1}{2}\overrightarrow{\boldsymbol{q}}^*\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{q}}, \tag{2.84}$$

where $\overrightarrow{\boldsymbol{q}}^*$ stands for the action of the orthogonal projector $\overrightarrow{\boldsymbol{q}}$ onto $\mathscr{S}$ on the bilinear form $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{q}$. This action extends $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{q}$, which is defined a priori only for vectors of $T_p\mathscr{S}$ by Eq. (2.74), to all vectors of $T_p\mathscr{M}$, for any $p \in \mathscr{S}$, via

$$\forall (\boldsymbol{u},\boldsymbol{v}) \in T_p\mathscr{M}\times T_p\mathscr{M}, \quad \overrightarrow{\boldsymbol{q}}^*\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{q}\,(\boldsymbol{u},\boldsymbol{v}) = \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\,\boldsymbol{q}\left(\overrightarrow{\boldsymbol{q}}(\boldsymbol{u}), \overrightarrow{\boldsymbol{q}}(\boldsymbol{v})\right). \tag{2.85}$$

Since $\boldsymbol{q}$ is symmetric, it is clear from the above definition that $\boldsymbol{\Theta}$ is a field on $\mathscr{S}$ of symmetric bilinear forms.

Thanks to the identity (2.77), we can use for $\boldsymbol{q}$ the spacetime extension (2.43) and write the index-notation version of the definition (2.84) as

$$\Theta_{\alpha\beta} = \frac{1}{2} q^{\mu}{}_{\alpha} q^{\nu}{}_{\beta} \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, q_{\mu\nu}. \tag{2.86}$$

Expressing the Lie derivative in terms of the covariant derivative $\boldsymbol{\nabla}$ via Eq. (A.87) of Appendix A and using formula (2.43) for $\boldsymbol{q}$, we get

$$\begin{aligned}\Theta_{\alpha\beta} &= \frac{1}{2} q^\mu{}_\alpha q^\nu{}_\beta \left(\ell^\sigma \nabla_\sigma q_{\mu\nu} + q_{\sigma\nu}\nabla_\mu \ell^\sigma + q_{\mu\sigma}\nabla_\nu \ell^\sigma\right) \\ &= \frac{1}{2} q^\mu{}_\alpha q^\nu{}_\beta \Big[\ell^\sigma \left(\nabla_\sigma \ell_\mu k_\nu + \ell_\mu \nabla_\sigma k_\nu + \nabla_\sigma k_\mu \ell_\nu + k_\mu \nabla_\sigma \ell_\nu\right) \\ &\qquad + \nabla_\mu \ell_\nu + k_\nu \underbrace{\ell_\sigma \nabla_\mu \ell^\sigma}_{0} + k_\sigma \ell_\nu \nabla_\mu \ell^\sigma + \nabla_\nu \ell_\mu + \ell_\mu k_\sigma \nabla_\nu \ell^\sigma + k_\mu \underbrace{\ell_\sigma \nabla_\nu \ell^\sigma}_{0}\Big].\end{aligned}$$

Since $q^\mu{}_\alpha \ell_\mu = 0$ and $q^\mu{}_\alpha k_\mu = 0$, the above expression simplifies to

$$\Theta_{\alpha\beta} = q^\mu{}_\alpha q^\nu{}_\beta \nabla_\mu \ell_\nu. \tag{2.87}$$

Let us substitute (2.50) for the projector $\overrightarrow{\boldsymbol{q}}$:

$$\Theta_{\alpha\beta} = \left(\delta^\mu{}_\alpha + \ell^\mu k_\alpha + k^\mu \ell_\alpha\right)\left(\delta^\nu{}_\beta + \ell^\nu k_\beta + k^\nu \ell_\beta\right) \nabla_\mu \ell_\nu.$$

Expanding and simplifying (in particular via $\ell^\nu \nabla_\mu \ell_\nu = 0$) yields

$$\boxed{\nabla_\alpha \ell_\beta = \Theta_{\alpha\beta} + \omega_\alpha \ell_\beta - \ell_\alpha k^\mu \nabla_\mu \ell_\beta}, \tag{2.88}$$

where we have let appear the 1-form $\boldsymbol{\omega}$ defined by

$$\omega_\alpha := -k^\mu \nabla_\nu \ell_\mu \Pi^\nu{}_\alpha = -k^\mu \nabla_\alpha \ell_\mu - k^\mu k^\nu \nabla_\mu \ell_\nu \, \ell_\alpha, \tag{2.89}$$

where $\boldsymbol{\Pi}$ is the projector onto $\mathscr{H}$ along $\boldsymbol{k}$:

$$\Pi^\alpha{}_\beta = \delta^\alpha{}_\beta + k^\alpha \ell_\beta. \tag{2.90}$$

Indeed, for any vector $\boldsymbol{v}$ tangent to $\mathscr{H}$, one has $\boldsymbol{\Pi}(\boldsymbol{v}) = \boldsymbol{v}$ since $\Pi^\alpha{}_\mu v^\mu = v^\alpha + k^\alpha \ell_\mu v^\mu$ with $\ell_\mu v^\mu = 0$, while $\boldsymbol{\Pi}(\boldsymbol{k}) = 0$ since $\Pi^\alpha{}_\mu k^\mu = k^\alpha + k^\alpha \ell_\mu k^\mu$ with $\ell_\mu k^\mu = -1$. Consequently, the action of $\boldsymbol{\omega}$ on vectors tangent to $\mathscr{H}$ takes a simple form:

$$\forall \boldsymbol{v} \in T_p\mathscr{H}, \quad \langle \boldsymbol{\omega}, \boldsymbol{v}\rangle = -\boldsymbol{k} \cdot \boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{\ell}. \tag{2.91}$$

Remark 7: Thanks to the projector $\boldsymbol{\Pi}$ involved in its definition, the 1-form $\boldsymbol{\omega}$ does not depend on the extension of the vector field $\boldsymbol{\ell}$ away from $\mathscr{H}$. The same property holds for $\boldsymbol{\Theta}$. On the contrary, the tensor field $\boldsymbol{\nabla}\underline{\boldsymbol{\ell}}$, which appears in the left-hand side of formula (2.88) and in the last term of its right-hand side, depends on the extension of $\boldsymbol{\ell}$ away from $\mathscr{H}$.

By comparing (2.78) and (2.86), we notice that the trace of $\boldsymbol{\Theta}$ is nothing but the expansion $\theta_{(\boldsymbol{\ell})}$:

$$\boxed{\theta_{(\boldsymbol{\ell})} = g^{\mu\nu}\Theta_{\mu\nu} = q^{\mu\nu}\Theta_{\mu\nu} = \Theta^\mu{}_\mu}. \tag{2.92}$$

The trace-free part of $\boldsymbol{\Theta}$ is called the ***shear tensor*** of $\mathscr{S}$:

$$\boxed{\boldsymbol{\sigma} := \boldsymbol{\Theta} - \frac{1}{n-2}\theta_{(\boldsymbol{\ell})}\, \boldsymbol{q}}, \tag{2.93}$$

or, in index notation:

$$\sigma_{\alpha\beta} = \Theta_{\alpha\beta} - \frac{1}{n-2}\theta_{(\boldsymbol{\ell})}\, q_{\alpha\beta}. \tag{2.94}$$

Remark 8: The $1/(n-2)$ factor arises from the trace of $\boldsymbol{q}$, which is $n-2$. This follows immediately from $\boldsymbol{q}$ being a metric tensor on the $(n-2)$-dimensional manifold $\mathscr{S}$; this can also be recovered from the spacetime extension (2.43) of $\boldsymbol{q}$:

$$q^\mu{}_\mu = \underbrace{\delta^\mu{}_\mu}_{n} + 2\underbrace{\ell^\mu k^\mu}_{-1} = n-2.$$

By construction, we have thus

$$\sigma^\mu{}_\mu = g^{\mu\nu}\sigma_{\mu\nu} = q^{\mu\nu}\sigma_{\mu\nu} = 0. \tag{2.95}$$

Note that $\boldsymbol{\Theta}$ and $\boldsymbol{\sigma}$ are tensor fields tangent to $\mathscr{S}$, in the sense that

$$\boxed{\forall \boldsymbol{v} \in T_p^\perp\mathscr{S}, \quad \boldsymbol{\Theta}(\boldsymbol{v}, .) = \boldsymbol{\sigma}(\boldsymbol{v}, .) = 0}, \tag{2.96}$$

with the important special cases $\boldsymbol{v} = \boldsymbol{\ell}$ and $\boldsymbol{v} = \boldsymbol{k}$.

Example 21 (light cone): Let us consider the light cone in Minkowski spacetime described in terms of the spherical coordinates introduced in Example 15. Since the coordinates (t, θ, φ) are adapted to the vector field $\boldsymbol{\ell}$ (i.e. the θ and φ are constant along the field lines of $\boldsymbol{\ell}$ on $\mathscr{H}$ and $\boldsymbol{\ell} = \partial/\partial t$ in these coordinates, in other words, $\ell^\alpha = (1, 0, 0)$), we have [cf. formula (A.88) in Appendix A]

$$\mathcal{L}_{\boldsymbol{\ell}}\, q_{ab} = \frac{\partial}{\partial t} q_{ab} = \frac{\partial}{\partial r} q_{ab},$$

where the second equality follows from $t = r$ on $\mathscr{H}$. Given that $q_{ab} = \mathrm{diag}(r^2, r^2\sin^2\theta)$ [cf. Eq. (2.46)], we obtain

$$\mathcal{L}_{\boldsymbol{\ell}}\, q_{ab} = \begin{pmatrix} 2r & 0 \\ 0 & 2r\sin^2\theta \end{pmatrix} = \frac{2}{r}\, q_{ab}.$$

Hence (2.84) yields

$$\boldsymbol{\Theta} = \frac{1}{r}\, \boldsymbol{q}.$$

Taking the trace, we get immediately $\theta_{(\boldsymbol{\ell})} = 2/r$, i.e. we recover the result of Examples 17 and 19. From (2.93), we get a vanishing shear:

$$\boldsymbol{\sigma} = 0.$$

Example 22 (Schwarzschild horizon): The Lie derivative of $\boldsymbol{q}$, as given by Eq. (2.49), along $\boldsymbol{\ell}$ is (cf. Appendix D for the computation):

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, \boldsymbol{q} = 2r\frac{r-2m}{r+2m}\left(\mathbf{d}\theta \otimes \mathbf{d}\theta + \sin^2\theta\, \mathbf{d}\varphi \otimes \mathbf{d}\varphi\right) = \frac{2}{r}\frac{r-2m}{r+2m}\, \boldsymbol{q}.$$

Since $\overrightarrow{q}^* \boldsymbol{q} = \boldsymbol{q}$, Eq. (2.84) yields

$$\boldsymbol{\Theta} = \frac{r - 2m}{r(r+2m)}\, \boldsymbol{q}.$$

This formula is valid for any hypersurface of the $\mathscr{H}_u$ family. For the specific case of the Schwarzschild horizon $\mathscr{H}$, $r = 2m$ and it reduces to

$$\boldsymbol{\Theta} = 0. \tag{2.97}$$

2.4 Null Raychaudhuri equation

Let us derive an evolution equation for the expansion $\theta_{(\boldsymbol{\ell})}$; it is quite natural to consider the evolution along the null generators of $\mathscr{H}$, i.e. to evaluate the quantity $\boldsymbol{\nabla}_{\boldsymbol{\ell}} \theta_{(\boldsymbol{\ell})}$, all the more that $\boldsymbol{\ell}$ is by hypothesis future-directed. The starting point is the Ricci identity [Eq. (A.99) in Appendix A] applied to $\boldsymbol{\ell}$:

$$(\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha)\, \ell^\gamma = R^\gamma{}_{\mu\alpha\beta}\, \ell^\mu,$$

where $R^\gamma{}_{\mu\alpha\beta}$ is the Riemann tensor of the metric $\boldsymbol{g}$. Taking the trace on the indices (α, γ) and relabelling $\beta \to \alpha$ yields

$$\nabla_\mu \nabla_\alpha \ell^\mu - \nabla_\alpha \nabla_\mu \ell^\mu = R_{\mu\alpha} \ell^\mu,$$

where $R_{\mu\alpha} = R^\sigma{}_{\mu\sigma\alpha}$ is the Ricci tensor of $\boldsymbol{g}$. Substituting Eq. (2.88) for $\nabla_\alpha \ell^\mu$ and $\theta_{(\boldsymbol{\ell})} + \kappa$ for $\nabla_\mu \ell^\mu = \boldsymbol{\nabla} \cdot \boldsymbol{\ell}$ [cf. Eq. (2.81)] yields

$$\nabla_\mu \left(\Theta_\alpha{}^\mu + \omega_\alpha \ell^\mu - \ell_\alpha k^\nu \nabla_\nu \ell^\mu \right) - \nabla_\alpha \left(\theta_{(\boldsymbol{\ell})} + \kappa \right) = R_{\mu\alpha} \ell^\mu.$$

Expanding the left-hand side and using again Eqs. (2.81) and (2.88) results in

$$\begin{aligned} \nabla_\mu \Theta^\mu{}_\alpha + \ell^\mu \nabla_\mu \omega_\alpha - \nabla_\alpha \left(\theta_{(\boldsymbol{\ell})} + \kappa \right) + \left(\theta_{(\boldsymbol{\ell})} + \kappa \right) \omega_\alpha - \Theta_{\alpha\mu} k^\nu \nabla_\nu \ell^\mu & \\ - \left(\omega_\mu k^\nu \nabla_\nu \ell^\mu + \nabla_\mu k^\nu \nabla_\nu \ell^\mu + k^\nu \nabla_\mu \nabla_\nu \ell^\mu \right) \ell_\alpha & = R_{\mu\alpha} \ell^\mu. \end{aligned} \tag{2.98}$$

The above relation is a 1-form identity. Applying it to the vector field $\boldsymbol{\ell}$ (i.e. contracting with ℓ^α), we get, since $\ell_\nu \ell^\nu = 0$,

$$\ell^\nu \nabla_\mu \Theta^\mu{}_\nu + \ell^\nu \ell^\mu \nabla_\mu \omega_\nu - \ell^\mu \nabla_\mu \left(\theta_{(\boldsymbol{\ell})} + \kappa \right) + \left(\theta_{(\boldsymbol{\ell})} + \kappa \right) \omega_\mu \ell^\mu = R_{\mu\nu} \ell^\mu \ell^\nu. \tag{2.99}$$

Now, using $\Theta^\mu{}_\nu \ell^\nu = 0$ [Eq. (2.96)] and Eq. (2.88), we can write

$$\begin{aligned} \ell^\nu \nabla_\mu \Theta^\mu{}_\nu &= \nabla_\mu \underbrace{(\Theta^\mu{}_\nu \ell^\nu)}_{0} - \Theta^\mu{}_\nu \nabla_\mu \ell^\nu = -\Theta^{\mu\nu} \nabla_\mu \ell_\nu = -\Theta^{\mu\nu} \left(\Theta_{\mu\nu} + \omega_\mu \ell_\nu - \ell_\mu k^\sigma \nabla_\sigma \ell_\nu \right) \\ &= -\Theta^{\mu\nu} \Theta_{\mu\nu}. \end{aligned}$$

On the other side,

$$\ell^\nu \ell^\mu \nabla_\mu \omega_\nu = \ell^\mu \nabla_\mu (\omega_\nu \ell^\nu) - \omega_\nu \underbrace{\ell^\mu \nabla_\mu \ell^\nu}_{\kappa \ell^\nu} = \ell^\mu \nabla_\mu (\omega_\nu \ell^\nu) - \kappa \omega_\nu \ell^\nu,$$

Accordingly Eq. (2.99) becomes

$$-\Theta^{\mu\nu}\Theta_{\mu\nu} + \ell^\mu \nabla_\mu(\omega_\nu \ell^\nu) - \ell^\mu \nabla_\mu \left(\theta_{(\boldsymbol{\ell})} + \kappa\right) + \theta_{(\boldsymbol{\ell})}\omega_\mu \ell^\mu = R_{\mu\nu}\ell^\mu \ell^\nu .$$

The term $\omega_\mu \ell^\mu$, which appears twice in this equation, takes a simple form:

$$\omega_\mu \ell^\mu = \kappa. \tag{2.100}$$

Indeed, from the definition (2.89) of the 1-form $\boldsymbol{\omega}$,

$$\omega_\mu \ell^\mu = -k^\nu \underbrace{\ell^\mu \nabla_\mu \ell_\nu}_{\kappa \ell_\nu} - k^\rho k^\sigma \nabla_\rho \ell_\sigma \underbrace{\ell_\mu \ell^\mu}_{0} = -\kappa \underbrace{k^\nu \ell_\nu}_{-1} = \kappa. \tag{2.101}$$

Therefore, the two derivatives $\ell^\mu \nabla_\mu(\omega_\nu \ell^\nu)$ and $-\ell^\mu \nabla_\mu \kappa$ cancel out and one is left with

$$-\Theta^{\mu\nu}\Theta_{\mu\nu} - \ell^\mu \nabla_\mu \theta_{(\boldsymbol{\ell})} + \kappa \theta_{(\boldsymbol{\ell})} = R_{\mu\nu}\ell^\mu \ell^\nu . \tag{2.102}$$

The first term in the left-hand side can be re-expressed by the decomposition (2.93) of $\boldsymbol{\Theta}$ in terms of the shear tensor and the trace term:

$$\begin{aligned}
\Theta_{\mu\nu}\Theta^{\mu\nu} &= \left(\sigma_{\mu\nu} + \frac{1}{n-2}\,\theta_{(\boldsymbol{\ell})} q_{\mu\nu}\right)\left(\sigma^{\mu\nu} + \frac{1}{n-2}\,\theta_{(\boldsymbol{\ell})} q^{\mu\nu}\right) \\
&= \sigma_{\mu\nu}\sigma^{\mu\nu} + \frac{2}{n-2}\,\theta_{(\boldsymbol{\ell})} \underbrace{q^{\mu\nu}\sigma_{\mu\nu}}_{0} + \frac{1}{(n-2)^2}\,\theta_{(\boldsymbol{\ell})}^2 \underbrace{q_{\mu\nu}q^{\mu\nu}}_{n-2} \\
&= \sigma_{\mu\nu}\sigma^{\mu\nu} + \frac{1}{n-2}\,\theta_{(\boldsymbol{\ell})}^2, \\
&= \sigma_{ab}\sigma^{ab} + \frac{1}{n-2}\,\theta_{(\boldsymbol{\ell})}^2.
\end{aligned}$$

Hence Eq. (2.102) can be rewritten as

$$\boxed{\boldsymbol{\nabla}_{\boldsymbol{\ell}}\,\theta_{(\boldsymbol{\ell})} = \kappa \theta_{(\boldsymbol{\ell})} - \frac{1}{n-2}\,\theta_{(\boldsymbol{\ell})}^2 - \sigma_{ab}\sigma^{ab} - \boldsymbol{R}(\boldsymbol{\ell},\boldsymbol{\ell})}. \tag{2.103}$$

Since $\boldsymbol{\ell}$ is future-directed, this is an evolution equation for $\theta_{(\boldsymbol{\ell})}$. It is known as the ***null Raychaudhuri equation***.

Remark 1: Actually Eq. (2.103) is a particular case of what is generally called the null Raychaudhuri equation, namely the case where the vorticity of the vector field $\boldsymbol{\ell}$ vanishes. This appends because $\boldsymbol{\ell}$ is hypersurface-orthogonal, i.e. is normal to the hypersurface $\mathscr{H}$. The general case regards a generic ***congruence*** of null geodesics, i.e. a family of null geodesics, one, and exactly one, through each point of $\mathscr{M}$. A normal vector field $\boldsymbol{\ell}$ tangent to the geodesics of the congruence has a priori some vorticity $\boldsymbol{w}$ and a term $+w_{ab}w^{ab}$ must be added in the right-hand side of Eq. (2.103) (see e.g. Eq. (4.35) of [146]).

If the spacetime $(\mathscr{M}, \boldsymbol{g})$ is ruled by general relativity, i.e. if $\boldsymbol{g}$ obeys Einstein equation (1.39), we may express the term involving the Ricci tensor in terms of the total energy-momentum tensor $\boldsymbol{T}$:

$$\boldsymbol{R}(\boldsymbol{\ell},\boldsymbol{\ell}) = \frac{2}{n-2}\,\Lambda \underbrace{\boldsymbol{g}(\boldsymbol{\ell},\boldsymbol{\ell})}_{0} + 8\pi\left[\boldsymbol{T}(\boldsymbol{\ell},\boldsymbol{\ell}) - \frac{1}{n-2}\,T \underbrace{\boldsymbol{g}(\boldsymbol{\ell},\boldsymbol{\ell})}_{0}\right] = 8\pi \boldsymbol{T}(\boldsymbol{\ell},\boldsymbol{\ell}).$$

The null Raychaudhuri equation becomes then

$$\boxed{\boldsymbol{\nabla}_\ell\, \theta_{(\boldsymbol{\ell})} = \kappa\theta_{(\boldsymbol{\ell})} - \frac{1}{n-2}\,\theta_{(\boldsymbol{\ell})}^2 - \sigma_{ab}\sigma^{ab} - 8\pi \boldsymbol{T}(\boldsymbol{\ell},\boldsymbol{\ell})}. \tag{2.104}$$

Remark 2: Since $\theta_{(\boldsymbol{\ell})}$ is a scalar field on $\mathscr{H}$, $\boldsymbol{\nabla}_\ell\, \theta_{(\boldsymbol{\ell})}$ can be replaced by the Lie derivative $\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, \theta_{(\boldsymbol{\ell})}$ in the left-hand side of the Raychaudhuri equation.

Remark 3: The cosmological constant Λ does not appear in the null Raychaudhuri equation (2.104).

Example 23 (light cone): Let us check the null Raychaudhuri equation on the light cone in Minkowski spacetime. From Example 9, we have $\kappa = 0$, while from Example 21, we have $\boldsymbol{\sigma} = 0$, hence $\sigma_{ab}\sigma^{ab} = 0$. Moreover, the Ricci tensor of Minkowski spacetime vanishes identically. The null Raychaudhuri equation reduces then to

$$\boldsymbol{\nabla}_\ell\, \theta_{(\boldsymbol{\ell})} = -\frac{1}{2}\,\theta_{(\boldsymbol{\ell})}^2,$$

where we have set $n = 4$. Now, from Example 17, we have $\theta_{(\boldsymbol{\ell})} = 2/r$. Since, in the present case $\boldsymbol{\nabla}_\ell\, \theta_{(\boldsymbol{\ell})} = \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, \theta_{(\boldsymbol{\ell})} = \partial\theta_{(\boldsymbol{\ell})}/\partial r = -2/r^2$, we conclude that the null Raychaudhuri equation is satisfied (as it should!).

Example 24 (Schwarzschild horizon): For the Schwarzschild horizon $\mathscr{H}$, the null Raychaudhuri equation is trivially satisfied, i.e. each of its terms vanishes identically: $\theta_{(\boldsymbol{\ell})} = 0$ on $\mathscr{H}$ [Eq. (2.72)], which implies $\boldsymbol{\nabla}_\ell\, \theta_{(\boldsymbol{\ell})} = 0$ since $\boldsymbol{\ell}$ is tangent to $\mathscr{H}$, $\boldsymbol{\sigma} = 0$ since $\boldsymbol{\Theta} = 0$ [Eq. (2.97)] and the Ricci tensor of the metric (2.5) is zero (cf. Sec. D.3.1 in Appendix D).

Chapter 3

The concept of black hole 2: Non-expanding horizons and Killing horizons

Contents

3.1 Introduction

Having discussed in depth the geometry of null hypersurfaces in Chap. 2 we move forward to distinguish a null hypersurface representing a black hole event horizon from, let us say, that representing a mere future light cone. We do it here for black holes *in equilibrium*. Indeed, for such objects, it is quite natural to assume a vanishing expansion. This leads us to the concept of *non-expanding horizon* (Sec. 3.2). A special kind of these objects is that of *Killing horizons* (Sec. 3.3). Actually, we shall see in Chap. 5 that the event horizon of a black hole in equilibrium must be a Killing horizon.

3.2 Non-expanding horizons

3.2.1 Motivation and definition

In Chap. 2, the null hypersurfaces have been introduced as boundaries of black holes, from the "no-escape" aspect of the naive definition given in Sec. 2.2.1. To enforce the "localized" facet of the definition, we could demand that the cross-sections are closed

(compact without boundary[1]) and have a constant area, i.e. a vanishing expansion. Hence the definition:

A ***non-expanding horizon*** is a null hypersurface $\mathscr{H}$ having the topology (2.32):

$$\mathscr{H} \simeq \mathbb{R} \times \mathscr{S}, \tag{3.1}$$

where $\mathscr{S}$ is a closed manifold of dimension $n-2$, and such that the expansion of $\mathscr{H}$ along any null normal $\boldsymbol{\ell}$ vanishes identically:

$$\theta_{(\boldsymbol{\ell})} = 0. \tag{3.2}$$

Remark 1: Note that, given the scaling law (2.82), if $\theta_{(\boldsymbol{\ell})} = 0$ for some normal $\boldsymbol{\ell}$, then $\theta_{(\boldsymbol{\ell}')} = 0$ for any other normal $\boldsymbol{\ell}'$. Hence the definition of a non-expanding horizon does not depend on the choice of the null normal.

As we shall discuss in detail in Chap. 5, this definition captures only the event horizon of black holes in equilibrium. For a black hole out of equilibrium, one has generically $\theta_{(\boldsymbol{\ell})} > 0$.

Example 1 (Schwarzschild horizon): In view of Eq. (2.72), we may assert that the Schwarzschild horizon considered in Examples 3, 7, 10, 12, 16, 18, 20, 22 and 24 of Chap. 2 is a non-expanding horizon.

Example 2 (null hyperplane and light cone as counter-examples): The null hyperplane and light cone in Minkowski spacetime considered in the examples of Chap. 2 are excluded by the above definition, having non-compact cross-sections (null hyperplane) or nonzero expansion (light cone).

Historical note: The concept of non-expanding horizon has been introduced by Petr Hájiček in 1973 under the name of *totally geodesic null hypersurface* [138] or *perfect horizon* [139, 140]. The terminology *non-expanding horizon* is due to Abhay Ashtekar, Stephen Fairhurst and Badri Krishnan in 2000 [12] (see also [11]).

3.2.2 Invariance of the area

Given a cross-section $\mathscr{S}$ of $\mathscr{H}$, the area of $\mathscr{S}$, with respect to the spacetime metric $\boldsymbol{g}$, is [cf. Eqs. (2.60) and (2.66)]

$$A = \int_{\mathscr{S}} {}^{\mathscr{S}}\boldsymbol{\epsilon}(\mathrm{d}\boldsymbol{x}_{(2)}, \ldots, \mathrm{d}\boldsymbol{x}_{(n-1)}) = \int_{\mathscr{S}} \sqrt{q}\, \mathrm{d}x^2 \cdots \mathrm{d}x^{n-1}, \tag{3.3}$$

where $x^a = (x^2, \ldots, x^{n-1})$ is a coordinate system on $\mathscr{S}$ and q is the determinant with respect to these coordinates of the Riemannian metric $\boldsymbol{q}$ induced by $\boldsymbol{g}$ on $\mathscr{S}$.

A direct consequence of the definition of a non-expanding horizon is that A does not depend on the choice of the cross-section $\mathscr{S}$.

[1] Cf. the discussion in Sec. 2.3.4.

Proof. Let $\mathscr{S}'$ be a second cross-section of $\mathscr{H}$ and let $\boldsymbol{\ell}$ be a field of null normals of $\mathscr{H}$. Along the null geodesic generators of $\mathscr{H}$, we can always choose a parameter λ associated with $\boldsymbol{\ell}$ (i.e. such that $\boldsymbol{\ell} = \mathrm{d}/\mathrm{d}\lambda$ along a given null geodesic generator) such that $\lambda = 0$ on $\mathscr{S}$. By the very definition of a cross-section, any null geodesic generator $\mathscr{L}$ of $\mathscr{H}$ intersects $\mathscr{S}'$ at a single point. Let λ_0 be the value of λ at this point. We may then introduce a new parameter along $\mathscr{L}$ as follows:

$$\lambda' = \frac{\lambda}{\lambda_0}.$$

If we repeat this for all null geodesic generators of $\mathscr{H}$, we obtain a parametrization of all the null geodesic generators that satisfies $\lambda' = 0$ on $\mathscr{S}$ and $\lambda' = 1$ on $\mathscr{S}'$. Let $\boldsymbol{\ell}' = \mathrm{d}/\mathrm{d}\lambda'$ be the (null) tangent vector associated with λ'. We may then say that the cross-section $\mathscr{S}'$ is deduced from $\mathscr{S}$ by the Lie dragging of $\mathscr{S}$ along $\boldsymbol{\ell}'$ by a parameter $\delta\lambda' = 1$. More precisely, we may consider that $\mathscr{S}'$ is deduced from $\mathscr{S}$ by a continuous deformation, represented by a 1-parameter family $(\mathscr{S}_{\lambda'})$ of cross-sections such that $\mathscr{S}_0 = \mathscr{S}$ and $\mathscr{S}_1 = \mathscr{S}'$. Associated with this family is a real-valued function $\lambda' \mapsto A(\lambda')$ given the area of each element $\mathscr{S}_{\lambda'}$. By the very definition of the expansion along $\boldsymbol{\ell}'$ [Eq. (2.59)], we have then

$$\frac{\mathrm{d}A}{\mathrm{d}\lambda'} = \int_{\mathscr{S}_{\lambda'}} \theta_{(\boldsymbol{\ell}')}\, \delta A.$$

If $\mathscr{H}$ is a non-expanding horizon, then $\theta_{(\boldsymbol{\ell}')} = 0$ and it follows that $A(\lambda')$ is a constant function. Hence the area of $\mathscr{S}'$ is equal to that of $\mathscr{S}$. □

Given that the quantity A defined by (3.3) takes a unique value whatever the cross-section $\mathscr{S}$, we call it the ***area of the non-expanding horizon*** $\mathscr{H}$.

Example 3 (Schwarzschild horizon): The area of the Schwarzschild horizon is readily computed from the metric (2.35): $q_{ab}\mathrm{d}x^a\mathrm{d}x^b = 4m^2\left(\mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}^2\varphi\right)$; we get

$$A = 16\pi m^2.$$

3.2.3 Trapped surfaces

If there exists some natural concept of *outer*/*inner* for $\mathscr{H}$, for instance the outer region being the one that contains an asymptotically flat end, and if the transverse null normals $\boldsymbol{k}$ to cross-sections point to the inner region, then the property $\theta_{(\boldsymbol{\ell})} = 0$ means that any cross-section $\mathscr{S}$ of the non-expanding horizon $\mathscr{H}$ is a ***marginally outer trapped surface*** (often abridged as ***MOTS***). This definition is due to Hawking [145], an ***outer trapped surface*** would be one for which $\theta_{(\boldsymbol{\ell})} \leq 0$.

The MOTS definition is related to, but distinct from, the definition of a marginally trapped surface by Penrose [207]: a $(n-2)$-dimensional submanifold $\mathscr{S}$ of $\mathscr{M}$ is a ***trapped surface*** iff (i) $\mathscr{S}$ is closed (i.e. compact without boundary), (ii) $\mathscr{S}$ is spacelike and (iii)

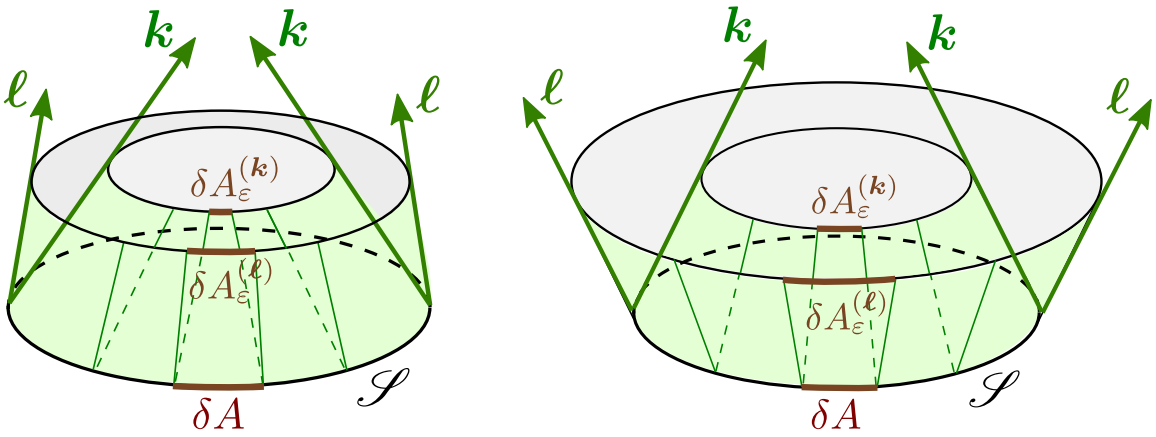


Figure 3.1: Trapped surface (left) and untrapped surface (right).

the two systems of null geodesics emerging orthogonally from $\mathscr{S}$ converge locally at $\mathscr{S}$, i.e. they have negative expansions:

$$\theta_{(\boldsymbol{\ell})} < 0 \quad \text{and} \quad \theta_{(\boldsymbol{k})} < 0, \tag{3.4}$$

where the expansion along $\boldsymbol{k}$ is defined in the same way as that along $\boldsymbol{\ell}$ [cf. Eq. (2.83)]:

$$\theta_{(\boldsymbol{k})} := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon} \frac{\delta A_\varepsilon^{(\boldsymbol{k})} - \delta A}{\delta A} = \frac{1}{2} \boldsymbol{\mathcal{L}}_{\boldsymbol{k}} \ln q = \frac{1}{2} q^{\mu\nu} \boldsymbol{\mathcal{L}}_{\boldsymbol{k}} q_{\mu\nu} = q^{\mu\nu} \nabla_\mu k_\nu, \tag{3.5}$$

$\delta A_\varepsilon^{(\boldsymbol{k})}$ begin the area of the surface element that is deduced from the surface element of area δA on $\mathscr{S}$ by the Lie dragging along $\boldsymbol{k}$ by a parameter ε (cf. Fig. 3.1 left). The limit case $\theta_{(\boldsymbol{\ell})} = 0$ and $\theta_{(\boldsymbol{k})} < 0$ correspond to the so-called ***marginally trapped surface***.

In flat spacetime (Minkowski), given any closed spacelike surface, one has $\theta_{(\boldsymbol{\ell})} > 0$ and $\theta_{(\boldsymbol{k})} < 0$ (cf. Fig. 3.1 right), so there is no trapped surface.

Remark 2: The hypothesis of *closed* surface is crucial in the definition of a trapped surface. For instance, there are non-closed spacelike surfaces in Minkowski spacetime having $\theta_{(\boldsymbol{\ell})} < 0$ and $\theta_{(\boldsymbol{k})} < 0$. A well known such example is the intersection of two past null cones (see e.g. Fig. 5.13 of [110]).

Cross-sections of a non-expanding horizon are usually marginally trapped surfaces (cf. the example below). However, there exist some pathological situations for which $\theta_{(\boldsymbol{k})} > 0$ at some points of $\mathscr{S}$ [116].

Example 4: Let us consider a cross-section $\mathscr{S}$ of the Schwarzschild horizon as defined in Example 12 of Chap. 2. Computing $q^{\mu\nu}\nabla_\mu k_\nu$ from the components k_ν given by (2.48) we get (cf. Sec. D.3.1 for the computation)

$$\theta_{(\boldsymbol{k})} = -\frac{r+2m}{r^2}.$$

In particular, on $\mathscr{S}$ ($r = 2m$),

$$\theta_{(\boldsymbol{k})} = -\frac{1}{m}.$$

Hence $\theta_{(\boldsymbol{k})} < 0$. Since we had already $\theta_{(\boldsymbol{\ell})} = 0$ [cf. Eq. (2.72)], we conclude that $\mathscr{S}$ is a marginally trapped surface. This could also have been inferred from Fig. 2.11, since according to the metric (2.49), the area of the cross-sections of $\mathscr{H}$ is nothing but $4\pi r^2$ and $\boldsymbol{k}$ points to decreasing values of r, while, on $\mathscr{H}$, $\boldsymbol{\ell}$ points to a fixed value of r.

3.2.4 Vanishing of the deformation rate tensor

If $\mathscr{H}$ is a non-expanding horizon, we may set $\theta_{(\boldsymbol{\ell})} = 0$ in the null Raychaudhuri equation (2.104); it reduces then to

$$\sigma_{ab}\sigma^{ab} + 8\pi \boldsymbol{T}(\boldsymbol{\ell}, \boldsymbol{\ell}) = 0. \tag{3.6}$$

Now, the first term is always non-negative:

$$\sigma_{ab}\sigma^{ab} \geq 0. \tag{3.7}$$

Proof. Since $\boldsymbol{\sigma}$ is symmetric, it can be diagonalized in an orthonormal basis of $\boldsymbol{q}$: $\sigma_{ab} = \mathrm{diag}(s_1, \ldots, s_{n-2})$. Moreover, $\boldsymbol{q}$ being a Riemannian metric, we have, in the same basis, $q^{ab} = \mathrm{diag}(1, \ldots, 1)$. Since $\sigma^{ab} = q^{am} q^{bn} \sigma_{mn}$, we conclude that

$$\sigma_{ab}\sigma^{ab} = s_1^2 + \cdots + s_{n-2}^2 \geq 0. \tag{3.8}$$

□

Regarding the second term in (3.6), it is quite natural to assume that matter and non-gravitational fields, represented by the total energy-momentum tensor $\boldsymbol{T}$, obey the ***null energy condition***, namely that

$$\boldsymbol{T}(\boldsymbol{\ell}, \boldsymbol{\ell}) \geq 0 \quad \text{for any null vector } \boldsymbol{\ell}. \tag{3.9}$$

This condition is pretty weak and is satisfied by

- vacuum: $\boldsymbol{T} = 0$;
- any "reasonable" matter model, such as a perfect fluid with a proper energy density ε and pressure p satisfying[2] $\varepsilon + p \geq 0$;
- any electromagnetic field;
- any real or complex scalar field;
- "dark energy" modelled by $\boldsymbol{T} = -\frac{\Lambda}{8\pi}\boldsymbol{g}$.

Note also that the null energy condition is implied by the so-called ***weak energy condition***, which states that

$$\boldsymbol{T}(\boldsymbol{u}, \boldsymbol{u}) \geq 0 \quad \text{for any timelike vector } \boldsymbol{u}. \tag{3.10}$$

[2] Indeed, from the form $\boldsymbol{T} = (\varepsilon + p)\underline{\boldsymbol{u}} \otimes \underline{\boldsymbol{u}} + p\boldsymbol{g}$ of the energy-momentum tensor of a perfect fluid, one has $\boldsymbol{T}(\boldsymbol{\ell}, \boldsymbol{\ell}) = (\varepsilon + p)(\boldsymbol{u} \cdot \boldsymbol{\ell})^2$ with $(\boldsymbol{u} \cdot \boldsymbol{\ell})^2 \geq 0$.

The null energy condition follows from the weak energy condition by continuity. Selecting for $\boldsymbol{u}$ the 4-velocity of an observer, we see that the weak energy condition has a simple physical interpretation: the energy density as measured by any observer is non-negative.

Given (3.7) and (3.9), Eq. (3.6) implies both

$$\sigma_{ab}\sigma^{ab} = 0 \tag{3.11}$$

and

$$\boldsymbol{T}(\boldsymbol{\ell},\boldsymbol{\ell}) = 0. \tag{3.12}$$

The identity $\sigma_{ab}\sigma^{ab} = 0$ is possible only if each of the s_i's in (3.8) is zero. Hence we have necessarily

$$\boldsymbol{\sigma} = 0. \tag{3.13}$$

Since we had already $\theta_{(\boldsymbol{\ell})} = 0$ (non-expanding horizon), this implies that the full deformation rate tensor vanishes identically [cf. Eq. (2.93)]:

$$\boldsymbol{\Theta} = 0. \tag{3.14}$$

In view of (2.84), this is equivalent to

$$\overrightarrow{\boldsymbol{q}}^* \boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, \boldsymbol{q} = 0. \tag{3.15}$$

We conclude that, provided that the null energy condition holds, the whole metric (and not only the area element ${}^{\mathscr{S}}\boldsymbol{\epsilon}$, as a mere $\theta_{(\boldsymbol{\ell})} = 0$ would suggest) of any cross-section of a non-expanding horizon is invariant along the null geodesic generators.

Example 5 (Schwarzschild horizon): We had already noticed that, for the Schwarzschild horizon, $\boldsymbol{\Theta} = 0$ [Eq. (2.97) in Example 22 of Chap. 2].

3.2.5 Induced affine connection

Since $\mathscr{H}$ is a null hypersurface, the "metric" $\boldsymbol{g}|_{\mathscr{H}}$ induced on it by the spacetime metric $\boldsymbol{g}$ is degenerate. As a consequence, there is a priori no unique connection on $\mathscr{H}$ associated with it. However, when $\mathscr{H}$ is a non-expanding horizon and the null energy condition holds on $\mathscr{H}$, so that $\boldsymbol{\Theta} = 0$, the spacetime connection $\boldsymbol{\nabla}$ induces a unique connection ${}^{\mathscr{H}}\boldsymbol{\nabla}$ on $\mathscr{H}$ as follows. Let $\boldsymbol{u}$ and $\boldsymbol{v}$ be two vector fields on $\mathscr{H}$. We have, using (2.88) to express $\nabla_\nu \ell_\mu$ in terms of $\Theta_{\nu\mu}$:

$$\begin{aligned} \ell_\mu u^\nu \nabla_\nu v^\mu &= u^\nu \nabla_\nu (\underbrace{\ell_\mu v^\mu}_{0}) - v^\mu u^\nu \nabla_\nu \ell_\mu \\ &= -\underbrace{\Theta_{\nu\mu}}_{0} v^\mu u^\nu - \omega_\nu u^\nu \underbrace{\ell_\mu v^\mu}_{0} + v^\mu \underbrace{u^\nu \ell_\nu}_{0} k^\sigma \nabla_\sigma \ell_\mu = 0. \end{aligned}$$

Hence $\boldsymbol{\ell}$ is orthogonal to the vector field $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}$. It follows immediately that $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}$ is tangent to $\mathscr{H}$. We conclude that the operator

$$\begin{aligned} {}^{\mathscr{H}}\boldsymbol{\nabla} \;:\; \mathfrak{X}(\mathscr{H}) \times \mathfrak{X}(\mathscr{H}) &\longrightarrow \mathfrak{X}(\mathscr{H}) \\ (\boldsymbol{u},\boldsymbol{v}) &\longmapsto \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}, \end{aligned} \tag{3.16}$$

where $\mathfrak{X}(\mathcal{H})$ is the space of vector fields on $\mathcal{H}$, is well-defined (i.e. ${}^{\mathcal{H}}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}$ does belong to $\mathfrak{X}(\mathcal{H})$). Moreover, this operator fulfills all the properties of an affine connection (cf. Sec. A.4.1), since $\boldsymbol{\nabla}$ does. We naturally call ${}^{\mathcal{H}}\boldsymbol{\nabla}$ the ***affine connection induced on*** $\mathcal{H}$ ***by*** $\boldsymbol{\nabla}$.

A geometrical consequence of the identity ${}^{\mathcal{H}}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v} = \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{v}$ is that $(\mathcal{H}, {}^{\mathcal{H}}\boldsymbol{\nabla})$ is a ***totally geodesic submanifold*** of $(\mathcal{M}, \boldsymbol{g})$: i.e. any geodesic of $(\mathcal{H}, {}^{\mathcal{H}}\boldsymbol{\nabla})$ is also a geodesic of $(\mathcal{M}, \boldsymbol{g})$ (cf. the historical note on page 58).

The covariant derivative of the null normal $\boldsymbol{\ell}$ with respect to the affine connection ${}^{\mathcal{H}}\boldsymbol{\nabla}$ takes a rather simple form:

$$\boxed{{}^{\mathcal{H}}\boldsymbol{\nabla}\boldsymbol{\ell} = \boldsymbol{\ell} \otimes {}^{\mathcal{H}}\boldsymbol{\omega}}, \tag{3.17}$$

where ${}^{\mathcal{H}}\boldsymbol{\omega}$ is the tensor field on $\mathcal{H}$ defined as the restriction of the 1-form $\boldsymbol{\omega}$ introduced in Sec. 2.3.6 to vectors tangent to $\mathcal{H}$ [cf. Eq. (2.91)]. In other words, ${}^{\mathcal{H}}\boldsymbol{\omega}$ is the pullback $\iota^*\boldsymbol{\omega}$ of $\boldsymbol{\omega}$ by the inclusion map $\iota : \mathcal{H} \to \mathcal{M}$ (cf. Sec. A.4.2 in Appendix A).

Proof. By definition of the covariant derivative with respect to an affine connection [cf. Eq. (A.60) in Appendix A], ${}^{\mathcal{H}}\boldsymbol{\nabla}\boldsymbol{\ell}$ is a tensor field of type $(1,1)$ on $\mathcal{H}$, the action of which on any pair $(\boldsymbol{a}, \boldsymbol{u})$ formed by a 1-form $\boldsymbol{a}$ and a vector field $\boldsymbol{u}$ on $\mathcal{H}$ is

$${}^{\mathcal{H}}\boldsymbol{\nabla}\boldsymbol{\ell}(\boldsymbol{a}, \boldsymbol{u}) = \langle \boldsymbol{a}, {}^{\mathcal{H}}\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\ell} \rangle.$$

In view of (3.16) and using Eq. (2.88) to express $\boldsymbol{\nabla}\boldsymbol{\ell}$, we get

$$\begin{aligned} {}^{\mathcal{H}}\boldsymbol{\nabla}\boldsymbol{\ell}(\boldsymbol{a}, \boldsymbol{u}) &= \langle \boldsymbol{a}, \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{\ell} \rangle = a_\mu u^\nu \nabla_\nu \ell^\mu \\ &= a_\mu u^\nu \Big(\underbrace{\Theta^\mu{}_\nu}_{0} + \omega_\nu \ell^\mu - \ell_\nu k^\rho \nabla_\rho \ell^\mu \Big) = a_\mu \ell^\mu \, \omega_\nu u^\nu - \underbrace{\ell_\nu u^\nu}_{0} a_\mu k^\rho \nabla_\rho \ell^\mu \\ &= \langle \boldsymbol{a}, \boldsymbol{\ell} \rangle \, \langle \boldsymbol{\omega}, \boldsymbol{u} \rangle. \end{aligned}$$

Given the definition of a tensor product and the fact that $\boldsymbol{u}$ is tangent to $\mathcal{H}$, this proves (3.17). □

A priori the 1-form ${}^{\mathcal{H}}\boldsymbol{\omega}$ depends upon the choice of the cross-section $\mathcal{S}$ of $\mathcal{H}$, via the vector $\boldsymbol{k}$ involved in Eq. (2.91): $\langle {}^{\mathcal{H}}\boldsymbol{\omega}, \boldsymbol{v} \rangle = -\boldsymbol{k} \cdot \boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{\ell}$ for any $\boldsymbol{v} \in T_p\mathcal{H}$. Formula (3.17) shows that for a non-expanding horizon, this is not the case: ${}^{\mathcal{H}}\boldsymbol{\omega}$ is a quantity intrinsic to $\mathcal{H}$ and to the value of $\boldsymbol{\ell}$ on $\mathcal{H}$. Moreover, under a change of null normal, $\boldsymbol{\ell} \mapsto \boldsymbol{\ell}' = \alpha\boldsymbol{\ell}$, it remains constant up to the addition of an exact 1-form: ${}^{\mathcal{H}}\boldsymbol{\omega}' = {}^{\mathcal{H}}\boldsymbol{\omega} + \mathbf{d}\ln\alpha$. The 1-form ${}^{\mathcal{H}}\boldsymbol{\omega}$ is usually called the ***rotation 1-form*** or ***connection 1-form*** of $(\mathcal{H}, \boldsymbol{\ell})$ [11, 125].

3.2.6 Going further

See Refs. [13, 125, 157] for more about non-expanding horizons, in particular for a subclass of them called *isolated horizons*.

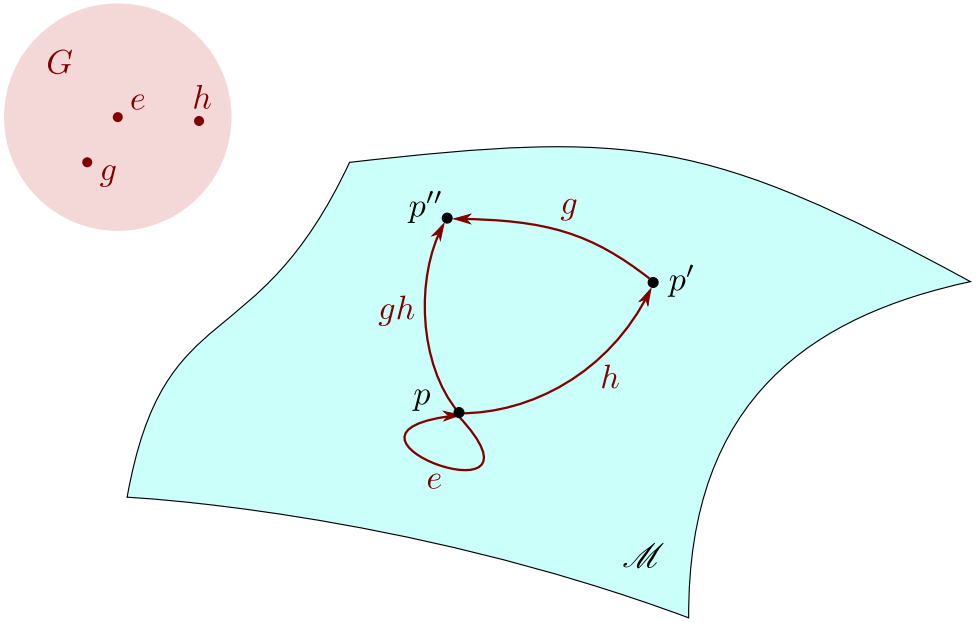


Figure 3.2: Group action of G on $\mathscr{M}$.

3.3 Killing horizons

A special kind of non-expanding horizons, which is of primordial importance for the theory of stationary black holes, is that of Killing horizons with closed-manifold cross-sections. Defining a Killing horizon requires the concepts of *1-dimensional group of isometries* and *Killing vector*, which we discuss first.

3.3.1 Spacetime symmetries

Symmetries of spacetime are described in a coordinate-independent way by means of a (symmetry) group acting on the spacetime manifold $\mathscr{M}$. Through this action, each transformation belonging to the group displaces points within $\mathscr{M}$ and one demands that the metric $\boldsymbol{g}$ is invariant under such displacement. More precisely, given a group G, a ***group action*** of G on $\mathscr{M}$ is a map[3]

$$\begin{array}{rccl} \Phi: & G\times\mathscr{M} & \longrightarrow & \mathscr{M} \\ & (g,p) & \longmapsto & \Phi(g,p) =: \Phi_g(p) \end{array} \tag{3.18}$$

such that (cf. Fig. 3.2)

- $\forall p \in \mathscr{M},\ \Phi_e(p) = p$, where e is the identity element of G;
- $\forall (g,h) \in G^2,\ \forall p \in \mathscr{M},\ \Phi_g(\Phi_h(p)) = \Phi_{gh}(p)$, where gh stands for the product of g by h according to G's group law.

[3] Do no confuse the generic element g of group G with the metric tensor $\boldsymbol{g}$.

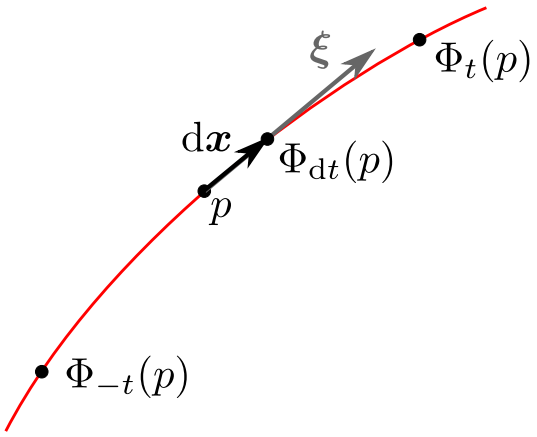


Figure 3.3: Orbit of a point p under the action Φ of a 1-dimensional Lie group, parameterized by $t \in \mathbb{R}$. The vector $\boldsymbol{\xi} = \mathrm{d}\boldsymbol{x}/\mathrm{d}t$ is the group generator associated with this parameter.

The ***orbit*** of a point $p \in \mathscr{M}$ is the set $\{g(p),\ g \in G\} \subset \mathscr{M}$, i.e. the set of points which are connected to p by some group transformation. One says that p is a ***fixed point*** of the group action if its orbit is reduced to $\{p\}$.

An important class of group actions are those for which G is a 1-dimensional *Lie group*, i.e. a "continuous" group (actually a "differentiable" group). Then around e, the elements of G can be labelled by a parameter $t \in \mathbb{R}$, such that $g_{t=0} = e$. It is then common to use the shorthand notation

$$\Phi_t := \Phi_{g_t}. \tag{3.19}$$

Because G is a 1-dimensional Lie group, the orbit of a given point $p \in \mathscr{M}$ under the group action is then either $\{p\}$ (when p is fixed point of the group action) or a curve of $\mathscr{M}$. In the latter case, t is a natural parameter along the curve (cf. Fig. 3.3). The tangent vector corresponding to that parameter is called the ***generator of the group*** G (associated with the t-parametrization). At each point p of an orbit, it is given by

$$\boldsymbol{\xi} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}t}, \tag{3.20}$$

where $\mathrm{d}\boldsymbol{x}$ is the infinitesimal vector connecting the point p to the point $\Phi_{\mathrm{d}t}(p)$ (cf. Sec. A.2.3 and Fig. 3.3). We have then

> The group action limited to infinitesimal transformations of parameter $\mathrm{d}t$ around the identity ($\mathrm{d}t = 0$) amounts to translations by the infinitesimal vector $\mathrm{d}t\,\boldsymbol{\xi}$.

A 1-dimensional Lie group G is said to be a ***symmetry group*** of the spacetime $(\mathscr{M}, \boldsymbol{g})$ if there is an action Φ of G on $\mathscr{M}$ such that for any value of the parameter t of G, Φ_t is an ***isometry*** of $(\mathscr{M}, \boldsymbol{g})$, i.e. Φ_t preserves the "distances" and more generally the "scalar products" on $(\mathscr{M}, \boldsymbol{g})$, in the following sense: for any $p \in \mathscr{M}$ and any pair of points (q, r) infinitely close to p, we shall have

$$\boldsymbol{g}|_{\Phi_t(p)}\,(\mathrm{d}\boldsymbol{x}', \mathrm{d}\boldsymbol{y}') = \boldsymbol{g}|_p\,(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{y}), \tag{3.21}$$

with the infinitesimal displacement vectors $\mathrm{d}\boldsymbol{x} := \overrightarrow{pq}$, $\mathrm{d}\boldsymbol{y} := \overrightarrow{pr}$, $\mathrm{d}\boldsymbol{x}' := \overrightarrow{\Phi_t(p)\Phi_t(q)}$ and $\mathrm{d}\boldsymbol{y}' := \overrightarrow{\Phi_t(p)\Phi_t(r)}$ (cf. Sec. 1.2). Now, by definition, $\mathrm{d}\boldsymbol{x}'$ is nothing but the pushforward of the vector $\mathrm{d}\boldsymbol{x} \in T_p\mathscr{M}$ to the tangent space $T_{\Phi_t(p)}\mathscr{M}$ by the map Φ_t (cf. Sec. A.4.2 of Appendix A), and similarly $\mathrm{d}\boldsymbol{y}'$ is the pushforward of $\mathrm{d}\boldsymbol{y}$ by Φ_t:

$$\mathrm{d}\boldsymbol{x}' = \Phi_t^*(\mathrm{d}\boldsymbol{x}) \quad \text{and} \quad \mathrm{d}\boldsymbol{y}' = \Phi_t^*(\mathrm{d}\boldsymbol{y}).$$

By rescaling by infinitely small parameters (using the bilinearity of $\boldsymbol{g}$), it is clear that (3.21) holds for finite vectors as well, so that we may say that Φ_t is an isometry of $(\mathscr{M}, \boldsymbol{g})$ iff

$$\forall p \in \mathscr{M},\ \forall (\boldsymbol{u}, \boldsymbol{v}) \in (T_p\mathscr{M})^2, \quad \boldsymbol{g}|_{\Phi_t(p)}\left(\Phi_t^*\boldsymbol{u}, \Phi_t^*\boldsymbol{v}\right) = \boldsymbol{g}|_p\left(\boldsymbol{u}, \boldsymbol{v}\right), \tag{3.22}$$

where $\Phi_t^*\boldsymbol{u}$ (resp. $\Phi_t^*\boldsymbol{v}$) is the pushforward of the vector $\boldsymbol{u} \in T_p\mathscr{M}$ (resp. $\boldsymbol{v} \in T_p\mathscr{M}$) to the tangent space $T_{\Phi_t(p)}\mathscr{M}$ by Φ_t [cf. Eq. (A.76)]. Given the definition (A.83) of the pullback of a bilinear form, we may reexpress the isometry condition (3.22) in terms of the pullback of $\boldsymbol{g}$ by Φ_t:

$$\Phi_t^*\boldsymbol{g} = \boldsymbol{g}. \tag{3.23}$$

According the definition (A.84) of the Lie derivative, we have

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\xi}}\, \boldsymbol{g} := \lim_{t\to 0} \frac{1}{t}\left(\Phi_t^*\boldsymbol{g} - \boldsymbol{g}\right). \tag{3.24}$$

If G is a symmetry group of $(\mathscr{M}, \boldsymbol{g})$ with generator $\boldsymbol{\xi}$, the isometry condition (3.23) leads then to $\boldsymbol{\mathcal{L}}_{\boldsymbol{\xi}}\, \boldsymbol{g} = 0$. The reverse is true by integration. Hence we conclude:

A 1-dimensional Lie group G is a symmetry group of the spacetime $(\mathscr{M}, \boldsymbol{g})$ iff the Lie derivative of the metric tensor along a generator $\boldsymbol{\xi}$ of G vanishes identically:

$$\boxed{\boldsymbol{\mathcal{L}}_{\boldsymbol{\xi}}\, \boldsymbol{g} = 0}. \tag{3.25}$$

The vector field $\boldsymbol{\xi}$ is then called a ***Killing vector*** of $(\mathscr{M}, \boldsymbol{g})$.

Expressing the Lie derivative via Eq. (A.87) of Appendix A, we see immediately that Eq. (3.25) is equivalent to the so-called ***Killing equation***:

$$\boxed{\nabla_\alpha \xi_\beta + \nabla_\beta \xi_\alpha = 0}. \tag{3.26}$$

If terms of the components $g_{\alpha\beta}$ of $\boldsymbol{g}$ with respect to coordinates $(x^\alpha) = (t, x^1, \dots, x^{n-1})$ adapted to the Killing vector $\boldsymbol{\xi}$, i.e. such that $\boldsymbol{\xi} = \boldsymbol{\partial}_t$, the isometry condition (3.25) is equivalent to

$$\frac{\partial g_{\alpha\beta}}{\partial t} = 0. \tag{3.27}$$

Proof. This is a direct consequence of the identity (A.88). □

3.3.2 Definition and examples of Killing horizons

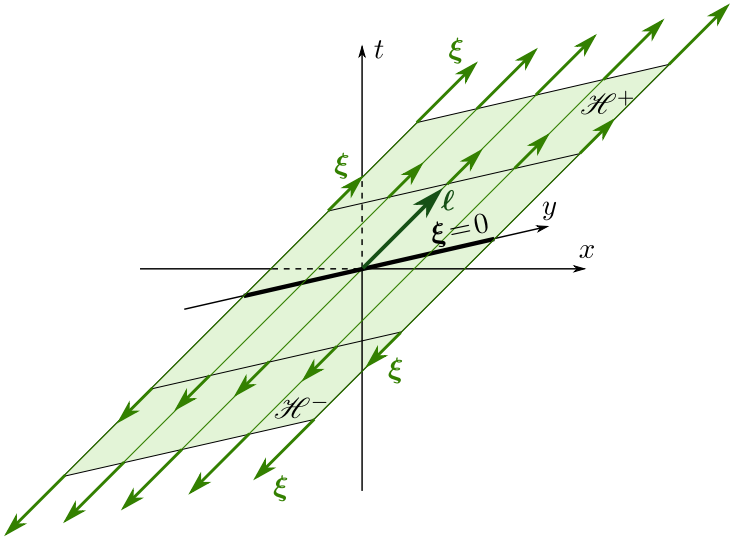


Figure 3.4: Null half-hyperplanes $\mathscr{H}^+$ and $\mathscr{H}^-$ as Killing horizons for the Killing vector field $\boldsymbol{\xi} = x\boldsymbol{\partial}_t + t\boldsymbol{\partial}_x$ generating Lorentz boosts in Minkowski spacetime. The green lines are the null geodesic generators of $\mathscr{H}$, while the thick black line (actually a 2-plane) marks the location where $\boldsymbol{\xi}$ vanishes.

A ***Killing horizon*** is a null hypersurface $\mathscr{H}$ in a spacetime $(\mathscr{M}, \boldsymbol{g})$ admitting a Killing vector field $\boldsymbol{\xi}$ such that, on $\mathscr{H}$, $\boldsymbol{\xi}$ is normal to $\mathscr{H}$.

Thus the existence of a Killing horizon requires that the spacetime $(\mathscr{M}, \boldsymbol{g})$ has some continuous symmetry (usually stationarity), namely that it is invariant under the action of a 1-parameter group, as described in Sec. 3.3.1. A definition equivalent to the above one is then:

A ***Killing horizon*** is a null hypersurface $\mathscr{H}$ whose null geodesic generators are orbits of a 1-parameter group of isometries of $(\mathscr{M}, \boldsymbol{g})$.

Remark 1: The above definition implies that the Killing vector field $\boldsymbol{\xi}$ is null and non-vanishing on $\mathscr{H}$:

$$\boldsymbol{\xi} \cdot \boldsymbol{\xi}|_{\mathscr{H}} = 0 \qquad \text{and} \qquad \boldsymbol{\xi}|_{\mathscr{H}} \neq 0. \tag{3.28}$$

Indeed, if $\boldsymbol{\xi}$ is vanishing at some point of $\mathscr{H}$, it cannot be considered as a normal vector to $\mathscr{H}$.

We shall see in Chap. 5 that in a stationary spacetime, a black hole event horizon must be a Killing horizon.

Example 6 (null hyperplane as a translation-Killing horizon): Let us consider the null hyperplane of Minkowski spacetime $\mathscr{H}$ discussed in Examples 1, 5 and 8 of Chap. 2. $\mathscr{H}$ is defined by the equation $t = x$. The vector field

$$\boldsymbol{\xi} := \boldsymbol{\partial}_t + \boldsymbol{\partial}_x \tag{3.29}$$

is a Killing vector of Minkowski spacetime: $\boldsymbol{\xi}$ is the generator of translations in the direction $\boldsymbol{\partial}_t + \boldsymbol{\partial}_x$, and these translations constitute a 1-dimensional subgroup of the Poincaré group —

the symmetry group of Minkowski spacetime. We note that $\boldsymbol{\xi}$ coincides with the null vector $\boldsymbol{\ell}$ defined by Eq. (2.12). Since $\boldsymbol{\ell}$ is normal to $\mathscr{H}$, we conclude immediately that $\mathscr{H}$ is a Killing horizon with respect to $\boldsymbol{\xi}$.

Example 7 (null hyperplane as a boost-Killing horizon): Let us consider the same null hyperplane $\mathscr{H}$ as above, but with another Killing vector of Minkowski spacetime:

$$\boldsymbol{\xi} := x\boldsymbol{\partial}_t + t\boldsymbol{\partial}_x. \tag{3.30}$$

This vector is indeed the generator of the 1-parameter group of Lorentz boosts in the (t, x) plane. On $\mathscr{H}$ we have (cf. Fig. 3.4):

$$\boldsymbol{\xi} \overset{\mathscr{H}}{=} t(\boldsymbol{\partial}_t + \boldsymbol{\partial}_x) \overset{\mathscr{H}}{=} t\,\boldsymbol{\ell},$$

where $\boldsymbol{\ell}$ is the null normal to $\mathscr{H}$ defined by Eq. (2.12) and the notation $\overset{\mathscr{H}}{=}$ means that the equality holds only on $\mathscr{H}$. We conclude that $\boldsymbol{\xi}$ is a normal to the null hypersurface $\mathscr{H}$ as soon as $t \neq 0$. Therefore, we may split $\mathscr{H} \setminus \{t = 0\}$ in two open half-hyperplanes:

$$\mathscr{H}^+ := \{p \in \mathscr{H}, \quad t(p) > 0\} \quad \text{and} \quad \mathscr{H}^- := \{p \in \mathscr{H}, \quad t(p) < 0\}, \tag{3.31}$$

so that each of them is a Killing horizon with respect to $\boldsymbol{\xi}$ (cf. Fig. 3.4).

Example 8 (null hyperplane as a null-rotation-Killing horizon): Another example of Killing horizon is still provided by the null hyperplane $\mathscr{H}$ considered above, but this time with the Killing vector

$$\boldsymbol{\xi} := y(\boldsymbol{\partial}_t + \boldsymbol{\partial}_x) + (t - x)\boldsymbol{\partial}_y. \tag{3.32}$$

This vector is indeed the generator of null rotations leaving the plane $\mathrm{Span}(\boldsymbol{\ell}, \boldsymbol{\partial}_z)$ strictly invariant (cf. e.g. Sec. 6.4.5 of Ref. [123]), $\boldsymbol{\ell}$ being the null normal of $\mathscr{H}$ defined by Eq. (2.12). These null rotations form a 1-dimensional subgroup of the Lorentz group, and thereby a symmetry group of Minkowski spacetime. It is also immediate to check that the vector defined by (3.32) obeys Killing equation (3.26). On $\mathscr{H}$, $t - x = 0$, so that (3.32) reduces to

$$\boldsymbol{\xi} \overset{\mathscr{H}}{=} y(\boldsymbol{\partial}_t + \boldsymbol{\partial}_x) \overset{\mathscr{H}}{=} y\,\boldsymbol{\ell}.$$

It follows that $\boldsymbol{\xi}$ is a null normal to $\mathscr{H}$ as soon as $y \neq 0$. We may then split $\mathscr{H} \setminus \{y = 0\}$ in two open half-hyperplanes:

$$\mathscr{H}_1 := \{p \in \mathscr{H}, \quad y(p) < 0\} \quad \text{and} \quad \mathscr{H}_2 := \{p \in \mathscr{H}, \quad y(p) > 0\},$$

each of them being a Killing horizon with respect to $\boldsymbol{\xi}$ (cf. Fig. 3.5).

Example 9 (light cone as a counter-example): The future light cone introduced in Example 2 of Chap. 2 is *not* a Killing horizon of Minkowski spacetime: it is invariant under the action of the Lorentz group, but its null generators are not invariant under the action of a single 1-dimensional subgroup of the Lorentz group. Actually the future light cone is an example of a more general structure, which Carter has termed a ***local isometry horizon*** [40, 43]: a null hypersurface that is invariant under some group G of isometries (here: the Lorentz group) and such that each null geodesic generator is an orbit of some 1-dimensional subgroup of G, this subgroup being not necessarily the same from one null generator to the next (here: using Minkowskian spherical coordinates (t, r, θ, φ), the null geodesic generator through the point of

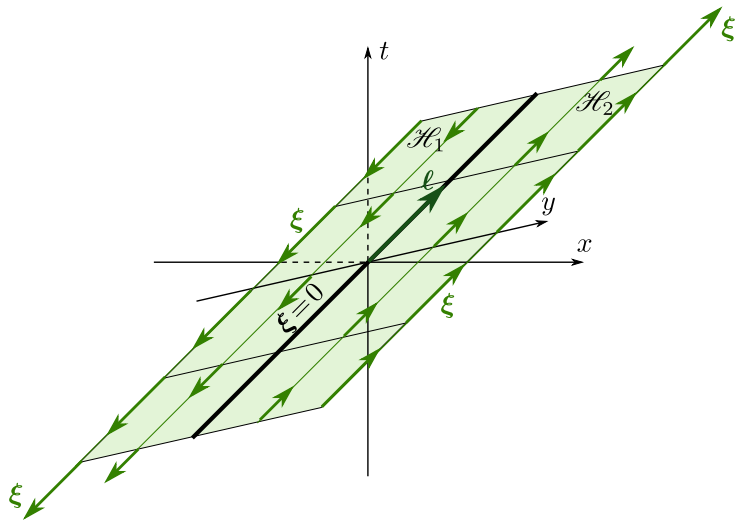


Figure 3.5: Null half-hyperplanes $\mathscr{H}_1$ and $\mathscr{H}_2$ as Killing horizons for the Killing vector field $\boldsymbol{\xi} = y(\boldsymbol{\partial}_t + \boldsymbol{\partial}_x) + (t - x)\boldsymbol{\partial}_y$ generating null rotations in Minkowski spacetime. The green lines are the null geodesic generators of $\mathscr{H}$, while the thick black line (actually a 2-plane) marks the location where $\boldsymbol{\xi}$ vanishes.

coordinates $(1, 1, \theta_0, \varphi_0)$ is the orbit of this point under the subgroup of boosts in the plane $(\theta, \varphi) = (\theta_0, \varphi_0)$). A Killing horizon is a local isometry horizon for which $\dim G = 1$.

Example 10 (Schwarzschild horizon): Given the expression (2.15) for the null normal $\boldsymbol{\ell}$ of the family of hypersurfaces $\mathscr{H}_u$ and the fact that the Schwarzschild horizon $\mathscr{H}$ is defined by $r = 2m$, we have

$$\boldsymbol{\ell} \overset{\mathscr{H}}{=} \boldsymbol{\partial}_t. \tag{3.33}$$

Now the vector field $\boldsymbol{\partial}_t$ is clearly a Killing vector of metric $\boldsymbol{g}$ as given by (2.5), since none of the metric components $g_{\alpha\beta}$ depends upon t. Hence (3.33) shows that the Schwarzschild horizon is a Killing horizon. By the way, Eq. (3.33) was our motivation for the choice of the null normal $\boldsymbol{\ell}$ performed in Example 7 of Chap. 2.

Historical note: The concept of Killing horizon has been introduced by Brandon Carter in 1966 [39, 40] and developed in an article published in 1969 [43]. The properties of Killing horizons have been studied in detail by Robert H. Boyer, in an article prepared posthumously from his notes by J. Ehlers and J.L. Stachel and published in 1969 [29], leading to the concept of *bifurcate Killing horizon*, to be discussed in Sec. 5.4 (cf. the historical note on page 117).

3.3.3 Killing horizons as non-expanding horizons

Let $\mathscr{H}$ be a Killing horizon with cross-sections that are closed manifolds, i.e. the topology of $\mathscr{H}$ is (2.32). Let us select the null normal $\boldsymbol{\ell}$ that coincides with the Killing vector $\boldsymbol{\xi}$ on $\mathscr{H}$: $\boldsymbol{\ell} \overset{\mathscr{H}}{=} \boldsymbol{\xi}$. Equation (3.25) then implies:

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, \boldsymbol{g} \overset{\mathscr{H}}{=} 0.$$

Let $\mathscr{S}$ be a cross-section of $\mathscr{H}$; since $\boldsymbol{q}$ is the metric induced by $\boldsymbol{g}$ on $\mathscr{S}$, we deduce immediately that

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, \boldsymbol{q} = 0.$$

From the definition (2.84), it follows that the expansion rate tensor of $\mathscr{S}$ vanishes identically:

$$\boxed{\boldsymbol{\Theta} = 0}. \tag{3.34}$$

In particular we have

$$\theta_{(\boldsymbol{\ell})} = 0.$$

We conclude that

Any Killing horizon with closed-manifold cross-sections is a non-expanding horizon.

Moreover, (3.34) shows that $\boldsymbol{\Theta}$ vanishes for all Killing horizons, while to get the same result on a generic non-expanding horizon, one has to assume that the null energy condition holds on $\mathscr{H}$.

3.3.4 Expressions of the non-affinity coefficient

Let κ be the non-affinity coefficient (cf. Sec. 2.3.3 and B.2.2) of the null normal $\boldsymbol{\ell}$ coinciding with the Killing vector $\boldsymbol{\xi}$ on a Killing horizon $\mathscr{H}$. According to the definition (2.23), we have

$$\boldsymbol{\nabla}_{\boldsymbol{\xi}}\, \boldsymbol{\xi} \overset{\mathscr{H}}{=} \kappa\, \boldsymbol{\xi}. \tag{3.35}$$

The metric dual of this relation is $\xi^\mu \nabla_\mu \xi_\alpha \overset{\mathscr{H}}{=} \kappa\, \xi_\alpha$. Using Killing equation (3.26) under the form $\nabla_\mu \xi_\alpha = -\nabla_\alpha \xi_\mu$, we get

$$\xi^\mu \nabla_\alpha \xi_\mu \overset{\mathscr{H}}{=} -\kappa\, \xi_\alpha.$$

Now $\xi^\mu \nabla_\alpha \xi_\mu = 1/2\, \nabla_\alpha (\xi_\mu \xi^\mu)$. Hence

$$\nabla_\alpha (\xi_\mu \xi^\mu) \overset{\mathscr{H}}{=} -2\kappa\, \xi_\alpha. \tag{3.36}$$

Since $\xi_\mu \xi^\mu = \boldsymbol{\xi} \cdot \boldsymbol{\xi}$ is a scalar field, we may replace the covariant derivative by the differential:

$$\boxed{\mathbf{d}(\boldsymbol{\xi} \cdot \boldsymbol{\xi}) \overset{\mathscr{H}}{=} -2\kappa\, \underline{\boldsymbol{\xi}}}. \tag{3.37}$$

Another interesting relation is obtained from the Frobenius theorem applied to $\boldsymbol{\xi}$. Indeed, since on $\mathscr{H}$, $\boldsymbol{\xi}$ is normal to a hypersurface ($\mathscr{H}$), the Frobenius theorem in its dual formulation (see e.g. Theorem B.3.2 in Wald's textbook [256] or Theorem C.2 in Straumann's textbook [237]) states that there exists a 1-form $\boldsymbol{a}$ such that

$$\mathbf{d}\underline{\boldsymbol{\xi}} \overset{\mathscr{H}}{=} \boldsymbol{a} \wedge \underline{\boldsymbol{\xi}}, \tag{3.38}$$

or equivalently

$$\nabla_\alpha \xi_\beta - \nabla_\beta \xi_\alpha \overset{\mathscr{H}}{=} a_\alpha\, \xi_\beta - a_\beta\, \xi_\alpha. \tag{3.39}$$

Remark 2: In the case of the vector $\boldsymbol{\ell}$, which is normal to $\mathscr{H}$ by definition, the Frobenius identity is Eq. (2.17): $\nabla_\alpha \ell_\beta - \nabla_\beta \ell_\alpha = \nabla_\alpha \rho\, \ell_\beta - \nabla_\beta \rho\, \ell_\alpha$. Since $\boldsymbol{\ell} \overset{\mathscr{H}}{=} \boldsymbol{\xi}$, we may write

$$\nabla_\alpha \ell_\beta - \nabla_\beta \ell_\alpha \overset{\mathscr{H}}{=} \nabla_\alpha \rho\, \xi_\beta - \nabla_\beta \rho\, \xi_\alpha .$$

But in general, $\nabla_\alpha \ell_\beta \neq \nabla_\alpha \xi_\beta$ on $\mathscr{H}$, since $\boldsymbol{\ell}$ and $\boldsymbol{\xi}$ do not coincide outside $\mathscr{H}$. Accordingly, one cannot identify the left-hand side of the above equation with the left-hand side of Eq. (3.39), so that the 1-form $\boldsymbol{a}$ is *not* $\boldsymbol{\nabla}\rho$.

Thanks to the Killing equation (3.26), we may reshape (3.39) to

$$2\nabla_\alpha \xi_\beta \overset{\mathscr{H}}{=} a_\alpha\, \xi_\beta - a_\beta\, \xi_\alpha . \tag{3.40}$$

Contracting this relation with $\boldsymbol{\xi}$, we get

$$2\xi^\mu \nabla_\mu \xi_\alpha \overset{\mathscr{H}}{=} a_\mu \xi^\mu\, \xi_\alpha - \underbrace{\xi_\mu \xi^\mu}_{\overset{\mathscr{H}}{=} 0}\, a_\alpha .$$

In view of Eq. (3.35), the left-hand side of this equation is $2\kappa\xi_\alpha$. Hence we obtain

$$a_\mu \xi^\mu \overset{\mathscr{H}}{=} 2\kappa . \tag{3.41}$$

Besides, taking the square of (3.40) leads to

$$\begin{aligned} 4\nabla_\mu \xi_\nu\, \nabla^\mu \xi^\nu &\overset{\mathscr{H}}{=} (a_\mu \xi_\nu - a_\nu \xi_\mu)\,(a^\mu \xi^\nu - a^\nu \xi^\mu) \\ &\overset{\mathscr{H}}{=} a_\mu a^\mu \underbrace{\xi_\nu \xi^\nu}_{\overset{\mathscr{H}}{=} 0} - \underbrace{a_\mu \xi^\mu}_{2\kappa}\underbrace{a_\nu \xi^\nu}_{2\kappa} - \underbrace{a_\nu \xi^\nu}_{2\kappa}\underbrace{a_\mu \xi^\mu}_{2\kappa} + a_\nu a^\nu \underbrace{\xi_\mu \xi^\mu}_{\overset{\mathscr{H}}{=} 0} \\ &\overset{\mathscr{H}}{=} -8\kappa^2 , \end{aligned}$$

where we have used Eq. (3.41). Hence

$$\boxed{\kappa^2 \overset{\mathscr{H}}{=} -\frac{1}{2} \nabla_\mu \xi_\nu \nabla^\mu \xi^\nu} . \tag{3.42}$$

This is an explicit expression of κ in terms of the Killing vector field $\boldsymbol{\xi}$. However, in actual calculations, it is generally preferable to employ (3.37) for evaluating κ, because the latter does not involve the computation of any covariant derivative, contrary to (3.42).

3.3.5 The zeroth law of black hole mechanics

We are going to derive a result of great importance for black hole physics, namely the non-affinity coefficient κ discussed above is constant on a Killing horizon, provided some mild energy condition holds.

Let us denote by $\boldsymbol{\ell}$ the null normal to $\mathscr{H}$ that coincides with the Killing vector field: $\boldsymbol{\ell} \overset{\mathscr{H}}{=} \boldsymbol{\xi}$. The vector field $\boldsymbol{\ell}$ is then a symmetry generator on $\mathscr{H}$, which implies

$$\boldsymbol{\mathcal{L}}_{\boldsymbol{\ell}}\, \kappa = 0 . \tag{3.43}$$

This means that κ is constant along the field lines of $\boldsymbol{\ell}$ (i.e. the null geodesic generators of $\mathscr{H}$). It could however vary from a field line to another one. To show that this is not the case, let us consider a cross-section $\mathscr{S}$ of $\mathscr{H}$ and project the contracted Ricci identity (2.98) onto it, via the orthogonal projector $\overrightarrow{\boldsymbol{q}}$ introduced in Sec. 2.3.4:

$$\nabla_\mu \Theta^\mu{}_\nu q^\nu{}_\alpha + \ell^\mu \nabla_\mu \omega_\nu q^\nu{}_\alpha - \nabla_\nu \left(\theta_{(\boldsymbol{\ell})} + \kappa\right) q^\nu{}_\alpha + \left(\theta_{(\boldsymbol{\ell})} + \kappa\right) \omega_\nu q^\nu{}_\alpha - \Theta_{\alpha\mu} k^\nu \nabla_\nu \ell^\mu = R_{\mu\nu} \ell^\mu q^\nu{}_\alpha,$$

where we have used $\Theta_{\nu\mu} q^\nu{}_\alpha = \Theta_{\alpha\mu}$ and $\ell_\nu q^\nu{}_\alpha = 0$. Now, since $\mathscr{H}$ is a Killing horizon, we have $\boldsymbol{\Theta} = 0$ [cf. Eq. (3.34)] and in particular $\theta_{(\boldsymbol{\ell})} = 0$. Accordingly, the above equation reduces to

$$\ell^\mu \nabla_\mu \omega_\nu q^\nu{}_\alpha - \nabla_\nu \kappa \, q^\nu{}_\alpha + \kappa \, \omega_\nu q^\nu{}_\alpha = R_{\mu\nu} \ell^\mu q^\nu{}_\alpha. \tag{3.44}$$

Let us express $\ell^\mu \nabla_\mu \omega_\nu$ in terms of the Lie derivative of $\boldsymbol{\omega}$ along $\boldsymbol{\ell}$ via formula (A.87) of Appendix A:

$$\ell^\mu \nabla_\mu \omega_\nu q^\nu{}_\alpha = \left(\mathcal{L}_{\boldsymbol{\ell}} \, \omega_\nu - \omega_\mu \nabla_\nu \ell^\mu\right) q^\nu{}_\alpha. \tag{3.45}$$

Now, since $\boldsymbol{\ell}$ is a symmetry generator on $\mathscr{H}$, we have

$$\mathcal{L}_{\boldsymbol{\ell}} \, {}^{\mathscr{H}}\boldsymbol{\omega} = 0, \tag{3.46}$$

where ${}^{\mathscr{H}}\boldsymbol{\omega}$ is the restriction of $\boldsymbol{\omega}$ to vectors tangent to $\mathscr{H}$, which has been shown to be a geometric quantity intrinsic to $\mathscr{H}$ and $\boldsymbol{\ell}$ in Sec. 3.2.5 and therefore has to obey the spacetime symmetry generated by $\boldsymbol{\xi} \overset{\mathscr{H}}{=} \boldsymbol{\ell}$. Any vector field $\boldsymbol{v}$ defined on $\mathscr{H}$ but not necessarily tangent to $\mathscr{H}$ can be split in a unique way into a part tangent to $\mathscr{H}$, $\boldsymbol{v}^\parallel$ say, and a part along the transverse null vector field $\boldsymbol{k}$ according to[4]

$$\boldsymbol{v} = \boldsymbol{v}^\parallel - \langle \underline{\boldsymbol{\ell}}, \boldsymbol{v} \rangle \boldsymbol{k},$$

so that

$$\langle \mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{\omega}, \boldsymbol{v} \rangle = \langle \mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{\omega}, \boldsymbol{v}^\parallel \rangle - \langle \underline{\boldsymbol{\ell}}, \boldsymbol{v} \rangle \langle \mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{\omega}, \boldsymbol{k} \rangle.$$

The first term in the right-hand side vanishes as a consequence of (3.46); indeed, the Leibniz rule allows us to write

$$\langle \mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{\omega}, \boldsymbol{v}^\parallel \rangle = \mathcal{L}_{\boldsymbol{\ell}} \, \langle \boldsymbol{\omega}, \boldsymbol{v}^\parallel \rangle - \langle \boldsymbol{\omega}, \mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{v}^\parallel \rangle = \mathcal{L}_{\boldsymbol{\ell}} \, \langle {}^{\mathscr{H}}\boldsymbol{\omega}, \boldsymbol{v}^\parallel \rangle - \langle {}^{\mathscr{H}}\boldsymbol{\omega}, \mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{v}^\parallel \rangle = \langle \underbrace{\mathcal{L}_{\boldsymbol{\ell}} \, {}^{\mathscr{H}}\boldsymbol{\omega}}_{0}, \boldsymbol{v}^\parallel \rangle = 0,$$

where the second equality follows from the fact that $\mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{v}^\parallel$ is a vector field tangent to $\mathscr{H}$ by the very definition of a Lie derivative (cf. Sec. A.4.2), $\boldsymbol{\ell}$ being tangent to $\mathscr{H}$. Hence the symmetry property (3.46) translates to the Lie derivative of $\boldsymbol{\omega}$ along $\boldsymbol{\ell}$ being proportional to the 1-form $\underline{\boldsymbol{\ell}}$:

$$\mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{\omega} = -\langle \mathcal{L}_{\boldsymbol{\ell}} \, \boldsymbol{\omega}, \boldsymbol{k} \rangle \, \underline{\boldsymbol{\ell}}. \tag{3.47}$$

Using this formula into Eq. (3.45) leads to

$$\ell^\mu \nabla_\mu \omega_\nu q^\nu{}_\alpha = -k^\mu \mathcal{L}_{\boldsymbol{\ell}} \, \omega_\mu \underbrace{\ell_\nu q^\nu{}_\alpha}_{0} - \omega_\mu \nabla_\nu \ell^\mu q^\nu{}_\alpha = -\omega_\mu \nabla_\nu \ell^\mu q^\nu{}_\alpha.$$

[4] $\boldsymbol{v}^\parallel = \boldsymbol{\Pi}(\boldsymbol{v})$, where $\boldsymbol{\Pi}$ is the the projector onto $\mathscr{H}$ along $\boldsymbol{k}$, as expressed by Eq. (2.90).

Accordingly, Eq. (3.44) becomes successively

$$\begin{aligned}
&-\omega_\mu \nabla_\nu \ell^\mu q^\nu{}_\alpha - \nabla_\nu \kappa\, q^\nu{}_\alpha + \kappa\, \omega_\nu q^\nu{}_\alpha = R_{\mu\nu}\ell^\mu q^\nu{}_\alpha \\
&-\omega_\mu \left(\Theta_\nu{}^\mu + \omega_\nu \ell^\mu - \ell_\nu k^\sigma \nabla_\sigma \ell^\mu\right) q^\nu{}_\alpha - \nabla_\nu \kappa\, q^\nu{}_\alpha + \kappa\, \omega_\nu q^\nu{}_\alpha = R_{\mu\nu}\ell^\mu q^\nu{}_\alpha \\
&-\omega_\mu \underbrace{\Theta_\alpha{}^\mu}_{0} - \underbrace{\omega_\mu \ell^\mu}_{\kappa} \omega_\nu q^\nu{}_\alpha - \nabla_\nu \kappa\, q^\nu{}_\alpha + \kappa\, \omega_\nu q^\nu{}_\alpha = R_{\mu\nu}\ell^\mu q^\nu{}_\alpha \\
&-\nabla_\nu \kappa\, q^\nu{}_\alpha = R_{\mu\nu}\ell^\mu q^\nu{}_\alpha,
\end{aligned}$$

where we have used (2.88) to get the second line, the identity $\ell_\nu q^\nu{}_\alpha = 0$ to get the third one and (2.101) to substitute κ for $\omega_\mu \ell^\mu$. In the above equation appears the covariant derivative of κ along $\mathscr{S}$, which we denote by ${}^{\mathscr{S}}\boldsymbol{D}$:

$${}^{\mathscr{S}}D_\alpha \kappa := \nabla_\nu \kappa\, q^\nu{}_\alpha. \tag{3.48}$$

Using the Einstein equation (1.39), we may then rewrite the above relation as

$${}^{\mathscr{S}}D_\alpha \kappa = -\frac{2}{n-2}\Lambda \underbrace{g_{\mu\nu}\ell^\mu q^\nu{}_\alpha}_{\ell^\mu q_{\mu\alpha}=0} - 8\pi\left(T_{\mu\nu}\ell^\mu q^\nu{}_\alpha - \frac{1}{n-2} T \underbrace{g_{\mu\nu}\ell^\mu q^\nu{}_\alpha}_{\ell^\mu q_{\mu\alpha}=0}\right),$$

i.e.

$${}^{\mathscr{S}}D_\alpha \kappa = -8\pi T_{\mu\nu}\ell^\mu q^\nu{}_\alpha. \tag{3.49}$$

To go further, we shall assume that matter and the non-gravitational fields obey the ***null dominant energy condition***:

$$\begin{aligned}
\boldsymbol{W} := -\overrightarrow{\boldsymbol{T}}(\boldsymbol{\ell}, .) \quad & \text{is future-directed null or timelike} \\
& \text{for any future-directed null vector } \boldsymbol{\ell}.
\end{aligned} \tag{3.50}$$

In the above equation, $\overrightarrow{\boldsymbol{T}}(\boldsymbol{\ell}, .)$ stands for the vector field that is the metric dual of the 1-form $\boldsymbol{T}(\boldsymbol{\ell}, .)$; in index notation,

$$W^\alpha = -g^{\alpha\nu} T_{\mu\nu} \ell^\mu = -T_\mu{}^\alpha \ell^\mu.$$

Note that the null dominant energy condition implies the null energy condition discussed in Sec. 3.2.4, since

$$\boldsymbol{T}(\boldsymbol{\ell}, \boldsymbol{\ell}) = -\boldsymbol{W} \cdot \boldsymbol{\ell} \geq 0,$$

the inequality holding because both $\boldsymbol{W}$ and $\boldsymbol{\ell}$ are future-directed.

The null dominant energy condition is implied by continuity by the ***dominant energy condition***:

$$\begin{aligned}
\boldsymbol{W} := -\overrightarrow{\boldsymbol{T}}(\boldsymbol{u}, .) \quad & \text{is future-directed null or timelike} \\
& \text{for any future-directed timelike vector } \boldsymbol{u}.
\end{aligned} \tag{3.51}$$

Physically, the dominant energy condition states that, with respect to any observer (represented by its 4-velocity $\boldsymbol{u}$, which is future-directed timelike), the energy of matter and non-gravitational fields, moves at a speed at most equal to c.

We note that in the right-hand side of (3.49), there appears the orthogonal projection of $\boldsymbol{W}$ onto $\mathscr{S}$ (more precisely its metric dual). If we assume the null dominant energy condition, the null energy condition holds and we have, according to (3.12),

$$\boldsymbol{\ell} \cdot \boldsymbol{W} = -\boldsymbol{T}(\boldsymbol{\ell}, \boldsymbol{\ell}) = 0,$$

This implies that the vector $\boldsymbol{W}$ is tangent to $\mathscr{H}$. The latter being a null hypersurface, $\boldsymbol{W}$ must then be either collinear to $\boldsymbol{\ell}$ or spacelike (cf. the lemma in Sec. 2.3.4). Now, according to the null dominant energy condition, $\boldsymbol{W}$ cannot be spacelike. We conclude that $\boldsymbol{W}$ is collinear to $\boldsymbol{\ell}$. Consequently its orthogonal projection onto $\mathscr{S}$ is zero:

$$q^{\alpha}{}_{\nu} W^{\nu} = -q^{\alpha}{}_{\nu} T_{\mu}{}^{\nu} \ell^{\mu} = 0.$$

Hence the right-hand side of (3.49) vanishes identically and we are left with

$${}^{\mathscr{S}}D_{\alpha}\kappa = 0.$$

This means that κ is constant over $\mathscr{S}$. Given that κ is constant along each null geodesic generator of $\mathscr{H}$, this completes the demonstration that κ is constant over $\mathscr{H}$. More precisely, we have shown that

If matter and non-gravitational fields obey the null dominant energy condition on the Killing horizon $\mathscr{H}$, then the non-affinity coefficient κ of the null normal coinciding with the Killing vector field on $\mathscr{H}$ is constant over $\mathscr{H}$:

$$\boxed{\kappa = \text{const.}} \tag{3.52}$$

In the context of Killing horizons, the non-affinity coefficient κ is called the horizon's ***surface gravity***, for a reason to be detailed in Sec. 3.3.7, and the result (3.52) is known as the ***zeroth law of black hole mechanics***. More precisely, the latter states that the surface gravity of a black hole in equilibrium is constant and we shall see in Chap. 5 that the event horizon of a black hole in equilibrium is a Killing horizon.

Example 11 (null hyperplane as a translation-Killing horizon): For the null hyperplane $\mathscr{H}$ considered in Example 6 as a Killing horizon with respect to the translation group along its normal, we have $\kappa = 0$, as already noticed in Example 8 of Chap. 2 [Eq. (2.28)], which is obviously constant over $\mathscr{H}$.

Example 12 (null hyperplane as a boost-Killing horizon): Let us consider each of the null half-hyperplanes $\mathscr{H}^+$ and $\mathscr{H}^-$ of Example 7, which are Killing horizons with respect to the boost Killing vector $\boldsymbol{\xi} = x\boldsymbol{\partial}_t + t\boldsymbol{\partial}_x$. On $\mathscr{H}^+$, the future-directed null normal coinciding with this Killing vector is $\boldsymbol{\ell}^+ = t\,\boldsymbol{\ell}$, $\boldsymbol{\ell}$ being the geodesic null normal defined by $\boldsymbol{\ell} := \boldsymbol{\partial}_t + \boldsymbol{\partial}_x$ [cf. Eq. (2.12)]. Using $\kappa_{\boldsymbol{\ell}} = 0$ and the scaling law (2.27), we get the non-affinity coefficient of $\boldsymbol{\ell}^+$ as $\kappa_+ = \boldsymbol{\nabla}_{\boldsymbol{\ell}} t = \partial_t t + \partial_x t$, i.e.

$$\kappa_+ = 1.$$

On $\mathscr{H}^-$, $\boldsymbol{\xi}$ is past-directed (cf. Fig. 3.4). Sticking to future-directed null normals, we shall then consider $\mathscr{H}^-$ as a Killing horizon with respect to the Killing vector field $-\boldsymbol{\xi}$. The future-directed

null normal coinciding with $-\boldsymbol{\xi}$ on $\mathscr{H}^-$ is then $\boldsymbol{\ell}^- = -t\,\boldsymbol{\ell}$, from which we deduce the non-affinity coefficient of $\boldsymbol{\ell}^-$: $\kappa_- = \boldsymbol{\nabla}_{\boldsymbol{\ell}}(-t) = \partial_t(-t) + \partial_x(-t)$, i.e.

$$\kappa_- = -1.$$

We check that κ_+ (resp. κ_-) is constant over the Killing horizon $\mathscr{H}^+$ (resp. $\mathscr{H}^-$), in agreement with the result above.

Example 13 (null hyperplane as a null-rotation-Killing horizon): In Example 8, we have introduced the Killing horizons $\mathscr{H}_1$ and $\mathscr{H}_2$ with respect to the null-rotation Killing vector $\boldsymbol{\xi} = y(\boldsymbol{\partial}_t + \boldsymbol{\partial}_x) + (t-x)\boldsymbol{\partial}_y$ of Minkowski spacetime. On $\mathscr{H}_1$, $\boldsymbol{\xi}$ is past-directed (cf. Fig. 3.5), so that we shall actually consider $\mathscr{H}_1$ as a Killing horizon with respect to the Killing vector field $-\boldsymbol{\xi}$. The future-directed null normal coinciding with $-\boldsymbol{\xi}$ on $\mathscr{H}_1$ is then $\boldsymbol{\ell}_1 = -y\,\boldsymbol{\ell}$. Since it is clearly constant along the null geodesic generators of $\mathscr{H}_1$, we have $\boldsymbol{\nabla}_{\boldsymbol{\ell}_1}\boldsymbol{\ell}_1 = 0$, hence the associated non-affinity coefficient vanishes:

$$\kappa_1 = 0.$$

On $\mathscr{H}_2$, $\boldsymbol{\xi}$ is future-directed (cf. Fig. 3.5) and the null normal coinciding with it is $\boldsymbol{\ell}_2 = y\,\boldsymbol{\ell}$, whose non-affinity coefficient is

$$\kappa_2 = 0.$$

Example 14 (Schwarzschild and Kerr horizons): We have found in Example 10 of Chap. 2 [cf. Eq. (2.31)] that on a Schwarzschild horizon:

$$\kappa = \frac{1}{4m},$$

which is clearly constant. But this last feature is rather trivial since the Schwarzschild horizon is spherically symmetric, so that no dependence of κ on θ nor φ could have been expected. A much less trivial example is that of the event horizon of a Kerr black hole, which we shall discuss in Chap. 10. This horizon is only axisymmetric, so that a priori κ could depend on θ. But it does not, as we shall see in Sec. 10.5.4:

$$\kappa = \frac{\sqrt{m^2 - a^2}}{2m(m + \sqrt{m^2 - a^2})},$$

where (m, a) are the two constant parameters of the Kerr solution. Note that for $a = 0$, we recover the Schwarzschild value: $\kappa = 1/(4m)$.

Historical note: The constancy of κ for a Killing horizon has been proven by Stephen Hawking in his lecture at the famous Les Houches School of Summer 1972 [145] (p. 43). It has also been proven without requiring the dominant energy condition, but assuming axisymmetry by Brandon Carter in his lecture at the same Les Houches School [46] (Theorem 8, p. 167). A third proof of the constancy of κ using the dominant energy condition has also been given in 1973 by James Bardeen, Brandon Carter and Stephen Hawking in their seminal article *The Four Laws of Black Hole Mechanics* [18].

3.3.6 Classification of Killing horizons

Since κ is constant on $\mathscr{H}$ (assuming the dominant energy condition), we may use it to classify Killing horizons in two categories, depending whether κ vanishes or not:

- if $\kappa = 0$, the Killing vector $\boldsymbol{\xi}$ is then a geodesic vector on $\mathscr{H}$ and $\mathscr{H}$ is called a ***degenerate Killing horizon***;
- if $\kappa \neq 0$, $\boldsymbol{\xi}$ is only a pregeodesic vector on $\mathscr{H}$ (cf. Sec. B.2.2) and $\mathscr{H}$ is called a ***non-degenerate Killing horizon***.

Example 15: In Minkowski spacetime, the null hyperplane as a translation-Killing horizon (Example 11) and the two half-hyperplanes as null-rotation-Killing horizons (Example 13) are degenerate Killing horizons, while the two half-hyperplanes as boost-Killing horizons (Example 12) are non-degenerate. From the values of κ given in Example 14, we see that the Schwarzschild horizon and the Kerr horizon for $a < m$ are non-degenerate Killing horizons, while the Kerr horizon for $a = m$ is a degenerate one.

3.3.7 Interpretation of κ as a "surface gravity"

In this section, we assume that $\mathscr{H}$ is a non-degenerate Killing horizon, i.e. that $\kappa \neq 0$. Let $p \in \mathscr{H}$ and $\boldsymbol{v} \in T_p\mathscr{M}$ be a vector *transverse* to $\mathscr{H}$, i.e. not tangent to $\mathscr{H}$. According to Eq. (3.37), we have

$$\boldsymbol{\nabla}_{\boldsymbol{v}}(\boldsymbol{\xi}\cdot\boldsymbol{\xi}) = -2\kappa\, \boldsymbol{\xi}\cdot\boldsymbol{v}.$$

The right-hand side of this expression does not vanish, because $\kappa \neq 0$ and $\boldsymbol{\xi}\cdot\boldsymbol{v} \neq 0$ (since $\boldsymbol{v}$ is not tangent to $\mathscr{H}$). Hence we have

$$\boldsymbol{\nabla}_{\boldsymbol{v}}(\boldsymbol{\xi}\cdot\boldsymbol{\xi}) \neq 0.$$

In other words, the derivative of the scalar square $\boldsymbol{\xi}\cdot\boldsymbol{\xi}$ along any direction transverse to $\mathscr{H}$ does not vanish. Since $\boldsymbol{\xi}\cdot\boldsymbol{\xi} = 0$ on $\mathscr{H}$, we conclude that, in the vicinity of $\mathscr{H}$, $\boldsymbol{\xi}\cdot\boldsymbol{\xi} < 0$ on one side of $\mathscr{H}$ and $\boldsymbol{\xi}\cdot\boldsymbol{\xi} > 0$ on the other side:

> In the vicinity of a non-degenerate Killing horizon $\mathscr{H}$, the Killing vector field $\boldsymbol{\xi}$ is timelike on one side of $\mathscr{H}$, null on $\mathscr{H}$ and spacelike on the other side.

Let us focus on the region in the vicinity of $\mathscr{H}$ where $\boldsymbol{\xi}$ is timelike. There we define the "norm" of $\boldsymbol{\xi}$ by

$$\boxed{V := \sqrt{-\boldsymbol{\xi}\cdot\boldsymbol{\xi}}}. \tag{3.53}$$

We have $V > 0$ and the square of the gradient of V provides a new expression of κ:

$$\boxed{\kappa^2 = \lim_{\mathscr{H}} \nabla_\mu V \nabla^\mu V}, \tag{3.54}$$

where $\lim_{\mathscr{H}}$ stands for the limit as one approaches $\mathscr{H}$ from the timelike side, which implies $V \to 0$.

Proof. Let us consider the 3-form $\boldsymbol{\omega}$ defined by

$$\begin{aligned}\omega_{\alpha\beta\gamma} &:= \xi_{[\alpha}\nabla_\beta\xi_{\gamma]} \\ &= \frac{1}{6}\left[\xi_\alpha\left(\nabla_\beta\xi_\gamma - \nabla_\gamma\xi_\beta\right) + \xi_\beta\left(\nabla_\gamma\xi_\alpha - \nabla_\alpha\xi_\gamma\right) + \xi_\gamma\left(\nabla_\alpha\xi_\beta - \nabla_\beta\xi_\alpha\right)\right], \end{aligned} \tag{3.55}$$

the second line being simply the explicit expression of the full antisymmetrization of $\xi_\alpha\nabla_\beta\xi_\gamma$, which is denoted by square brackets in the first line. Killing equation (3.26) enables us to simplify each term inside parentheses in (3.55), yielding

$$\omega_{\alpha\beta\gamma} = \frac{1}{3}\left(\xi_\alpha\nabla_\beta\xi_\gamma + \xi_\beta\nabla_\gamma\xi_\alpha + \xi_\gamma\nabla_\alpha\xi_\beta\right). \tag{3.56}$$

The "square" of $\boldsymbol{\omega}$ is then

$$\begin{aligned}\omega_{\mu\nu\rho}\,\omega^{\mu\nu\rho} &= \frac{1}{9}\Big(\xi_\mu\nabla_\nu\xi_\rho\,\xi^\mu\nabla^\nu\xi^\rho + \xi_\mu\nabla_\nu\xi_\rho\,\xi^\nu\nabla^\rho\xi^\mu + \xi_\mu\nabla_\nu\xi_\rho\,\xi^\rho\nabla^\mu\xi^\nu \\ &\quad + \xi_\nu\nabla_\rho\xi_\mu\,\xi^\mu\nabla^\nu\xi^\rho + \xi_\nu\nabla_\rho\xi_\mu\,\xi^\nu\nabla^\rho\xi^\mu + \xi_\nu\nabla_\rho\xi_\mu\,\xi^\rho\nabla^\mu\xi^\nu \\ &\quad + \xi_\rho\nabla_\mu\xi_\nu\,\xi^\mu\nabla^\nu\xi^\rho + \xi_\rho\nabla_\mu\xi_\nu\,\xi^\nu\nabla^\rho\xi^\mu + \xi_\rho\nabla_\mu\xi_\nu\,\xi^\rho\nabla^\mu\xi^\nu\Big).\end{aligned}$$

Now in the first line,

$$\xi_\mu\nabla_\nu\xi_\rho\,\xi^\mu\nabla^\nu\xi^\rho = \xi_\mu\xi^\mu\nabla_\nu\xi_\rho\nabla^\nu\xi^\rho = -V^2\nabla_\nu\xi_\rho\nabla^\nu\xi^\rho = -V^2\nabla_\mu\xi_\nu\nabla^\mu\xi^\nu \tag{3.57}$$

and (using Killing equation (3.26))

$$\xi_\mu\nabla_\nu\xi_\rho\,\xi^\nu\nabla^\rho\xi^\mu = \xi_\mu\nabla^\rho\xi^\mu\,\xi^\nu\nabla_\nu\xi_\rho = -\xi_\mu\nabla^\rho\xi^\mu\,\xi^\nu\nabla_\rho\xi_\nu = -\frac{1}{4}\nabla^\rho V^2\,\nabla_\rho V^2 = -V^2\nabla^\rho V\,\nabla_\rho V. \tag{3.58}$$

Actually, we notice that each line is made of one term of type (3.57) and two terms of type (3.58). Hence

$$\omega_{\mu\nu\rho}\,\omega^{\mu\nu\rho} = -\frac{V^2}{3}\left(\nabla_\mu\xi_\nu\nabla^\mu\xi^\nu + 2\nabla_\mu V\nabla^\mu V\right). \tag{3.59}$$

On $\mathscr{H}$, each of the terms inside parentheses in Eq. (3.55) can be expressed thanks to the Frobenius identity (3.39):

$$\omega_{\alpha\beta\gamma} \overset{\mathscr{H}}{=} \frac{1}{6}\left[\xi_\alpha\left(a_\beta\xi_\gamma - a_\gamma\xi_\beta\right) + \xi_\beta\left(a_\gamma\xi_\alpha - a_\alpha\xi_\gamma\right) + \xi_\gamma\left(a_\alpha\xi_\beta - a_\beta\xi_\alpha\right)\right].$$

We notice that all terms in the right-hand side canceal two by two, yielding

$$\omega_{\alpha\beta\gamma} \overset{\mathscr{H}}{=} 0. \tag{3.60}$$

Equation (3.60) is actually nothing but a variant of Frobenius theorem, expressing the fact that the vector field $\boldsymbol{\xi}$ is hypersurface-orthogonal on $\mathscr{H}$ (see e.g. Eq. (B.3.6) in Wald's

textbook [256]). Let us evaluate the gradient of the square (3.59) and take the limit on $\mathscr{H}$:

$$\nabla_\alpha\omega_{\mu\nu\rho}\underbrace{\omega^{\mu\nu\rho}}_{\to 0}+\underbrace{\omega_{\mu\nu\rho}}_{\to 0}\nabla_\alpha\omega^{\mu\nu\rho} = -\frac{1}{3}\underbrace{\nabla_\alpha V^2}_{\to 2\kappa\xi_\alpha}\Big(\underbrace{\nabla_\mu\xi_\nu\nabla^\mu\xi^\nu}_{\to -2\kappa^2}+2\nabla_\mu V\nabla^\mu V\Big) -\frac{1}{3}\underbrace{V^2}_{\to 0}\nabla_\alpha\left(\nabla_\mu\xi_\nu\nabla^\mu\xi^\nu+2\nabla_\mu V\nabla^\mu V\right),$$

where we have used Eq. (3.37) in the form $\nabla_\alpha V^2 \overset{\mathscr{H}}{=} 2\kappa\xi_\alpha$, as well as expression (3.42) of κ^2. Hence we are left with

$$\kappa\left(\nabla_\mu V\nabla^\mu V-\kappa^2\right)\xi_\alpha \longrightarrow 0 \quad \text{on } \mathscr{H}.$$

Now, by the very definition of a Killing horizon, $\xi_\alpha \neq 0$ on $\mathscr{H}$. Moreover, $\mathscr{H}$ being a non-degenerate Killing horizon, we have $\kappa \neq 0$ as well. The above limit is then equivalent to (3.54). □

In the region where $\boldsymbol{\xi}$ is timelike, the vector field

$$\boldsymbol{u} := \frac{1}{V}\boldsymbol{\xi} \tag{3.61}$$

is a future-directed unit timelike vector field. It is future-directed because by convention[5] $\boldsymbol{\xi}$ is future-directed null on $\mathscr{H}$ and by continuity this orientation must be preserved in the region where $\boldsymbol{\xi}$ is timelike. The unit vector field $\boldsymbol{u}$ can be then considered as the 4-velocity of an observer $\mathscr{O}$, whose worldline of is a field line of $\boldsymbol{\xi}$, i.e. an orbit of the isometry group generated by $\boldsymbol{\xi}$. One may call $\mathscr{O}$ a ***stationary observer*** since the spacetime geometry is not changing along its worldline. The 4-acceleration of $\mathscr{O}$ is

$$\begin{aligned}\boldsymbol{a} &:= \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u}\\ &= \boldsymbol{\nabla}_{V^{-1}\boldsymbol{\xi}}\left(V^{-1}\boldsymbol{\xi}\right) = V^{-1}\boldsymbol{\nabla}_{\boldsymbol{\xi}}\left(V^{-1}\boldsymbol{\xi}\right) = V^{-1}\left[-V^{-2}(\boldsymbol{\nabla}_{\boldsymbol{\xi}}V)\,\boldsymbol{\xi}+V^{-1}\boldsymbol{\nabla}_{\boldsymbol{\xi}}\boldsymbol{\xi}\right].\end{aligned}$$

Now, since $\boldsymbol{\xi}$ is a symmetry generator, $\boldsymbol{\nabla}_{\boldsymbol{\xi}}V = 0$. This can be shown explicitly by means of Killing equation (3.26):

$$\boldsymbol{\nabla}_{\boldsymbol{\xi}}V = \xi^\mu\nabla_\mu(\sqrt{-\xi_\nu\xi^\nu}) = -\frac{1}{2\sqrt{-\xi_\nu\xi^\nu}}\xi^\mu\nabla_\mu(\xi_\nu\xi^\nu) = -\frac{1}{V}\underbrace{\xi^\mu\xi^\nu\nabla_\mu\xi_\nu}_{0} = 0.$$

We have thus

$$\boldsymbol{a} = \frac{1}{V^2}\boldsymbol{\nabla}_{\boldsymbol{\xi}}\boldsymbol{\xi}, \tag{3.62}$$

from which

$$a_\alpha = \frac{1}{V^2}\xi^\mu\nabla_\mu\xi_\alpha.$$

[5] Were $\boldsymbol{\xi}$ past-directed, we could always consider the Killing field $-\boldsymbol{\xi}$ instead.

Thanks to Killing equation (3.26), we may rewrite this relation as

$$a_\alpha = -\frac{1}{V^2}\,\xi^\mu \nabla_\alpha \xi_\mu = -\frac{1}{2V^2}\,\nabla_\alpha(\xi_\mu \xi^\mu) = \frac{1}{2V^2}\,\nabla_\alpha V^2 = \frac{1}{2}\,\nabla_\alpha \ln V^2 = \nabla_\alpha \ln V,$$

hence

$$\boldsymbol{a} = \vec{\boldsymbol{\nabla}} \ln V. \tag{3.63}$$

The norm of $\boldsymbol{a}$, which is always a spacelike vector (since the unit character of $\boldsymbol{u}$ implies $\boldsymbol{u}\cdot\boldsymbol{a} = 0$), is

$$a := \sqrt{\boldsymbol{a}\cdot\boldsymbol{a}} = \frac{1}{V}\sqrt{\nabla_\mu V \nabla^\mu V}. \tag{3.64}$$

Given the result (3.54), we get an expression of κ involving a:

$$\boxed{\kappa = \lim_{\mathscr{O}\to\mathscr{H}} Va}, \tag{3.65}$$

where $\mathscr{O}\to\mathscr{H}$ means that the limit is achieved by choosing the worldline of observer $\mathscr{O}$ arbitrarily close to $\mathscr{H}$. Since $V \to 0$ as one approaches $\mathscr{H}$, it follows that

$$\lim_{\mathscr{O}\to\mathscr{H}} a = +\infty. \tag{3.66}$$

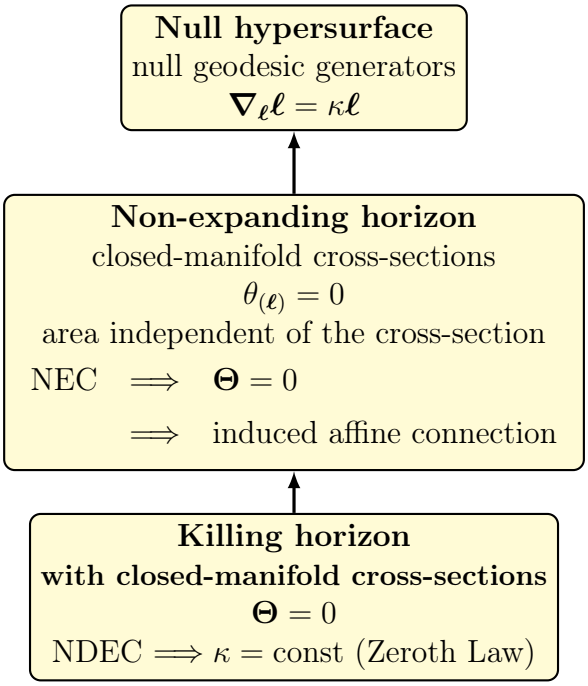
This means that the acceleration felt by observer $\mathscr{O}$ (the "gravity") diverges as $\mathscr{O}$ is placed more and more close to $\mathscr{H}$. In that sense, the *physical* surface gravity of $\mathscr{H}$ is infinite. But Eq. (3.65) shows that the rescaled acceleration Va remains finite as one approaches $\mathscr{H}$, and tends to κ. It is this quantity that is named the ***surface gravity*** of the Killing horizon $\mathscr{H}$.

Remark 3: As stressed above, the surface gravity κ is not the actual gravity a measured *locally*, i.e. by an observer at rest with respect to $\mathscr{H}$ and infinitely close to it. However, κ can be interpreted as a physical force (per unit mass) measured by a *distant* observer, at least in the special case of a Schwarzschild black hole, for which $\boldsymbol{\xi}$ is timelike in the entire region outside the Killing horizon[6]. In this case, one can identify κ to the force exerted by an observer "at infinity" to hold in place a particle of unit mass close to $\mathscr{H}$ by means of an infinitely long massless string (see e.g. Sec. 5.2.4 of Poisson's textbook [214]).

3.4 Summary

Here is an inheritance diagram summarizing the main results of this chapter. The vertical arrow means "is a", i.e. the element at the bottom of the arrow is a special case of the element at the top of the arrow. "NEC" stands for "Null Energy Condition" and "NDEC" for "Null Dominant Energy condition".

[6] This is not true for a rotating Kerr black hole: $\boldsymbol{\xi}$ becomes null at some "light-cylinder" outside $\mathscr{H}$ and is then spacelike away from it, cf. Eq. (10.73), where $\boldsymbol{\xi}$ is denoted by $\boldsymbol{\chi}$.


Null hypersurface
null geodesic generators
$\boldsymbol{\nabla}_{\boldsymbol{\ell}}\boldsymbol{\ell} = \kappa\boldsymbol{\ell}$
Non-expanding horizon
closed-manifold cross-sections
$\theta_{(\boldsymbol{\ell})} = 0$
area independent of the cross-section
NEC $\Longrightarrow$ $\boldsymbol{\Theta} = 0$
$\Longrightarrow$ induced affine connection
Killing horizon
with closed-manifold cross-sections
$\boldsymbol{\Theta} = 0$
NDEC $\Longrightarrow$ κ = const (Zeroth Law)

Chapter 4

The concept of black hole 3: The global view

Contents

4.1 Introduction

Having attempted in Chaps. 2 and 3 to characterize a black hole by the local properties of its boundary, we turn now to the general definition of a black hole. As it could have been anticipated from the naive "definition" given in Sec. 2.2.1, the mathematically meaningful definition of a black hole cannot be local: it has to take into account the full spacetime structure, in particular its future asymptotics. Indeed, to conclude firmly that a null geodesic has escaped, one has to wait until the "end of time"...

In this chapter, we therefore consider the global spacetime picture to arrive at the general definition of a black hole in Sec. 4.4. This amounts to focusing on the spacetime asymptotics, which can be seen as the region where the "distant observers" live and may, or may not, receive light rays from some "central region". This far-away structure is best described in terms of the so-called *conformal completion*, which brings the spacetime infinity(ies) to a finite distance in another manifold — a technique initiated by R. Penrose [205, 206] (see Ref. [108] for a review). We start by investigating the conformal completion of the simplest spacetime: Minkowski spacetime.

4.2 Conformal completion of Minkowski spacetime

In this section $(\mathscr{M}, \boldsymbol{g})$ is the 4-dimensional Minkowski spacetime, i.e. $\mathscr{M}$ is a smooth manifold diffeomorphic to $\mathbb{R}^4$ and $\boldsymbol{g}$ is the metric tensor whose expression in terms of some global coordinates $(x^\alpha) = (t, x, y, z)$ implementing the diffeomorphism to $\mathbb{R}^4$ (i.e. *Minkowskian coordinates*) is

$$g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{4.1}$$

4.2.1 Finite-range coordinates on Minkowski spacetime

Since we would like to deal with the "far" region, it is natural to introduce $r := \sqrt{x^2 + y^2 + z^2}$ and the associated spherical coordinates $(x^\alpha) = (t, r, \theta, \varphi)$, which are related to the Minkowskian ones by

$$\begin{cases} x = r \sin\theta \cos\varphi \\ y = r \sin\theta \sin\varphi \\ z = r \cos\theta. \end{cases} \tag{4.2}$$

The coordinates (t, r, θ, φ) span $\mathbb{R} \times (0, +\infty) \times (0, \pi) \times (0, 2\pi)$; they do not cover the whole manifold $\mathscr{M}$ as a regular chart (cf. Sec. A.2.1 of Appendix A), but only $\mathscr{M} \setminus \Pi$, where Π is the closed half hyperplane defined by $y = 0$ and $x \geq 0$. Once expressed in terms of the spherical coordinates, the Minkowski metric (4.1) takes the form

$$g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}t^2 + \mathrm{d}r^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right). \tag{4.3}$$

Let us introduce the null coordinate system (u, v, θ, φ) where u and v are respectively the retarded and advanced time defined by (cf. Fig. 4.1)

$$\begin{cases} u = t - r \\ v = t + r \end{cases} \iff \begin{cases} t = \frac{1}{2}(v + u) \\ r = \frac{1}{2}(v - u). \end{cases} \tag{4.4}$$

The metric tensor takes then the shape

$$g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}u\, \mathrm{d}v + \frac{1}{4}(v - u)^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right). \tag{4.5}$$

The coordinates (u, v) span the half part of $\mathbb{R}^2$ defined by $u < v$. In order to have coordinates within a finite range, let us consider their arctangents (cf. Fig. 4.2):

$$\begin{cases} U = \arctan u \\ V = \arctan v \end{cases} \iff \begin{cases} u = \tan U \\ v = \tan V. \end{cases} \tag{4.6}$$

Then the coordinates (U, V) span the half part of $(-\pi/2, \pi/2) \times (-\pi/2, \pi/2)$ defined by $U < V$ (since arctan is a monotonically increasing function, cf. Fig. 4.2):

$$-\frac{\pi}{2} < U < \frac{\pi}{2}, \quad -\frac{\pi}{2} < V < \frac{\pi}{2}, \quad \text{and} \quad U < V. \tag{4.7}$$

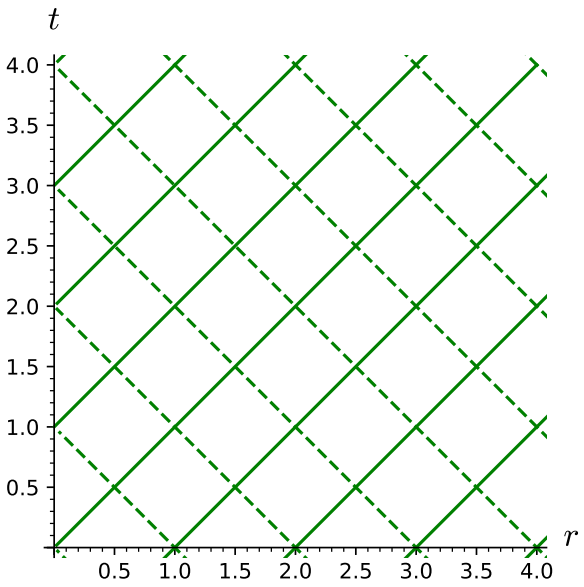


Figure 4.1: Lines of constant null coordinates u (solid) and v (dashed) in terms of the coordinates (t, r).

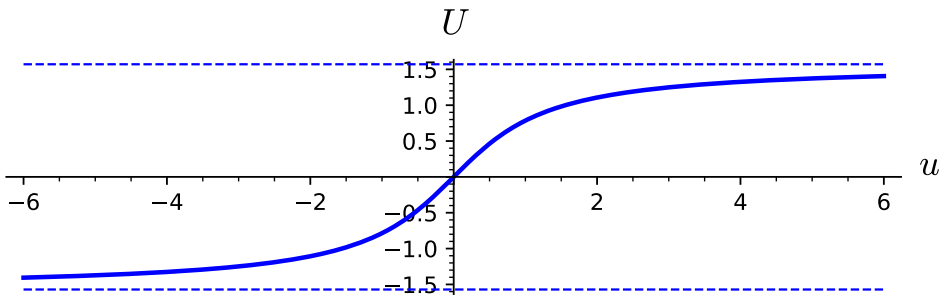


Figure 4.2: The arctangent function mapping $\mathbb{R}$ to $(-\pi/2, \pi/2)$.

Since

$$\mathrm{d}u = \frac{\mathrm{d}U}{\cos^2 U}, \quad \mathrm{d}v = \frac{\mathrm{d}V}{\cos^2 V} \quad \text{and} \quad \tan V - \tan U = \frac{\sin(V-U)}{\cos U \cos V},$$

the Minkowski metric (4.5) is expressed in terms of the coordinates $(x^\alpha) = (U, V, \theta, \varphi)$ as[1]

$$g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \frac{1}{4\cos^2 U \cos^2 V}\left[-4\mathrm{d}U\,\mathrm{d}V + \sin^2(V-U)\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)\right]. \tag{4.8}$$

Remark 1: The retarded/advanced times u and v have the dimension of a time, or of a length in the $c = 1$ units that we are using. Therefore, one should introduce some length scale, ℓ_0 say,

[1] See also the SageMath notebook D.2.1.

before taking their arctangent and rewrite (4.6) as

$$\begin{cases} U = \arctan(u/\ell_0) \\ V = \arctan(v/\ell_0) \end{cases} \iff \begin{cases} u = \ell_0 \tan U \\ v = \ell_0 \tan V. \end{cases}$$

The coordinates (U, V) are dimensionless and a global factor ℓ_0^2 should be introduced in the right-hand side of Eq. (4.8). However, the length scale ℓ_0 plays no essential role and, to keep simple notations, it is omitted in what follows. In other words, we are using units for which $\ell_0 = 1$.

4.2.2 Conformal metric

In the right-hand side of (4.8), the terms in square brackets defines a metric $\tilde{\boldsymbol{g}}$ such that

$$\boxed{\tilde{\boldsymbol{g}} = \Omega^2 \boldsymbol{g}}, \tag{4.9}$$

where Ω is the scalar field $\mathscr{M} \to \mathbb{R}$ obeying

$$\Omega = 2 \cos U \cos V \tag{4.10a}$$

$$= \frac{2}{\sqrt{1+u^2}\sqrt{1+v^2}} \tag{4.10b}$$

$$= \frac{2}{\sqrt{(t-r)^2+1}\sqrt{(t+r)^2+1}}. \tag{4.10c}$$

We notice on (4.10b) and (4.10c) that the function Ω never vanishes on $\mathscr{M}$, so that the bilinear form $\tilde{\boldsymbol{g}}$ defined by (4.9) constitutes a well-behaved metric on $\mathscr{M}$. Moreover, since $\Omega^2 > 0$, $\tilde{\boldsymbol{g}}$ has the same signature as $\boldsymbol{g}$, i.e. $(-,+,+,+)$. The specific expression of $\tilde{\boldsymbol{g}}$ is deduced from (4.8) and (4.10a):

$$\tilde{g}_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = -4 \mathrm{d}U \, \mathrm{d}V + \sin^2(V-U) \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\varphi^2 \right). \tag{4.11}$$

In view of (4.9), one says that the metric $\tilde{\boldsymbol{g}}$ is ***conformal to*** the metric $\boldsymbol{g}$, or equivalently, that the metrics $\boldsymbol{g}$ and $\tilde{\boldsymbol{g}}$ are ***conformally related***, or that $\tilde{\boldsymbol{g}}$ arises from $\boldsymbol{g}$ via a ***conformal transformation***. The scalar field Ω is called the ***conformal factor***.

A key property of a conformal transformation is to preserve orthogonality relations, since (4.9) clearly implies, at any point $p \in \mathscr{M}$,

$$\forall (\boldsymbol{u}, \boldsymbol{v}) \in T_p\mathscr{M} \times T_p\mathscr{M}, \quad \tilde{\boldsymbol{g}}(\boldsymbol{u}, \boldsymbol{v}) = 0 \iff \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = 0.$$

In particular, null vectors for $\tilde{\boldsymbol{g}}$ coincide with null vectors for $\boldsymbol{g}$:

$$\forall \boldsymbol{\ell} \in T_p\mathscr{M}, \quad \tilde{\boldsymbol{g}}(\boldsymbol{\ell}, \boldsymbol{\ell}) = 0 \iff \boldsymbol{g}(\boldsymbol{\ell}, \boldsymbol{\ell}) = 0.$$

Consequently the light cones of $(\mathscr{M}, \boldsymbol{g})$ and $(\mathscr{M}, \tilde{\boldsymbol{g}})$ are identical, which implies that $(\mathscr{M}, \boldsymbol{g})$ and $(\mathscr{M}, \tilde{\boldsymbol{g}})$ have the same causal structure. Moreover, since $\Omega^2 > 0$, the spacelike and timelike characters of vectors is preserved as well:

$$\begin{aligned} \forall \boldsymbol{v} \in T_p\mathscr{M}, \quad & \boldsymbol{v} \text{ spacelike for } \tilde{\boldsymbol{g}} \iff \boldsymbol{v} \text{ spacelike for } \boldsymbol{g} \\ & \boldsymbol{v} \text{ timelike for } \tilde{\boldsymbol{g}} \iff \boldsymbol{v} \text{ timelike for } \boldsymbol{g}. \end{aligned} \tag{4.12}$$

It follows that a curve $\mathscr{L}$ is timelike (resp. null, spacelike) for $\tilde{\boldsymbol{g}}$ iff $\mathscr{L}$ is timelike (resp. null, spacelike) for $\boldsymbol{g}$. Similarly, a hypersurface Σ is timelike (resp. null, spacelike) for $\tilde{\boldsymbol{g}}$ iff Σ is timelike (resp. null, spacelike) for $\boldsymbol{g}$.

What about geodesics? Let us first recall that a null curve is not necessarily a null geodesic (cf. Remark 2 on page 35 and Appendix B), so that one cannot deduce from the above results that conformal transformations preserve null geodesics. However, this turns out to be true:

A smooth curve $\mathscr{L}$ in $\mathscr{M}$ is a null geodesic for $\tilde{\boldsymbol{g}}$ iff $\mathscr{L}$ is a null geodesic for $\boldsymbol{g}$.

To prove it, it suffices to write explicitly the geodesic equation [Eq. (B.10)] and to express the Christoffel symbols of $\tilde{\boldsymbol{g}}$ in terms of those of $\boldsymbol{g}$ and the derivatives of Ω (see e.g. Appendix D of Wald's textbook [256] for details).

On the contrary, conformal transformations preserve neither the timelike geodesics nor the spacelike ones.

The coordinates (U, V) are of null type; let us consider instead the "time+space" coordinates (τ, χ) defined by[2]

$$\begin{cases} \tau = V + U \\ \chi = V - U \end{cases} \iff \begin{cases} U = \frac{1}{2}(\tau - \chi) \\ V = \frac{1}{2}(\tau + \chi). \end{cases} \tag{4.13}$$

Given (4.7), the range of these new coordinates is

$$0 < \chi < \pi \quad \text{and} \quad \chi - \pi < \tau < \pi - \chi. \tag{4.14}$$

In other words, if we draw the Minkowski spacetime in the (τ, χ) plane, it takes the shape of a half-diamond, as depicted in Fig. 4.3.

By combining (4.4) (4.6) and (4.13), we get the link between (t, r) and (τ, χ):

$$\begin{cases} \tau = \arctan(t + r) + \arctan(t - r) \\ \chi = \arctan(t + r) - \arctan(t - r) \end{cases} \iff \begin{cases} t = \dfrac{\sin\tau}{\cos\tau + \cos\chi} \\ r = \dfrac{\sin\chi}{\cos\tau + \cos\chi}. \end{cases} \tag{4.15}$$

We may use these relations to draw the lines $t = \text{const}$ and $r = \text{const}$ in Fig. 4.3.

The expression of the conformal factor in the coordinates $(\tau, \chi, \theta, \varphi)$ is easily deduced from (4.10a) and (4.13):

$$\Omega = \cos\tau + \cos\chi. \tag{4.16}$$

4.2.3 Conformal completion

The expression of the conformal metric in terms of the coordinates $(x^\alpha) = (\tau, \chi, \theta, \varphi)$ is easily deduced from that in terms of (U, V, θ, φ) as given by (4.11):

$$\tilde{g}_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}\tau^2 + \mathrm{d}\chi^2 + \sin^2\chi\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right). \tag{4.17}$$

[2] Notice the similarity with (4.4) up to some 1/2 factors.

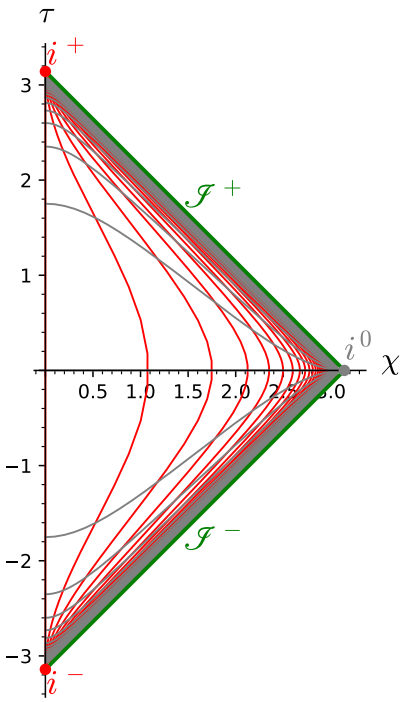


Figure 4.3: Conformal diagram of Minkowski spacetime. Constant-r curves are drawn in red, while constant-t ones are drawn in grey. *[Figure generated by the notebook D.2.1]*

Restricting to a $\tau = \text{const}$ hypersurface, i.e. setting $\mathrm{d}\tau = 0$, we recognize the standard metric of the hypersphere $\mathbb{S}^3$ in the hyperspherical coordinates (χ, θ, φ). Moreover, we notice that the full metric (4.17) is perfectly regular even if we relax the condition on τ in (4.14), i.e. if we let τ span the entire $\mathbb{R}$. We may then consider the manifold

$$\mathscr{E} = \mathbb{R} \times \mathbb{S}^3 \tag{4.18}$$

and $\tilde{\boldsymbol{g}}$ as the Lorentzian metric on $\mathscr{E}$ given by (4.17). The Lorentzian manifold $(\mathscr{E}, \tilde{\boldsymbol{g}})$ is nothing but the ***Einstein static universe***, also called the ***Einstein cylinder***, a static solution of the Einstein equation (1.38) with $\Lambda > 0$ and some pressureless matter of uniform density $\rho = \Lambda/(4\pi)$. We have thus an embedding[3] of Minkowski spacetime into the Einstein cylinder:

$$\Phi : \ \mathscr{M} \longrightarrow \mathscr{E} \tag{4.19}$$

and this embedding is a conformal isometry from $(\mathscr{M}, \boldsymbol{g})$ to $(\Phi(\mathscr{M}), \tilde{\boldsymbol{g}})$. In the following, we shall identify $\Phi(\mathscr{M})$ and $\mathscr{M}$, i.e. use the same symbol $\mathscr{M}$ to denote the subset of $\mathscr{E}$ that is the image of $\mathscr{M}$ via the embedding (4.19).

Since $\mathscr{E}$ and $\mathscr{M}$ have the same dimension, $\mathscr{M}$ is an open subset of $\mathscr{E}$. Its closure $\overline{\mathscr{M}}$ in $\mathscr{E}$ is (cf. Figs. 4.3 and 4.4)

$$\overline{\mathscr{M}} = \mathscr{M} \cup \mathscr{I}^+ \cup \mathscr{I}^- \cup \{i^0\} \cup \{i^+\} \cup \{i^-\}, \tag{4.20}$$

where

[3] Cf. Sec. A.2.7 of Appendix A

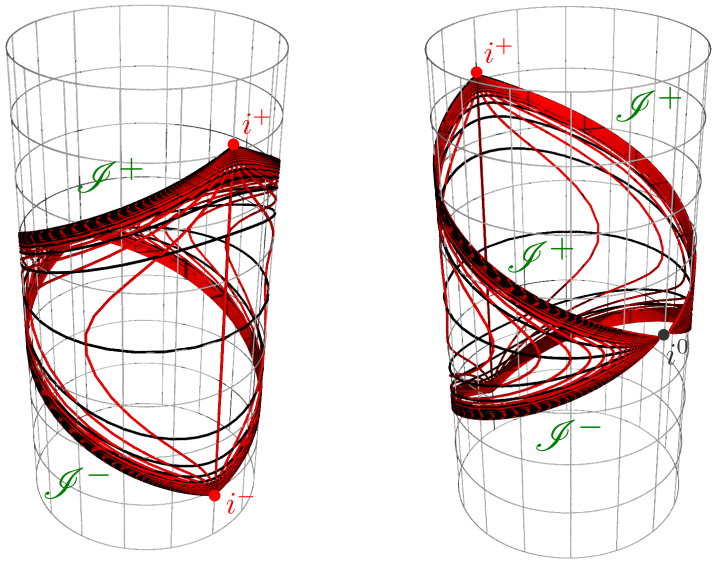


Figure 4.4: Two views of the Einstein cylinder $\mathscr{E}$, with the conformal embedding of Minkowski spacetime in it. Due to the dimensional reduction $4 \to 2$ for the drawing, the $\mathbb{S}^3$ sections of the cylinder are depicted as horizontal circles. The red curves are the same constant-r curves as in Fig. 4.3, while the black curves are the same constant-t curves as those drawn in grey in Fig. 4.3. *[Figure generated by the notebook D.2.1]*

- $\mathscr{I}^+$ is the hypersurface of $\mathscr{E}$ defined by $\tau = \pi - \chi$ and $0 < \tau < \pi$;
- $\mathscr{I}^-$ is the hypersurface of $\mathscr{E}$ defined by $\tau = \chi - \pi$ and $-\pi < \tau < 0$;
- i^0 is the point of $\mathscr{E}$ defined by $\tau = 0$ and $\chi = \pi$;
- i^+ is the point of $\mathscr{E}$ defined by $\tau = \pi$ and $\chi = 0$;
- i^- is the point of $\mathscr{E}$ defined by $\tau = -\pi$ and $\chi = 0$.

It is customary to pronounce $\mathscr{I}$ as "scri", for *script i*.

Remark 2: On $\mathbb{S}^3$, the hyperspherical coordinates (χ, θ, φ) are singular at $\chi = 0$ and $\chi = \pi$, so that setting $\chi = 0$ (or $\chi = \pi$) defines a unique point of $\mathbb{S}^3$, whatever the value of (θ, φ). Note also that the vertical left boundary of the diamond drawn in Fig. 4.3, i.e. the segment defined by $\tau \in (-\pi, \pi)$ and $\chi = 0$, is *not* a part of the boundary of $\mathscr{M}$ but merely reflect the coordinate singularity at $\chi = 0$, in the same way that the left vertical boundary of Fig. 4.1 is not a boundary of Minkowski spacetime but is due to the coordinate singularity at $r = 0$. Note by the way that $\chi = 0$ implies $r = 0$ via (4.15).

Let

$$\mathscr{I} := \mathscr{I}^+ \cup \mathscr{I}^- \tag{4.21}$$

and

$$\tilde{\mathscr{M}} := \mathscr{M} \cup \mathscr{I}. \tag{4.22}$$

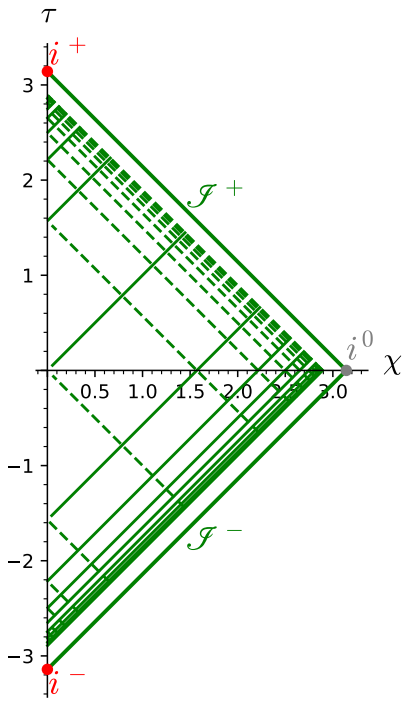


Figure 4.5: Null radial geodesics in the conformal diagram of Minkowski spacetime. The solid green lines are null geodesics $u = \text{const}$ for 17 values of u uniformly spanning $[-8, 8]$, while the dashed green lines are null geodesics $v = \text{const}$ for 17 values of v uniformly spanning $[-8, 8]$. *[Figure generated by the notebook D.2.1]*

$\tilde{\mathscr{M}}$ is naturally a smooth manifold with boundary[4] and its boundary is $\mathscr{I}$:

$$\partial\tilde{\mathscr{M}} = \mathscr{I}. \tag{4.23}$$

Remark 3: Because the closure $\overline{\mathscr{M}}$ is self-intersecting at the point i^0 (cf. Fig. 4.4), it is not a manifold with boundary: no open neighbourhood of i^0 is homeomorphic to a neighbourhood of $\mathbb{H}^4 = \mathbb{R}^3 \times [0, +\infty)$, as the definition of a manifold with boundary would require, cf. Sec. A.2.2. At the points i^+ and i^-, $\overline{\mathscr{M}}$ can be considered as a topological manifold with boundary, but not as a *smooth* manifold with boundary. Hence, the three points i^0, i^+ and i^- are excluded from the definition of the manifold with boundary $\tilde{\mathscr{M}}$.

The hypersurface $\mathscr{I}^+$ is the location of $\tilde{\mathscr{M}}$ where all radial null geodesics terminate, while $\mathscr{I}^-$ is the location of $\tilde{\mathscr{M}}$ where all these geodesics originate (cf. Fig. 4.5). For this reason $\mathscr{I}^+$ is called the ***future null infinity*** of $(\mathscr{M}, \boldsymbol{g})$ and $\mathscr{I}^-$ the ***past null infinity*** of $(\mathscr{M}, \boldsymbol{g})$. On the other side, any timelike geodesic of $(\mathscr{M}, \boldsymbol{g})$ originates at i^- and ends at i^+ (cf. Fig. 4.3), while any spacelike geodesic of $(\mathscr{M}, \boldsymbol{g})$ originates at i^0 and terminates there (after having completed a closed path on $\mathbb{S}^3$ (cf. Fig. 4.4). The point i^+ is then called the ***future timelike infinity*** of $(\mathscr{M}, \boldsymbol{g})$, i^- the ***past timelike infinity*** of $(\mathscr{M}, \boldsymbol{g})$ and i^0 the ***spacelike infinity*** of $(\mathscr{M}, \boldsymbol{g})$.

[4] Cf. Sec. A.2.2 for the precise definition.

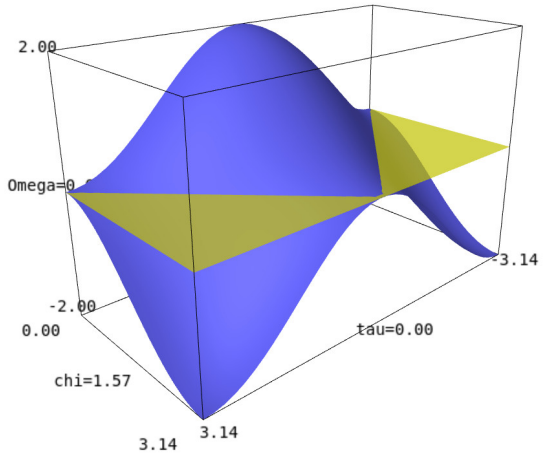


Figure 4.6: Conformal factor Ω as a function of (τ, χ) [cf. Eq. (4.16)]. Only the part above the yellow horizontal plane ($\Omega = 0$) is physical. *[Figure generated by the notebook D.2.1]*

As it is clear on the conformal diagram of Fig. 4.3, both $\mathscr{I}^+$ and $\mathscr{I}^-$ are null hypersurfaces of $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$.

It is precisely because Ω vanishes (cf. Fig. 4.6) at the boundary

$$\overline{\mathscr{M}} \setminus \mathscr{M} = \mathscr{I}^+ \cup \mathscr{I}^- \cup \{i^0\} \cup \{i^+\} \cup \{i^-\} \tag{4.24}$$

that the conformal transformation (4.9) brings infinity of Minkowski space to a finite distance.

Historical note: The idea of using a conformal transformation to treat infinity as a boundary "at a finite distance" has been put forward by Roger Penrose in 1963 [205] and expanded in 1964 in the seminal paper [206], where Penrose constructed the conformal completion of Minkowski spacetime as a part of the Einstein cylinder. In particular, Fig. 3 of Ref. [206] is equivalent to Fig. 4.4.

4.3 Conformal completions and asymptotic flatness

Having investigated the asymptotic structure of Minkowski spacetime via a conformal completion, let us use the latter to define spacetimes that "look like" Minkowski spacetime asymptotically. A first step is the concept of conformal completion.

4.3.1 Conformal completion

A spacetime $(\mathscr{M}, \boldsymbol{g})$ admits a ***conformal completion*** iff there exists a Lorentzian manifold with boundary $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ equipped with a smooth non-negative scalar field $\Omega : \tilde{\mathscr{M}} \to \mathbb{R}^+$ such that

1. $\tilde{\mathscr{M}} = \mathscr{M} \cup \mathscr{I}$, with $\mathscr{I} := \partial\tilde{\mathscr{M}}$ (the boundary of $\tilde{\mathscr{M}}$);

2. on $\mathscr{M}$, $\tilde{\boldsymbol{g}} = \Omega^2 \boldsymbol{g}$;
3. on $\mathscr{I}$, $\Omega = 0$;
4. on $\mathscr{I}$, $\mathbf{d}\Omega \neq 0$.

$\mathscr{I}$ is called the ***conformal boundary*** of $(\mathscr{M}, \boldsymbol{g})$ within the conformal completion $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$.

Condition 1 expresses that $\mathscr{M}$ has been endowed with some boundary. A rigorous formulation of it would be via an embedding $\Phi : \mathscr{M} \to \tilde{\mathscr{M}}$, as in Eq. (4.19), so that $\tilde{\mathscr{M}} = \Phi(\mathscr{M}) \cup \mathscr{I}$. However, as above, we identify $\Phi(\mathscr{M})$ with $\mathscr{M}$ and therefore simply write $\tilde{\mathscr{M}} = \mathscr{M} \cup \mathscr{I}$. Conditions 2 and 3 express that the boundary of $\mathscr{M}$, which "lies at an infinite distance" with respect to $\boldsymbol{g}$, has been brought to a finite distance with respect to $\tilde{\boldsymbol{g}}$. Indeed, in terms of length elements, condition 2 implies

$$\mathrm{d}s^2 = \frac{1}{\Omega^2}\, \mathrm{d}\tilde{s}^2$$

with $1/\Omega^2 \to +\infty$ as one approaches $\mathscr{I}$ (condition 3). Finally, condition 4 ensures that $\mathscr{I}$ is a regular hypersurface of $\tilde{\mathscr{M}}$. It is of course fulfilled by Minkowski spacetime, as we can check graphically on Fig. 4.6: the graph of Ω has no horizontal slope at $\mathscr{I}$.

Remark 1: The statement that $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ is a Lorentzian manifold with boundary implies that $\tilde{\boldsymbol{g}}$ is smooth everywhere on $\tilde{\mathscr{M}}$, including at the boundary $\mathscr{I}$.

Remark 2: The conformal boundary $\mathscr{I}$ is not part of the physical spacetime $\mathscr{M}$, but only of the conformal completion $\tilde{\mathscr{M}}$.

Remark 3: One often speaks about *conformal compactification* instead of *conformal completion*, but in general $\tilde{\mathscr{M}}$ is not a compact manifold. For instance, because we omitted the points i^+, i^- and i^0, the completion $\tilde{\mathscr{M}}$ of Minkowski spacetime defined by Eq. (4.22) is not compact.

Example 1 (conformal completion of $\mathbf{AdS}_4$ spacetime): The 4-dimensional ***anti-de Sitter spacetime*** is $(\mathscr{M}, \boldsymbol{g})$ with $\mathscr{M} \simeq \mathbb{R}^4$ and $\boldsymbol{g}$ is the metric whose components in the so-called *global coordinates* $(x^\alpha) = (\tau, \rho, \theta, \varphi)$ are given by

$$g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = \ell^2 \left[-\cosh^2\rho\, \mathrm{d}\tau^2 + \mathrm{d}\rho^2 + \sinh^2\rho \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2 \right) \right], \tag{4.25}$$

where ℓ is a positive constant. Note that τ spans $\mathbb{R}$, ρ spans $(0, +\infty)$, while (θ, φ) are standard polar coordinates on $\mathbb{S}^2$: $\theta \in (0, \pi)$ and $\varphi \in (0, 2\pi)$. The metric (4.25) is a solution of the Einstein equation (1.38) with the negative cosmological constant $\Lambda = -3/\ell^2$ and $\boldsymbol{T} = 0$ (vacuum). Using coordinates $(x^\alpha) = (\tau, \chi, \theta, \varphi)$ with $\chi = \mathrm{atan}(\sinh\rho) \in (0, \pi/2)$, one gets

$$g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = \frac{\ell^2}{\cos^2\chi} \left[-\mathrm{d}\tau^2 + \mathrm{d}\chi^2 + \sin^2\chi \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2 \right) \right]. \tag{4.26}$$

Defining $\Omega := \ell^{-1} \cos\chi = (\ell \cosh\rho)^{-1}$, we notice that a conformal completion of $(\mathscr{M}, \boldsymbol{g})$ is $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ where (i) $\tilde{\mathscr{M}}$ is the part $\chi \leq \pi/2$ of the Einstein cylinder[5] introduced in Sec. 4.2.3 and

[5] Recall that on the Einstein cylinder the range of χ is $(0, \pi)$, cf. Eq. (4.14).

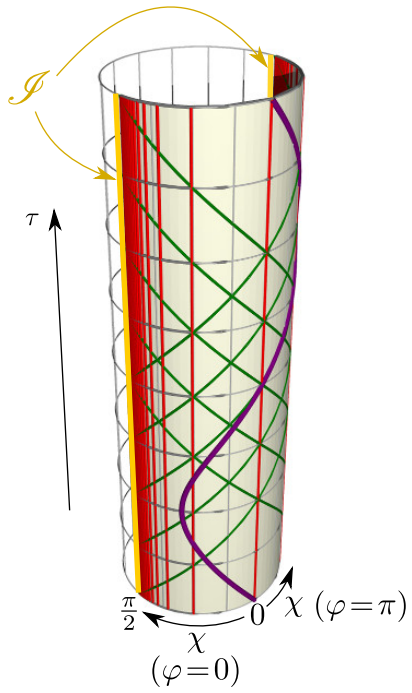


Figure 4.7: Conformal completion of AdS_4 spacetime, depicted on the Einstein cylinder. The conformal boundary $\mathscr{I}$ is shown in yellow, red lines are lines $\chi = \text{const}$ (uniformly sampled in terms of $\tan\chi = \sinh\rho$), green curves are radial null geodesics and the purple curve is a radial timelike geodesic, bouncing back and forth around $\chi = 0$. *The SageMath notebook (cf. Appendix D) generating this figure is available at* https://nbviewer.jupyter.org/github/sagemanifolds/SageManifolds/blob/master/Notebooks/SM_anti_de_Sitter.ipynb.

(ii) $\tilde{\boldsymbol{g}}$ is the metric (4.17). The boundary $\mathscr{I} = \partial\tilde{\mathscr{M}}$ is then the hypersurface $\chi = \pi/2$ of the Einstein cylinder (cf. Fig. 4.7); $\mathscr{I}$ is spanned by the coordinates (τ, θ, φ) and its topology is that of a 3-dimensional cylinder: $\mathscr{I} \simeq \mathbb{R} \times \mathbb{S}^2$. We notice that conditions 3 and 4 of the definition of a conformal completion are satisfied: $\Omega = \ell^{-1}\cos\chi = 0$ at $\mathscr{I}$ and $\mathbf{d}\Omega = -\ell^{-1}\sin\chi\,\mathbf{d}\chi = -\ell^{-1}\mathbf{d}\chi \neq 0$ at $\mathscr{I}$. The metric induced by $\tilde{\boldsymbol{g}}$ on $\mathscr{I}$ is obtained by setting $\chi = \pi/2$ in (4.17): $\mathrm{d}\tilde{s}^2 = -\mathrm{d}\tau^2 + \mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2$. This 3-metric is clearly Lorentzian, which shows that $\mathscr{I}$ is a timelike hypersurface of $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$.

The above example shows that $\mathscr{I}$ is not necessarily a null hypersurface, as it is for Minkowski spacetime (cf. Sec. 4.2.3). Actually the hypersurface-type of $\mathscr{I}$ is determined by the cosmological constant, as follows:

If the spacetime dimension obeys[a] $n \geq 3$ and $\boldsymbol{g}$ is a solution of Einstein equation with a cosmological constant Λ [Eq. (1.38)] and the trace T of the energy-momentum tensor tends to zero in the vicinity of $\mathscr{I}$ (i.e. when $\Omega \to 0$), then

- $\mathscr{I}$ is a null hypersurface of $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ iff $\Lambda = 0$;
- $\mathscr{I}$ is a spacelike hypersurface of $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ iff $\Lambda > 0$;
- $\mathscr{I}$ is a timelike hypersurface of $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ iff $\Lambda < 0$.

[a] Cf. Remark 1 in Sec. 1.5.

Proof. It follows from $\tilde{\boldsymbol{g}} = \Omega^2 \boldsymbol{g}$ that the Ricci scalars $\tilde{R}$ and R of respectively $\tilde{\boldsymbol{g}}$ and $\boldsymbol{g}$ are related by[6]

$$\Omega^2 \tilde{R} = R - (n-1)\left(2\Omega\, \tilde{g}^{\mu\nu}\tilde{\nabla}_\mu\tilde{\nabla}_\nu\Omega - n\, \tilde{g}^{\mu\nu}\partial_\mu\Omega\partial_\nu\Omega\right), \tag{4.27}$$

where $n = \dim \mathscr{M}$ and $\tilde{\nabla}$ stands for the Levi-Civita connection of $\tilde{\boldsymbol{g}}$. Using the trace of the Einstein equation (1.39) to express R, we get

$$\Omega^2 \tilde{R} = \frac{2}{n-2}\left(n\Lambda - 8\pi T\right) - (n-1)\left(2\Omega\, \tilde{g}^{\mu\nu}\tilde{\nabla}_\mu\tilde{\nabla}_\nu\Omega - n\, \tilde{g}^{\mu\nu}\partial_\mu\Omega\partial_\nu\Omega\right)$$

This equation is a priori valid in $\mathscr{M} = \tilde{\mathscr{M}} \setminus \mathscr{I}$ only. Taking the limit $\Omega \to 0$ and assuming that $T \to 0$ in that limit, we get, by continuity, an identity on $\mathscr{I}$:

$$\tilde{g}^{\mu\nu}\partial_\mu\Omega\partial_\nu\Omega \overset{\mathscr{I}}{=} -\frac{2}{(n-1)(n-2)}\Lambda. \tag{4.28}$$

Since $\mathscr{I}$ corresponds to a constant value of the scalar field Ω ($\Omega = 0$), the left-hand side of this equation is nothing but the scalar square $\tilde{\boldsymbol{g}}(\boldsymbol{n},\boldsymbol{n})$ of the vector $\boldsymbol{n}$ normal to $\mathscr{I}$ defined as the dual with respect to $\tilde{\boldsymbol{g}}$ of the 1-form $\mathbf{d}\Omega$: $n^\alpha = \tilde{g}^{\alpha\mu}\partial_\mu\Omega$ (remember that by hypothesis 4 in the definition of a conformal completion, $\mathbf{d}\Omega$ is non-vanishing on $\mathscr{I}$, so that $\boldsymbol{n}$ is a valid normal vector to $\mathscr{I}$). Equation (4.28) implies that the sign of $\tilde{\boldsymbol{g}}(\boldsymbol{n},\boldsymbol{n})$ is the opposite of that of Λ. Given the link between the type of a hypersurface and the type of its normal (cf. Sec. 2.2.2), this completes the proof. □

The definition of black hole shall involve a subcategory of conformal completions:

Let $(\mathscr{M}, \boldsymbol{g})$ be a time-orientable[a] spacetime admitting a conformal completion $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$. One says that $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ is a ***conformal completion at null infinity*** of $(\mathscr{M}, \boldsymbol{g})$ iff the boundary $\mathscr{I} := \partial\tilde{\mathscr{M}}$ obeys

$$\mathscr{I} = \mathscr{I}^+ \cup \mathscr{I}^-, \tag{4.29}$$

with $\mathscr{I}^+$ (resp. $\mathscr{I}^-$) being never intersected by any past-directed (resp. future-directed) causal curve originating in $\mathscr{M}$. Let us recall that a ***causal curve*** is curve whose tangent vectors are nowhere spacelike. As for Minkowksi spacetime, we shall call $\mathscr{I}^+$ the ***future null infinity*** and $\mathscr{I}^-$ the ***past null infinity*** of $(\mathscr{M}, \boldsymbol{g})$.

[a] Cf. Sec. 1.2.2.

[6] This relation is easily established by starting from Eq. (2.30) of Hawking & Ellis textbook [146] or Eq. (2.19) on page 645 of Choquet-Bruhat one [51] and inverting the roles of $\tilde{\boldsymbol{g}}$ and $\boldsymbol{g}$, thereby substituting Ω^{-1} for Ω.

Remark 4: The above definition of $\mathscr{I}^+$ and $\mathscr{I}^-$ does not impose that these two objects are null hypersurfaces of $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$. This is true for Minkowski spacetime, but cannot hold for spacetimes with a non-zero cosmological constant, as shown above. In particular, the following example exhibits spacelike $\mathscr{I}^+$ and $\mathscr{I}^-$.

Example 2 (Conformal completion of dS_4 spacetime): The 4-dimensional ***de Sitter spacetime*** is $(\mathscr{M}, \boldsymbol{g})$ with $\mathscr{M} \simeq \mathbb{R} \times \mathbb{S}^3$ and $\boldsymbol{g}$ is the metric whose components in the so-called *global coordinates* $(x^\alpha) = (t, \chi, \theta, \varphi)$ are given by

$$g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = \ell^2 \left[-\mathrm{d}t^2 + \cosh^2 t \left(\mathrm{d}\chi^2 + \sin^2\chi \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\varphi^2 \right) \right) \right], \tag{4.30}$$

where ℓ is a positive constant. Note that t spans $\mathbb{R}$ while (χ, θ, φ) are standard polar coordinates on $\mathbb{S}^3$: $\chi \in (0, \pi)$, $\theta \in (0, \pi)$ and $\varphi \in (0, 2\pi)$. The metric (4.30) is a solution of the Einstein equation (1.38) with the positive cosmological constant $\Lambda = 3/\ell^2$ and $\boldsymbol{T} = 0$ (vacuum). Using coordinates $(x^\alpha) = (\tau, \chi, \theta, \varphi)$ with $\tau = 2\mathrm{atan}(\tanh(t/2)) \in (-\pi/2, \pi/2)$, one gets

$$g_{\mu\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu = \frac{\ell^2}{\cos^2\tau} \left[-\mathrm{d}\tau^2 + \mathrm{d}\chi^2 + \sin^2\chi \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\varphi^2 \right) \right]. \tag{4.31}$$

Defining $\Omega := \ell^{-1} \cos\tau = (\ell \cosh t)^{-1}$, we notice that a conformal completion of $(\mathscr{M}, \boldsymbol{g})$ is $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ where (i) $\tilde{\mathscr{M}}$ is the part $-\pi/2 \leq \tau \leq \pi/2$ of the Einstein cylinder introduced in Sec. 4.2.3 and (ii) $\tilde{\boldsymbol{g}}$ is the metric (4.17). The boundary $\mathscr{I} = \partial\tilde{\mathscr{M}}$ has two connected components: $\mathscr{I}^+$, which is the hypersurface $\tau = \pi/2$ of $\tilde{\mathscr{M}}$, and $\mathscr{I}^-$, which is the hypersurface $\tau = -\pi/2$. Both $\mathscr{I}^+$ and $\mathscr{I}^-$ are spanned by the coordinates (χ, θ, φ) and their topology is that of $\mathbb{S}^3$. We notice that conditions 3 and 4 of the definition of a conformal completion are satisfied: $\Omega = \ell^{-1}\cos\tau = 0$ at $\mathscr{I}$ and $\mathbf{d}\Omega = -\ell^{-1}\sin\tau \, \mathbf{d}\tau = \pm\ell^{-1}\mathbf{d}\tau \neq 0$ at $\mathscr{I}$. The metric induced by $\tilde{\boldsymbol{g}}$ on $\mathscr{I}$ is obtained by setting $\tau = \pm\pi/2$ in (4.17): $\mathrm{d}\tilde{s}^2 = \mathrm{d}\chi^2 + \sin^2\chi \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\varphi^2 \right)$. This 3-metric is clearly Riemannian (this is actually the standard round metric of $\mathbb{S}^3$), which shows that $\mathscr{I}$ is a spacelike hypersurface of $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$. This of course agrees with the general property established above for spacetimes with a positive cosmological constant. Finally, it is clear that any causal curve originating in $\mathscr{M}$ that intersects $\mathscr{I}^+$ must approach $\tau = \pi/2$ from below, i.e. cannot be past-directed. Similarly any causal curve originating in $\mathscr{M}$ that intersects $\mathscr{I}^-$ must approach $\tau = -\pi/2$ from above, i.e. cannot be future-directed. We conclude that $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ is a conformal completion at null infinity of the de Sitter spacetime.

4.3.2 Asymptotic flatness

Penrose [206, 209] has defined a spacetime $(\mathscr{M}, \boldsymbol{g})$ to be ***asymptotically simple*** iff there exists a conformal completion $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ of $(\mathscr{M}, \boldsymbol{g})$ such that every null geodesic in $\mathscr{M}$ has two endpoints in $\mathscr{I}$.

The last condition, which is verified by Minkowski spacetime (cf. Fig. 4.5), de Sitter spacetime and anti-de Sitter spacetime (cf. the null geodesics in Fig. 4.7), is rather restrictive. In particular, it excludes black hole spacetimes, since, almost by definition, the latter contain null geodesics that have no endpoint on $\mathscr{I}^+$, having only a past endpoint on $\mathscr{I}^-$, as far as $\mathscr{I}$ is concerned. To cope with these spacetimes, Penrose has also introduced the following definition [209]: a spacetime $(\mathscr{M}, \boldsymbol{g})$ is ***weakly asymptotically simple*** iff there exists an open subset $\mathscr{U}$ of $\mathscr{M}$ and an asymptotically simple spacetime $(\mathscr{M}_0, \boldsymbol{g}_0)$

with an open neighbourhood $\mathscr{U}_0$ of $\mathscr{I}_0 = \partial\tilde{\mathscr{M}}_0$ in $\tilde{\mathscr{M}}_0$ such that $(\mathscr{U}_0 \cap \mathscr{M}_0, \boldsymbol{g}_0)$ is isometric to $(\mathscr{U}, \boldsymbol{g})$.

Remark 5: For a given weakly asymptotically simple spacetime, there may be different (non overlapping) regions $\mathscr{U}$ satisfying the above property. For instance we shall see in Chap. 10 that there are an infinite series of them in the Kerr spacetime.

Finally one says that a spacetime $(\mathscr{M}, \boldsymbol{g})$ is ***asymptotically flat*** (or more precisely ***weakly asymptotically simple and empty*** [146]) iff $(\mathscr{M}, \boldsymbol{g})$ is weakly asymptotically simple and the Ricci tensor of $\boldsymbol{g}$ vanishes in an open neighbourhood of $\mathscr{I}$: $\boldsymbol{R} = 0$.

Example 3: The de Sitter and anti-de Sitter spacetimes are asymptotically simple but are not asymptotically flat.

Penrose [208] (see also [108]) has shown that if $(\mathscr{M}, \boldsymbol{g})$ is asymptotically simple and empty, that the Weyl tensor of $\boldsymbol{g}$ (cf. Sec. A.5.4) vanishes at $\mathscr{I}$. Since the Ricci tensor is zero, this implies that the full Riemann curvature tensor vanishes at $\mathscr{I}$ [cf. Eq. (A.114)], hence the qualifier *asymptotically flat*.

The following property holds:

> The conformal boundary $\mathscr{I}$ of an asymptotically flat spacetime $(\mathscr{M}, \boldsymbol{g})$ is a null hypersurface of the conformal completion $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$.

Proof. Consider Eq. (4.27). Near $\mathscr{I}$, we have $R = 0$ by the very definition of asymptotic flatness. The limit $\Omega \to 0$ results then in $\tilde{g}^{\mu\nu}\partial_\mu\Omega\partial_\nu\Omega \overset{\mathscr{I}}{=} 0$, which, following the argument in the proof on p. 92, implies that $\mathscr{I}$ is a null hypersurface. □

4.4 Black holes

4.4.1 Preliminaries regarding causal structure

Before we proceed to the precise definition of a black hole, let us introduce some concepts regarding the causal structure of a given time-orientable spacetime $(\mathscr{M}, \boldsymbol{g})$. For any subset S of $\mathscr{M}$, one defines

- the ***chronological future of*** S as the set $I^+(S)$ of all points of $\mathscr{M}$ that can be reached from a point of S by a future-directed timelike curve of nonzero extent;
- the ***causal future of*** S as the set $J^+(S)$ of all points that either are in S or can be reached from a point of S by a future-directed causal curve;
- the ***chronological past of*** S as the set $I^-(S)$ of all points of $\mathscr{M}$ that can be reached from a point of S by a past-directed timelike curve of nonzero extent;
- the ***causal past of*** S as the set $J^-(S)$ of all points that either are in S or can be reached from a point of S by a past-directed causal curve.

From the above definitions, one has always $S \subset J^\pm(S)$ and $I^\pm(S) \subset J^\pm(S)$.

Remark 1: One has not necessarily $S \subset I^\pm(S)$. For instance, if the spacetime does not contain any closed timelike curve, one has $S \cap I^\pm(S) = \varnothing$ for $S = \{p\}$ with p any point of $\mathscr{M}$.

Here are some basic topological properties of the future and past sets defined above (see e.g. § 6.2 of [146] or Chap. 14 of [199] for proofs):

- $I^\pm(S)$ is always an open subset[7] of $\mathscr{M}$, while $J^\pm(S)$ is not necessarily a closed subset.
- The interior of $J^\pm(S)$ is $I^\pm(S)$:
$$\operatorname{int} J^\pm(S) = I^\pm(S). \tag{4.32}$$
- Both sets have the same closure:
$$\overline{J^\pm(S)} = \overline{I^\pm(S)}. \tag{4.33}$$
- It follows from (4.32) and (4.33) that both sets share the same boundary:
$$\partial J^\pm(S) = \partial I^\pm(S). \tag{4.34}$$

4.4.2 General definition of a black hole

We are now in position to give the general definition of a black hole. We shall do it for a spacetime $(\mathscr{M}, \boldsymbol{g})$ that admits a conformal completion at null infinity as defined in Sec. 4.3 and thus possesses a future null infinity $\mathscr{I}^+$. Moreover, we shall assume that $\mathscr{I}^+$ is ***complete***: if $\mathscr{I}^+$ is a null hypersurface, which occurs if $(\mathscr{M}, \boldsymbol{g})$ is asymptotically flat (cf. Sec. 4.3.2), this means that $\mathscr{I}^+$ is generated by complete[8] null geodesics. The completeness condition is imposed to avoid "spurious" black holes, such as black holes in Minkowski space (cf. Remark 2 below). The neighbourhood of $\mathscr{I}^+$ in $\tilde{\mathscr{M}}$ can then be considered as the infinitely far region reached by outgoing null geodesics. If a null geodesic does not reach this region, it can be considered as being trapped somewhere else in spacetime: this "somewhere else" constitutes the black hole region.

Let $(\mathscr{M}, \boldsymbol{g})$ be a spacetime with a conformal completion at null infinity such that $\mathscr{I}^+$ is complete; the ***black hole region***, or simply ***black hole***, is the set of points of $\mathscr{M}$ that are not in the causal past of the future null infinity:

$$\boxed{\mathscr{B} := \mathscr{M} \setminus \left(J^-(\mathscr{I}^+) \cap \mathscr{M}\right)}. \tag{4.35}$$

(cf. Fig. 4.8).

The black hole region is thus the set of points of $\mathscr{M}$ from which no future-directed causal curve in $\tilde{\mathscr{M}}$ reaches $\mathscr{I}^+$. Of course, it may be that $\mathscr{B} = \varnothing$, in which case one says that the spacetime $(\mathscr{M}, \boldsymbol{g})$ contains no black hole.

[7] This property is a direct consequence of Lemma 1 in Sec. 4.4.3 below.

[8] Let us recall that a *complete* geodesic is an inextendible (i.e. maximal) geodesic, whose affine parameters range through the whole of $\mathbb{R}$, cf. Sec. B.3.2 in Appendix B.

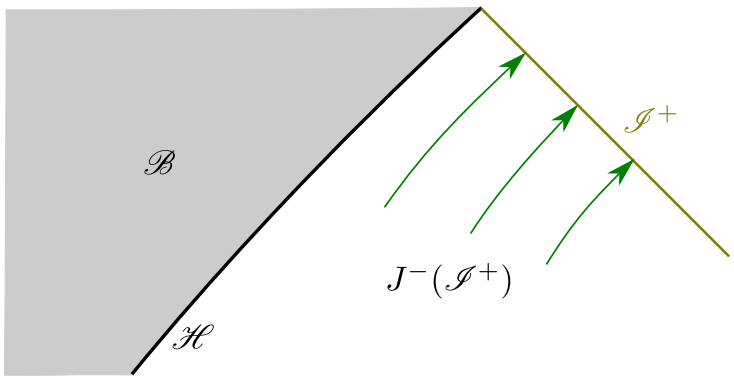


Figure 4.8: The black hole region $\mathscr{B}$ defined as the complement of the causal past of the future null infinity, $J^-(\mathscr{I}^+)$.

Example 4: The Minkowski spacetime contains no black hole, for all future-directed null geodesics terminate at $\mathscr{I}^+$ (cf. Fig. 4.5). More generally, any asymptotically simple spacetime contains no black hole.

Example 5: The prototype of a black hole is the *Schwarzschild black hole*; it will be shown in Sec. 6.4 that the Schwarzschild spacetime contains a region $\mathscr{B}$ that fulfills the above definition of a black hole. The corresponding event horizon is nothing but the Schwarzschild horizon considered in the examples of Chaps. 2 and 3.

Remark 2: If we release the assumption of $\mathscr{I}^+$-completeness in the above definition, we may end up with unphysical or "spurious" black holes. For instance, let us consider the conformal completion of Minkowski spacetime $(\mathscr{M}, \boldsymbol{g})$ resulting from its embedding in the Einstein cylinder $(\mathscr{E}, \tilde{\boldsymbol{g}})$, as in Sec. 4.2.3, keeping the same $\mathscr{I}^-$ but defining $\mathscr{I}^+$ as the hypersurface of $\mathscr{E}$ given by $\tau = \pi - \chi$ and $0 < \tau < \pi/2$, instead of $0 < \tau < \pi$ in Sec. 4.2.3. The manifold with boundary $\tilde{\mathscr{M}} := \mathscr{M} \cup \mathscr{I}^+ \cup \mathscr{I}^-$, equipped with the Einstein cylinder metric $\tilde{\boldsymbol{g}}$, is then a conformal completion of $(\mathscr{M}, \boldsymbol{g})$ at null infinity. With such a $\mathscr{I}^+$, the black hole region defined by (4.35) is non-empty, as shown in Fig. 4.9.

Remark 3: Some authors (in particular Hawking and Ellis [146]) define a *black hole* as a connected component of $S(\tau) \cap \mathscr{B}$, where $S(\tau)$ is a spacelike hypersurface that is a slice of the future development of a partial Cauchy surface[9] $S(0)$ such that the closure in $\tilde{\mathscr{M}}$ of the domain of dependence of $S(0)$ contains $\mathscr{I}^+$. According to such a definition, a black hole is a $(n-1)$-dimensional object, while the black hole $\mathscr{B}$ defined above is a n-dimensional object.

If $\mathscr{B} \neq \varnothing$, the boundary $\mathscr{H}$ of the black hole region is called the ***future event horizon*** (or simply the ***event horizon*** when no ambiguity may arise):

$$\boxed{\mathscr{H} := \partial\mathscr{B}}. \tag{4.36}$$

By plugging expression (4.35) for $\mathscr{B}$ in the standard identity $\partial\mathscr{B} = \overline{\mathscr{B}} \cap \overline{\mathscr{M} \setminus \mathscr{B}}$, we get an equivalent expression for $\mathscr{H}$:

$$\mathscr{H} = \overline{\mathscr{M} \setminus (J^-(\mathscr{I}^+) \cap \mathscr{M})} \cap \overline{(J^-(\mathscr{I}^+) \cap \mathscr{M})} = \partial(J^-(\mathscr{I}^+) \cap \mathscr{M}).$$

[9] The concepts of *partial Cauchy surface* and *future development* are defined in Sec. 10.8.3.

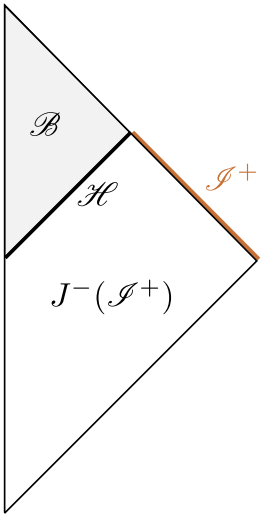


Figure 4.9: Spurious black hole region $\mathscr{B}$ in Minkowski spacetime resulting from a conformal completion with a non-complete $\mathscr{I}^+$. Compare with Fig. 4.5.

Now, the boundary of $J^-(\mathscr{I}^+)$ in $\tilde{\mathscr{M}}$ is $\partial J^-(\mathscr{I}^+) = \partial(J^-(\mathscr{I}^+) \cap \mathscr{M}) \cup \mathscr{I}^+$, so that $\partial J^-(\mathscr{I}^+) \cap \mathscr{M} = \partial(J^-(\mathscr{I}^+) \cap \mathscr{M})$; hence

$$\boxed{\mathscr{H} = \partial J^-(\mathscr{I}^+) \cap \mathscr{M}}. \tag{4.37}$$

In words: the future event horizon $\mathscr{H}$ is the part of the boundary of the causal past of the future null infinity $\mathscr{I}^+$ that lies in $\mathscr{M}$ (cf. Fig. 4.8). Note that thanks to identity (4.34), we can write as well

$$\mathscr{H} = \partial I^-(\mathscr{I}^+) \cap \mathscr{M}. \tag{4.38}$$

White hole

By inverting past and future in the black hole definition (4.35), one defines the ***white hole region*** of a spacetime $(\mathscr{M}, \boldsymbol{g})$ with a conformal completion at null infinity as the complement within $\mathscr{M}$ of the causal future of the past null infinity $\mathscr{I}^-$:

$$\boxed{\mathscr{W} := \mathscr{M} \setminus (J^+(\mathscr{I}^-) \cap \mathscr{M})}. \tag{4.39}$$

The white hole region is thus the set of points of $\mathscr{M}$ from which no past-directed causal curve in $\tilde{\mathscr{M}}$ reaches $\mathscr{I}^-$. The boundary of white hole region is called the ***past event horizon***:

$$\boxed{\mathscr{H}^- := \partial\mathscr{W} = \partial J^+(\mathscr{I}^-) \cap \mathscr{M}}. \tag{4.40}$$

Example 6: We shall see an example of white hole in the maximal extension of Schwarzschild spacetime, to be discussed in Chap. 9 (cf. Sec. 9.4.4).

The ***domain of outer communications*** is the part $\langle\langle\mathscr{M}\rangle\rangle$ of $\mathscr{M}$ that lies neither in the black hole region nor in the white hole one:

$$\langle\langle\mathscr{M}\rangle\rangle := \mathscr{M} \setminus (\mathscr{B} \cup \mathscr{W}) = \big(J^-(\mathscr{I}^+) \cap J^+(\mathscr{I}^-)\big) \cap \mathscr{M}. \tag{4.41}$$

The last equality, which is a direct consequence of the definitions of $\mathscr{B}$ and $\mathscr{W}$, shows that the domain of outer communications is the set of points from which it is possible to send a signal to and to receive a signal from arbitrary far regions. It also follows immediately from the definitions of the two event horizons that the boundary of the domain of outer communications is their union:

$$\partial\langle\langle\mathscr{M}\rangle\rangle = \mathscr{H} \cup \mathscr{H}^-. \tag{4.42}$$

Historical note: The term *event horizon* has been introduced by Wolfgang Rindler in 1956 [223] in the context of a single observer moving in some cosmological spacetime. Regarding the name *black hole*, the standard story is that it has been coined by John A. Wheeler in the end of 1967, following a suggestion shouted from the audience during one of his conference. However, a recent study [150] reveals that the expression *black hole* circulated as early as 1963 in the first Texas Symposium on Relativistic Astrophysics held in Dallas, while discussing the discovery of quasars, and could have been forged by Robert Dicke in some lecture given in 1961. The term *black hole* superseded the previous names *frozen star*, *collapsed star*, or *astre occlus* (the latter appearing along *black holes* in the title of the proceedings of the famous Les Houches summer school of 1972 [87]). The expression *domain of outer communications* is due to Brandon Carter (1971) [44].

4.4.3 Properties of the future event horizon

Having defined a black hole in full generality, let us derive the main properties of the black hole boundary: the event horizon $\mathscr{H}$.

Property 1:

$\mathscr{H}$ is an ***achronal set***, i.e. no pair of points of $\mathscr{H}$ can be connected by a timelike curve of $\mathscr{M}$.

Note that in the definition of an achronal set, it is not demanded that the timelike curve lies entirely in the set (for instance, the set can be discrete, so that no curve whatsoever lies in it). Accordingly, an equivalent statement of Property 1 is: no timelike curve of $\mathscr{M}$ encounters $\mathscr{H}$ at more than one point.

Proof. Let us assume the negation of Property 1, i.e. that there exists two points in $\mathscr{H}$ which are connected by a timelike curve $\mathscr{L}$. Let us call p and q these two points, with q in the future of p (cf. Fig. 4.10). We shall then use the following property:

Lemma 1: One can "move the ends" of any timelike curve "a little bit" and still get a timelike curve. More precisely, if two points $p, q \in \mathscr{M}$ are connected by a timelike curve, there exists a neighbourhood U of p and a neighbourhood V of q such that any point $p' \in U$ can be connected to any point $q' \in V$ by a timelike curve.

Proof of Lemma 1. This is more or less evident on a spacetime diagram (cf. Fig. 4.11) and a formal proof can be found as Lemma 3 in Chap. 14 of O'Neill's textbook [199]. □

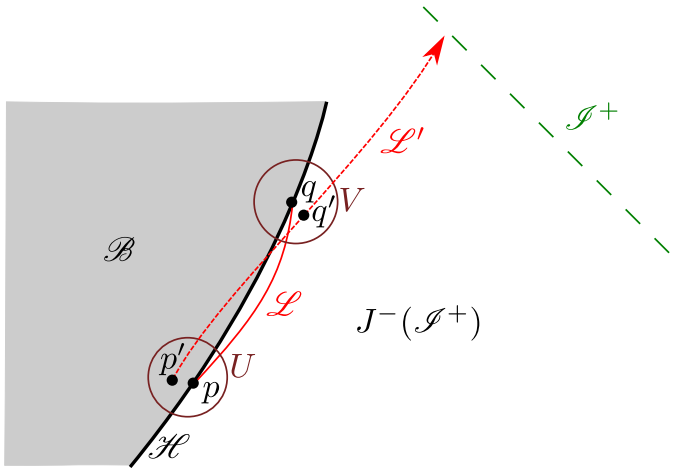


Figure 4.10: Proving that $\mathscr{H}$ is achronal.

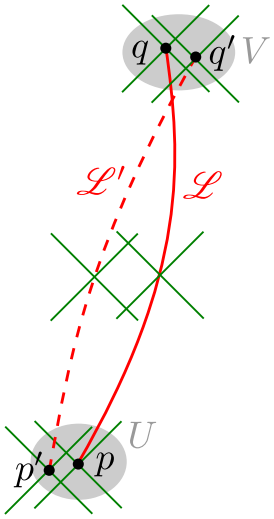


Figure 4.11: Lemma 1: moving slightly the ends p and q of a timelike curve $\mathscr{L}$ yields another timelike curve $\mathscr{L}'$.

Applying Lemma 1, let us choose $p' \in U \cap \mathscr{B}$ and $q' \in V \cap J^-(\mathscr{I}^+)$. Such a choice is always possible since p and q lie on the boundary between $\mathscr{B}$ and $J^-(\mathscr{I}^+)$ (cf. Fig. 4.10). Since $q' \in J^-(\mathscr{I}^+)$, the timelike curve linking p' and q' can then be extended in the future into a causal curve $\mathscr{L}'$ reaching $\mathscr{I}^+$. This implies $p' \in J^-(\mathscr{I}^+)$, which contradicts $p' \in \mathscr{B}$. □

Property 2:

$\mathscr{H}$ is a topological manifold of dimension $n-1$, n being the spacetime dimension.

Proof. Let $p \in \mathscr{H}$ and U some open neighbourhood of p where one can define a normal coordinate system (x^α). We have then $\boldsymbol{\partial}_0$ timelike, $\boldsymbol{\partial}_i$ spacelike for $i \in \{1, \dots, n-1\}$ and

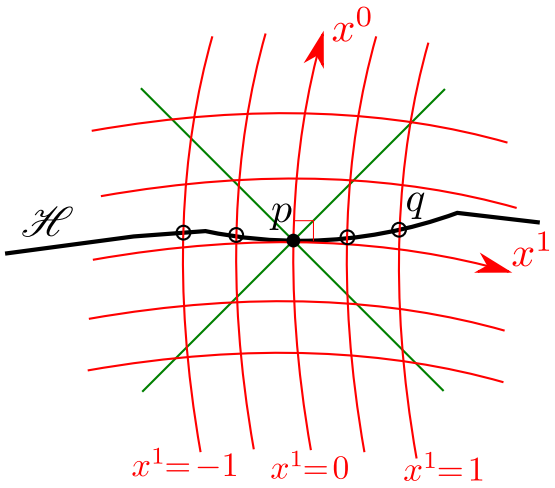


Figure 4.12: Proving that $\mathscr{H}$ is a topological manifold of dimension $n-1$.

$\boldsymbol{g}(\boldsymbol{\partial}_0, \boldsymbol{\partial}_i) = 0$. Let us consider a curve in U defined by $x^1 = a_1$,..., $x^{n-1} = a_{n-1}$, where a_1, ..., a_{n-1} are $n-1$ constants. This curve is timelike, since it has $\boldsymbol{\partial}_0$ as a tangent vector (cf. Fig. 4.12). It therefore intersects $\mathscr{H}$ at a single point q, for $\mathscr{H}$ is achronal (Property 1). Let us then give the coordinates $(y^i) = (a_1, \ldots, a_{n-1})$ to q. By varying $(a_1, \ldots, a_{n-1})$, we get a homeomorphism from $U \cap \mathscr{H}$ to an open subset of $\mathbb{R}^{n-1}$. □

Remark 4: Generically, the topological manifold $\mathscr{H}$ is not a smooth manifold, for it contains some points (the crossovers defined below) at which it is not differentiable. Actually $\mathscr{H}$ is slightly more than a mere topological submanifold of $\mathscr{M}$: it is a *Lipschitz submanifold* of $\mathscr{M}$. The latter is intermediate between a topological submanifold, i.e. a submanifold of class C^0 (continuous), and a differentiable submanifold of class C^1. On $U \cap \mathscr{H}$, the function x^0 is a Lipschitz function of the coordinates (y^i): $\left|x^0(y^i) - x^0(y'^i)\right| < K\sqrt{\sum_i (y^i - y'^i)^2}$. This follows from the achronal character of $\mathscr{H}$: the points of coordinates (y^i) and (y'^i) cannot have a too large separation in terms of x^0, otherwise they would be timelike separated. Hence, one says that $\mathscr{H}$ is a ***Lipschitz submanifold*** of $\mathscr{M}$. The notation C^{1-} (i.e. a kind of intermediate between C^0 and C^1) is generally used to denote Lipschitz submanifolds.

Property 3 (Penrose 1968 [209]):

$\mathscr{H}$ is ruled by a family of null geodesics that (i) either lie entirely in $\mathscr{H}$ or never leave $\mathscr{H}$ when followed into the future from the point where they arrived in $\mathscr{H}$, and (ii) have no endpoint in the future. Moreover, there is exactly one null geodesic through each point of $\mathscr{H}$, except at special points where null geodesics enter in contact with $\mathscr{H}$, which are called ***crossovers***. A special case of crossover, called ***caustic***, is a point where neighbouring null geodesics focus and converge while entering on $\mathscr{H}$.

In particular, once a null geodesic has merged with $\mathscr{H}$ (at a point where it may intersect other null geodesics), it will stay forever on $\mathscr{H}$ and will never intersect any other null

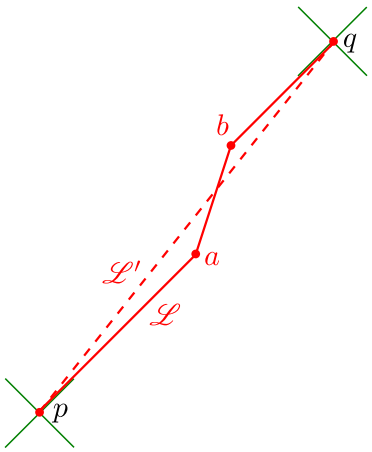


Figure 4.13: Lemma 2: A *causal* curve $\mathscr{L}$ containing a timelike segment (between a and b on the figure) can be deformed into a *timelike* curve $\mathscr{L}'$ with the ends kept fixed (dashed curve).

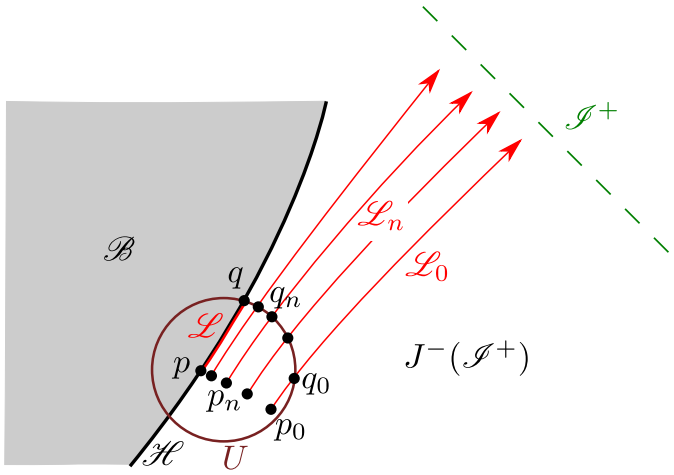


Figure 4.14: Causal curve $\mathscr{L}$ connecting p to q obtained as a limit of causal curves in $J^-(\mathscr{I}^+)$.

geodesic of the family ruling $\mathscr{H}$. These null geodesics are called the ***generators of*** $\mathscr{H}$. The set of all crossovers is called the ***crease set*** [233, 234, 32].

Proof. The following proof is adapted from that presented in Box 34.1 of MTW [190]. It relies on the following lemma:

Lemma 2: Let $\mathscr{L}$ be a causal curve connecting two points p and q of $\mathscr{M}$. If $\mathscr{L}$ contains a timelike segment, then there exists an entirely timelike curve connecting p and q.

Proof of Lemma 2. We shall provide only a graphical "proof", based on the spacetime diagram of Fig. 4.13. The causal curve $\mathscr{L}$ may have parts where it is null (segments pa and bq in Fig. 4.13); these parts are drawn with an angle of incline $\theta = \pm 45°$. If $\mathscr{L}$ contains a timelike segment (as ab in Fig. 4.13), i.e. a segment with $|\theta| > 45°$, it can be deformed, while keeping the same ends, to a curve with $|\theta| > 45°$ everywhere, i.e. to a timelike curve. □

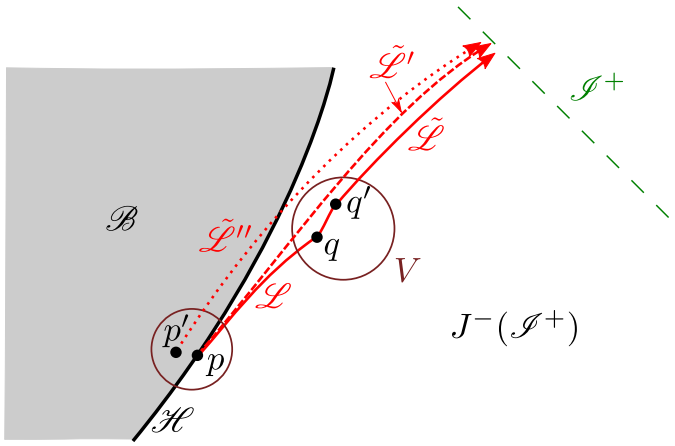


Figure 4.15: Proving by contradiction that q lies in $\mathscr{H}$.

Let $p \in \mathscr{H}$ and let U be some convex open neighbourhood of p. Since p lies in the boundary of $J^-(\mathscr{I}^+)$, it is always possible to consider a sequence of points $(p_n)_{n\in\mathbb{N}}$ converging toward p and such that $\forall n \in \mathbb{N}$, $p_n \in U \cap J^-(\mathscr{I}^+)$ (cf. Fig. 4.14). Since $p_n \in J^-(\mathscr{I}^+)$, there exists a future-directed causal curve $\mathscr{L}_n$ from p_n to $\mathscr{I}^+$ for each $n \in \mathbb{N}$. The neighbourhood U being convex, each $\mathscr{L}_n$ intersects its boundary ∂U at a unique point, q_n say: $\{q_n\} = \mathscr{L}_n \cap \partial U$ (cf. Fig. 4.14). Since ∂U is compact, the sequence $(q_n)_{n\in\mathbb{N}}$ admits a subsequence, $(q_{f(n)})_{n\in\mathbb{N}}$ say (f being an increasing function $\mathbb{N} \to \mathbb{N}$), that converges to some limit point q. Since from any point $p_{f(n)}$ arbitrary close to p, there is the causal curve $\mathscr{L}_{f(n)}$ to the point $q_{f(n)}$ arbitrary close to q, one can show that there exists a causal curve $\mathscr{L}$ connecting p to q (cf. Fig. 4.14; see e.g. Lemma 6.2.1 of Hawking & Ellis textbook [146] for a precise demonstration).

As the limit of points in $J^-(\mathscr{I}^+)$, q lies in the closure $\overline{J^-(\mathscr{I}^+)} = J^-(\mathscr{I}^+) \cup \mathscr{H}$, $\mathscr{H}$ being the boundary of $J^-(\mathscr{I}^+)$. Let us show by contradiction that actually $q \in \mathscr{H}$. If we assume $q \notin \mathscr{H}$, then necessarily $q \in J^-(\mathscr{I}^+)$. There exists then an open neighbourhood V of q such that $V \subset J^-(\mathscr{I}^+)$ (cf. Fig. 4.15). Let us choose $q' \in V$ such that q is connected to q' via a timelike curve. We may then extend $\mathscr{L}$ to a causal curve $\tilde{\mathscr{L}}$ from p to $\mathscr{I}^+$ via q and q' (cf. Fig. 4.15). Since $\tilde{\mathscr{L}}$ contains a timelike segment (between q and q'), we may invoke Lemma 2 to deform it into a timelike curve $\tilde{\mathscr{L}}'$ between p and $\mathscr{I}^+$. Then, by Lemma 1, one can "move the past end" of $\tilde{\mathscr{L}}'$ to get a new timelike curve $\tilde{\mathscr{L}}''$ linking an event $p' \in \mathscr{B}$ close to p to $\mathscr{I}^+$ (dotted curve in Fig. 4.15), which is impossible by the very definition of the black hole region $\mathscr{B}$. Hence $q \in \mathscr{H}$.

The causal curve $\mathscr{L}$ connecting p to q cannot be timelike since p and q are both in $\mathscr{H}$, which is achronal (Property 1). If $\mathscr{L}$ would contain a timelike segment, then by Lemma 2, it could be deformed into a timelike curve between p and q, which again would contradict the achronal character of $\mathscr{H}$. Hence $\mathscr{L}$ is necessarily a null curve. Moreover, it is a geodesic. Indeed, let us assume it is not. There is then some non-geodesic null segment of $\mathscr{L}$, ab say. Now, as shown in Sec. B.4.3 of Appendix B, a curve from a to b is a geodesic iff any of its parametrizations $P : [\lambda_a, \lambda_b] \to \mathscr{M}$, $\lambda \mapsto P(\lambda) \in \mathscr{L}$ is a stationary

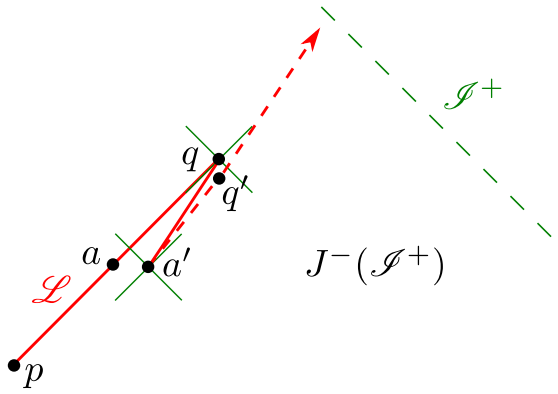


Figure 4.16: Proving that $\mathscr{L}$ lies entirely in $\mathscr{H}$.

point of the action

$$E_{(a,b)}(P) := \int_{\lambda_p}^{\lambda_q} \boldsymbol{g}(\boldsymbol{v},\boldsymbol{v})\,\mathrm{d}\lambda,$$

where $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ is the tangent vector associated with P. For the null segment ab of $\mathscr{L}$, we have $E_{(a,b)}(P) = 0$. Since ab is assumed to be not geodesic, it is not a stationary point of $E_{(a,b)}(P)$, which implies that there exists a nearby curve from a to b with $E_{(a,b)}(P) < 0$, i.e. there exists a curve from a to b with a timelike part. It follows that p and q can be connected by a causal curve with a timelike segment. By Lemma 2, this curve can be deformed into a timelike curve from p to q, which contradicts the achronal character of $\mathscr{H}$. Hence $\mathscr{L}$ is a null geodesic.

At this stage, we have shown that given $p \in \mathscr{H}$, there exists a future-directed null geodesic $\mathscr{L}$ connecting p to another point $q \in \mathscr{H}$. There remains to show that $\mathscr{L}$ lies entirely in $\mathscr{H}$. Let us start by showing that $\mathscr{L} \subset \overline{J^-(\mathscr{I}^+)}$. Let a be a generic point of $\mathscr{L}$ between p and q. Since $\mathscr{L}$ is null, there exists a point a' arbitrary close to a such that a' is connected to q by a future-directed timelike curve (cf. Fig. 4.16). Thanks to Lemma 1 and the property $q \in \overline{J^-(\mathscr{I}^+)}$, we may find a point $q' \in J^-(\mathscr{I}^+)$ close to q such that a' is connected to q' by a future-directed timelike curve. Since $q' \in J^-(\mathscr{I}^+)$, such a curve can be extended to a causal curve to $\mathscr{I}^+$ (the dashed curve in Fig. 4.16); hence $a' \in J^-(\mathscr{I}^+)$. Since a' is arbitrary close to a, we conclude that $a \in \overline{J^-(\mathscr{I}^+)}$. This property being valid for any point $a \in \mathscr{L}$, we have shown in fact that $\mathscr{L} \subset \overline{J^-(\mathscr{I}^+)} = J^-(\mathscr{I}^+) \cup \mathscr{H}$. Now it is easy to show that any point a of $\mathscr{L}$ actually lies in $\mathscr{H}$ by repeating exactly the same reasoning as that employed above to show that $q \in \mathscr{H}$, by replacing q by a. We therefore conclude that $\mathscr{L}$ lies entirely in $\mathscr{H}$.

Given a point $p \in \mathscr{H}$, we have thus constructed a future-directed null geodesic $\mathscr{L}$ lying entirely in $\mathscr{H}$ and connecting p to another point $q \in \mathscr{H}$. One can repeat the construction from the point q to get another future-directed null geodesic $\mathscr{L}' \subset \mathscr{H}$ connecting q to another point $q' \in \mathscr{H}$. Now $\mathscr{L}$ and $\mathscr{L}'$ must be two segments of the same null geodesic $\mathscr{L} \cup \mathscr{L}'$ by the following lemma:

Lemma 3: If from a point $q \in \mathscr{H}$, there exists a past-directed null geodesic $\mathscr{L} \subset \mathscr{H}$ and a future-directed null geodesic $\mathscr{L}' \subset \mathscr{H}$, then necessarily $\mathscr{L}$ and $\mathscr{L}'$ have collinear tangent

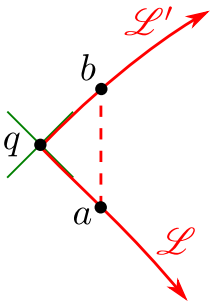


Figure 4.17: Proof of Lemma 3.

vectors at their common point q. It follows that $\mathscr{L}$ (with a time-reversed parametrization) and $\mathscr{L}'$ are two segments of a same null geodesic through q.

Proof of Lemma 3. Assume that $\mathscr{L}$ and $\mathscr{L}'$ have non-collinear tangent vectors at q. Then, in the vicinity of q, one can find a point $a \in \mathscr{L}$ and a point $b \in \mathscr{L}'$ such that a and b can be connected by a timelike curve (cf. Fig. 4.17). Since $\mathscr{L} \subset \mathscr{H}$ and $\mathscr{L}' \subset \mathscr{H}$, we have $a \in \mathscr{H}$ and $b \in \mathscr{H}$ and therefore we get a contradiction with $\mathscr{H}$ being achronal. □

Thanks to Lemma 3, we conclude that $\mathscr{L}'$ extends $\mathscr{L}$ to null geodesic $\mathscr{L} \cup \mathscr{L}'$ entirely lying in $\mathscr{H}$. By iterating, we conclude that the null geodesic $\mathscr{L}$ through p can be extended indefinitely into the future. Moreover, it can never leave $\mathscr{H}$. Indeed, if it leaves $\mathscr{H}$ at some point q, by the same procedure used above for p, one could construct a future-directed null geodesic $\mathscr{L}' \subset \mathscr{H}$ starting from q and Lemma 3 would imply that the extension of $\mathscr{L}$ outside $\mathscr{H}$ has to coincide with $\mathscr{L}'$, which is in contradiction with $\mathscr{L}' \subset \mathscr{H}$.

Another direct consequence of Lemma 3 is that no two distinct null generators may intersect at a point $p \in \mathscr{H}$, except if their segments in the past of p lie outside $\mathscr{H}$. This completes the proof of Property 3. □

Some features of Property 3 are illustrated in Fig. 4.18, which displays the null geodesic generators in a numerical simulation of the head-on collision of two black holes by Matzner et al. (1995) [185]. Note that new null geodesics enter the event horizon at the "crotch" of the "pair of pants".

The head-on black hole merger has been also computed by Cohen et al. (2009) [65], with an increase numerical accuracy (cf. Fig. 4.19). Cross-sections of the event horizon $\mathscr{H}$ (cf. Sec. 2.3.4) are depicted in Fig. 4.20. The same figure shows also how some null geodesics will reach $\mathscr{H}$ to become null generators.

Finally, Fig. 4.21 shows a cross-section of the event horizon computed by Cohen et al. (2012) [64] in some inspiralling binary black hole merger. The black hole spacetime itself has been computed as a solution of the vacuum Einstein equation by Scheel at al. [227]; it corresponds to 16 inspiralling orbits of a equal-mass binary black hole with vanishing initial spins.

Generically, for a binary black hole merger, the crease set forms a 2-dimensional subset of the event horizon $\mathscr{H}$ and is bounded by the set of caustic points, which forms a 1-dimensional subset of $\mathscr{H}$ [233, 234, 154, 64].

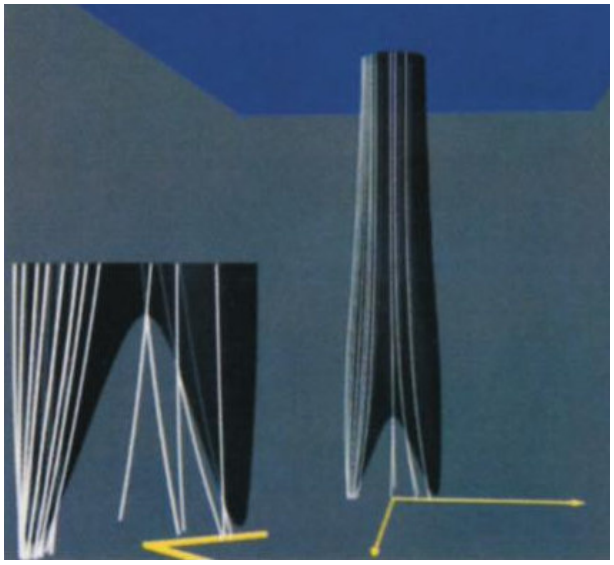


Figure 4.18: Spacetime diagram of the event horizon corresponding to the head-on merger of two black holes as computed by Matzner et al. (1995) [185]. The white curves are some null geodesic generators; the left picture is a zoom of the merger region, with the crease set (source: Fig. 4 of Ref. [185]; ©1995 American Association for the Advancement of Science).

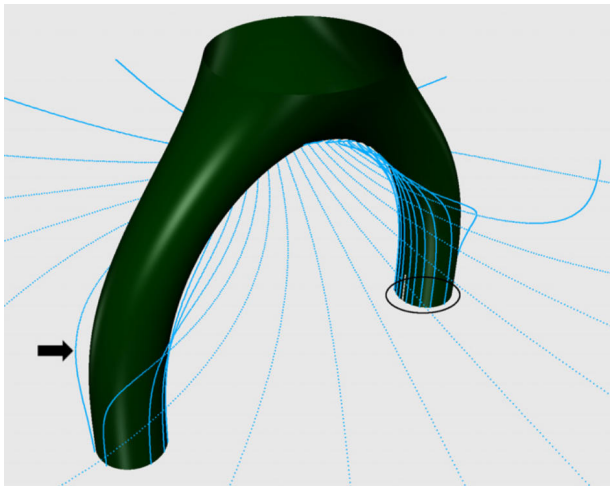


Figure 4.19: Spacetime diagram showing the event horizon in the head-on merger of two black holes, as computed by Cohen et al. (2009) [65]. The blue curves are null geodesics that will eventually become null generators of the event horizon; those arising from regions close to the event horizon are marked by the arrow and the black ellipse (source: Fig. 15 of Ref. [65]; ©2009 IOP Publishing Ltd).

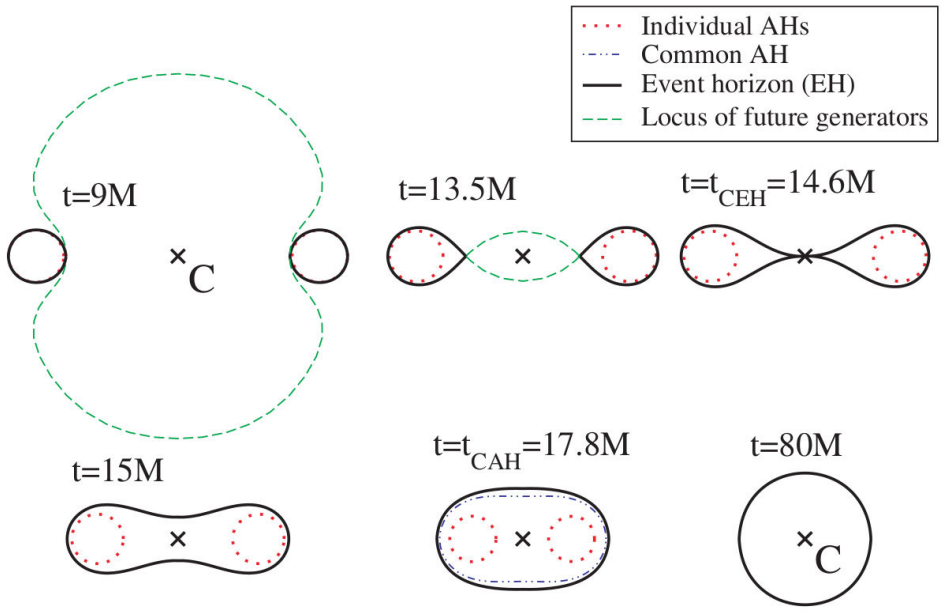


Figure 4.20: Cross-sections (at various coordinate times t) of the event horizon $\mathscr{H}$ corresponding to the head-on merger of two black holes as computed by Cohen et al. (2009) [65] and displayed in Fig. 4.19. Each figure is a 2D cut of a hypersurface Σ_t defined by a constant value of the coordinate time t, expressed in units of the sum M of the initial irreducible masses of each black hole (cf. Sec. ??). The whole 3D hypersurface Σ_t can be reconstructed by rotation around the collision axis. t_{CEH} (for "Common Event Horizon") is the coordinate time at which the cross-section of $\mathscr{H}$ becomes a connected 2-surface. The cross-sections of $\mathscr{H}$ are displayed in black, while the green dashed curves denote the set of the intersections with Σ_t of the null geodesics that will become null generators of $\mathscr{H}$ through the cusps in the "individual" event horizons. The red and blue dashed curves denotes apparent horizons (cf. Sec. ??). (source: Fig. 1 of Ref. [65]; ©2009 IOP Publishing Ltd).

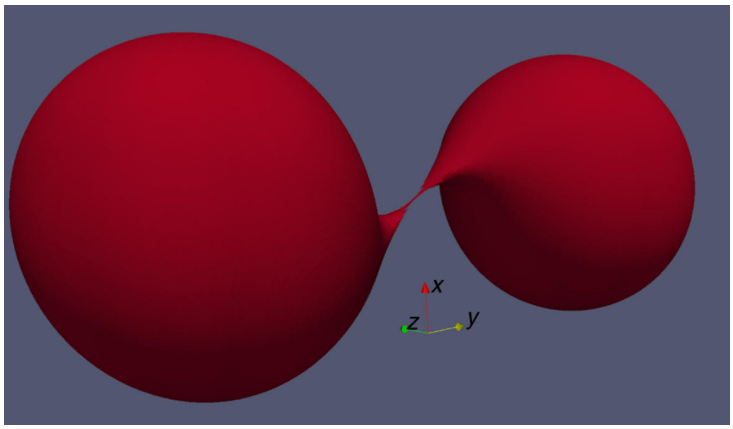


Figure 4.21: Cross-section of the event horizon $\mathscr{H}$ of the inspiralling merger of two black holes as computed by Cohen et al. (2012) [64]. The x and y axes define the orbital plane. This cross-section is the first connected one in the slicing of $\mathscr{H}$ by surfaces of constant coordinate time t (source: Fig. 2 of Ref. [64]; ©2012 American Physical Society).

Property 4:

Wherever it is smooth, $\mathscr{H}$ is a null hypersurface.

Proof. Let us assume that $\mathscr{H}$ is smooth in some open subset U. By Property 2, $\mathscr{H}$ is then a smooth hypersurface in U. According to Property 3, there is a null geodesic lying in $\mathscr{H}$ through any point of $\mathscr{H} \cap U$. This implies null tangent vectors at any point of $\mathscr{H} \cap U$, so that, in U, $\mathscr{H}$ must be either a null hypersurface or a timelike one. But $\mathscr{H}$ is achronal by Property 1 and therefore cannot be timelike. Hence, $\mathscr{H}$ is a null hypersurface in U. □

When $\mathscr{H}$ is smooth, its generators, as defined in Property 3, are then nothing but the null-hypersurface generators as defined in Sec. 2.3.3.

Remark 5: Properties 1 to 4 are not specific to black hole horizons: they are actually valid for any boundary $\partial J^-(S)$ of the causal past of a given set $S \subset \mathscr{M}$. They are also valid for the boundary $\partial J^+(S)$ of the causal future of S, modulo the relevant changes future $\leftrightarrow$ past in Property 3.

It can be shown that event horizons are smooth almost everywhere: the only location where they are not differentiable is the crease set, i.e. the set of points where null geodesics cross each other while arriving at $\mathscr{H}$ and becoming null generators.

Chapter 5

Stationary black holes

Contents

5.1 Introduction

5.2 Definition and first properties

5.2.1 Stationarity and staticity

We shall define a spacetime $(\mathscr{M}, \boldsymbol{g})$ to be ***stationary*** iff (i) it is invariant under the action of the translation group $(\mathbb{R}, +)$ and (ii) the orbits of the group action are everywhere timelike curves or (ii') $(\mathscr{M}, \boldsymbol{g})$ admits a conformal completion (cf. Sec. 4.3) and the orbits of the group action are timelike in the vicinity of the conformal boundary $\mathscr{I}$. It is equivalent to say that there exists a Killing vector field $\boldsymbol{\xi}$ (the generator of the translation group, cf. Sec. 3.3.1) that is timelike everywhere or at least in the vicinity of $\mathscr{I}$ when there exists a conformal completion.

Remark 1: Some authors (e.g. Carter [46]) call such spacetimes *pseudo-stationary*, keeping the qualifier *stationary* for the case where the Killing field $\boldsymbol{\xi}$ is timelike in all $\mathscr{M}$, i.e. obeying the point (ii) above. As we are going to see, when $\mathscr{M}$ contains a black hole, $\boldsymbol{\xi}$ cannot be timelike everywhere, so only *pseudo-stationarity* in the above sense is relevant for them. Our terminology follows that of Chruściel, Lopes Costa & Heusler [60] and Choquet-Bruhat [51].

A stronger notion than stationarity is that of *staticity*: an asymptotically flat spacetime $(\mathscr{M}, \boldsymbol{g})$ is called ***static*** iff (i) it is stationary and (ii) the Killing vector field $\boldsymbol{\xi}$

generating the stationary action is orthogonal to a family of hypersurfaces (one says that $\boldsymbol{\xi}$ is ***hypersurface-orthogonal***). Via the Frobenius theorem (cf. Sec. 2.3.3), the hypersurface-orthogonal condition is equivalent to the existence of a 1-form $\boldsymbol{\omega}$ such that

$$\mathbf{d}\underline{\boldsymbol{\xi}} = \boldsymbol{\omega} \wedge \underline{\boldsymbol{\xi}}, \tag{5.1}$$

where $\underline{\boldsymbol{\xi}}$ is the metric dual of $\boldsymbol{\xi}$ (cf. Sec. A.3.3). Equation (5.1) is equivalent to

$$\underline{\boldsymbol{\xi}} \wedge \mathbf{d}\underline{\boldsymbol{\xi}} = 0, \tag{5.2}$$

or, in terms of components (expressing the exterior derivative $\mathbf{d}\underline{\boldsymbol{\xi}}$ in terms of the spacetime Levi-Civita connection $\boldsymbol{\nabla}$):

$$\xi_{[\alpha}\nabla_{\beta}\xi_{\gamma]} = 0. \tag{5.3}$$

Then, one can show (see e.g. Sec. 2.9 of Straumann's textbook [237]), that locally, there exists a coordinate system $(x^\alpha) = (t, x^1, \ldots, x^{n-1})$ such that

$$\boldsymbol{\xi} = \boldsymbol{\partial}_t \qquad \text{and} \qquad \underline{\boldsymbol{\xi}} = (\boldsymbol{\xi}\cdot\boldsymbol{\xi})\,\mathbf{d}t. \tag{5.4}$$

The second relation implies that $\boldsymbol{\xi}$ is orthogonal to the hypersurfaces $t = \text{const}$. This orthogonality property translates to the metric components in the coordinates (x^α) as $g_{0i} = 0$ for $i \in \{1, \ldots, n-1\}$, so that one may write

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = V\mathrm{d}t^2 + g_{ij}\mathrm{d}x^i\mathrm{d}x^j, \tag{5.5}$$

where $V = \boldsymbol{\xi}\cdot\boldsymbol{\xi}$ and g_{ij} are functions of $(x^1, \ldots, x^{n-1})$ only. It is clear that the metric (5.5) is invariant[1] in the transformation $t \mapsto -t$. One says that a static spacetime is ***time-reflection symmetric***.

Remark 2: The same comment as in Remark 1 above can be made: some authors would call *static spacetimes* only those in which the hypersurface-orthogonal Killing vector is everywhere timelike, whereas our definition is weaker, the timelike character being demanded only in the vicinity of the spacetime "boundary", i.e. the boundary of a conformal completion.

5.2.2 Black holes in stationary spacetimes

Let us consider a spacetime $(\mathscr{M}, \boldsymbol{g})$ that contains a black hole, as defined in Sec. 4.4.2. In particular, $(\mathscr{M}, \boldsymbol{g})$ admits a future null infinity $\mathscr{I}^+$ and a past null infinity $\mathscr{I}^-$. Furthermore, we assume that $(\mathscr{M}, \boldsymbol{g})$ is stationary, in the sense defined above. Since $(\mathscr{M}, \boldsymbol{g})$ is invariant under the action of the isometry group $(\mathbb{R}, +)$, so is $\mathscr{I}^+$ (under some proper extension of $\boldsymbol{\xi}$ to the conformal completion $\tilde{\mathscr{M}}$) and therefore its causal past $J^-(\mathscr{I}^+)$. As the boundary of $J^-(\mathscr{I}^+)$ inside $\mathscr{M}$, the event horizon $\mathscr{H}$ must therefore be invariant under the action of the isometry group. Note that this means that $\mathscr{H}$ is invariant *as a whole*, not that each point of $\mathscr{H}$ is invariant (i.e. is a fixed point) under the group action. Let us assume that $\mathscr{H}$ is smooth (which sounds likely in a stationary context; a rigorous proof can be found in [58]), it is then a null hypersurface (Property 4 in Sec. 4.4.3). Now, $\mathscr{H}$ is globally invariant if, and only if, the generator $\boldsymbol{\xi}$ of the isometry group is tangent to $\mathscr{H}$. Since a timelike vector cannot be tangent to a null hypersurface (cf. the lemma in Sec. 2.3.4), we conclude that

[1] Would (5.5) have contained a non-vanishing $g_{0i}\,\mathrm{d}t\,\mathrm{d}x^i$ term, this would not have been the case.

In a stationary spacetime containing a black hole, the stationary Killing vector field $\boldsymbol{\xi}$ is tangent to the event horizon $\mathscr{H}$, which implies that $\boldsymbol{\xi}$ is either null or spacelike on $\mathscr{H}$.

5.3 The event horizon as a Killing horizon

Let us discuss successively the two allowed types for the stationary Killing vector $\boldsymbol{\xi}$ on $\mathscr{H}$: null and spacelike.

5.3.1 Null stationary Killing field on $\mathscr{H}$: the staticity theorem

By the lemma of Sec. 2.3.4, if the Killing vector field $\boldsymbol{\xi}$ is null on $\mathscr{H}$, it is necessarily tangent to the null geodesic generators of $\mathscr{H}$ and therefore collinear to the null normals $\boldsymbol{\ell}$ of $\mathscr{H}$. From the definition given in Sec. 3.3.2, it follows immediately that $\mathscr{H}$ is a Killing horizon (with respect to the Killing field $\boldsymbol{\xi}$). In dimension $n = 4$ and using the Einstein equation, D. Sudarsky and R.M. Wald (1992) [239] have then proven that $\boldsymbol{\xi}$ must be hypersurface-orthogonal not only on $\mathscr{H}$, but everywhere, i.e. that the spacetime $(\mathscr{M}, \boldsymbol{g})$ is *static*, according to the above definition. For this reason, Sudarski & Wald's result is often called the ***staticity theorem***.

Having that $(\mathscr{M}, \boldsymbol{g})$ is static, we can go further and apply the

Israel uniqueness theorem:

If $(\mathscr{M}, \boldsymbol{g})$ is a n-dimensional static spacetime containing a black hole, with $\boldsymbol{g}$ solution of the vacuum Einstein equation, then the domain of outer communications of $\mathscr{M}$ is isometric to the domain of outer communications of a n-dimensional Schwarzschild spacetime.

This theorem has been proved in 1967 by W. Israel [155], and improved latter by many authors, in particular by P. Chruściel & G. Galloway (2010) [59], who removed the hypothesis of analyticity (cf. Remark 3 in Sec. A.2.1). A demonstration of Israel's theorem can be found in Sec. 8.2 of Straumann's textbook [237].

So basically, in dimension $n = 4$ (i.e. when the staticity theorem applies), all stationary vacuum black holes with the stationary Killing field $\boldsymbol{\xi}$ null on $\mathscr{H}$ are nothing but Schwarzschild black holes, which we will study in detail in Chaps. 6 and 9.

5.3.2 Spacelike stationary Killing field on $\mathscr{H}$: the strong rigidity theorem

When $\boldsymbol{\xi}$ is spacelike on $\mathscr{H}$, it obviously cannot be collinear to any null normal $\boldsymbol{\ell}$ of $\mathscr{H}$. Assuming that $\mathscr{H}$ has cross-sections of spherical topology, we observe that, with respect to the null geodesic generators of $\mathscr{H}$, the field lines of $\boldsymbol{\xi}$ form some helices, as depicted in Fig. 5.1a. By reciprocity, with respect to the field lines of $\boldsymbol{\xi}$, the null geodesic generators

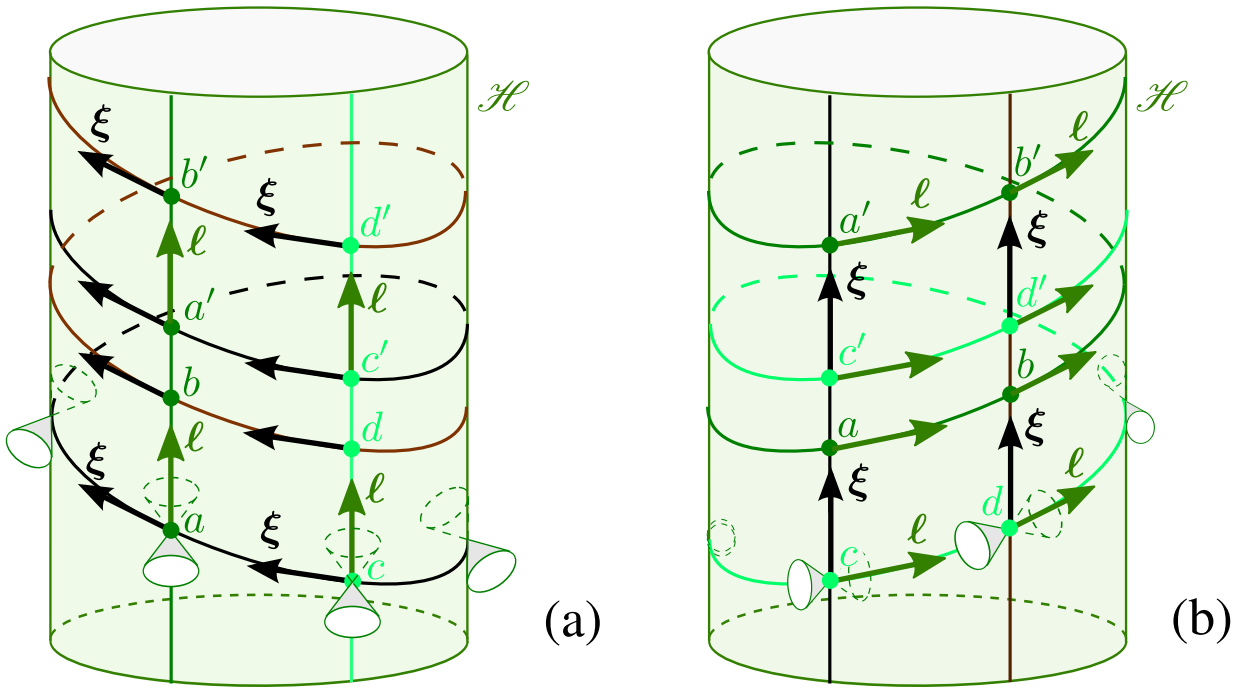


Figure 5.1: Event horizon $\mathscr{H}$ with a stationary Killing vector field $\boldsymbol{\xi}$ spacelike on it: **(a)** Representation with the null geodesic generators of $\mathscr{H}$ drawn as vertical lines; two of them are actually depicted, in dark green and light green respectively, with a null normal $\boldsymbol{\ell}$ along them; besides, two field lines of $\boldsymbol{\xi}$ (orbits of the isometry group) are depicted, in black and brown respectively. **(b)** Representation with the field lines of $\boldsymbol{\xi}$ as vertical lines. The color code is the same as in (a) and labelled points (a, b, etc.) help to identify the two figures. A few light cones are drawn in each figure; note that $\boldsymbol{\xi}$, being spacelike, is always outside of them, while the null normal $\boldsymbol{\ell}$ is always tangent to them.

form some helices as well, as depicted in Fig. 5.1b): observe that Fig. 5.1b can be obtained from Fig. 5.1a by "untwisting" the field lines of $\boldsymbol{\xi}$.

Since asymptotically the field lines of $\boldsymbol{\xi}$ are worldlines of inertial observers, Fig. 5.1b leads us to say (in loose terms at this stage) that the event horizon $\mathscr{H}$ "is rotating", all the more that we have seen above that when the null generators coincide with the field lines of $\boldsymbol{\xi}$, the black hole is static, i.e. non-rotating.

Since the Killing field $\boldsymbol{\xi}$ is not null on $\mathscr{H}$, we cannot say a priori that $\mathscr{H}$ is a Killing horizon. However, it turns out that this is indeed the case, according to a famous result by S.W. Hawking (1972) [144, 146], known as the ***strong rigidity theorem***. Assuming $n = 4$ and the metric $\boldsymbol{g}$ obeying the vacuum Einstein equation, Hawking was able to show that there exists a second Killing vector field, $\boldsymbol{\chi}$ say, which is null on $\mathscr{H}$. Hence $\mathscr{H}$ is a Killing horizon in this case as well, albeit not with respect to the stationary Killing vector field $\boldsymbol{\xi}$.

Hawking's result has been extended to dimensions $n \geq 4$ by V. Moncrief and J. Isenberg (2008) [191], under the hypotheses that $\mathscr{H}$ has cross-sections that are compact and transverse to $\boldsymbol{\xi}$ (see also Theorem 8.1 p. 470 of Choquet-Bruhat's textbook [51]). Both Hawking's result and Moncrief & Isenberg's one rely on the rather strong assumption that $\mathscr{M}$ and $\mathscr{H}$ are (real) *analytic* manifolds (cf. Remark 3 in Sec. A.2.1), with $\boldsymbol{g}$ being an

analytic field. On physical grounds, it would be desirable to assume only *smooth* manifolds and fields. Recently, S. Alexakis, A.D. Ionescu and S. Klainerman [9] (2014) have succeeded in proving the strong rigidity theorem without the analyticity assumption, but only for slowly rotating black holes.

Since we have two Killing vectors, $\boldsymbol{\xi}$ and $\boldsymbol{\chi}$, we may form any linear combination of them with constant coefficients and still get a Killing vector. For instance, if Ω_H is a non-zero constant, the vector field $\boldsymbol{\eta}$ defined by

$$\boldsymbol{\eta} = \frac{1}{\Omega_H}(\boldsymbol{\chi} - \boldsymbol{\xi}) \quad \Longleftrightarrow \quad \boldsymbol{\chi} = \boldsymbol{\xi} + \Omega_H \boldsymbol{\eta}, \tag{5.6}$$

is a Killing vector field on $\mathscr{M}$. One can show (see e.g. [54] for a rigorous proof) that Ω_H and some constant rescaling of $\boldsymbol{\chi}$ can be chosen so that $\boldsymbol{\eta}$ is a spacelike vector field whose field lines are closed, with 2π-periodicity in terms of the parameter φ associated to $\boldsymbol{\eta}$ (i.e. $\boldsymbol{\eta} = \mathrm{d}/\mathrm{d}\varphi$ along the field lines), and such that $\boldsymbol{\eta}$ vanishes on a timelike 2-dimensional surface, called the ***rotation axis***. It follows that the isometry group whose generator is $\boldsymbol{\eta}$ is the rotation group SO(2). In other words, the spacetime $(\mathscr{M}, \boldsymbol{g})$ is ***axisymmetric*** in addition to be stationary. The constant Ω_H is then called the ***black hole rotation velocity***.

By the very definition of stationarity, the Killing vector field $\boldsymbol{\xi}$ is timelike in the vicinity of $\mathscr{I}^+$ and $\mathscr{I}^-$. If $\boldsymbol{\xi}$ is spacelike on $\mathscr{H}$, as assumed in this section, by continuity it must be spacelike in some part of the domain of outer communications $\langle\langle \mathscr{M} \rangle\rangle$ near $\mathscr{H}$. The simplest configuration is then when $\boldsymbol{\xi}$ is spacelike in some connected region $\mathscr{G} \subset \langle\langle \mathscr{M} \rangle\rangle$ around $\mathscr{H}$, null at the boundary of $\mathscr{G}$ and timelike outside $\mathscr{G}$ up to $\mathscr{I}^+$ and $\mathscr{I}^-$. The subset $\mathscr{G}$ is called the ***ergoregion*** and its boundary $\mathscr{E} := \partial\mathscr{G}$ the ***ergosphere***. We shall discuss it further in connection with the Penrose process in Chap. 10.

5.4 Bifurcate Killing horizons

5.4.1 Definition and first properties

Let $(\mathscr{M}, \boldsymbol{g})$ be a n-dimensional spacetime endowed with a Killing vector field $\boldsymbol{\xi}$. A ***bifurcate Killing horizon*** is the union

$$\mathscr{H} = \mathscr{H}_1 \cup \mathscr{H}_2, \tag{5.7}$$

such that

- $\mathscr{H}_1$ and $\mathscr{H}_2$ are two null hypersurfaces;
- $\mathscr{S} := \mathscr{H}_1 \cap \mathscr{H}_2$ is a spacelike $(n-2)$-surface;
- each of the sets $\mathscr{H}_1 \setminus \mathscr{S}$ and $\mathscr{H}_2 \setminus \mathscr{S}$ has two connected components, which are Killing horizons[a] with respect to $\boldsymbol{\xi}$.

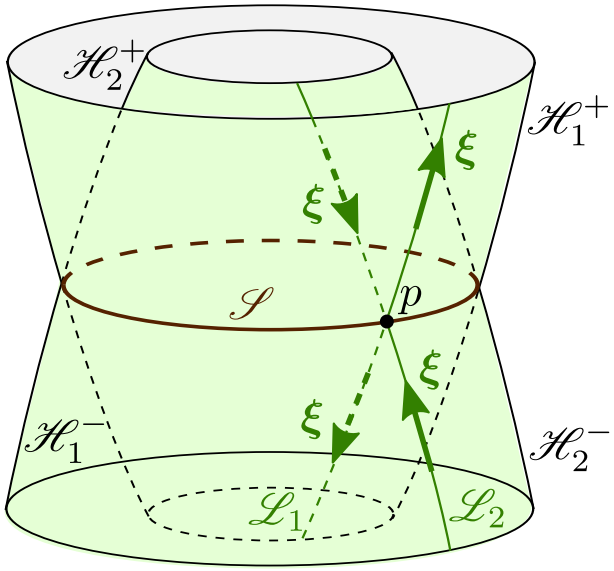


Figure 5.2: Bifurcate Killing horizon $\mathscr{H}_1 \cup \mathscr{H}_2$ with respect to the Killing vector field $\boldsymbol{\xi}$; $\mathscr{S}$ is the bifurcation surface. $\mathscr{L}_1$ and $\mathscr{L}_2$ are null geodesic generators of respectively $\mathscr{H}_1$ and $\mathscr{H}_2$, which cross each other at the point $p \in \mathscr{S}$.

The $(n-2)$-dimensional submanifold $\mathscr{S}$ is called the ***bifurcation surface*** of $\mathscr{H}$.

[a] Cf. Sec. 3.3.2 for the definition of a Killing horizon.

Hence we may say that a bifurcate Killing horizon is formed by four Killing horizons, $\mathscr{H}_1^+$, $\mathscr{H}_1^-$, $\mathscr{H}_2^+$ and $\mathscr{H}_2^-$ say, which are merged together at the bifurcation surface $\mathscr{S}$ (cf. Fig. 5.2), in such a way that

$$\mathscr{H}_1 = \mathscr{H}_1^- \cup \mathscr{S} \cup \mathscr{H}_1^+ \quad \text{and} \quad \mathscr{H}_2 = \mathscr{H}_2^- \cup \mathscr{S} \cup \mathscr{H}_2^+$$

are null hypersurfaces.

A first property of bifurcate Killing horizons is

The Killing vector field vanishes on the bifurcation surface:

$$\boxed{\boldsymbol{\xi}|_{\mathscr{S}} = 0}. \tag{5.8}$$

Proof. Let $p \in \mathscr{S}$ and let us assume that $\boldsymbol{\xi}|_p \neq 0$. Let $\mathscr{L}_1$ (resp. $\mathscr{L}_2$) be the null geodesic generator of $\mathscr{H}_1$ (resp. $\mathscr{H}_2$) that intersects $\mathscr{S}$ at p (cf. Fig. 5.2). Since $\mathscr{S}$ is spacelike, $\mathscr{L}_1$ and $\mathscr{L}_2$ are unique. By definition of a Killing horizon, $\boldsymbol{\xi}$ is tangent to $\mathscr{L}_1 \cap \mathscr{H}_1^+$ and to $\mathscr{L}_1 \cap \mathscr{H}_1^-$, i.e. to $\mathscr{L}_1 \setminus \{p\}$. If $\boldsymbol{\xi}|_p \neq 0$, then by continuity, $\boldsymbol{\xi}$ is a (non-vanishing) tangent vector field all along $\mathscr{L}_1$. Similarly, $\boldsymbol{\xi}$ is tangent to all $\mathscr{L}_2$. At their intersection point p, the geodesics $\mathscr{L}_1$ and $\mathscr{L}_2$ have thus a common tangent vector, namely $\boldsymbol{\xi}|_p$. The geodesic uniqueness theorem (cf. Sec. B.3.2 in Appendix B) yields then $\mathscr{L}_1 = \mathscr{L}_2$. Then $\mathscr{L}_1 \subset \mathscr{H}_1 \cap \mathscr{H}_2 = \mathscr{S}$. But since $\mathscr{S}$ is spacelike and $\mathscr{L}_1$ is null, we reach a contradiction. Hence we must have $\boldsymbol{\xi}|_p = 0$. □

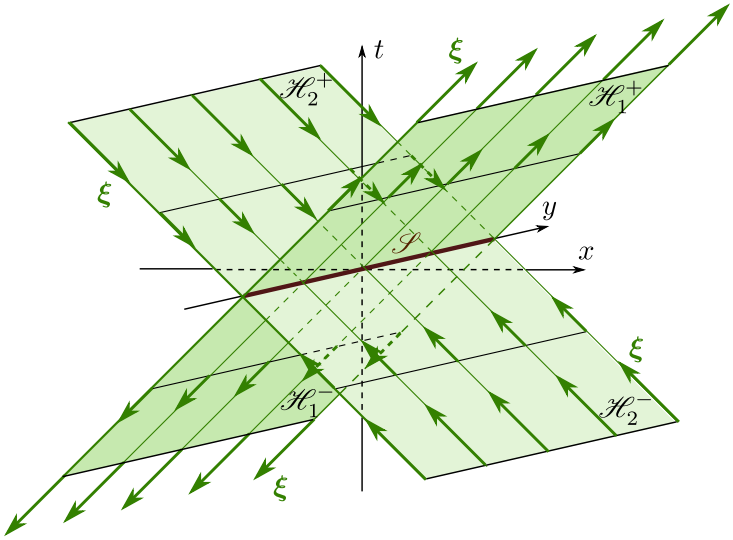


Figure 5.3: Bifurcate Killing horizon $\mathscr{H}_1 \cup \mathscr{H}_2$ with respect to the Killing vector field $\boldsymbol{\xi}$ generating Lorentz boosts in the plane (t, x) of Minkowski spacetime. The dimension along z having been suppressed, the bifurcation surface $\mathscr{S}$ appears as a line, while it is actually a 2-plane.

Remark 1: Having a Killing vector field that vanishes somewhere (here $\mathscr{S}$) is not the sign of any pathology: it simply means that the points of $\mathscr{S}$ are fixed points of the isometries generated by $\boldsymbol{\xi}$, since setting $\boldsymbol{\xi} = 0$ in Eq. (3.20) leads to $\mathrm{d}\boldsymbol{x} = 0$, i.e. to $\Phi_{\mathrm{d}t}(p) = p$.

Remark 2: Contrary to what the name may suggest, a bifurcate Killing horizon is *not* a Killing horizon, for the latter, as defined in Sec. 3.3.2, is a regular (i.e. embedded) hypersurface of $\mathscr{M}$ (cf. Sec. A.2.7 in Appendix A), while the union of two hypersurfaces is not in general a hypersurface. Moreover on a Killing horizon, the Killing vector field is nowhere vanishing [cf. Eq. (3.28)], while on a bifurcate Killing horizon, it is vanishing at the bifurcation surface.

Example 1 (bifurcate Killing horizon w.r.t. Lorentz boost): Let us consider the boost Killing vector in Minkowski spacetime as given by Eq. (3.30): $\boldsymbol{\xi} := x\boldsymbol{\partial}_t + t\boldsymbol{\partial}_x$ and let us take for $\mathscr{H}_1$ the null hyperplane of equation $t = x$ considered in Example 7 in Chap. 3 and denoted there by $\mathscr{H}$. The two half-hyperplanes defined by Eq. (3.31) are then the Killing horizons $\mathscr{H}_1^+$ and $\mathscr{H}_1^-$. The union $\mathscr{H}_1 \cup \mathscr{H}_2$, where $\mathscr{H}_2$ is the null hyperplane of equation $t = -x$ is a bifurcate Killing horizon with respect to $\boldsymbol{\xi}$, with the 2-plane of equation $t = 0$ and $x = 0$ as bifurcation surface (cf. Fig. 5.3). Note that on $\mathscr{H}_1$, the Killing vector $\boldsymbol{\xi}$ points away from $\mathscr{S}$, while on $\mathscr{H}_2$, it points towards $\mathscr{S}$.

5.4.2 Non-degenerate Killing horizons and Boyer theorem

Let us consider a Killing horizon $\mathscr{H}$ with respect to some Killing vector field $\boldsymbol{\xi}$. As shown in Sec. 3.3.5, modulo some mild energy condition (the null dominant energy condition), the surface gravity of $\mathscr{H}$, i.e. the non-affinity coefficient κ of $\boldsymbol{\xi}$ on $\mathscr{H}$, is constant over $\mathscr{H}$ (the zeroth law of black hole mechanics). In what follows, we consider the case where $\kappa \neq 0$, i.e. $\mathscr{H}$ is a non-degenerate Killing horizon (cf. Sec. 3.3.6).

Let us assume that $\boldsymbol{\xi}$ is future-directed on $\mathscr{H}$; if not, we can always consider $\mathscr{H}$ as a Killing horizon with respect to $-\boldsymbol{\xi}$. Let t be the parameter associated with $\boldsymbol{\xi}$ along the null geodesic generators of $\mathscr{H}$, i.e. $\boldsymbol{\xi} = \mathrm{d}/\mathrm{d}t$ along any null geodesic generator $\mathscr{L}$. Since $\kappa \neq 0$, t is not an affine parameter of $\mathscr{L}$. The null vector field $\boldsymbol{\ell}$ defined on $\mathscr{H}$ by

$$\boldsymbol{\ell} = \mathrm{e}^{-\kappa t}\, \boldsymbol{\xi} \quad \Longleftrightarrow \quad \boldsymbol{\xi} = \mathrm{e}^{\kappa t}\, \boldsymbol{\ell} \tag{5.9}$$

is a geodesic vector field and the affine parameter associated with it is

$$\lambda = \frac{\mathrm{e}^{\kappa t}}{\kappa} + \lambda_0, \tag{5.10}$$

where λ_0 is some constant.

Proof. We have

$$\boldsymbol{\nabla}_{\boldsymbol{\ell}}\, \boldsymbol{\ell} = \boldsymbol{\nabla}_{\mathrm{e}^{-\kappa t}\boldsymbol{\xi}} \left(\mathrm{e}^{-\kappa t}\, \boldsymbol{\xi}\right) = \mathrm{e}^{-\kappa t} \boldsymbol{\nabla}_{\boldsymbol{\xi}} \left(\mathrm{e}^{-\kappa t}\, \boldsymbol{\xi}\right) = \mathrm{e}^{-\kappa t} \big[\big(\underbrace{\boldsymbol{\nabla}_{\boldsymbol{\xi}} \mathrm{e}^{-\kappa t}}_{\mathrm{d}\mathrm{e}^{-\kappa t}/\mathrm{d}t} \big) \boldsymbol{\xi} + \mathrm{e}^{-\kappa t} \underbrace{\boldsymbol{\nabla}_{\boldsymbol{\xi}}\, \boldsymbol{\xi}}_{\kappa \boldsymbol{\xi}} \big] = 0.$$

Hence $\boldsymbol{\ell}$ is a geodesic vector. Besides, along any null generator of $\mathscr{H}$, one has [cf. Eq. (A.8)]

$$\boldsymbol{\xi}(\lambda) = \frac{\mathrm{d}\lambda}{\mathrm{d}t} = \mathrm{e}^{\kappa t} \underbrace{\boldsymbol{\ell}(\lambda)}_{1} = \mathrm{e}^{\kappa t},$$

which yields Eq. (5.10). □

Let us assume $\kappa > 0$. Let $\mathscr{L}$ be a null geodesic generator of the Killing horizon $\mathscr{H}$. $\mathscr{L}$ can be parameterized by t, the corresponding tangent vector being $\boldsymbol{\xi}$. When t spans the whole interval $(-\infty, +\infty)$, Eq. (5.10) implies that λ spans the interval $(\lambda_0, +\infty)$ only. Since λ is an affine parameter of $\mathscr{L}$, this means that $\mathscr{L}$ is an *incomplete* geodesic (cf. Sec. B.3.2). Moreover, Eq. (5.9) leads to

$$\boldsymbol{\xi} \to 0 \quad \text{when} \quad t \to -\infty \qquad (\kappa > 0). \tag{5.11}$$

In other words, the Killing vector field $\boldsymbol{\xi}$ vanishes and the null geodesic $\mathscr{L}$ stops at the "edge" of $\mathscr{H}$ corresponding to $t \to -\infty$. If there is no singularity there, $\mathscr{L}$ can be extended to $\lambda \in (-\infty, \lambda_0]$, giving rise to a complete null geodesic $\tilde{\mathscr{L}}$. This operation can be performed for all the null geodesic generators of $\mathscr{H}$ and we have the freedom to choose the same value of λ_0 in Eq. (5.10) for all of them. In this process, one gives birth to a null hypersurface, $\tilde{\mathscr{H}}$ say, which contains $\mathscr{H}$. Let $\mathscr{S} \subset \tilde{\mathscr{H}}$ be the set of points of affine parameter $\lambda = \lambda_0$ along all the extended null geodesics $\tilde{\mathscr{L}}$. $\mathscr{S}$ is clearly a cross-section of $\tilde{\mathscr{H}}$ (cf. Sec. 2.3.4); it is then a spacelike $(n-2)$-dimensional surface. $\mathscr{S}$ constitutes the past boundary of $\mathscr{H}$, i.e. the boundary corresponding to $t \to -\infty$. Since $\boldsymbol{\xi}$ is a smooth vector field on $\mathscr{M}$, Eq. (5.11) implies that $\boldsymbol{\xi}$ vanishes on $\mathscr{S}$. In other words, $\mathscr{S}$ is a set of fixed points for the isometry group generated by $\boldsymbol{\xi}$ (cf. Remark 1 above). Let us denote by $\mathscr{H}^-$ the subset of $\tilde{\mathscr{H}}$ generated by the segments $\lambda < \lambda_0$ of the null geodesics $\tilde{\mathscr{L}}$: $\mathscr{H}^- = \tilde{\mathscr{H}} \setminus (\mathscr{H} \cup \mathscr{S})$. $\mathscr{H}^-$ is clearly a null hypersurface. Since $\mathscr{S}$ is spacelike and

$(n-2)$-dimensional, there are, at each point $p \in \mathcal{S}$, only two null directions normal to $\mathcal{S}$ (cf. Sec. 2.3.4). One of them is along $\boldsymbol{\ell}$. The set of all null geodesics departing from $\mathcal{S}$ along the other null direction forms a null hypersurface, $\mathcal{H}_2^+$ say, in the future of $\mathcal{S}$ and another null hypersurface, $\mathcal{H}_2^-$ say, in the past of $\mathcal{S}$. By studying the behaviour of a Killing vector field around the set of its fixed points (here $\mathcal{S}$), Boyer [29] has shown that in the current setting (i.e. $\mathcal{S}$ spacelike), $\boldsymbol{\xi}$ acts *locally* as the generator of Lorentz boosts in Minkowski spacetime and $\mathcal{S}$ is the bifurcation surface of a bifurcate Killing horizon similar to that of Example 1 (cf. Fig. 5.3). More precisely, Boyer proved the following theorem [29]: a Killing horizon $\mathcal{H}$ is contained in a bifurcate Killing horizon if and only if $\mathcal{H}$ contains at least one incomplete, extendable, null geodesic generator. The last property is guaranteed by $\kappa \neq 0$, as we have seen. It follows that $\mathcal{H}^-$, $\mathcal{H}_2^+$ and $\mathcal{H}_2^-$ are three Killing horizons, so that $\mathcal{H} \cup \mathcal{H}^- \cup \mathcal{H}_2^+ \cup \mathcal{H}_2^- \cup \mathcal{S}$ is a bifurcate Killing horizon.

If $\kappa < 0$, we see from Eq. (5.10) that while t spans the whole interval $(-\infty, +\infty)$, the affine parameter λ spans the interval $(-\infty, \lambda_0)$ only. Moreover, Eq. (5.9) leads to

$$\boldsymbol{\xi} \to 0 \quad \text{when} \quad t \to +\infty \qquad (\kappa < 0). \tag{5.12}$$

The reasoning developed for $\kappa > 0$ can be then applied mutatis mutandis, leading to a bifurcate Killing horizon with a bifurcation surface $\mathcal{S}$ that is the future boundary of $\mathcal{H}$. Hence we conclude

> The null geodesic generators of a non-degenerate Killing horizon $\mathcal{H}$ are incomplete; if they can be extended, $\mathcal{H}$ is contained in a bifurcate Killing horizon, the bifurcation surface of which is the past (resp. future) boundary of $\mathcal{H}$ if $\kappa > 0$ (resp. $\kappa < 0$).

Remark 3: For a degenerate Killing horizon, the problem of extension disappears, since t is then an affine parameter of the null generators. Consequently if t spans the whole interval $(-\infty, \infty)$, the null generators are complete geodesics. One can still have $\boldsymbol{\xi} \to 0$ at some boundary of $\mathcal{H}$, but this is a null boundary, not a spacelike one, and it does not correspond to a bifurcation surface. An example is the Killing horizon with respect to a null-rotation Killing vector in Minkowski spacetime, exhibited as Examples 8 and 13 in Chap. 3, p. 68 and 75 respectively (cf. Fig. 3.5): $\boldsymbol{\xi} = 0$ on the null 2-plane of equation $t = x$, $y = 0$.

Historical note: The concept of bifurcate Killing horizons has been introduced by Robert H. Boyer (1932-1966), a young American mathematical physicist just appointed to the University of Liverpool. Sadly, Boyer was killed, among 14 victims, in a mass murder that occurred in the University of Texas at Austin on 1 August 1966. His last notes, containing the definition of bifurcate Killing horizon and the proof of the above mentioned theorem, have been transformed to an article by J. Ehlers and J.L. Stachel and published in 1969 [29].

5.5 The no-hair theorem

In dimension $n = 4$, one can go much further then just claiming that the event horizon of a stationary black hole must be a Killing horizon (Sec. 5.3). One has indeed the

Carter-Robinson theorem (Carter 1971 [44], Robinson 1975 [224]): any stationary and axisymmetric 4-dimensional asymptotically flat black hole spacetime $(\mathscr{M}, \boldsymbol{g})$ that is solution of the vacuum Einstein equation with a connected regular event horizon $\mathscr{H}$ and no closed timelike curve outside it has a domain of outer communications that is isometric to the domain of outer communications of the Kerr spacetime.

Remark 1: In their original works, Carter and Robinson assumed that $\mathscr{H}$ is a *non-degenerate* Killing horizon, i.e. that the non-affinity coefficient κ associated with the Killing vector $\boldsymbol{\chi}$ is non-zero (cf. Sec. 3.3.6). However, this non-degeneracy hypothesis can be released [62] (see [60] for an extended discussion).

Remark 2: The causality condition (absence of closed timelike curves in the black hole exterior), which is one of the assumptions of Carter's theorem (cf. [49] for a discussion), does not appear in Israel's theorem (Sec. 5.3.1) because a static spacetime, which by definition has hypersurface-orthogonal timelike curves, cannot contain any closed timelike curve.

By combining the staticity, Israel, strong rigidity and Carter-Robinson theorems, one arrives at the famous ***no-hair theorem***:

Any spacetime $(\mathscr{M}, \boldsymbol{g})$ that

- is 4-dimensional
- is asymptotically flat
- is stationary
- is a solution of the vacuum Einstein equation
- contains a black hole with a connected regular horizon
- does not contain any closed timelike curve in the domain of outer communications
- is analytic

has a domain of outer communications that is isometric to the domain of outer communications of the Kerr spacetime.

Historical note: See the historical account by Carter [49].

Chapter 6

Schwarzschild black hole

Contents

6.1 Introduction

After having discussed stationary black holes in Chap. 5, we examine here the simplest of them: the Schwarzschild black hole. Let us recall that the prime importance of this object in general relativity stems from the no-hair theorem (Sec. 5.5), which, in the non-rotating case, implies that any static black hole in an asymptotically flat 4-dimensional spacetime must be a Schwarzschild black hole.

In this chapter, we derive the Schwarzschild metric as a solution of the Einstein equation, possibly with a non-vanishing cosmological constant (Sec. 6.2); we then explore it by means of the Eddington-Finkelstein coordinates, which have the advantage to be regular on the horizon (Sec. 6.3). Finally, in Sec. 6.4, we check formally that the Schwarzschild spacetime has a region that obeys the general definition of a black hole given in Sec. 4.4.2. The maximal extension of the Schwarzschild spacetime and its bifurcate Killing horizon is discussed Chap. 9, after two chapters (Chap. 7 and 8) devoted to geodesics of the Schwarzschild spacetime.

6.2 The Schwarzschild-(anti-)de Sitter solution

6.2.1 Vacuum Einstein equation with a cosmological constant

Let us search for a static and spherically symmetric solution of the Einstein equation (1.38) in a vacuum 4-dimensional spacetime $(\mathcal{M}, \boldsymbol{g})$ with some arbitrary cosmological

constant Λ. Setting $\boldsymbol{T} = 0$ in Eq. (1.38) yields the equation to solve:

$$\boldsymbol{R} + \left(\Lambda - \frac{1}{2} R\right) \boldsymbol{g} = 0, \tag{6.1}$$

$\boldsymbol{R}$ being the Ricci tensor of $\boldsymbol{g}$ and $R := g^{\mu\nu} R_{\mu\nu}$ its trace with respect to $\boldsymbol{g}$, i.e. the so-called Ricci scalar (cf. Sec. A.5.3 in Appendix A). Let us first note that Eq. (6.1) implies a constraint on R. Indeed, the trace of Eq. (6.1) with respect to $\boldsymbol{g}$ is

$$R + \left(\Lambda - \frac{1}{2} R\right) \times 4 = 0,$$

hence

$$\boxed{R = 4\Lambda}. \tag{6.2}$$

In particular R is constant. Inserting this value back into (6.1), we get

$$\boxed{\boldsymbol{R} = \Lambda \boldsymbol{g}}. \tag{6.3}$$

Since this equation yields (6.2) as well, we conclude that it is equivalent to (6.1).

6.2.2 Static and spherically symmetric metric

Let us assume that the spacetime $(\mathscr{M}, \boldsymbol{g})$ is *static*, in the sense defined in Sec. 5.2.1: the translation group $(\mathbb{R}, +)$ is a isometry group of $(\mathscr{M}, \boldsymbol{g})$ (cf. Sec. 3.3.1), with orbits that are timelike, at least near some conformal boundary (stationarity property) and hypersurface-orthogonal (staticity property). Let us denote by $\boldsymbol{\xi}$ the associated Killing vector field (unique up to some constant rescaling), i.e. the generator of the isometry group $(\mathbb{R}, +)$ (cf. Sec. 3.3.1).

We may foliate $\mathscr{M}$ by a 1-parameter family of hypersurfaces $(\Sigma_t)_{t\in\mathbb{R}}$, such that $\boldsymbol{\xi}$ is normal to all Σ_t's and t is a parameter associated to $\boldsymbol{\xi}$:

$$\boldsymbol{\xi}(t) = 1 \tag{6.4}$$

or equivalently,

$$\langle \mathbf{d}t, \boldsymbol{\xi} \rangle = 1.$$

In addition to being static, we assume that $(\mathscr{M}, \boldsymbol{g})$ is ***spherically symmetric***, i.e. that it is invariant under the action of the rotation group SO(3), whose orbits are spacelike 2-spheres (cf. Sec. 3.3.1). Let $\mathscr{S}$ be some generic orbit 2-sphere. The static Killing vector field $\boldsymbol{\xi}$ must be orthogonal to $\mathscr{S}$, otherwise the orthogonal projection of $\boldsymbol{\xi}$ onto $\mathscr{S}$ would define some privileged direction on $\mathscr{S}$, which is incompatible with spherical symmetry. The orthogonality of $\boldsymbol{\xi}$ and $\mathscr{S}$ implies that $\mathscr{S} \subset \Sigma_t$. Let $(x^a) = (\theta, \varphi)$ be spherical coordinates on $\mathscr{S}$. The (Riemannian) metric $\boldsymbol{q}$ induced by $\boldsymbol{g}$ on $\mathscr{S}$ is given by

$$q_{ab}\, \mathrm{d}x^a\, \mathrm{d}x^b = r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right). \tag{6.5}$$

The positive coefficient r^2 in front of the standard spherical element must be constant over $\mathscr{S}$, by virtue of spherical symmetry. The area of $\mathscr{S}$ is then $A = 4\pi r^2$. For this reason,

r is called the ***areal radius*** of $\mathscr{S}$. Letting $\mathscr{S}$ vary, r can be considered as a scalar field on $\mathscr{M}$. If $\mathbf{d}r \neq 0$, we may use it as a coordinate. Since $\mathscr{S} \subset \Sigma_t$, (r, θ, φ) is a coordinate system on each hypersurface Σ_t. The set (t, r, θ, φ), where t is adapted to $\boldsymbol{\xi}$ thanks to (6.4), is then a spacetime coordinate system and, by construction, the expression of the metric tensor with respect to this system is

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -A(r)\,\mathrm{d}t^2 + B(r)\,\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right). \tag{6.6}$$

Note that this is a special case of the general static metric element (5.5) and that Eq. (5.4) holds:

$$\boldsymbol{\xi} = \boldsymbol{\partial}_t. \tag{6.7}$$

In particular, $g_{tt} = -A(r)$ and $g_{rr} = B(r)$ do not depend on t as a result of the spacetime stationarity, while $g_{tr} = g_{t\theta} = g_{t\varphi} = 0$ expresses the orthogonality of $\boldsymbol{\xi}$ and Σ_t, i.e. the spacetime staticity. The coordinates (t, r, θ, φ) are called ***areal coordinates***, reflecting the fact that r is the areal radius.

6.2.3 Solving Einstein equation

The Christoffel symbols of the metric (6.6) with respect to the areal coordinates are (cf. Sec. D.3.2 for the computation):

$$\begin{gathered}
\Gamma^t{}_{tr} = \Gamma^t{}_{rt} = \frac{1}{2A}\frac{\mathrm{d}A}{\mathrm{d}r} \qquad \Gamma^r{}_{tt} = \frac{1}{2B}\frac{\mathrm{d}A}{\mathrm{d}r} \qquad \Gamma^r{}_{rr} = \frac{1}{2B}\frac{\mathrm{d}B}{\mathrm{d}r} \qquad \Gamma^r{}_{\theta\theta} = -\frac{r}{B} \\
\Gamma^r{}_{\varphi\varphi} = -\frac{r\sin^2\theta}{B} \qquad \Gamma^\theta{}_{r\theta} = \Gamma^\theta{}_{\theta r} = \frac{1}{r} \qquad \Gamma^\theta{}_{\varphi\varphi} = -\sin\theta\cos\theta \\
\Gamma^\varphi{}_{r\varphi} = \Gamma^\varphi{}_{\varphi r} = \frac{1}{r} \qquad \Gamma^\varphi{}_{\theta\varphi} = \Gamma^\varphi{}_{\varphi\theta} = \frac{1}{\tan\theta},
\end{gathered} \tag{6.8}$$

the Christoffel symbols not listed above being zero.

The tt component of the Einstein equation (6.1) leads to (cf. Sec. D.3.2 for the computation)

$$r\frac{\mathrm{d}B}{\mathrm{d}r} - B + (1 - \Lambda r^2)B^2 = 0, \tag{6.9}$$

while the rr component leads to

$$r\frac{\mathrm{d}A}{\mathrm{d}r} + A - (1 - \Lambda r^2)AB = 0. \tag{6.10}$$

Finally, the $\theta\theta$ and $\varphi\varphi$ components lead to the same equation:

$$2\frac{\mathrm{d}^2A}{\mathrm{d}r^2} + \frac{2}{r}\frac{\mathrm{d}A}{\mathrm{d}r} - \frac{1}{B}\left(\frac{\mathrm{d}A}{\mathrm{d}r} + \frac{2A}{r}\right)\frac{\mathrm{d}B}{\mathrm{d}r} - \frac{1}{A}\left(\frac{\mathrm{d}A}{\mathrm{d}r}\right)^2 + 4\Lambda AB = 0. \tag{6.11}$$

All the other components of the Einstein equation (6.1) are identically zero.

Adding Eq. (6.9) multiplied by A to Eq. (6.10) multiplied by B yields

$$B\frac{\mathrm{d}A}{\mathrm{d}r} + A\frac{\mathrm{d}B}{\mathrm{d}r} = \frac{\mathrm{d}}{\mathrm{d}r}(AB) = 0.$$

The solution of this equation is obviously $A(r)B(r) = C$, where C is a constant. Without any loss of generality, we may choose $C = 1$. Indeed, substituting $C/B(r)$ for $A(r)$ in Eq. (6.6) results in

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\frac{C}{B(r)}\,\mathrm{d}t^2 + B(r)\,\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right).$$

Assuming $C > 0$, the change of variable $t' = \sqrt{C}t$, which is equivalent to changing the stationary Killing vector from $\boldsymbol{\xi}$ to $\boldsymbol{\xi}' = 1/\sqrt{C}\,\boldsymbol{\xi}$, yields

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\frac{1}{B(r)}\,\mathrm{d}t'^2 + B(r)\,\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right),$$

which is exactly the solution corresponding to $C = 1$. Hence from now on, we set $C = 1$, i.e.

$$B(r) = \frac{1}{A(r)}. \tag{6.12}$$

Substituting this expression in Eq. (6.10) yields an ordinary differential equation for $A(r)$:

$$r\frac{\mathrm{d}A}{\mathrm{d}r} + A - 1 + \Lambda r^2 = 0,$$

the solution of which is

$$A(r) = 1 - \frac{2m}{r} - \frac{\Lambda}{3}\,r^2, \tag{6.13}$$

where m is a constant. The general static and spherically symmetric solution of the vacuum Einstein equation (6.1) is therefore

$$\boxed{g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\left(1 - \frac{2m}{r} - \frac{\Lambda}{3}\,r^2\right)\mathrm{d}t^2 + \left(1 - \frac{2m}{r} - \frac{\Lambda}{3}\,r^2\right)^{-1}\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)}. \tag{6.14}$$

It is called the ***Kottler metric*** (cf. the historical note below). The ***Schwarzschild metric*** is the particular case $\Lambda = 0$. If $\Lambda > 0$, (6.14) is called the ***Schwarzschild-de Sitter metric***, often abridged as ***Schwarzschild-dS metric***, while if $\Lambda < 0$, it is called the ***Schwarzschild-anti-de Sitter metric***, often abridged as ***Schwarzschild-AdS metric***.

In the rest of this chapter, we will focus on the Schwarzschild metric, i.e. on the version $\Lambda = 0$ of Eq. (6.14):

$$\boxed{g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\left(1 - \frac{2m}{r}\right)\mathrm{d}t^2 + \left(1 - \frac{2m}{r}\right)^{-1}\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)}. \tag{6.15}$$

The areal coordinates (t, r, θ, φ) are then called the ***Schwarzschild-Droste coordinates***[1].

[1] In the literature they are often referred to as simply ***Schwarzschild coordinates***; we follow here Deruelle & Uzan [85, 86].

Since $A(r) = 1 - 2m/r$ and $B(r) = (1 - 2m/r)^{-1}$ for the Schwarzschild metric, the non-vanishing Christoffel symbols (6.8) become[2]

$$\begin{aligned}
&\Gamma^t{}_{tr} = \Gamma^t{}_{rt} = \frac{m}{r(r-2m)} \qquad \Gamma^r{}_{tt} = \frac{m(r-2m)}{r^3} \qquad \Gamma^r{}_{rr} = -\frac{m}{r(r-2m)} \\
&\Gamma^r{}_{\theta\theta} = 2m - r \qquad \Gamma^r{}_{\varphi\varphi} = (2m-r)\sin^2\theta \qquad \Gamma^\theta{}_{r\theta} = \Gamma^\theta{}_{\theta r} = \frac{1}{r} \\
&\Gamma^\theta{}_{\varphi\varphi} = -\sin\theta\cos\theta \qquad \Gamma^\varphi{}_{r\varphi} = \Gamma^\varphi{}_{\varphi r} = \frac{1}{r} \qquad \Gamma^\varphi{}_{\theta\varphi} = \Gamma^\varphi{}_{\varphi\theta} = \frac{1}{\tan\theta}.
\end{aligned} \tag{6.16}$$

6.2.4 The mass parameter

The Schwarzschild metric (6.15) depends on a single parameter: m. This parameter has a direct physical interpretation: it is the ***gravitational mass*** (or simply ***mass***) that is felt by an observer located at large values of r. Indeed, we will see in Chap. 7 that an observer on a circular orbit at a large value of r has an orbital period T obeying Kepler's third law: $T^2 = 4\pi^2 r^3/m$. Without waiting for Chap. 7, we may notice that for $r \gg |m|$, the line element (6.15) takes the standard weak-field form (see e.g. [38, 190]):

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \simeq -\left(1 + 2\Phi(r)\right)\,\mathrm{d}t^2 + \left(1 - 2\Phi(r)\right)\,\mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right), \tag{6.17}$$

where $\Phi(r) := -m/r$ is the Newtonian gravitational potential outside a spherically symmetric body of mass m.

Historical note: The Schwarzschild metric (6.15) is actually the first non-trivial (i.e. different from Minkowski metric) solution of Einstein equation ever found. It has been obtained by the astrophysicist Karl Schwarzschild in the end of 1915 [229], only a few weeks after the publication of the articles funding general relativity by Albert Einstein. It is also quite remarkable that Schwarzschild found the solution while serving in the German army at the Russian front. Unfortunately, he died from a rare skin disease a few months later. The way Schwarzschild proceeded was quite different from that exposed above: instead of the coordinates (t, r, θ, φ) named today after him, he used the coordinates (t, x^1, x^2, φ) where $x^1 = r_*^3/3$, with $r_*^3 = r^3 - 8m^3$, and $x^2 = -\cos\theta$. Such a choice was made to enforce $\det(g_{\alpha\beta}) = -1$, a condition prescribed by Einstein in an early version of general relativity, which had been presented on 18 November 1915 and on which Schwarzschild was working. Only in the final version, published on 25 November 1915, did Einstein relax the condition $\det(g_{\alpha\beta}) = -1$, allowing for full covariance. Schwarzschild however exhibited the famous line element (6.15), via what he called the "auxiliary quantity" $r = (r_*^3 + 8m^3)^{1/3}$. For him, the "center", namely the location of the "point mass" generating the field, was at $r_* = 0$, i.e. at $r = 2m$. Independently of Schwarzschild, Johannes Droste, then PhD student of Hendrik Lorentz, arrived at the solution (6.15) in May 1916 [96]. Contrary to Schwarzschild, Droste performed the computation with a spherical coordinate system, $(t, \bar{r}, \theta, \varphi)$, yet distinct from the standard "Schwarzschild-Droste" coordinates (t, r, θ, φ) by the fact that the radial coordinate $\bar{r}$ was not chosen to be the areal radius, but instead a coordinate for which $g_{\bar{r}\bar{r}} = 1$. At the end, by a change of variable, Droste exhibited the line element (6.15). The generalization to a non-vanishing cosmological constant, i.e. Eq. (6.14), has been obtained by

[2] See also the notebook D.3.3 for a check.

Friedrich Kottler in 1918 [168] and, independently, by Hermann Weyl in 1919 [262]. We refer to Eisenstaedt's article [99] for a detailed account of the early history of the Schwarzschild solution.

6.2.5 The Schwarzschild-Droste domain

We immediately notice on (6.15) that the metric components are singular at $r = 0$ and $r = 2m$. Accordingly, the Schwarzschild-Droste coordinates (t, r, θ, φ) cover the following subset of $\mathscr{M}$, which we call the ***Schwarzschild-Droste domain***:

$$\mathscr{M}_{\rm SD} := \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm II}, \tag{6.18a}$$

$$\mathscr{M}_{\rm I} := \mathbb{R} \times (2m, +\infty) \times \mathbb{S}^2, \tag{6.18b}$$

$$\mathscr{M}_{\rm II} := \mathbb{R} \times (0, 2m) \times \mathbb{S}^2, \tag{6.18c}$$

with the coordinate t spanning $\mathbb{R}$, the coordinate r spanning $(2m, +\infty)$ on $\mathscr{M}_{\rm I}$ and $(0, 2m)$ on $\mathscr{M}_{\rm II}$, and the coordinates (θ, φ) constituting a standard spherical chart of $\mathbb{S}^2$. Note that $\mathscr{M}_{\rm SD}$ is a disconnected open subset of the full spacetime manifold $\mathscr{M}$ (to be specified later), whose connected components are $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm II}$.

Remark 1: To cover entirely $\mathbb{S}^2$ in a regular way, one needs a second chart, in addition to (θ, φ); this is related to the standard singularities of spherical coordinates at $\theta = 0$ and $\theta = \pi/2$. It is fully understood that the metric $\boldsymbol{g}$, as expressed by (6.15), is fully regular on $\mathbb{S}^2$. The fact that $\det(g_{\alpha\beta}) = -r^2 \sin^2\theta$ is zero at $\theta = 0$ and $\theta = \pi/2$ reflects merely the coordinate singularity of the (θ, φ) chart there. We shall not discuss this coordinate singularity any further.

The boundary value $r_{\rm S} := 2m$ of r between $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm II}$ is conventionaly called the ***Schwarzschild radius***. A more appropriate name would have been the *Schwarzschild areal radius*, for r does not describe a *radius* (in the sense of a distance from some "origin") but rather an *area*, as discussed in Sec. 6.2.2.

A first property of the Schwarzschild metric is that $\mathscr{M}_{\rm I}$ has an asymptotically flat end: it is clear on (6.15) that the metric $\boldsymbol{g}$ tends to Minkowski metric (4.3) when $r \to +\infty$ [see also Eq. (6.17)].

Besides, in region $\mathscr{M}_{\rm II}$, we notice on (6.15) that $g_{tt} > 0$. Since $g_{tt} = \boldsymbol{g}(\boldsymbol{\partial}_t, \boldsymbol{\partial}_t)$, this implies that the Killing vector field $\boldsymbol{\xi} = \boldsymbol{\partial}_t$ is spacelike. Hence, $(\mathscr{M}_{\rm II}, \boldsymbol{g})$ is not static, in the sense defined in Sec. 6.2.2: the translation group $(\mathbb{R}, +)$ is still an isometry group of $(\mathscr{M}_{\rm II}, \boldsymbol{g})$, but its orbits are spacelike curves. We note that $g_{rr} < 0$ in $\mathscr{M}_{\rm II}$, so that the metric (6.15) keeps a Lorentzian signature, as it should! In other words, in $\mathscr{M}_{\rm II}$, t becomes a space coordinate and r a time coordinate. Accordingly, the axes of the light cones in Fig. 6.1 are horizontal lines for $r < 2m$.

6.3 Radial null geodesics and Eddington-Finkelstein coordinates

6.3.1 Radial null geodesics

Let us search for the null geodesics of the Schwarzschild metric (6.15) that are radial, i.e. along which $\theta = \text{const}$ and $\varphi = \text{const}$. They are found by setting $\mathrm{d}\theta = 0$ and $\mathrm{d}\varphi = 0$ in

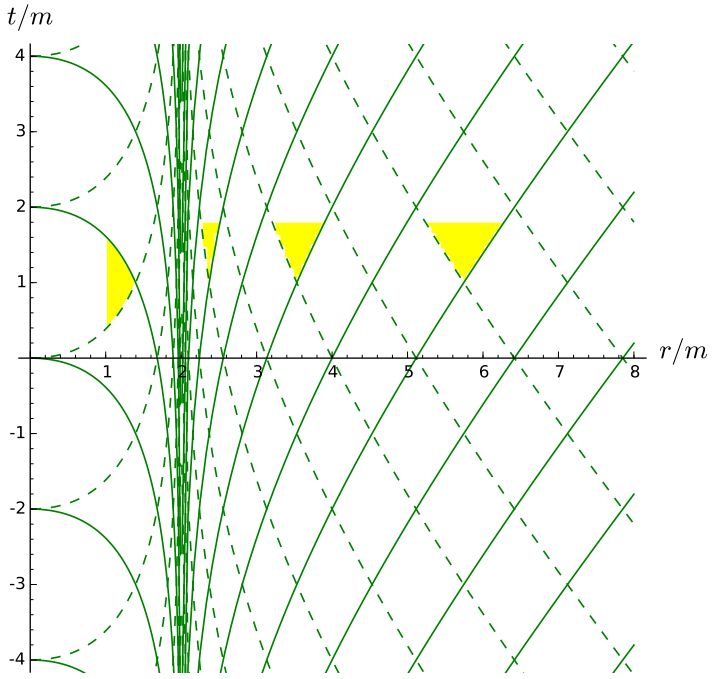


Figure 6.1: Radial null geodesics of Schwarzschild spacetime, plotted in terms of Schwarzschild-Droste coordinates (t, r): the solid (resp. dashed) lines correspond to outgoing (resp. ingoing) geodesics, as given by Eq. (6.22) (resp. Eq. (6.23)). The interiors of some future light cones are depicted in yellow.

(6.15) and searching for $\mathrm{d}s^2 = g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = 0$:

$$\mathrm{d}s^2 = 0 \iff \mathrm{d}t^2 = \frac{\mathrm{d}r^2}{\left(1 - \frac{2m}{r}\right)^2}. \tag{6.19}$$

Hence the radial null geodesics are governed by

$$\mathrm{d}t = \pm \frac{\mathrm{d}r}{1 - \frac{2m}{r}}. \tag{6.20}$$

This equation is easily integrated:

$$t = \pm r \pm 2m \ln \left| \frac{r}{2m} - 1 \right| + \text{const.} \tag{6.21}$$

We have thus two families of curves, one for each choice of sign in $\pm$:

- the ***outgoing radial null geodesics***, whose equation is

$$t = r + 2m \ln \left| \frac{r}{2m} - 1 \right| + u, \tag{6.22}$$

  where u is a constant;

- the ***ingoing radial null geodesics***, whose equation is

$$t = -r - 2m \ln \left| \frac{r}{2m} - 1 \right| + v, \tag{6.23}$$

where v is a constant.

By introducing the ***tortoise coordinate***

$$r_* := r + 2m \ln \left| \frac{r}{2m} - 1 \right|, \tag{6.24}$$

one may rewrite the above equations as

$$t = r_* + u \tag{6.25}$$

$$t = -r_* + v. \tag{6.26}$$

The parameter u appears then as a *retarded time*: $u = t - r_*$ and v as an *advanced time*: $v = t + r_*$.

Strictly speaking, we have found radial null *curves* only, i.e. solutions of Eq. (6.19). Since not all null curves are null geodesics[3], there remains to prove that the curves defined by (6.22) and (6.23) obey the geodesic equation [Eq. (B.10) in Appendix B]:

$$\frac{\mathrm{d}^2 x^\alpha}{\mathrm{d}\lambda^2} + \Gamma^\alpha{}_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = 0, \tag{6.27}$$

where λ is an affine parameter (cf. Sec. B.2.1). Let us check that (6.27) is satisfied by choosing $\lambda = r$. For the curves defined by (6.22), we have

$$x^\alpha(r) = \left(r + 2m \ln \left| \frac{r}{2m} - 1 \right| + u,\ r,\ \theta,\ \varphi \right).$$

Hence

$$\frac{\mathrm{d}x^\alpha}{\mathrm{d}r} = \left(\frac{r}{r-2m}, 1, 0, 0 \right) \qquad \text{and} \qquad \frac{\mathrm{d}^2 x^\alpha}{\mathrm{d}r^2} = \left(-\frac{2m}{(r-2m)^2}, 0, 0, 0 \right).$$

Given the Christoffel symbols (6.16), it is then a simple exercise to show that Eq. (6.27) is satisfied. The same property holds for the family (6.23). Hence we conclude

> The radial null geodesics in the Schwarzschild-Droste domain are ruled by Eqs. (6.22)-(6.23). Moreover, the areal radius r is an affine parameter along them.

The two families of radial null geodesics are depicted in Fig. 6.1. The singularity of Schwarzschild-Droste coordinates at the Schwarzschild radius $r = 2m$ appears clearly on this figure.

Remark 1: Despite their name, geodesics of the outgoing family are actually *ingoing* in the region $r < 2m$, in the sense that r is decreasing along them when moving towards the future. Indeed, as noticed in Sec. 6.2.5, for $r < 2m$, r is a timelike coordinate in the system (t, r, θ, φ), i.e. $\boldsymbol{\partial}_r$ is a timelike vector and we shall see in Sec. 6.3.6 that $-\boldsymbol{\partial}_r$ is oriented towards the future (cf. the "tilted" light cone in Fig. 6.1).

[3] A famous counterexample is the null helix in Minkowski spacetime, cf. Remark 2 on p. 35.

6.3.2 Eddington-Finkelstein coordinates

The parameter v introduced in Eq. (6.23) can be seen as a label for the ingoing radial null geodesics: each of these curves is entirely identified by the data (v, θ, φ), which remains fixed along it. Let us promote v to a spacetime coordinate, instead of t, i.e. let us consider the coordinate system (v, r, θ, φ) with the relation to Schwarzschild-Droste coordinates (t, r, θ, φ) governed by Eq. (6.23):

$$v = t + r + 2m \ln \left| \frac{r}{2m} - 1 \right| . \tag{6.28}$$

It follows immediately that

$$\mathrm{d}v = \mathrm{d}t + \mathrm{d}r + \frac{\mathrm{d}r}{r/2m - 1} = \mathrm{d}t + \frac{\mathrm{d}r}{1 - 2m/r},$$

i.e.

$$\mathrm{d}t = \mathrm{d}v - \frac{\mathrm{d}r}{1 - 2m/r}. \tag{6.29}$$

Taking the square gives

$$\mathrm{d}t^2 = \mathrm{d}v^2 - \frac{2}{1 - 2m/r}\, \mathrm{d}v\, \mathrm{d}r + \frac{1}{(1 - 2m/r)^2}\, \mathrm{d}r^2.$$

Substituting this expression for $\mathrm{d}t^2$ in Eq. (6.15) yields the metric components with respect to the coordinates $(x^{\hat{\alpha}}) := (v, r, \theta, \varphi)$:

$$\boxed{g_{\hat{\mu}\hat{\nu}}\, \mathrm{d}x^{\hat{\mu}}\, \mathrm{d}x^{\hat{\nu}} = -\left(1 - \frac{2m}{r}\right) \mathrm{d}v^2 + 2\, \mathrm{d}v\, \mathrm{d}r + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)}. \tag{6.30}$$

The coordinates $(x^{\hat{\alpha}}) = (v, r, \theta, \varphi)$ are called the ***null ingoing Eddington-Finkelstein (NIEF) coordinates***. The qualifier *null* stems from the fact that r is a null coordinate in this system, i.e. the vector $\boldsymbol{\partial}_r$ of the coordinate basis associated with (v, r, θ, φ) is a null vector, as it follows from $g_{rr} = 0$ in Eq. (6.30).

To deal with a "standard" time + space coordinate system instead of a null one, let us set

$$\boxed{\tilde{t} := v - r} \iff \boxed{v = \tilde{t} + r} \tag{6.31}$$

and define the ***ingoing Eddington-Finkelstein (IEF) coordinates*** to be

$$(x^{\tilde{\alpha}}) := (\tilde{t}, r, \theta, \varphi). \tag{6.32}$$

Remark 2: From (6.31), v appears as the "time" $\tilde{t}$ "advanced" by r, while from (6.26), v is the "time" t "advanced" by r_*.

The relation between the ingoing Eddington-Finkelstein coordinates $(\tilde{t}, r, \theta, \varphi)$ and the Schwarzschild-Droste ones (t, r, θ, φ) is obtained by combining Eqs. (6.28) and (6.31):

$$\boxed{\tilde{t} = t + 2m \ln \left| \frac{r}{2m} - 1 \right|}. \tag{6.33}$$

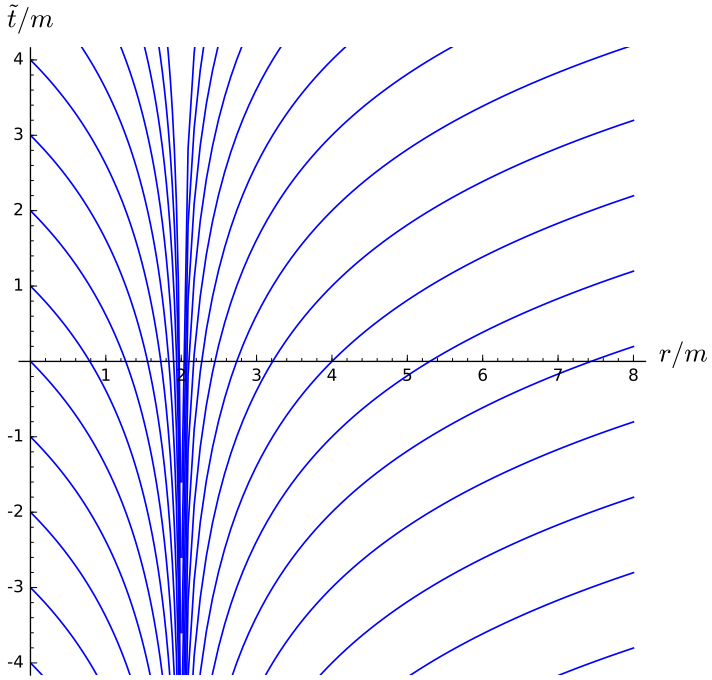


Figure 6.2: Hypersurfaces of constant Schwarzschild-Droste coordinate t, drawn in terms of the ingoing Eddington-Finkelstein coordinates $(\tilde{t}, r)$. Since the dimensions along θ and φ are not represented, these 3-dimensional surfaces appear as curves.

The hypersurfaces $t = \text{const}$ are plotted in Fig. 6.2, in terms of the IEF coordinates.

From (6.31), we have $\mathrm{d}v = \mathrm{d}\tilde{t} + \mathrm{d}r$. Substituting into (6.30) yields

$$\boxed{g_{\tilde{\mu}\tilde{\nu}}\, \mathrm{d}x^{\tilde{\mu}}\, \mathrm{d}x^{\tilde{\nu}} = -\left(1 - \frac{2m}{r}\right) \mathrm{d}\tilde{t}^2 + \frac{4m}{r}\, \mathrm{d}\tilde{t}\, \mathrm{d}r + \left(1 + \frac{2m}{r}\right) \mathrm{d}r^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)}. \tag{6.34}$$

We check that $g_{\tilde{t}\tilde{t}} < 0$ in $\mathscr{M}_{\mathrm{I}}$, hence $\tilde{t}$ is a timelike coordinate there. In $\mathscr{M}_{\mathrm{II}}$, $g_{\tilde{t}\tilde{t}} > 0$, so that $\tilde{t}$ becomes spacelike there, as for the Schwarzschild-Droste coordinate t (cf. Sec. 6.2.5). However, we have $g_{rr} = 1+2m/r > 0$ everywhere, so that r remains a spacelike coordinate (for the IEF system) in $\mathscr{M}_{\mathrm{II}}$, contrary to what happens with Schwarzschild-Droste coordinates (cf. Sec. 6.2.5).

Remark 3: The above example shows that the property of being timelike, null or spacelike is not intrinsic to a given coordinate (here r). It is instead a property of the whole coordinate system under consideration. This is understandable since r spacelike means that the line along which r varies while the three other coordinates (x^0, x^2, x^3) are kept constant is a spacelike curve. For the Schwarzschild-Droste system $(x^0, x^2, x^3) = (t, \theta, \varphi)$, while for the NIEF system $(x^0, x^2, x^3) = (v, \theta, \varphi)$ and for the IEF system $(x^0, x^2, x^3) = (\tilde{t}, \theta, \varphi)$. Hence the three sets of r-lines differ. Equivalently, the coordinate vectors $\boldsymbol{\partial}_r$ tangent to the three kinds of r-lines are different:

$$\left.\frac{\partial}{\partial r}\right|_{t,\theta,\varphi} \neq \left.\frac{\partial}{\partial r}\right|_{v,\theta,\varphi} \neq \left.\frac{\partial}{\partial r}\right|_{\tilde{t},\theta,\varphi}.$$

To avoid any ambiguity, we shall denote by $\boldsymbol{\partial}_{\tilde{r}}$ the coordinate vector of the IEF frame and by $\boldsymbol{\partial}_r$ the coordinate vector of the Schwarzschild-Droste frame:

$$\boldsymbol{\partial}_{\tilde{r}} := \left.\frac{\partial}{\partial r}\right|_{\tilde{t},\theta,\varphi} \qquad \text{and} \qquad \boldsymbol{\partial}_r := \left.\frac{\partial}{\partial r}\right|_{t,\theta,\varphi}. \tag{6.35}$$

The relation between the two vectors is given by the chain rule:

$$\left.\frac{\partial}{\partial r}\right|_{\tilde{t},\theta,\varphi} = \left.\frac{\partial}{\partial t}\right|_{r,\theta,\varphi} \underbrace{\left.\frac{\partial t}{\partial r}\right|_{\tilde{t},\theta,\varphi}}_{\left(1-\frac{r}{2m}\right)^{-1}} + \left.\frac{\partial}{\partial r}\right|_{t,\theta,\varphi} \underbrace{\left.\frac{\partial r}{\partial r}\right|_{\tilde{t},\theta,\varphi}}_{1} + \left.\frac{\partial}{\partial \theta}\right|_{t,r,\varphi} \underbrace{\left.\frac{\partial \theta}{\partial r}\right|_{\tilde{t},\theta,\varphi}}_{0} + \left.\frac{\partial}{\partial \varphi}\right|_{t,r,\theta} \underbrace{\left.\frac{\partial \varphi}{\partial r}\right|_{\tilde{t},\theta,\varphi}}_{0},$$

where (6.33) has been used to evaluate $\partial t/\partial r|_{\tilde{t},\theta,\varphi}$. Hence

$$\boldsymbol{\partial}_{\tilde{r}} = \boldsymbol{\partial}_r + \left(1 - \frac{r}{2m}\right)^{-1} \boldsymbol{\partial}_t. \tag{6.36}$$

On the other hand, we deduce from (6.33) that

$$\left.\frac{\partial}{\partial \tilde{t}}\right|_{r,\theta,\varphi} = \left.\frac{\partial}{\partial t}\right|_{r,\theta,\varphi}, \tag{6.37}$$

which implies:

$$\boldsymbol{\partial}_{\tilde{t}} = \boldsymbol{\partial}_t. \tag{6.38}$$

In particular, the vector $\boldsymbol{\partial}_{\tilde{t}}$ of the IEF frame coincides with the Killing vector $\boldsymbol{\xi}$:

$$\boxed{\boldsymbol{\partial}_{\tilde{t}} = \boldsymbol{\xi}}. \tag{6.39}$$

Remark 4: The result (6.39) is not surprising since the metric components (6.34) are independent from $\tilde{t}$. This implies $\boldsymbol{\partial}_{\tilde{t}} = \alpha\boldsymbol{\xi}$, where α is a constant. Since $\tilde{t} \sim t$ when $r \to +\infty$, we get $\alpha = 1$.

Remark 5: The IEF-coordinates line element (6.34) can be recast in the following remarkable form:

$$g_{\tilde{\mu}\tilde{\nu}}\, \mathrm{d}x^{\tilde{\mu}}\, \mathrm{d}x^{\tilde{\nu}} = \underbrace{-\mathrm{d}\tilde{t}^2 + \mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)}_{f_{\tilde{\mu}\tilde{\nu}}\, \mathrm{d}x^{\tilde{\mu}}\, \mathrm{d}x^{\tilde{\nu}}} + \underbrace{\frac{2m}{r}\left(\mathrm{d}\tilde{t} + \mathrm{d}r\right)^2}_{k_{\tilde{\mu}}\mathrm{d}x^{\tilde{\mu}}\, k_{\tilde{\nu}}\mathrm{d}x^{\tilde{\nu}}}, \tag{6.40}$$

where the $f_{\tilde{\mu}\tilde{\nu}}$'s are the components of the (flat) Minkowski metric expressed in terms of the spherical coordinates $(\tilde{t}, r, \theta, \varphi)$ and the $k_{\tilde{\mu}}$'s are the components of a 1-form dual to a null vector:

$$\underline{\boldsymbol{k}} = \sqrt{\frac{2m}{r}}\, \mathbf{d}(\tilde{t} + r) = \sqrt{\frac{2m}{r}}\, \mathbf{d}v.$$

The fact that $\boldsymbol{k}$ is a null vector follows from $g^{\tilde{\mu}\tilde{\nu}} k_{\tilde{\mu}} k_{\tilde{\nu}} = 0$, which is easily deduced from $k_{\tilde{\mu}} = \sqrt{2m/r}(1, 1, 0, 0)$ and expression (6.42) for $g^{\tilde{\mu}\tilde{\nu}}$ below. A metric of the type (6.40) is said to be a ***Kerr-Schild metric*** (cf. Appendix C for details).

6.3.3 The Schwarzschild horizon

Contrary to the Schwarzschild-Droste components (6.15), the metric components (6.34) are regular as $r \to 2m$. Hence (6.34) defines a regular non-degenerate metric on the whole ***ingoing Eddington-Finkelstein domain***

$$\mathscr{M}_{\mathrm{IEF}} := \mathbb{R} \times (0, +\infty) \times \mathbb{S}^2, \tag{6.41}$$

with the coordinate $\tilde{t}$ spanning $\mathbb{R}$, the coordinate r spanning $(0, +\infty)$ and the coordinates (θ, φ) forming the standard spherical chart of $\mathbb{S}^2$. The components of the inverse metric with respect to the ingoing Eddington-Finkelstein coordinates are

$$g^{\tilde{\alpha}\tilde{\beta}} = \begin{pmatrix} -\left(1 + \frac{2m}{r}\right) & \frac{2m}{r} & 0 & 0 \\ \frac{2m}{r} & 1 - \frac{2m}{r} & 0 & 0 \\ 0 & 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & 0 & \frac{1}{r^2 \sin^2\theta} \end{pmatrix}. \tag{6.42}$$

In particular, the components $g^{\tilde{\alpha}\tilde{\beta}}$ are regular at $r = 2m$.

The IEF domain is an extension of the Schwarzschild-Droste domain introduced in Sec. 6.2.5:

$$\mathscr{M}_{\mathrm{IEF}} = \mathscr{M}_{\mathrm{SD}} \cup \mathscr{H} = \mathscr{M}_{\mathrm{I}} \cup \mathscr{M}_{\mathrm{II}} \cup \mathscr{H}, \tag{6.43}$$

where $\mathscr{H}$ is the subset of $\mathscr{M}_{\mathrm{IEF}}$ defined by $r = 2m$. Note that $\mathscr{H}$ has the topology

$$\mathscr{H} \simeq \mathbb{R} \times \mathbb{S}^2 \tag{6.44}$$

and that $(\tilde{t}, \theta, \varphi)$ is a coordinate system on $\mathscr{H}$. Actually $\mathscr{H}$ is nothing but what has been called the ***Schwarzschild horizon*** in the examples of Chaps. 2 and 3. Indeed, the metric (6.34) is nothing but the metric (2.5) introduced in Example 3 of Chap. 2 (p. 29), up to the change of notation $\tilde{t} \leftrightarrow t$ (compare (2.6) and (6.42) as well). We have thus the fundamental result, the proof of which is given in Example 10 of Chap. 3 (p. 69):

> $\mathscr{H}$ is a Killing horizon, the null normal of which is $\boldsymbol{\xi}$.

In particular, $\mathscr{H}$ is a null hypersurface, whose null geodesic generators admit $\boldsymbol{\xi} = \boldsymbol{\partial}_{\tilde{t}}$ as tangent vector. It is a non-expanding horizon, whose area, as defined in Sec. 3.2.2, is (cf. Example 3 of Chap. 3, p. 59)

$$A = 16\pi m^2. \tag{6.45}$$

$\mathscr{H}$ is depicted in Fig. 2.7. We shall see in Sec. 6.4 that $\mathscr{H}$ is actually a black hole event horizon in Schwarzschild spacetime.

6.3.4 Coordinate singularity vs. curvature singularity

The above considerations show that the divergence of the metric component g_{rr} in (6.15) when $r \to 2m$ reflects a pathology of Schwarzschild-Droste coordinates and not a singularity in the metric tensor $\boldsymbol{g}$ by itself: $(\mathscr{M}_{\rm IEF}, \boldsymbol{g})$ is perfectly regular spacetime, including at the Schwarzschild radius $r = 2m$. The bad behaviour of Schwarzschild-Droste coordinates is obvious in Fig. 6.2: the hypersurfaces $t =$ const fail to provide a regular slicing of spacetime. This pathology is called a ***coordinate singularity***, since it is intrinsic a given coordinate system (here the Schwarzschild-Droste one).

Another pathology appears in the metric components in both the Schwarzschild-Droste coordinates and the ingoing Eddington-Finkelstein ones: g_{tt} and $\tilde{g}_{\tilde{t}\tilde{t}}$ diverge when $r \to 0$. This type of singularity cannot be removed by a coordinate transformation. Indeed, the ***Kretschmann scalar***, defined as the following "square" of the Riemann curvature tensor

$$K := R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}, \tag{6.46}$$

is (cf. Sec. D.3.3 for the computation)

$$K = \frac{48m^2}{r^6}. \tag{6.47}$$

Hence $K \to +\infty$ when $r \to 0$. Since K is a scalar field, its value is independent of any coordinate system used to express it. Hence the divergence of K reflects a pathology of the Riemann tensor per se: it is called a ***curvature singularity***.

Historical note: Eddington-Finkelstein coordinates have been introduced by Arthur Eddington in 1924 [97]. More precisely, Eddington introduced the *outgoing* version of these coordinates, while we have focused above on the *ingoing* version. Indeed Eddington's Eq. (2) is $\tilde{t} = t - 2m\ln(r-m)$, which mainly differs from our Eq. (6.33) by the minus sign in front of the logarithm[4], which means that Eddington's time coordinate is actually $\tilde{t} = u + r$, instead of $\tilde{t} = v - r$ (our Eq. (6.31)). Eddington used his transformation to get the Kerr-Schild form (6.40) of Schwarzschild metric, with $(\mathrm{d}\tilde{t} + \mathrm{d}r)^2$ replaced by $(\mathrm{d}\tilde{t} - \mathrm{d}r)^2$ due to the change ingoing $\leftrightarrow$ outgoing. For a modern reader, it is quite surprising that Eddington did not point out that the metric components w.r.t. $(\tilde{t}, r, \theta, \varphi)$ are regular at $r = 2m$. Actually the main purpose of Eddington's article [97] was elsewhere, in the comparison of general relativity to an alternative theory proposed in 1922 by the mathematician Alfred N. Whitehead (see e.g. [118]). Only in 1958 did David Finkelstein reintroduce the Eddington transformation to show that the Schwarzschild metric is analytic over the whole domain $r \in (0, +\infty)$ [106]. Meanwhile the regularity of Schwarzschild metric at $r = 2m$ had been proven by Georges Lemaître in 1932 [178], via another coordinate system (see [101] for a detailed discussion), as well as by John L. Synge in 1950 [240], by means of yet another coordinate system (cf. the historical note on p. 231).

Remark 6: In the literature, the terminology *Eddington-Finkelstein coordinates* is often used for the coordinates (v, r, θ, φ) (or (u, r, θ, φ)), i.e. for what we have called the *null Eddington-Finkelstein coordinates*, and the regularity of the metric tensor at $r = 2m$ is demonstrated by

[4] The other differences with (6.33) are a constant additive term and a misprint in Eddington's formula: the term $\ln(r-m)$ should be replaced by $\ln(r-2m)$.

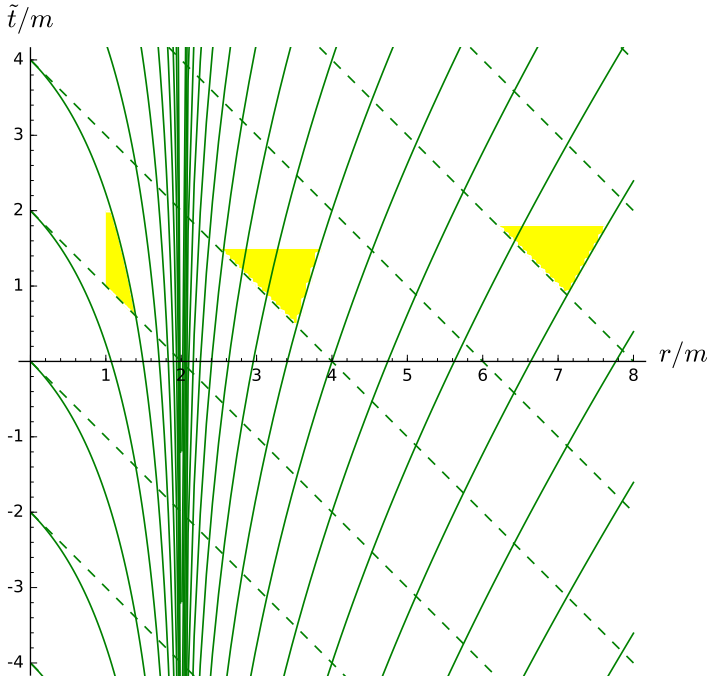


Figure 6.3: Radial null geodesics of Schwarzschild spacetime, plotted in terms of ingoing Eddington-Finkelstein coordinates $(\tilde{t}, r)$: the solid (resp. dashed) lines correspond to outgoing (resp. ingoing) geodesics, as given by Eq. (6.49) (resp. Eq. (6.48)). The interiors of some future light cones are depicted in yellow.

considering the components (6.30). However, neither Eddington [97] nor Finkelstein [106] considered this null version: they used coordinates $(\tilde{t}, r, \theta, \varphi)$, where $\tilde{t}$ is timelike and they exhibited (the outgoing version of) the metric components (6.34). Hence our terminology is more faithful to history. Moreover, focusing on (v, r, θ, φ) may give the false impression to a novice reader that it is necessary to introduce some null coordinate to establish the regularity of the metric tensor at $r = 2m$, while the timelike coordinate $\tilde{t}$ does the job very well.

6.3.5 Radial null geodesics in terms of the Eddington-Finkelstein coordinates

By construction, the equation of the ingoing radial null geodesics in terms of the IEF coordinates is very simple:

$$\tilde{t} = -r + v, \tag{6.48}$$

where the constant $v \in \mathbb{R}$ labels the geodesic. The equation of the outgoing radial null geodesics is obtained by combining (6.22) and (6.33):

$$\tilde{t} = r + 4m \ln \left| \frac{r}{2m} - 1 \right| + u, \tag{6.49}$$

where the constant $u \in \mathbb{R}$ labels the geodesic. The radial null geodesics are depicted in Fig. 6.3 in terms of the IEF coordinates.

6.3.6 Time orientation of the spacetime manifold

From now on, we consider as *Schwarzschild spacetime* $(\mathscr{M}, \boldsymbol{g})$ the spacetime whose manifold is the largest one considered so far, i.e. the ingoing Eddington-Finkelstein domain:

$$\boxed{\mathscr{M} := \mathscr{M}_{\mathrm{IEF}} = \mathscr{M}_{\mathrm{I}} \cup \mathscr{H} \cup \mathscr{M}_{\mathrm{II}}}. \tag{6.50}$$

We have then $\mathscr{M} = \mathbb{R} \times (0, +\infty) \times \mathbb{S}^2$ [Eq. (6.41)]. Note that we shall extend this spacetime in Chap. 9.

We have seen in Sec. 6.3.1 that r is an affine parameter along the radial null geodesics. We may then choose $\lambda = -r$ as an affine parameter along the ingoing ones; according to Eq. (6.48), the equation of the ingoing radial null geodesics becomes then

$$\tilde{t}(\lambda) = \lambda + v, \quad r(\lambda) = -\lambda, \quad \theta(\lambda) = \theta_0, \quad \varphi(\lambda) = \varphi_0.$$

The tangent vector $\boldsymbol{k}$ associated with this parametrization has components

$$k^{\tilde{\alpha}} = \frac{\mathrm{d}x^{\tilde{\alpha}}}{\mathrm{d}\lambda} = (1, -1, 0, 0)$$

with respect to the IEF coordinates, i.e.

$$\boldsymbol{k} = \boldsymbol{\partial}_{\tilde{t}} - \boldsymbol{\partial}_{\tilde{r}}. \tag{6.51}$$

$\boldsymbol{k}$ is a nonzero null vector field defined on the whole manifold $\mathscr{M}$. It may therefore be used to set the time orientation of the Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$ (cf. Sec. 1.2.2). Since for $r \to +\infty$, $\boldsymbol{k}$ clearly points towards increasing $\tilde{t}$, we declare that $\boldsymbol{k}$ defines the *future* direction:

> The time orientation of the Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$ is such that the null vector $\boldsymbol{k}$ defined by Eq. (6.51) is everywhere future-directed.

The above choice induces a time orientation of the subdomains $\mathscr{M}_{\mathrm{I}}$ and $\mathscr{M}_{\mathrm{II}}$ of $\mathscr{M}$. In Sec. 6.2.5, we have noticed that the coordinate vector $\boldsymbol{\partial}_r$ of Schwarzschild-Droste coordinates is timelike in $\mathscr{M}_{\mathrm{II}}$; according to Lemma 2 of Sec. 1.2.2, we may then get its time orientation form the scalar product $\boldsymbol{k} \cdot \boldsymbol{\partial}_r$. Given Eqs. (6.36) and (6.38), we have

$$\boldsymbol{\partial}_r = -\left(1 - \frac{2m}{r}\right)^{-1} \boldsymbol{\partial}_{\tilde{t}} + \boldsymbol{\partial}_{\tilde{r}}.$$

Via (6.51) and (6.34), we deduce then that

$$\boldsymbol{k} \cdot \boldsymbol{\partial}_r = \frac{r}{2m}\left(1 - \frac{r}{2m}\right)^{-1} > 0 \quad \text{in } \mathscr{M}_{\mathrm{II}}.$$

In view of Eq. (1.4c) in Lemma 2 of Sec. 1.2.2, we conclude:

In region $\mathscr{M}_{\rm II}$, the vector $\boldsymbol{\partial}_r$ of Schwarzschild-Droste coordinates is a past-directed timelike vector.

This explains why the future null cones are along $-\boldsymbol{\partial}_r$ in Fig. 6.1 for $r < 2m$.

A corollary is:

In region $\mathscr{M}_{\rm II}$, r must decrease towards the future along any null or timelike worldline.

Proof. Let $\mathscr{L}$ be a causal curve in region $\mathscr{M}_{\rm II}$ and λ a parameter along $\mathscr{L}$ increasing towards the future. The associated tangent vector $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ is then future-directed. According to the above result, $-\boldsymbol{\partial}_r$ is a future-directed timelike vector in $\mathscr{M}_{\rm II}$, so that we can apply Lemma 1 of Sec. 1.2.2 with $\boldsymbol{u} = -\boldsymbol{\partial}_r$ and get $\boldsymbol{g}(-\boldsymbol{\partial}_r, \boldsymbol{v}) < 0$. Now, using the Schwarzschild-Droste components (6.15), we have

$$\boldsymbol{g}(-\boldsymbol{\partial}_r, \boldsymbol{v}) = -g_{r\mu} v^\mu = -g_{rr} v^r = -g_{rr} \frac{\mathrm{d}r}{\mathrm{d}\lambda} = \left(\frac{2m}{r} - 1 \right)^{-1} \frac{\mathrm{d}r}{\mathrm{d}\lambda}.$$

Since $2m/r - 1 > 0$ in $\mathscr{M}_{\rm II}$, $\boldsymbol{g}(-\boldsymbol{\partial}_r, \boldsymbol{v}) < 0$ is thus equivalent to $\mathrm{d}r/\mathrm{d}\lambda < 0$, which proves that r is decreasing along $\mathscr{L}$ as λ increases. □

Thus not only an observer in $\mathscr{M}_{\rm II}$ cannot cross $\mathscr{M}_{\rm II}$'s outer boundary $\mathscr{H}$ to reach $\mathscr{M}_{\rm I}$, $\mathscr{H}$ being a null hypersurface, but he is forced to move to decreasing r until he reaches the curvature singularity at $r \to 0$. We shall study this motion in detail in Sec. 7.3.2.

6.4 Black hole character

We have already seen in Sec. 6.3.3 that $\mathscr{H}$ is a Killing horizon. In particular, it is a null hypersurface, and thereby a one-way membrane (cf. Sec. 2.2.2). Since $\mathscr{H}$ is the boundary of $\mathscr{M}_{\rm II}$, we conclude that no particle nor electromagnetic signal may emerge from $\mathscr{M}_{\rm II}$ (this is pretty clear by looking to null geodesics on Fig. 6.3). Hence, with respect to the "outside" world, represented by the asymptotically flat region $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ is a black hole.

It would be satisfactory though to check that $\mathscr{M}_{\rm II}$ fulfills the formal definition of a black hole region that we have given in Sec. 4.4.2. The first step is to define a conformal completion at null infinity $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ of the Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$, as defined by Eq. (6.50). To this aim, let us start from the null ingoing Eddington-Finkelstein coordinates $(x^{\hat{\alpha}}) = (v, r, \theta, \varphi)$ introduced in Sec. 6.3.2; they cover entirely $\mathscr{M}$ and the metric tensor $\boldsymbol{g}$ is expressed in terms of them by Eq. (6.30). Performing the change of coordinates $(x^{\hat{\alpha}}) = (v, r, \theta, \varphi) \mapsto (x^{\alpha'}) = (v, x, \theta, \varphi)$ with

$$x = 1 - \frac{2m}{r} \iff r = \frac{2m}{1-x}, \qquad x \in (-\infty, 1), \tag{6.52}$$

we deduce from (6.30) that

$$g_{\mu'\nu'}\, \mathrm{d}x^{\mu'}\, \mathrm{d}x^{\nu'} = -x\, \mathrm{d}v^2 + \frac{4m}{(1-x)^2}\, \mathrm{d}v\, \mathrm{d}x + \frac{4m^2}{(1-x)^2} \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2 \right). \tag{6.53}$$

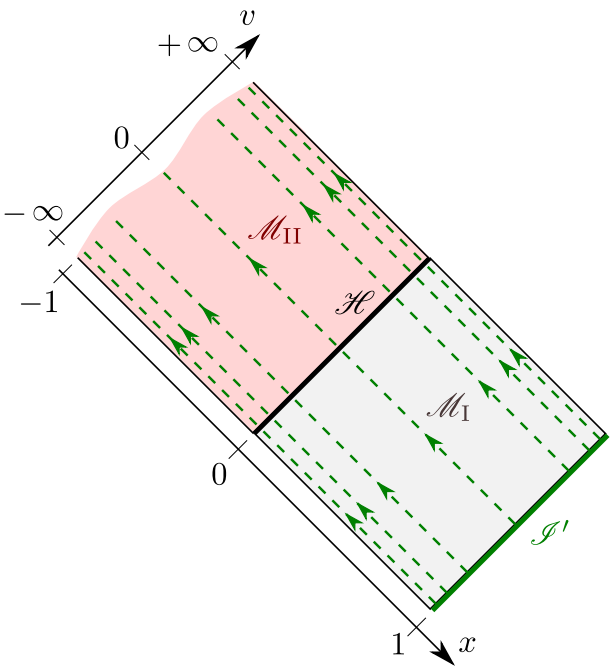


Figure 6.4: Manifold with boundary $\mathscr{M}' = \mathscr{M}_{\mathrm{II}} \cup \mathscr{H} \cup \mathscr{M}_{\mathrm{I}} \cup \mathscr{I}'$, drawn in terms of the coordinates x and (a compactified version of) v. The dashed lines are the ingoing radial null geodesics (as in Fig. 6.3), the arrows marking the future orientation.

Defining

$$\Omega := 1 - x = \frac{2m}{r}, \tag{6.54}$$

we may rewrite the metric tensor as

$$\boldsymbol{g} = \Omega^{-2} \tilde{\boldsymbol{g}}, \tag{6.55}$$

with $\tilde{\boldsymbol{g}}$ given by

$$\tilde{g}_{\mu'\nu'} \, \mathrm{d}x^{\mu'} \, \mathrm{d}x^{\nu'} = -x(1-x)^2 \, \mathrm{d}v^2 + 4m \, \mathrm{d}v \, \mathrm{d}x + 4m^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\varphi^2 \right). \tag{6.56}$$

Since (v, x, θ, φ) is a global coordinate system on $\mathscr{M}$ (up to the trivial coordinate singularities of (θ, φ)), we can identify $\mathscr{M}$ to the following open subset of $\mathbb{R}^2 \times \mathbb{S}^2$:

$$\mathscr{M} = \mathbb{R} \times (-\infty, 1) \times \mathbb{S}^2, \tag{6.57}$$

with v spanning $\mathbb{R}$, x spanning $(-\infty, 1)$ and (θ, φ) spanning $\mathbb{S}^2$. We can then extend $\mathscr{M}$ to the manifold with boundary[5]

$$\mathscr{M}' := \mathbb{R} \times (-\infty, 1] \times \mathbb{S}^2. \tag{6.58}$$

Notice the change $(-\infty, 1) \to (-\infty, 1]$ with respect to (6.57), which means that $x = 1$ is an allowed value on $\mathscr{M}'$; it actually defines the boundary of $\mathscr{M}'$, $\mathscr{I}'$ say. According to

[5] Cf. Sec. A.2.2 for the definition of a *manifold with boundary*.

(6.52), $\mathscr{I}'$ corresponds to $r \to +\infty$. A view of the manifold $\mathscr{M}'$ is provided in Fig. 6.4. We note that the conformal metric (6.56) can be extended to the boundary $\mathscr{I}'$, yielding a regular metric. Indeed, the determinant of the metric components (6.56) is

$$\det\left(\tilde{g}_{\alpha'\beta'}\right) = -64m^6 \sin^2\theta,$$

which does not vanish at $x = 1$ (except at the trivial coordinate singularity $\theta = 0$ or $\theta = \pi$), showing that $\tilde{\boldsymbol{g}}$ is a non-degenerate symmetric bilinear form at $\mathscr{I}'$ and hence a well-defined metric on all $\mathscr{M}'$. Furthermore we have $\Omega > 0$ on $\mathscr{M}$ and $\Omega = 0$ at $\mathscr{I}'$ [set $x = 1$ in Eq. (6.54)], as well as

$$\mathbf{d}\Omega = -\mathbf{d}x \neq 0. \tag{6.59}$$

Hence $(\mathscr{M}', \tilde{\boldsymbol{g}})$ obeys all the conditions listed in Sec. 4.3 to be a *conformal completion* of $(\mathscr{M}, \boldsymbol{g})$. However, it is not a proper *conformal completion at null infinity*, as defined in Sec. 4.3 and required in the black hole definition of Sec. 4.4.2. Indeed, any part of $\mathscr{I}'$ is intersected by a past-directed null geodesic (cf. Fig. 6.4): a generic point of $\mathscr{I}'$ has coordinates $(v, x, \theta, \varphi) = (v_0, 1, \theta_0, \varphi_0)$ and is the past end point of the ingoing radial null geodesic defined by $(v, \theta, \varphi) = (v_0, \theta_0, \varphi_0)$. So $\mathscr{I}'$ does not contain any future null infinity part ($\mathscr{I}^+$). Actually, we shall see below that $\mathscr{I}'$ is entirely a past null infinity ($\mathscr{I}^-$). Therefore, we shall extend $\mathscr{M}'$ to include some $\mathscr{I}^+$ part. To achieve this, we shall construct $\mathscr{I}^+$ as the set of endpoints of the *outgoing* radial null geodesics in $\mathscr{M}_{\rm I}$. In terms of the null ingoing Eddington-Finkelstein coordinates (v, r, θ, φ), the equation of these geodesics is obtained by combining (6.49) and (6.48):

$$v = 2r + 4m \ln\left|\frac{r}{2m} - 1\right| + u, \tag{6.60}$$

where $u \in \mathbb{R}$ is a constant parameter along a given geodesic. We notice that on $\mathscr{M}_{\rm I}$, we may use $(x^{\check{\alpha}}) = (u, r, \theta, \varphi)$ as a coordinate system, naturally called the ***null outgoing Eddington-Finkelstein coordinates***. Since (6.60) implies

$$\mathrm{d}v = \mathrm{d}u + \frac{2}{1 - 2m/r}\,\mathrm{d}r,$$

we easily deduce from (6.30) the metric components in these coordinates:

$$g_{\check{\mu}\check{\nu}}\,\mathrm{d}x^{\check{\mu}}\,\mathrm{d}x^{\check{\nu}} = -\left(1 - \frac{2m}{r}\right)\mathrm{d}u^2 - 2\,\mathrm{d}u\,\mathrm{d}r + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right). \tag{6.61}$$

Remark 1: Contrary to (v, r, θ, φ), the coordinates (u, r, θ, φ) do not cover all $\mathscr{M} = \mathscr{M}_{\rm IEF}$, but only $\mathscr{M}_{\rm I}$. This is graphically evident from Fig. 6.3, where the outgoing radial null geodesics, which are labelled by u, accumulate on $\mathscr{H}$ as $u \to +\infty$ from the $\mathscr{M}_{\rm I}$ side.

On $\mathscr{M}_{\rm I}$, let us perform the change of coordinates $(x^{\check{\alpha}}) = (u, r, \theta, \varphi) \to (x^{\alpha''}) = (u, x, \theta, \varphi)$, where x is related to r by the same formula as (6.52), except that on $\mathscr{M}_{\rm I}$, x's range is $(0, 1)$ only. We deduce from (6.61) and (6.52) the expression of $\boldsymbol{g}$ in terms of the coordinates (u, x, θ, φ):

$$g_{\mu''\nu''}\,\mathrm{d}x^{\mu''}\,\mathrm{d}x^{\nu''} = -x\,\mathrm{d}u^2 - \frac{4m}{(1-x)^2}\,\mathrm{d}u\,\mathrm{d}x + \frac{4m^2}{(1-x)^2}\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right). \tag{6.62}$$

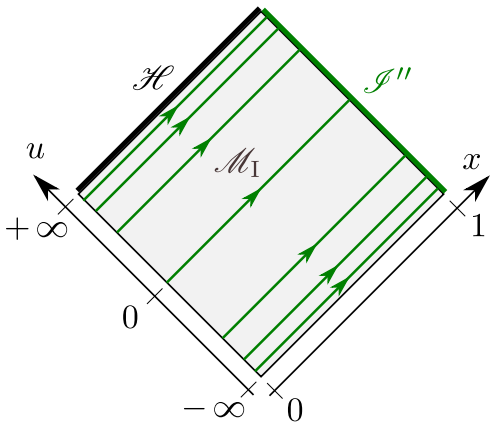


Figure 6.5: Manifold with boundary $\mathscr{M}_{\rm I}'' = \mathscr{M}_{\rm I} \cup \mathscr{I}''$, drawn in terms of the coordinates x and (a compactified version of) u. The green solid lines are the outgoing radial null geodesics (as in Fig. 6.3), the arrows marking the future orientation. Note that $\mathscr{H}$, which is drawn on this figure, is not part of $\mathscr{M}_{\rm I}''$.

Let us identify $\mathscr{M}_{\rm I}$ with the following open subset of $\mathbb{R}^2 \times \mathbb{S}^2$:

$$\mathscr{M}_{\rm I} = \mathbb{R} \times (0,1) \times \mathbb{S}^2, \tag{6.63}$$

with u spanning $\mathbb{R}$, x spanning $(0,1)$ and (θ, φ) spanning $\mathbb{S}^2$. Similarly to what we did above for $\mathscr{M}$, we may then extend $\mathscr{M}_{\rm I}$ to the manifold with boundary

$$\mathscr{M}_{\rm I}'' := \mathbb{R} \times (0,1] \times \mathbb{S}^2. \tag{6.64}$$

The boundary of $\mathscr{M}_{\rm I}''$, $\mathscr{I}''$ say, lies at $x = 1$ (cf. Fig. 6.5). It shall not be confused with the boundary of $\mathscr{M}_{\rm I}$ as a submanifold of $\mathscr{M}'$, which is $\mathscr{I}'$. The difference arises from the fact that u diverges (to $-\infty$) when one approaches $\mathscr{I}'$ in $\mathscr{M}'$, so that u cannot be used as a coordinate on $\mathscr{M}'$. This is clear on the relation (6.60) between u, v and r, which, once re-expressed in terms of x, becomes

$$u = v - 4m \left[\frac{1}{1-x} + \ln\left(\frac{x}{1-x} \right) \right]. \tag{6.65}$$

For a fixed value of v in $\mathscr{M}'$, this relation yields indeed diverging values of u at two places:

- $x \to 0^+$ (the horizon $\mathscr{H}$): $u \to +\infty$;
- $x \to 1^-$ (the boundary $\mathscr{I}'$): $u \to -\infty$.

Reciprocally, for a fixed value of u, relation (6.65) implies that v diverges (to $+\infty$) when $x \to 1^-$, which shows that $\mathscr{I}''$ is not included in $\mathscr{M}'$.

The conformal metric $\tilde{\boldsymbol{g}}$ on $\mathscr{M}_{\rm I}''$ is given by

$$\tilde{g}_{\mu''\nu''}\, \mathrm{d}x^{\mu''}\, \mathrm{d}x^{\nu''} = -x(1-x)^2\, \mathrm{d}u^2 - 4m\, \mathrm{d}u\, \mathrm{d}x + 4m^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2 \right). \tag{6.66}$$

We notice that it is regular and non-degenerate in all $\mathscr{M}_{\rm I}''$, including on $\mathscr{I}''$ ($x = 1$), and that on the submanifold $\mathscr{M}_{\rm I}$, it is related to the physical metric $\boldsymbol{g}$ by $\tilde{\boldsymbol{g}} = \Omega^2 \boldsymbol{g}$, with the

scalar field Ω taking the same expression in terms of x as that introduced in Eq. (6.54): $\Omega = 1 - x$.

The conformal completion of $(\mathcal{M}, \boldsymbol{g})$ including both $\mathscr{I}'$ (as $\mathscr{I}^-$) and $\mathscr{I}''$ (as $\mathscr{I}^+$) is constructed as follows. Let

$$\tilde{\mathcal{M}} = \mathcal{M}' \cup \mathcal{M}''_{\rm I}. \tag{6.67}$$

We endow $\tilde{\mathcal{M}}$ with two coordinate charts:

$$\begin{array}{rccc} \Phi_1: & \mathcal{M}' & \longrightarrow & \mathbb{R}\times(-\infty,1]\times\mathbb{S}^2 \\ & p & \longmapsto & (v,x,\theta,\varphi) \end{array} \quad \text{and} \quad \begin{array}{rccc} \Phi_2: & \mathcal{M}''_{\rm I} & \longrightarrow & \mathbb{R}\times(0,1]\times\mathbb{S}^2 \\ & p & \longmapsto & (u,x,\theta,\varphi) \end{array} \tag{6.68}$$

and define the intersection of the two chart codomains:

$$\mathcal{M}' \cap \mathcal{M}''_{\rm I} = \{p \in \mathcal{M}',\ x(p) \in (0,1)\} = \{p \in \mathcal{M}''_{\rm I},\ x(p) \in (0,1)\}, \tag{6.69}$$

along with the transition map implementing (6.65):

$$\begin{array}{rccc} \Phi_2 \circ \Phi_1^{-1}: & \mathbb{R}\times(0,1)\times\mathbb{S}^2 & \longrightarrow & \mathbb{R}\times(0,1)\times\mathbb{S}^2 \\ & (v,x,\theta,\varphi) & \longmapsto & \left(u = v - 4m\left[\frac{1}{1-x} + \ln\left(\frac{x}{1-x}\right)\right],\ x,\ \theta,\ \varphi\right), \end{array} \tag{6.70}$$

The above construction makes $\tilde{\mathcal{M}}$ a manifold with boundary (cf. Fig. 6.6), the boundary being

$$\mathscr{I} = \mathscr{I}^+ \cup \mathscr{I}^-, \tag{6.71}$$

with

$$\mathscr{I}^+ := \{p \in \mathcal{M}''_{\rm I},\ x(p) = 1\} \qquad \text{and} \qquad \mathscr{I}^- := \{p \in \mathcal{M}',\ x(p) = 1\}. \tag{6.72}$$

We then endow $\tilde{\mathcal{M}}$ with a Lorentzian metric $\tilde{\boldsymbol{g}}$, whose expression is given by (6.56) on $\mathcal{M}'$ and by (6.66) on $\mathcal{M}''_{\rm I}$. By construction, $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ is then a conformal completion at null infinity of the Schwarzschild spacetime $(\mathcal{M}, \boldsymbol{g})$, the conformal factor Ω being given by (6.54) in both charts $(\mathcal{M}', \Phi_1)$ and $(\mathcal{M}''_{\rm I}, \Phi_2)$: $\Omega = 1 - x$. In particular, it is clear that no past-directed causal curve originating in $\mathcal{M}$ intersects $\mathscr{I}^+$ and that no future-directed causal curve originating in $\mathcal{M}$ intersects $\mathscr{I}^-$. We also check immediately that $\mathscr{I}^+$ and $\mathscr{I}^-$ are null hypersurfaces with respect to the metric $\tilde{\boldsymbol{g}}$: both hypersurfaces are defined by $x = 1$, so that the induced metric on them, as deduced from (6.56) and (6.66), is

$$\tilde{\boldsymbol{g}}|_{\mathscr{I}^\pm} = 4m^2\left(\mathbf{d}\theta \otimes \mathbf{d}\theta + \sin^2\theta\, \mathbf{d}\varphi \otimes \mathbf{d}\varphi\right), \tag{6.73}$$

which is clearly degenerate (along the u direction for $\mathscr{I}^+$ and along the v direction for $\mathscr{I}^-$).

As it is clear from Fig. 6.6, $\mathcal{M}_{\rm I}$ is the interior of the causal past of $\mathscr{I}^+$ within $\mathcal{M}$:

$$\mathcal{M}_{\rm I} = \mathrm{int}\left(J^-(\mathscr{I}^+) \cap \mathcal{M}\right). \tag{6.74}$$

In view of the formal definition (4.35), we conclude that

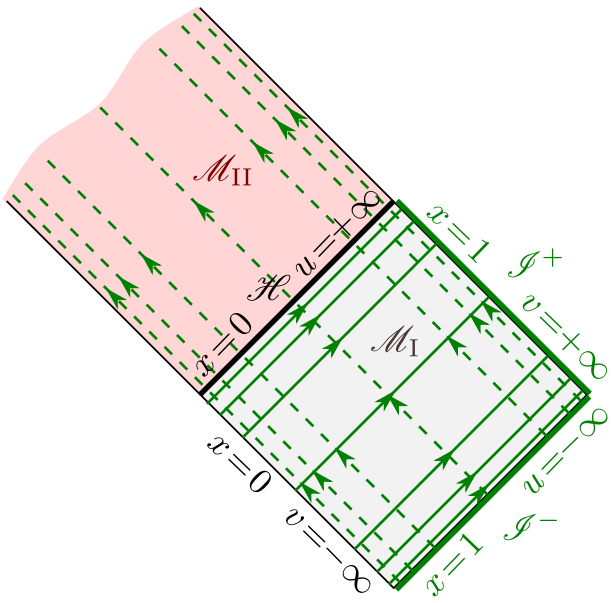


Figure 6.6: Schematic view of the manifold with boundary $\tilde{\mathscr{M}}$, which defines a conformal completion at null infinity of Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$. *NB:* contrary to Figs. 6.4 and 6.5, this figure is not drawn on some specific coordinate system. As in Figs. 6.3, 6.4 and 6.5, the green solid (resp. dashed) lines are the outgoing (resp. ingoing) radial null geodesics, the arrows marking the future orientation.

The Schwarzschild spacetime $(\mathscr{M} = \mathscr{M}_{\rm IEF}, \boldsymbol{g})$ has a black hole region $\mathscr{B}$, the interior of which is $\mathscr{M}_{\rm II}$; the event horizon is nothing but the Schwarzschild horizon $\mathscr{H}$ discussed in Sec. 6.3.3.

Remark 2: As stated at the beginning of this section, the null character of the boundary $\mathscr{H}$ between $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm II}$ and the fact that $\mathscr{M}_{\rm II}$ never intersect the asymptotically flat region $r \to +\infty$, was sufficient to claim that $\mathscr{M}_{\rm II}$ represents what by any means should be called a *black hole region*. Therefore, we can view the above demonstration more as a "sanity check" of the formal definition of a black hole given in Sec. 4.4.2: this definition would not have been acceptable if it would not apply to the Schwarzschild spacetime.

Remark 3: The above construction of the conformal completion at null infinity $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ involves two coordinate charts, (v, x, θ, φ) and (u, x, θ, φ), with two different domains, $\mathscr{M}'$ and $\mathscr{M}''_{\rm I}$. As will be discussed in Chap. 9, one may construct a conformal completion with a single chart, as in the Minkowski case, but its relation with the coordinates introduced so far is quite involved. In particular the standard compactification of Kruskal-Szekeres coordinates, which is used in many textbooks to construct the Carter-Penrose diagram of Schwarzschild spacetime, does *not* provide any conformal completion, as it will be discussed in Sec. 9.4.2.

Chapter 7

Geodesics in Schwarzschild spacetime: generic and timelike cases

Contents

7.1 Introduction

We have already investigated some geodesics in Schwarzschild spacetime in Chap. 6, namely the radial null geodesics (Sec. 6.3.1). Here, we perform an extensive study. After having established the main properties of generic causal (timelike or null) geodesics in Sec. 7.2, we investigate timelike geodesics in Sec. 11.4. They are of great physical importance, since they represent orbits of planets or stars around the black hole, as well as worldlines of intrepid observers freely falling into the black hole. The study of null geodesics, which govern images received by observers, is deferred to Chap. 8.

Before studying this chapter, the reader might want to have a look at Appendix B, which recaps the properties of geodesics in manifolds equipped with a metric. It could also be worth to read again Sec. 1.3 about worldlines of particles.

7.2 Geodesic motion

Let $\mathscr{L}$ be a geodesic[1] of Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$. We shall assume that $\mathscr{L}$ is causal, i.e. either timelike or null[2]. It therefore can be considered as the worldline of some particle $\mathscr{P}$, either massive ($\mathscr{L}$ timelike) or massless ($\mathscr{L}$ null). As recalled in

[1] The definition and basic properties of geodesics are recalled in Appendix B; see also Sec. 1.3.2.
[2] As shown in Sec. B.2.1, a geodesic cannot be partly timelike and partly null.

Sec. 1.3, the worldline of a particle $\mathscr{P}$ is a geodesic if, and only if, $\mathscr{P}$ is submitted only to gravitation, i.e. $\mathscr{P}$ is in free fall.

7.2.1 First integrals of motion

The Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$ is static and spherically symmetric; the Killing vector $\boldsymbol{\xi}$ associated with the staticity (cf. Sec. 6.2.2) and the Killing vector $\boldsymbol{\eta}$ associated with the rotation symmetry along some axis, give birth to two conserved quantities along $\mathscr{L}$:

Denoting by $\boldsymbol{p}$ the 4-momentum of particle $\mathscr{P}$ (cf. Sec. 1.3), the scalar products

$$\boxed{E := -\boldsymbol{\xi}\cdot\boldsymbol{p} = -\boldsymbol{g}(\boldsymbol{\xi},\boldsymbol{p})} \tag{7.1a}$$

$$\boxed{L := \boldsymbol{\eta}\cdot\boldsymbol{p} = \boldsymbol{g}(\boldsymbol{\eta},\boldsymbol{p})}, \tag{7.1b}$$

are constant along the geodesic $\mathscr{L}$. The scalar E is called $\mathscr{P}$'s ***conserved energy*** or ***energy at infinity***, while L is called $\mathscr{P}$'s ***conserved angular momentum*** or ***angular momentum at infinity***.

Proof. The 4-momentum $\boldsymbol{p}$ is a tangent vector associated with an affine parameter of $\mathscr{L}$, i.e. it obeys the geodesic equation (1.9). The constancy of E and L follow then from the generic property (B.42) of geodesics in presence of a spacetime symmetry. □

In coordinates (t, r, θ, φ) adapted to the spacetime symmetries, i.e. coordinates such that $\boldsymbol{\xi} = \boldsymbol{\partial}_t$ and $\boldsymbol{\eta} = \boldsymbol{\partial}_\varphi$, for instance the Schwarzschild-Droste coordinates or the Eddington-Finkelstein ones, one can rewrite (7.1) in terms of the components $p_t = g_{t\mu}\,p^\mu$ and $p_\varphi = g_{\varphi\mu}\,p^\mu$ of the 1-form $\underline{\boldsymbol{p}}$ associated to $\boldsymbol{p}$ by metric duality:

$$E = -p_t \tag{7.2a}$$

$$L = p_\varphi \tag{7.2b}$$

Indeed, in such a coordinate system, $\xi^\mu = \delta^\mu{}_t$ and $\eta^\mu = \delta^\mu{}_\varphi$, so that $E = -g_{\mu\nu}\,\xi^\mu p^\nu = -g_{t\nu}\,p^\nu = -p_t$ and $L = g_{\mu\nu}\,\eta^\mu p^\nu = g_{\varphi\nu}\,p^\nu = p_\varphi$.

It is worth stressing that E is not a genuine energy, i.e. it is not an energy measured by some observer. Indeed, the latter is defined by Eq. (1.21), which resembles Eq. (7.1a) but differs from it by $\boldsymbol{\xi}$ not being a unit vector in general: $\boldsymbol{\xi}\cdot\boldsymbol{\xi} \neq -1$. In other words, $\boldsymbol{\xi}$ cannot be interpreted as the 4-velocity of some observer, so that the quantity E defined by (7.1a) cannot be a *physically measured* particle energy. It is only in the asymptotic region, where $\boldsymbol{\xi}\cdot\boldsymbol{\xi} = g_{tt} \to -1$, that $\boldsymbol{\xi}$ is eligible as a 4-velocity, hence the name *energy at infinity*. Note that this name is commonly used, even in the particle $\mathscr{P}$ never visits the asymptotic region. Similarly, L is not some (component of a) genuine angular momentum. Only in the asymptotic region do we have

$$L \simeq g_{\varphi\varphi}\,p^\varphi \simeq r^2\sin^2\theta\,p^\varphi \simeq r^2\sin^2\theta\,P^\varphi \simeq r\sin\theta\,P^{(\varphi)}, \tag{7.3}$$

where $P^{(\varphi)}$ is the azimuthal component of the momentum $\boldsymbol{P}$ of particle $\mathscr{P}$ as measured by an asymptotic inertial observer $\mathscr{O}$ (cf. Sec. 1.4), i.e. the component of $\boldsymbol{P}$ along $\boldsymbol{e}_{(\varphi)}$ in the

orthonormal basis $(\boldsymbol{e}_{(r)}, \boldsymbol{e}_{(\theta)}, \boldsymbol{e}_{(\varphi)})$, with $\boldsymbol{e}_{(\varphi)} = (r\sin\theta)^{-1}\boldsymbol{\partial}_\varphi$. In view of (7.3), we may say that L is the angular momentum about the symmetry axis $\theta = 0$ that $\mathscr{O}$ would attribute to particle $\mathscr{P}$ if the latter would move close to him. Equivalently, in a Cartesian coordinate system defined by $(x, y, z) = (r\sin\theta\cos\varphi, r\sin\theta\sin\varphi, r\cos\theta)$, L is the component $L^z_{\rm tot}$ of the total angular momentum of $\mathscr{P}$ as measured by $\mathscr{O}$:

$$\boldsymbol{L}_{\rm tot} := \boldsymbol{r} \times \boldsymbol{P} = L^x_{\rm tot}\, \boldsymbol{\partial}_x + L^y_{\rm tot}\, \boldsymbol{\partial}_y + L\, \boldsymbol{\partial}_z. \tag{7.4}$$

From its very definition, Eq. (7.1a), E is a positive quantity as soon as the geodesic $\mathscr{L}$ has some part in $\mathscr{M}_{\rm I}$, i.e. some part with $r > 2m$:

$$\mathscr{L} \cap \mathscr{M}_{\rm I} \neq \varnothing \quad \Longrightarrow \quad E > 0. \tag{7.5}$$

Proof. In $\mathscr{M}_{\rm I}$, the Killing vector $\boldsymbol{\xi}$ is timelike and future-directed. The 4-momentum $\boldsymbol{p}$ is either timelike or null and always future-directed. By Eq. (1.3a) in Lemma 1 of Sec. 1.2.2, one has then necessarily $\boldsymbol{\xi}\cdot\boldsymbol{p} < 0$; hence Eq. (7.1a) implies $E > 0$ in $\mathscr{M}_{\rm I}$. Since E is constant along $\mathscr{L}$, it follows that $E > 0$ everywhere. □

Remark 1: If the geodesic $\mathscr{L}$ is confined to $\mathscr{M}_{\rm II}$, i.e. to the black hole region (cf. Sec. 6.4), where $\boldsymbol{\xi}$ is spacelike (cf. Sec. 6.2.5), it is possible to have $E \leq 0$, since the scalar product of $\boldsymbol{p}$ with a spacelike vector can take any value.

Remark 2: The Killing vector $\boldsymbol{\eta}$ being always spacelike, the scalar product $\boldsymbol{g}(\boldsymbol{\eta}, \boldsymbol{p})$ can a priori take any real value, and thus there is no constraint on the sign of L.

To be specific, let us describe Schwarzschild spacetime in terms of the Schwarzschild-Droste coordinates (t, r, θ, φ) introduced in Sec. 6.2.3. Without any loss of generality, we may choose these coordinates so that at $t = 0$, the particle $\mathscr{P}$ is located in the equatorial plane $\theta = \pi/2$ and the spatial projection of the worldline $\mathscr{L}$ lies in that plane, i.e. $\boldsymbol{p}$ has no component along $\boldsymbol{\partial}_\theta$:

$$\boldsymbol{p} \overset{t=0}{=} p^t \boldsymbol{\partial}_t + p^r \boldsymbol{\partial}_r + p^\varphi \boldsymbol{\partial}_\varphi. \tag{7.6}$$

Now, for $t > 0$, if the geodesic $\mathscr{L}$ were departing from $\theta = \pi/2$, this would constitute some breaking of spherical symmetry, making a difference between the "Northern" hemisphere and the "Southern" one. Hence[3] $\mathscr{L}$ must stay at $\theta = \pi/2$, which implies

$$\boxed{p^\theta = 0}. \tag{7.7}$$

We conclude that

[3] More rigorously, Eq. (7.7) can be derived from the geodesic equation (1.9): given the expression of the Christoffel symbols of $\boldsymbol{g}$ in Schwarzschild-Droste coordinates (cf. Sec. D.3.2), Eq. (1.9) yields

$$\frac{{\rm d}p^\theta}{{\rm d}\lambda} + \frac{2}{r} p^r p^\theta - \sin\theta\cos\theta\, (p^\varphi)^2 = 0,$$

where λ is the affine parameter of $\mathscr{L}$ associated with $\boldsymbol{p}$, so that $p^\theta = {\rm d}\theta/{\rm d}\lambda$. Whatever the values of $r(\lambda)$, $p^r(\lambda)$ and $p^\varphi(\lambda)$, the solution of this ordinary differential equation with the initial conditions $p^\theta = 0$ and $\cos\theta = 0$ is $p^\theta = 0$ for all values of λ.

A geodesic $\mathscr{L}$ of Schwarzschild spacetime is necessarily confined to a timelike hypersurface. Without any loss of generality, we can choose Schwarzschild-Droste coordinates (t, r, θ, φ) such that this hypersurface is the "equatorial hyperplane" $\theta = \pi/2$. Then the component p^θ of the 4-momentum of the particle having $\mathscr{L}$ as worldline vanishes identically [Eq. (7.7)].

Let us denote by μ the mass of particle $\mathscr{P}$, with possibly $\mu = 0$ if $\mathscr{P}$ is a photon. The scalar square of the 4-momentum $\boldsymbol{p}$ is then [cf. Eq. (1.8)]

$$\boldsymbol{g}(\boldsymbol{p}, \boldsymbol{p}) = -\mu^2. \tag{7.8}$$

So μ^2 can be seen as an integral of motion.

7.2.2 Equations of motion and generic properties

Contemplating Eqs. (7.1a), (7.1b), (7.7) and (7.8), we realize that we have four first integral of motions. The problem is then completely integrable. More specifically, let λ be the affine parameter along the geodesic $\mathscr{L}$ associated with the 4-momentum $\boldsymbol{p}$ [cf. Eq. (B.2)]:

$$\boxed{\boldsymbol{p} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}}, \tag{7.9}$$

where $\mathrm{d}\boldsymbol{x}$ is the infinitesimal displacement along $\mathscr{L}$ corresponding to the parameter change $\mathrm{d}\lambda$. Note that λ is dimensionless and necessarily increases towards the future[4], since $\boldsymbol{p}$ is by definition future-oriented (cf. Sec. 1.3.1). In terms of the components with respect to Schwarzschild-Droste coordinates, this yields

$$\dot{t} := \frac{\mathrm{d}t}{\mathrm{d}\lambda} = p^t, \qquad \dot{r} := \frac{\mathrm{d}r}{\mathrm{d}\lambda} = p^r, \qquad \dot{\theta} := \frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = p^\theta, \qquad \dot{\varphi} := \frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = p^\varphi. \tag{7.10}$$

In the present case, where $\theta(\lambda) = \pi/2$, we have of course $\dot{\theta} = 0$, in agreement with Eq. (7.7). Given the components (6.15) of Schwarzschild metric with respect to the Schwarzschild-Droste coordinates, Eq. (7.1a) can be written as

$$E = -g_{t\mu}p^\mu = -g_{tt}p^t = -g_{tt}\dot{t} = \left(1 - \frac{2m}{r}\right)\dot{t},$$

hence

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\lambda} = E\left(1 - \frac{2m}{r}\right)^{-1}}. \tag{7.11}$$

Similarly, Eq. (7.1b) becomes

$$L = g_{\varphi\mu}p^\mu = g_{\varphi\varphi}p^\varphi = g_{\varphi\varphi}\dot{\varphi} = r^2\sin^2\theta\,\dot{\varphi}.$$

[4] Let us recall that Schwarzschild spacetime is time-oriented, cf. Sec. 6.3.6.

Since $\theta = \pi/2$, we get

$$\boxed{\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{L}{r^2}}. \tag{7.12}$$

We have already noticed that the sign of L is unconstrained (Remark 2 on p. 143). The above equation shows that it corresponds to the increase ($L > 0$) or decrease ($L < 0$) of φ along the geodesic $\mathscr{L}$. In other words, we deduce from Eq. (7.12) that

Along any timelike or null geodesic of Schwarzschild spacetime, the azimuthal coordinate φ is either constant ($L = 0$) or increases (resp. decreases) monotonically ($L > 0$) (resp. $L < 0$).

The last unexploited first integral of motion is Eq. (7.8); it yields

$$-\left(1 - \frac{2m}{r}\right)(\dot{t})^2 + \left(1 - \frac{2m}{r}\right)^{-1}(\dot{r})^2 + r^2(\dot{\theta})^2 + r^2\sin^2\theta(\dot{\varphi})^2 = -\mu^2.$$

Using (7.11), (7.12), as well as $\dot{\theta} = 0$ and $\theta = \pi/2$, we get

$$-E^2\left(1 - \frac{2m}{r}\right)^{-1} + \left(1 - \frac{2m}{r}\right)^{-1}(\dot{r})^2 + \frac{L^2}{r^2} = -\mu^2,$$

which can be recast as

$$\boxed{\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 - \frac{2\mu^2 m}{r} + \frac{L^2}{r^2}\left(1 - \frac{2m}{r}\right) = E^2 - \mu^2}. \tag{7.13}$$

To summarize, the geodesic motion in Schwarzschild spacetime is governed by Eqs. (7.11), (7.12) and (7.13), where $r = r(\lambda)$ and μ, E and L are constants. This constitutes a system of 3 differential equations for the 3 unknown functions $t(\lambda)$, $r(\lambda)$ and $\varphi(\lambda)$. We observe that Eq. (7.13) is decoupled from the other two equations. The task is then to first solve this equation for $r(\lambda)$ and to inject the solution into Eqs. (7.11) and (7.12), which can then be integrated separately.

A constraint to keep in mind is that the 4-momentum vector $\boldsymbol{p}$, whose components are related to the solution $(t(\lambda), r(\lambda), \varphi(\lambda))$ by Eq. (7.10), has to be a future-directed causal vector. In $\mathscr{M}_{\mathrm{I}}$, as we have seen above, this is guaranteed by choosing $E > 0$ [cf. Eq. (7.5)]. In $\mathscr{M}_{\mathrm{II}}$, a future-directed timelike vector is $-\boldsymbol{\partial}_r$ (cf. Sec. 6.3.6). According to Eq. (1.3a) in Lemma 1 of Sec. 1.2.2, we have then $\boldsymbol{p}$ future-directed iff $-\boldsymbol{\partial}_r \cdot \boldsymbol{p} < 0$, i.e. iff

$$\left(\frac{2m}{r} - 1\right)^{-1} p^r < 0.$$

Since $2m/r - 1 > 0$ in $\mathscr{M}_{\mathrm{II}}$, this is equivalent to $p^r < 0$, i.e. to $\mathrm{d}r/\mathrm{d}\lambda < 0$. Hence

In the black hole region $\mathscr{M}_{\rm II}$, i.e. for $r < 2m$, the solution $r(\lambda)$ of Eq. (7.13) must be a strictly decreasing function of λ.

Actually, we recover the result stated for any causal worldline (not necessarily a geodesic) in Sec. 6.3.6.

Remark 3: We have derived the system of Eqs. (7.11), (7.12) and (7.13) without invoking explicitly the famous *geodesic equation*, i.e. Eq. (B.10) in Appendix B. This is because we had enough first integrals of the second-order differential equation (B.10) to completely reduce it to a system of first order equations.

7.2.3 Trajectories in the orbital plane

If the conserved angular momentum vanishes, $L = 0$, the equation of motion (7.12) implies that $\varphi = \text{const} = \varphi_0$. The geodesic $\mathscr{L}$ is then confined to the 2-dimensional timelike surface $(\theta, \varphi) = (\pi/2, \varphi_0)$, which is spanned by the coordinates (t, r). One says that $\mathscr{L}$ is a ***radial geodesic***.

In the remainder of this section, we discuss the opposite case, namely we assume

$$L \neq 0. \tag{7.14}$$

We have stressed above that φ is then a strictly monotonic function of λ, increasing (resp. decreasing) continuously along $\mathscr{L}$ for $L > 0$ (resp. $L < 0$). Consequently,

Along any timelike or null geodesic with $L \neq 0$, φ can be chosen as a parameter, provided one does not restrict its range to $(0, 2\pi)$.

Contrary to λ, φ is not in general an affine parameter of $\mathscr{L}$. Indeed, the dependency $r = r(\lambda)$ in Eq. (7.12) does not correspond to an affine relation between φ and λ, except for $r(\lambda) = \text{const}$ (case of circular orbits).

Let us use Eq. (7.12) to write $\mathrm{d}r/\mathrm{d}\lambda = \mathrm{d}r/\mathrm{d}\varphi \times \mathrm{d}\varphi/\mathrm{d}\lambda = L/r^2 \mathrm{d}r/\mathrm{d}\varphi$ and substitute this expression into the equation of motion (7.13). We get

$$\frac{1}{r^4}\left(\frac{\mathrm{d}r}{\mathrm{d}\varphi}\right)^2 = \frac{2m}{r^3} - \frac{1}{r^2} + 2\left(\frac{\mu}{L}\right)^2 \frac{m}{r} + \left(\frac{E}{L}\right)^2 - \left(\frac{\mu}{L}\right)^2. \tag{7.15}$$

To simplify this equation, it is natural to introduce the dimensionless variable

$$\boxed{u := \frac{m}{r}} \tag{7.16}$$

instead of r. We then get

$$\boxed{\left(\frac{\mathrm{d}u}{\mathrm{d}\varphi}\right)^2 = 2u^3 - u^2 + 2\left(\frac{m\mu}{L}\right)^2 u + \left(\frac{mE}{L}\right)^2 - \left(\frac{m\mu}{L}\right)^2}. \tag{7.17}$$

This differential equation determines entirely the (r,φ)-part of the geodesic $\mathscr{L}$, which we shall call the ***trajectory of $\mathscr{L}$ in the orbital plane***. The term "orbital plane" is a slight abuse of language for the 2-dimensional *surface* $(t,\theta)=(t_0,\pi/2)$, where t_0 is a constant.

In general Eq. (7.17) is not solvable in terms of elementary functions. The exceptions are circular orbits ($u=$ const, the constant being one of the roots of the cubic polynomial in u in the right-hand side) and the critical null geodesics, which we shall discuss in Sec. 8.3.2. For the generic case, exact solutions are expressible in terms of some non-elementary special functions; there are basically two strategies:

- The first one is to invoke the *Weierstrass elliptic function*[5] $\wp(z;\omega_1,\omega_2)$, which is a doubly-periodic meromorphic function of the complex variable z, of periods $\omega_1\in\mathbb{C}$ and $\omega_2\in\mathbb{C}$. Among the many properties of this function, the one relevant here is that $\wp$ is a solution of the differential equation

$$\left(\wp'(z)\right)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3, \tag{7.18}$$

  where g_2 and g_3 are two constants entirely determined by the periods ω_1 and ω_2 of $\wp$. Indeed, via the change of variables $v:=u-1/6$ and $\tilde{\varphi}:=\varphi/\sqrt{2}$, Eq. (7.17) is equivalent to

$$\left(\frac{\mathrm{d}v}{\mathrm{d}\tilde{\varphi}}\right)^2 = 4v^3 - \left[\frac{1}{3} - 4\left(\frac{m\mu}{L}\right)^2\right]v + 2\left(\frac{mE}{L}\right)^2 - \frac{4}{3}\left(\frac{m\mu}{L}\right)^2 - \frac{1}{27}, \tag{7.19}$$

  which is obviously of type (7.18) (no square in the cubic polynomial). The solution is thus

$$u = \frac{m}{r} = \wp\left(\frac{\varphi}{\sqrt{2}} + C;\omega_1,\omega_2\right) + \frac{1}{6}, \tag{7.20}$$

  where $C\in\mathbb{C}$ is a constant and ω_1 and ω_2 are determined by m, μ, E and L (see e.g. Ref. [117] for more details regarding this method applied to null geodesics).

- The second approach consists in noticing that the method of separation of variables can easily be applied to Eq. (7.17), leading to

$$\varphi = \pm\int_{u_0}^{u}\frac{\mathrm{d}\bar{u}}{\sqrt{2\bar{u}^3 - \bar{u}^2 + 2\left(\frac{m\mu}{L}\right)^2\bar{u} + \left(\frac{mE}{L}\right)^2 - \left(\frac{m\mu}{L}\right)^2}} + \varphi_0, \tag{7.21}$$

  where u_0 and φ_0 are two constants, and the $\pm$ sign can be $+$ on some parts of the geodesic $\mathscr{L}$ and $-$ on some other parts of $\mathscr{L}$. The integral in right-hand side is expressible in terms of the so-called *incomplete elliptic integrals of the first kind*. We shall detail such a technique for null geodesics in Sec. 8.3. Note that this approach leads to $\varphi=\varphi(r)$, whereas the method involving the Weierstrass function leads to the "polar equation" form: $r=r(\varphi)$. It is possible though to get the polar form by invoking the inverses of elliptic integrals, namely *Jacobi elliptic functions*.

[5] The character $\wp$ is a kind of calligraphic lowercase p, which is standard to denote this function.

Once the solution $r = r(\varphi)$ of Eq. (7.17) has been obtained, it can be injected into the equation for $t = t(\varphi)$ that can be deduced from the equations of motions (7.11)-(7.12):

$$\frac{\mathrm{d}t}{\mathrm{d}\varphi} = \frac{E}{L}\frac{r(\varphi)^3}{r(\varphi) - 2m}. \tag{7.22}$$

This is an ordinary differential equation for $t = t(\varphi)$, the solution of which amounts to finding a primitive with respect to φ of the right-hand side. Unfortunately, this is not an easy task in general, the function $r(\varphi)$ being quite involved, except for circular orbits ($r(\varphi) = \text{const}$).

In what follows, we discuss separately the resolution of the system (7.11)-(7.13) or of Eq. (7.17) for timelike geodesics (Sec. 11.4) and for null geodesics (Chap. 8).

7.3 Timelike geodesics

7.3.1 Effective potential

When the geodesic $\mathscr{L}$ is timelike, it is natural to use the proper time τ as an affine parameter along it, instead of the parameter λ associated with the 4-momentum $\boldsymbol{p}$. Since the tangent vector associated with τ is the 4-velocity $\boldsymbol{u}$ (cf. Sec. 1.3.3) and $\boldsymbol{p}$ and $\boldsymbol{u}$ are related by Eq. (1.16): $\boldsymbol{p} = \mu\,\boldsymbol{u}$, we get $\mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda = \mu\,\mathrm{d}\boldsymbol{x}/\mathrm{d}\tau$, from which we infer the relation between τ and λ:

$$\tau = \mu\lambda, \tag{7.23}$$

up to some additive constant. This is of course a special case of the generic relation (B.3) between two affine parameters of the same geodesic. Equation (7.13) becomes then

$$\boxed{\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + V_\ell(r) = \frac{\varepsilon^2 - 1}{2}}, \tag{7.24}$$

where

$$\boxed{V_\ell(r) := -\frac{m}{r} + \frac{\ell^2}{2r^2}\left(1 - \frac{2m}{r}\right)} \tag{7.25}$$

and ε and ℓ are respectively the ***specific conserved energy*** and ***specific conserved angular momentum*** of particle $\mathscr{P}$:

$$\boxed{\varepsilon := \frac{E}{\mu} = -\boldsymbol{\xi}\cdot\boldsymbol{u}} \quad \text{and} \quad \boxed{\ell := \frac{L}{\mu} = \boldsymbol{\eta}\cdot\boldsymbol{u}}, \tag{7.26}$$

where $\boldsymbol{u}$ is the 4-velocity of $\mathscr{P}$ and the second equalities result from definitions (7.1) and the relation $\boldsymbol{p} = \mu\,\boldsymbol{u}$ [Eq. (1.16)]. Note that ε is dimensionless (in units $c = 1$) and that it shares the same positiveness property (7.5) as E, namely ε is positive as soon as the timelike geodesic $\mathscr{L}$ has some part in $\mathscr{M}_{\rm I}$, i.e. some part with $r > 2m$:

$$\mathscr{L} \cap \mathscr{M}_{\rm I} \neq \varnothing \quad \Longrightarrow \quad \varepsilon > 0. \tag{7.27}$$

On the contrary, ℓ can be either positive, zero or negative, depending on the variation of φ along $\mathscr{L}$, as was already noticed above for L.

We note that Eq. (7.24) has the shape of the first integral of the 1-dimensional motion of a non-relativist particle in the potential V_ℓ (called hereafter the ***effective potential***), the term $1/2\,(\mathrm{d}r/\mathrm{d}\tau)^2$ being interpreted as the kinetic energy per unit mass, $V_\ell(r)$ as the potential energy per unit mass and the constant right-hand side $(\varepsilon^2-1)/2$ as the total mechanical energy per unit mass.

Remark 1: The effective potential (7.25) differs from its non-relativistic (Newtonian) counterpart only by the factor $1-2m/r$ instead of 1. This difference plays an important role for small values of r, leading to some orbital instability, as we shall see in Sec. 7.3.3.

In $\mathscr{M}_{\rm I}$, where the Killing vector $\boldsymbol{\xi}$ is timelike, we may introduce the ***static observer*** $\mathscr{O}$, whose 4-velocity $\boldsymbol{u}_{\mathscr{O}}$ is collinear to $\boldsymbol{\xi}$:

$$\boldsymbol{u}_{\mathscr{O}} = \left(1-\frac{2m}{r}\right)^{-1/2} \boldsymbol{\xi}, \tag{7.28}$$

the proportionality coefficient ensuring that $\boldsymbol{u}_{\mathscr{O}} \cdot \boldsymbol{u}_{\mathscr{O}} = -1$ given that $\boldsymbol{\xi}\cdot\boldsymbol{\xi} = g_{tt} = -(1-2m/r)$. We have then, from (7.26),

$$\varepsilon = -\left(1-\frac{2m}{r}\right)^{1/2} \boldsymbol{u}_{\mathscr{O}} \cdot \boldsymbol{u} = \Gamma \left(1-\frac{2m}{r}\right)^{1/2}, \tag{7.29}$$

where $\Gamma = -\boldsymbol{u}_{\mathscr{O}}\cdot\boldsymbol{u}$ is the Lorentz factor of $\mathscr{P}$ with respect to $\mathscr{O}$ (cf. Sec. 1.4; in particular Eq. (1.31)). We may express Γ in terms of the norm v of the velocity of $\mathscr{P}$ with respect to $\mathscr{O}$, according to Eq. (1.33): $\Gamma = (1-v^2)^{-1/2}$ and get

$$\varepsilon = \left(1-v^2\right)^{-1/2} \left(1-\frac{2m}{r}\right)^{1/2}. \tag{7.30}$$

In the region $r \gg m$, we may perform a first order expansion, assuming that $\mathscr{P}$ moves at nonrelativistic velocity with respect to $\mathscr{O}$ ($v \ll 1$), thereby obtaining:

$$\boxed{\varepsilon - 1 \simeq \frac{1}{2}v^2 - \frac{m}{r}} \qquad (r \gg m \quad \text{and} \quad v \ll 1)\,. \tag{7.31}$$

We recognize in the right-hand side the *Newtonian mechanical energy per unit mass* of particle $\mathscr{P}$ with respect to observer $\mathscr{O}$, who can then considered as an inertial observer, $v^2/2$ being the kinetic energy per unit mass and $-m/r$ the gravitational potential energy per unit mass.

The profile of $V_\ell(r)$ for selected values of ℓ is plotted in Figs. 7.1 and 7.2 . Its extrema are given by $\mathrm{d}V_\ell/\mathrm{d}r = 0$, which is equivalent to

$$mr^2 - \ell^2 r + 3\ell^2 m = 0.$$

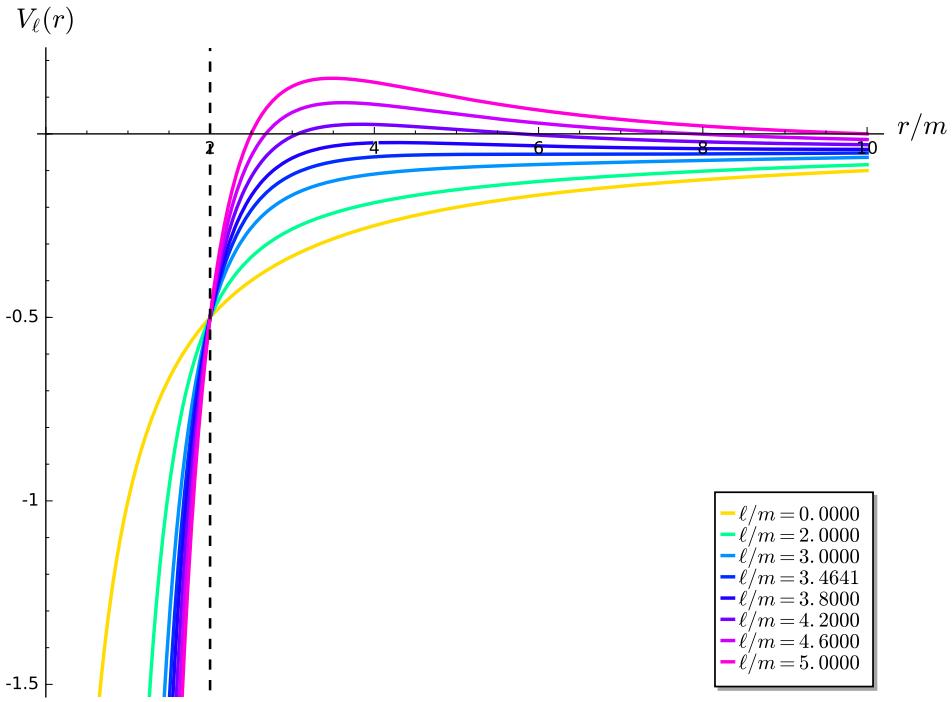


Figure 7.1: Effective potential $V_\ell(r)$ governing the r-part of the motion along a timelike geodesic in Schwarzschild spacetime via Eq. (7.24). The vertical dashed line marks $r = 2m$, i.e. the location of the event horizon. The numerical value $\ell/m = 3.4641$ is that of the critical specific angular momentum (7.32).

This quadratic equation admits real roots iff $|\ell| \geq \ell_{\rm crit}$, with

$$\boxed{\ell_{\rm crit} = 2\sqrt{3}\, m \simeq 3.464102\, m}. \tag{7.32}$$

For $|\ell| \geq \ell_{\rm crit}$, the two roots are

$$r_{\rm max} = \frac{\ell}{2m}\left(\ell - \sqrt{\ell^2 - \ell_{\rm crit}^2}\right) \qquad \text{and} \qquad r_{\rm min} = \frac{\ell}{2m}\left(\ell + \sqrt{\ell^2 - \ell_{\rm crit}^2}\right), \tag{7.33}$$

corresponding respectively to a maximum of V_ℓ and a minimum of V_ℓ, hence the indices "max" and "min". Note that $r_{\rm max} \leq r_{\rm min}$. In the marginal case $|\ell| = \ell_{\rm crit}$, the two roots coincide and correspond to an inflection point of V_ℓ (the circled dot in Fig. 7.2).

For $|\ell| < \ell_{\rm crit}$, there is no extremum and V_ℓ is a strictly increasing function of r.

To get a full solution in terms of the Schwarzschild-Droste coordinates, once Eq. (7.24) is solved for $r(\tau)$, one has still to solve Eqs. (7.11) and (7.12), which can be rewritten in terms of the proper time τ as

$$\frac{{\rm d}t}{{\rm d}\tau} = \varepsilon\left(1 - \frac{2m}{r(\tau)}\right)^{-1}, \tag{7.34}$$

$$\frac{{\rm d}\varphi}{{\rm d}\tau} = \frac{\ell}{r(\tau)^2}. \tag{7.35}$$

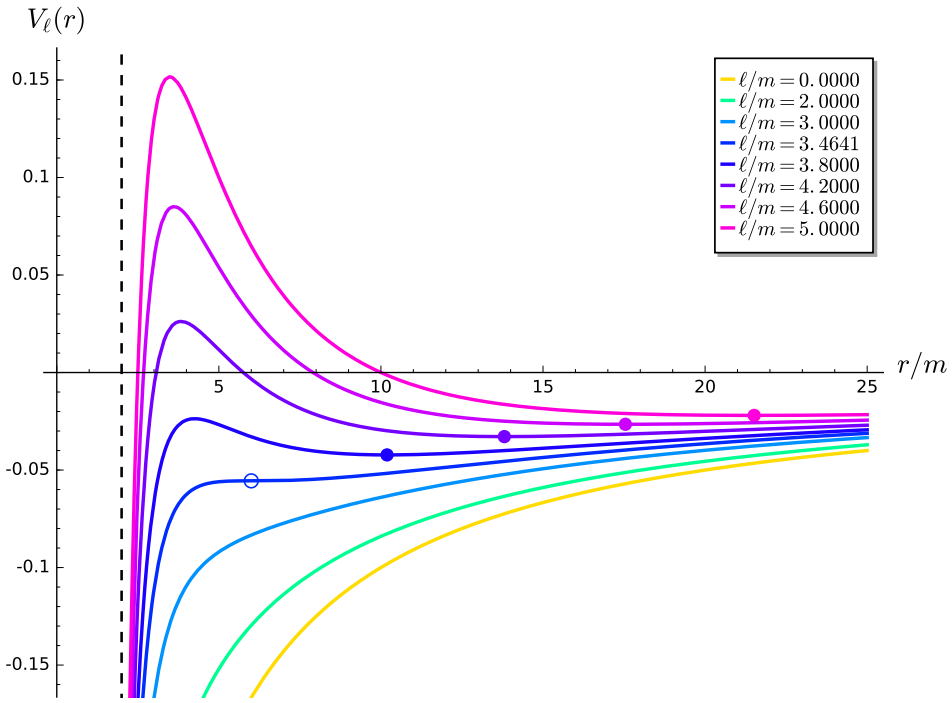


Figure 7.2: Same as Fig. 7.1, but with a zoom in along the y-axis and a zoom out along the x-axis. The dots mark the mimima of V_ℓ, locating stable circular orbits.

7.3.2 Radial free fall

Generic case

The radial geodesics correspond to a vanishing conserved angular momentum: $\ell = 0$. Indeed, setting $\ell = 0$ in Eq. (7.12) yields $\varphi = \text{const}$, which defines a purely radial trajectory in the plane $\theta = \pi/2$. The effective potential (7.25) reduces then to $V_\ell(r) = -m/r$, so that the equation of radial motion (7.24) becomes

$$\frac{1}{2}\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 - \frac{m}{r} = \frac{\varepsilon^2 - 1}{2}. \tag{7.36}$$

This equation is identical to that governing radial free fall in the gravitational field generated by a mass m in Newtonian gravity. The solution is well known and depends on the sign of the "mechanical energy" in the right-hand side, i.e. of the position of ε with respect to 1:

- if $\varepsilon > 1$, the solution is given in parameterized form (parameter η) by

$$\begin{cases} \tau = \dfrac{m}{(\varepsilon^2-1)^{3/2}}\left(\sinh\eta - \eta\right) + \tau_0 \\ r = \dfrac{m}{\varepsilon^2 - 1}\left(\cosh\eta - 1\right), \end{cases} \tag{7.37}$$

- if $\varepsilon = 1$, the solution is

$$r(\tau) = \left(\frac{9m}{2}(\tau - \tau_0)^2\right)^{1/3}, \tag{7.38}$$

- if $\varepsilon < 1$, the solution is given in parameterized form (parameter η) by

$$\begin{cases} \tau = \dfrac{m}{(1-\varepsilon^2)^{3/2}}\,(\eta + \sin\eta) + \tau_0 \\ r = \dfrac{m}{1-\varepsilon^2}\,(1 + \cos\eta)\,, \end{cases} \tag{7.39}$$

In the above formulas, τ_0 is a constant; for $\varepsilon \geq \mu$, τ_0 is the value of τ for which $r \to 0$, while for $\varepsilon < 1$, it is the value of τ at which r takes its maximal value.

Radial free fall from rest

Let us focus on the radial free fall from rest, starting at some position $r = r_0$ at $\tau = 0$. Starting from rest means $\mathrm{d}r/\mathrm{d}\tau = 0$ at $\tau = 0$. The equation of radial motion (7.36) leads then to $-m/r_0 = (\varepsilon^2 - 1)/2$, or equivalently

$$\varepsilon^2 = 1 - \frac{2m}{r_0}. \tag{7.40}$$

The right-hand side of this equation must be non-negative. This implies $r_0 \geq 2m$. We recover the fact that one cannot be momentarily at rest (in terms of r) if $r_0 < 2m$, for r has to decrease along any causal geodesic in the black hole region $\mathscr{M}_{\mathrm{II}}$ (cf. Sec. 7.2.2).

Equation (7.40) implies $\varepsilon < 1$, i.e. $E < \mu$. The solution is thus given by Eq. (7.39); expressing $1 - \varepsilon^2$ in it via (7.40), we get

$$\boxed{\begin{cases} \tau = \sqrt{\dfrac{r_0^3}{8m}}\,(\eta + \sin\eta) \\ r = \dfrac{r_0}{2}\,(1 + \cos\eta) \end{cases}} \qquad 0 \leq \eta \leq \pi, \tag{7.41}$$

where the range of η is such that $r = r_0$ for $\tau = 0$ ($\eta = 0$) and r decays to 0 when $\eta \to \pi$. The function $r(\tau)$ resulting from (7.41) is depicted in Fig. 7.3.

The solution for $t = t(\tau)$ is obtained by combining $\mathrm{d}t/\mathrm{d}\tau$ as expressed by (7.34) and $\mathrm{d}\tau/\mathrm{d}\eta$ deduced from (7.41):

$$\frac{\mathrm{d}\tau}{\mathrm{d}\eta} = \sqrt{\frac{r_0^3}{8m}}\,(1 + \cos\eta) = \sqrt{\frac{r_0}{2m}}\,r.$$

We get

$$\frac{\mathrm{d}t}{\mathrm{d}\eta} = \frac{\mathrm{d}t}{\mathrm{d}\tau}\frac{\mathrm{d}\tau}{\mathrm{d}\eta} = \varepsilon\sqrt{\frac{r_0}{2m}}\,r\left(1 - \frac{2m}{r}\right)^{-1} = \sqrt{\frac{r_0}{2m} - 1}\;\frac{r^2}{r - 2m},$$

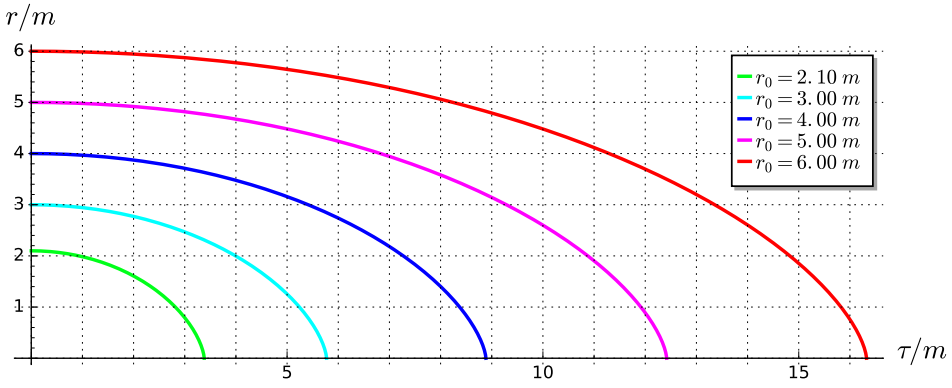


Figure 7.3: Coordinate r as a function of the proper time τ for the radial free fall from rest, for various initial values r_0 of r.

where we have used (7.40) and $\varepsilon > 0$ [Eq. (7.27)] to write $\varepsilon = \sqrt{1 - 2m/r_0}$. Substituting r from Eq. (7.41), we get

$$\frac{\mathrm{d}t}{\mathrm{d}\eta} = \frac{r_0}{2}\sqrt{\frac{r_0}{2m} - 1}\, \frac{(1+\cos\eta)^2}{1 + \cos\eta - 4m/r_0}.$$

This equation can be integrated to (cf. the SageMath computation in Sec. D.3.5)

$$\boxed{t = 2m\left\{\sqrt{\frac{r_0}{2m} - 1}\left[\eta + \frac{r_0}{4m}(\eta + \sin\eta)\right] + \ln\left|\frac{\sqrt{\frac{r_0}{2m} - 1} + \tan\frac{\eta}{2}}{\sqrt{\frac{r_0}{2m} - 1} - \tan\frac{\eta}{2}}\right|\right\}}, \tag{7.42}$$

where we have assumed $t = 0$ at $\tau = 0$ ($\eta = 0$).

The solution of the radial free fall starting from rest at $r = r_0$ is thus given in parametric form by Eqs. (7.41) and (7.42) and is represented in the left panel of Fig. 7.4. It has been obtained in the Schwarzschild-Droste coordinates (t, r, θ, φ), which are singular at the event horizon $\mathscr{H}$. So, one might wonder if such a solution can describe the full infall, with the crossing of $\mathscr{H}$. In particular, we notice that the differential equation for $t(\tau)$, Eq. (7.34), is singular at $r(\tau) = 2m$, i.e. on $\mathscr{H}$. The solution $t(\eta)$, as given by Eq. (7.42), is singular at $\eta = \eta_\mathrm{h}$, where

$$\eta_\mathrm{h} := 2\,\mathrm{atan}\sqrt{\frac{r_0}{2m} - 1} \tag{7.43}$$

is precisely the value of η yielding $r = 2m$ in Eq. (7.41) [to see it, rewrite the second part of Eq. (7.41) as $r = r_0\cos^2(\eta/2) = r_0/(1 + \tan^2(\eta/2))$]. This singularity of $t(\eta)$ appears also clearly on Fig. 7.4 (left panel). On the other hand, the equation for r, Eq. (7.36), does not exhibit any pathology at $r = 2m$, nor its solution (7.42). Actually, had we started from the ingoing Eddington-Finkelstein (IEF) coordinates $(\tilde{t}, r, \theta, \varphi)$, instead of the Schwarzschild-Droste ones, we would have found[6] exactly the same solution for $r(\tau)$ (which is not

[6] *Exercice:* do it!

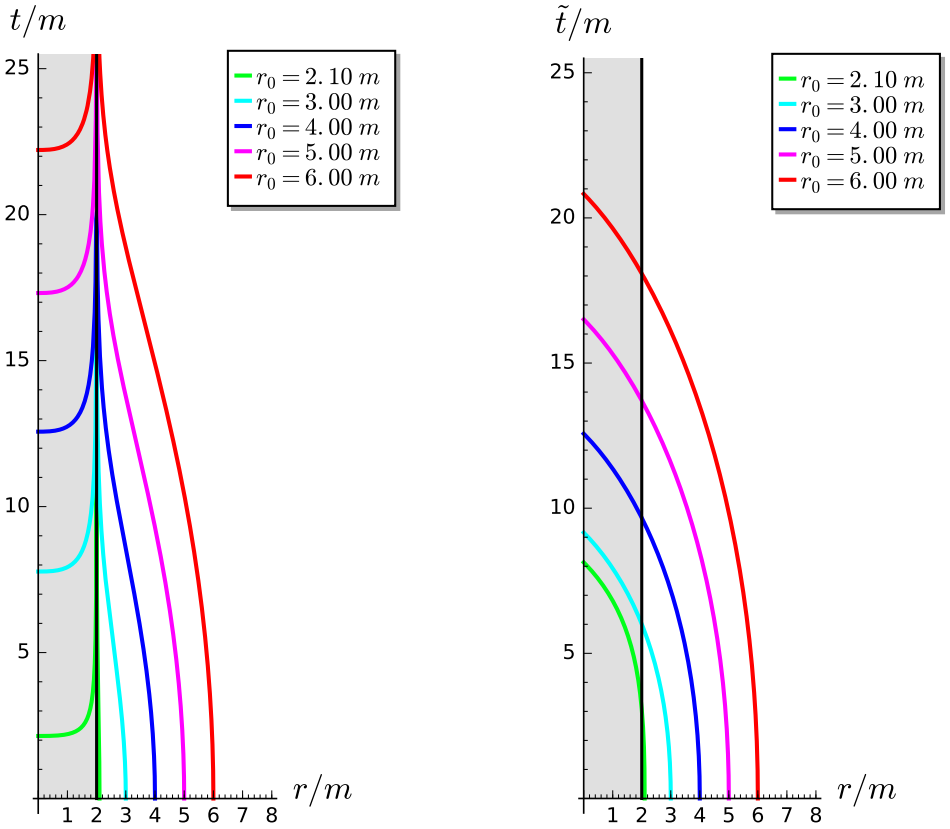


Figure 7.4: Radial free fall from rest, viewed in Schwarzschild-Droste coordinates (t, r) (left) and in the ingoing Eddington-Finkelstein coordinates $(\tilde{t}, r)$ (right), for various values r_0 of the coordinate r at $\tau = 0$. The grey area is the black hole region $\mathscr{M}_{\rm II}$.

surprising since r, considered as a scalar field on $\mathscr{M}$, is perfectly regular at $\mathscr{H}$). The solution for $\tilde{t}(\tau)$ can be deduced from that for $t(\tau)$ by the coordinate transformation law (6.33). Noticing that and $r/(2m) = \cos^2(\eta/2)/\cos^2(\eta_{\rm h}/2)$, we get

$$\begin{aligned} \frac{r}{2m} - 1 &= \frac{\cos^2(\eta/2)}{\cos^2(\eta_{\rm h}/2)} - 1 = \cos^2(\eta/2)\left(\frac{1}{\cos^2(\eta_{\rm h}/2)} - \frac{1}{\cos^2(\eta/2)}\right) \\ &= \cos^2(\eta/2)\left(\tan^2(\eta_{\rm h}/2) - \tan^2(\eta/2)\right). \end{aligned} \tag{7.44}$$

Using this identity as well as (7.43) to express $\sqrt{r_0/(2m) - 1}$ in Eq. (7.42), the transformation law (6.33) yields

$$\begin{aligned} \tilde{t} &= 2m\left\{\sqrt{\frac{r_0}{2m} - 1}\left[\eta + \frac{r_0}{4m}(\eta + \sin\eta)\right] + \ln\left|\frac{\tan\frac{\eta_{\rm h}}{2} + \tan\frac{\eta}{2}}{\tan\frac{\eta_{\rm h}}{2} - \tan\frac{\eta}{2}}\cos^2\frac{\eta}{2}\left(\tan^2\frac{\eta_{\rm h}}{2} - \tan^2\frac{\eta}{2}\right)\right|\right\} \\ &= 2m\left\{\sqrt{\frac{r_0}{2m} - 1}\left[\eta + \frac{r_0}{4m}(\eta + \sin\eta)\right] + \ln\left|\cos^2\frac{\eta}{2}\left(\tan\frac{\eta_{\rm h}}{2} + \tan\frac{\eta}{2}\right)^2\right|\right\} \\ &= 2m\left\{\sqrt{\frac{r_0}{2m} - 1}\left[\eta + \frac{r_0}{4m}(\eta + \sin\eta)\right] + 2\ln\left(\cos\frac{\eta}{2}\tan\frac{\eta_{\rm h}}{2} + \sin\frac{\eta}{2}\right)\right\}. \end{aligned}$$

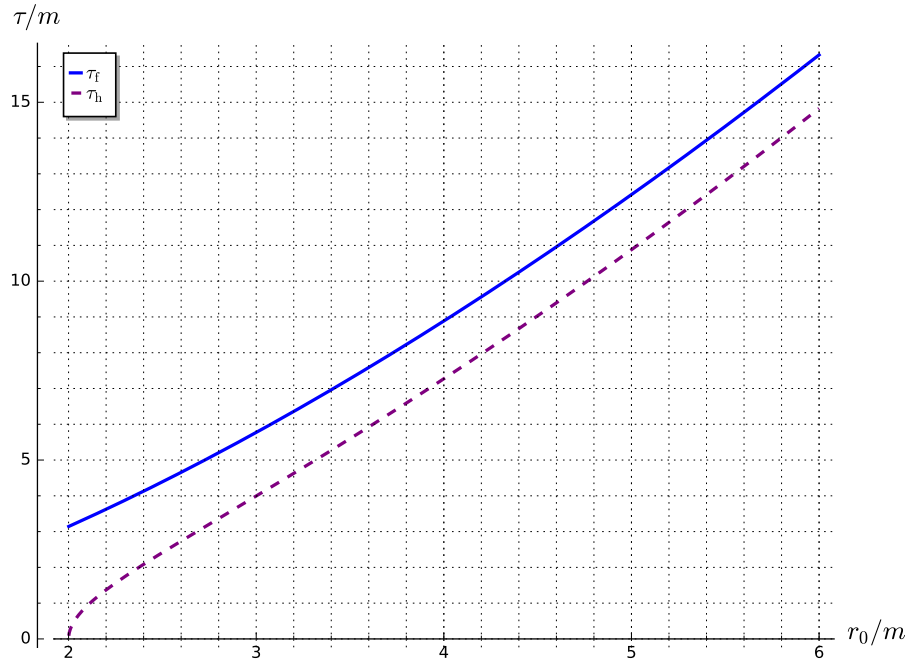


Figure 7.5: Elapsed proper time to reach the event horizon ($\tau_{\rm h}$, dashed curve) and the central singularity ($\tau_{\rm f}$, solid curve), as a function of the initial value of r for a radial free fall from rest.

From this expression, we have $\tilde{t} = 4m \ln\tan(\eta_{\rm h}/2)$ for $\eta = 0$. Now, we can change the origin of the IEF coordinate $\tilde{t}$ to ensure $\tilde{t} = 0$ for $\eta = 0$, i.e. $\tau = 0$. We get then

$$\boxed{\tilde{t} = 2m\left\{\sqrt{\frac{r_0}{2m}-1}\left[\eta + \frac{r_0}{4m}(\eta+\sin\eta)\right] + 2\ln\left[\cos\frac{\eta}{2} + \left(\frac{r_0}{2m}-1\right)^{-1/2}\sin\frac{\eta}{2}\right]\right\}}. \quad (7.45)$$

This expression is perfectly regular for all values of η in $[0,\pi]$, reflecting the fact that the ingoing Eddington-Finkelstein coordinates cover all $\mathscr{M}$ in a regular way. The radial free fall solution in terms of $(\tilde{t}, r)$ is represented in the right panel of Fig. 7.4. We note the smooth crossing of the event horizon $\mathscr{H}$.

In view of Eq. (7.41), we may say that the radial infall starts at $\eta = 0$, for which $\tau = 0$ and $r = r_0$, and terminates at $\eta = \pi$, for which $r = 0$, which means that the particle hits the curvature singularity (cf. Sec. 6.3.4). The final value of the particle's proper time is obtained by setting $\eta = \pi$ in Eq. (7.41):

$$\boxed{\tau_{\rm f} = \frac{\pi}{2}\sqrt{\frac{r_0^3}{2m}}}. \quad (7.46)$$

Similarly, the final value of $\tilde{t}$ is obtained by setting $\eta = \pi$ in (7.45):

$$\boxed{\tilde{t}_{\rm f} = 2m\left[\pi\sqrt{\frac{r_0}{2m}-1}\left(\frac{r_0}{4m}+1\right) - \ln\left(\frac{r_0}{2m}-1\right)\right]}. \quad (7.47)$$

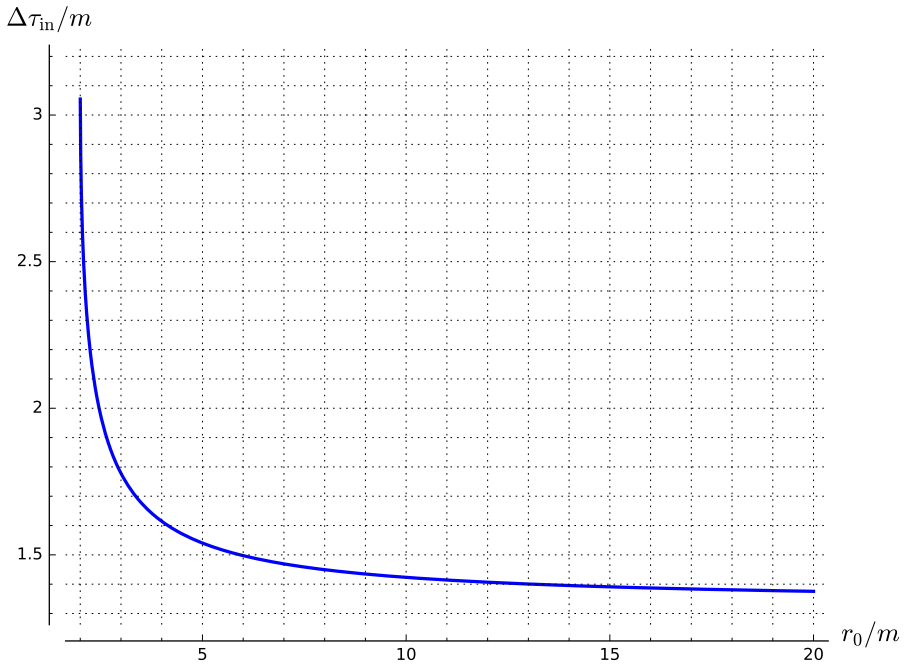


Figure 7.6: Proper time spent inside the black hole region as a function of the initial value of r for a radial free fall from rest. Note that $r_0 = 2m$ does not correspond to any asymptote but to the finite value $\Delta\tau_{\rm in} = \pi\, m$ with a vertical tangent. On the other side, there is an horizontal asymptote $\Delta\tau_{\rm in} \to 4m/3$ for $r_0 \to +\infty$.

As noticed above, the event horizon $\mathscr{H}$ is crossed at $\eta = \eta_{\rm h}$; via (7.41) and (7.43), this corresponds to the following value of the proper time:

$$\boxed{\tau_{\rm h} = \sqrt{\frac{r_0^3}{2m}}\left[\operatorname{atan}\sqrt{\frac{r_0}{2m}-1} + \sqrt{\frac{2m}{r_0}\left(1-\frac{2m}{r_0}\right)}\right]}, \tag{7.48}$$

while (7.45) leads to the following value of the IEF coordinate $\tilde{t}$:

$$\boxed{\tilde{t}_{\rm h} = 2m\left[2\left(1+\frac{r_0}{4m}\right)\sqrt{\frac{r_0}{2m}-1}\operatorname{atan}\sqrt{\frac{r_0}{2m}-1} + \frac{r_0}{2m} - 1 - \ln\frac{r_0}{2m}\right]}. \tag{7.49}$$

The variation of $\tau_{\rm h}$ and $\tau_{\rm f}$ with r_0 are depicted in Fig. 7.5 and numerical values for $r_0 = 6m$ and standard astrophysical black holes are provided in Table 7.1.

The proper time spent inside the black hole is

$$\boxed{\Delta\tau_{\rm in} = \tau_{\rm f} - \tau_{\rm h} = \sqrt{\frac{r_0^3}{2m}}\left[\frac{\pi}{2} - \operatorname{atan}\sqrt{\frac{r_0}{2m}-1} - \sqrt{\frac{2m}{r_0}\left(1-\frac{2m}{r_0}\right)}\right]}. \tag{7.50}$$

It varies between πm ($r_0 \to 2m$) and $4m/3$ ($r_0 \to +\infty$) (cf. Fig. 7.6 and Sec. D.3.5 for the computation of $\lim_{r_0\to+\infty}\Delta\tau_{\rm in}$). Numerical values for astrophysical black holes are provided in Table 7.1.

m	$r_S = 2m$	τ_h	τ_f	$\Delta\tau_{in}$
15 $M_\odot$ (Cyg X-1)	44.3 km	1.10 ms	1.21 ms	0.11 ms
4.3 10^6 $M_\odot$ (Sgr A*)	12.7 10^6 km = 0.085 au	5 min 14 s	5 min 46 s	32 s
6 10^9 $M_\odot$ (M87*)	118 au	5.07 days	5.58 days	12 h 17 min

Table 7.1: Proper time to reach the event horizon (τ_h) and the central curvature singularity (τ_f), as well as elapsed proper time inside the black hole region ($\Delta\tau_{in}$), when freely falling from rest at $r_0 = 6m$. The numerical values are given for various black hole masses m, corresponding to astrophysical objects: the stellar black hole Cygnus X-1 [201, 121], the supermassive black hole at the center of our galaxy (Sagittarius A*) [115, 159] and the supermassive black hole M87* in the nucleus of the galaxy Messier 87 [114, 5].

7.3.3 Circular orbits

Circular orbits are defined as timelike geodesics with $r = \text{const}$. We have then $dr/d\tau = 0$ and $d^2r/d\tau^2 = 0$, so that Eq. (7.24) implies

$$V_\ell(r) = \frac{\varepsilon^2 - 1}{2} \tag{7.51a}$$

$$\frac{dV_\ell}{dr} = 0. \tag{7.51b}$$

Given the expression (7.25) of V_ℓ, Eq. (7.51b) is equivalent to

$$mr^2 - \ell^2 r + 3\ell^2 m = 0. \tag{7.52}$$

As already noticed in Sec. 7.3.1, this quadratic equation in r admits two real roots iff $|\ell| \geq \ell_{crit}$, with $\ell_{crit} = 2\sqrt{3}\,m$ [Eq. (7.32)], which are

$$\boxed{r^\pm_{circ}(\ell) = \frac{\ell}{2m}\left(\ell \pm \sqrt{\ell^2 - \ell^2_{crit}}\right)}. \tag{7.53}$$

$r^+_{circ}(\ell)$ corresponds to a minimum of the effective potential V_ℓ and thus to a stable orbit (see the dots in Fig. 7.2), while $r^-_{circ}(\ell)$ corresponds to a maximum of V_ℓ and thus to an unstable orbit. When ℓ varies from ℓ_{crit} to $+\infty$, $r^+_{circ}(\ell)$ increases from $6m$ to $+\infty$, while $r^-_{circ}(\ell)$ decreases from $6m$ to $3m$ (cf. Fig. 7.7). We conclude that

> Circular orbits in Schwarzschild spacetime exist for all values of $r > 3m$. Those with $r < 6m$ are unstable and those with $r > 6m$ are stable. The marginal case $r = 6m$ is called the ***innermost stable circular orbit***, often abridged as ***ISCO***.

Remark 2: In the Newtonian spherical gravitational field generated by a point mass m, there is no unstable orbit, and thus no ISCO. The existence of unstable orbits in the relativistic case can be understood by the extra term in the effective potential $V_\ell(r)$ (cf. Remark 1 on p. 149),

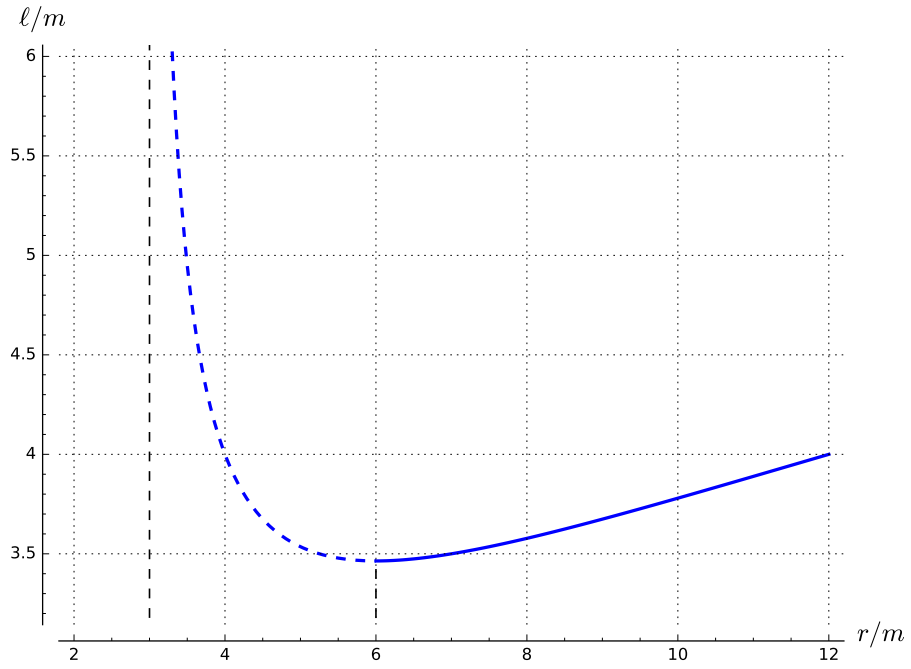


Figure 7.7: Specific conserved angular momentum $\ell = L/\mu$ on circular orbits as a function of the orbit circumferential radius r. The dashed part of the curve corresponds to unstable orbits ($r = r^-_{\rm circ}(\ell)$, as given by Eq. (7.53)), while the solid part corresponds to stable orbits ($r = r^+_{\rm circ}(\ell)$). The minimal value of ℓ is $\ell_{\rm crit} = 2\sqrt{3}\,m \simeq 3.46\,m$.

which adds the attractive part $-\ell^2 m/r^3$ to the two terms constituting the Newtonian potential: $-m/r$ (attractive) and $\ell^2/(2r^2)$ (repulsive). The latter is responsible for the infinite "centrifugal barrier" at small r in the Newtonian problem, leading always to a minimum of $V_\ell(r)$ and thus to a stable circular orbit. In the relativistic case, for r small enough, the attractive term, which is $O(r^{-3})$, dominates over the centrifugal one, which is only $O(r^{-2})$. Equivalently, we may say that the "centrifugal barrier" is weakened by the factor $1 - 2m/r$ (cf. the expression (7.25) of $V_\ell(r)$) and ceases to exist for small values of $|\ell|$ (i.e. $|\ell| < \ell_{\rm crit}$).

From Eq. (7.52), we can easily express ℓ as a function of r on a circular orbit:

$$\boxed{|\ell| = r\sqrt{\frac{m}{r-3m}}}. \tag{7.54}$$

This function is represented in Fig. 7.7 (for $\ell > 0$).

If we substitute (7.54) for ℓ in the expression (7.25) of V_ℓ and use Eq. (7.51a), we obtain the value of the specific conserved energy along a circular orbit, in terms of r:

$$\boxed{\varepsilon = \frac{r-2m}{\sqrt{r(r-3m)}}}. \tag{7.55}$$

This function is represented in Fig. 7.8. The minimal value of ε is achieved for $r = 6m$, i.e. at the ISCO:

$$\boxed{\min \varepsilon = \frac{2\sqrt{2}}{3} \simeq 0.9428}. \tag{7.56}$$

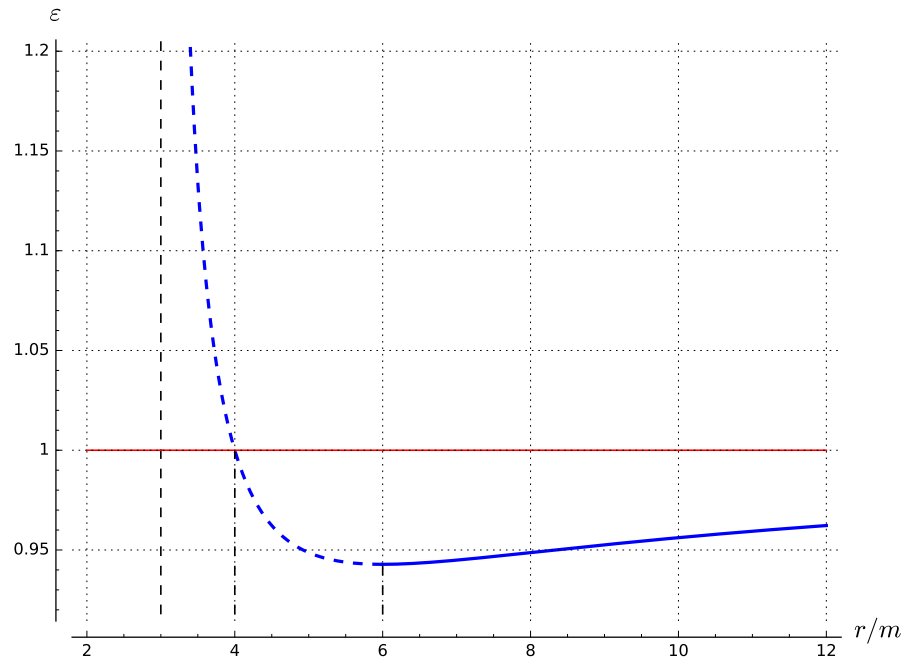


Figure 7.8: Specific conserved energy $\varepsilon = E/\mu$ on circular orbits as a function of the orbit circumferential radius r. The dashed part of the curve corresponds to unstable orbits, while the solid part corresponds to stable ones. The horizontal red line $\varepsilon = 1$ marks the limit of bound orbits.

From Fig. 7.8, we notice that

$$r > 4m \iff \varepsilon < 1 \iff E < \mu. \tag{7.57}$$

This corresponds to ***bound orbits***, i.e. to geodesics that, if slightly perturbed, cannot reach the asymptotically flat region $r \gg 2m$, since $E \geq \mu$ there. Indeed, when $r \to +\infty$, the Killing vector $\boldsymbol{\xi}$ can be interpreted as the 4-velocity of some asymptotically inertial observer (at rest with respect to the black hole) and E is the particle energy measured by that observer; the famous Einstein relation (1.30) is then $E = \Gamma\mu$, where Γ is the Lorentz factor of the particle with respect to the observer. Since $\Gamma \geq 1$ [Eq. (1.33)], we have obviously[7] $E \geq \mu$. For this reason, the circular orbit at $r = 4m$ is called the ***marginally bound circular orbit***. Note that the marginally bound circular orbit is unstable, since it has $r < 6m$.

The track of circular orbits in the (ℓ, ε) plane is depicted in Fig. 7.9. The ISCO, which is a minimum for both ε and ℓ, appears as a cusp point.

The ***angular velocity*** of a circular orbit $\mathscr{L}$ is defined by

$$\Omega := \left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{\mathscr{L}} = \frac{u^\varphi}{u^t}, \tag{7.58}$$

where $u^\varphi = \mathrm{d}\varphi/\mathrm{d}\tau$ and $u^t = \mathrm{d}t/\mathrm{d}\tau$ are the only nonzero components w.r.t. Schwarzschild-Droste coordinates of the 4-velocity $\boldsymbol{u}$ along the worldline $\mathscr{L}$. It follows from (7.58) that

[7] Similarly, the radial-motion solutions (7.37)-(7.38), which allow for $r \to +\infty$, have $E \geq \mu$, while the solution (7.39), which is relevant for a free fall from rest, has $E < \mu$.

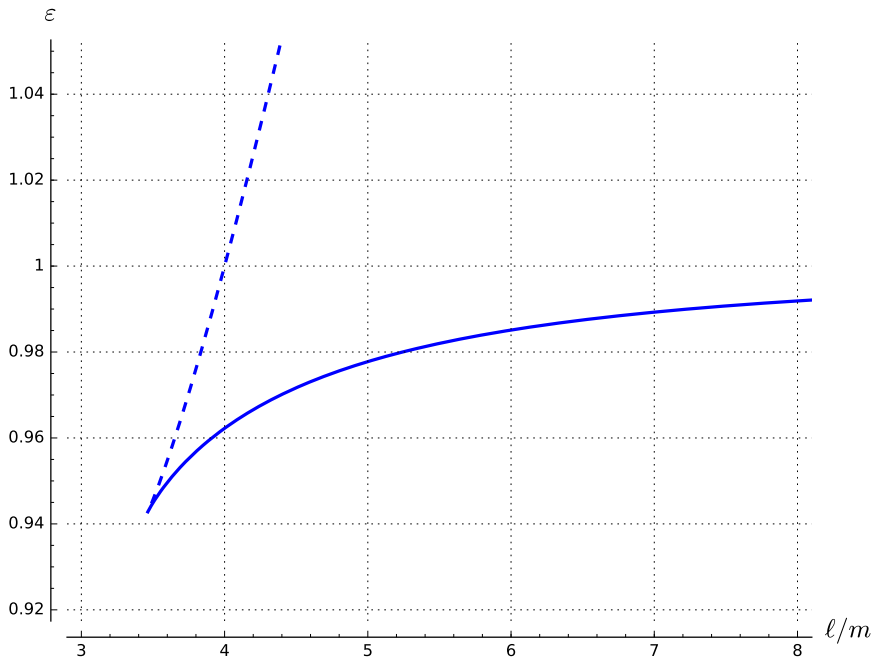


Figure 7.9: Circular orbits in the (ℓ, ε) plane. The solid (resp. dashed) curve corresponds to stable (resp. unstable) orbits. The ISCO is located at the cusp point.

Ω enters into the linear combination of the two Killing vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ expressing the 4-velocity on a circular orbit according to

$$\boldsymbol{u} = u^t \left(\boldsymbol{\xi} + \Omega \boldsymbol{\eta}\right). \tag{7.59}$$

We have the following nice physical interpretation:

The quantity Ω defined by Eq. (7.58) is nothing but the angular velocity of the orbiting particle $\mathscr{P}$ monitored by a infinitely distant static observer $\mathscr{O}$.

Proof. Suppose that $\mathscr{O}$ is located at fixed coordinates $(r, \theta, \varphi) = (r_{\mathscr{O}}, \pi/2, 0)$ with $r_{\mathscr{O}} \gg m$ and that $\mathscr{P}$ emits a photon at the event $(t_1, r, \pi/2, 0)$ along a *radial* null geodesic. This photon is received by $\mathscr{O}$ at $t = t'_1$. After one orbit, at the event $(t_2, r, \pi/2, 2\pi)$, $\mathscr{P}$ emits a second photon in the radial direction, which is received at $t = t'_2$ by $\mathscr{O}$. According to the definition (7.58) of Ω, we have

$$2\pi = \Omega(t_2 - t_1).$$

On the other hand, since $r_{\mathscr{O}} \gg m$, the proper time of $\mathscr{O}$ is t, so that the angular velocity measured by $\mathscr{O}$ is

$$\Omega_{\mathscr{O}} = \frac{2\pi}{t'_2 - t'_1}.$$

Now, since t is the coordinate associated to the spacetime invariance by time translation (stationarity), we have necessarily $t'_2 - t_2 = t'_1 - t_1$ (the increment in coordinate t for the

m	$r_{\rm ISCO} = 6m$	$\dfrac{\Omega_{\rm ISCO}}{2\pi}$	$T_{\rm ISCO}$	$T_{\mathscr{P},\rm ISCO}$
$15\,M_\odot$ (Cyg X-1)	133 km	147 Hz	6.80 ms	4.81 ms
$4.3\,10^6\,M_\odot$ (Sgr A*)	$38.1\,10^6$ km (0.255 au)	$5.11\,10^{-4}$ Hz	32 min 37 s	23 min 4 s
$6\,10^9\,M_\odot$ (M87*)	355 au	$3.66\,10^{-7}$ Hz	31 d 15 h	22 d 9 h

Table 7.2: Values of various quantities at the ISCO for masses m of some astrophysical black holes (see Table 7.1 for details): areal radius r, orbital frequency $\Omega_{\rm ISCO}/(2\pi)$, orbital period seen from infinity $T_{\rm ISCO}$ and orbital period measured by the orbiting observer/particle $T_{\mathscr{P},\rm ISCO}$.

second signal is the same as for the first one), so that $t'_2 - t'_1 = t_2 - t_1$. Accordingly, the above two equations combine to $\Omega_{\mathscr{O}} = \Omega$. □

By combining Eqs. (7.34) and (7.35), we get

$$\Omega = \frac{1}{r^2}\left(1 - \frac{2m}{r}\right)\frac{\ell}{\varepsilon}.$$

Substituting expression (7.54) for ℓ and expression (7.55) for ε, we obtain

$$\boxed{\Omega = \pm\sqrt{\frac{m}{r^3}}}, \tag{7.60}$$

with the + (resp. −) sign for $\ell > 0$ (resp. $\ell < 0$).

Remark 3: This formula is identical to that of Newtonian gravity (Kepler's third law for circular orbits) for all values of r. This is a mere coincidence, valid only for Schwarzschild-Droste coordinates. Only for $r \gg m$, i.e. in the weak-field limit, this agreement is physically meaningful; it can be then used to interpret the parameter m as the *gravitational mass* of Schwarzschild spacetime, as mentioned in Sec. 6.2.4.

Remark 4: Ω is not the orbital angular frequency experienced by the particle/observer $\mathscr{P}$ on the circular orbit $\mathscr{L}$, because the proper time of $\mathscr{P}$ is τ and not t. The actual orbital frequency measured by $\mathscr{P}$ is

$$\Omega_{\mathscr{P}} = \frac{\mathrm{d}t}{\mathrm{d}\tau}\Omega = u^t\Omega,$$

with $u^t = \mathrm{d}t/\mathrm{d}\tau$ obtained from (7.34) and (7.55): $u^t = \sqrt{r/(r-3m)}$. Hence

$$\Omega_{\mathscr{P}} = \sqrt{\frac{r}{r-3m}}\,\Omega = \pm\frac{1}{r}\sqrt{\frac{m}{r-3m}}. \tag{7.61}$$

Note that $|\Omega_{\mathscr{P}}| > |\Omega|$; in particular, at the ISCO ($r = 6m$), $\Omega_{\mathscr{P}} = \sqrt{2}\Omega$. The orbital period measured by $\mathscr{P}$ is $T_{\mathscr{P}} = 2\pi/|\Omega_{\mathscr{P}}|$. Some ISCO values of $T_{\mathscr{P}}$ for astrophysical black holes are provided in Table 7.2.

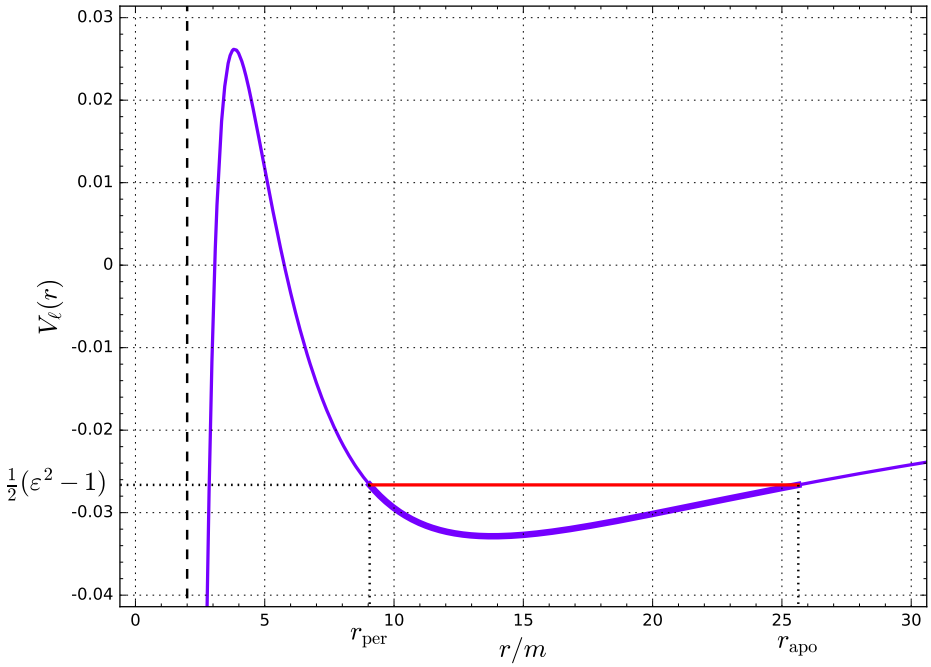


Figure 7.10: Effective potential $V_\ell(r)$ for $\ell = 4.2m$ (one of the values displayed in Figs. 7.1 and 7.2). The horizontal red line marks $V_\ell(r) = (\varepsilon^2-1)/2$ with $\varepsilon = 0.973$, leading to $r_{\rm per} = 9.058\,m$ and $r_{\rm apo} = 25.634\,m$. The corresponding orbit is shown in Fig. 7.11.

At the ISCO, $r = 6m$ and formula (7.60) yields (for orbits with positive ℓ)

$$\boxed{\Omega_{\rm ISCO} = \frac{1}{6\sqrt{6}\,m}}. \tag{7.62}$$

Numerical values of $\Omega_{\rm ISCO}$ (actually the frequency $\Omega_{\rm ISCO}/(2\pi)$, which is more relevant from an observational point of view) are provided in Table 7.2.

7.3.4 Other orbits

Let us relax the assumption $r = \text{const}$ and consider generic orbits $\mathscr{L}$ obeying

$$|\ell| > \ell_{\rm crit} \qquad \text{and} \qquad 0 < \varepsilon < 1. \tag{7.63}$$

The first condition ensures that the effective potential $V_\ell(r)$ takes the shape of a well in the region $r > 2m$ (cf. Fig. 7.10) and the second one that the particle $\mathscr{P}$ is trapped in this well. Indeed, $0 < \varepsilon < 1$ makes the right-hand of Eq. (7.24) negative, so that the region $r \to +\infty$, where $V_\ell(r) \to 0$, cannot be reached. We have also argued in Sec. 7.3.3 that $\varepsilon < 1$ is forbidden in the region $r \to +\infty$ on physical grounds [cf. the discussion below Eq. (7.57)].

In the potential well, the r-coordinate along $\mathscr{L}$ varies between two extrema: a minimum $r_{\rm per}$, for ***periastron*** (or ***pericenter*** or ***periapsis***), and a maximum $r_{\rm apo}$, for ***apoastron*** (or ***apocenter*** or ***apoapsis***) (cf. Figs. 7.10 and 7.11). Being extrema of $r(\tau)$, the

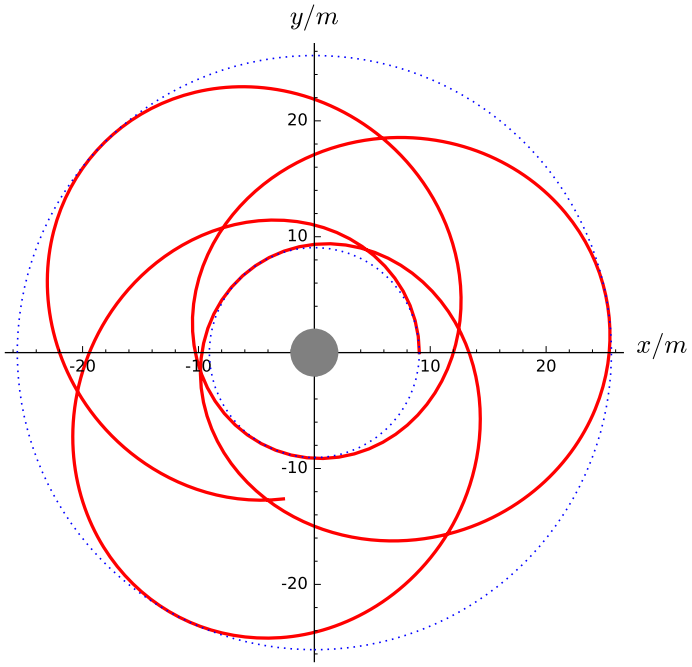


Figure 7.11: Timelike geodesic with $\varepsilon = 0.973$ and $\ell = 4.2m$ (same values as in Fig. 7.10), plotted in terms of the coordinates $(x, y) := (r\cos\varphi, r\sin\varphi)$. The dotted circles correspond to $r = r_{\rm per}$ (periastron) and $r = r_{\rm apo}$ (apoastron). The grey disk indicates the black hole region $r < 2m$. *[Figure produced with the notebook D.3.6]*

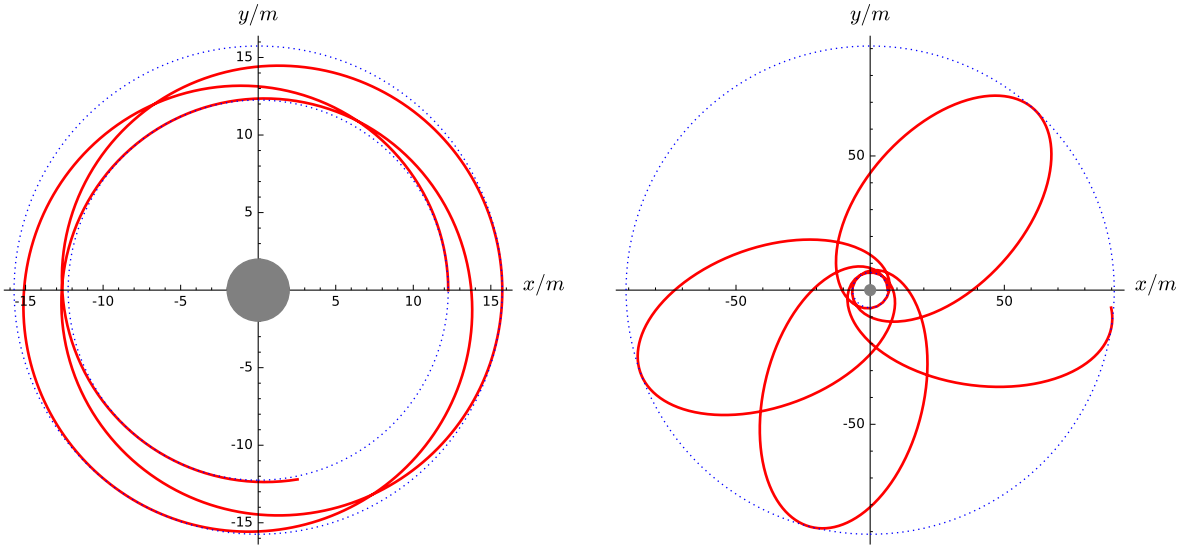


Figure 7.12: Timelike geodesics with the same value of ℓ as in Fig. 7.11 ($\ell = 4.2m$), but for different values of ε: $\varepsilon = 0.967$ (left) and $\varepsilon = 0.990$ (right). Note that the left and right figures have different scales. *[Figure produced with the notebook D.3.6]*

values of $r_{\rm per}$ and $r_{\rm apo}$ are obtained by setting $\mathrm{d}r/\mathrm{d}\tau = 0$ in Eq. (7.24), which leads to

$$\left(1 - \frac{2m}{r}\right)\left(1 + \frac{\ell^2}{r^2}\right) = \varepsilon^2. \tag{7.64}$$

This is a cubic equation in r^{-1}, which has three real positive roots, corresponding to the three intersections of the curve $V_\ell(r)$ with the horizontal line at $(\varepsilon^2 - 1)/2$ in Fig. 7.10. However, the smaller root has to be disregarded as a periastron since it would lead to a motion with $V_\ell(r) > (\varepsilon^2 - 1)/2$, which is forbidden by Eq. (7.24).

We get, from Eqs. (7.24)-(7.25),

$$\frac{\mathrm{d}r}{\mathrm{d}\tau} = \pm\sqrt{\varepsilon^2 - \left(1 - \frac{2m}{r}\right)\left(1 + \frac{\ell^2}{r^2}\right)}. \tag{7.65}$$

The equation governing the trajectory of $\mathscr{L}$ in the orbital is Eq. (7.17), which we can recast as

$$\frac{\mathrm{d}u}{\mathrm{d}\varphi} = \pm\sqrt{2u^3 - u^2 + 2\left(\frac{m}{\ell}\right)^2 u + \left(\frac{m\varepsilon}{\ell}\right)^2 - \left(\frac{m}{\ell}\right)^2}. \tag{7.66}$$

Let us recall that $u := m/r$ and that methods for solving this differential equation have been briefly discussed in Sec. 7.2.3.

Remark 5: Far from the black hole, i.e. in the region $r \gg m$, one can easily recover the Newtonian orbits from Eq. (7.66). Indeed, according to Eq. (7.31), $\varepsilon = 1 + \varepsilon_0$, where $\varepsilon_0 = v^2/2 - m/r$ is the Newtonian mechanical energy per unit mass. It obeys $|\varepsilon_0| \ll 1$, so that $\varepsilon^2 \simeq 1 + 2\varepsilon_0$. Moreover, for $r \gg m$, $u \ll 1$ and we can neglect the u^3 term in front of the u^2 one in Eq. (7.66). Hence Eq. (7.66) reduces to

$$\frac{\mathrm{d}u}{\mathrm{d}\varphi} \simeq \pm\sqrt{2\left(\frac{m}{\ell}\right)^2 \varepsilon_0 - u^2 + 2\left(\frac{m}{\ell}\right)^2 u}. \tag{7.67}$$

Let us introduce the constants

$$p := \frac{\ell^2}{m} \qquad \text{and} \qquad e := \sqrt{1 + 2\frac{\varepsilon_0 \ell^2}{m^2}}. \tag{7.68}$$

Then Eq. (7.67) can be rewritten as

$$\frac{\mathrm{d}\varphi}{\mathrm{d}u} = \pm\frac{\frac{p}{me}}{\sqrt{1 - \left(\frac{\frac{p}{m}u - 1}{e}\right)^2}},$$

which is readily integrated into

$$\varphi = \pm\arccos\left(\frac{\frac{p}{m}u - 1}{e}\right) + \varphi_0,$$

where φ_0 is a constant. We have then $\frac{p}{m}u = 1 + e\cos(\varphi - \varphi_0)$, or equivalently,

$$r = \frac{p}{1 + e\cos(\varphi - \varphi_0)} \tag{7.69}$$

Assuming a bound orbit, we have $\varepsilon < 1$, which implies $\varepsilon_0 < 0$ and, via Eq. (7.68), $e < 1$. We recognize then in (7.69) the equation of an ellipse of eccentricity e and semi-latus rectum p. Hence Keplerian orbits are recovered for $r \gg m$, as they should.

Generic bound orbits differ from the Keplerian ellipses by the fact that the variation of φ between two successive ***periastron passages***, i.e. two events along the worldline of $\mathscr{P}$ for which $r = r_{\rm per}$, is strictly larger than 2π. This phenomenon is called ***periastron advance*** and causes the orbits to be not closed, as illustrated in Figs. 7.11 and 7.12.

Historical note: The equations of geodesic motion (7.11)-(7.13), as well as Eq. (7.17) for the trajectories in the orbital plane, have been first given by Karl Schwarzschild himself in January 1916 in the very same article [229] in which he presented his famous solution[8]. Schwarzschild discussed only the weak field limit of these equations, to recover the Mercury's perihelion advance computed by Einstein.

It is quite remarkable that the general solution to the geodesic motion in Schwarzschild spacetime[9] has been given as early as May 1916 by Johannes Droste, in the same article [96] in which he derived the Schwarzschild solution, independently of Karl Schwarzschild (cf. historical note on p. 123). Droste derived the equations of geodesic motion in Schwarzschild-Droste coordinates (t, r, θ, φ), using t as the parameter along the geodesics, as well as the equation governing the trajectories in the orbital plane[10]. He gave the solutions for the trajectories in terms of the Weierstrass elliptic function $\wp$ (cf. Sec. 7.2.3). He classified the solutions in terms of the roots of the cubic polynomial in v that appears in the right-hand side of Eq. (7.19). Droste noticed that it takes an infinite amount of coordinate time t for a particle to reach $r = 2m$ (cf. left panel of Fig. 7.4) and he concluded incorrectly that "a moving particle outside the sphere $r = 2m$ can never pass that sphere". He missed that this is only a coordinate effect, reflecting the pathology of Schwarzschild-Droste coordinates at the horizon. One shall keep in mind that in 1916, general relativity was just in its infancy and disentangling coordinate artefacts from physical effects was not so obvious, especially regarding time. One can be amused by the fact that the first analysis of the static black hole of general relativity made the black hole appear, not as an object from which no particle may escape, but as an object into which no particle may penetrate...

Finally, it is worth mentioning two early detailed studies of geodesic motion of massive particles in Schwarzschild spacetime: one by by Carlo De Jans in 1923 [83] and other one by Yusuke Hagihara in 1931 [137]. For a detailed account about the history of geodesic motion in Schwarzschild spacetime see Ref. [100].

[8] Equation (7.17) for the trajectories is Eq. (18) in the article [229], the link between Schwarzschild's notations and ours being $R = r$, $x = u/m$, $c = L$, $1 = E$, $h = \mu^2$.

[9] More precisely: in the $\mathscr{M}_{\rm I}$ region of Schwarzschild spacetime.

[10] The link between Droste's notations and ours is $\alpha = 2m$, $x = 2u$, $z = 2u - 1/3$, $A = \mu^2/E^2$ and $B = L/E$.

Chapter 8

Null geodesics and images in Schwarzschild spacetime

Contents

8.1 Introduction

Having investigated the properties of generic causal geodesics in Schwarzschild spacetime in Chap. 7, we focus here on null geodesics. The main interest of null geodesics is of course that they are the carriers of information from the surroundings of the black hole to some observer, in particular in the form of images. We start by studying the general property of null geodesics in Sec. 8.2, distinguishing the radial geodesics, from the non-radial ones and focussing on the r-motion. Then in Sec. 8.3, we study of the φ-motion and show that the trajectory of a given null geodesic in the equatorial plane is fully integrable, via elliptic integrals. In Sec. 8.4 we compute the asymptotic direction taken by a null geodesic emitted at a given point; this prepares the discussion of images in Sec. 8.5. The study of black hole images has received a tremendous boost after the release of the first observed image by the Event Horizon Telescope collaboration in 2019, that of the black hole M87* [5]. However, we differ the discussion of this image to the chapter regarding null geodesics around a rotating black hole (Chap. 12), since the black hole spin plays some role in the images and M87* is expected to be a fast rotator.

8.2 Main properties of null geodesics

We use the same notations as in Sec. 7.2 of the preceding chapter: $(\mathscr{M}, \boldsymbol{g})$ is Schwarzschild spacetime, (t, r, θ, φ) are Schwarzschild-Droste coordinates and $\mathscr{P}$ is a particle, the world-line of which is a geodesic $\mathscr{L}$. The particle $\mathscr{P}$ is characterized by its 4-momentum $\boldsymbol{p}$ and E and L stand for the conserved energy and conserved angular momentum along $\mathscr{L}$ [cf. Eq. (7.1)].

In all this chapter, we consider that $\mathscr{L}$ is a *null* geodesic, or equivalently that $\mathscr{P}$ is a *massless* particle, typically a photon. As in Chap. 7, we assume that the coordinates (θ, φ) are chosen so that $\mathscr{L}$ lies in the hyperplane $\theta = \pi/2$. Let us recall that a geodesic of Schwarzschild spacetime lies necessarily in some hypersurface, which can be chosen to be the hyperplane $\theta = \pi/2$ without any loss of generality (cf. Sec. 7.2.1).

8.2.1 Equations to be solved

In terms of Schwarzschild-Droste coordinates, the geodesic motion of $\mathscr{P}$ is governed by Eqs. (7.11), (7.12) and (7.13) with $\mu = 0$ (massless particle):

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\lambda} = E\left(1 - \frac{2m}{r}\right)^{-1}}, \tag{8.1}$$

$$\boxed{\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{L}{r^2}}, \tag{8.2}$$

$$\boxed{\left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^2 + \frac{L^2}{r^2}\left(1 - \frac{2m}{r}\right) = E^2}. \tag{8.3}$$

Let us recall that the variable λ with respect to which these differential equations hold is the affine parameter associated with the 4-momentum $\boldsymbol{p}$ of $\mathscr{P}$: $\boldsymbol{p} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ [Eq. (7.9)] and that λ increases towards the future, $\boldsymbol{p}$ being always future-oriented (cf. Sec. 7.2.2).

8.2.2 Radial null geodesics

Let us first discuss the case of radial geodesics, which are characterized by $L = 0$. Equation (8.2) yields immediately $\varphi = \text{const}$, while Eq. (8.3) simplifies drastically:

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda} = \pm E, \tag{8.4}$$

the solution of which is immediate:

$$r = \pm E\lambda + r_0, \tag{8.5}$$

where r_0 is some constant. Moreover, writing $\mathrm{d}t/\mathrm{d}\lambda = \mathrm{d}t/\mathrm{d}r \times \mathrm{d}r/\mathrm{d}\lambda$ and combining Eqs. (8.1) and (8.4), we get

$$\frac{\mathrm{d}t}{\mathrm{d}r} = \pm\left(1 - \frac{2m}{r}\right)^{-1}. \tag{8.6}$$

We recognize the equation governing the radial null geodesics obtained in Sec. 6.3.1 [Eq. (6.20)], the solution of which is given by Eq. (6.21):

$$\boxed{t = \pm r \pm 2m \ln \left| \frac{r}{2m} - 1 \right| + \text{const}}. \tag{8.7}$$

Remark 1: Since λ is an affine parameter and E is constant, Eq. (8.5) shows explicitly that r is another affine parameter along radial null geodesics of Schwarzschild spacetime — a feature that we had already obtained in Sec. 6.3.1.

Some radial null geodesics of Schwarzschild spacetime are plotted in Fig. 6.1 in terms of the Schwarzschild-Droste coordinates and in Fig. 6.3 in terms of the ingoing Eddington-Finkelstein coordinates.

8.2.3 Generic null geodesics: effective potential

Having discussed the case $L = 0$ in the preceding section, we focus now on the case $L \neq 0$ ("generic" case). We may then consider along the geodesic $\mathscr{L}$ the affine parameter

$$\boxed{\tilde{\lambda} := |L| \lambda}, \tag{8.8}$$

instead of λ, the latter being the affine parameter associated with the 4-momentum $\boldsymbol{p}$ via Eq. (7.9). Since L is a non-vanishing constant along a geodesic, the above formula does define a new affine parameter. The absolute value ensures that $\tilde{\lambda}$ is increasing towards the future, as λ, whatever the sign of L. Note that $\tilde{\lambda}$ has the same dimension as L, i.e. a squared length, given that λ is dimensionless (cf. Sec. 7.2.2). In terms of $\tilde{\lambda}$, the differential system (8.1)-(8.3) becomes

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\tilde{\lambda}} = b^{-1} \left(1 - \frac{2m}{r} \right)^{-1}}, \tag{8.9}$$

$$\boxed{\frac{\mathrm{d}\varphi}{\mathrm{d}\tilde{\lambda}} = \frac{\epsilon_L}{r^2}}, \tag{8.10}$$

$$\boxed{\left(\frac{\mathrm{d}r}{\mathrm{d}\tilde{\lambda}} \right)^2 + U(r) = \frac{1}{b^2}}, \tag{8.11}$$

where

$$\boxed{b := \frac{|L|}{E}}, \tag{8.12}$$

$$\epsilon_L := \frac{L}{|L|} = \mathrm{sgn}\, L \tag{8.13}$$

and

$$\boxed{U(r) := \frac{1}{r^2} \left(1 - \frac{2m}{r} \right)}. \tag{8.14}$$

As for the timelike case (cf. Eq. (7.24)), we note that Eq. (8.11) has the shape of the first integral of the 1-dimensional motion of a non-relativist particle in the ***effective potential*** $U(r)$. The main difference is that $U(r)$ does not depend on L, contrary to the effective potential $V_\ell(r)$ for timelike geodesics. Actually, $U(r)$ is a function of r and the black hole mass m only. We thus arrive at

A null geodesic $\mathscr{L}$ is entirely characterized by the constants b and $\epsilon_L = \pm 1$. As soon as $\mathscr{L}$ has some part in region $\mathscr{M}_{\rm I}$, then $E > 0$ [cf. Eq. (7.5)], so that b, as defined by Eq. (8.12), is finite and positive. It has has the dimension of a length and can be interpreted as the ***impact parameter*** in the case of a particle arising from infinity.

Proof. Let us consider that the massless particle $\mathscr{P}$ arises from a point of coordinates $\theta = \pi/2$, $r = r_0 \gg m$ and $\varphi = \varphi_0$. Let us introduce Cartesian coordinates (x, y) in the plane $\theta = \pi/2$: $x := r\cos\varphi$ and $y := r\sin\varphi$. Without any loss of generality, we may consider $|\varphi_0| \ll 1$. For large values of r, the motion of $\mathscr{P}$ is then essentially along the x-axis, with φ remaining small: $y \simeq y_0$, where $y_0 = r_0 \sin\varphi_0 \simeq r_0\varphi_0$. The quantity $|y_0|$ is the impact parameter. Since $y = r\sin\varphi \simeq r\varphi$, we get from $y \simeq y_0$ the following relation

$$\varphi \simeq \frac{y_0}{r} \qquad (r \gg m).$$

Deriving with respect to t leads to

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} \simeq -\frac{y_0}{r^2}\frac{\mathrm{d}r}{\mathrm{d}t} \simeq \frac{y_0}{r^2} \qquad (r \gg m), \tag{8.15}$$

where we have used $\mathrm{d}r/\mathrm{d}t = -1$ for $r \gg m$. This last property is easy to admit since $r \gg m$ corresponds to the flat part of spacetime; however, it can be derived rigorously by combining Eqs. (8.9) and (8.11) to evaluate $\mathrm{d}r/\mathrm{d}t$. Now by combining Eqs. (8.9) and (8.10), we have

$$\frac{\mathrm{d}\varphi}{\mathrm{d}t} = \epsilon_L \frac{b}{r^2}\left(1 - \frac{2m}{r}\right). \tag{8.16}$$

For $r \gg m$, the right-hand side reduces to $\epsilon_L b/r^2$, so that the comparison with (8.15) yields $\epsilon_L b = y_0$, i.e. $b = |y_0|$, which proves that b is the impact parameter of the massless particle $\mathscr{P}$ with respect to the central black hole. □

The effective potential $U(r)$ is plotted in Fig. 8.1. It has no minimum and a maximum at $r = 3m$, which is

$$U_{\rm max} = \frac{1}{27m^2}. \tag{8.17}$$

This extremum is a stationary position in r: it corresponds thus to a circular orbit, usually called the ***circular photon orbit*** The set of all photon orbits (one per choice of equatorial plane) is often called the ***photon sphere***. However, since it corresponds to a *maximum* of the effective potential, the circular orbit at $r = 3m$ is an unstable orbit. So one should not imagine that a Schwarzschild black hole is surrounded by any spherical shell of photons...

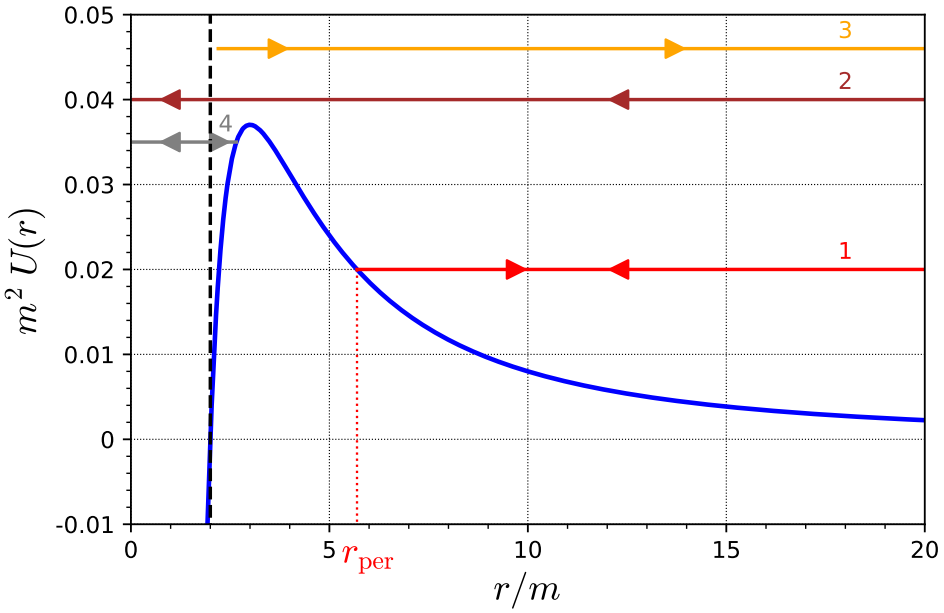


Figure 8.1: Effective potential $U(r)$ (rescaled by m^2 to make it dimensionless) governing the r-part of the motion along a null geodesic in Schwarzschild spacetime [Eq. (8.14)]. The vertical dashed line marks $r = 2m$, i.e. the location of the event horizon. The horizontal lines marked "1" to "4" correspond to the r-motion of four null geodesics (cf. Sec. 8.2.4 for details); their trajectories in the equatorial plane are depicted in Fig. 8.2. *[Figure produced with the notebook D.3.7]*

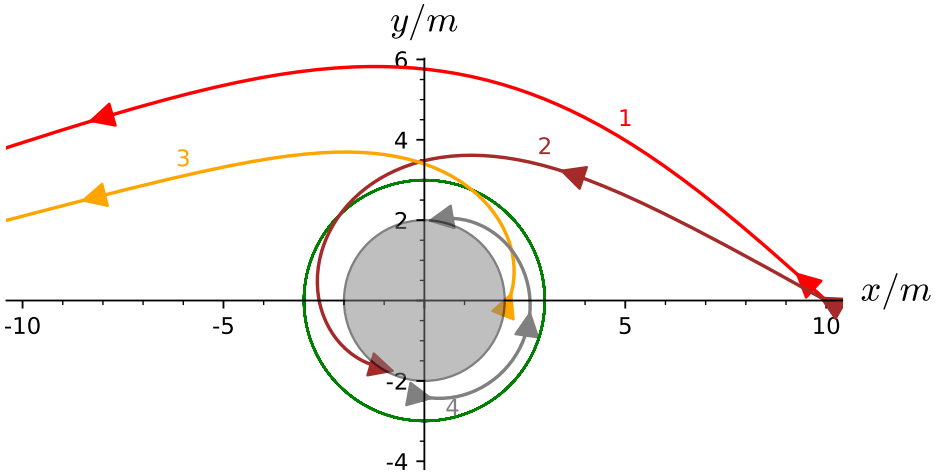


Figure 8.2: Selected null geodesics in the equatorial plane of Schwarzschild spacetime, plotted in terms of the coordinates $(x, y) := (r\cos\varphi, r\sin\varphi)$, with the black hole region $r < 2m$ depicted as a grey disk. The green geodesic is the photon circular orbit at $r = 3m$. The red geodesic (label "1") starts at $r_0 = 10m$, $\varphi_0 = 0$, with the impact parameter $b = 7.071\,m$ ($b^{-2} = 0.02\,m^{-2}$). The brown geodesic (label "2") starts at $r_0 = 10m$, $\varphi_0 = 0$ with $b = 5m$ ($b^{-2} = 0.04\,m^{-2}$). The orange geodesic (label "3") starts outward at $r_0 = 2.1m$, $\varphi_0 = 0$ with $b = 4.663m$ ($b^{-2} = 0.046\,m^{-2}$). The grey geodesic (label "4") starts outward at $r_0 = 2.4m$, $\varphi_0 = -\pi/2$ with $b = 5.345m$ ($b^{-2} = 0.035\,m^{-2}$). The r-motion of these four gedeosics is depicted in Fig. 8.1. *[Figure produced with the notebook D.3.8]*

8.2.4 Radial behaviour of null geodesics

For the sake of concreteness, in this section and the remaining ones, we refer to the massless particle $\mathscr{P}$ as a "photon". In most applications, in particular the astrophysical ones, $\mathscr{P}$ will indeed be a photon. But one shall keep in mind that all results are valid for any other massless particle.

Since the effective potential $U(r)$ has no minimum, it does not offer any potential well, as it is clear from Fig. 8.1. This is in sharp contrast with the effective potential $V_\ell(r)$ for massive particle (compare Fig. 7.10). Hence there are no bound orbits for photons[1].

We can infer various types of photon worldlines from Fig. 8.1. In view of the "first integral" (8.11), each photon worldline can be represented by a horizontal line of ordinate b^{-2} in this figure, which must lie above the curve $U(r)$ by the positive quantity $(\mathrm{d}r/\mathrm{d}\tilde{\lambda})^2$. The region under the curve $U(r)$ is thus excluded.

For an initially inward photon, i.e. a photon emitted with $\mathrm{d}r/\mathrm{d}\tilde{\lambda} < 0$ from a position $r = r_{\rm em}$, there are two possibilities, depending on the values of $r_{\rm em}$ and of the impact parameter b:

- if $r_{\rm em} > 3m$ and b is large enough to fulfil $b^{-2} < U_{\rm max}$ (e.g. trajectory no. 1 in Fig. 8.1), the photon "bounces" on the potential barrier constituted by $U(r)$ at some minimal value $r_{\rm p}$ of r — the ***periastron***, which is given by $U(r_{\rm p}) = b^{-2}$, or equivalently by
$$r_{\rm p}^3 - b^2\, r_{\rm p} + 2mb^2 = 0, \quad r_{\rm p} > 3m. \tag{8.18}$$
Equation (8.11) implies then
$$\left.\frac{\mathrm{d}r}{\mathrm{d}\tilde{\lambda}}\right|_{r=r_{\rm p}} = 0. \tag{8.19}$$
Actually, $\mathrm{d}r/\mathrm{d}\tilde{\lambda}$ changes sign at $r = r_{\rm p}$ and the photon subsequently moves away from the black hole for ever (cf. geodesic no. 1 in Fig. 8.2). We may call such a worldline a ***scattering trajectory***.

- if $r_{\rm em} < 3m$ (e.g. trajectory no. 4 in Fig. 8.1) or b is small enough to fulfil $b^{-2} > U_{\rm max}$ (e.g. trajectory no. 2 in Fig. 8.1) the photon is not halted by the potential barrier constituted by $U(r)$. It then reaches arbitrary small values of r and is eventually absorbed by the black hole ($r < 2m$) (cf. geodesic no. 2 in Fig. 8.2).

For an initially outward photon, i.e. a photon emitted with $\mathrm{d}r/\mathrm{d}\tilde{\lambda} > 0$, one has necessarily $r_{\rm em} > 2m$ according to the result obtained in Sec. 7.2.2 and there are then two possible outcomes:

- if $r_{\rm em} > 3m$ (e.g. trajectory no. 1 in Fig. 8.1) or b is small enough to fulfil $b^{-2} > U_{\rm max}$ (e.g. trajectory no. 3 in Figs. 8.1), the photon escapes to infinity (cf. geodesic no. 3 in Fig. 8.2);

[1] unless one counts as "bound" a worldline that terminates in the black hole region

- if $2m < r_{\rm em} < 3m$ and $b^{-2} < U_{\max}$ (e.g. trajectory no. 4 in Fig. 8.1), the photon "bounces" on the left side of the potential barrier, reaching a maximal value $r_{\rm a}$ of r — the ***apoastron***, which is given by $U(r_{\rm a}) = b^{-2}$, or equivalently by

$$r_{\rm a}^3 - b^2\, r_{\rm a} + 2mb^2 = 0, \quad r_{\rm a} < 3m. \tag{8.20}$$

  The photon moves subsequently towards the black hole and is absorbed by it (cf. geodesic no. 4 in Fig. 8.2).

The critical value of the impact parameter b separating the cases discussed above is determined by $b_{\rm c}^{-2} = U_{\max}$. Given the value (8.17) of $U_{\max}$, we get

$$\boxed{b_{\rm c} = 3\sqrt{3}\, m \simeq 5.196152\, m}. \tag{8.21}$$

It follows from the above discussion that

> Along any null geodesic of Schwarzschild spacetime, the areal coordinate r either is a monotonic function or has a single turning point. In the latter case, if the turning point corresponds to a minimum of r (a *periastron*, which can occur only for $r > 3m$), the null geodesic escapes to infinity, while if the turning point corresponds to a maximum of r (an *apoastron*, which can occur only for $2m < r < 3m$), the null geodesic terminates at the central singularity ($r = 0$).

Note that for $r < 2m$, i.e. in the black hole region, r is always a monotonic function, since it has been demonstrated in Sec. 7.2.2 that $r(\lambda)$ is strictly decreasing. The impossibility of a turning point for $r < 2m$ can also be graphically inferred from Fig. 8.1, which shows that the effective potential $U(r)$ is negative for $r < 2m$, preventing the turning point condition $U(r) = b^{-2}$ to hold.

For $b > b_{\rm c}$, the explicit expression of the periastron radius $r_{\rm p}$ (or the apoastron radius $r_{\rm a}$) is obtained by solving the cubic equation (8.18) (or (8.20), which is the same cubic equation, except for the range of the solution). Fortunately Eq. (8.18) is a depressed[2] cubic equation, which makes it simpler to solve. For $b > b_{\rm c}$, its discriminant $-(4p^3 + 27q^2)$ is positive, which implies that it admits three distinct real roots. Two of them are positive and are precisely $r_{\rm p}$ and $r_{\rm a}$. The third root is negative, since the product of the roots is $-2mb^2 < 0$; it has therefore no physical significance. The roots of the generic depressed cubic equation $x^3 + px + q = 0$ can be expressed via Viète's formulas:

$$x_k = 2\sqrt{-\frac{p}{3}}\cos\left[\frac{1}{3}\arccos\left(\frac{3q}{2p}\sqrt{-\frac{3}{p}}\right) + \frac{2k\pi}{3}\right], \qquad k \in \{0,1,2\}. \tag{8.22}$$

In the present case, $p = -b^2$, $q = 2mb^2$ and $r_{\rm p}$ (resp. $r_{\rm a}$) corresponds to $k = 0$ (resp. $k = 2$), so that we obtain

$$\boxed{r_{\rm p} = \frac{2b}{\sqrt{3}}\cos\left[\frac{\pi}{3} - \frac{1}{3}\arccos\left(\frac{b_{\rm c}}{b}\right)\right]}. \tag{8.23}$$

[2] A *depressed* cubic equation is a polynomial equation of the type $x^3 + px + q = 0$.

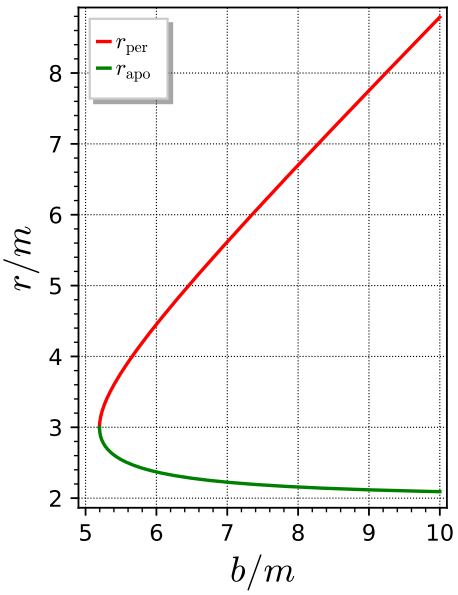


Figure 8.3: Radial coordinate of the periastron, $r_{\rm p}$, and of the apoastron, $r_{\rm a}$, along null geodesics with $b > b_{\rm c} \simeq 5.196\,m$. Note that a given null geodesic has either a periastron or an apoastron, but not both. *[Figure produced with the notebook D.3.9]*

$$\boxed{r_{\rm a} = \frac{2b}{\sqrt{3}} \cos\left[\frac{5\pi}{3} - \frac{1}{3}\arccos\left(\frac{b_{\rm c}}{b}\right)\right]}. \tag{8.24}$$

As a check, we note that for $b \gg b_{\rm c}$, Eq. (8.23) yields $r_{\rm p} \simeq 2b/\sqrt{3}\cos\left(\pi/3 - 1/3 \times \pi/2\right) \simeq 2b/\sqrt{3}\cos(\pi/6) \simeq b$, as expected. Indeed, $b \gg b_{\rm c}$ implies $b \gg m$, so that the photon stays far from the black hole, a regime in which the impact parameter coincides with the distance of closest approach: $b \simeq r_{\rm p}$.

The variation of $r_{\rm p}$ and $r_{\rm a}$ with b is plotted in Fig. 8.3. Note that $r_{\rm p} = r_{\rm a} = 3m$ (the photon orbit) at the limit $b = b_{\rm c}$.

Remark 2: It is clear on Fig. 8.3 that, for the same value of b, $r_{\rm a} \leq r_{\rm p}$, which may seem contradictory with the apoastron corresponding to the largest distance from the center and the periastron to the distance of closest approach. However, one shall keep in mind that the apoastron and the periastron always refer to different null geodesics: geodesics with an apoastron lie below the photon sphere ($r = 3m$), while those with a periastron are always outside it. A null geodesic can of course cross the photon sphere (examples are geodesics 2 and 3 on Figs. 8.1 and 8.2), but then it has neither an apoastron nor a periastron.

8.3 Trajectories of null geodesics in the equatorial plane

In all this section, we assume $L \neq 0$, i.e. we consider non-radial null geodesics, the radial case having been discussed in Sec. 8.2.2.

Let us first recall the generic property of causal geodesics in Schwarzschild spacetime established in Sec. 7.2.2 : for $L \neq 0$, the azimuthal angle coordinate φ has no turning

point: it is either always increasing towards the future ($L > 0$) or always decreasing towards the future ($L < 0$).

Remark 1: We recover the above property from Eq. (8.10): $\mathrm{d}\varphi/\mathrm{d}\tilde{\lambda} = \epsilon_L/r^2$, given that $\epsilon_L = \mathrm{sgn}\, L$ and $\tilde{\lambda}$ increases towards the future.

8.3.1 Differential equation and fundamental cubic polynomial

The equation governing the (r, φ)-part of a null geodesic, named *trajectory in the orbital plane* in Sec. 7.2.3, is the generic differential equation (7.17) with μ (the particle's mass) set to zero and the ratio E^2/L^2 replaced by $1/b^2$ [cf. Eq. (8.12)]:

$$\boxed{\left(\frac{\mathrm{d}u}{\mathrm{d}\varphi}\right)^2 = P_b(u)} \tag{8.25}$$

where $u := m/r$ [Eq. (7.16)] and $P_b(u)$ stands for the cubic polynomial

$$\boxed{P_b(u) := 2u^3 - u^2 + \frac{m^2}{b^2}}. \tag{8.26}$$

Remark 2: Of course, Eq. (8.25) can be recovered by combining Eqs. (8.10), (8.11) and (8.14).

The graph of the polynomial P_b is shown in Fig. 8.4 for selected values of b. Note that along a null geodesic, one must have $u > 0$ and $P_b(u) \geq 0$. The first condition follows from the very definition of u as m/r, while the second one is a direct consequence of Eq. (8.25).

By Eq. (8.25), the zeros of the polynomial P_b correspond to points at which $\mathrm{d}u/\mathrm{d}\varphi = 0$. They are thus stationary points for u, and hence for r. This can only occur at the circular photon orbit $r = \mathrm{const} = 3m$ (then $b = b_\mathrm{c}$) or at the periastron or apoastron discussed in Sec. 8.2.4 (then $b > b_\mathrm{c}$). The zeros of P_b are governed by its discriminant, which is

$$\Delta = 4\frac{m^4}{b^4}\left(\frac{b^2}{m^2} - 27\right). \tag{8.27}$$

One may distinguish three cases:

- $\Delta > 0 \iff b > \sqrt{27}m = b_\mathrm{c}$: P_b has three distinct real zeros (cf. the green and cyan curves in Fig. 8.4); one of them, u_n say, is negative, and hence unphysical (since $u := m/r > 0$), while the two other ones are positive, being nothing but

$$u_\mathrm{p} = \frac{m}{r_\mathrm{p}} \qquad \text{and} \qquad u_\mathrm{a} = \frac{m}{r_\mathrm{a}}, \tag{8.28}$$

  where r_p and r_a are the periastron and apoastron radii given by Eqs. (8.23)-(8.24).

- $\Delta = 0 \iff b = b_\mathrm{c}$: P_b has a double zero: $u_\mathrm{a} = u_\mathrm{p} = 1/3$ and a negative zero: $u_\mathrm{n} = -1/6$ (cf. red curve in Fig. 8.4); only the first one is physical and correspond to $r = 3m$.

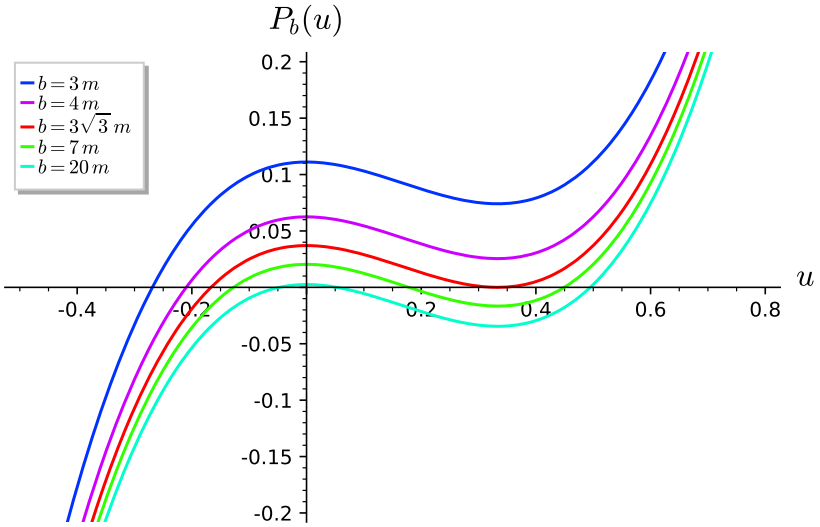


Figure 8.4: Graph of the cubic polynomial $P_b(u) = 2u^3 - u^2 + m^2/b^2$ [Eq. (8.26)]. *[Figure produced with the notebook D.3.9]*

- $\Delta < 0 \iff b < b_\mathrm{c}$: P_b has one real negative zero, u_n, and two complex zeros (cf. blue and magenta curves in Fig. 8.4), so that $P_b(u)$ never vanishes for physical values of u, which are real positive.

Since they will be required in what follows, let us find explicit expressions for the real zeros of the polynomial P_b. The equation $P_b(u) = 0$ can be brought to a depressed form (no square term) by the change of variable $u = x + 1/6$. We get

$$x^3 - \frac{1}{12}x + \frac{m^2}{2b^2} - \frac{1}{108} = 0. \tag{8.29}$$

We shall discuss separately the cases $b > b_\mathrm{c}$ and $b < b_\mathrm{c}$.

Zeros of P_b for $b \geq b_\mathrm{c}$

For $b > b_\mathrm{c}$, we can express the solutions of Eq. (8.29) via Viète's formulas (8.22) with $p = -1/12$ and $q = m^2/(2b^2) - 1/108$. We then get the zeros of P_b as

$$u_k = \frac{1}{3}\cos\left[\frac{1}{3}\arccos\left(1 - 2\frac{b_\mathrm{c}^2}{b^2}\right) + \frac{2k\pi}{3}\right] + \frac{1}{6}, \qquad k \in \{0, 1, 2\}.$$

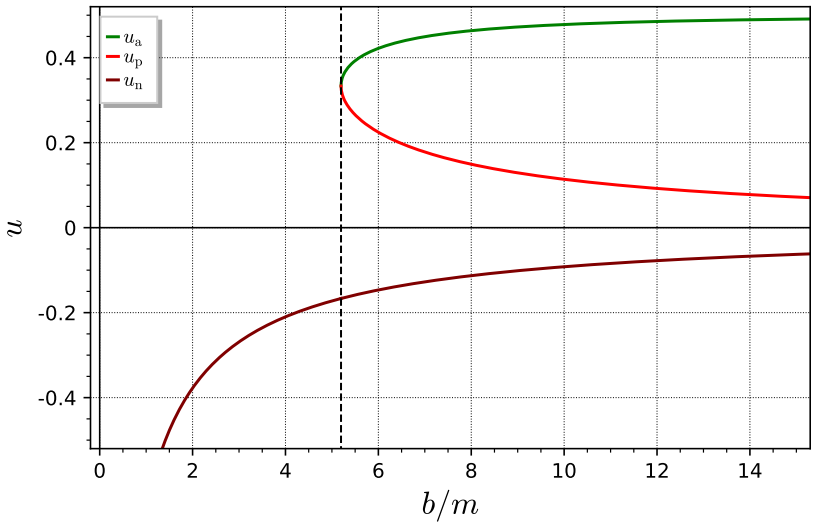


Figure 8.5: Real zeros u_n, u_p and u_a of the cubic polynomial $P_b(u) = 2u^3 - u^2 + m^2/b^2$ as functions of b. The vertical dashed line marks the critical value $b = b_c \simeq 5.196\,m$. *[Figure produced with the notebook D.3.9]*

Using the identity $\arccos(1-2x^2) = 2\arcsin x$ and noticing that $u_1 < 0$, which implies $u_1 = u_n$, and $u_0 \geq u_2$, which implies $u_0 = u_a$ and $u_2 = u_p$, we arrive at

$$\boxed{u_n = \frac{1}{3}\cos\left[\frac{2}{3}\arcsin\left(\frac{b_c}{b}\right) + \frac{2\pi}{3}\right] + \frac{1}{6}} \qquad (b \geq b_c) \tag{8.30a}$$

$$\boxed{u_p = \frac{1}{3}\cos\left[\frac{2}{3}\arcsin\left(\frac{b_c}{b}\right) + \frac{4\pi}{3}\right] + \frac{1}{6}} \qquad (b \geq b_c) \tag{8.30b}$$

$$\boxed{u_a = \frac{1}{3}\cos\left[\frac{2}{3}\arcsin\left(\frac{b_c}{b}\right)\right] + \frac{1}{6}} \qquad (b \geq b_c). \tag{8.30c}$$

These zeros are plotted in terms of b on Fig. 8.5. Note the ordering

$$-\frac{1}{6} \leq u_n < 0 < u_p \leq \frac{1}{3} \leq u_a < \frac{1}{2}, \tag{8.31}$$

with the inequalities $\leq$ being saturated for $b = b_c$. Note also that

$$\lim_{b\to+\infty} u_n = 0, \qquad \lim_{b\to+\infty} u_p = 0, \qquad \text{and} \qquad \lim_{b\to+\infty} u_a = \frac{1}{2}. \tag{8.32}$$

It is easy to express two of the zeros in terms of the third one. For instance, let us pick u_p. From $P_b(u_p) = 0$, we have immediately $b^2/m^2 = -2u_p^3 + u_p^2$. If we substitute this expression for b^2/m^2 into $P_b(u)$, we get

$$\begin{aligned} P_b(u) = 0 \quad &\iff \quad 2u^3 - u^2 - 2u_p^3 + u_p^2 = 0 \\ &\iff \quad (u-u_p)\left[2(u^2 + u_p u + u_p^2) - (u + u_p)\right] = 0 \\ &\iff \quad (u-u_p)\left[2u^2 + (2u_p - 1)u + u_p(2u_p - 1)\right] = 0. \end{aligned}$$

The two zeros different from u_p of this equation, namely u_n and u_a, must then obey

$$2u^2 + (2u_\mathrm{p} - 1)u + u_\mathrm{p}(2u_\mathrm{p} - 1) = 0.$$

Solving this quadratic equation leads to the sought expressions:

$$u_\mathrm{n} = \frac{1}{4}\left(1 - 2u_\mathrm{p} - \sqrt{(1-2u_\mathrm{p})(1+6u_\mathrm{p})}\right) \tag{8.33a}$$

$$u_\mathrm{a} = \frac{1}{4}\left(1 - 2u_\mathrm{p} + \sqrt{(1-2u_\mathrm{p})(1+6u_\mathrm{p})}\right). \tag{8.33b}$$

Real zero of P_b for $b < b_\mathrm{c}$

As mentioned above, for $b < b_\mathrm{c}$, P_b has only one real zero, u_n, which is negative. Its value is obtained by means of *Viète's substitution*, which consists in setting $x = w + 1/(36w)$ in Eq. (8.29), thereby turning it into a quadratic equation for w^3. Solving the latter yields

$$\boxed{u_\mathrm{n} = \frac{1}{6}\left[1 - \left(\frac{b_\mathrm{c}}{b} - \sqrt{\frac{b_\mathrm{c}^2}{b^2} - 1}\right)^{2/3} - \left(\frac{b_\mathrm{c}}{b} - \sqrt{\frac{b_\mathrm{c}^2}{b^2} - 1}\right)^{-2/3}\right]} \qquad (b < b_\mathrm{c}). \tag{8.34}$$

u_n is plotted as a function of b in the left part of Fig. 8.5. Note that

$$u_\mathrm{n} < -\frac{1}{6}, \qquad \lim_{b\to b_\mathrm{c}^-} u_\mathrm{n} = -\frac{1}{6} \qquad \text{and} \qquad \lim_{b\to 0} u_\mathrm{n} = -\infty. \tag{8.35}$$

The last limit results from the expansion of Eq. (8.34), which yields $u_\mathrm{n} \sim -(2b_\mathrm{c}/b)^{2/3}/6$ when $b \to 0$. Observe also from Fig. 8.5 the smooth transition with the value of u_n for $b \geq b_\mathrm{c}$, which is given by Eq. (8.30a).

Historical note: As mentioned in the historical note on p. 165, the equations of geodesic motion in Schwarzschild metric have been first derived by Karl Schwarzschild in January 1916 [229] and Johannes Droste in May 1916 [96], but these two authors focussed on the *timelike* case (orbit of a massive particle or a "planet"). It seems that the first explicit writing of the equations governing *null* geodesics is due to Ludwig Flamm in September 1916 [107]. In particular, Flamm derived[3] Eqs. (8.1), (8.2) and a combination of Eqs. (8.1) and (8.3). He also derived Eq. (8.25) governing the trajectories in the equatorial plane, as well as Eq. (8.18) giving the periastron. Regarding the solutions of these equations, he got only approximate ones, to the second order in m/b and recover Einstein's famous result for the deflection of light by the Sun. For details about early studies of null geodesics in Schwarzschild spacetime, see Ref. [100].

8.3.2 Critical null geodesics

For $b = b_\mathrm{c} = 3\sqrt{3}\,m$, the differential equation (8.25) can be integrated by means of elementary functions. Indeed, one has then $P_b(u) = (u - 1/3)^2(2u + 1/3)$ and Eq. (8.25)

[3] The link between Flamm's notations and ours is $\alpha = 2m$, $R = r$, $x = u/m$ and $\Delta = b$.

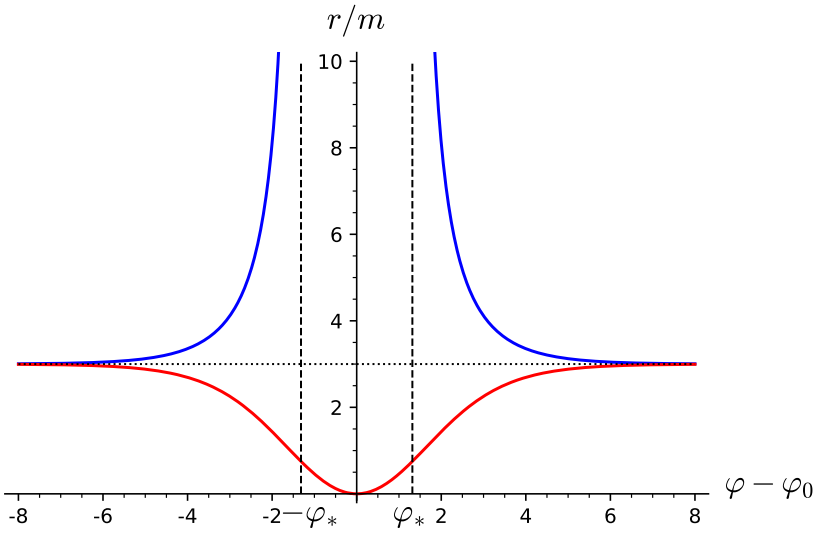


Figure 8.6: r as a function of φ along a null geodesic with an impact parameter equal to the critical one: $b = b_c = 3\sqrt{3}\,m$; the blue curve is for a critical null geodesic with $r > 3m$ [Eq. (8.40)], while the red one regards $r < 3m$ [Eq. (8.42)]. *[Figure produced with the notebook D.3.10]*

is equivalent to

$$\frac{\mathrm{d}\varphi}{\mathrm{d}u} = \pm \frac{1}{\left|u - \frac{1}{3}\right| \sqrt{2u + \frac{1}{3}}}. \tag{8.36}$$

This equation can be easily integrated by noticing that

$$\frac{1}{(1-x)\sqrt{x}} = \begin{cases} 2\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\operatorname{artanh}\sqrt{x}\right) & \text{for } x \in (0,1) \\ 2\dfrac{\mathrm{d}}{\mathrm{d}x}\left(\operatorname{arcoth}\sqrt{x}\right) & \text{for } x \in (1,+\infty). \end{cases} \tag{8.37}$$

We thus perform the change of variable $x = 2u + 1/3$ and treat separately two cases : $u < 1/3 \iff x \in (1/3, 1)$ and $u > 1/3 \iff x \in (1, +\infty)$. We call geodesics in the first case ***external critical null geodesics***, since $u < 1/3 \iff r > 3m$, and those in the second case ***internal critical null geodesics***, since $u > 1/3 \iff r < 3m$. Note that the qualifiers *external* and *internal* refer to the photon sphere $r = 3m$ discussed in Sec. 8.2.3, and *not* to the black hole region ($r < 2m$).

External critical null geodesics

For $u < 1/3$, $x \in (1/3, 1)$, so that the first line of Eq.(8.37) is relevant and Eq. (8.36) is integrated to

$$\varphi = \pm 2 \operatorname{artanh} \sqrt{2u + \frac{1}{3}} + \varphi_0, \tag{8.38}$$

where φ_0 is some integration constant. This relation is easily inverted to

$$u = \frac{1}{2} \tanh^2 \left(\frac{\varphi - \varphi_0}{2} \right) - \frac{1}{6}. \tag{8.39}$$

Moving back to $r = m/u$, we obtain the equation of the external critical null geodesic in polar form:

$$\boxed{r = \frac{2m}{\tanh^2\left(\frac{\varphi-\varphi_0}{2}\right) - \frac{1}{3}}}. \tag{8.40}$$

The constant φ_0 can be related to the asymptotic value φ_∞ of φ when $r \to +\infty$ by setting $u = 0$ in Eq. (8.39); we get, using the identity $\operatorname{artanh} x = 1/2 \ln[(1+x)/(1-x)]$,

$$\varphi_0 = \varphi_\infty \pm \varphi_*, \quad \text{with} \quad \varphi_* := \ln\left(\frac{\sqrt{3}+1}{\sqrt{3}-1}\right) \simeq 1.316958. \tag{8.41}$$

The function $r(\varphi)$, as given by Eq. (8.40), is depicted on Fig. 8.6 (blue curve). The region $\varphi_0 - \varphi_* < \varphi < \varphi_0 + \varphi_*$ is excluded, since Eq. (8.40) would yield $r < 0$. For $\varphi > \varphi_0 + \varphi_*$, $r(\varphi)$ is a decaying function, which corresponds to the plus sign in Eq. (8.38) and to the minus sign in Eq. (8.41): $\varphi_0 = \varphi_\infty - \varphi_*$. Figure 8.7 shows such a null geodesic with $\varphi_\infty = 0$, which implies $\varphi_0 = -\varphi_*$ and $\varphi > 0$. When $\varphi \to +\infty$, the geodesic rolls up indefinitely onto the photon orbit discussed in Sec. 8.2.3; this behaviour corresponds to the horizontal asymptote at $r = 3m$ in the right part of Fig. 8.6. Note that the geodesic approaches very fast the photon orbit, only a single path round to it being graphically visible in Fig. 8.7. This is because the asymptotic expansion of relation (8.40) is $r \sim 3m(1 + 6\mathrm{e}^{-\varphi})$ when $\varphi \to +\infty$.

It is worth stressing that Fig. 8.7 describes both (i) the trace in the (x, y)-plane of a geodesic with $L > 0$ (so that φ increases towards the future, cf. Eq. (8.2)) arising from $r \to +\infty$ and spiralling inwards to the photon orbit and (ii) the trace of a geodesic with $L < 0$ (so that φ decays towards the future) arising from $r = r_{\mathrm{em}} > 3m$, $\varphi = \varphi_{\mathrm{em}} > 0$, spiralling outwards and escaping to $r \to +\infty$ as $\varphi \to 0$. Had we restored the t dimension perpendicular to the (x, y)-plane in a 3d plot, these two geodesics would have clearly appeared distinct.

On the contrary, for $\varphi < \varphi_0 - \varphi_*$, $r(\varphi)$ is an increasing function, as it is clear on the left part of the blue curve in Fig. 8.6. This corresponds to the minus sign in Eq. (8.38) and to the plus sign in Eq. (8.41): $\varphi_0 = \varphi_\infty + \varphi_*$. A null geodesic of this type is depicted in Fig. 8.8. It has $\varphi_\infty = 0$, so that $\varphi_0 = \varphi_*$ and $\varphi < 0$. As for Fig. 8.7, the curve depicted in Fig. 8.8 can be interpreted as the trace in the (x, y)-plane of two distinct geodesics: one with $L < 0$ arising from $r \to +\infty$ and spiralling inwards to the photon orbit when $\varphi \to -\infty$ and another one with $L > 0$ arising from $r = r_{\mathrm{em}} > 3m$, $\varphi = \varphi_{\mathrm{em}} < 0$, spiralling outwards and escaping to $r \to +\infty$ as $\varphi \to 0$.

Internal critical null geodesics

Let us now consider the "internal" case: $r < 3m \iff u > 1/3$. This implies $x \in (1, +\infty)$ in Eq. (8.37), so that Eq. (8.36) is integrated to

$$\boxed{r = \frac{2m}{\coth^2\left(\frac{\varphi-\varphi_0}{2}\right) - \frac{1}{3}}}. \tag{8.42}$$

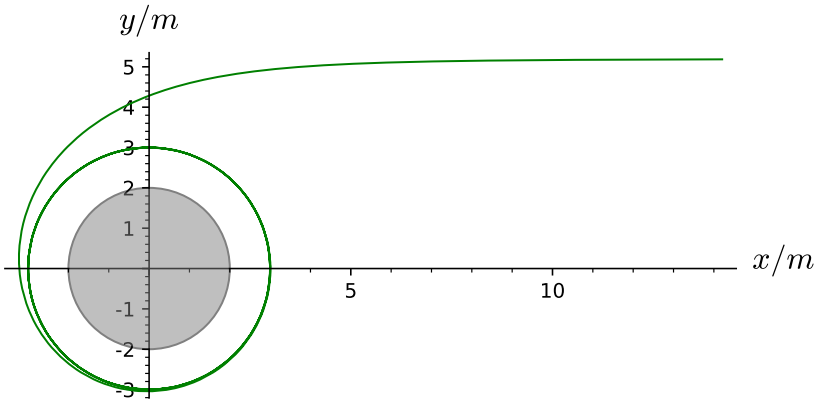


Figure 8.7: Trace in the equatorial plane spanned by the coordinates $x := r\cos\varphi$, $y := r\sin\varphi$ of an external critical null geodesic ($b = 3\sqrt{3}\,m \simeq 5.196m$ and $r > 3m$) with $\varphi_\infty = 0$. It obeys Eq. (8.40) with $\varphi_0 = -\varphi_*$ and $\varphi \in (0, +\infty)$ (right branch of the blue curve in Fig. 8.6). *[Figure produced with the notebook D.3.10]*

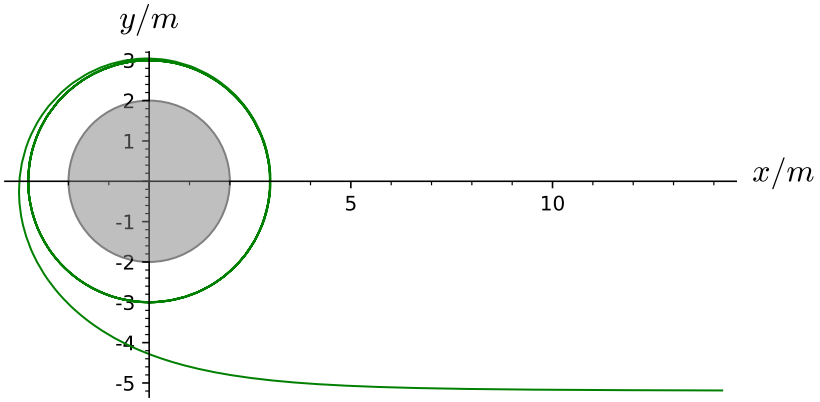


Figure 8.8: Same as Fig. 8.7, but for $\varphi_0 = \varphi_*$ and $\varphi \in (-\infty, 0)$ (left branch of the blue curve in Fig. 8.6). *[Figure produced with the notebook D.3.10]*

Again, φ_0 is an integration constant. But this time, it cannot be determined by the value of φ when $r \to +\infty$ since we are in the case $r < 3m$. The function $r(\varphi)$ given by Eq. (8.42) is plotted as the red curve in Fig. 8.6. Contrary to external critical null geodesics, there is no exclusion interval for φ. However, φ cannot range from $-\infty$ to $+\infty$ along a given geodesic. Indeed, for $\varphi = \varphi_0$, one gets $r = 0$, which corresponds to the curvature singularity of the Schwarzschild black hole. This is necessarily a termination point for any geodesic. The maximum range of φ along an internal critical null geodesic is therefore either $(-\infty, \varphi_0)$ or $(\varphi_0, +\infty)$.

An internal critical null geodesic with $\varphi_0 = 0$ and $\varphi \in (-\infty, 0)$ is depicted in the left panel of Fig. 8.9. For $\varphi \to -\infty$, it rolls up indefinitely onto the photon orbit ($r = 3m$) from below. Again the approach to the photon orbit is exponentially fast, the asymptotic expansion of relation (8.42) being $r \sim 3m(1 - 6\mathrm{e}^{\varphi})$ when $\varphi \to -\infty$. The curve plotted in Fig. 8.9 actually represents the trace in the equatorial plane of two distinct geodesics:

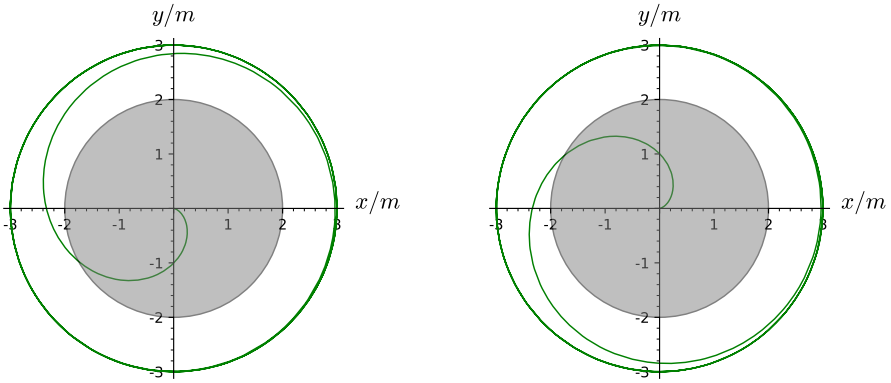


Figure 8.9: Trace in the equatorial plane of two internal critical null geodesics ($b = 3\sqrt{3}\,m \simeq 5.196m$ and $r < 3m$). They both obey Eq. (8.42) with $\varphi_0 = 0$. The left panel corresponds to $\varphi \in (-\infty, 0)$ (left part of the red curve in Fig. 8.6), while the right one is for $\varphi \in (0, +\infty)$ (right part of the red curve in Fig. 8.6). *[Figure produced with the notebook D.3.10]*

(i) a geodesic with $L > 0$ arising from $r = r_{\rm em} < 3m$, $\varphi = \varphi_{\rm em} < 0$ and inspiralling to the central singularity as $\varphi \to 0$ and (ii) a geodesic with $L < 0$ arising from $r = r_{\rm em} \in (2m, 3m)$, $\varphi = \varphi_{\rm em} < 0$ and spiralling outward to the photon orbit as $\varphi \to -\infty$. Note that for (ii), the emission point must fulfil $r_{\rm em} > 2m$, i.e. must lie outside the black hole region. Indeed, as shown in Sec. 7.2.2, along any geodesic, r must decrease towards the future in the black hole region. Hence no null geodesic can be spiralling outwards there.

Remark 3: Actually, an outward spiralling critical null geodesic can emerge from the region $r < 2m$ when the latter corresponds to the *white hole* region in the extended Schwarzschild spacetime that will be discussed in Chap. 9. This explains why nothing in Eq. (8.42) and Fig. 8.9 seems to prevent this behavior.

The right panel of Fig. 8.9 corresponds to an internal critical null geodesic with $\varphi_0 = 0$ and $\varphi \in (0, \infty)$. For $\varphi \to +\infty$, it rolls up indefinitely onto the photon orbit ($r = 3m$) from below. Again, the plotted curve describes two cases: (i) a geodesic with $L < 0$ arising from $r = r_{\rm em} < 3m$, $\varphi = \varphi_{\rm em} > 0$ and inspiralling to the central singularity as $\varphi \to 0$ and (ii) a geodesic with $L > 0$ arising from $r = r_{\rm em} \in (2m, 3m)$, $\varphi = \varphi_{\rm em} > 0$ and spiralling outward to the photon orbit as $\varphi \to +\infty$.

Historical note: The photon circular orbit at $r = 3m$ has been exhibited by David Hilbert in December 1916 [152, 153]. Hilbert also discussed the critical null geodesics and their spirals around the photon orbit, interpreting the latter as a Poincaré limit cycle of the dynamical system governed by Eq. (8.25). Hilbert pointed out that incoming null geodesics with an impact parameter $b > b_{\rm c}$ are deflected (possibly looping around the photon orbit if b is close to $b_{\rm c}$) and escape to infinity, while those with $b < b_{\rm c}$ cross the photon orbit and terminate on "the circle $r = 2m$". Hilbert claimed that null geodesics stop there, because their (coordinate!) velocity vanishes[4]. We recover here the interpretation of the $r = 2m$ (coordinate) singularity as an impenetrable sphere advanced by Droste while discussing timelike geodesics (cf. the historical

[4]The coordinate velocity for incoming radial null geodesics is given by Eq. (8.6): $dr/dt = -1 + 2m/r$, which clearly vanishes at $r = 2m$.

note on p. 165). The distinction between coordinate effects and physical ones was definitely not clear in the early days of general relativity, even for great minds like Hilbert! The same analysis and conclusions are found in the general relativity treatise by Max von Laue published in 1921 [174].

8.3.3 Null geodesics with $b > b_{\rm c}$ and $r > 3m$

For $b \neq b_{\rm c}$, the differential equation (8.25) cannot be integrated in terms of elementary functions. As we are going to see, it is integrable though in terms of standard functions of mathematical physics, namely elliptic integrals of the first kind. As a first step, we rewrite Eq. (8.25) as

$$\frac{\mathrm{d}\varphi}{\mathrm{d}u} = \epsilon_L \epsilon_{\rm in} \frac{1}{\sqrt{P_b(u)}}, \tag{8.43}$$

where $\epsilon_L = \pm 1$ is the sign of the conserved angular momentum L [cf. Eq. (8.13)] and $\epsilon_{\rm in} = \pm 1$ is defined by

$$\epsilon_{\rm in} := \mathrm{sgn}\left(\frac{\mathrm{d}u}{\mathrm{d}\tilde{\lambda}}\right), \tag{8.44}$$

i.e. $\epsilon_{\rm in} = +1$ if u increases towards the future along $\mathscr{L}$, or equivalently if r decreases towards the future along $\mathscr{L}$ (*inward* motion, hence the index "in"), and $\epsilon_{\rm in} = -1$ otherwise. Let us recall that by vertue of the equation of motion (8.10), ϵ_L gives the sign of $\mathrm{d}\varphi/\mathrm{d}\tilde{\lambda}$, so that the sign of $\mathrm{d}\varphi/\mathrm{d}u = \mathrm{d}\varphi/\mathrm{d}\tilde{\lambda} \times (\mathrm{d}u/\mathrm{d}\tilde{\lambda})^{-1}$ is $\epsilon_L\epsilon_{\rm in}$, which justifies the factor $\epsilon_L\epsilon_{\rm in}$ in front of the positive quantity $1/\sqrt{P_b(u)}$ in the right-hand side of Eq. (8.43).

In this section, we consider null geodesics with $b > b_{\rm c}$ and outside the photon sphere, i.e. geodesics similar to that labelled 1 in Figs. 8.1 and 8.2. Each of these geodesics has a periastron, at $u = u_{\rm p}$, and obeys $u \to 0$ ($r \to +\infty$) for both $\tilde{\lambda} \to -\infty$ and $\tilde{\lambda} \to +\infty$. The general solution to the differential equation (8.43) can be written as

$$\varphi = \varphi_{\rm p} + \epsilon_L \epsilon_{\rm in} \int_{u_{\rm p}}^{u} \frac{\mathrm{d}\bar{u}}{\sqrt{P_b(\bar{u})}}, \tag{8.45}$$

where $\varphi_{\rm p}$ is the value of φ at the periastron. We shall rewrite it as

$$\boxed{\varphi = \varphi_{\rm p} - \epsilon_L \epsilon_{\rm in} \Phi_b(u)}, \tag{8.46}$$

where we have introduced the function

$$\boxed{\Phi_b(u) := \int_u^{u_{\rm p}} \frac{\mathrm{d}\bar{u}}{\sqrt{P_b(\bar{u})}} = \int_u^{u_{\rm p}} \frac{\mathrm{d}\bar{u}}{\sqrt{2\bar{u}^3 - \bar{u}^2 + (m/b)^2}}} \quad (b > b_{\rm c}). \tag{8.47}$$

Since $u_{\rm p}$ is the function of b given by Eq. (8.30b), the above relation defines uniquely a function of u and b, which we consider as a function of u parameterized by b. Note that, since $u \leq u_{\rm p}$ (by definition of the periastron), one has $\Phi_b(u) \geq 0$. To evaluate $\Phi_b(u)$, we rewrite it in terms of the three zeros $(u_{\rm n}, u_{\rm p}, u_{\rm a})$ of P_b:

$$\Phi_b(u) = \int_u^{u_{\rm p}} \frac{\mathrm{d}\bar{u}}{\sqrt{2(\bar{u} - u_{\rm n})(u_{\rm p} - \bar{u})(u_{\rm a} - \bar{u})}}. \tag{8.48}$$

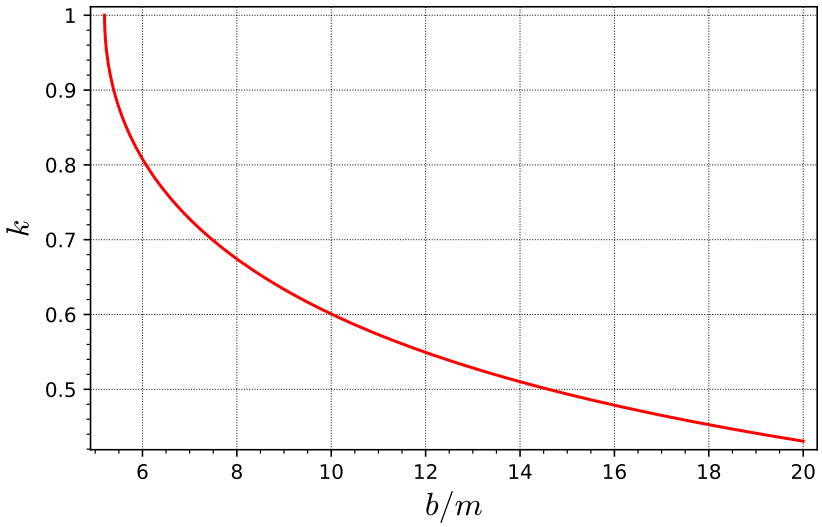


Figure 8.10: Modulus k of the elliptic integrals F and K that are involved in expression (8.53) for $\Phi_b(u)$. *[Figure produced with the notebook D.3.9]*

Let us perform the change of variable[5]

$$t := \frac{\bar{u} - u_\mathrm{n}}{u_\mathrm{p} - u_\mathrm{n}} \quad \Longleftrightarrow \quad \bar{u} = (u_\mathrm{p} - u_\mathrm{n})t + u_\mathrm{n}.$$

We get

$$\Phi_b(u) = \frac{1}{\sqrt{2(u_\mathrm{a} - u_\mathrm{n})}} \int_{\frac{u-u_\mathrm{n}}{u_\mathrm{p}-u_\mathrm{n}}}^{1} \frac{\mathrm{d}t}{\sqrt{t(1-t)(1-k^2 t)}}, \tag{8.49}$$

where k is the following constant:

$$\boxed{k := \sqrt{\frac{u_\mathrm{p} - u_\mathrm{n}}{u_\mathrm{a} - u_\mathrm{n}}} = \frac{\sqrt{2}}{\sqrt{\sqrt{3}\cot\left(\frac{2}{3}\arcsin\left(\frac{b_\mathrm{c}}{b}\right)\right) + 1}}} \qquad (b > b_\mathrm{c}), \tag{8.50}$$

the second equality following from the expressions (8.30) of u_n, u_p and u_a in terms of b.

We can simplify further the integral via a second change of variable:

$$t = \sin^2\vartheta \iff \vartheta = \arcsin\sqrt{t}.$$

Note that $0 < u \leq u_\mathrm{p}$ implies $0 < t \leq 1$, so that $0 < \vartheta \leq \pi/2$. Since $\mathrm{d}t = 2\sin\vartheta\cos\vartheta\,\mathrm{d}\vartheta = 2\sqrt{t(1-t)}\,\mathrm{d}\vartheta$, we arrive immediately at

$$\Phi_b(u) = \frac{\sqrt{2}}{\sqrt{u_\mathrm{a} - u_\mathrm{n}}} \int_{\phi_b(u)}^{\pi/2} \frac{\mathrm{d}\vartheta}{\sqrt{1 - k^2\sin^2\vartheta}}, \tag{8.51}$$

[5]It should be clear that the variable t introduced here has nothing to do with the Schwarzschild-Droste coordinate t.

where[6]

$$\boxed{\phi_b(u) := \arcsin\sqrt{\frac{u-u_{\rm n}}{u_{\rm p}-u_{\rm n}}}} \qquad (b > b_{\rm c}). \tag{8.52}$$

By splitting the integral according to $\int_\phi^{\pi/2} = \int_0^{\pi/2} - \int_0^\phi$, we rewrite (8.51) as

$$\boxed{\Phi_b(u) = \frac{\sqrt{2}}{\sqrt{u_{\rm a}-u_{\rm n}}}\left[K(k) - F\left(\phi_b(u),\ k\right)\right]} \qquad (b > b_{\rm c}), \tag{8.53}$$

where $F(\phi, k)$ is the ***incomplete elliptic integral of the first kind*** [34, 127, 2]:

$$\boxed{F(\phi,k) := \int_0^\phi \frac{{\rm d}\vartheta}{\sqrt{1-k^2\sin^2\vartheta}}} \tag{8.54}$$

and $K(k)$ is the ***complete elliptic integral of the first kind***:

$$\boxed{K(k) := \int_0^{\frac{\pi}{2}} \frac{{\rm d}\vartheta}{\sqrt{1-k^2\sin^2\vartheta}} = F\left(\frac{\pi}{2}, k\right)}. \tag{8.55}$$

The notation $F(\phi, k)$ is the most common one in the literature [34, 127], but one may encounter as well $F(\phi|m)$ for $F(\phi, k)$ with $m = k^2$ [2]. The parameter k is called the ***modulus*** of the elliptic integral. From its expression (8.50), we see that k is a function of b. It is plotted in Fig. 8.10. Given the ordering (8.31) and the limits (8.32), we deduce from expression (8.50) that

$$0 < k < 1, \tag{8.56}$$

with

$$\lim_{b\to b_{\rm c}^+} k = 1 \qquad \text{and} \qquad \lim_{b\to+\infty} k = 0. \tag{8.57}$$

Given the range $(0, u_{\rm p})$ of u, we deduce from expression (8.52) that

$$0 < \arcsin\sqrt{\frac{u_{\rm n}}{u_{\rm n}-u_{\rm p}}} < \phi_b(u) \leq \frac{\pi}{2}, \tag{8.58}$$

with $\phi_b(u_{\rm p}) = \pi/2$.

The function $\Phi_b(u)$ is plotted in Fig. 8.11 for various values of b. Note that, by construction [cf. Eq. (8.47)], $\Phi_b(u) \geq 0$ and $\Phi_b(u) = 0 \iff u = u_{\rm p}$. Note also that the closer b is from $b_{\rm c} \simeq 5.1961452$, the larger the amplitude of $\Phi_b(u)$. This point will be discussed further in Secs. 8.3.6 and 8.4.

Remark 4: One can express $\Phi_b(u)$ in terms of a single elliptic integral thanks to the following property of incomplete elliptic integrals of the first kind[7]:

$$\tan\psi\tan\phi = \frac{1}{\sqrt{1-k^2}} \implies F(\psi,k) = K(k) - F(\phi,k). \tag{8.59}$$

[6] Do no confuse the letter ϕ with the symbol used for the coordinate φ.

[7] See e.g. Eq. (117.01) of Ref. [34] with $k' = \sqrt{1-k^2}$ or Eq. (17.4.13) of Ref. [2] with $\sin\alpha = k$.

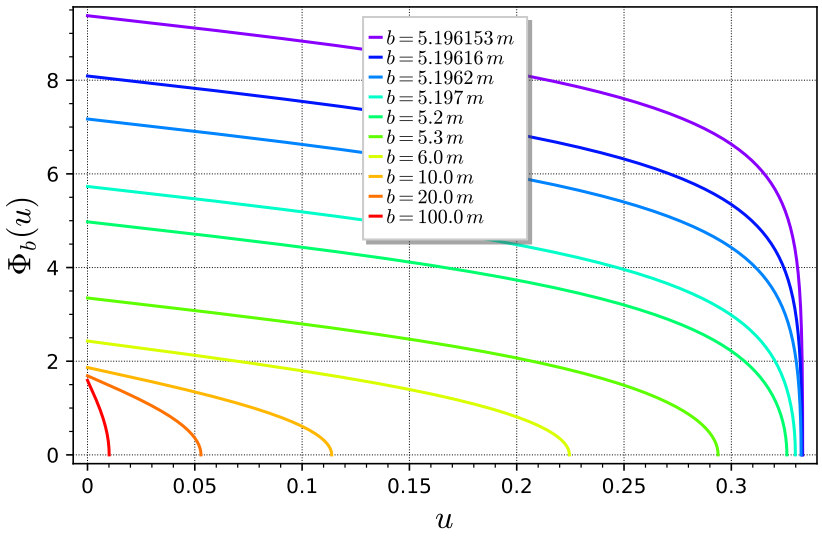


Figure 8.11: Function $\Phi_b(u)$ defined by Eq. (8.47) and evaluated via the elliptic integral expression (8.52)-(8.53), for selected values of $b > b_c$. For each value of b, the range of u is $(0, u_p]$, with the inverse periastron radius $u_p = u_p(b)$ given by Eq. (8.30b). *[Figure produced with the notebook D.3.11]*

Then by setting

$$\psi_b(u) := \arcsin\left(\frac{1}{k}\sqrt{\frac{u_p - u}{u_a - u}}\right), \tag{8.60}$$

we can rewrite Eq. (8.53) as

$$\Phi_b(u) = \frac{\sqrt{2}}{\sqrt{u_a - u_n}}\, F\left(\psi_b(u),\ k\right). \tag{8.61}$$

We prefer however the form (8.53) of $\Phi_b(u)$ because (i) expression (8.52) for $\phi_b(u)$ is simpler than expression (8.60) for $\psi_b(u)$ and (ii) the form (8.53) is better adapted to the study of the limit $b \to b_c$, to be discussed in Sec. 8.4.1.

Remark 5: The trajectory of a null geodesic in the plane $\theta = \pi/2$ as given by Eqs. (8.46) and (8.53) is of the form $\varphi = \varphi(r)$ on each of the two arcs with respect to the periastron. One can invert this relation to express the trajectory in the polar form $r = r(\varphi)$. This is performed thanks to the inverse of the incomplete elliptic integral $F(\phi, k)$, which is the ***Jacobi elliptic sine*** $\mathrm{sn}(x, k)$, defined by $\sin\phi = \mathrm{sn}(F(\phi, k), k)$. We deduce then from Eqs. (8.60)-(8.61) that

$$\frac{u_p - u}{u_a - u} = k^2\, \mathrm{sn}\left(\frac{\sqrt{u_a - u_n}}{\sqrt{2}}\Phi_b(u),\ k\right). \tag{8.62}$$

We refer the reader to Refs. [35, 192] for more details.

8.3.4 Null geodesics with $b > b_c$ and $r < 3m$

Let us consider now null geodesics still with $b > b_c$ but located below the photon sphere, i.e. geodesics similar to that labelled 4 in Figs. 8.1 and 8.2. Each of these geodesics has

an apoastron, at $u = u_{\rm a}$, where $u_{\rm a}$ is the function of b given by Eq. (8.30c). In other words, i.e. one has $u \geq u_{\rm a}$ along the geodesic.

The treatment is similar to that of Sec. 8.3.3, but with $u_{\rm a}$ playing the role of $u_{\rm p}$.

To be detailed later...

8.3.5 Null geodesics with $b < b_{\rm c}$

A null geodesic $\mathscr{L}$ with an impact parameter $b < b_{\rm c}$ has neither a periastron nor an apoastron, since P_b has no zero in the physical range $(0, +\infty)$ of u in that case (cf. Sec. 8.3.1 and Fig. 8.4). The only real zero of P_b is $u_{\rm n} < 0$, which is the function of b given by Eq. (8.34). Examples of such geodesics are those labelled 2 and 3 in Figs. 8.1 and 8.2. Having no periastron or apoastron implies that r, and hence u, is a monotonous function along $\mathscr{L}$. We shall then distinguish two cases:

- $\mathscr{L}$ is ***ingoing***: r decreases all along $\mathscr{L}$, from $+\infty$ to 0 as $\tilde{\lambda}$ varies from $-\infty$ to some value $\tilde{\lambda}_0$ where $\mathscr{L}$ hits the curvature singularity at $r = 0$:

$$\lim_{\tilde{\lambda}\to-\infty} r = +\infty \text{ and } \lim_{\tilde{\lambda}\to\tilde{\lambda}_0} r = 0 \quad \Longleftrightarrow \quad \lim_{\tilde{\lambda}\to-\infty} u = 0 \text{ and } \lim_{\tilde{\lambda}\to\tilde{\lambda}_0} u = +\infty. \tag{8.63}$$

- $\mathscr{L}$ is ***outgoing***: r increases all along $\mathscr{L}$; as shown in Sec. 7.2.2, this cannot happen in the region $r < 2m$ (the black hole interior). Hence, we must have r varying from $2m$ to $+\infty$, with $\tilde{\lambda}$ ranging from some finite value, $\tilde{\lambda}_1$ say, to $+\infty$:

$$\lim_{\tilde{\lambda}\to\tilde{\lambda}_1} r = 2m \text{ and } \lim_{\tilde{\lambda}\to+\infty} r = +\infty \quad \Longleftrightarrow \quad \lim_{\tilde{\lambda}\to\tilde{\lambda}_1} u = \frac{1}{2} \text{ and } \lim_{\tilde{\lambda}\to+\infty} u = 0. \tag{8.64}$$

That the range of $\tilde{\lambda}$ for outgoing geodesics is $(\lambda_1, +\infty)$ and not the whole real line can be understood by considering the limit $b \to 0$. We are then in the case of radial null geodesics, for which it has been proved that r is an affine parameter (cf. Secs. 6.3.1 and 8.2.2). Since this particular affine parameter obviously takes a finite value at $r = 2m$, any other affine parameter must take a finite value as well. Actually, we shall see in Chap. 9 (cf. Sec. 9.3.1) that this $r = 2m$ limit in the past of outgoing null geodesics does not correspond to the black hole horizon but to the event horizon of a white hole, which is a located in a part of the spacetime that is not covered by the Schwarzschild-Droste coordinates considered here.

Since $P_b(u_{\rm n}) = 0$, we have $m^2/b^2 = -2u_{\rm n}^3 + u_{\rm n}^2$, so that we may write

$$P_b(u) = 2(u^3 - u_{\rm n}^3) - (u^2 - u_{\rm n}^2) = (u - u_{\rm n})\left[2(u^2 + uu_{\rm n} + u_{\rm n}^2) - (u + u_{\rm n})\right].$$

After a slight rearrangement of the term in square brackets, we arrive at

$$P_b(u) = 2(u - u_{\rm n})\left[(u - u_0)^2 + (u_* - u_{\rm n})^2 - (u_0 - u_{\rm n})^2\right], \tag{8.65}$$

with

$$\boxed{u_0 := \frac{1}{4} - \frac{u_{\rm n}}{2}} \quad \text{and} \quad \boxed{u_* := \sqrt{u_{\rm n}(3u_{\rm n} - 1)} + u_{\rm n}}. \tag{8.66}$$

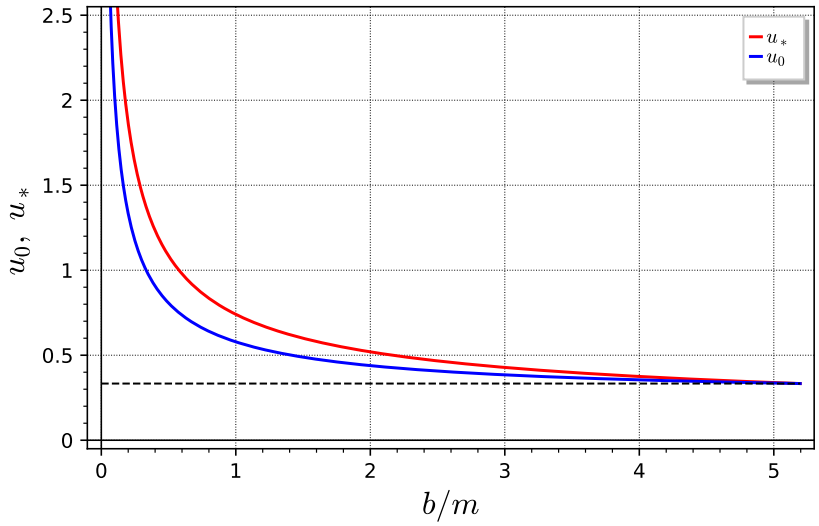


Figure 8.12: Parameters u_0 and u_*, defined by Eq. (8.66), as functions of b. The horizontal dashed line marks $u = 1/3$. *[Figure produced with the notebook D.3.12]*

Note that u_0 and u_* are functions of b, via the expression (8.34) of u_n. They are plotted in Fig. 8.12. It is clear from this figure that $u_* > u_0$; consequently the term inside the square brackets in Eq. (8.65) is always positive, in agreement with u_n being the only real zero of P_b for $b < b_\mathrm{c}$. Furthermore, according to the limits (8.35), we have

$$\lim_{b\to 0} u_* = +\infty \qquad \text{and} \qquad \lim_{b\to b_\mathrm{c}^-} u_* = \frac{1}{3}. \tag{8.67}$$

The general solution of the differential equation (8.43) can be written as

$$\boxed{\varphi = \varphi_* - \epsilon_L \epsilon_\mathrm{in} \Phi_b(u)}, \tag{8.68}$$

where

$$\boxed{\Phi_b(u) := \int_u^{u_*} \frac{\mathrm{d}\bar{u}}{\sqrt{P_b(\bar{u})}} = \int_u^{u_*} \frac{\mathrm{d}\bar{u}}{\sqrt{2\bar{u}^3 - \bar{u}^2 + (m/b)^2}}} \quad (b < b_\mathrm{c}) \tag{8.69}$$

and φ_* is the value of φ at $u = u_*$. Since

$$\lim_{b\to b_\mathrm{c}^-} u_* = \frac{1}{3} = \lim_{b\to b_\mathrm{c}^+} u_\mathrm{p},$$

there is a kind of continuity of (8.69) with the definition (8.47) of $\Phi_b(u)$ for $b > b_\mathrm{c}$, despite the fact that $\Phi_b(u)$ is not defined for $b = b_\mathrm{c}$ (more precisely, as we shall see later, $\lim_{b\to b_\mathrm{c}^-} \Phi_b(u) = +\infty$ for $u < u_*$ and $\lim_{b\to b_\mathrm{c}^+} \Phi_b(u) = +\infty$ for $u < u_\mathrm{p}$).

To evaluate $\Phi_b(u)$, we shall use expression (8.65) for $P_b(u)$:

$$\Phi_b(u) = \frac{1}{\sqrt{2}} \int_u^{u_*} \frac{\mathrm{d}\bar{u}}{\sqrt{(\bar{u} - u_\mathrm{n})\left[(\bar{u} - u_0)^2 + (u_* - u_\mathrm{n})^2 - (u_0 - u_\mathrm{n})^2\right]}}.$$

Performing the change of variable

$$t := \frac{u_* - \bar{u}}{u_* + \bar{u} - 2u_{\rm n}} \quad \Longleftrightarrow \quad \bar{u} = (u_* - u_{\rm n})\frac{1-t}{1+t} + u_{\rm n} \tag{8.70}$$

yields

$$\Phi_b(u) = \int_0^{\frac{u_*-u}{u_*+u-2u_{\rm n}}} \frac{{\rm d}t}{\sqrt{1-t^2}\sqrt{(u_*-u_{\rm n})(1+t^2)-(u_0-u_{\rm n})(1-t^2)}}. \tag{8.71}$$

Given the range $(0,+\infty)$ for u, the range of t is

$$-1 < t \leq \frac{u_*}{u_* - 2u_{\rm n}} < 1.$$

To proceed, we shall distinguish the cases $u \leq u_*$ and $u > u_*$. The first case corresponds to $0 \leq t < \frac{u_*}{u_*-2u_{\rm n}} < 1$ and we perform the change of variable $t = \cos\vartheta$ with $\vartheta \in (0, \pi/2)$ in the integral (8.71), yielding

$$\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_{\rm n})}} \int_{\phi_b(u))}^{\pi/2} \frac{{\rm d}\vartheta}{\sqrt{1 - k^2\sin^2\vartheta}},$$

with

$$\boxed{\phi_b(u) := \arccos\left(\frac{|u_* - u|}{u_* + u - 2u_{\rm n}}\right)} \qquad (b < b_{\rm c}). \tag{8.72}$$

and[8]

$$\boxed{k := \sqrt{\frac{u_* - 5u_{\rm n}/2 + 1/4}{2(u_* - u_{\rm n})}}} \qquad (b < b_{\rm c}). \tag{8.73}$$

The absolute value in the right-hand side of Eq. (8.72) is not necessary in the present case since $u \geq u_*$. We keep it because the same expression will be used below for the case $u > u_*$. To let appear the incomplete elliptic integral of the first kind (8.54), let us rewrite the above expression as

$$\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_{\rm n})}}\left[\int_0^{\pi/2} \frac{{\rm d}\vartheta}{\sqrt{1 - k^2\sin^2\vartheta}} - \int_0^{\phi_b(u)} \frac{{\rm d}\vartheta}{\sqrt{1 - k^2\sin^2\vartheta}}\right].$$

In view of respectively (8.55) and (8.54), the first integral is $K(k)$, while the second one is $F(\phi_b(u), k)$. We have thus

$$\boxed{\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_{\rm n})}}\left[K(k) - F\left(\phi_b(u),\, k\right)\right]} \qquad (u \leq u_*,\ b < b_{\rm c}). \tag{8.74}$$

As for the case $b > b_{\rm c}$, the modulus k of the elliptic integrals K and F is a function of b only, which is given by combining Eqs. (8.73), (8.66) and (8.34). It is plotted in Fig. 8.13. Note that its range is pretty limited:

$$\underbrace{\frac{1}{2}\sqrt{2+\sqrt{3}}}_{\simeq 0.96592} < k < 1. \tag{8.75}$$

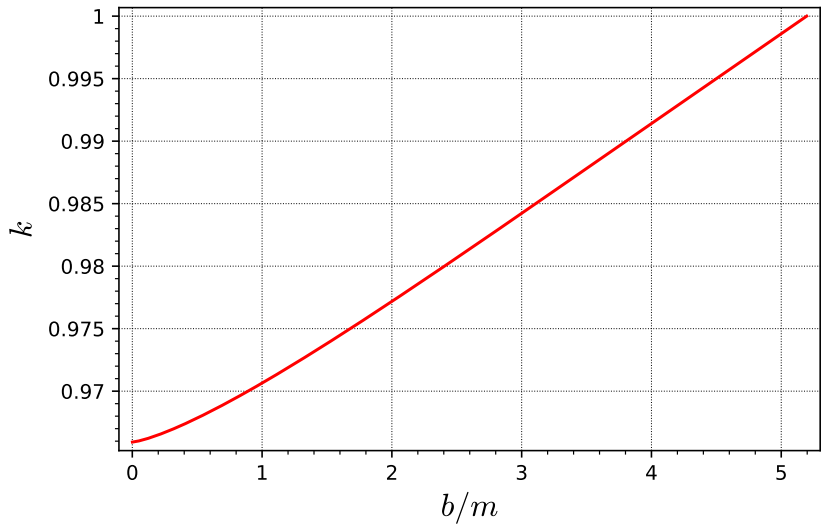


Figure 8.13: Modulus k of the elliptic integrals F and K that are involved in expressions (8.74) and (8.76) for $\Phi_b(u)$. *[Figure produced with the notebook D.3.12]*

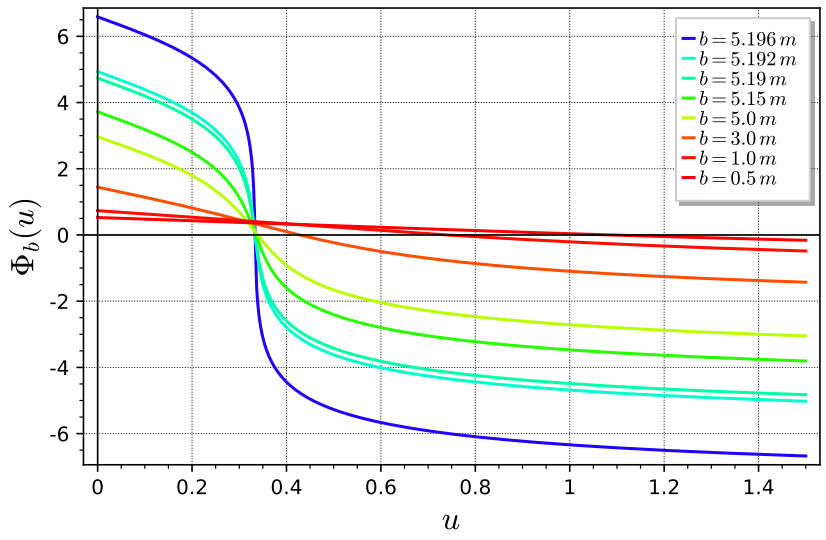


Figure 8.14: Function $\Phi_b(u)$ defined by Eq. (8.69) and evaluated via the elliptic integral expressions (8.74) and (8.76), for selected values of $b < b_c$. *[Figure produced with the notebook D.3.12]*

Let us now turn to the case $u > u_*$. It implies $-1 < t < 0$, so that we perform the change of variable $t = -\cos\vartheta$ in order to keep $\vartheta \in (0, \pi/2)$. We obtain then

$$\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_{\rm n})}} \int_{\pi/2}^{\phi_b(u)} \frac{\mathrm{d}\vartheta}{\sqrt{1 - k^2 \sin^2\vartheta}},$$

where $\phi_b(u)$ and k are given by Eqs. (8.72) and (8.73). We conclude that

$$\boxed{\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_{\rm n})}} \left[F\left(\phi_b(u),\, k\right) - K(k)\right]} \qquad (u > u_*,\ b < b_{\rm c}). \tag{8.76}$$

The function Φ_b evaluated via Eqs. (8.74) and (8.76) is plotted in Fig. 8.14 for various values of b. Note that, by construction [cf. Eq. (8.69)], $\Phi_b(u) = 0 \iff u = u_*$. Note also that, as in the case $b > b_{\rm c}$ (Sec. 8.3.3), the closer b is from $b_{\rm c} \simeq 5.1961452$, the larger the amplitude of $\Phi_b(u)$.

8.3.6 Deflection angle and winding number

Let us consider a null geodesic $\mathscr{L}$ with $b > b_{\rm c}$ arising from $r \to +\infty$, i.e. belonging to the family studied in Sec. 8.3.3. $\mathscr{L}$ reaches some periastron and departs to $r \to +\infty$. The total change of φ along the geodesic history is

$$\Delta\varphi = \varphi_\infty - \varphi_{-\infty}, \tag{8.77}$$

with $\varphi_\infty := \lim_{\tilde\lambda\to+\infty} \varphi$ and $\varphi_{-\infty} := \lim_{\tilde\lambda\to-\infty} \varphi$. Given that $\lim_{\tilde\lambda\to\pm\infty} u = 0$, Eq. (8.46) with respectively $\epsilon_{\rm in} = +1$ and $\epsilon_{\rm in} = -1$ leads to

$$\varphi_{-\infty} = \varphi_{\rm p} - \epsilon_L \Phi_b(0) \qquad \text{and} \qquad \varphi_\infty = \varphi_{\rm p} + \epsilon_L \Phi_b(0). \tag{8.78}$$

Hence

$$\Delta\varphi = 2\epsilon_L \Phi_b(0). \tag{8.79}$$

Note that one can have $|\Delta\varphi| > 2\pi$; in such a case, the null geodesic is winding around the black hole before leaving to infinity. We therefore introduce the ***winding number*** $n \in \mathbb{Z}$ by

$$\Delta\varphi =: \overline{\Delta\varphi} + 2\pi n \quad \text{with} \quad \begin{cases} \overline{\Delta\varphi} \in [0, 2\pi) & \text{if } L > 0 \\ \overline{\Delta\varphi} \in (-2\pi, 0] & \text{if } L < 0. \end{cases} \tag{8.80}$$

Note that $n \geq 0$ for $L > 0$ and $n \leq 0$ for $L < 0$. We then define the ***deflection angle*** Θ by

$$\Theta := \overline{\Delta\varphi} - \epsilon_L \pi. \tag{8.81}$$

Let us recall that $\epsilon_L = \pm 1$ is the sign of the conserved angular momentum L [cf. Eq. (8.13)]. The $-\epsilon_L\pi$ term in the above equation is chosen so that $\Theta = 0$ in flat spacetime (no deflection of light). By construction, the range of Θ is

$$-\pi \leq \Theta \leq \pi. \tag{8.82}$$

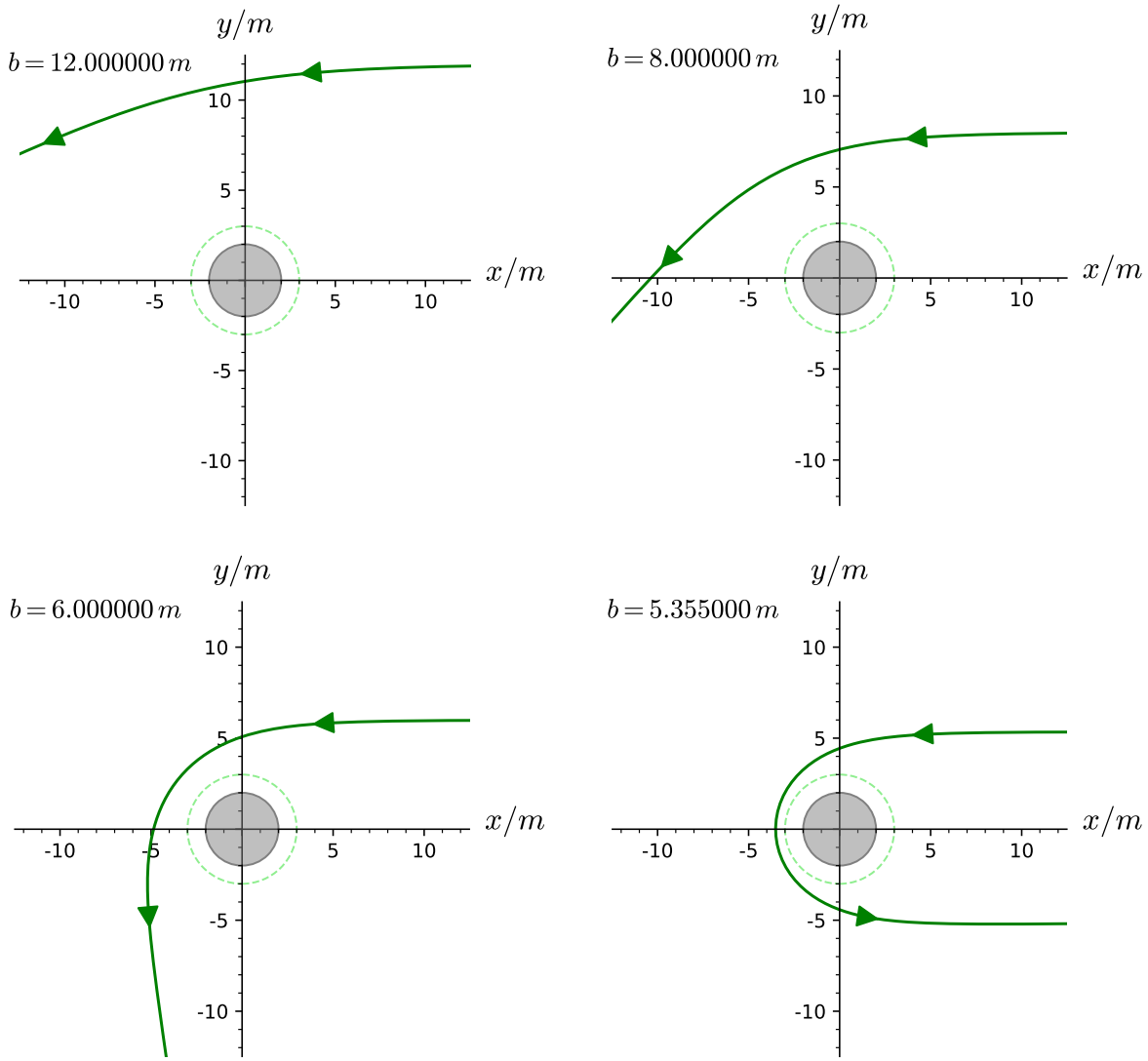


Figure 8.15: Null geodesics in the plane $\theta = \pi/2$ of Schwarzschild spacetime, plotted in terms of the coordinates $(x, y) := (r\cos\varphi, r\sin\varphi)$. All geodesics arise from $x \gg m$ with trajectories initially parallel to the x-axis; they differ by the value of the impact parameter b. The grey disk marks the black hole region $r < 2m$, while the dashed green circle indicates the photon orbit at $r = 3m$. *[Figure produced with the notebook D.3.8]*

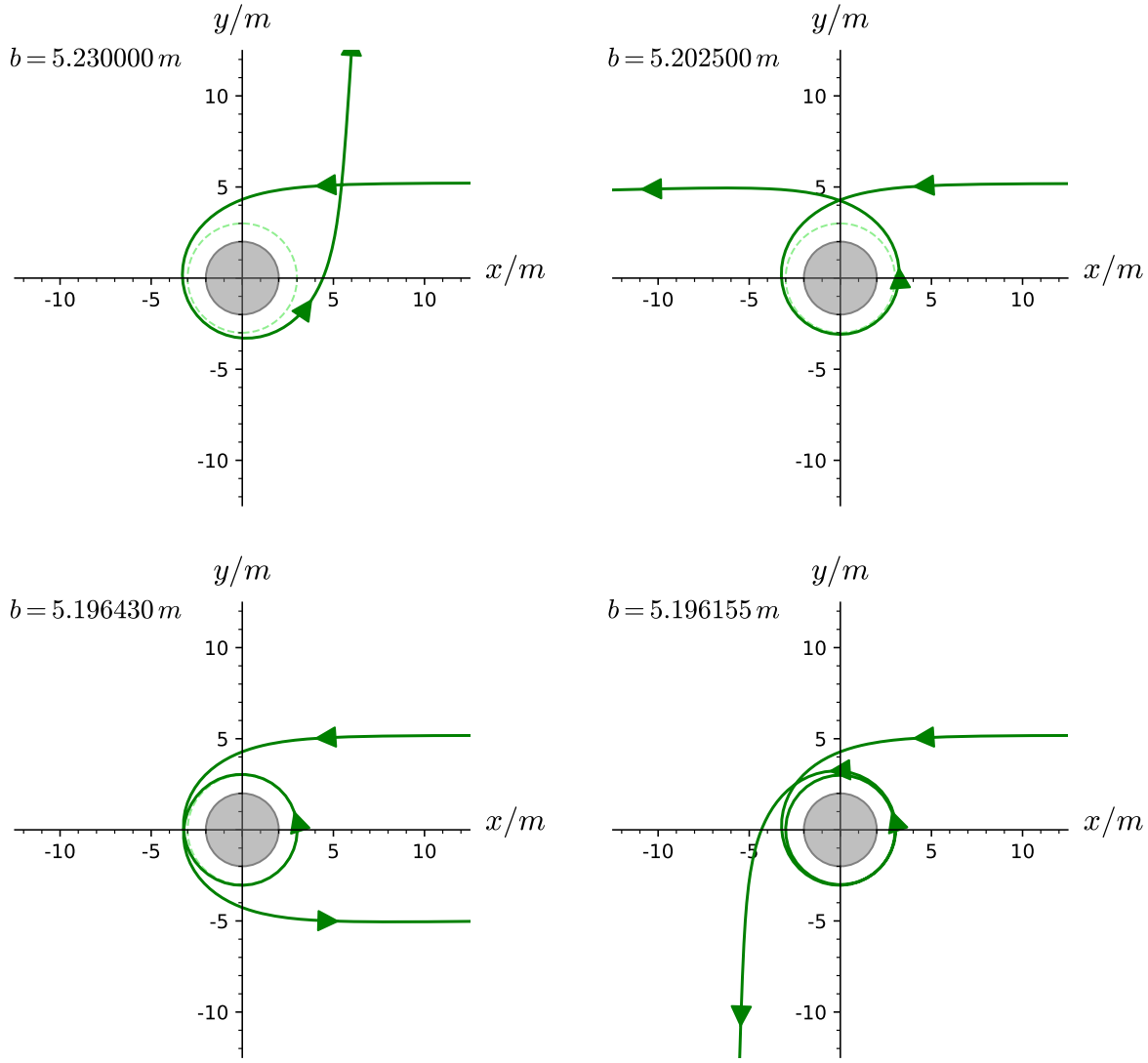


Figure 8.16: Same as Fig. 8.15 but for values of b closer to $b_c \simeq 5.196152\,m$. *[Figure produced with the notebook D.3.8]*

The above concepts are illustrated in Figs. 8.15-8.16, which show the trajectories of null geodesics arising from infinity along the direction $\varphi = 0$ (i.e. having $\varphi_{-\infty} = 0$), for various values of the impact parameter b decaying from $b = 12m$ to $b_{\rm c}$. They have been computed by means of the geodesic integrator of SageMath (cf. Appendix D), instead of making use of the elliptic-integral expression (8.53). All these geodesics have $L > 0$ and hence $\epsilon_L = +1$. For $b = 12m$ (upper left panel of Fig. 8.15), the geodesic suffers only some moderate bending: the deflection angle is $\Theta \simeq \pi/6$ and the winding number is $n = 0$. We recover here the standard ***deflection of light*** by massive bodies in general relativity. For $b = 8m$ and $6m$, the bending is more pronounced, exceeding $\Theta = \pi/2$ for $b = 6m$, still with $n = 0$. For $b = 5.355\,m$ (lower right panel of Fig. 8.15), the deflection angle is $\Theta \simeq \pi$, i.e. the photon goes back in the direction from which it was coming.

When the impact parameter becomes even closer to the critical value $b_{\rm c} \simeq 5.196152\,m$ [Eq. (8.21)], the null geodesic starts to wind around the black hole before escaping to infinity (Fig. 8.16). For $b = 5.230\,m$ (upper left panel of Fig. 8.16), the winding number is $n = 1$ and the deflection angle is $\Theta \simeq -\pi/2$. For $b = 5.2025\,m$ (upper right panel of Fig. 8.16), one has $n = 1$ and $\Theta = 0$. For $b = 5.19643\,m$ (lower left panel of Fig. 8.16), one has $n = 1$ and $\Theta \simeq \pi$ and for $b = 5.196155\,m$ (lower right panel of Fig. 8.16), one has $n = 2$ and $\Theta \simeq \pi/2$. Note that the winding is taking place almost at the photon circular orbit ($r = 3m$). That after a few turns the null geodesic departs to infinity corroborates the fact that the photon orbit is unstable (cf. Sec. 8.2.3).

We can understand the winding phenomenon around the photon circular orbit for b close to $b_{\rm c}$ without investigating the properties of elliptic integrals. Indeed, by combining Eqs. (8.79) and (8.48), we get

$$\Delta\varphi = \epsilon_L \sqrt{2} \int_0^{u_{\rm p}} \frac{{\rm d}u}{\sqrt{(u_{\rm p} - u)(u_{\rm a} - u)(u - u_{\rm n})}}. \tag{8.83}$$

Given the ordering (8.31), each of the three factors under the square root is positive on the integration range $(0, u_{\rm p})$. For $b \neq b_{\rm c}$, one has $u_{\rm a} \neq u_{\rm p}$ (cf. Fig. 8.5) and the only diverging term in the integrand of (8.83) is $1/\sqrt{u_{\rm p} - u}$, which diverges at the integral boundary $u = u_{\rm p}$. However, the integral

$$\int_0^{u_{\rm p}} \frac{{\rm d}u}{\sqrt{u_{\rm p} - u}}$$

is finite, being equal to $2\sqrt{u_{\rm p}}$, so that $\Delta\varphi$ remains finite. When $b \to b_{\rm c}$, $u_{\rm a} \to u_{\rm p}$ and the integral (8.83) has a behaviour similar to

$$\int_0^{u_{\rm p}} \frac{{\rm d}u}{\sqrt{(u_{\rm p} - u)^2}} = \int_0^{u_{\rm p}} \frac{{\rm d}u}{u_{\rm p} - u}.$$

Since the latter is a diverging integral, we conclude that

$$\boxed{\Delta\varphi \to \pm\infty \quad \text{when} \quad b \to b_{\rm c}}. \tag{8.84}$$

We recover the behaviour observed for $b = b_{\rm c}$ in Sec. 8.3.2: for an external critical null geodesic, $\Delta\varphi$ is infinite, the geodesic spiralling indefinitely around the photon orbit (cf. Fig. 8.7). We shall refine (8.84) in Sec. 8.4.1 [Eq. (8.102)].

[8] To get expression (8.73), we have used Eq. (8.66) to get rid of u_0.

8.4 Asymptotic direction from some emission point

To discuss images in Sec. 8.5, we shall need the change in φ between some arbitrary point on a null geodesic (the "emission" point) and a point far away from the black hole (the "reception" point). The total change $\Delta\varphi$ along all the geodesic history considered in Sec. 8.3.6 is then a special case: that for which the emission point is infinitely far from the black hole.

8.4.1 Asymptotic direction for $b > b_c$

Let us consider a null geodesic $\mathscr{L}$ with $b > b_c$ and an event of (finite) affine parameter $\tilde{\lambda} = \tilde{\lambda}_{em}$ on $\mathscr{L}$, which we shall call the *emission point*. We are interested in null geodesics which reach the asymptotic flat region $r \to +\infty$ for $\tilde{\lambda} \to +\infty$. This implies that the emission point is located outside the photon sphere, i.e. obeys

$$r_{em} := r(\tilde{\lambda}_{em}) > 3m. \tag{8.85}$$

Indeed, we have seen in Sec. 8.3.4 that null geodesics with $b > b_c$ and emitted under the photon sphere never cross the later (this is also obvious from the effective potential profile, as plotted in Fig. 8.1, cf. trajectory no. 4). Our aim is to relate the value φ_∞ of φ for $\tilde{\lambda} \to +\infty$ (the asymptotic direction) to its value φ_{em} at $\tilde{\lambda} = \tilde{\lambda}_{em}$.

If the emission point is past $\mathscr{L}$'s periastron, i.e. if $\tilde{\lambda}_{em} > \tilde{\lambda}_p$, φ_{em} is related to φ_p by Eq. (8.46) with $\epsilon_{in} = -1$:

$$\varphi_{em} = \varphi_p + \epsilon_L \Phi_b(u_{em}), \tag{8.86}$$

where $u_{em} := m/r_{em}$. On the other side, the asymptotic value φ_∞ is related to φ_p by Eq. (8.78). By combining these two relations to eliminate φ_p, we get

$$\varphi_\infty = \varphi_{em} + \epsilon_L \left[\Phi_b(0) - \Phi_b(u_{em})\right]. \tag{8.87}$$

Substituting expression (8.53) for Φ_b yields

$$\boxed{\varphi_\infty = \varphi_{em} + \epsilon_L \frac{\sqrt{2}}{\sqrt{u_a - u_n}} \left[F\left(\phi_b(u_{em}), k\right) - F\left(\phi_b(0), k\right)\right]} \qquad \begin{matrix}(b > b_c)\\ (\tilde{\lambda}_{em} > \tilde{\lambda}_p)\end{matrix}, \tag{8.88}$$

where, according to Eq. (8.52),

$$\phi_b(u_{em}) = \arcsin\sqrt{\frac{u_{em} - u_n}{u_p - u_n}} \qquad \text{and} \qquad \phi_b(0) = \arcsin\sqrt{\frac{|u_n|}{u_p - u_n}}. \tag{8.89}$$

If the emission point is located prior to the periastron, i.e. if $\tilde{\lambda}_{em} < \tilde{\lambda}_p$, φ_{em} is related to φ_p by Eq. (8.46) with $\epsilon_{in} = +1$:

$$\varphi_{em} = \varphi_p - \epsilon_L \Phi_b(u_{em}) \tag{8.90}$$

and we get, by combining with Eq. (8.78),

$$\varphi_\infty = \varphi_{em} + \epsilon_L \left[\Phi_b(0) + \Phi_b(u_{em})\right]. \tag{8.91}$$

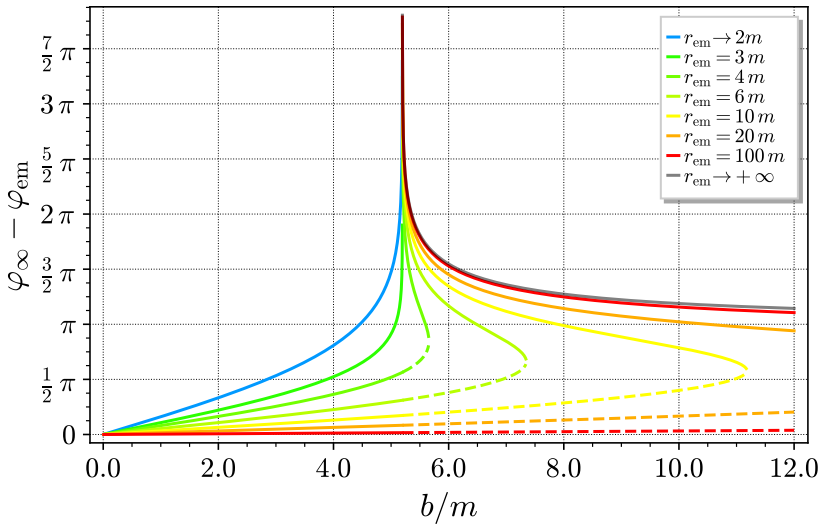


Figure 8.17: Total change $\varphi_\infty - \varphi_{\rm em}$ in φ from the emission point at $(r, \varphi) = (r_{\rm em}, \varphi_{\rm em})$ as a function of the impact parameter b of the null geodesic, assuming $\epsilon_L = +1$. The dashed curves correspond to trajectories with $b > b_{\rm c}$ that do not pass through the periastron. For each value of $r_{\rm em} > 3m$, the maximum of b is given by $r_{\rm p}(b) = r_{\rm em}$, i.e. by $b^2 = r_{\rm em}^3/(r_{\rm em} - 2m)$ [cf. Eq. (8.18) and Fig. 8.3]. *[Figure produced with the notebook D.3.11]*

Equation (8.53) then yields

$$\boxed{\varphi_\infty = \varphi_{\rm em} + \epsilon_L \frac{\sqrt{2}}{\sqrt{u_{\rm a} - u_{\rm n}}} \left[2K(k) - F\left(\phi_b(u_{\rm em}), k\right) - F\left(\phi_b(0), k\right)\right]} \qquad \begin{matrix}(b > b_{\rm c}) \\ (\tilde{\lambda}_{\rm em} < \tilde{\lambda}_{\rm p})\end{matrix}. \tag{8.92}$$

$\varphi_\infty - \varphi_{\rm em}$ is plotted as a function of b in Fig. 8.17 for various values of $r_{\rm em} = m/u_{\rm em}$, with the dashed curves corresponding to the case $\tilde{\lambda}_{\rm em} > \tilde{\lambda}_{\rm p}$. We notice that for $b \gg m$ (cf. the red curve near $b = 12m$), one has either $\varphi_\infty - \varphi_{\rm em} \simeq 0$ (dashed red curve: $\tilde{\lambda}_{\rm em} > \tilde{\lambda}_{\rm p}$) or $\varphi_\infty - \varphi_{\rm em} \simeq \pi$ (solid red curve: $\tilde{\lambda}_{\rm em} < \tilde{\lambda}_{\rm p}$): null geodesics with large impact parameters suffer almost no deflection, as expected. Another striking feature of Fig. 8.17 is the divergence of $\varphi_\infty - \varphi_{\rm em}$ when b tends to $b_{\rm c}$. Let us examine this in details.

Limit $b \to b_{\rm c}^+$

From Eqs. (8.30), we have the following limits:

$$\lim_{b \to b_{\rm c}^+} u_{\rm n} = -\frac{1}{6} \qquad \text{and} \qquad \lim_{b \to b_{\rm c}^+} u_{\rm p} = \lim_{b \to b_{\rm c}^+} u_{\rm a} = \frac{1}{3}. \tag{8.93}$$

It follows then from expressions (8.89) that

$$\lim_{b \to b_{\rm c}^+} \phi_b(u_{\rm em}) = \arcsin \sqrt{2u_{\rm em} + \frac{1}{3}} \qquad \text{and} \qquad \lim_{b \to b_{\rm c}^+} \phi_b(0) = \arcsin\left(\frac{1}{\sqrt{3}}\right). \tag{8.94}$$

Besides, $\lim_{b\to b_c^+} k = 1$ [cf. Eq. (8.57)] and, for any $\phi \in [0, \pi/2)$, the definition (8.54) of F leads to

$$F(\phi, 1) = \int_0^\phi \frac{\mathrm{d}\vartheta}{\sqrt{1-\sin^2\vartheta}} = \int_0^\phi \frac{\mathrm{d}\vartheta}{\cos\vartheta} = \ln\left(\frac{1+\sin\phi}{\cos\phi}\right). \tag{8.95}$$

The limits (8.94) result then in

$$\lim_{b\to b_c^+} F\left(\phi_b(u_{\mathrm{em}}), k\right) = \ln\left(\frac{\sqrt{3}+\sqrt{6u_{\mathrm{em}}+1}}{\sqrt{2(1-3u_{\mathrm{em}})}}\right) \quad \text{and} \quad \lim_{b\to b_c^+} F\left(\phi_b(0), k\right) = \ln\left(\frac{\sqrt{3}+1}{\sqrt{2}}\right). \tag{8.96}$$

For $\tilde{\lambda}_{\mathrm{em}} > \tilde{\lambda}_{\mathrm{p}}$, inserting these formulas into (8.88) leads to

$$\boxed{\varphi_\infty \underset{b\to b_c^+}{\sim} \varphi_{\mathrm{em}} + \epsilon_L \ln\left(\frac{2+3u_{\mathrm{em}}+\sqrt{3(6u_{\mathrm{em}}+1)}}{(2+\sqrt{3})(1-3u_{\mathrm{em}})}\right)} \qquad (\tilde{\lambda}_{\mathrm{em}} > \tilde{\lambda}_{\mathrm{p}}). \tag{8.97}$$

Note that the constraint (8.85) is equivalent to $1 - 3u_{\mathrm{em}} > 0$, so that the above formula is well posed.

For the case $\tilde{\lambda}_{\mathrm{em}} < \tilde{\lambda}_{\mathrm{p}}$, φ_∞ is given by Eq. (8.92) and we need to determine the behaviour of $K(k)$ to conclude. Actually, when $k \to 1$, $K(k)$ is diverging with the following behaviour[9]:

$$\lim_{k\to 1}\left[K(k) - \ln\left(\frac{4}{\sqrt{1-k^2}}\right)\right] = 0. \tag{8.98}$$

To relate $\sqrt{1-k^2}$ to $b - b_{\mathrm{c}}$, let us introduce the small parameter $\varepsilon > 0$ such that

$$u_{\mathrm{p}} =: \frac{1}{3} - \varepsilon. \tag{8.99}$$

The relation $P_b(u_{\mathrm{p}}) = 0$, once expanded to second order in ε^2, leads to the following relation between $b - b_{\mathrm{c}}$ and ε:

$$\frac{b-b_{\mathrm{c}}}{m} = \frac{81\sqrt{3}}{2}\varepsilon^2 + O(\varepsilon^3). \tag{8.100}$$

Besides, from the expressions (8.33) of u_{n} and u_{a} in terms of u_{p}, we get, still to the second order in ε,

$$u_{\mathrm{n}} = -\frac{1}{6} + 2\varepsilon^2 + O(\varepsilon^3) \qquad \text{and} \qquad u_{\mathrm{a}} = \frac{1}{3} + \varepsilon - 2\varepsilon^2 + O(\varepsilon^3).$$

Substituting these values in the expression (8.50) of the modulus k and expanding to the first order in ε leads to

$$k = 1 - 2\varepsilon + O(\varepsilon^2),$$

from which

$$\sqrt{1-k^2} \underset{\varepsilon\to 0}{\sim} \sqrt{1-(1-4\varepsilon)} \underset{\varepsilon\to 0}{\sim} 2\sqrt{\varepsilon}.$$

[9] See e.g. Eq. (112.1) of Ref. [34] with $k' = \sqrt{1-k^2}$ or Eq. (17.3.26) of Ref. [2] with $m = k^2$ and $m_1 = 1 - k^2$.

The property (8.98) then leads to

$$2K(k) \underset{\varepsilon\to 0}{\sim} 2\ln\left(\frac{2}{\sqrt{\varepsilon}}\right) \underset{\varepsilon\to 0}{\sim} \ln\left(\frac{4}{\varepsilon}\right).$$

Using this result, as well as (8.96), in Eq. (8.92) yields

$$\begin{aligned}\varphi_\infty &\underset{b\to b_{\rm c}^+}{\sim} \varphi_{\rm em} + \epsilon_L \times 2\left[\ln\left(\frac{4}{\varepsilon}\right) - \ln\left(\frac{\sqrt{3}+\sqrt{6u_{\rm em}+1}}{\sqrt{2(1-3u_{\rm em})}}\right) - \ln\left(\frac{\sqrt{3}+1}{\sqrt{2}}\right)\right] \\ &\underset{b\to b_{\rm c}^+}{\sim} \varphi_{\rm em} + \epsilon_L \ln\left(\frac{16}{\varepsilon^2}\frac{2(1-3u_{\rm em})}{(\sqrt{3}+\sqrt{6u_{\rm em}+1})^2}\frac{2}{(\sqrt{3}+1)^2}\right).\end{aligned}$$

Note that we have used the limits (8.93) to evaluate the prefactor in Eq. (8.92) as $\sqrt{2/(u_{\rm a}-u_{\rm n})} \sim \sqrt{2/(1/3-(-1/6))} = 2$. Expressing ε^2 in terms of $b-b_{\rm c}$ via Eq. (8.100) leads to the final formula:

$$\boxed{\varphi_\infty \underset{b\to b_{\rm c}^+}{\sim} \varphi_{\rm em} + \epsilon_L \ln\left(\frac{648(2\sqrt{3}-3)(1-3u_{\rm em})}{2+3u_{\rm em}+\sqrt{3(1+6u_{\rm em})}} \times \frac{m}{b-b_{\rm c}}\right)} \qquad (\tilde{\lambda}_{\rm em} < \tilde{\lambda}_{\rm p}). \tag{8.101}$$

The prefactor of $m/(b-b_{\rm c})$ in the logarithm is plotted as a function of $r_{\rm em} = m/u_{\rm em}$ in Fig. 8.22 below.

The value of the total change $\Delta\varphi$ along the complete geodesic history [cf. Eq. (8.77)] is obtained by taking the limit $u_{\rm em} \to 0$ (i.e. $r_{\rm em} \to +\infty$) in this formula. We get

$$\boxed{\Delta\varphi \underset{b\to b_{\rm c}^+}{\sim} \epsilon_L \ln\left(\frac{648(7\sqrt{3}-12)m}{b-b_{\rm c}}\right)}. \tag{8.102}$$

We recover the result (8.84): $\Delta\varphi$ is diverging when $b \to b_{\rm c}^+$. Moreover, Eq. (8.102) specifies this divergence as logarithmic in $b-b_{\rm c}$. More generally, Eq. (8.101) shows that $\varphi_\infty - \varphi_{\rm em}$ diverges logarithmically in $b-b_{\rm c}$ when $b \to b_{\rm c}^+$, whatever the position of the emission point.

Let us express $\Delta\varphi$ is terms of the deflection angle Θ and the winding number n through Eqs. (8.80) and (8.81):

$$\Delta\varphi = \Theta - \epsilon_L\pi + 2\pi n. \tag{8.103}$$

We can then deduce from (8.102) that

$$\boxed{b-b_{\rm c} \underset{b\to b_{\rm c}^+}{\sim} \underbrace{648(7\sqrt{3}-12)\mathrm{e}^{-\pi}}_{\simeq 3.482284}\, m\, \mathrm{e}^{-\epsilon_L\Theta-2\pi|n|}}. \tag{8.104}$$

In the above writing, we have used the fact that the sign of n is the same as that of L, so that $\epsilon_L n = |n|$.

Remark 1: A check of Eq. (8.102) is obtained by comparing it with Eq. (268) in Chap. 3 of Ref. [50], where $\delta D = b-b_{\rm c}$ and Θ is the same as ours, or with Eq. (7.4.54) of Ref. [111], where $\Delta\ell = (b-b_{\rm c})/(2m)$.

Historical note: Equation (8.104) has been first derived by Charles Galton Darwin – the grandson of the famous naturalist Charles Robert Darwin – in 1959 [79] (cf. Eqs. (31) and (32) of Ref. [79], where $\mu = \Theta$).

8.4.2 Asymptotic direction for $b < b_{\rm c}$

When the impact parameter b is lower than the critical value, the null geodesic $\mathscr{L}$ has no periastrion and $\varphi_{\rm em}$ and φ_∞ are given by Eq. (8.68) with $\epsilon_{\rm in} = -1$ (outgoing motion):

$$\varphi_{\rm em} = \varphi_* + \epsilon_L \Phi_b(u_{\rm em}) \qquad \text{and} \qquad \varphi_\infty = \varphi_* + \epsilon_L \Phi_b(0). \tag{8.105}$$

We have then

$$\varphi_\infty = \varphi_{\rm em} + \epsilon_L \left[\Phi_b(0) - \Phi_b(u_{\rm em})\right]. \tag{8.106}$$

$\Phi_b(0)$ is given by Eq. (8.74):

$$\Phi_b(0) = \frac{1}{\sqrt{2(u_* - u_{\rm n})}} \left[K(k) - F\left(\phi_b(0),\, k\right)\right]. \tag{8.107}$$

If $u_{\rm em} < u_*$, $\Phi_b(u_{\rm em})$ is given by Eq. (8.74) as well, so that Eq. (8.106) becomes

$$\boxed{\varphi_\infty = \varphi_{\rm em} + \frac{\epsilon_L}{\sqrt{2(u_* - u_{\rm n})}} \left[F\left(\phi_b(u_{\rm em}),\, k\right) - F\left(\phi_b(0),\, k\right)\right]} \qquad \begin{array}{c}(b < b_{\rm c}) \\ (u_{\rm em} < u_*)\end{array}, \tag{8.108}$$

with $\phi_b(u_{\rm em})$ and $\phi_b(0)$ given by Eq. (8.72):

$$\phi_b(u_{\rm em}) = \arccos\left(\frac{|u_* - u_{\rm em}|}{u_* + u_{\rm em} - 2u_{\rm n}}\right) \qquad \text{and} \qquad \phi_b(0) = \arccos\left(\frac{u_*}{u_* - 2u_{\rm n}}\right). \tag{8.109}$$

If $u_{\rm em} > u_*$, then $\Phi_b(u_{\rm em})$ is given by Eq. (8.76). Combining with Eq. (8.107), we can then write Eq. (8.106) as

$$\boxed{\varphi_\infty = \varphi_{\rm em} + \frac{\epsilon_L}{\sqrt{2(u_* - u_{\rm n})}} \left[2K(k) - F\left(\phi_b(u_{\rm em}),\, k\right) - F\left(\phi_b(0),\, k\right)\right]} \qquad \begin{array}{c}(b < b_{\rm c}) \\ (u_{\rm em} > u_*)\end{array}, \tag{8.110}$$

where $\phi_b(u_{\rm em})$ is still given by Eq. (8.109). $\varphi_\infty - \varphi_{\rm em}$ is plotted as a function of b in Fig. 8.17 for various values of $r_{\rm em} = m/u_{\rm em}$. For $r_{\rm em} \gg m$ (cf. the red curve $r_{\rm em} = 100\,m$ for $b < b_{\rm c}$), we note that $\varphi_\infty - \varphi_{\rm em} \simeq 0$, as expected. As for the case $b > b_{\rm c}$, we also note that $\varphi_\infty - \varphi_{\rm em}$ is diverging when b tends to $b_{\rm c}$. Let us quantify this diverging behaviour:

Limit $b \to b_{\rm c}^-$

When $b \to b_{\rm c}^-$, we have the following limits [cf. Eqs. (8.35) and (8.67)]:

$$\lim_{b \to b_{\rm c}^-} u_{\rm n} = -\frac{1}{6} \qquad \text{and} \qquad \lim_{b \to b_{\rm c}^-} u_* = \frac{1}{3}.$$

Consequently, Eq. (8.109) yields

$$\lim_{b\to b_{\rm c}^-}\phi_b(u_{\rm em})=\arccos\left(\frac{|1-3u_{\rm em}|}{2+3u_{\rm em}}\right)\quad\text{and}\quad\lim_{b\to b_{\rm c}^-}\phi_b(0)=\arccos\left(\frac{1}{2}\right)=\frac{\pi}{3}. \tag{8.111}$$

Moreover, from expression (8.73) for k and the above limits for u_* and $u_{\rm n}$, we have (see also Fig. 8.13)

$$\lim_{b\to b_{\rm c}^-}k=1.$$

Given the values (8.111) and expression (8.95) for the elliptic integral F when $k=1$, we get

$$\lim_{b\to b_{\rm c}^-}F\left(\phi_b(u_{\rm em}),\,k\right)=\ln\left(\frac{2+3u_{\rm em}+\sqrt{3(1+6u_{\rm em})}}{|1-3u_{\rm em}|}\right) \tag{8.112a}$$

$$\lim_{b\to b_{\rm c}^-}F\left(\phi_b(0),\,k\right)=\ln(2+\sqrt{3}). \tag{8.112b}$$

When $u_{\rm em}<u_*$, inserting these formulas into Eq. (8.108) leads to

$$\boxed{\varphi_\infty\underset{b\to b_{\rm c}^-}{\sim}\varphi_{\rm em}+\epsilon_L\ln\left(\frac{2+3u_{\rm em}+\sqrt{3(6u_{\rm em}+1)}}{(2+\sqrt{3})(1-3u_{\rm em})}\right)}\qquad\left(u_{\rm em}<\frac{1}{3}\right). \tag{8.113}$$

We have written $u_{\rm em}<1/3$ because $u_*\to 1/3$ when $b\to b_{\rm c}^-$. For this reason, we also get rid of the absolute value around $1-3u_{\rm em}$. Note that the value of φ_∞ given by Eq. (8.113) is identical to that given by Eq. (8.97), which was obtained for $b\to b_{\rm c}^+$ and an emission point beyond the periastron. This is not surprising if one invokes the continuity around $b=b_{\rm c}$ of null geodesics as regards their part beyond the periastron for $b>b_{\rm c}$ and outside $r=3m$ for $b<b_{\rm c}$. Indeed, the geodesics with $b\to b_{\rm c}^+$ differ significantly from those with $b\to b_{\rm c}^-$ only on parts including the periastron. This continuity appears clearly on Fig. 8.17: the curves with $b<b_{\rm c}$ and $r_{\rm em}>3m$ have a continuous prolongation with the dashed curves, which are the curves with $b>b_{\rm c}$ without any periastrion on the path from the emission point to infinity.

For $u_{\rm em}>u_*$, we shall use formula (8.110), with $K(k)$ having the diverging behaviour (8.98) since $\lim_{b\to b_{\rm c}^-}k=1$. To express $K(k)$ in terms of $|b-b_{\rm c}|$, let us introduce the small parameter $\varepsilon>0$ such that

$$u_{\rm n}=-\frac{1}{6}-\varepsilon. \tag{8.114}$$

The relation $P_b(u_{\rm n})=0$, once expanded to first order in ε, leads then to the following relation between $b_{\rm c}-b$ and ε:

$$\frac{b_{\rm c}-b}{m}=\frac{81\sqrt{3}}{4}\varepsilon. \tag{8.115}$$

Besides, by expanding formula (8.66) at first order in ε, we get

$$u_*=\frac{1}{3}+\varepsilon+O(\varepsilon^2).$$

Substituting this value, as well as (8.114), into expression (8.73) for k yields

$$k = 1 - \frac{\varepsilon}{4} + O(\varepsilon^2),$$

so that

$$\sqrt{1-k^2} \underset{\varepsilon\to 0}{\sim} \sqrt{1-\left(1-\frac{\varepsilon}{2}\right)} \underset{\varepsilon\to 0}{\sim} \sqrt{\frac{\varepsilon}{2}}$$

The property (8.98) then leads to

$$2K(k) \underset{\varepsilon\to 0}{\sim} 2\ln\left(\frac{4\sqrt{2}}{\sqrt{\varepsilon}}\right) \underset{\varepsilon\to 0}{\sim} \ln\left(\frac{32}{\varepsilon}\right).$$

Substituting this expression for $K(k)$, as well as Eq. (8.112) for $F\left(\phi_b(u_{\rm em}),\, k\right)$ and $F\left(\phi_b(0),\, k\right)$, into Eq. (8.110) results in

$$\begin{aligned}\varphi_\infty &\underset{b\to b_{\rm c}^-}{\sim} \varphi_{\rm em} + \frac{\epsilon_L}{\sqrt{2\left(\frac{1}{3}+\frac{1}{6}\right)}}\left[\ln\left(\frac{32}{\varepsilon}\right) - \ln\left(\frac{2+3u_{\rm em}+\sqrt{3(1+6u_{\rm em})}}{3u_{\rm em}-1}\right) - \ln(2+\sqrt{3})\right] \\ &\underset{b\to b_{\rm c}^-}{\sim} \varphi_{\rm em} + \epsilon_L \ln\left(\frac{32}{\varepsilon}\frac{(2-\sqrt{3})(3u_{\rm em}-1)}{2+3u_{\rm em}+\sqrt{3(1+6u_{\rm em})}}\right).\end{aligned}$$

Finally, using (8.115) to let appear $b_{\rm c} - b$ instead of ε, we get

$$\boxed{\varphi_\infty \underset{b\to b_{\rm c}^-}{\sim} \varphi_{\rm em} + \epsilon_L \ln\left(\frac{648(2\sqrt{3}-3)(3u_{\rm em}-1)}{2+3u_{\rm em}+\sqrt{3(1+6u_{\rm em})}} \times \frac{m}{b_{\rm c}-b}\right)} \qquad \left(u_{\rm em} > \frac{1}{3}\right). \quad (8.116)$$

Hence we recover a logarithmic divergence of $\varphi_\infty - \varphi_{\rm em}$ when b tends to $b_{\rm c}$ by lower values.

Remark 2: The prefactor of $m/(b_{\rm c}-b)$ in formula (8.116) is exactly the opposite of the prefactor of $m/(b-b_{\rm c})$ in formula (8.101). It is plotted as a function of $r_{\rm em} = m/u_{\rm em}$ in Fig. 8.22 below.

The upper bound on $u_{\rm em}$ is $1/2$, for this corresponds to a source just outside the black hole event horizon at $r = 2m$. Let us evaluate φ_∞ in this limit; setting $u_{\rm em} \to 1/2$ in Eq. (8.116) results in, after simplification,

$$\varphi_\infty \underset{b\to b_{\rm c}^-}{\sim} \varphi_{\rm em} + \epsilon_L \ln\left(\frac{648(26\sqrt{3}-45)m}{b_{\rm c}-b}\right) \qquad \left(u_{\rm em} \to \frac{1}{2}\right). \quad (8.117)$$

Remark 3: As a check, Eq. (8.117) agrees with Eq. (4) in Ref. [129].

8.5 Images

Being the worldlines of photons, null geodesics are the key ingredient in determining images as seen by some observer of emitting material around a black hole. Computing such images is of great interest, especially after the first image of a black hole vicinity obtained by the Event Horizon Telescope team in 2019 [5, 36]. As stated in Sec. 8.1, we shall differ the discussion of that image to the chapter dealing with rotating black holes (Sec. 12.5.3).

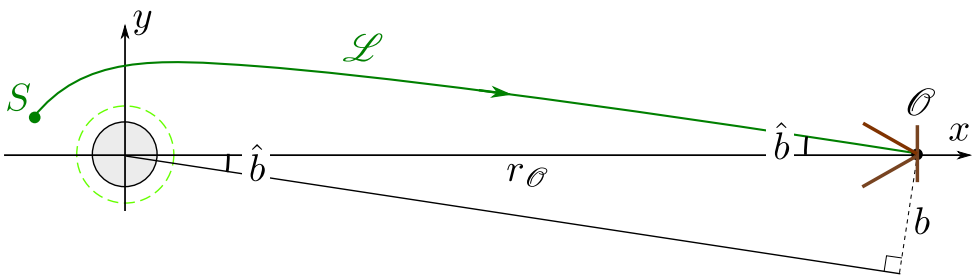


Figure 8.18: Link between the observation angle $\hat{b}$ and the impact parameter b for an asymptotic observer $\mathscr{O}$ located at $r_{\mathscr{O}} \gg m$.

8.5.1 The asymptotic observer

Let us consider some "far-away" observer, i.e. an observer $\mathscr{O}$ located at $r = r_{\mathscr{O}} \gg m$. Without any loss of generality, we may assume that $\mathscr{O}$ is located at $\theta = \pi/2$ and $\varphi = 0$ (cf. Fig. 8.18). Furthermore, we suppose that $\mathscr{O}$ is equipped with an optical device (telescope) pointing in the direction from $\mathscr{O}$ to the black hole, which is the x-axis in terms of the coordinates $(x, y) := (r\cos\varphi, r\sin\varphi)$. Images are formed by null geodesics reaching $\mathscr{O}$'s screen with a angle $\hat{b}$ with respect to the telescope axis (in $\mathscr{O}$'s frame) within the telescope aperture. The angle $\hat{b}$ is actually related to the geodesic impact parameter b by $\sin\hat{b} = b/r_{\mathscr{O}}$ (cf. Fig. 8.18), which, for large $r_{\mathscr{O}}$, can be rewritten as

$$\hat{b} = \frac{b}{r_{\mathscr{O}}}. \tag{8.118}$$

A second parameter characterizing any incoming null geodesic is a polar angle $\alpha \in [0, 2\pi)$ so that $(\hat{b}, \alpha)$ are spherical coordinates on $\mathscr{O}$'s celestial sphere, the North pole ($\hat{b} = 0$) of which coinciding with the direction to the black hole. The image on $\mathscr{O}$'s screen at the telescope output may be then parameterized by polar coordinates $(\hat{b}, \alpha)$, with $\hat{b}$ representing the distance to the screen's center. Other projections from the celestial sphere to the planar screen may be considered but this is rather unimportant here, given the small field of view of the telescope. In what follows, we shall use directly b, instead of $\hat{b}$, as the radial coordinate in the screen plane. According to Eq. (8.118), this amounts to drop the constant scale factor $1/r_{\mathscr{O}}$.

Moreover, we shall consider the asymptotic observer limit: $r_{\mathscr{O}} \to \infty$. In that limit, null geodesics that reach $\mathscr{O}$ must have

$$\varphi_\infty = 0 \mod 2\pi. \tag{8.119}$$

8.5.2 Images of a point source

Qualitative analysis from a concrete example

To understand the formation of images on $\mathscr{O}$'s screen, a bunch of null geodesics with $\varphi_\infty = 0$ is drawn in Fig. 8.19. $\mathscr{O}$ is located in the far right of this figure and all the drawn geodesics eventually hit $\mathscr{O}$'s screen. Let us consider a luminous point source S. From the

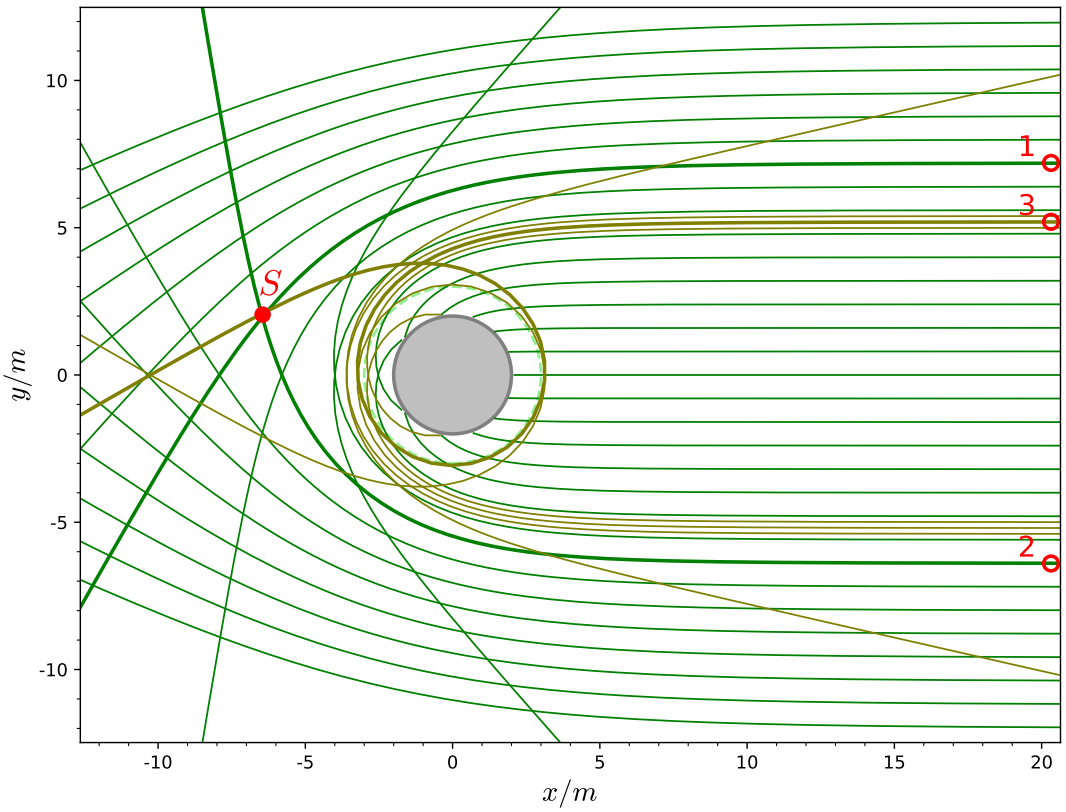


Figure 8.19: Null geodesics in Schwarzschild spacetime with $\varphi_\infty = 0$ and various values of the impact parameter b. The green ones have b ranging from 0 to $12m$, by steps of $0.8m$, while the olive ones have b close to b_c, namely $b/m \in \{5.0, 5.2, 5.4\}$. S is a luminous point source and three of its images for an observer at $y = 0$ and $x \to +\infty$ are obtained by following the drawn null geodesics through S. *[Figure produced with the notebook D.3.13]*

drawing of Fig. 8.19, three null geodesics of the $\varphi_\infty = 0$ family goes through S. They give rise to three distinct images of S on $\mathscr{O}$'s screen, which are represented by the open circles labelled 1 to 3 in the right part of the figure. These three geodesics can be distinguished by their deflection angle Θ and their winding number n (cf. Sec. 8.3.6). The geodesic giving image 1 is the less deflected one: it has $\Theta \simeq -\pi/4$ and $n = 0$, while that of image 2 has $\Theta \simeq 2\pi/3$ and $n = 0$. The geodesic giving image 3 is much more deflected it makes a full turn around the black hole before reaching $\mathscr{O}$; it's winding number is $n = -1$ and it has $\Theta \simeq -\pi/6$. We note that the more deflected the geodesic, the closer the image with respect to the circle $b = b_c$ on $\mathscr{O}$'s screen.

For the sake of clarity, only three images of the source S have been depicted in Fig. 8.19, but we are going to see that there is actually an infinite number of them: two per value of $|n|$ (the number of turns around the black hole) — one with $L > 0$ and one with $L < 0$.

Remark 1: In Fig. 8.19, the source S is rather close to the black hole but it is worth to stress that the multiple character of the images of a given source is not due to the proximity between the source and the black hole. Multiple images are formed for any source, even very far ones. For instance, the source can lie *behind* the observer, i.e. one can have $x_S > x_{\mathscr{O}}$. This is clear if

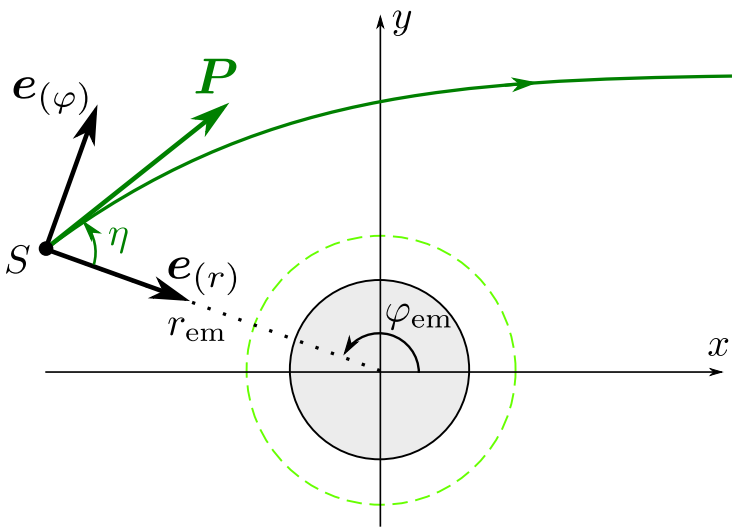


Figure 8.20: Part $(\boldsymbol{e}_{(r)}, \boldsymbol{e}_{(\varphi)})$ of the vector frame of the emitter $\mathscr{O}_{\rm em}$ and emission angle η. The coordinates (x, y) are defined in terms of the Schwarzschild-Droste coordinates by $x = r\cos\varphi$ and $y = r\sin\varphi$. $\boldsymbol{P}$ is the photon momentum as measured by $\mathscr{O}_{\rm em}$.

we consider the lower-right panel of Fig. 8.15 (case $b = 5.355\,m$): a source with $x_S > x_{\mathscr{O}} \gg m$ and $y = 5.355\,m$ gives an image on $\mathscr{O}$'s screen at $b = 5.355\,m$.

Link between the emission angle and the impact parameter

In the above discussion, we have parameterized the null geodesics arriving on the observer's screen by the impact parameter b. Let us relate the latter to the emission angle in the source frame. To this aim, we consider that the point source S, as depicted in Fig. 8.19, is actually the trace in the plane of the figure of the worldline of a static observer equipped with a source of light, who we shall call the *emitter* and denote by $\mathscr{O}_{\rm em}$. Furthermore, we suppose that the rest frame of $\mathscr{O}_{\rm em}$ is an orthonormal tetrad $\left(\boldsymbol{u}_{\rm em}, \boldsymbol{e}_{(r)}, \boldsymbol{e}_{(\theta)}, \boldsymbol{e}_{(\varphi)}\right)$ such that the first spacelike vector, $\boldsymbol{e}_{(r)}$, is always pointing towards the black hole (cf. Fig. 8.20). Note that the first vector of the tetrad is necessarily the 4-velocity $\boldsymbol{u}_{\rm em}$ of $\mathscr{O}_{\rm em}$. More precisely, we set

$$\boldsymbol{u}_{\rm em} = \left(1 - \frac{2m}{r_{\rm em}}\right)^{-1/2} \boldsymbol{\partial}_t \tag{8.120a}$$

$$\boldsymbol{e}_{(r)} = -\left(1 - \frac{2m}{r_{\rm em}}\right)^{1/2} \boldsymbol{\partial}_r \tag{8.120b}$$

$$\boldsymbol{e}_{(\theta)} = -r_{\rm em}^{-1} \boldsymbol{\partial}_\theta \tag{8.120c}$$

$$\boldsymbol{e}_{(\varphi)} = -r_{\rm em}^{-1} \boldsymbol{\partial}_\varphi, \tag{8.120d}$$

where $(\boldsymbol{\partial}_t, \boldsymbol{\partial}_r, \boldsymbol{\partial}_\theta, \boldsymbol{\partial}_\varphi)$ is the natural basis associated with Schwarzschild-Droste coordinates and $r_{\rm em}$ is the (constant) r-coordinate of $\mathscr{O}_{\rm em}$, and hence of $\mathscr{O}_{\rm em}$. It is immediate that $\left(\boldsymbol{u}_{\rm em}, \boldsymbol{e}_{(r)}, \boldsymbol{e}_{(\theta)}, \boldsymbol{e}_{(\varphi)}\right)$ is an orthonormal tetrad with respect to the Schwarzschild met-

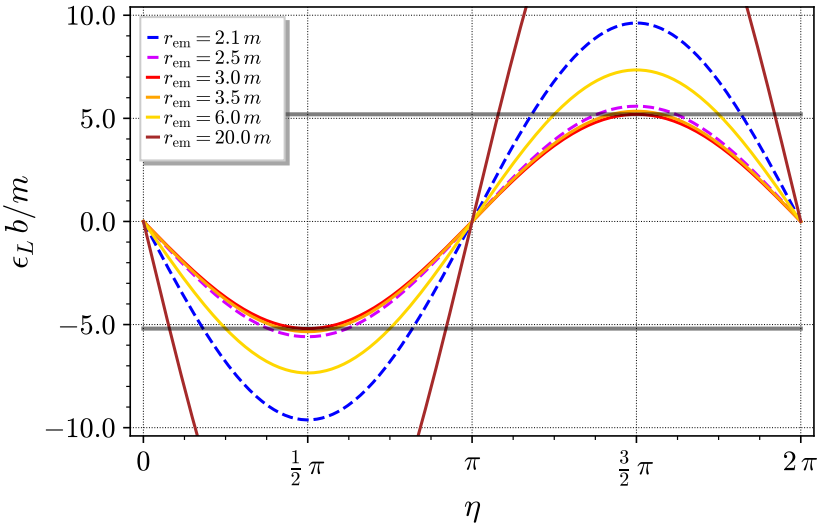


Figure 8.21: Impact parameter b of null geodesics (multiplied by $\epsilon_L = \pm 1$) as a function of the emission angle η in the emitter's rest frame, for various values of the r-coordinate $r_{\rm em}$ of the static emitter. Dashed curves correspond to values of $r_{\rm em}$ lower than $3m$. The two horizontal grey lines marks $b = b_{\rm c} \simeq 5.196\,m$. *[Figure produced with the notebook D.3.14]*

ric (6.15). Moreover, having a 4-velocity $\boldsymbol{u}_{\rm em}$ collinear to the Killing vector $\boldsymbol{\partial}_t$ makes the observer $\mathscr{O}_{\rm em}$ static [cf. Eq. (7.28)].

The 4-momentum $\boldsymbol{p}$ of a photon emitted by $\mathscr{O}_{\rm em}$ in the equatorial plane is given by Eqs. (7.10) and (8.1)-(8.3):

$$\boldsymbol{p} = E\left(1 - \frac{2m}{r}\right)^{-1} \boldsymbol{\partial}_t \pm E\sqrt{1 - b^2U(r)}\,\boldsymbol{\partial}_r + \epsilon_L E \frac{b}{r^2}\,\boldsymbol{\partial}_\varphi, \tag{8.121}$$

where we have used Eq. (8.14) to let appear $U(r)$ and Eqs. (8.12)-(8.13) to express L in terms of E and b. Taking the value of $\boldsymbol{p}$ at the emission point S and using (8.120) to express $\boldsymbol{\partial}_t$, $\boldsymbol{\partial}_r$ and $\boldsymbol{\partial}_\varphi$ in terms of $\mathscr{O}_{\rm em}$'s orthonormal tetrad, we get

$$\boldsymbol{p} = \varepsilon_{\rm em}\left(\boldsymbol{u}_{\rm em} + \boldsymbol{n}\right), \tag{8.122}$$

where

$$\varepsilon_{\rm em} := -\boldsymbol{u}_{\rm em}\cdot\boldsymbol{p} = E\left(1 - \frac{2m}{r_{\rm em}}\right)^{-1/2} \tag{8.123}$$

is the energy of the photon as measured by $\mathscr{O}_{\rm em}$ (cf. Sec. 1.4) and

$$\boldsymbol{n} := \pm\sqrt{1 - b^2U(r_{\rm em})}\,\boldsymbol{e}_{(r)} - \epsilon_L b\sqrt{U(r_{\rm em})}\,\boldsymbol{e}_{(\varphi)} \tag{8.124}$$

is a unit spacelike vector orthogonal to $\boldsymbol{u}_{\rm em}$: $\boldsymbol{n}\cdot\boldsymbol{n} = 1$ and $\boldsymbol{u}_{\rm em}\cdot\boldsymbol{n} = 0$.

In view of (8.122) and $\boldsymbol{u}_{\rm em}\cdot\boldsymbol{n} = 0$, the linear momentum of the photon as measured by $\mathscr{O}_{\rm em}$ is $\boldsymbol{P} = \varepsilon_{\rm em}\boldsymbol{n}$ [cf. Eq. (1.20) and Fig. 8.20]. We conclude that the unit vector $\boldsymbol{n}$

gives the direction of emission of the photon in $\mathscr{O}_{\rm em}$'s rest frame. Let us then define the ***emission angle*** η by

$$\boldsymbol{n} = \cos\eta\, \boldsymbol{e}_{(r)} + \sin\eta\, \boldsymbol{e}_{(\varphi)}. \tag{8.125}$$

Equation (8.124) yields immediately

$$\cos\eta = \pm\sqrt{1 - b^2 U(r_{\rm em})} \qquad \text{and} \qquad \sin\eta = -\epsilon_L b\sqrt{U(r_{\rm em})}, \tag{8.126}$$

from which we can express the impact parameter b and in terms of the emission angle η:

$$\boxed{b = -\epsilon_L \frac{\sin\eta}{\sqrt{U(r_{\rm em})}}}. \tag{8.127}$$

Remark 2: Equation (8.126) is well posed because $U(r) \geq 0$ for $r > 2m$ [cf. Eq. (8.14)] and the equation of motion (8.11) implies $b^{-2} - U(r_{\rm em}) \geq 0$, from which we get $0 \leq b^2 U(r_{\rm em}) \leq 1$.

The variation of b in terms of η and $r_{\rm em}$, as given by formula (8.127), is depicted in Fig. 8.21. We notice that the smallest range of b is $[0, b_{\rm c})$ and is reached for $r_{\rm em} = 3m$, which is not surprising in view of (8.127) since the maximum of U precisely occurs at $r = 3m$ (cf. Fig. 8.1).

Infinite sequence of images

It appears clearly from Fig. 8.21 that, whatever the location $r_{\rm em}$ of the emitter, including locations inside the photon sphere ($r_{\rm em} < 3m$), there always exist four emission angles η giving birth to impact parameters b in the vicinity of the critical value $b_{\rm c}$. These four angles can be gathered in two pairs of angles of opposite directions: $(\eta_1, \eta_1 + \pi)$ and $(\eta_2, \eta_2 + \pi)$. In each pair, only a single emission direction is prior to the periastron for null geodesics with $b > b_{\rm c}$ and only a single emission direction is outgoing and reaches infinity for $b < b_{\rm c}$. In other words, on the four values of η, only two of them define geodesics that may reach the asymptotic observer $\mathscr{O}$.

Let $\varphi_{\rm em} \in [0, 2\pi)$ be the φ coordinate of the emitter $\mathscr{O}_{\rm em}$. A null geodesic emitted by $\mathscr{O}_{\rm em}$ reaches the asymptotic observer $\mathscr{O}$, who is located at $\varphi = 0$, iff $\varphi_\infty = 0 \mod 2\pi$ [Eq. (8.119)], i.e. iff

$$\varphi_\infty = 2\pi\epsilon_L n, \qquad n \in \mathbb{N}. \tag{8.128}$$

The above formula takes into account the fact that the sign of φ_∞ is that of L, hence the factor ϵ_L and $n \geq 0$. From Fig. 8.17, we see that values $n \geq 1$ can be reached iff either (i) $r_{\rm em} > 3m$, b is close to $b_{\rm c}$ from above and $\tilde{\lambda}_{\rm em} < \tilde{\lambda}_{\rm p}$ or (ii) $r_{\rm em} < 3m$ and b is close to $b_{\rm c}$ from below. In the first case, we may use the approximate formula (8.101) with $u_{\rm em} = m/r_{\rm em}$, while in the second case formula (8.116) is relevant. Inserting (8.128) in either of these formulas, we get

$$2\pi\epsilon_L n = \varphi_{\rm em} + \epsilon_L \ln\left(A(r_{\rm em})\frac{m}{b - b_{\rm c}}\right), \tag{8.129}$$

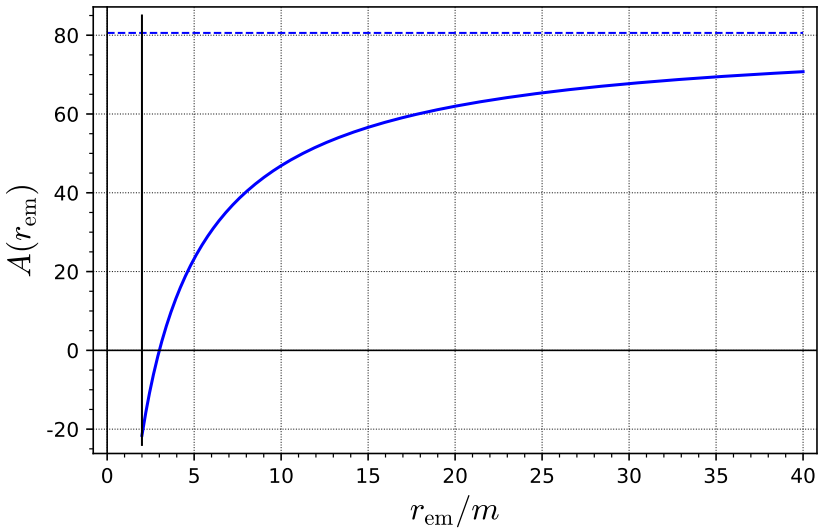


Figure 8.22: Function $A(r_{\rm em})$ defined by Eq. (8.130). The dashed line marks the horizontal asymptote $A = 648(7\sqrt{3} - 12) \simeq 80.58246$. *[Figure produced with the notebook D.3.14]*

with (cf. Remark 2, page 201)

$$\boxed{A(r_{\rm em}) := \frac{648(2\sqrt{3} - 3)(r_{\rm em} - 3m)}{2r_{\rm em} + 3m + \sqrt{3r_{\rm em}(r_{\rm em} + 6m)}}}. \tag{8.130}$$

Note that $A(r_{\rm em}) > 0$ for $r_{\rm em} > 3m$ and $A(r_{\rm em}) < 0$ for $r_{\rm em} < 3m$. Taking into account that $b - b_{\rm c} > 0$ for case (i) and $b - b_{\rm c} < 0$ for case (ii), we check that the argument of the logarithm in Eq. (8.129) is always positive. Equation (8.129) can be solved for b:

$$b = b_{\rm c} + mA(r_{\rm em})\mathrm{e}^{\epsilon_L \varphi_{\rm em} - 2\pi n},$$

which we may write $b = b_n^+$ or b_n^-, denoting the solution with $\epsilon_L = +1$ by b_n^+ and that with $\epsilon_L = -1$ by b_n^-:

$$\boxed{b_n^+ = b_{\rm c} + mA(r_{\rm em})\,\mathrm{e}^{\varphi_{\rm em} - 2\pi n}}, \quad n \in \mathbb{N}^* \tag{8.131a}$$

$$\boxed{b_n^- := b_{\rm c} + mA(r_{\rm em})\,\mathrm{e}^{-\varphi_{\rm em} - 2\pi n}}, \quad n \in \mathbb{N}^*. \tag{8.131b}$$

The function $A(r_{\rm em})$ is plotted in Fig. 8.22. It varies monotonically between

$$\lim_{r_{\rm em}\to 2m} A(r_{\rm em}) = \underbrace{-648(26\sqrt{3} - 45)}_{\simeq -21.59201} \quad \text{and} \quad \lim_{r_{\rm em}\to+\infty} A(r_{\rm em}) = \underbrace{648(7\sqrt{3} - 12)}_{\simeq 80.58246}.$$

One shall recall that formulas (8.131) have been derived under the assumption $|b - b_{\rm c}| \ll m$. Given the amplitude of $A(r_{\rm em})$ shown in Fig. 8.22 and the value $\mathrm{e}^{-2\pi} \simeq 1.9 \times 10^{-3}$, this is valid for any $n \geq 1$, except maybe for b_n^+ with $n = 1$ and $\varphi_{\rm em}$ large (i.e. close to 2π).

Via Eq. (8.127), the two sequences (8.131) of the impact parameter b correspond to two sequences of the emission angle η:

$$\sin\eta_n^+ := -\sqrt{U(r_{\rm em})}\left(b_{\rm c} + mA(r_{\rm em})\mathrm{e}^{\varphi_{\rm em}-2\pi n}\right), \quad n\in\mathbb{N}^* \tag{8.132a}$$

$$\sin\eta_n^- := \sqrt{U(r_{\rm em})}\left(b_{\rm c} + mA(r_{\rm em})\mathrm{e}^{-\varphi_{\rm em}-2\pi n}\right), \quad n\in\mathbb{N}^*. \tag{8.132b}$$

Note that $0 \leq b_{\rm c}\sqrt{U(r_{\rm em})} \leq 1$ and $b_{\rm c}\sqrt{U(r_{\rm em})} = 1 \iff r_{\rm em} = 3m$. As discussed above (cf. Fig. 8.21), for $r_{\rm em} \neq 3m$, there are always some solutions for η_n^+ and η_n^- for n sufficiently large (in practice, $n \geq 1$). We therefore conclude:

Any static point source S give birth to two infinite sequences of images on the screen of the asymptotic observer $\mathscr{O}$, at the effective distances from the screen's center $(b_n^+)_{n\in\mathbb{N}^*}$ and $(b_n^-)_{n\in\mathbb{N}^*}$, as given by Eq. (8.131). If S, $\mathscr{O}$ and the black hole are not aligned, i.e. if $\varphi_{\rm em} \notin \{0,\pi\}$, all the images of S are aligned on $\mathscr{O}$'s screen: they all lie in the plane defined by S, $\mathscr{O}$ and the black hole (plane of Fig. 8.19). If S is located outside the photon sphere ($r_{\rm em} > 3m$), all images lie outside the circle $b = b_{\rm c}$ on $\mathscr{O}$'s screen, while if S is located inside the photon sphere ($r_{\rm em} < 3m$), all images lie inside that circle. In both cases, the images of the sequence $(b_n^+)_{n\in\mathbb{N}^*}$ (positive L) are located on the opposite side of those of the sequence $(b_n^-)_{n\in\mathbb{N}^*}$ (negative L). The higher n, the closer the images b_n^+ and b_n^- to the circle $b = b_{\rm c}$. Actually, according to the laws (8.131), each sequence converge exponentially fast to that circle. From n to $n+1$, the distance to the circle is reduced by a factor $\mathrm{e}^{-2\pi} \simeq 1.9\times10^{-3}$.

Remark 3: The alignment of the images is a direct consequence of the spherical symmetry of Schwarzschild spacetime.

Remark 4: The reader may be puzzled by the compatibility between an infinite number of images from a single source and the conservation of energy. There is actually no issue here because one can show that the images are fainter as n increases. So in practice, only a few images would be visible.

8.5.3 Aligned source and Einstein rings

In the special case where the source is located on the x-axis, the images form a series of concentric circles accumulating from above on the circle $b = b_{\rm c}$. This is illustrated by the source A in Fig. 8.23. Four geodesics through the source are drawn, given birth to four images that are symmetric with respect to the x-axis, labelled 1, 3 and 2, 4 respectively. Since in this case the plane of the figure is not privileged (one cannot speak about the plane defined by A, $\mathscr{O}$ and the black hole, since they are aligned); we deduce by rotation about the x-axis that A generates images all along circles centered on the origin in $\mathscr{O}$'s screen. Images 1 and 2 in Fig. 8.23 lie actually on the same circle, as well as images 3 and 4. As for the non-aligned case, Fig. 8.23 shows only a limited number of images, but there is actually an infinite number of such images: the full image of A on $\mathscr{O}$'s screen is composed by a infinite sequence $(\mathscr{C}_n)_{n\in\mathbb{N}}$ of concentric circles that accumulate exponentially fast onto

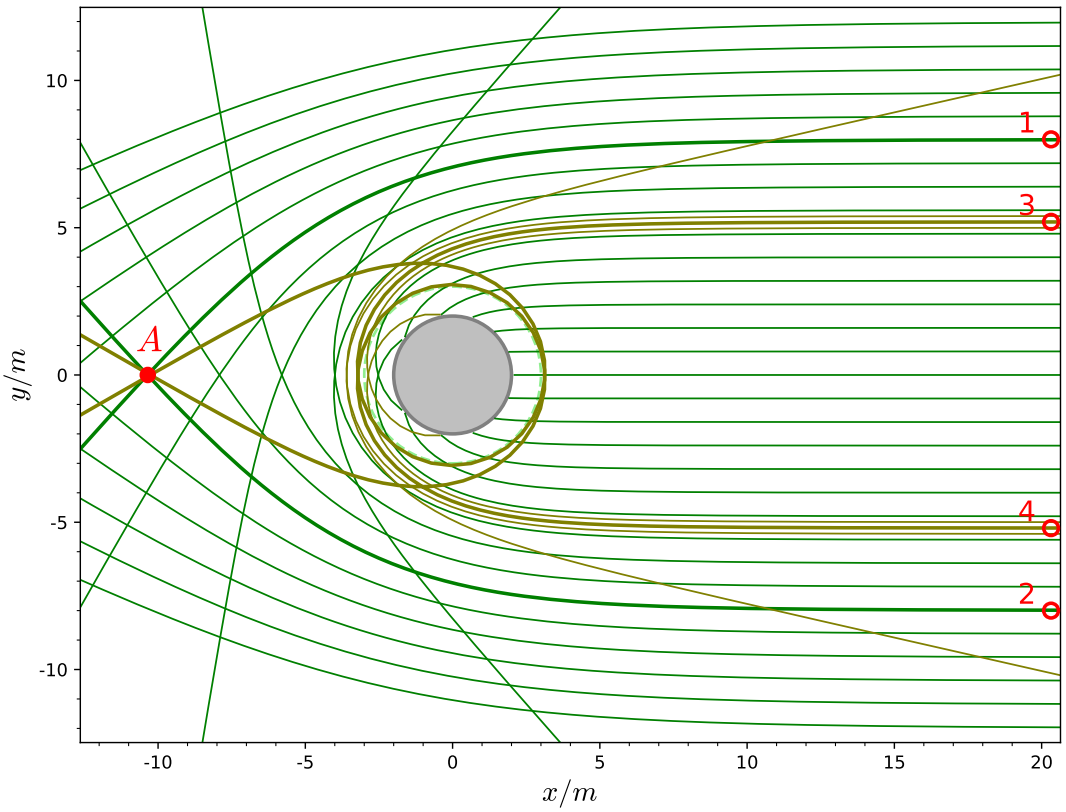


Figure 8.23: Same as Fig. 8.19 but underlining four null geodesics from a point source A located on the x-axis, i.e. aligned with the black hole and the observer on the far right. *[Figure produced with the notebook D.3.13]*

the circle of radius $b = b_{\rm c}$. Each circle $\mathscr{C}_n$ has a radius b_n that is given by Eq. (8.131) with $\varphi_{\rm em} = 0$ or π (case of Fig. 8.23), so that $b_n^+ = b_n^- =: b_n$ (case $\varphi_{\rm em} = 0$) or $b_n^+ = b_{n-1}^- =: b_n$ (case $\varphi_{\rm em} = \pi$). The outermost circle, $\mathscr{C}_0$, is called the ***Einstein ring of*** A. It is the only significant ring in astronomical images involving *gravitational lensing* by a massive foreground source (not necessarily a black hole), which are in the context $|x_A| \gg m$ and $b \gg m$. We may call $\mathscr{C}_n$ the ***n*$^{\rm th}$ *Einstein ring of*** A. For $n \geq 1$, $\mathscr{C}_n$ is also called a ***relativistic Einstein ring of*** A. Indeed, these rings exist only in the case of a very relativistic central object, so that photons can wind around it. On the contrary $\mathscr{C}_0$ exists even for non-relativistic objects. In particular, in astronomical observations of Einstein rings performed up to now, the central deflecting object is a galaxy or a galaxy cluster, both being highly non-relativistic (very low compactness).

Historical note: The infinite sequence of images of a point source located at $r > 3m$, with accumulation just outside the circle $b = b_{\rm c}$, has been predicted in 1959 by Charles Galton Darwin [79] (cf. historical note on page 199). Darwin has also derived the sequence of relativistic Einstein rings for aligned sources.

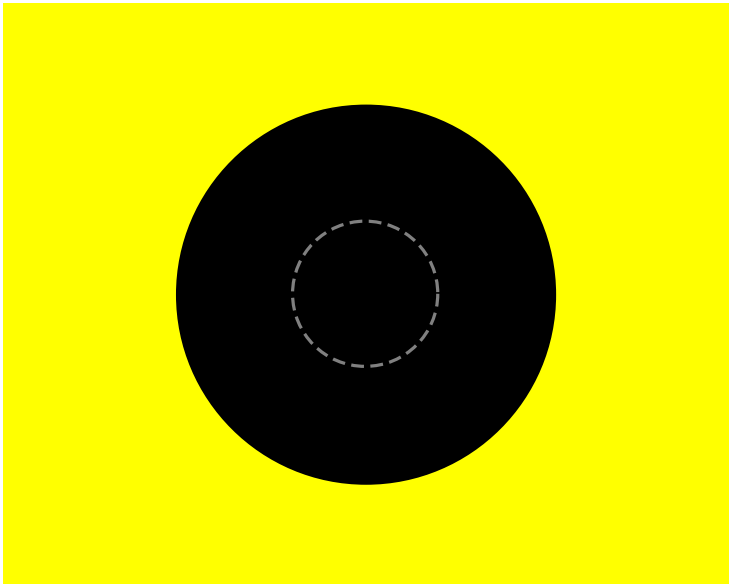


Figure 8.24: Shadow of a Schwarzschild black hole. The dashed grey circle delimitates the dark disk that would appear on the screen of an observer looking at a black sphere of radius $r = 2m$ in Minkowski spacetime. The ratio between the radii of the two disks is $b_c/2m = 3\sqrt{3}/2 \simeq 2.60$.

8.5.4 Black hole shadow

Let us determine the image on observer $\mathscr{O}$'s screen when all the sources of light are very far, both from the black hole and from $\mathscr{O}$. We may think of many stars shining on the celestial sphere. To simplify the problem, we shall assume that the black hole and the observer are surrounded by a distant sphere $\mathscr{S}$ that is uniformly bright. In such a setting, all values of $b > b_c$ on $\mathscr{O}$'s screen, whatever the polar angle α, correspond to a null geodesic that originates from the shining sphere $\mathscr{S}$, while this is not possible for $b \leq b_c$. This is clear on Fig. 8.19: when traced backward from the right ($\mathscr{O}$ position), (i) null geodesics with $b > b_c$ eventually end up far away from the black hole, i.e. necessarily on $\mathscr{S}$, (ii) null geodesics with $b < b_c$ end up infinitely close to the black hole horizon and (iii) null geodesics with $b = b_c$ roll up indefinitely around the photon sphere (cf. Figs. 8.7 and 8.8).

The same conclusion can be reached by considering the effective potential diagram in Fig. 8.1. For concreteness, imagine that $\mathscr{O}$ is located at $r = 15m$ and that the shining sphere $\mathscr{S}$ has a coordinate radius $r = 20m$. These values allow us to put $\mathscr{O}$ and $\mathscr{S}$ in Fig. 8.1 according to our settings ($\mathscr{S}$ is encompassing both $\mathscr{O}$ and the black hole), but one shall keep in mind that both $\mathscr{O}$ and $\mathscr{S}$ are assumed to lie at much larger values of r (fulfilling $r \gg m$). Only two kinds of geodesic trajectories in Fig. 8.1 diagram can reach $r = 15m$ ($\mathscr{O}$'s screen) with increasing values of r (the direction of observation): those of type 1 and those of type 3. The latter ones originate necessarily from a region between the black hole horizon and $\mathscr{O}$'s location, i.e. a region that does not intersect $\mathscr{S}$. On the contrary, all the geodesics of type 1 must have had $r = 20m$ in their past history, i.e. we may consider that they all have been emitted by $\mathscr{S}$ inwards and have bounced on the potential barrier before reaching $\mathscr{O}$'s screen in their outward motion.

Both reasonings, that based on Fig. 8.19 and that based on Fig. 8.1, led us to conclude that

> For a distant observer, the image of a uniformly bright sphere $\mathscr{S}$ surrounding the black hole and the observer is a bright area fulfilling the observer's screen, except for a central black disk of radius $b = b_c$ (cf. Fig. 8.24). This disk is called the ***black hole shadow***.

Historical note: In his 1921 treatise [174], Max von Laue noted that an emitting sphere of radius r_0 such that $2m < r_0 < 3m$ will appear as having a radius of $3\sqrt{3}m$ ($= b_c$) to a distant observer, independently of the value of r_0. Maybe the first mention of a shadow lies in the article by Charles Galton Darwin in 1959 [79]. Actually, Darwin imagined that the source of the Schwarzschild metric is a massive point particle, called by him the "sun", and wrote *"suppose that there is a roughly uniform star-field all round the sky, and consider what will be seen and mapped from the telescope... To the area inside this circle* (the circle $b = b_c$) *only the sun itself can contribute light. The most obvious assumption is that its rays would emerge in straight lines, so that there would be a brilliant point of light surrounded by blackness"*. In the full black hole context, the first computation of the shadow has been presented by James M. Bardeen at the famous Les Houches summer school of 1972 [17]; the shadow is called the *apparent shape of the black hole* by Bardeen and the computation has been performed for the Kerr black hole, which generalizes the Schwarzschild black hole to the rotating case (cf. Chap. 10); this Kerr black hole shadow will be discussed in details in Sec. 12.4.

To go further, see the review article [71].

8.5.5 Image of an accretion disk

A realistic "source of light" in the vicinity of a black hole is an accretion disk [1]. Actually most of astronomical observations of black holes in the electromagnetic domain are measurements about an accretion disk, either a spectrum or an image, like the image of M87* released in 2019 by the Event Horizon Telescope collaboration [5]. The matter (mostly hydrogen) constituting the accretion disk arises either from a companion star (stellar-mass black hole in a binary system, like Cyg X-1, cf. Table 7.1) or from gas clouds in a galactic center (supermassive black holes, like Sgr A* and M87*, cf. Table 7.1).

Figure 8.25 shows various views of an accretion disk around a Schwarzschild black hole as seen by a distant observer. The accretion disk is a model developed by Novikov & Thorne [198] and Page & Thorne [202], lying in some plane around the black hole. It is geometrically thin but optically thick. The inner radius of the disk is the innermost stable circular orbit (ISCO) at $r = 6m$ (cf. Sec. 7.3.3). The images shown in Fig. 8.25 have been computed by the open-source ray-tracing code Gyoto [253] (cf. Appendix E). This code integrates the null geodesic equations backward, starting from the observer's screen. For Fig. 8.25, the observer is located at $r = 1000\,m$ and at various inclination angles ι with respect to the accretion disk, ranging from $\iota = 0$ (disk seen face-on) to $\iota = 3/2$ (disk seen almost edge-on). The colors in Fig. 8.25 encode the flux at a fixed wavelength.

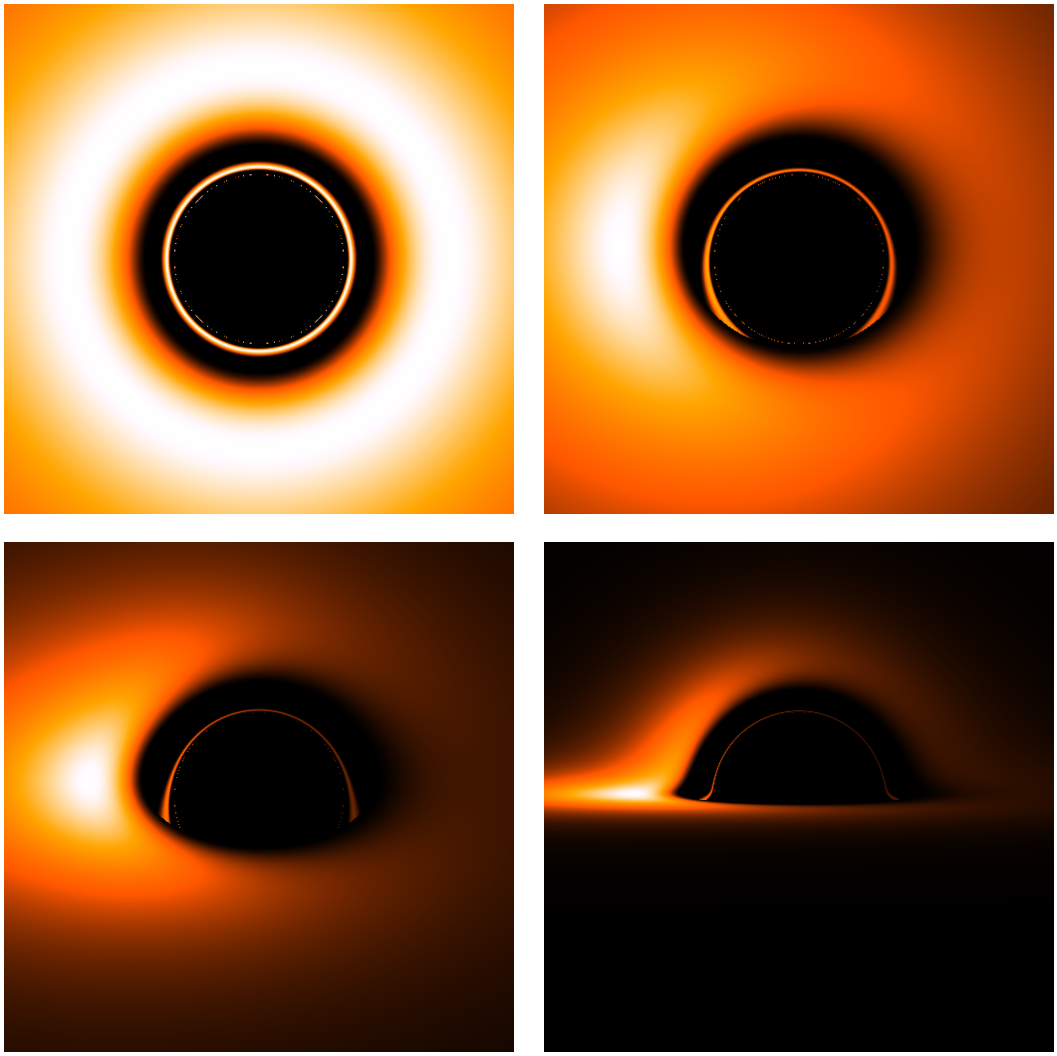


Figure 8.25: Image of an accretion disk around a Schwarzschild black hole, as seen by a distant observer, with various inclination angles: $\iota = 0$ (upper left), $\iota = \pi/6$ (upper right), $\iota = \pi/3$ (lower left) and $\iota = 3/2$ (lower right). *[Figure produced by Gyoto with the input files given in Sec. E.2.1]*

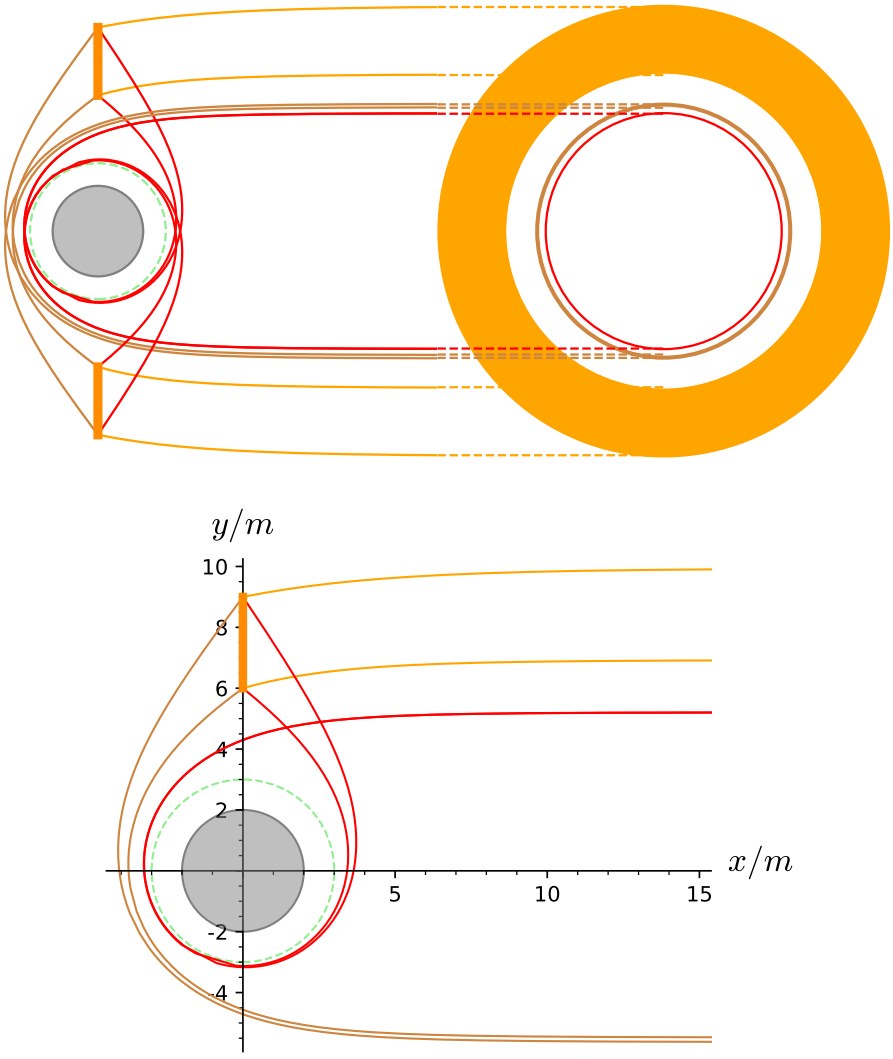


Figure 8.26: Formation of the primary (orange), secondary (brown) and tertiary (red) images of an accretion disk when viewed face-on. The figure plane is orthogonal to the disk plane, the intersection between the two planes being the y-axis, so that disk is depicted by the thick segments located at $\varphi = \pi/2$ and $\varphi = 3\pi/2$. As in Fig. 8.25, the inner boundary of the disk is $r_{\rm in} = 6\,m$ (the ISCO), but the outer boundary has been truncated at $r_{\rm out} = 9\,m$ for pedagogical purposes. The observer is located at $x \gg m$ and $y = 0$. The dashed green circle is the photon orbit at $r = 3\,m$. The lower panel shows the geodesics emerging from the inner and outer edges of the upper part of the disk. *[Figure produced with the notebook D.3.15]*

Let us first discuss the face-on view (upper-left panel of Fig. 8.25). To interpret it, some null geodesics generating it have been depicted in Fig. 8.26. One may distinguish three images of the accretion disk:

- the *primary image*, which is formed by the outermost bright part; it is generated by null geodesics that suffers a small deflection from the disk to the observer; more precisely, they are arising from the side of the disk facing the observer and have $\Delta\varphi = \pm\pi/2$ in their trajectory plane (cf. orange curves in Fig. 8.26); the inner boundary of the primary image is the image of the ISCO under these geodesics; it is located at $b = 6.932\,m$ on the observer's screen;

- the *secondary image*, which is the bright yellow ring separated from the inner boundary of the primary image by the thick black annulus; this image is generated by null geodesics that perform half a turn around the photon sphere before reaching the observer; more precisely, they are arising from the side of the disk *opposite* to the observer and have $\Delta\varphi = \pm 3\pi/2$ in their trajectory plane (cf. brown curves in Fig. 8.26); the inner boundary of the secondary image is located at $b = 5.479\,m$ on the observer's screen;

- the *tertiary image*, which is the thin faint innermost ring; this image is generated by null geodesics that perform a full turn around the photon sphere before reaching the observer; more precisely, they are arising from the side of the disk facing the observer and have $\Delta\varphi = \pm 5\pi/2$ in their trajectory plane (cf. red curves in Fig. 8.26); the inner boundary of the tertiary image is located at $b = 5.208\,m$ on the observer's screen.

As discussed in Sec. 8.5.2, there are actually an infinite number of images of the accretion disk but they are more and more faint and only three of them are visible in Fig. 8.25. Moreover these images are formed by exponentially thinner rings and accumulate near the circle $b = b_{\rm c}$. In Figs. 8.25 and 8.26, the tertiary image corresponds to $n = 2$ for $\epsilon_L = 1$ and $n = 1$ for $\epsilon_L = -1$ in Eq. (8.128) [10], so that Eqs. (8.131) with $\varphi_{\rm em} = \pi/2$ ($b = b_1^-$) and $\varphi_{\rm em} = 3\pi/2$ ($b = b_2^+$) yield $(b - b_{\rm c})/b_{\rm c} = A(r_{\rm em}){\rm e}^{-5\pi/2}/(3\sqrt{3})$. With $r_{\rm em} = 6m$, we have $A(r_{\rm em}) \simeq 30.3$, so that $(b - b_{\rm c})/b_{\rm c} \simeq 0.0013$ for the tertiary image. This we may conclude that, up to a relative accuracy of 10^{-3}, the interior of the tertiary image corresponds to the black hole shadow as defined in Sec. 8.5.4,

For a nonzero inclination angle, the images shown in Fig. 8.25 exhibit some asymmetry between the left and the right. This is due to the rotation of the disk at relativistic speed, which generates a strong Doppler boost, making the left part of the disk, which moves towards the observer, much brighter than the right part, which recedes from the observer. Note that for the three views with $\iota \neq 0$, the primary, secondary and tertiary images are still present but are no longer circular rings. Note that the almost edge-on view is drastically different from what would get in flat (Minkowski) spacetime, since for the latter the accretion disk would appear as a very thin ellipsoidal shape elongated along the horizontal axis.

[10] This asymmetry of n with respect to ϵ_L is due to the convention $\varphi_{\rm em} \in [0, 2\pi)$; had we chosen $\varphi_{\rm em} \in (-\pi, \pi]$, it would not be present.

Historical note: The first computation of the image of an accretion disk around a Schwarzschild black hole has been performed by Jean-Pierre Luminet in 1979 [179], (see Ref. [180] for an historical account). The accretion disk was the same Page-Thorne model as that considered here and Luminet obtained an image pretty close to that of the last quadrant of Fig 8.25, the difference lying in the inclination angle: $\iota = 80^\circ \simeq 1.40$ rad versus $\iota = 1.50$ rad in Fig 8.25. In 1991, Jean-Alain Marck computed the first movie of an observer plunging into a Schwarzschild black endowed with an accretion disk [182, 183]. In 2007, Alain Riazuelo computed very precise images of a Schwarzschild black hole in front of a realistic stellar field [220, 221], which illustrate magnificently the multiple character of images of stars (Sec. 8.5.2) and the concept of black hole shadow (Sec. 8.5.4) (cf. Appendix F).

Chapter 9

Maximal extension of Schwarzschild spacetime

Contents

9.1 Introduction

In the preceding chapters, the Schwarzschild spacetime was considered to be $(\mathscr{M}_{\rm IEF}, \boldsymbol{g})$, where the manifold $\mathscr{M}_{\rm IEF}$ is covered by the ingoing Eddington-Finkelstein coordinates. It turns out that this spacetime can be smoothly (actually analytically) extended to a larger spacetime, $(\mathscr{M}, \boldsymbol{g})$ say, which on its turn cannot be extended, i.e. it is maximal. To construct $(\mathscr{M}, \boldsymbol{g})$, we first introduce the so-called *Kruskal-Szekeres coordinates* on $\mathscr{M}_{\rm IEF}$ in Sec. 9.2. These coordinates restore some symmetry between the ingoing radial null geodesics and the outgoing ones: both types of geodesics appear as straight lines in a spacetime diagram built on Kruskal-Szekeres coordinates, with a slope $+1$ for the outgoing family and -1 for the ingoing one. Then it appears clearly that the outgoing radial null geodesics are artificially halted[1] at some past boundary of $\mathscr{M}_{\rm IEF}$, while no curvature singularity is located there. This calls for an extension of the spacetime, which is performed in Sec. 9.3, thanks to Kruskal-Szekeres coordinates. This extension is maximal; in particular, all incomplete geodesics are so because they either terminate to or emanate from a curvature singularity. In Sec. 9.4, we construct the Carter-Penrose diagram of the

[1] Technically, one says that they are *incomplete* geodesics, cf. Sec. B.3.2.

maximal Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$, via two compactified versions of the Kruskal-Szekeres coordinates. We show that the simplest one, which is standard in the litterature, does not lead to a regular conformal completion of $(\mathscr{M}, \boldsymbol{g})$, as defined in Sec. 4.3. The second version, built on Frolov-Novikov compactified coordinates, achieves this goal. We use this completion to show explicitly that the maximal spacetime contains a white hole, in addition to the black one. In Sec. 9.5, we investigate the hypersurfaces of constant Kruskal-Szekeres time, which connect two asymptotically flat regions of $(\mathscr{M}, \boldsymbol{g})$ through the co-called *Einstein-Rosen bridge*. Finally, Sec. 9.6 discusses the physical relevance of the maximally extended Schwarzschild spacetime.

9.2 Kruskal-Szekeres coordinates

9.2.1 Definition

On the open set $\mathscr{M}_{\rm I}$, let us consider the "double-null" coordinate system $x^{\hat{\alpha}} = (u, v, \theta, \varphi)$. It is related to Schwarzschild-Droste coordinates (t, r, θ, φ) by Eqs. (6.22)-(6.23):

$$\left\{ \begin{array}{l} u = t - r - 2m \ln \left| \frac{r}{2m} - 1 \right| \\ \\ v = t + r + 2m \ln \left| \frac{r}{2m} - 1 \right| \end{array} \right. \iff \left\{ \begin{array}{l} t = \frac{1}{2}(u+v) \\ \\ r + 2m \ln \left| \frac{r}{2m} - 1 \right| = \frac{1}{2}(v-u). \end{array} \right. \tag{9.1}$$

Despite one cannot express explicitly r in terms of (u, v), the function $r \mapsto r + 2m \ln \left| \frac{r}{2m} - 1 \right|$ is invertible on $(2m, +\infty)$ (cf. Fig. 9.1), so that (9.1) does define a coordinate system on $\mathscr{M}_{\rm I}$. The range of (u, v) is $\mathbb{R}^2$.

The above relations imply

$$\mathrm{d}u = \mathrm{d}t - \frac{\mathrm{d}r}{1 - \frac{2m}{r}} \qquad \text{and} \qquad \mathrm{d}v = \mathrm{d}t + \frac{\mathrm{d}r}{1 - \frac{2m}{r}}.$$

Hence

$$\mathrm{d}u\, \mathrm{d}v = \mathrm{d}t^2 - \frac{\mathrm{d}r^2}{\left(1 - \frac{2m}{r}\right)^2}.$$

The line element (6.15) becomes then

$$\boxed{g_{\hat{\mu}\hat{\nu}}\, \mathrm{d}x^{\hat{\mu}}\, \mathrm{d}x^{\hat{\nu}} = -\left(1 - \frac{2m}{r}\right) \mathrm{d}u\, \mathrm{d}v + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)}. \tag{9.2}$$

In this formula, r is to be considered as a function of (u, v), given by (9.1).

The metric components (9.2) are regular on $\mathscr{M}_{\rm I}$. Having a look at Fig. 9.1, we realize that we cannot extend this coordinate system to include the Schwarzschild horizon $\mathscr{H}$, since $r \to 2m$ is equivalent to $v - u \to -\infty$: if u (resp. v) were taking a finite value on $\mathscr{H}$, we would have $v \to -\infty$ (resp. $u \to +\infty$). This impossibility of extending to $\mathscr{H}$ is also reflected by the fact that

$$\det\left(g_{\hat{\alpha}\hat{\beta}}\right) = -\frac{1}{4}\left(1 - \frac{2m}{r}\right)^2 r^4 \sin^2\theta$$

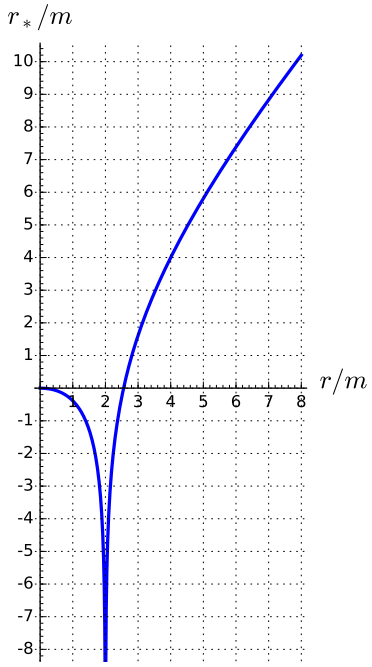


Figure 9.1: Function $r_*(r) = r + 2m \ln\left|\frac{r}{2m} - 1\right|$ (the tortoise coordinate, cf. Eq. (6.24)). It relates r to (u, v) via $r_*(r) = (u - v)/2$ [Eq. (9.1)].

vanishes for $r \to 2m$, which would make $\boldsymbol{g}$ a degenerate bilinear form at $r = 2m$, while it is not of course.

Instead of (u, v), let us use on $\mathscr{M}_{\mathrm{I}}$ the coordinates (U, V) defined by

$$\begin{cases} U := -\mathrm{e}^{-u/4m} \\ V := \mathrm{e}^{v/4m}. \end{cases} \tag{9.3}$$

Since the range of (u, v) is $\mathbb{R}^2$, the range of U is $(-\infty, 0)$ and that of V is $(0, +\infty)$. We have

$$\mathrm{d}U = \frac{1}{4m}\,\mathrm{e}^{-u/4m}\,\mathrm{d}u \qquad \text{and} \qquad \mathrm{d}V = \frac{1}{4m}\,\mathrm{e}^{v/4m}\,\mathrm{d}v,$$

hence

$$\mathrm{d}u\,\mathrm{d}v = 16m^2 \mathrm{e}^{(u-v)/4m}\,\mathrm{d}U\,\mathrm{d}V.$$

Now, on $\mathscr{M}_{\mathrm{I}}$, $r > 2m$ and (9.1) yields

$$r + 2m \ln\left(\frac{r}{2m} - 1\right) = \frac{1}{2}(v - u) \quad \Longrightarrow \quad \mathrm{e}^{r/2m}\left(\frac{r}{2m} - 1\right) = \mathrm{e}^{(v-u)/4m}, \tag{9.4}$$

so that

$$\mathrm{d}u\,\mathrm{d}v = 16m^2\,\mathrm{e}^{-r/2m}\left(\frac{r}{2m} - 1\right)^{-1}\mathrm{d}U\,\mathrm{d}V = \frac{32m^3}{r}\,\mathrm{e}^{-r/2m}\left(1 - \frac{2m}{r}\right)^{-1}\mathrm{d}U\,\mathrm{d}V.$$

Substituting this expression in (9.2) yields the expression of the metric components with respect to coordinates $X^{\hat{\alpha}} := (U, V, \theta, \varphi)$:

$$\boxed{g_{\hat{\mu}\hat{\nu}}\, \mathrm{d}X^{\hat{\mu}}\, \mathrm{d}X^{\hat{\nu}} = -\frac{32m^3}{r}\, \mathrm{e}^{-r/2m}\, \mathrm{d}U\, \mathrm{d}V + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)}. \tag{9.5}$$

In this formula, r has to be considered as a function of (U, V), whose implicit expression is found by combining (9.3) and (9.4):

$$\boxed{\mathrm{e}^{r/2m} \left(\frac{r}{2m} - 1\right) = -UV}. \tag{9.6}$$

Remark 1: This relation takes a very simple form in terms of the tortoise coordinate (cf. Eq. (6.24)):

$$\mathrm{e}^{r_*/2m} = -UV. \tag{9.7}$$

We notice that the factor $(1 - 2m/r)$ has disappeared in the line element (9.5), which becomes perfectly regular as $r \to 2m$.

We read on (9.5) that $g_{UU} = 0$ and $g_{VV} = 0$. Hence (U, V) is a double-null coordinate system, as much as (u, v). To cope with a timelike-spacelike coordinate system instead, let us introduce on $\mathscr{M}_{\mathrm{I}}$ the pair (T, X) such that U is T retarded by X and V is T advanced by X:

$$\left\{ \begin{array}{l} U = T - X \\ V = T + X \end{array} \right. \iff \left\{ \begin{array}{l} T = \frac{1}{2}(U + V) \\ X = \frac{1}{2}(V - U) \end{array} \right. \tag{9.8}$$

Since the range of U on $\mathscr{M}_{\mathrm{I}}$ is $(-\infty, 0)$ and that of V is $(0, +\infty)$, the range of (T, X) is ruled by $T < X$, $T > -X$ and $X > 0$. In other words, the coordinates (T, X) span the following quarter of $\mathbb{R}^2$ (cf. Fig. 9.2):

$$\mathscr{M}_{\mathrm{I}} : \quad X > 0 \quad \text{and} \quad -X < T < X. \tag{9.9}$$

The coordinates $X^{\alpha} := (T, X, \theta, \varphi)$ are called the ***Kruskal-Szekeres coordinates***.

We have $\mathrm{d}U\, \mathrm{d}V = (\mathrm{d}T - \mathrm{d}X)(\mathrm{d}T + \mathrm{d}X) = \mathrm{d}T^2 - \mathrm{d}X^2$, so that the metric components with respect to the Kruskal-Szekeres coordinates are easily deduced from the line element (9.5):

$$\boxed{g_{\mu\nu}\, \mathrm{d}X^{\mu}\, \mathrm{d}X^{\nu} = \frac{32m^3}{r}\, \mathrm{e}^{-r/2m} \left(-\mathrm{d}T^2 + \mathrm{d}X^2\right) + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)}. \tag{9.10}$$

Here r is to be considered as a function of (T, X), which is implicitly defined by

$$\boxed{\mathrm{e}^{r/2m} \left(\frac{r}{2m} - 1\right) = X^2 - T^2}. \tag{9.11}$$

This relation is a direct consequence of (9.6) and (9.8). We may rewrite it as

$$F\left(\frac{r}{2m}\right) = X^2 - T^2, \tag{9.12}$$

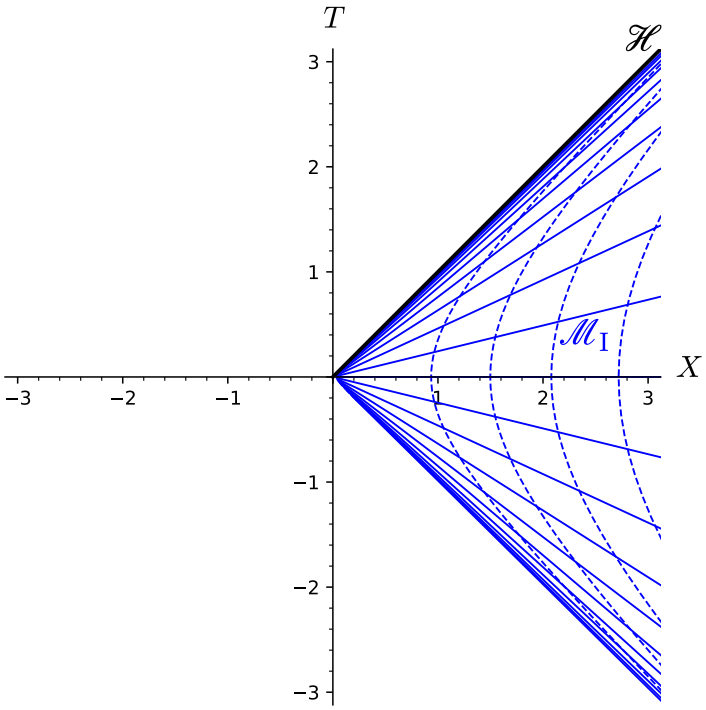


Figure 9.2: Submanifold $\mathscr{M}_{\mathrm{I}}$ in the Kruskal-Szekeres coordinates (T, X): $\mathscr{M}_{\mathrm{I}}$ is covered by the Schwarzschild-Droste grid (in blue): the solid lines have $t = \text{const}$ (spaced apart by $\delta t = m$), while the dashed curves have $r = \text{const}$ (spaced apart by $\delta r = m/2$). *[Figure generated by the notebook D.3.16]*

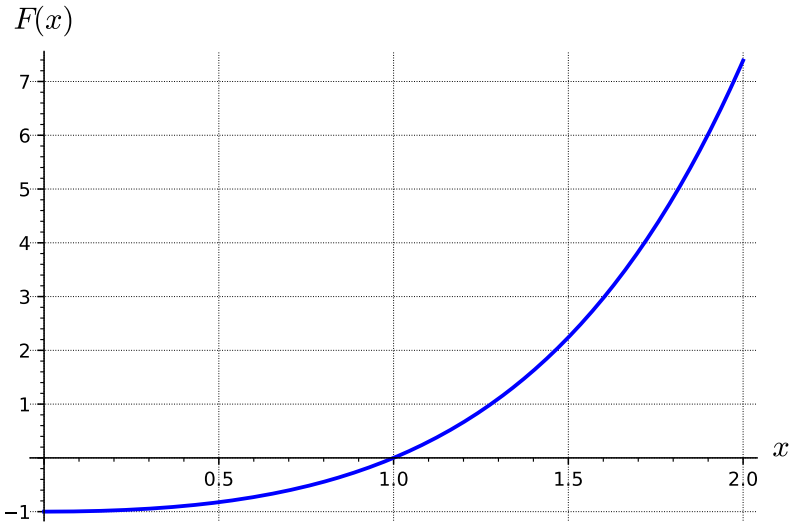


Figure 9.3: Function $F : x \mapsto \mathrm{e}^x(x-1)$, yielding $X^2 - T^2 = F(r/2m)$, cf. Eq. (9.11).

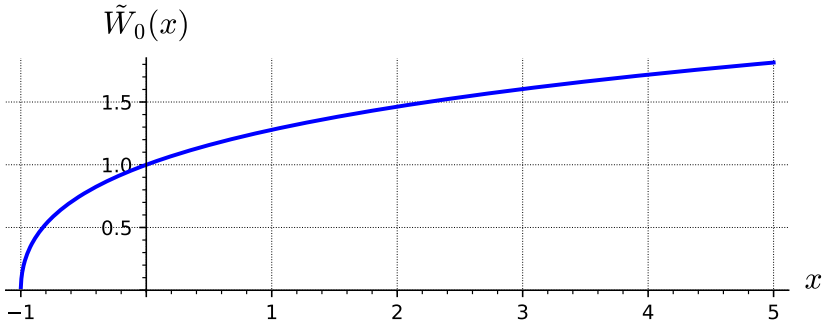


Figure 9.4: Rescaled Lambert function $\tilde{W}_0$, defined by (9.16) and obeying $\mathrm{e}^{\tilde{W}_0(x)}(\tilde{W}_0(x)-1)=x$.

where F is the function defined by

$$\begin{array}{rccc} F: & (0,+\infty) & \longrightarrow & (-1,+\infty) \\ & x & \longmapsto & \mathrm{e}^x(x-1). \end{array} \tag{9.13}$$

The graph of F is shown in Fig. 9.3. We see clearly that F is a bijective map. In particular, F induces a bijection between $(1,+\infty)$ (the range of $r/2m$ on $\mathscr{M}_{\mathrm{I}}$) and $(0,+\infty)$ (the range of X^2-T^2 on $\mathscr{M}_{\mathrm{I}}$, according to (9.9)). The inverse of F can be expressed in terms of the ***Lambert function*** W_0, which is defined as the inverse of $x\mapsto x\mathrm{e}^x$:

$$\begin{array}{rccc} W_0: & (-1/\mathrm{e},+\infty) & \longrightarrow & (-1,+\infty) \\ & x & \longmapsto & y \text{ such that } y\mathrm{e}^y=x. \end{array} \tag{9.14}$$

Noticing that

$$F(x)=\mathrm{e}^x(x-1)=\mathrm{e}\times(x-1)\mathrm{e}^{x-1}$$

we may write

$$F^{-1}=\tilde{W}_0, \tag{9.15}$$

where $\tilde{W}_0$ is the rescaled Lambert function defined by

$$\boxed{\tilde{W}_0(x):=W_0\left(\frac{x}{\mathrm{e}}\right)+1}. \tag{9.16}$$

Note that $\tilde{W}_0$ is a bijection $(-1,+\infty)\to(0,+\infty)$, which obeys

$$\mathrm{e}^{\tilde{W}_0(x)}\left(\tilde{W}_0(x)-1\right)=x. \tag{9.17}$$

Its graph is shown in Fig. 9.4.

Using $F^{-1}=\tilde{W}_0$, we may invert the relation (9.11) to $r=2m\tilde{W}_0(X^2-T^2)$. Noticing that $2m/r\,\mathrm{e}^{-r/2m}=(X^2-T^2+\mathrm{e}^{r/2m})^{-1}$ [cf. Eq. (9.11)], we may eliminate r from the

expression (9.10) of the metric components in Kruskal-Szekeres coordinates:

$$\boxed{\begin{aligned} g_{\mu\nu}\,\mathrm{d}X^\mu\,\mathrm{d}X^\nu &= 4m^2\Big\{ \frac{4}{X^2-T^2+\mathrm{e}^{\tilde{W}_0(X^2-T^2)}}\left(-\mathrm{d}T^2+\mathrm{d}X^2\right) \\ &\qquad + \tilde{W}_0(X^2-T^2)^2\left(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\varphi^2\right)\Big\}. \end{aligned}} \tag{9.18}$$

The relation between the Kruskal-Szekeres coordinates and the Schwarzschild-Droste ones is obtained by combining (9.8), (9.3) and (9.1):

$$\begin{aligned} T &= \frac{1}{2}(U+V) = \frac{1}{2}\left(\mathrm{e}^{v/4m}-\mathrm{e}^{-u/4m)}\right) = \frac{1}{2}\left(\mathrm{e}^{(t+r_*)/4m}-\mathrm{e}^{(r_*-t)/4m}\right) \\ &= \mathrm{e}^{r_*/4m}\sinh\left(\frac{t}{4m}\right), \end{aligned}$$

where r_* is related to r by (6.24). Similarly

$$X = \mathrm{e}^{r_*/4m}\cosh\left(\frac{t}{4m}\right).$$

In particular, we have

$$\frac{T}{X} = \tanh\left(\frac{t}{4m}\right). \tag{9.19}$$

From Eq. (6.24), we have

$$\mathrm{e}^{r_*/4m} = \mathrm{e}^{r/4m}\sqrt{\frac{r}{2m}-1}.$$

We may summarize the above relations as follows:

$$\mathscr{M}_{\mathrm{I}}: \quad \boxed{\begin{cases} T = \mathrm{e}^{r/4m}\sqrt{\frac{r}{2m}-1}\sinh\left(\frac{t}{4m}\right) \\ X = \mathrm{e}^{r/4m}\sqrt{\frac{r}{2m}-1}\cosh\left(\frac{t}{4m}\right) \end{cases}} \iff \boxed{\begin{cases} t = 2m\,\ln\left(\frac{X+T}{X-T}\right) \\ r = 2m\tilde{W}_0(X^2-T^2). \end{cases}} \tag{9.20}$$

Note that we have used the identity $\operatorname{artanh} x = 1/2\ln[(1+x)/(1-x)]$. The curves of constant t and constant r in the (T,X) plane are drawn in Fig. 9.2. The fact that the curves of constant t are straight lines from the origin follow immediately from Eq. (9.19).

Remark 2: Given the properties of the cosh and sinh functions, it is clear on these expressions that the constraints (9.9) are satisfied.

Remark 3: In line element (9.18) the metric components g_{TT} and g_{XX} depend on both X and T; this shows that neither $\boldsymbol{\partial}_T$ nor $\boldsymbol{\partial}_X$ coincide with a Killing vector. In other words, the coordinates (T,X) are not adapted to the spacetime symmetries, contrary to the Schwarzschild-Droste coordinates or to the Eddington-Finkelstein ones.

9.2.2 Extension to the IEF domain

We notice that the metric components (9.10) are perfectly regular at $r = 2m$. Therefore, the Kruskal-Szekeres coordinates can be extended to cover the Schwarzschild horizon $\mathscr{H}$. Actually they can be extended to all values of $r \in (0, 2m]$, i.e. to the whole domain of the ingoing Eddington-Finkelstein coordinates: the manifold $\mathscr{M}_{\rm IEF}$ introduced in Sec. 6.3.3: $\mathscr{M}_{\rm IEF} = \mathscr{M}_{\rm I} \cup \mathscr{H} \cup \mathscr{M}_{\rm II}$. Let us show this in detail. Back on $\mathscr{M}_{\rm I}$, we can express the IEF coordinate $\tilde{t}$ in terms of (T, X) by combining $\tilde{t} = v - r$ [Eq. (6.31)], $v = 4m \ln V$ [Eq. (9.3)] and $V = T + X$ [Eq. (9.8)]:

$$\tilde{t} = 4m \ln(T + X) - r. \tag{9.21}$$

The above relation is a valid expression as long as $T + X > 0$. Besides, we already noticed that the function F defined by (9.13) is a bijection from the range of $r/2m$ on $\mathscr{M}_{\rm IEF}$, i.e. $(0, +\infty)$, to $(-1, +\infty)$, with the $(0, +\infty)$ part of the latter interval representing the range of $X^2 - T^2$ on $\mathscr{M}_{\rm I}$. We may use these properties to extend the Kruskal-Szekeres coordinates to all $\mathscr{M}_{\rm IEF}$ by requiring

$$\tilde{t} = 4m \ln(T + X) - r \tag{9.22a}$$

$$\underbrace{\mathrm{e}^{r/2m} \left(\frac{r}{2m} - 1 \right)}_{F(r/2m)} = X^2 - T^2. \tag{9.22b}$$

The range of the coordinates (T, X) on $\mathscr{M}_{\rm IEF}$ is then ruled by

$$\mathscr{M}_{\rm IEF} : \quad T + X > 0 \quad \text{and} \quad X^2 - T^2 > -1,$$

which can be rewritten as

$$\mathscr{M}_{\rm IEF} : \quad -X < T < \sqrt{X^2 + 1}. \tag{9.23}$$

We deduce from (9.22) that

$$\left\{ \begin{array}{rcl} X + T & = & \mathrm{e}^{(\tilde{t}+r)/4m} \\ X - T & = & \mathrm{e}^{(r-\tilde{t})/4m} \left(\frac{r}{2m} - 1 \right). \end{array} \right. \tag{9.24}$$

Hence the relation between the ingoing Eddington-Finkelstein coordinates and the Kruskal-Szekeres ones on $\mathscr{M}_{\rm IEF}$:

$$\boxed{\left\{ \begin{array}{l} T = \mathrm{e}^{r/4m} \left[\cosh\left(\frac{\tilde{t}}{4m} \right) - \frac{r}{4m} \mathrm{e}^{-\tilde{t}/4m} \right] \\ X = \mathrm{e}^{r/4m} \left[\sinh\left(\frac{\tilde{t}}{4m} \right) + \frac{r}{4m} \mathrm{e}^{-\tilde{t}/4m} \right] \end{array} \right.} \iff \boxed{\left\{ \begin{array}{l} \tilde{t} = 2m \left[2 \ln(T + X) - \tilde{W}_0(X^2 - T^2) \right] \\ r = 2m \tilde{W}_0(X^2 - T^2) \end{array} \right.} \tag{9.25}$$

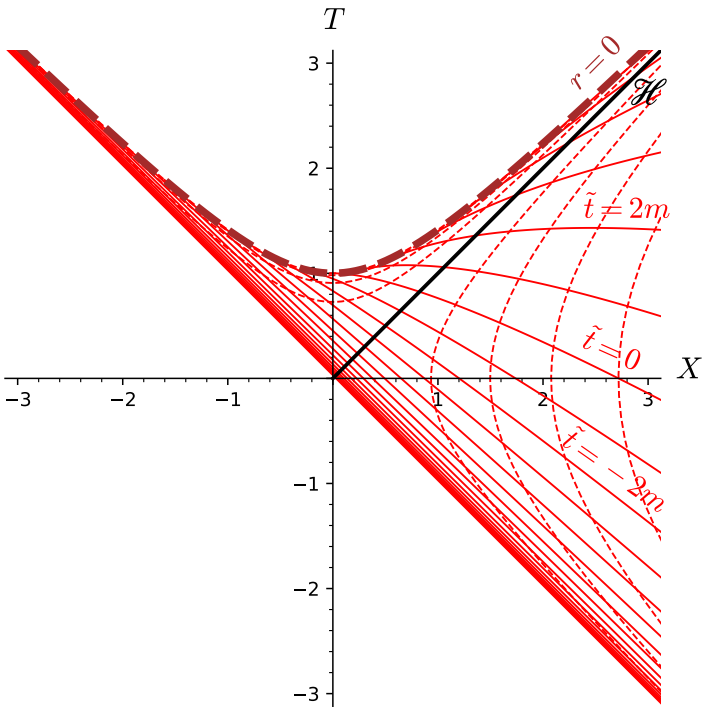


Figure 9.5: Domain of ingoing Eddington-Finkelstein coordinates, $\mathscr{M}_{\mathrm{IEF}} = \mathscr{M}_{\mathrm{I}} \cup \mathscr{H} \cup \mathscr{M}_{\mathrm{II}}$, depicted in terms of the Kruskal-Szekeres coordinates (T, X): the solid red curves have $\tilde{t} = \text{const}$ (spaced apart by $\delta\tilde{t} = m$), while the dashed red curves have $r = \text{const}$ (spaced apart by $\delta r = m/2$). *[Figure generated by the notebook D.3.16]*

The various subsets of $\mathscr{M}_{\mathrm{IEF}}$ correspond then to the following coordinate ranges (cf. Fig. 9.5):

$$\mathscr{M}_{\mathrm{I}}: \quad X > 0 \quad \text{and} \quad -X < T < X \tag{9.26a}$$

$$\mathscr{H}: \quad X > 0 \quad \text{and} \quad T = X \tag{9.26b}$$

$$\mathscr{M}_{\mathrm{II}}: \quad |X| < T < \sqrt{X^2 + 1}. \tag{9.26c}$$

Since the relation between IEF coordinates and Kruskal-Szekeres ones is the same in $\mathscr{M}_{\mathrm{II}}$ as in $\mathscr{M}_{\mathrm{I}}$ (being given by (9.25) in both cases), we conclude that the expression (9.10) of the metric components with respect to Kruskal-Szekeres coordinates is valid in all $\mathscr{M}_{\mathrm{IEF}}$.

Let us determine the relation between the Kruskal-Szekeres coordinates and the Schwarzschild-Droste ones in $\mathscr{M}_{\mathrm{II}}$. Since $r < 2m$ in $\mathscr{M}_{\mathrm{II}}$, Eq. (6.33) gives

$$\mathscr{M}_{\mathrm{II}}: \quad \mathrm{e}^{\tilde{t}/4m} = \mathrm{e}^{t/4m}\sqrt{1 - \frac{r}{2m}},$$

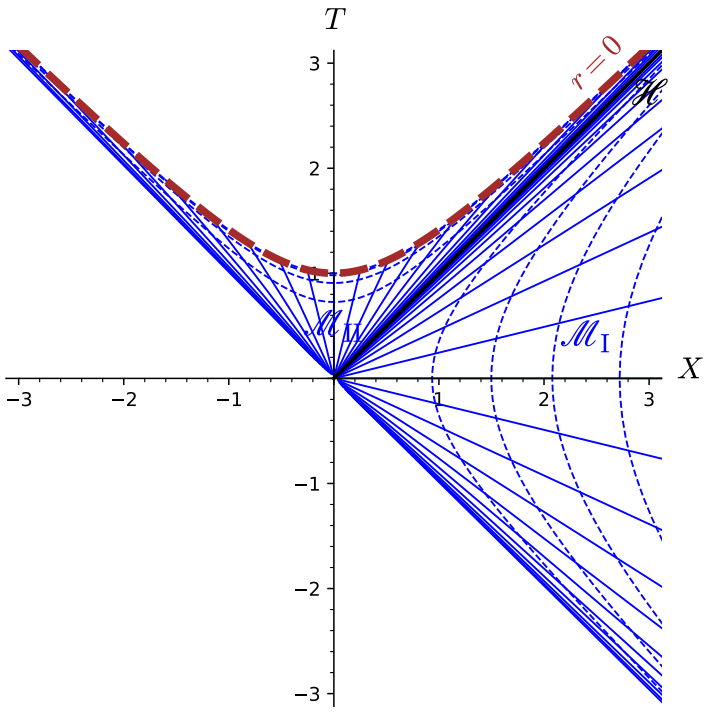


Figure 9.6: Schwarzschild-Droste coordinates in $\mathscr{M}_{\rm SD} = \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm II}$ depicted in terms of the Kruskal-Szekeres coordinates (T, X): the solid blue curves have $t = \text{const}$ (spaced apart by $\delta t = m$), while the dashed blue curves have $r = \text{const}$ (spaced apart by $\delta r = m/2$). *[Figure generated by the notebook D.3.16]*

so that (9.24) can be rewritten as

$$\mathscr{M}_{\rm II}: \quad \begin{cases} X + T &= \mathrm{e}^{(t+r)/4m}\sqrt{1-\dfrac{r}{2m}} \\ X - T &= -\mathrm{e}^{(r-t)/4m}\sqrt{1-\dfrac{r}{2m}}. \end{cases}$$

We obtain then

$$\mathscr{M}_{\rm II}: \quad \boxed{\begin{cases} T = \mathrm{e}^{r/4m}\sqrt{1-\frac{r}{2m}}\cosh\left(\frac{t}{4m}\right) \\ X = \mathrm{e}^{r/4m}\sqrt{1-\frac{r}{2m}}\sinh\left(\frac{t}{4m}\right) \end{cases}} \iff \boxed{\begin{cases} t = 2m\,\ln\left(\frac{T+X}{T-X}\right) \\ r = 2m\tilde{W}_0(X^2-T^2). \end{cases}} \tag{9.27}$$

This is to be compared with (9.20). The curves of constant t and constant r in the (T, X) plane are drawn in Fig. 9.6, which extends Fig. 9.2 to $\mathscr{M}_{\rm II}$.

As discussed in Sec. 6.3.4, one approaches a curvature singularity as $r \to 0$. According to (9.25) or (9.27), this corresponds to $X^2 - T^2 \to -1$ (see also Fig. 9.3), with $T > 0$. Hence, in the (T, X) plane, the curvature singularity is located at $T = \sqrt{X^2+1}$, i.e. at the upper branch of the hyperbola $T^2 - X^2 = 1$.

9.2.3 Radial null geodesics in Kruskal-Szekeres coordinates

By construction, the Kruskal-Szekeres coordinates (T, X, θ, φ) are adapted to the radial null geodesics. This is clear on the expression (9.10) of the metric tensor, where the (T, X) part is conformal to the flat metric $-\mathrm{d}T^2 + \mathrm{d}X^2$. Consequently the radial null geodesics are straight lines of slope $\pm 45°$ in the (T, X) plane (cf. Fig. 9.7):

- the ingoing radial null geodesics obey

$$T = -X + V, \tag{9.28}$$

  where V is a positive constant (the constraint $V > 0$ following from (9.23)), so that each geodesic of this family can be labelled by (V, θ, φ);

- the outgoing radial null geodesics obey

$$T = X + U, \tag{9.29}$$

  where U is an arbitrary real constant, so that each geodesic of this family can be labelled by (U, θ, φ).

In particular, the Schwarzschild horizon $\mathscr{H}$ is generated by the outgoing radial null geodesics having $U = 0$: Eqs. (9.29) and (9.11) clearly imply $r = 2m$ for $U = 0$, i.e. $X = T$. The outgoing radial null geodesics not lying on $\mathscr{H}$ have an equation in terms of the IEF coordinates given by Eq. (6.49): $\tilde{t} = r + 4m \ln|r/2m - 1| + u$, where the constant u is related to U by

$$U = -\mathrm{e}^{-u/4m} \quad \text{on } \mathscr{M}_{\mathrm{I}} \tag{9.30a}$$

$$U = 0 \quad \text{on } \mathscr{H} \tag{9.30b}$$

$$U = \mathrm{e}^{-u/4m} \quad \text{on } \mathscr{M}_{\mathrm{II}}. \tag{9.30c}$$

These relations are easily established by combining (6.49) and (9.25).

Remark 4: The relation $U = -\mathrm{e}^{-u/4m}$ introduced in Sec. 9.2.1 by Eq. (9.3) is thus valid only in $\mathscr{M}_{\mathrm{I}}$. On the contrary the relation $V = \mathrm{e}^{v/4m}$ is valid in all $\mathscr{M}_{\mathrm{IEF}}$.

9.3 Maximal extension

9.3.1 Construction

The spacetime $(\mathscr{M}_{\mathrm{IEF}}, \boldsymbol{g})$ is not geodesically complete (cf. Sec. B.3.2 in Appendix B). Indeed, let us consider the radial null geodesics discussed above. We have seen in Sec. 6.3.1 that r is an affine parameter along them, except for those that are null generators of $\mathscr{H}$ (the outgoing ones with $U = 0$). Now, for the ingoing radial null geodesics, r is decreasing towards the future and all of them terminate at $r = 0$ (the left end-point of the dashed lines in Fig. 9.7). They are thus incomplete geodesics. However, they cannot be extended

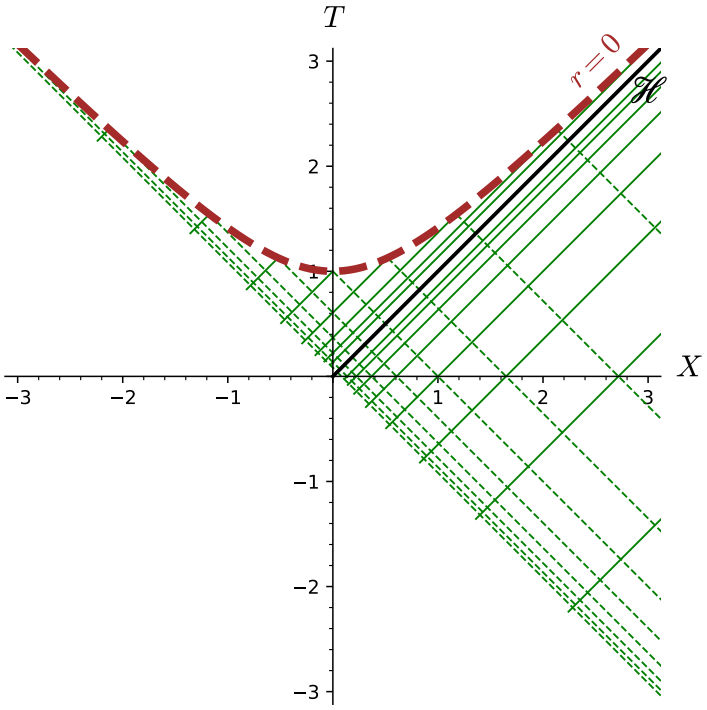


Figure 9.7: Radial null geodesics in $\mathscr{M}_{\rm IEF} = \mathscr{M}_{\rm I} \cup \mathscr{H} \cup \mathscr{M}_{\rm II}$ depicted in terms of the Kruskal-Szekeres coordinates (T, X): the solid lines correspond to the outgoing family, with u spanning $[-6m, 8m]$ (with steps $\delta u = 2m$), from the left to the right in $\mathscr{M}_{\rm II}$ and from the right to the left in $\mathscr{M}_{\rm I}$; the dashed lines correspond to the ingoing family, with v spanning $[-8m, 6m]$ (with steps $\delta v = 2m$) from the left to the right. *[Figure generated by the notebook D.3.16]*

to negative values of the affine parameter r by extending the spacetime since $r = 0$ marks a spacetime singularity (cf. Sec. 6.3.4).

On the other hand, the outgoing radial null geodesics are limited by the constraint $T + X > 0$, which corresponds to $r > 2m$ in $\mathscr{M}_{\rm I}$, with r increasing towards the future, and to $r < 2m$ in $\mathscr{M}_{\rm II}$, with r decreasing towards the future. Thus all outgoing radial null geodesics terminate towards the past at the finite value $2m$ of the affine parameter r (the left end point of the solid lines in Fig. 9.7) and are therefore incomplete geodesics. However, contrary to ingoing radial null geodesics, they can be extended since $r = 2m$ does not mark any spacetime singularity. More precisely, the limit at which outgoing radial null geodesics terminate is $T = -X$, which by virtue of (9.11) yields $r = 2m$. This does not correspond to the Schwarzschild horizon $\mathscr{H}$, since for the latter $T = X$, but rather to $\tilde{t} \to -\infty$, as it is clear when comparing Fig. 9.7 with Fig. 9.5.

Another hint regarding the extendability of $(\mathscr{M}_{\rm IEF}, \boldsymbol{g})$ is the fact that the Killing horizon $\mathscr{H}$ is non-degenerate, having a non-zero surface gravity (cf. Sec. 3.3.6); the latter has been computed in Example 10 of Chap. 2: $\kappa = 1/4m$. Now, we have seen in Sec. 5.4 that non-degenerate Killing horizons have incomplete null generators and, if they can be extended, they must be part of a bifurcate Killing horizon. In the present case, the null generators of $\mathscr{H}$ are nothing but outgoing radial null geodesics. They are thus as incomplete as those discussed above, i.e. those that admit r as an affine parameter.

The possibility of spacetime extension beyond $\mathscr{M}_{\rm IEF}$ is clear on the metric element (9.18): it is invariant by the transformation

$$\begin{array}{rccc} \Phi: & \mathbb{R}^2 & \longrightarrow & \mathbb{R}^2 \\ & (T,X) & \longmapsto & (-T,-X). \end{array} \tag{9.31}$$

Thus we may include the part $T+X<0$ by adding a copy of $\mathscr{M}_{\rm IEF}$, symmetric to the original one with respect to the "origin" $(T,X)=(0,0)$. The whole spacetime manifold is then the following open subset of $\mathbb{R}^2\times\mathbb{S}^2$:

$$\boxed{\mathscr{M} := \{p\in\mathbb{R}^2\times\mathbb{S}^2, \quad T^2(p)-X^2(p)<1\}}, \tag{9.32}$$

where (T,X,θ,φ) is the canonical coordinate system on $\mathbb{R}^2\times\mathbb{S}^2$, called in this context ***Kruskal-Szekeres coordinates***. The metric $\boldsymbol{g}$ on the whole $\mathscr{M}$ is then defined by (9.18):

$$\boxed{\begin{aligned} g_{\mu\nu}\,\mathrm{d}X^\mu\,\mathrm{d}X^\nu \;=\; 4m^2\Bigg\{ & \frac{4}{X^2-T^2+\mathrm{e}^{\tilde{W}_0(X^2-T^2)}}\left(-\mathrm{d}T^2+\mathrm{d}X^2\right) \\ & +\tilde{W}_0(X^2-T^2)^2\left(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\varphi^2\right)\Bigg\}, \end{aligned}} \tag{9.33}$$

where $\tilde{W}_0$ is the rescaled Lambert function defined by (9.16) (cf. Fig. 9.4); it is the inverse of the function $x\mapsto \mathrm{e}^x(x-1)$, which establishes a bijection from $(0,+\infty)$ to $(-1,+\infty)$.

Let us define the following open subsets of $\mathscr{M}$, which are respectively the images of $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm II}$ by the reflection through the origin (9.31):

$$\mathscr{M}_{\rm III}: \quad X<0 \quad\text{and}\quad X<T<-X \tag{9.34a}$$

$$\mathscr{M}_{\rm IV}: \quad -\sqrt{X^2+1}<T<-|X|. \tag{9.34b}$$

On $\mathscr{M}_{\rm III}\cup\mathscr{M}_{\rm IV}$, one may introduce coordinates (t',r',θ,φ) of Schwarzschild-Droste type; they are related to the Kruskal-Szekeres coordinates by formulas analogous to (9.20) and (9.27), simply changing T to $-T$ and X to $-X$:

$$\mathscr{M}_{\rm III}: \quad \boxed{\begin{cases} T=-\mathrm{e}^{r'/4m}\sqrt{\frac{r'}{2m}-1}\sinh\left(\frac{t'}{4m}\right) \\ X=-\mathrm{e}^{r'/4m}\sqrt{\frac{r'}{2m}-1}\cosh\left(\frac{t'}{4m}\right) \end{cases}} \iff \boxed{\begin{cases} t'=2m\,\ln\left(\frac{X+T}{X-T}\right) \\ r'=2m\tilde{W}_0(X^2-T^2). \end{cases}} \tag{9.35}$$

$$\mathscr{M}_{\rm IV}: \quad \boxed{\begin{cases} T=-\mathrm{e}^{r'/4m}\sqrt{1-\frac{r'}{2m}}\cosh\left(\frac{t'}{4m}\right) \\ X=-\mathrm{e}^{r'/4m}\sqrt{1-\frac{r'}{2m}}\sinh\left(\frac{t'}{4m}\right) \end{cases}} \iff \boxed{\begin{cases} t'=2m\,\ln\left(\frac{T+X}{T-X}\right) \\ r'=2m\tilde{W}_0(X^2-T^2). \end{cases}} \tag{9.36}$$

The extended Schwarzschild spacetime $(\mathscr{M},\boldsymbol{g})$ is depicted in Fig. 9.8, which is usually called a ***Kruskal diagram***. There are two curvature singularities, which formally are not

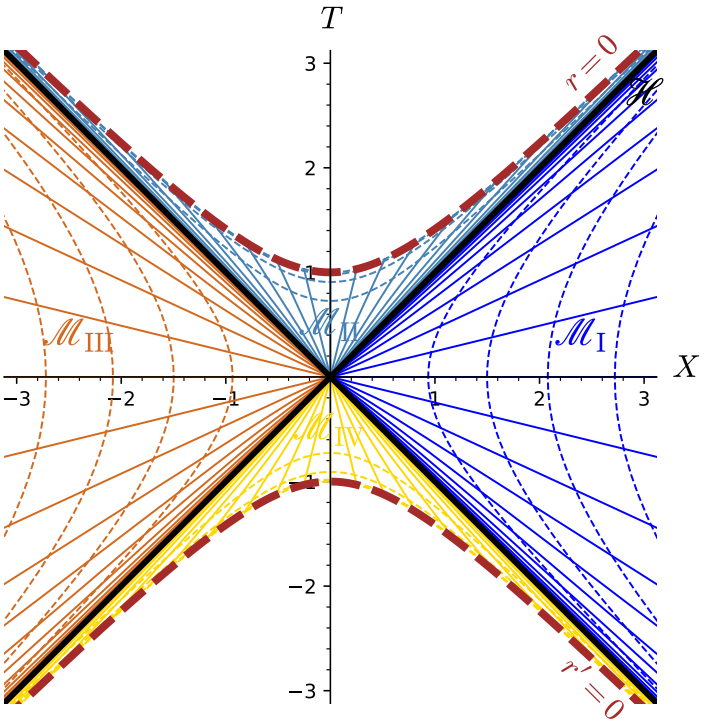


Figure 9.8: *Kruskal diagram:* Schwarzschild spacetime $\mathscr{M}$ depicted in terms of Kruskal-Szekeres coordinates (T, X). Each point in this diagram, including the one at $(T, X) = (0, 0)$, is actually a sphere $\mathbb{S}^2$, spanned by the coordinates (θ, φ). Solid lines denote the hypersurfaces $t = \text{const}$ in $\mathscr{M}_{\mathrm{I}}$ and $\mathscr{M}_{\mathrm{II}}$ and the hypersurfaces $t' = \text{const}$ in $\mathscr{M}_{\mathrm{III}}$ and $\mathscr{M}_{\mathrm{IV}}$, whiles dashed curves denote the hypersurfaces $r = \text{const}$ in $\mathscr{M}_{\mathrm{I}}$ and $\mathscr{M}_{\mathrm{II}}$ and the hypersurfaces $r' = \text{const}$ in $\mathscr{M}_{\mathrm{III}}$ and $\mathscr{M}_{\mathrm{IV}}$. The bifurcate Killing horizon is marked by thick black lines, while the singularities at $r = 0$ and $r' = 0$ are depicted by the heavy dashed brown curve. *[Figure generated by the notebook D.3.16]*

part of $\mathscr{M}$: the hypersurfaces $r = 0$ and $r' = 0$. As discussed in Sec. 9.2.3, the radial null geodesics appear as straight lines of slope $\pm 45°$ (+ for the outgoing family, and − for the ingoing one). As in $(\mathscr{M}_{\mathrm{IEF}}, \boldsymbol{g})$, they are still not complete but the only locations where they terminate are the curvature singularities at $r = 0$ (future end point) and $r' = 0$ (past end point). Therefore, they cannot be extended further. For this reason, $(\mathscr{M}, \boldsymbol{g})$ is called the ***maximal extension*** of Schwarzschild spacetime.

Remark 1: The extended manifold $\mathscr{M}$ is not just the union $\mathscr{M}_{\mathrm{I}} \cup \mathscr{M}_{\mathrm{II}} \cup \mathscr{M}_{\mathrm{III}} \cup \mathscr{M}_{\mathrm{IV}}$, since the latter does not contain the hypersurfaces $T = \pm X$ (cf. the strict inequalities in Eqs. (9.26) and (9.34)), which are parts of $\mathscr{M}$ according to the definition (9.32). Actually, we have

$$\mathscr{M} = \mathscr{M}_{\mathrm{I}} \cup \mathscr{M}_{\mathrm{II}} \cup \mathscr{M}_{\mathrm{III}} \cup \mathscr{M}_{\mathrm{IV}} \cup \hat{\mathscr{H}}, \tag{9.37}$$

where $\hat{\mathscr{H}}$ is the bifurcate Killing horizon, to be discussed in Sec. 9.3.3.

9.3.2 Global null coordinates

In Secs. 9.2.1 and 9.2.3, we have introduced on $\mathscr{M}_{\rm IEF}$ the null coordinates (U,V); they are related to the coordinates (T,X) by Eq. (9.8) (or equivalently Eqs. (9.28)-(9.29)), which we case use define (U,V) in all the maximal extension $\mathscr{M}$:

$$\left\{\begin{array}{l} U = T - X \\ V = T + X \end{array}\right. \quad \Longleftrightarrow \quad \left\{\begin{array}{l} T = \frac{1}{2}(U+V) \\ X = \frac{1}{2}(V-U) \end{array}\right. \tag{9.38}$$

The range of (U,V) of $\mathscr{M}$ is deduced from the constraint $T^2 - X^2 < 1$ [cf. Eq. (9.32)]: since $T^2 - X^2 = UV$, we get:

$$\mathscr{M}: \quad (U,V) \in \mathbb{R}^2 \quad \text{and} \quad UV < 1. \tag{9.39}$$

The expression of the metric tensor in terms of the null coordinates $X^{\hat\alpha} = (U,V,\theta,\varphi)$ is deduced from (9.33):

$$\boxed{g_{\hat\mu\hat\nu}\, \mathrm{d}X^{\hat\mu}\, \mathrm{d}X^{\hat\nu} = 4m^2 \left[\frac{4}{UV - \mathrm{e}^{\tilde W_0(-UV)}}\, \mathrm{d}U\, \mathrm{d}V + \tilde W_0(-UV)^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)\right].} \tag{9.40}$$

We can also rewrite it as (9.5):

$$\boxed{g_{\hat\mu\hat\nu}\, \mathrm{d}X^{\hat\mu}\, \mathrm{d}X^{\hat\nu} = -\frac{32m^3}{r}\, \mathrm{e}^{-r/2m}\, \mathrm{d}U\, \mathrm{d}V + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right),} \tag{9.41}$$

where r is the function of (U,V) given by

$$r = 2m\tilde W_0(-UV). \tag{9.42}$$

Note that the relation (9.6) between r and (U,V) holds in all $\mathscr{M}$:

$$\boxed{\mathrm{e}^{r/2m}\left(\frac{r}{2m} - 1\right) = -UV}. \tag{9.43}$$

Remark 2: In Sec. 9.3.1, we have distinguished the coordinate r in $\mathscr{M}_{\rm I}\cup\mathscr{M}_{\rm II}$ from the coordinate r' in $\mathscr{M}_{\rm III}\cup\mathscr{M}_{\rm IV}$. Here, r is the function (9.42) of (U,V), which has the same expression in $\mathscr{M}_{\rm I}\cup\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}\cup\mathscr{M}_{\rm IV}$. There is no need to make any distinction. Hence there is no mention of r' in (9.41).

Historical note: The Kruskal-Szekeres coordinates have been introduced in 1960 independently by Martin D. Kruskal [170] and George Szekeres [242]. Actually the coordinates introduced by Szekeres were[2] $(2T/\sqrt{\mathrm{e}}, 2X/\sqrt{\mathrm{e}})$. Both Kruskal and Szekeres have used these coordinates to construct the maximal extension of Schwarzschild spacetime. Its graphical representation in the (X,T) plane (the *Kruskal diagram*, cf. Fig. 9.8) has been presented by Kruskal (Fig. 2 of Ref. [170]). Actually, the maximal extension of Schwarzschild spacetime has been

[2] They are denoted by (v,u) in Szekeres' article [242].

first constructed by John L. Synge in 1950 [240]. He used coordinates (T', X') whose relation to Schwarzschild-Droste coordinates is more complicated than the Kruskal-Szekeres one: $T' = R(r)\sinh\left(\frac{t}{4m}\right)$ and $X' = R(r)\cosh\left(\frac{t}{4m}\right)$, with $R(r) := 2m\left[\operatorname{acosh}\sqrt{\frac{r}{2m}} + \sqrt{\frac{r}{2m}\left(\frac{r}{2m}-1\right)}\right]$; compare with (9.20). Albeit looking complicated, $R(r)$ is nothing but the primitive vanishing at $r = 2m$ of $r \mapsto \left(\frac{r}{2m}-1\right)^{-1/2}$. Interestingly, in his article [242], Szekeres says that the transformations (9.20) "are essentially due to Synge", probably because they differ only in the choice of the function $R(r)$, the latter being $R_{\text{KS}}(r) = \mathrm{e}^{r/4m}\sqrt{r/2m-1}$ for Kruskal-Szekeres coordinates. For this reason, both coordinate systems share some similarities: in Synge diagram (Figs. 8 and 9 in Ref. [240]), the bifurcate horizon appears as the two bisector lines $T' = \pm X'$ and the singularity $r = 0$ as the hyperbola $T'^2 - X'^2 = \pi^2 m^2$ (compare with $T^2 - X^2 = 1$ for Kruskal-Szekeres coordinates). A major difference is that Synge diagram is not "conformal": the radial null geodesics are generally not lines with $\pm 45°$ slope. Even, in some regions, the coordinate T' ceases to be timelike[3]. The maximal extension of Schwarzschild spacetime has also been found by Christian Fronsdal [112] in 1959, not via any explicit change of coordinates but rather via an isometric embedding of the spacetime in the 6-dimensional Minkowski spacetime.

9.3.3 Bifurcate Killing horizon

As discussed in Sec. 9.3, the Schwarzschild horizon $\mathscr{H}$ is a non-degenerate Killing horizon and therefore shall be part of a bifurcate Killing horizon (cf. Sec. 5.4) in the extended spacetime. The bifurcate Killing horizon, $\hat{\mathscr{H}}$ say, is easily found by considering the Killing vector field $\boldsymbol{\xi}$ in the maximal extension of Schwarzschild spacetime. The components of $\boldsymbol{\xi}$ w.r.t. to the Kruskal-Szekeres coordinates are obtained from the property $\boldsymbol{\xi} = \boldsymbol{\partial}_t$:

$$\xi^T = \frac{\partial T}{\partial t}, \quad \xi^X = \frac{\partial X}{\partial t}, \quad \xi^\theta = \frac{\partial \theta}{\partial t} = 0, \quad \xi^\varphi = \frac{\partial \varphi}{\partial t} = 0.$$

Given the coordinate transformation laws (9.20) and (9.27), we get in $\mathscr{M}_{\text{I}}$ and $\mathscr{M}_{\text{II}}$:

$$\xi^T = \frac{1}{4m} X, \quad \xi^X = \frac{1}{4m} T, \quad \xi^\theta = \xi^\varphi = 0.$$

Hence in $\mathscr{M}_{\text{I}} \cup \mathscr{M}_{\text{II}}$,

$$\boxed{\boldsymbol{\xi} = \frac{1}{4m}\left(X\,\boldsymbol{\partial}_T + T\,\boldsymbol{\partial}_X\right)}. \tag{9.44}$$

Now, this formula defines a smooth vector field in all $\mathscr{M}$. Moreover, in $\mathscr{M}_{\text{III}} \cup \mathscr{M}_{\text{IV}}$, this vector coincides with $\boldsymbol{\partial}_{t'}$ since $\xi^T = \partial T/\partial t'$ and $\xi^X = \partial X/\partial t'$, with the partial derivatives with respect to t' evaluated from (9.35)-(9.36). Hence the vector field $\boldsymbol{\xi}$ defined by (9.44) is a Killing vector field of maximal extension $(\mathscr{M}, \boldsymbol{g})$. This vector field is depicted in Fig. 9.9.

The bifurcate Killing horizon with respect to $\boldsymbol{\xi}$ that extends $\mathscr{H}$ is $\hat{\mathscr{H}} = \mathscr{H}_1 \cup \mathscr{H}_2$, where

- $\mathscr{H}_1$ is the null hypersurface $T = X$ (or equivalently $U = 0$);

[3] We refer the reader to Fig. 2 of Ref. [250] for a plot of Synge coordinates in terms of Kruskal-Szekeres ones

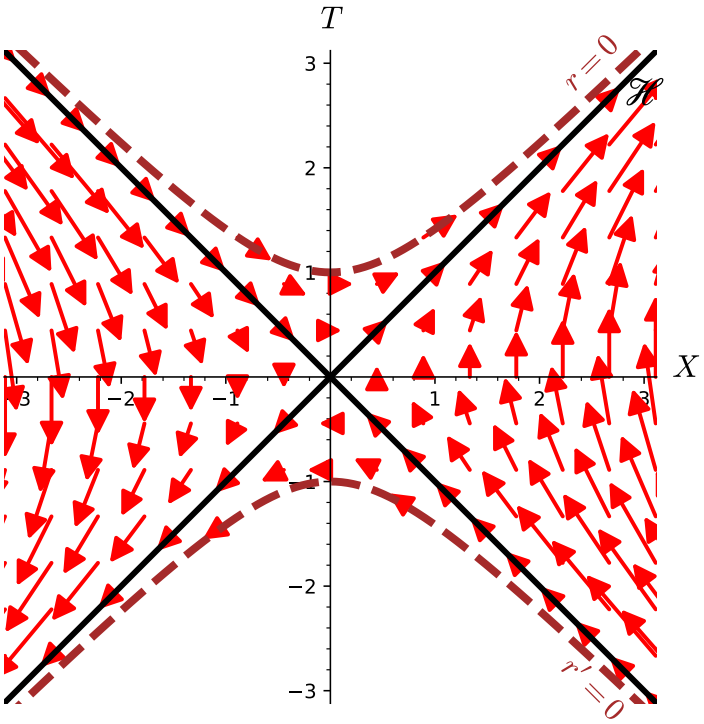


Figure 9.9: Killing vector field $\boldsymbol{\xi}$ on the extended Schwarzschild manifold. *[Figure generated by the notebook D.3.16]*

- $\mathscr{H}_2$ is the null hypersurface $T = -X$ (or equivalently $V = 0$).

The bifurcate Killing horizon $\hat{\mathscr{H}}$ is depicted in black in Fig. 9.9. The Schwarzschild horizon $\mathscr{H}$ is the part of $\mathscr{H}_1$ defined by $X > 0$. In terms of the null coordinates (U, V) introduced in Sec. 9.3.2, we have, given (9.38),

$$\hat{\mathscr{H}} : \quad U = 0 \quad \text{or} \quad V = 0 \tag{9.45a}$$

$$\mathscr{H} : \quad U = 0 \quad \text{and} \quad V > 0. \tag{9.45b}$$

The bifurcation surface is $\mathscr{S} = \mathscr{H}_1 \cap \mathscr{H}_2$, which is the 2-surface defined by $T = 0$ and $X = 0$, or equivalently by $U = 0$ and $V = 0$. It is a 2-sphere, since any fixed value of the pair (T, X) defines a 2-sphere, according to the definition of $\mathscr{M}$ as a part of $\mathbb{R}^2 \times \mathbb{S}^2$ [cf. Eq. (9.32)]. Accordingly, $\mathscr{S}$ is called the ***bifurcation sphere***. It is located at the center of Fig. 9.9. The areal radius of $\mathscr{S}$ is found by setting $\mathrm{d}T = 0$, $\mathrm{d}X = 0$ and $(T, X) = (0, 0)$ in the line element (9.33):

$$r_{\mathscr{S}}^2 = 4m^2 \tilde{W}_0(0)^2.$$

Since $\tilde{W}_0(0) = 1$ (cf. Fig. 9.4), we get

$$\boxed{r_{\mathscr{S}} = 2m}. \tag{9.46}$$

Moreover, setting $(T, X) = (0, 0)$ in Eq. (9.44), we recover the general property (5.8): the Killing vector field vanishes at the bifurcation sphere:

$$\boldsymbol{\xi}|_{\mathscr{S}} = 0. \tag{9.47}$$

9.4 Carter-Penrose diagram

9.4.1 First construction

To have a compact representation of the maximal extension of Schwarzschild spacetime, one can use the same trick as for Minkowski spacetime (cf. Sec. 4.2.1), namely employ the arctangent function to map the range $(-\infty, +\infty)$ of the null coordinates U and V to the interval $(-\pi/2, \pi/2)$, thereby defining the finite-range coordinates $(\hat{U}, \hat{V})$:

$$\begin{cases} \hat{U} = \arctan U \\ \hat{V} = \arctan V \end{cases} \iff \begin{cases} U = \tan \hat{U} \\ V = \tan \hat{V}. \end{cases} \tag{9.48}$$

The range of $(\hat{U}, \hat{V})$ is deduced from (9.39):

$$UV < 1 \iff \tan \hat{U} \, \tan \hat{V} < 1.$$

Since for $\hat{U}, \hat{V} \in (-\pi/2, \pi/2)$, we have $\cos \hat{U} > 0$ and $\cos \hat{V} > 0$, we may write

$$UV < 1 \iff \sin \hat{U} \, \sin \hat{V} < \cos \hat{U} \, \cos \hat{V} \iff \cos(\hat{U} + \hat{V}) > 0 \iff -\frac{\pi}{2} < \hat{U} + \hat{V} < \frac{\pi}{2}.$$

Hence the range of $(\hat{U}, \hat{V})$ on the maximal extension of Schwarzschild spacetime:

$$\mathscr{M} : \quad -\frac{\pi}{2} < \hat{U} < \frac{\pi}{2}, \quad -\frac{\pi}{2} < \hat{V} < \frac{\pi}{2} \quad \text{and} \quad -\frac{\pi}{2} < \hat{U} + \hat{V} < \frac{\pi}{2}. \tag{9.49}$$

Since (9.48) yields $\mathrm{d}U = \mathrm{d}\hat{U}/\cos^2 \hat{U}$ and $\mathrm{d}V = \mathrm{d}\hat{V}/\cos^2 \hat{V}$, we deduce immediately from (9.41) the expression of the metric tensor in terms of the coordinates $x^\alpha = (\hat{U}, \hat{V}, \theta, \varphi)$:

$$\boxed{g_{\mu\nu} \, \mathrm{d}x^\mu \, \mathrm{d}x^\nu = -\frac{32m^3}{r} \, \mathrm{e}^{-r/2m} \, \frac{\mathrm{d}\hat{U}}{\cos^2 \hat{U}} \, \frac{\mathrm{d}\hat{V}}{\cos^2 \hat{V}} + r^2 \left(\mathrm{d}\theta^2 + \sin^2 \theta \, \mathrm{d}\varphi^2 \right)}, \tag{9.50}$$

where [cf. Eq. (9.42)]

$$r = 2m \tilde{W}_0(-\tan \hat{U} \, \tan \hat{V}). \tag{9.51}$$

To depict $\mathscr{M}$, let us introduce "time+space" coordinates $(\hat{T}, \hat{X})$, which are related to $(\hat{U}, \hat{V})$ in exactly the same way as the coordinates (τ, χ) were related to the finite-range null coordinates (U, V) for Minkowski spacetime [cf. Eq. (4.13)]:

$$\begin{cases} \hat{T} = \hat{U} + \hat{V} \\ \hat{X} = \hat{V} - \hat{U} \end{cases} \iff \begin{cases} \hat{U} = \frac{1}{2}(\hat{T} - \hat{X}) \\ \hat{V} = \frac{1}{2}(\hat{T} + \hat{X}). \end{cases} \tag{9.52}$$

The range of $(\hat{T}, \hat{X})$ is deduced from (9.49):

$$\mathscr{M} : \quad -\frac{\pi}{2} < \hat{T} < \frac{\pi}{2}, \quad \hat{T} - \pi < \hat{X} < \hat{T} + \pi \quad \text{and} \quad -\hat{T} - \pi < \hat{X} < -\hat{T} + \pi. \tag{9.53}$$

Via (9.48) and (9.38), the relation between $(\hat{T}, \hat{X})$ and the Kruskal-Szekeres coordinates (T, X) is then the same as that between (τ, χ) and (t, r) for Minkowksi spacetime [Eq. (4.15)]:

$$\begin{cases} \hat{T} = \arctan(T+X) + \arctan(T-X) \\ \hat{X} = \arctan(T+X) - \arctan(T-X) \end{cases} \iff \begin{cases} T = \dfrac{\sin \hat{T}}{\cos \hat{T} + \cos \hat{X}} \\ X = \dfrac{\sin \hat{X}}{\cos \hat{T} + \cos \hat{X}}. \end{cases} \tag{9.54}$$

The maximal extension of Schwarzschild spacetime is depicted with respect to the coordinates $(\hat{T}, \hat{X})$ in Fig. 9.10. Such a plot is called a ***Carter-Penrose diagram*** (see the historical note p. 238). As the Kruskal diagram (Fig. 9.8), it has the property to display the radial null geodesics as straight lines with slope $\pm 45°$. This holds since $\hat{U}$ (resp. $\hat{V}$) is a function of U only (resp. V only), cf. Eq. (9.48), so that $\hat{U}$ (resp. $\hat{V}$) is constant on outgoing (resp. ingoing) radial null geodesics. In particular, the bifurcate Killing horizon and the Schwarzschild horizon are obtained for specific values of $\hat{U}$ and $\hat{V}$:

$$\hat{\mathscr{H}} : \quad \hat{U} = 0 \quad \text{or} \quad \hat{V} = 0 \tag{9.55a}$$

$$\mathscr{H} : \quad \hat{U} = 0 \quad \text{and} \quad \hat{V} > 0. \tag{9.55b}$$

These relations follow immediately from (9.45) and (9.48).

We have seen in Sec. 6.4 that the future null infinity $\mathscr{I}^+$ corresponds to $v \to +\infty$ and that the past null infinity $\mathscr{I}^-$ to $u \to -\infty$ (cf. Fig. 6.6). Since on $\mathscr{M}_{\rm I}$, $U = -\mathrm{e}^{-u/4m}$ and $V = \mathrm{e}^{v/4m}$ [cf. Eq. (9.3)], we may write equivalently:

$$\mathscr{I}^+ : \quad V \to +\infty \quad \text{and} \quad U \in (-\infty, 0) \tag{9.56a}$$

$$\mathscr{I}^- : \quad U \to -\infty \quad \text{and} \quad V \in (0, +\infty). \tag{9.56b}$$

In view of (9.48), we get then:

$$\mathscr{I}^+ : \quad \hat{V} \to \frac{\pi}{2} \quad \text{and} \quad \hat{U} \in \left(-\frac{\pi}{2}, 0\right) \tag{9.57a}$$

$$\mathscr{I}^- : \quad \hat{U} \to -\frac{\pi}{2} \quad \text{and} \quad \hat{V} \in \left(0, \frac{\pi}{2}\right). \tag{9.57b}$$

By symmetry, the extension $\mathscr{M}_{\rm III} \cup \mathscr{M}_{\rm IV}$ of Schwarzschild spacetime has the following null infinity:

$$\mathscr{I}'^+ : \quad \hat{U} \to \frac{\pi}{2} \quad \text{and} \quad \hat{V} \in \left(-\frac{\pi}{2}, 0\right) \tag{9.58a}$$

$$\mathscr{I}'^- : \quad \hat{V} \to -\frac{\pi}{2} \quad \text{and} \quad \hat{U} \in \left(0, \frac{\pi}{2}\right). \tag{9.58b}$$

9.4.2 Discussion: Carter-Penrose diagram and conformal completion

The Carter-Penrose diagram in Fig. 9.10 can be compared with the conformal diagram of Minkowski spacetime in Fig. 4.3. The right asymptotics of the Carter-Penrose diagram

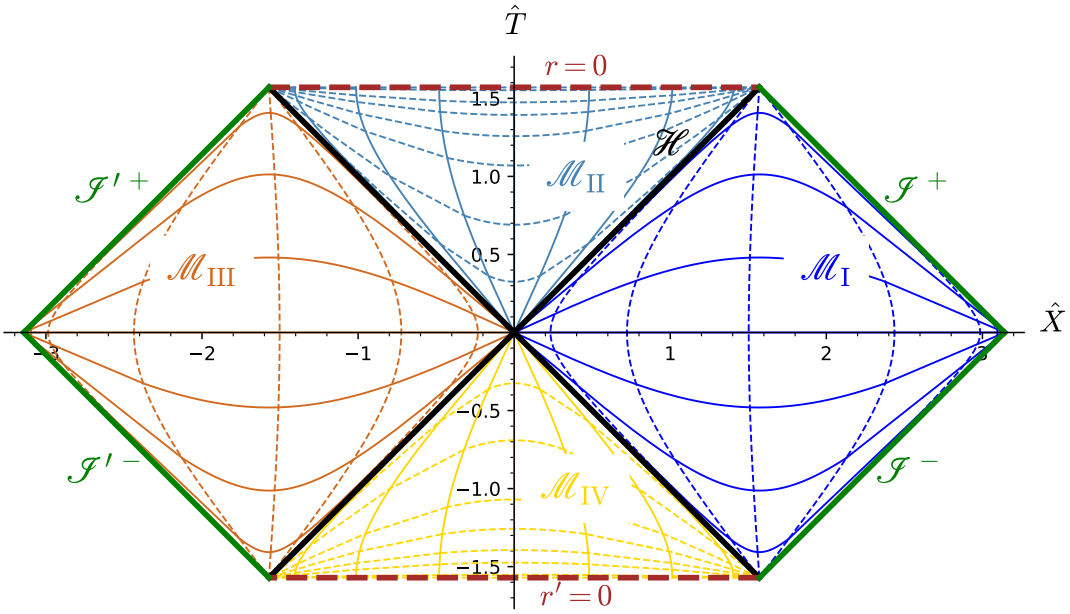


Figure 9.10: Carter-Penrose diagram of the Schwarzschild spacetime constructed with the compactified coordinates $(\hat{T}, \hat{X})$. Solid curves denote hypersurfaces of constant Schwarzschild-Droste coordinate t: in region $\mathscr{M}_{\rm I}$, from the $\hat{X}$-axis to the top: $t = 0$, $2m$, $5m$, $10m$, $20m$ and $50m$, the last two being barely visible; in region $\mathscr{M}_{\rm II}$, from the $\hat{T}$-axis to the right: $t = 0$, $2m$, $5m$, $10m$, $20m$ and $50m$, Dashed curves denote hypersurfaces of constant Schwarzschild-Droste coordinate r: in region $\mathscr{M}_{\rm I}$, from the left to the right: $r = 2.01m$, $2.1m$, $2.5m$ (almost vertical), $4m$, $8m$, $12m$, $20m$ and $100m$, the last three being barely visible; in region $\mathscr{M}_{\rm II}$, from the bottom to the top: $r = 1.98m$, $1.9m$, $1.7m$, $1.5m$, $1.25m$, m, $0.5m$ and $0.1m$. The color code is the same as in Fig. 9.8. *[Figure generated by the notebook D.3.17]*

(i.e. the part $\hat{X} > \pi/2$) looks similar to that of Minkowski conformal diagram. However, there is a difference: the coordinates $(\hat{T}, \hat{X})$ employed in the construction of the diagram of Fig. 9.10 are not related to any (regular) conformal completion — as defined in Sec. 4.3 — contrary to the coordinates (τ, χ) used for Minkowski spacetime.

To see this, let us rewrite the metric components (9.50) in a form that makes clear their behaviour near null infinity. Given (9.51) and (9.17), we have

$$\mathrm{e}^{r/2m}\left(\frac{r}{2m} - 1\right) = -\tan\hat{U}\tan\hat{V}, \tag{9.59}$$

from which we get

$$\frac{2m}{r}\mathrm{e}^{-r/2m} = -\frac{1 - 2m/r}{\tan\hat{U}\tan\hat{V}}.$$

Hence

$$\frac{2m}{r}\frac{\mathrm{e}^{-r/2m}}{\cos^2\hat{U}\cos^2\hat{V}} = -\frac{1 - 2m/r}{\sin\hat{U}\cos\hat{U}\sin\hat{V}\cos\hat{V}} = -\frac{4(1 - 2m/r)}{\sin 2\hat{U}\sin 2\hat{V}}.$$

Therefore, we may rewrite expression (9.50) for the metric tensor as

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = 64m^2\left(1 - \frac{2m}{r}\right)\frac{\mathrm{d}\hat{U}\,\mathrm{d}\hat{V}}{\sin 2\hat{U}\sin 2\hat{V}} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right), \tag{9.60}$$

with r given by (9.51). To get a conformal completion, we should write (cf. Sec. 4.3)

$$\boldsymbol{g} = \Omega^{-2}\tilde{\boldsymbol{g}}, \tag{9.61}$$

where $\Omega = 0$ and $\mathbf{d}\Omega \neq 0$ on the spacetime boundary $\mathscr{I}$ and $\tilde{\boldsymbol{g}}$ is a regular metric on the completion $\mathscr{M} \cup \mathscr{I}$. Since in Eq. (9.60), the term $\sin 2\hat{U} \sin 2\hat{V}$ vanishes at $\mathscr{I} = \mathscr{I}^+ \cup \mathscr{I}^- \cup \mathscr{I}'^+ \cup \mathscr{I}'^-$ [cf. Eqs. (9.57)-(9.58)], we would have, up to some constant factor,

$$\Omega = \sqrt{-\sin 2\hat{U} \sin 2\hat{V}}, \tag{9.62}$$

the minus sign taking into account that $\sin 2\hat{U} \sin 2\hat{V}$ approaches zero via negative values near $\mathscr{I}$. A first issue is that the square root in (9.62) makes Ω not differentiable on $\mathscr{I}$, where either $\sin 2\hat{U} = 0$ or $\sin 2\hat{V} = 0$. In other words, $\mathbf{d}\Omega$ is diverging on $\mathscr{I}$. Suppose we accept this and are ready to introduce a slight deviation (given that Ω^2, which is involved in (9.61), is smooth) from the definition given in Sec. 4.3. Then the conformal metric should be

$$\tilde{g}_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu = -64m^2 \left(1 - \frac{2m}{r}\right) \mathrm{d}\hat{U}\, \mathrm{d}\hat{V} - r^2 \sin 2\hat{U} \sin 2\hat{V} \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right). \tag{9.63}$$

Near $\mathscr{I}$, $r \to +\infty$ and we have $\tilde{g}_{\hat{U}\hat{V}} \to -32m^2$. On the contrary, $\tilde{g}_{\theta\theta}$ is of the type "$\infty \times 0$"; in order to determine its behavior, let us rewrite it as follows:

$$\tilde{g}_{\theta\theta} = -r^2 \sin 2\hat{U} \sin 2\hat{V} = -4r^2 \sin\hat{U} \sin\hat{V} \times \cos\hat{U} \cos\hat{V},$$

with $\cos\hat{U} \cos\hat{V}$ expressed via (9.59):

$$\cos\hat{U} \cos\hat{V} = -\sin\hat{U} \sin\hat{V} \frac{\mathrm{e}^{-r/2m}}{r/2m - 1}.$$

Hence

$$\tilde{g}_{\theta\theta} = 8m \sin^2\hat{U} \sin^2\hat{V} \frac{r\mathrm{e}^{-r/2m}}{1 - 2m/r}$$

and (9.63) becomes

$$\tilde{g}_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu = -64m^2 \left(1 - \frac{2m}{r}\right) \mathrm{d}\hat{U}\, \mathrm{d}\hat{V} + 8m \sin^2\hat{U} \sin^2\hat{V} \frac{r\mathrm{e}^{-r/2m}}{1 - 2m/r} \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right). \tag{9.64}$$

On this expression, we can read directly the value of the conformal metric at $\mathscr{I}$, where $r \to +\infty$, $2m/r \to 0$, $r\mathrm{e}^{-r/2m} \to 0$ and $\sin^2\hat{U} \to 1$ or $\sin^2\hat{V} \to 1$:

$$\tilde{g}_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu \overset{\mathscr{I}}{=} -64m^2 \mathrm{d}\hat{U}\, \mathrm{d}\hat{V}. \tag{9.65}$$

This bilinear form is clearly degenerate (cf. Sec. A.3.1). Therefore, $\tilde{\boldsymbol{g}}$ is not a regular metric on the whole manifold $\mathscr{M} \cup \mathscr{I}$. We conclude that (9.62)-(9.63) does not define a conformal completion of $(\mathscr{M}, \boldsymbol{g})$.

Historical note: The first compactified conformal diagram of the (maximal extension of) Schwarzschild spacetime has been constructed by Brandon Carter in 1966 [39], using the same coordinates $(\hat{T}, \hat{X})$ as here[4]: compare Fig. 9.10 with Fig. 1c of Ref. [39]. In his article, Carter notes that "the manner in which the distant flat-space parts (...) are compressed into finite parts of the (ξ, ψ) plane by the coordinate transformations recalls the conformal diagrams used by R. Penrose" in 1964 [206], which regard Minkowski and de Sitter spacetimes only. Hence it seems quite fair to call the graphical representation shown in Fig. 9.10 a *Carter-Penrose diagram*, and not merely a *Penrose diagram*, as often done in the literature.

9.4.3 A regular conformal completion based on Frolov-Novikov coordinates

In order to get a regular conformal completion of the maximally extended Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$, a finite-range coordinate system has been proposed by Frolov & Novikov [110]: instead of (9.48), the finite-range coordinates $(\tilde{U}, \tilde{V})$ are defined in terms of the null Kruskal-Szekeres coordinates (U, V) by

$$\begin{cases} \tilde{U} = \arctan(\operatorname{arsinh} U) \\ \tilde{V} = \arctan(\operatorname{arsinh} V) \end{cases} \iff \begin{cases} U = \sinh(\tan \tilde{U}) \\ V = \sinh(\tan \tilde{V}). \end{cases} \tag{9.66}$$

The range of $(\tilde{U}, \tilde{V})$ is deduced from (9.39):

$$\mathscr{M}: \quad -\frac{\pi}{2} < \tilde{U} < \frac{\pi}{2}, \quad -\frac{\pi}{2} < \tilde{V} < \frac{\pi}{2} \quad \text{and} \quad \sinh(\tan \tilde{U}) \sinh(\tan \tilde{V}) < 1. \tag{9.67}$$

Note that contrary to what happened for $(\hat{U}, \hat{V})$, these conditions do not yield to simple polygonal region in the $(\tilde{U}, \tilde{V})$ plane. The presence of the sinh function in the expression (9.66) of (U, V) in terms of $(\tilde{U}, \tilde{V})$ does not alter the values of the finite-range coordinates at null infinity, as compared to $(\hat{U}, \hat{V})$ [cf. (9.57)-(9.58)]:

$$\mathscr{I}^+: \quad \tilde{V} \to \frac{\pi}{2} \quad \text{and} \quad \tilde{U} \in \left(-\frac{\pi}{2}, 0\right) \tag{9.68a}$$

$$\mathscr{I}^-: \quad \tilde{U} \to -\frac{\pi}{2} \quad \text{and} \quad \tilde{V} \in \left(0, \frac{\pi}{2}\right) \tag{9.68b}$$

$$\mathscr{I}'^+: \quad \tilde{U} \to \frac{\pi}{2} \quad \text{and} \quad \tilde{V} \in \left(-\frac{\pi}{2}, 0\right) \tag{9.68c}$$

$$\mathscr{I}'^-: \quad \tilde{V} \to -\frac{\pi}{2} \quad \text{and} \quad \tilde{U} \in \left(0, \frac{\pi}{2}\right). \tag{9.68d}$$

We shall call $(\tilde{U}, \tilde{V}, \theta, \varphi)$ the ***Frolov-Novikov coordinates***.

From (9.66), we get

$$\mathrm{d}U = \frac{\cosh(\tan \tilde{U})}{\cos^2 \tilde{U}} \mathrm{d}\tilde{U} \qquad \text{and} \qquad \mathrm{d}V = \frac{\cosh(\tan \tilde{V})}{\cos^2 \tilde{V}} \mathrm{d}\tilde{V},$$

[4] $(\hat{T}, \hat{X})$ are denoted (ψ, ξ) by Carter [39].

so that the metric components in terms of the coordinates $x^\alpha = (\tilde U, \tilde V, \theta, \varphi)$ are easily deduced from (9.41)-(9.42):

$$\boxed{g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\frac{32m^3}{r}\,\mathrm{e}^{-r/2m}\,\frac{\cosh(\tan\tilde U)\cosh(\tan\tilde V)}{\cos^2\tilde U\cos^2\tilde V}\,\mathrm{d}\tilde U\,\mathrm{d}\tilde V + r^2\left(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\varphi^2\right)}, \tag{9.69}$$

where r is the function of $(\tilde U, \tilde V)$ given by

$$r = 2m\tilde W_0\Big(-\sinh(\tan\tilde U)\sinh(\tan\tilde V)\Big)\,. \tag{9.70}$$

As we did for $(\hat U, \hat V)$, let us rewrite (9.69) in a form that is better adapted to the null asymptotics. Given (9.70) and (9.17), we have

$$\mathrm{e}^{r/2m}\left(\frac{r}{2m}-1\right) = -\sinh(\tan\tilde U)\sinh(\tan\tilde V), \tag{9.71}$$

from which we get

$$\frac{2m}{r}\mathrm{e}^{-r/2m} = -\left(1-\frac{2m}{r}\right)\frac{1}{\sinh(\tan\tilde U)\sinh(\tan\tilde V)}.$$

Hence (9.69) becomes

$$\begin{aligned} g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu &= 16m^2\left(1-\frac{2m}{r}\right)\frac{\mathrm{d}\tilde U\,\mathrm{d}\tilde V}{\tanh(\tan\tilde U)\tanh(\tan\tilde V)\cos^2\tilde U\cos^2\tilde V} \\ &\quad + r^2\left(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\varphi^2\right). \end{aligned} \tag{9.72}$$

Given the values (9.68) of $\tilde U$ and $\tilde V$ near $\mathscr{I}$, $\tanh(\tan\tilde U)\tanh(\tan\tilde V)$ does not vanish there. A natural choice of conformal factor is then

$$\boxed{\Omega := \cos\tilde U\cos\tilde V}. \tag{9.73}$$

The corresponding conformal metric is

$$\boxed{\begin{aligned} \tilde g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu &= 16m^2\left(1-\frac{2m}{r}\right)\frac{\mathrm{d}\tilde U\,\mathrm{d}\tilde V}{\tanh(\tan\tilde U)\tanh(\tan\tilde V)} \\ &\quad + r^2\cos^2\tilde U\cos^2\tilde V\left(\mathrm{d}\theta^2+\sin^2\theta\,\mathrm{d}\varphi^2\right) \end{aligned}}. \tag{9.74}$$

Considering $(\tilde U, \tilde V, \theta, \varphi)$ as a canonical coordinate system on $\mathbb{R}^2\times\mathbb{S}^2$, we define the conformal completion manifold as

$$\begin{aligned} \tilde{\mathscr{M}} &:= \left\{p\in\mathbb{R}^2\times\mathbb{S}^2,\ (\tilde U(p),\tilde V(p))\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right)^2 \text{ and } \sinh(\tan\tilde U(p))\sinh(\tan\tilde V(p))<1\right\} \\ &\quad \cup\mathscr{I}^+\cup\mathscr{I}^-\cup\mathscr{I}'^+\cup\mathscr{I}'^-, \end{aligned} \tag{9.75}$$

with

$$\mathscr{I}^+ := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2,\ \tilde{V}(p) = \frac{\pi}{2} \text{ and } \tilde{U}(p) \in \left(-\frac{\pi}{2}, 0\right) \right\} \tag{9.76a}$$

$$\mathscr{I}^- := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2,\ \tilde{U}(p) = -\frac{\pi}{2} \text{ and } \tilde{V}(p) \in \left(0, \frac{\pi}{2}\right) \right\} \tag{9.76b}$$

$$\mathscr{I}'^+ := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2,\ \tilde{U}(p) = \frac{\pi}{2} \text{ and } \tilde{V}(p) \in \left(-\frac{\pi}{2}, 0\right) \right\} \tag{9.76c}$$

$$\mathscr{I}'^- := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2,\ \tilde{V}(p) = -\frac{\pi}{2} \text{ and } \tilde{U}(p) \in \left(0, \frac{\pi}{2}\right) \right\}. \tag{9.76d}$$

Note that the first line in (9.75) corresponds to $\mathscr{M}$, identified as a subset of $\mathbb{R}^2 \times \mathbb{S}^2$ [cf. Eq. (9.67)] and that the definitions of $\mathscr{I}^+$, $\mathscr{I}^-$, $\mathscr{I}'^+$ and $\mathscr{I}'^-$ are in agreement with (9.68). It is clear that $\tilde{\mathscr{M}}$ is a manifold with boundary and that

$$\partial\tilde{\mathscr{M}} = \mathscr{I} := \mathscr{I}^+ \cup \mathscr{I}^- \cup \mathscr{I}'^+ \cup \mathscr{I}'^-. \tag{9.77}$$

Moreover, the scalar field Ω defined by (9.73) satisfies $\Omega \geq 0$ on $\tilde{\mathscr{M}}$, along with $\Omega = 0$ on $\mathscr{I}$ and $\mathbf{d}\Omega \neq 0$ on $\mathscr{I}$. The last property follows from

$$\mathbf{d}\Omega = -\sin\tilde{U}\cos\tilde{V}\,\mathbf{d}\tilde{U} - \cos\tilde{U}\sin\tilde{V}\,\mathbf{d}\tilde{V},$$

which implies $\mathbf{d}\Omega|_{\mathscr{I}^+} = -\cos\tilde{U}\,\mathbf{d}\tilde{V} \neq 0$, $\mathbf{d}\Omega|_{\mathscr{I}^-} = \cos\tilde{V}\,\mathbf{d}\tilde{U} \neq 0$, $\mathbf{d}\Omega|_{\mathscr{I}'^+} = -\cos\tilde{V}\,\mathbf{d}\tilde{U} \neq 0$ and $\mathbf{d}\Omega|_{\mathscr{I}'^-} = \cos\tilde{U}\,\mathbf{d}\tilde{V} \neq 0$. Hence the conditions 1, 3 and 4 of the definition of a conformal completion given in Sec. 4.3 are fulfilled. There remains to check condition 2, namely that the tensor $\tilde{\boldsymbol{g}}$ defined by (9.74) is a regular metric on the whole $\tilde{\mathscr{M}}$. This was the main failing point in the attempt of Sec. 9.4.2. Since $\Omega^2 > 0$ on $\mathscr{M}$, $\tilde{\boldsymbol{g}}$ is well-behaved on $\mathscr{M}$. Let us thus examine its behaviour on $\mathscr{I}$. We shall focus on $\mathscr{I}^+$, the behaviour on the other parts of $\mathscr{I}$ being obtained by some trivial symmetry. As one approaches $\mathscr{I}^+$, $r \to +\infty$, $\tilde{V} \to \pi/2$ and $\tanh(\tan(\tilde{V})) \to 1$; accordingly we read from (9.74) that

$$\tilde{g}_{\tilde{U}\tilde{V}} \overset{\mathscr{I}^+}{=} \frac{8m^2}{\tanh(\tan\tilde{U})}.$$

Besides, we have $\tilde{g}_{\theta\theta} = r^2\cos^2\tilde{U}\cos^2\tilde{V}$, which is of the type "$+\infty \times 0$" near $\mathscr{I}^+$. Noticing that $\tan\tilde{V} \sim 1/\cos\tilde{V}$ when $\tilde{V} \to \pi/2$, we get from (9.71)

$$\sinh\left(\frac{1}{\cos\tilde{V}}\right) \sim -\frac{r\mathrm{e}^{r/2m}}{2m\sinh(\tan\tilde{U})} \quad \text{when} \quad \tilde{V} \to \frac{\pi}{2}.$$

Since $\operatorname{arsinh}(x) = \ln(x + \sqrt{x^2+1}) \sim \ln(2x)$ when $x \to +\infty$, we obtain

$$\begin{aligned} \frac{1}{\cos\tilde{V}} &\sim \ln\left(-\frac{r\mathrm{e}^{r/2m}}{m\sinh(\tan\tilde{U})}\right) = \frac{r}{2m} + \ln\left(\frac{r}{m}\right) - \ln\Big(-\sinh(\tan\tilde{U})\Big) \\ &\sim \frac{r}{2m} \quad \text{when} \quad \tilde{V} \to \frac{\pi}{2}. \end{aligned}$$

Hence $\cos^2\tilde{V} \sim 4m^2/r^2$ and $\tilde{g}_{\theta\theta} \sim 4m^2\cos^2\tilde{U}$. Gathering the above results, we have

$$\tilde{g}_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu \overset{\mathscr{I}^+}{=} 4m^2\left[\frac{4}{\tanh(\tan\tilde{U})}\,\mathrm{d}\tilde{U}\,\mathrm{d}\tilde{V} + \cos^2\tilde{U}\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)\right]. \tag{9.78}$$

Since $\cos^2 \tilde{U} \neq 0$ on $\mathscr{I}^+$ [cf. Eq. (9.76a)], this bilinear form is non-degenerate. Moreover, since $\tanh(\tan \tilde{U}) < 0$ on $\mathscr{I}^+$ [again by (9.76a)], its signature is $(-,+,+,+)$. We conclude that $\tilde{\boldsymbol{g}}$ is a well-behaved metric on the whole manifold $\tilde{\mathscr{M}}$. This completes the demonstration that

> The pair $(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$, with $\tilde{\mathscr{M}}$ defined by (9.75)-(9.76) and $\tilde{\boldsymbol{g}}$ defined by (9.74) is a conformal completion of the maximally extented Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$, the conformal factor being given by (9.73). The Frolov-Novikov coordinates $(\tilde{U}, \tilde{V}, \theta, \varphi)$ employed in this construction are related to the null Kruskal-Szekeres coordinates (U, V, θ, φ) by (9.66).

Remark 1: At first sight, the metric $\tilde{\boldsymbol{g}}$ given by (9.74) looks degenerate at the bifurcate Killing horizon $\mathscr{H}$, since $1 - 2m/r = 0$ there. But one shall not forget that on $\mathscr{H}$, which is defined by ($\tilde{U} = 0$ or $\tilde{V} = 0$), one has $\tanh(\tan \tilde{U}) \tanh(\tan \tilde{V}) = 0$, which compensate the vanishing of $1 - 2m/r$ in the term $\tilde{g}_{\tilde{U}\tilde{V}}$. Actually, to deal with $\tilde{\boldsymbol{g}}$ near $\mathscr{H}$, it is more appropriate to use the form that is deduced from (9.69) and (9.73):

$$\begin{aligned} \tilde{g}_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu &= -\frac{32m^3}{r}\, \mathrm{e}^{-r/2m}\, \cosh(\tan \tilde{U}) \cosh(\tan \tilde{V})\, \mathrm{d}\tilde{U}\, \mathrm{d}\tilde{V} \\ &\quad + r^2 \cos^2 \tilde{U} \cos^2 \tilde{V} \left(\mathrm{d}\theta^2 + \sin^2 \theta\, \mathrm{d}\varphi^2\right). \end{aligned} \tag{9.79}$$

Remark 2: The conformal completion constructed above cannot be analytically extended "beyond" $\mathscr{I}$, because the function $\tilde{V} \mapsto 1/\tanh(\tan \tilde{V})$, which appears in (9.74), is C^∞ but not analytic at $\tilde{V} = \pi/2$. It is possible to construct an analytic conformal completion, but it involves more complicated coordinate transformations. The latter start, not from the Kruskal-Szekeres coordinates (U, V), but from the null coordinates (u, v) defined by (9.1). We refer to the article [141] for details.

To depict $\tilde{\mathscr{M}}$, let us introduce "time+space" coordinates $(\tilde{T}, \tilde{X})$, which are related to $(\tilde{U}, \tilde{V})$ in exactly the same way as $(\hat{T}, \hat{X})$ are related to $(\hat{U}, \hat{V})$ [cf. Eq. (9.52)]:

$$\left\{ \begin{array}{l} \tilde{T} = \tilde{U} + \tilde{V} \\ \tilde{X} = \tilde{V} - \tilde{U} \end{array} \right. \iff \left\{ \begin{array}{l} \tilde{U} = \frac{1}{2}(\tilde{T} - \tilde{X}) \\ \tilde{V} = \frac{1}{2}(\tilde{T} + \tilde{X}). \end{array} \right. \tag{9.80}$$

The range of $(\tilde{T}, \tilde{X})$ is deduced from (9.67):

$$\begin{aligned} \mathscr{M}: \quad & -\pi < \tilde{T} - \tilde{X} < \pi, \quad -\pi < \tilde{T} + \tilde{X} < \pi \\ & \sinh[\tan((\tilde{T} - \tilde{X})/2)]\ \sinh[\tan((\tilde{T} + \tilde{X})/2)] < 1. \end{aligned} \tag{9.81}$$

The picture of $(\mathscr{M}, \boldsymbol{g})$ in the $(\tilde{T}, \tilde{X})$ plane is shown in Fig. 9.11. We shall call it a ***regular Carter-Penrose diagram*** of Schwarzschild spacetime. As the singular Carter-Penrose diagram of Fig. 9.10, it has the property to display the radial null geodesics as straight lines with slope $\pm 45°$, since $\tilde{U}$ (resp. $\tilde{V}$) is a function of U only (resp. V only) [cf.

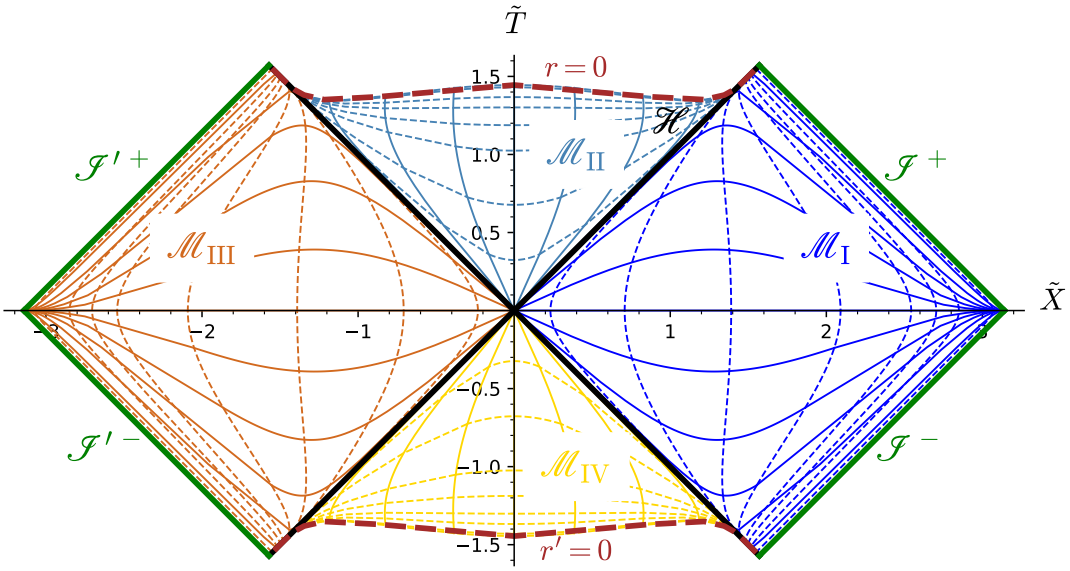


Figure 9.11: Carter-Penrose diagram of the Schwarzschild spacetime based on the Frolov-Novikov coordinates. Solid curves denote the same hypersurfaces of constant Schwarzschild-Droste coordinate t as in Fig. 9.10: in region $\mathscr{M}_{\rm I}$, from the $\hat{X}$-axis to the top: $t=0$, $2m$, $5m$, $10m$, $20m$ and $50m$; in region $\mathscr{M}_{\rm II}$, from the $\hat{T}$-axis to the right: $t=0$, $2m$, $5m$, $10m$, $20m$ and $50m$, Dashed curves denote the same hypersurfaces of constant Schwarzschild-Droste coordinate r as in Fig. 9.10: in region $\mathscr{M}_{\rm I}$, from the left to the right: $r=2.01m$, $2.1m$, $2.5m$, $4m$, $8m$, $12m$, $20m$ and $100m$; in region $\mathscr{M}_{\rm II}$, from the bottom to the top: $r=1.98m$, $1.9m$, $1.7m$, $1.5m$, $1.25m$, m, $0.5m$ and $0.1m$. The color code is the same as in Figs. 9.8 and 9.10. Contrary to the Carter-Penrose of Fig. 9.10, this one is associated to a regular conformal completion at null infinity of Schwarzschild spacetime. *[Figure generated by the notebook D.3.18]*

Eq. (9.66)]. In particular, the bifurcate Killing horizon and the Schwarzschild horizon are defined by:

$$\mathscr{H}: \quad \tilde{U}=0 \quad \text{or} \quad \tilde{V}=0 \iff \tilde{T}=\tilde{X} \quad \text{or} \quad \tilde{T}=-\tilde{X} \tag{9.82a}$$

$$\mathscr{H}: \quad \tilde{U}=0 \quad \text{and} \quad \tilde{V}>0 \iff \tilde{T}=\tilde{X} \quad \text{and} \quad \tilde{T}>0. \tag{9.82b}$$

These relations follow immediately from (9.45), (9.66) and (9.80).

At first sight, the main difference with the "standard" Carter-Penrose diagram of Fig. 9.10 is the more complicated shape of the boundary around the $\tilde{T}$-axis. This follows from the third condition in (9.81), which is more involved than the third condition in (9.53). Actually this boundary corresponds to the curvature singularity limit $r\to 0$ or $r'\to 0$. Indeed, from Eq. (9.43), we have

$$r=0 \iff UV=1. \tag{9.83}$$

In terms of the coordinates $(\hat{U},\hat{V})$, we have then [cf. Eq. (9.48)]

$$\begin{aligned} r=0 &\iff \tan\hat{U}\tan\hat{V}=1 \iff \sin\hat{U}\sin\hat{V}=\cos\hat{U}\cos\hat{V} \\ &\iff \cos\hat{U}\cos\hat{V}-\sin\hat{U}\sin\hat{V}=0 \iff \cos(\hat{U}+\hat{V})=0 \iff \hat{U}+\hat{V}=\pm\frac{\pi}{2}. \end{aligned}$$

Since $\hat{T} = \hat{U} + \hat{V}$, we get the simple relation

$$r = 0 \iff \hat{T} = \pm\frac{\pi}{2}. \tag{9.84}$$

On the contrary, in terms of the coordinates $(\tilde{U}, \tilde{V})$, Eq. (9.83) becomes [cf. Eq. (9.66)]

$$r = 0 \iff \sinh(\tan\tilde{U})\sinh(\tan\tilde{V}) = 1,$$

which yields to the complicated formula

$$r = 0 \iff \sinh[\tan((\tilde{T} - \tilde{X})/2)]\ \sinh[\tan((\tilde{T} + \tilde{X})/2)] = 1. \tag{9.85}$$

This explains the more complex boundary of Fig. 9.11 diagram with respect to Fig. 9.10 diagram.

Remark 3: The shape of the Carter-Penrose diagram in Frolov & Novikov's book (Fig. 5.2 of Ref. [110]; see also Fig. 10.6 of Ref. [111]) differs slightly from the diagram obtained here (Fig. 9.11). This is because the coordinates used by Frolov & Novikov are constructed from the Szekeres' version (up to a factor 2) of Kruskal-Szekeres coordinates: $T' = T/\sqrt{e}$ and $X' = X/\sqrt{e}$ (cf. the historical note on p. 231). Accordingly, in Frolov & Novikov's version, one shall replace the 1 in the right-hand side of Eq. (9.85) by $1/e$, yielding to a different shape of the boundary $r = 0$.

Remark 4: As noticed by Frolov and Novikov [110] (see their Sec. 5.1.3), one can perform some coordinate transformation from $(\tilde{U}, \tilde{V})$ to get a Carter-Penrose diagram with a straight line for the boundary $r = 0$.

Besides the shape of the boundary $r = 0$, another difference between the Carter-Penrose diagram based on Frolov-Novikov coordinates (Fig. 9.11) and the "standard" diagram of Fig. 9.10 is that the $t =$ const hypersurfaces of the former (solid curves in Fig. 9.11) are all tangent to the horizontal axis when $\tilde{X} \to \pm\pi$. On the contrary, the same hypersurfaces in Fig. 9.10 reach the point $(\hat{T}, \hat{X}) = (0, \pm\pi)$ with a finite slope. We note that in this respect, the Carter-Penrose diagram of Fig. 9.11 is similar to the conformal diagram of Minkowski spacetime, as shown in Fig. 4.3, and therefore display correctly the asymptotic flatness structure. The failure of diagram of Fig. 9.10 to reproduce this behavior reflects the fact that the coordinates $(\hat{T}, \hat{X})$ are singular on the boundary, as discussed in Sec. 9.4.2.

9.4.4 Black hole and white hole regions

$(\tilde{\mathscr{M}}, \tilde{\boldsymbol{g}})$ is not only a conformal completion of the maximally extended Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$, as established above, but it is a *conformal completion at null infinity*, in the sense defined in Sec. 4.3.1. Indeed, we can rewrite the boundary $\mathscr{I}$ of $\tilde{\mathscr{M}}$ expressed by Eq. (9.77) as

$$\mathscr{I} = \mathscr{I}_*^+ \cup \mathscr{I}_*^-, \tag{9.86}$$

with

$$\mathscr{I}_*^+ := \mathscr{I}^+ \cup \mathscr{I}'^+ \qquad \text{and} \qquad \mathscr{I}_*^- := \mathscr{I}^- \cup \mathscr{I}'^-. \tag{9.87}$$

It is clear that $\mathscr{I}_*^+$ is never intersected by any past-directed causal curve and that $\mathscr{I}_*^-$ is never intersected by any future-directed causal curve (cf. Fig. 9.11). Hence the writing (9.86) agrees with Eq. (4.29) in the definition of a conformal completion at null infinity. $\mathscr{I}_*^+$ is then the future null infinity of $(\mathscr{M}, \boldsymbol{g})$ and $\mathscr{I}_*^-$ its past null infinity.

We are thus in position to apply the definitions of a black hole and a white hole given in Sec. 4.4.2. The causal past of $\mathscr{I}_*^+$ is $J^-(\mathscr{I}_*^+) = \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III} \cup \mathscr{M}_{\rm IV}$ (cf. Fig. 9.11). In view of the definition (4.35), it follows that

> The maximal extension of Schwarzschild spacetime admits a black hole region, the interior of which is $\mathscr{M}_{\rm II}$. The black hole event horizon is the part of the bifurcate Killing horizon $\hat{\mathscr{H}}$ (cf. Sec. 9.3.3) that has $\tilde{T} > 0$.

The black hole region has thus the same interior $\mathscr{M}_{\rm II}$ as the black hole region of the Schwarzschild spacetime $(\mathscr{M}_{\rm IEF}, \boldsymbol{g})$ considered in Chap. 6. It differs only a larger boundary: it is not reduced to the Schwarzschild horizon $\mathscr{H}$, but contains the part $\tilde{X} < 0$ and $\tilde{T} > 0$ of $\hat{\mathscr{H}}$.

The novelty with respect to the original Schwarzschild spacetime $(\mathscr{M}_{\rm IEF}, \boldsymbol{g})$ is the existence of a white hole region. Indeed, the causal future of $\mathscr{I}_*^-$ is $J^+(\mathscr{I}_*^-) = \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm II} \cup \mathscr{M}_{\rm III}$ (cf. Fig. 9.11). In view of the definition (4.39), it follows that

> The maximal extension of Schwarzschild spacetime admits a white hole region, the interior of which is $\mathscr{M}_{\rm IV}$. The corresponding past event horizon is the part of the bifurcate Killing horizon $\hat{\mathscr{H}}$ (cf. Sec. 9.3.3) that has $\tilde{T} < 0$.

9.5 Einstein-Rosen bridge

To get some insight on the maximally extended Schwarzschild spacetime $(\mathscr{M}, \boldsymbol{g})$, let us examine the geometry of a slice of constant Kruskal-Szekeres time T.

9.5.1 Hypersurfaces of constant Kruskal-Szekeres time

Let Σ_{T_0} be a hypersurface of $\mathscr{M}$ defined in terms of the global Kruskal-Szekeres coordinates (T, X, θ, φ) by $T = T_0$, where $T_0 \in \mathbb{R}$ is a constant (cf. Fig. 9.12). The 3-tuple $(x^i) = (X, \theta, \varphi)$ is then a coordinate system on Σ_{T_0} subject to the constraint expressed in (9.32):

$$X^2 > T_0^2 - 1. \tag{9.88}$$

Consequently

- if $|T_0| < 1$, the hypersurface Σ_{T_0} is connected and diffeomorphic to $\mathbb{R} \times \mathbb{S}^2$, the coordinate X spanning $\mathbb{R}$ and (θ, φ) spanning $\mathbb{S}^2$.
- if $|T_0| \geq 1$, Σ_{T_0} has two connected components, defined by $X < -\sqrt{T_0^2 - 1}$ and $X > \sqrt{T_0^2 - 1}$ respectively (cf. Fig. 9.12). Each of them is diffeomorphic to $\mathbb{R} \times \mathbb{S}^2$.

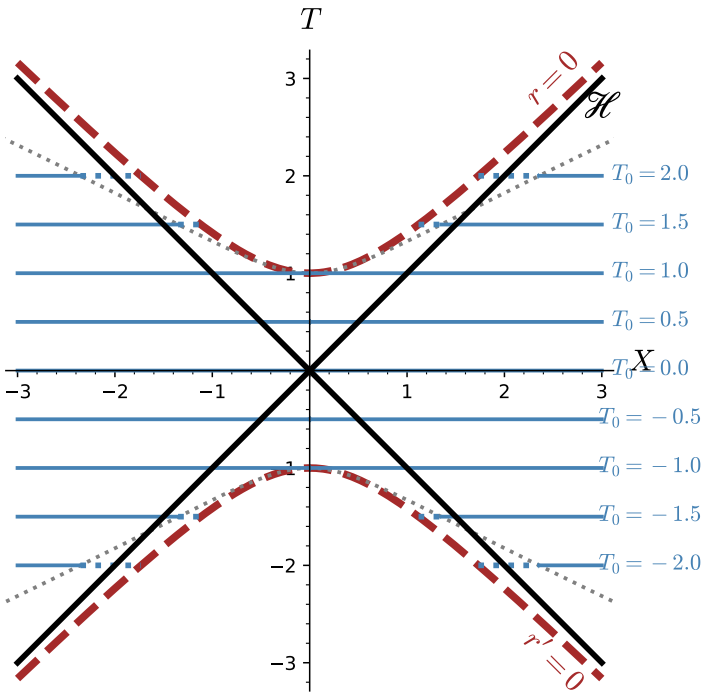


Figure 9.12: Kruskal diagram with the hypersurfaces Σ_{T_0} (defined by $T = T_0 = \text{const}$) as blue horizontal lines. For $|T_0| > 1$, the dotted part of Σ_{T_0} corresponds to a region that cannot be embedded isometrically in the Euclidean space. When T_0 varies, the limit of these regions form the grey dotted curve. *[Figure generated by the notebook D.3.19]*

For future convenience, we split Σ_{T_0} in two disjoint parts, according to the sign of X:

$$\Sigma_{T_0}^+ = \{p \in \Sigma_{T_0}, \quad X(p) \geq 0\} \quad \text{and} \quad \Sigma_{T_0}^- = \{p \in \Sigma_{T_0}, \quad X(p) < 0\}\,. \tag{9.89}$$

For $|T_0| < 1$, there is a slight asymmetry between the two parts: $\Sigma_{T_0}^+$ is a manifold with boundary (cf. Sec. A.2.2), the boundary corresponding to $X = 0$, while $\Sigma_{T_0}^-$ is not. For $|T_0| \geq 1$, $\Sigma_{T_0}^+$ and $\Sigma_{T_0}^-$ are nothing but the two connected components of Σ_{T_0}.

The geometry of Σ_{T_0} is defined by the metric $\boldsymbol{\gamma}$ induced on it by $\boldsymbol{g}$:

$$\gamma_{ij} \mathrm{d}x^i \mathrm{d}x^j = \frac{32m^3}{r} \mathrm{e}^{-r/2m}\, \mathrm{d}X^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right), \tag{9.90}$$

where r is the function of X defined by

$$r = r(X) = 2m\tilde{W}_0(X^2 - T_0^2). \tag{9.91}$$

The line element (9.90) is obtained by setting $T = T_0$ and $\mathrm{d}T = 0$ in (9.33). Since $r > 0$, the metric (9.90) is clearly positive definite, i.e. $\boldsymbol{\gamma}$ is a Riemannian metric and Σ_{T_0} is a spacelike hypersurface.

The graph of the function $r(X)$ is shown in Fig. 9.13. Once restricted to positive (resp. negative) values of X, this function is a bijection $(X_0, +\infty) \to (r_0, +\infty)$ (resp.

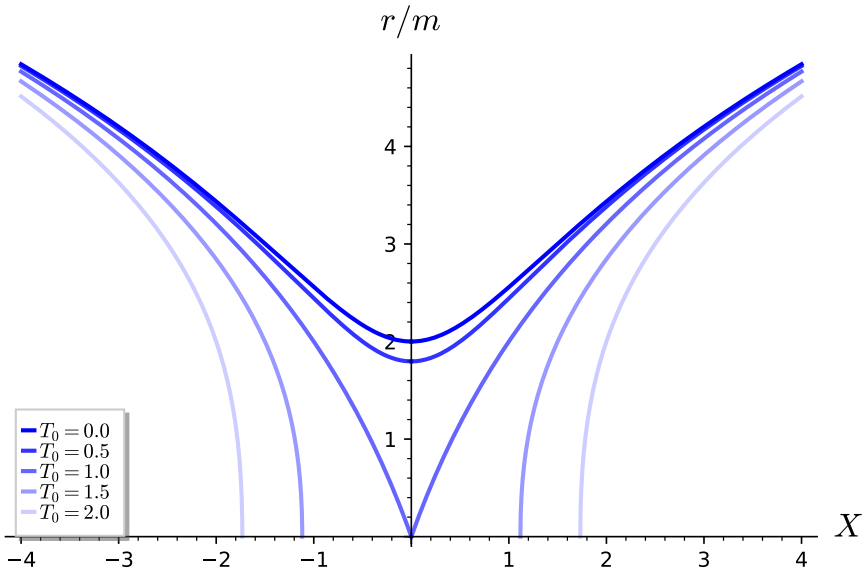


Figure 9.13: Function $r = r(X)$ on the hypersurface Σ_{T_0}, for the same values of T_0 as in Fig. 9.12. *[Figure generated by the notebook D.3.19]*

$(-\infty, -X_0) \to (r_0, +\infty)$), where

$$X_0 = \begin{cases} 0 & \text{if} \quad |T_0| < 1 \\ \sqrt{T_0^2 - 1} & \text{if} \quad |T_0| \geq 1. \end{cases} \quad \text{and} \quad r_0 = \begin{cases} 2m\tilde{W}_0(-T_0^2) & \text{if} \quad |T_0| < 1 \\ 0 & \text{if} \quad |T_0| \geq 1. \end{cases} \tag{9.92}$$

The inverse of this bijection is[5]

$$X = X(r) = \pm\sqrt{\mathrm{e}^{r/2m}\left(\frac{r}{2m} - 1\right) + T_0^2}, \tag{9.93}$$

with the $+$ sign on $\Sigma_{T_0}^+$ and the $-$ sign on $\Sigma_{T_0}^-$.

We may use the above bijection to introduce coordinates (r, θ, φ) instead of (X, θ, φ) on each of the two regions $\Sigma_{T_0}^+$ and $\Sigma_{T_0}^-$. Differentiating (9.93) leads to

$$\mathrm{d}X = \pm \frac{r\mathrm{e}^{r/2m}}{8m^2\sqrt{\mathrm{e}^{r/2m}\left(\frac{r}{2m} - 1\right) + T_0^2}}\, \mathrm{d}r.$$

Substituting in (9.90) we get the expression of the metric on Σ_{T_0} in terms of the coordinates $(x^i) = (r, \theta, \varphi)$:

$$\gamma_{ij}\, \mathrm{d}x^i \mathrm{d}x^j = \left(1 - \frac{2m}{r}\left(1 - T_0^2 \mathrm{e}^{-r/2m}\right)\right)^{-1} \mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right). \tag{9.94}$$

[5] Let us recall that $\tilde{W}_0^{-1}(x) = \mathrm{e}^x(x-1)$.

Remark 1: As a check of the above formula, we notice that for $T_0 = 0$ it reduces to the metric of a slice $t =$ const in Schwarzschild-Droste coordinates [set $\mathrm{d}t = 0$ in Eq. (6.15)]. This is correct since the positive-X half of the hypersurface $T = 0$ in Kruskal-Szekeres coordinates, i.e. Σ_0^+, coincides with the hypersurface $t = 0$ in Schwarzschild-Droste coordinates, as it can be seen by setting $T = 0$ in Eq. (9.20) (see also Fig. 9.12).

9.5.2 Isometric embedding in 3-dimensional Euclidean space

We may visualize the geometry of the spacelike hypersurface Σ_{T_0} via some isometric embedding of some 2-dimensional slice of it in the 3-dimensional Euclidean space $(\mathbb{R}^3, \boldsymbol{f})$, $\boldsymbol{f}$ being the standard flat (Euclidean) metric. By ***isometric embedding*** of a 2-dimensional Riemannian manifold $(\mathcal{S}, \boldsymbol{g})$ in $(\mathbb{R}^3, \boldsymbol{f})$, it is meant a *smooth embedding* $\Phi : \mathcal{S} \to \mathbb{R}^3$, as defined in Sec. A.2.7, such that the metric induced on $\Phi(\mathcal{S})$ by the Euclidean metric of $\mathbb{R}^3$ coincides with the original metric $\boldsymbol{g}$ on $\mathcal{S}$:

$$\forall p \in \mathcal{S}, \quad \forall (\boldsymbol{u}, \boldsymbol{v}) \in (T_p\mathcal{S})^2, \quad \boldsymbol{f}(\Phi^*(\boldsymbol{u}), \Phi^*(\boldsymbol{v})) = \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}), \tag{9.95}$$

where $\Phi^*(\boldsymbol{u})$ is the vector of $\mathbb{R}^3$ that is the "image of the vector $\boldsymbol{u}$ by Φ", i.e. the push-forward of $\boldsymbol{u}$ by Φ, as defined in Sec. A.4.2. Another phrasing of the isometry property (9.95) is: the pullback of $\boldsymbol{f}$ on $\mathcal{S}$ by Φ coincides with $\boldsymbol{g}$: $\Phi^*\boldsymbol{f} = \boldsymbol{g}$ [cf. Eq. (A.83)].

Taking into account the spherical symmetry of Σ_{T_0}, there is no loss of generality in choosing the equatorial plane $\theta = \pi/2$ as the 2-dimensional slice. We shall denote it by $\Sigma_{T_0}^{\mathrm{eq}}$. Coordinates on $\Sigma_{T_0}^{\mathrm{eq}}$ are $(x^a) = (X, \varphi)$, or on each of the two parts $\Sigma_{T_0}^{+,\mathrm{eq}}$ ($X \geq 0$) and $\Sigma_{T_0}^{-,\mathrm{eq}}$ ($X < 0$), $(x^a) = (r, \varphi)$. If $|T_0| < 1$, the topology of $\Sigma_{T_0}^{\mathrm{eq}}$ is $\mathbb{R} \times \mathbb{S}^1$, i.e. that of a cylinder, while for $|T_0| \geq 1$, it has two connected components, $\Sigma_{T_0}^{+,\mathrm{eq}}$ and $\Sigma_{T_0}^{-,\mathrm{eq}}$, each of them having the topology of a cylinder.

The metric induced by $\boldsymbol{g}$ on $\Sigma_{T_0}^{\mathrm{eq}}$, $\boldsymbol{q}$ say, is obtained by setting $\theta = \pi/2$ and $\mathrm{d}\theta = 0$ in Eq. (9.94):

$$q_{ab}\,\mathrm{d}x^a\mathrm{d}x^b = \left(1 - \frac{2m}{r}\left(1 - T_0^2 \mathrm{e}^{-r/2m}\right)\right)^{-1} \mathrm{d}r^2 + r^2\mathrm{d}\varphi^2. \tag{9.96}$$

Given the invariance in φ, it is quite natural to embed $(\Sigma_{T_0}^{\mathrm{eq}}, \boldsymbol{q})$ as a surface of revolution in the Euclidean space $(\mathbb{R}^3, \boldsymbol{f})$. Describing $\mathbb{R}^3$ with cylindrical coordinates $(x^i) = (r, z, \varphi)$, the Euclidean metric $\boldsymbol{f}$ is

$$f_{ij}\,\mathrm{d}x^i\mathrm{d}x^j = \mathrm{d}r^2 + \mathrm{d}z^2 + r^2\mathrm{d}\varphi^2. \tag{9.97}$$

A surface of revolution $\mathcal{S}$ in $\mathbb{R}^3$ is described by an equation of the type $z = Z(r)$. On such a surface, one has therefore $\mathrm{d}z = Z'(r)\,\mathrm{d}r$, so that the metric $\boldsymbol{h}$ induced by $\boldsymbol{f}$ on it is

$$h_{ab}\,\mathrm{d}x^a\mathrm{d}x^b = \left(1 + Z'(r)^2\right)\mathrm{d}r^2 + r^2\mathrm{d}\varphi^2. \tag{9.98}$$

Comparing (9.98) with (9.96), we see that a possible isometric embedding of $\Sigma_{T_0}^{\mathrm{eq}}$ into $\mathbb{R}^3$ is

$$\begin{array}{rccl} \Phi : & \Sigma_{T_0}^{\mathrm{eq}} & \longrightarrow & \mathbb{R}^3 \\ & (X, \varphi) & \longmapsto & (r, z, \varphi) = (r(X), \pm Z(r(X)), \varphi)\,, \end{array} \tag{9.99}$$

with (i) the sign $\pm$ being $+$ on $\Sigma_{T_0}^{+,\mathrm{eq}}$ and $-$ on $\Sigma_{T_0}^{-,\mathrm{eq}}$ and (ii) the function Z obeying

$$1 + Z'(r)^2 = \left(1 - \frac{2m}{r}\left(1 - T_0^2 \mathrm{e}^{-r/2m}\right)\right)^{-1}.$$

Thanks to Eq. (9.93), this expression can be recast as

$$Z'(r)^2 = \frac{1 - T_0^2 \mathrm{e}^{-r/2m}}{T_0^2 \mathrm{e}^{-r/2m} + \frac{r}{2m} - 1} = \frac{\mathrm{e}^{r/2m} - T_0^2}{X(r)^2}. \tag{9.100}$$

For $|T_0| < 1$, i.e. when $\Sigma_{T_0}^{\mathrm{eq}}$ is connected, the map (9.99) defines a smooth embedding if, and only if, at the boundary $X = 0$ between $\Sigma_{T_0}^{+,\mathrm{eq}}$ and $\Sigma_{T_0}^{-,\mathrm{eq}}$, the following holds:

$$Z(r(0)) = 0 \quad \text{and} \quad Z'(r(0)) = \infty. \tag{9.101}$$

The condition $Z(r(0)) = 0$ insures the continuity of the embedded surface $\Phi(\Sigma_{T_0}^{\mathrm{eq}})$, while $Z'(r(0)) = +\infty$ insures that it has a vertical tangent at the junction between $\Phi(\Sigma_{T_0}^{+,\mathrm{eq}})$ and $\Phi(\Sigma_{T_0}^{-,\mathrm{eq}})$, so that it is a smooth surface. Fortunately, the condition $Z'(r(0)) = \infty$ is automatically fulfilled from the second expression of $Z'(r)^2$ in (9.100): $Z'(r)^2$ clearly diverges at $X = 0$. Moreover, in order for the isometric embedding Φ to be well-defined, the right-hand side of (9.100) must be non-negative. Since the denominator of the last term is manifestly non-negative, the sign is determined by the numerator. Hence the condition $\mathrm{e}^{r/2m} \geq T_0^2$, or equivalently,

$$r \geq 4m \ln |T_0|. \tag{9.102}$$

For $|T_0| \leq 1$, this condition is always fulfilled, since $\ln |T_0| \leq 0$ and $r \geq 0$. For $|T_0| > 1$, it implies the existence of a minimal value of r,

$$r_{\mathrm{emb}}(T_0) := 4m \ln |T_0|, \tag{9.103}$$

such that the part of $\Sigma_{T_0}^{\mathrm{eq}}$ with $r < r_{\mathrm{emb}}(T_0)$ cannot be embedded isometrically in the Euclidean 3-space.

Remark 2: The above result should not be surprising since there is no guarantee that a 2-dimensional Riemannian manifold can be isometrically embedded in the 3-dimensional Euclidean space. The relevant theorem here is *Nash embedding theorem* [194], which states that any smooth Riemannian manifold of dimension n can be isometrically embedded in the Euclidean space $(\mathbb{R}^m, \boldsymbol{f})$, with $m \leq n(n+1)(3n+11)/2$. For $n = 2$, we get $m \leq 51$, so there is really no guarantee that $m = 3$ is sufficient...

Via (9.91) and the fact that the rescaled Lambert function $\tilde{W}_0$ is an increasing function (cf. Fig. 9.4), of inverse $F(x) = \mathrm{e}^x(x-1)$, the condition (9.102) can be turned into a condition on X:

$$|X| \geq X_{\mathrm{emb}}(T_0) := |T_0|\sqrt{2 \ln |T_0|}. \tag{9.104}$$

This limit is shown as the grey dotted curve in Fig. 9.12.

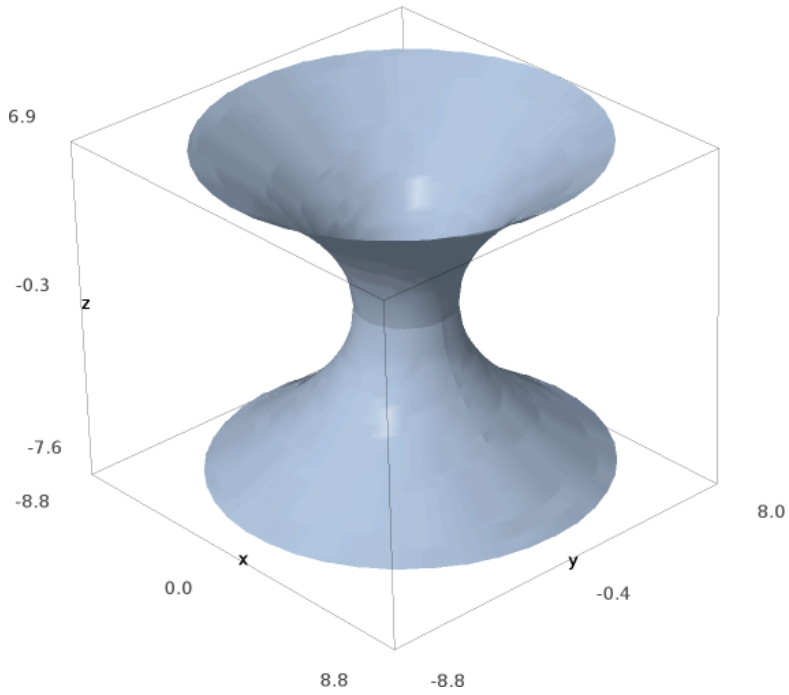


Figure 9.14: Flamm paraboloid: isometric embedding in the Euclidean $\mathbb{R}^3$ of the spacelike slice $T = 0$ and $\theta = \pi/2$ of Schwarzschild spacetime. The (x, y) coordinates are the standard Cartesian coordinates of $\mathbb{R}^3$ related to (r, φ) via $x = r\cos\varphi$ and $y = r\sin\varphi$. The labels are in units of m. *[Figure generated by the notebook D.3.19]*

Summarizing, the minimal value of r on the embedded surface is r_0 [cf. Eq. (9.92)] for $|T_0| \leq 1$ or $r_{\rm emb}(T_0)$ for $|T_0| > 1$:

$$r_{\rm min}(T_0) = \begin{cases} 2m\tilde{W}_0(-T_0^2) & \text{if} \quad |T_0| \leq 1 \\ 4m \ln |T_0| & \text{if} \quad |T_0| > 1. \end{cases} \tag{9.105}$$

Note that $r_{\rm min}(T_0)$ is a continuous function, with the peculiar values $r_{\rm min}(0) = 2m$ and $r_{\rm min}(1) = 0$. The embedding function $Z(r)$ is found by integration of $Z'(r)$, as given by (9.100), from $r_{\rm min}(T_0)$ to r:

$$Z(r) = 2m \int_{\frac{r_{\rm min}(T_0)}{2m}}^{\frac{r}{2m}} \sqrt{\frac{1 - T_0^2 \mathrm{e}^{-x}}{T_0^2 \mathrm{e}^{-x} + x - 1}}\, \mathrm{d}x. \tag{9.106}$$

The integral cannot be computed exactly in terms of elementary functions, except for $T_0 = 0$, where it reduces to

$$Z(r) = 2m \int_1^{\frac{r}{2m}} \frac{\mathrm{d}x}{\sqrt{x-1}} \qquad (T_0 = 0).$$

Hence

$$\boxed{Z(r) = 4m\sqrt{\frac{r}{2m} - 1}} \qquad (T_0 = 0). \tag{9.107}$$

According to (9.99), the whole surface $\Phi(\Sigma^{\rm eq}_{T_0})$ is obtained by considering $z = -Z(r)$ as well. The surface equation in terms of the cylindrical coordinates (r, z, φ) of $\mathbb{R}^3$ is then $z^2 = 16m^2(r/2m - 1)$, or

$$\boxed{z^2 = 8m(r - 2m)}. \tag{9.108}$$

We recognize the equation of a paraboloid of revolution around the z-axis. It is known as ***Flamm paraboloid*** [107] and is depicted in Fig. 9.14. Its topology is clearly that of a cylinder ($\mathbb{R} \times \mathbb{S}^1$). The geometry is different though: from top to bottom, the radius of the "cylinder" decreases to a minimal value, $r_{\rm min} = 2m$, and then increases. The "neck" around $r = r_{\rm min}$, or equivalently $X = 0$, is called the ***Einstein-Rosen bridge*** [98]. Contemplating the slice $T = 0$ in the Kruskal diagram of Fig. 9.12, we realize that this "bridge" connects the two asymptotically flat regions $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$. The Einstein-Rosen bridge is also called the ***Schwarzschild wormhole***. However, it is not a *traversable* wormhole: it is clear from the Kruskal diagram (Figs. 9.8 and 9.12) or the Carter-Penrose diagram (Fig. 9.11) that no timelike or null worldline can go from $\mathscr{M}_{\rm I}$ to $\mathscr{M}_{\rm III}$.

When $T_0 \neq 0$, the integral in (9.106) has to be computed numerically (see Sec. D.3.19 for the computation with SageMath). The resulting embedded surfaces $\Phi(\Sigma^{\rm eq}_{T_0})$ are shown in Fig. 9.15 for the values of T_0 involved in the Kruskal diagram of Fig. 9.12. When T_0 increases from 0, the "neck" becomes thinner and thinner. At $T_0 = 1$, it ceases to be connected. As mentioned above, for $T_0 > 1$, the surface $\Sigma^{\rm eq}_{T_0}$ can no longer be entirely isometrically embedded in the Euclidean 3-space. Hence the holes in the central parts of the surfaces for $T_0 = 1.5$ and $T_0 = 2$. These holes correspond to the dotted segments in Fig. 9.12 and their radii are given by Eq. (9.103). Note that the tangents to the embedded surfaces at their inner boundaries are horizontal.

The evolution of Σ_{T_0} as T_0 increases is not surprising if one remembers that the Kruskal-Szekeres time coordinate T is not associated with any timelike Killing vector of Schwarzschild spacetime. The sequence shown in Fig. 9.15 can be thought of as representing the dynamics of the Schwarzschild wormhole, in particular its "pinching-off" at $T_0 = 1$, which forbids any traveler to go through it.

Remark 3: We have restricted ourselves to slices $T =$ const of Schwarzschild spacetime, with the isometric embedding limitation for $|T| > 1$. We refer the reader to Ref. [66] for more general slices and the corresponding embedding diagrams.

Remark 4: There are many inexact plots of embeddings of spatial sections of Schwarzschild spacetime in the literature, including renown textbooks. A first common error is to draw the two ends of the embedded surface as asymptotic to flat planes, which a paraboloid is not (the vertical distance between the two ends grows unbounded, as $\sqrt{r}$, cf. Eq. (9.107) and Fig. 9.16). This is correct from a topological point of view, but not from the geometrical one, i.e. the embedding depicted in this way is not an isometry. Probably this results from some confusion with asymptotic flatness: it is true that the metric (9.96) tends to a flat metric when $r \to +\infty$, reflecting the asymptotic flatness of Schwarzschild spacetime, but the associated curvature does not decay fast enough to allow the embedded surface to be tangent to a plane. A second error regards the embeddings for $|T_0| > 1$, which are depicted as variants of that $T_0 = 1$ (cf. Fig. 9.15), with two spikes at $r = 0$, simply pushed apart. However, as discussed above, the isometric embeddings with $T =$ const cannot reach the region near $r = 0$ for $|T_0| > 1$.

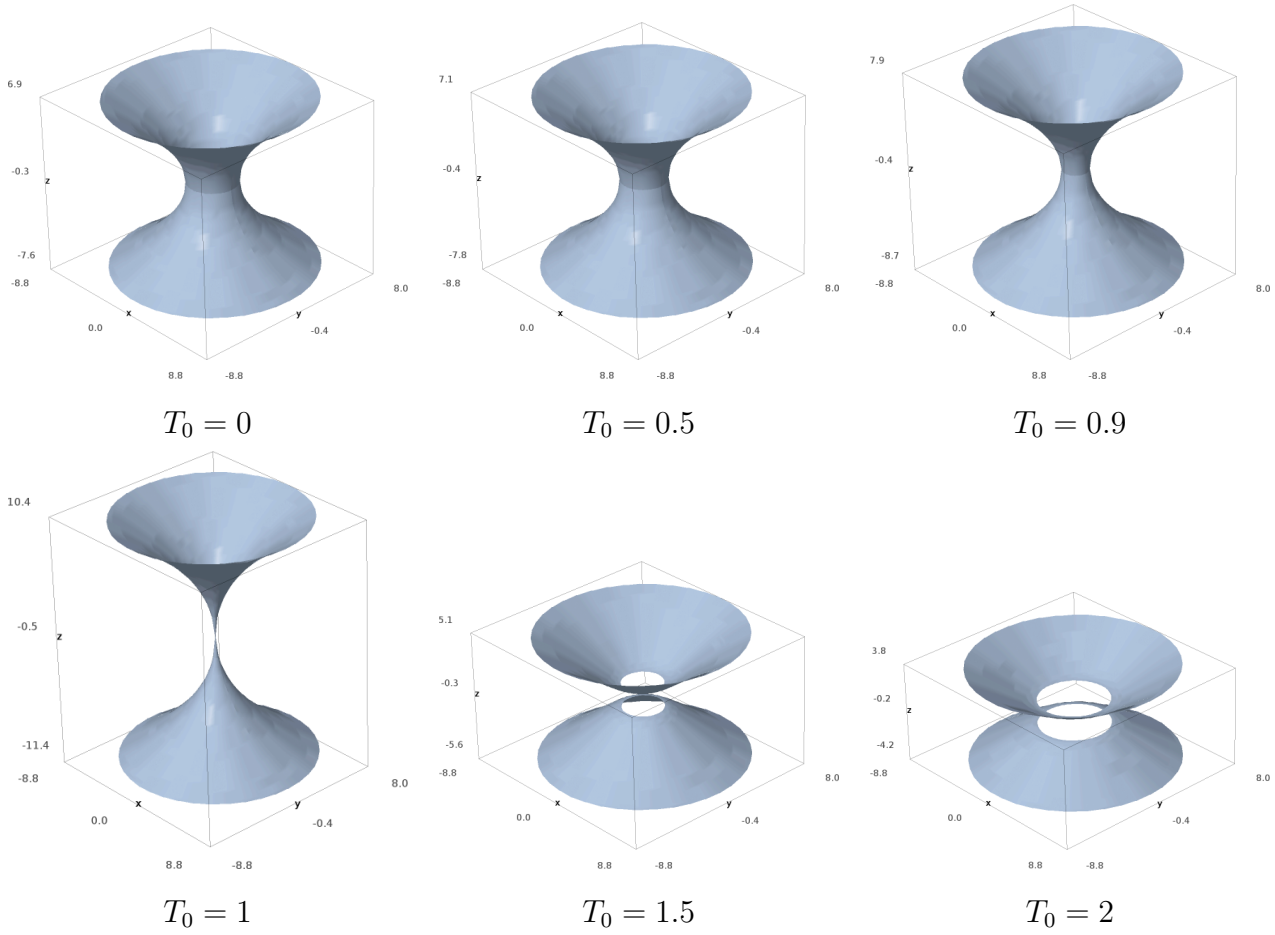


Figure 9.15: Sequence of isometric embeddings in the Euclidean space of spacelike slices of Schwarzschild spacetime defined by $T = T_0$ and $\theta = \pi/2$. The slices are those shown in the Kruskal diagram of Fig. 9.12 (except for $T_0 = 0.9$). The first embedding ($T_0 = 0$) is the Flamm paraboloid depicted in Fig. 9.14. In the disconnected case ($T_0 = 1.5$ and $T_0 = 2.0$), the distance between the upper and lower parts is arbitrary (chosen here to be $\Delta z = 1$). *[Figure generated by the notebook D.3.19]*

Historical note: In 1916, very soon after the publication of Schwarzschild solution [229], the Austrian physicist Ludwig Flamm (1885-1964) showed that the slice $t =$ const and $\theta = \pi/2$ in Schwarzschild-Droste coordinates (t, r, θ, φ) can be isometrically embedded in the Euclidean space as a paraboloid of revolution obeying Eq. (9.108) [107]. Let us recall that the positive-X part of the hypersurface $T = 0$ considered here coincides with the hypersurface $t = 0$ (cf. Remark 1 on p. 247). Although he draw the whole paraboloid (actually a parabola in a 2-dimensional plot — Fig. 2 of Ref. [107]), Flamm did not seem to have considered the negative-z part as physically relevant. In other words, he limited his considerations to $\mathscr{M}_{\rm I}$ and did not contemplate any bridge to the extension $\mathscr{M}_{\rm III}$.

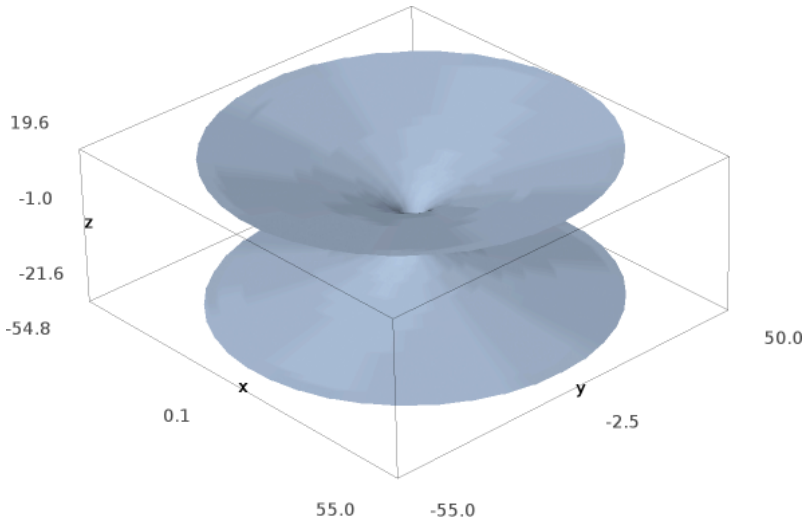


Figure 9.16: Same Flamm paraboloid as in Fig. 9.14 but seen from farther away. Despite being more and more flat, none of the two sheets is asymptotic to a plane. *[Figure generated by the notebook D.3.19]*

9.5.3 Isotropic coordinates

Let us consider a hypersurface of constant Schwarzschild-Droste time t in $\mathscr{M}_{\mathrm{I}}$. According to Eq. (9.19), this hypersurface obeys

$$T = \tanh\left(\frac{t}{4m}\right) X, \tag{9.109}$$

with $X > 0$, which implies that it is represented by a straight half-line from the origin in the Kruskal diagram (cf. Fig. 9.8). Similarly, a hypersurface of constant t' in $\mathscr{M}_{\mathrm{III}}$ obeys an equation identical to (9.109), except for t replaced by t' and $X < 0$ [cf. Eq. (9.35)]. Accordingly, for $t' = t$, the union of these two hypersurfaces forms a hypersurface of $\mathscr{M}$ ruled by Eq. (9.109), with $X < 0$ or $X > 0$. If we add the points $(T, X) = (0, 0)$ to it (i.e. the bifurcation sphere $\mathscr{S}$ (cf. Sec. 9.3.3), we obtain a connected hypersurface in which X takes all values in the range $(-\infty, +\infty)$. Let us call $\mathcal{S}_t$ this hypersurface. In other words, $\mathcal{S}_t$ is the hypersurface of $\mathscr{M}$ defined by Eq. (9.109) with $X \in \mathbb{R}$. Note that for $t = 0$, this hypersurface coincides with the hypersurface $T = 0$ introduced in Sec. 9.5.1: $\mathcal{S}_0 = \Sigma_0$. But for $t \neq 0$, $\mathcal{S}_t \neq \Sigma_T$.

There are two Schwarzschild-Droste coordinate systems on $\mathcal{S}_t$: (r, θ, φ) on $\mathcal{S}_t \cap \mathscr{M}_{\mathrm{I}}$ and (r', θ, φ) on $\mathcal{S}_t \cap \mathscr{M}_{\mathrm{III}}$, with both r and r' ranging $(2m, +\infty)$. Let us introduce on $\mathcal{S}_t$ a

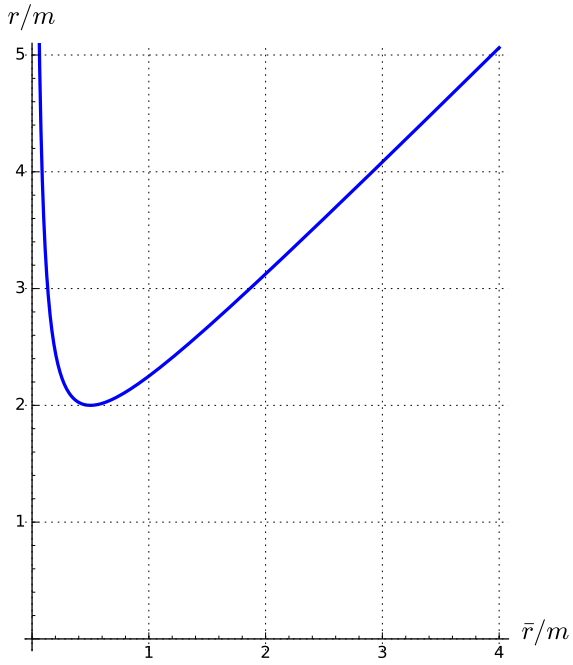


Figure 9.17: Areal radius r as a function of the isotropic coordinate $\bar{r}$.

third coordinate system $(x^{\bar{i}}) = (\bar{r}, \theta, \varphi)$ as follows:

$$\text{on } \mathcal{S}_t \cap \mathscr{M}_{\rm I}: \quad \bar{r} \in \left(\frac{m}{2}, +\infty\right), \quad r = \bar{r}\left(1 + \frac{m}{2\bar{r}}\right)^2 \tag{9.110a}$$

$$\iff \bar{r} = \frac{1}{2}\left(r - m + \sqrt{r(r-2m)}\right) \tag{9.110b}$$

$$\text{on } \mathcal{S}_t \cap \mathscr{M}_{\rm III}: \quad \bar{r} \in \left(0, \frac{m}{2}\right), \quad r' = \bar{r}\left(1 + \frac{m}{2\bar{r}}\right)^2 \tag{9.110c}$$

$$\iff \bar{r} = \frac{1}{2}\left(r' - m - \sqrt{r'(r'-2m)}\right) \tag{9.110d}$$

$$\text{on } \mathcal{S}_t \cap \mathscr{S}: \quad \bar{r} = \frac{m}{2}. \tag{9.110e}$$

The range of $\bar{r}$ is thus $(0, +\infty)$. The graph of the function $\bar{r} \mapsto \bar{r}(1 + m/(2\bar{r}))^2$ is depicted in Fig. 9.17. We can separate this graph in two parts: $\bar{r} \in (0, m/2)$ (the $\mathscr{M}_{\rm III}$ part) and $\bar{r} \in (m/2, +\infty)$ (the $\mathscr{M}_{\rm I}$ part). In each of these part, there is a one-to-one correspondence between $\bar{r}$ and r (or r'). Note that

$$\text{when } r \to +\infty, \quad \bar{r} \sim r \tag{9.111a}$$

$$\text{when } r' \to +\infty, \quad \bar{r} \sim \frac{m^2}{4r'}. \tag{9.111b}$$

When t varies, $\mathcal{S}_t$ constitute a foliation of

$$\mathscr{M}_{\rm iso} := \mathscr{M}_{\rm I} \cup \mathscr{S} \cup \mathscr{M}_{\rm III}. \tag{9.112}$$

This foliation is regular in both $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$, but is singular at the bifurcation sphere $\mathscr{S}$, since all the hypersurfaces $\mathcal{S}_t$ intersect there (cf. Fig. 9.8). We may then consider $(x^{\bar\alpha}) = (t, \bar r, \theta, \varphi)$ as a coordinate system on $\mathscr{M}_{\rm iso}$, which is regular on $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$, but is singular at $\mathscr{S}$, i.e. at $\bar r = m/2$. This system is called ***isotropic coordinates***.

From (9.110a), we get

$$\mathrm{d}r = \left(1 + \frac{m}{2\bar r}\right)\left(1 - \frac{m}{2\bar r}\right)\mathrm{d}\bar r$$

and

$$1 - \frac{2m}{r} = \left(\frac{1 - \frac{m}{2\bar r}}{1 + \frac{m}{2\bar r}}\right)^2 .$$

It is then immediate to deduce from (6.15) the expression of the metric tensor in terms of the isotropic coordinates $(x^{\bar\alpha}) = (t, \bar r, \theta, \varphi)$:

$$\boxed{g_{\bar\mu\bar\nu}\,\mathrm{d}x^{\bar\mu}\,\mathrm{d}x^{\bar\nu} = -\left(\frac{1 - \frac{m}{2\bar r}}{1 + \frac{m}{2\bar r}}\right)^2 \mathrm{d}t^2 + \left(1 + \frac{m}{2\bar r}\right)^4 \left[\mathrm{d}\bar r^2 + \bar r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)\right]} . \tag{9.113}$$

Since the relation between r' and $\bar r$ is identical to that between r and $\bar r$ [cf. Eqs. (9.110a) and (9.110c)], the above expression of $\boldsymbol{g}$ is valid on $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$. Note that all metric coefficients are regular on $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$ (except for the standard coordinate singularity of the spherical coordinates (θ, φ) for $\theta \in \{0, \pi\}$). On the contrary Eq. (9.113) yields $\det(g_{\bar\alpha\bar\beta}) = 0$ for $\bar r = m/2$, which reflects the fact that the isotropic coordinates are singular on the bifurcation sphere $\mathscr{S}$.

A remarkable feature of the line element (9.113) is that the spatial part is proportional to the line element of the flat metric $\boldsymbol{f}$ on the Euclidean 3-space:

$$f_{\bar i\bar j}\,\mathrm{d}x^{\bar i}\,\mathrm{d}x^{\bar j} = \mathrm{d}\bar r^2 + \bar r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right) . \tag{9.114}$$

In other words, the metric $\boldsymbol{\gamma}$ induced by $\boldsymbol{g}$ on $\mathcal{S}_t$ is *conformal* to the flat metric $\boldsymbol{f}$ (cf. Sec. 4.2.2):

$$\boldsymbol{\gamma} = \Psi^4\,\boldsymbol{f}, \tag{9.115}$$

with the conformal factor[6]

$$\Psi = 1 + \frac{m}{2\bar r}. \tag{9.116}$$

The conformally-flat feature explains the name *isotropic* given to the coordinates $(t, \bar r, \theta, \varphi)$.

Remark 5: Sometimes, isotropic coordinates are simply presented as coordinates deduced from the standard Schwarzschild-Droste ones by formula (9.110a). But much more than a mere change of the coordinate r to $\bar r$ is involved: the two coordinate systems do not cover the same part of the extended Schwarzschild spacetime: the Schwarzschild-Droste coordinates cover the region $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm II}$, while the isotropic coordinates cover the region $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$. In particular, the isotropic coordinates cannot be used to describe the black hole interior (i.e. $\mathscr{M}_{\rm II}$). This last feature can be inferred directly from the line element (9.113), which corresponds clearly to an everywhere timelike Killing vector $\boldsymbol{\partial}_t$ (because of the square in g_{tt}), while $\boldsymbol{\partial}_t$ is a spacelike vector field in $\mathscr{M}_{\rm II}$.

[6] Note a different convention with respect to Sec. 4.2.2: with respect to the latter, what would be called the *conformal factor* is the square of Ψ: $\Omega = \Psi^2$.

9.6 Physical relevance of the maximal extension

9.6.1 Naked singularity

Beside harboring a white hole (cf. Sec. 9.4.4), the maximal extension of Schwarzschild spacetime contains a ***naked singularity***, i.e. a curvature singularity that can be seen by arbitrarily far observers in the asymptotically flat regions $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$. Indeed, it is clear from the Carter-Penrose diagram of Fig. 9.11 that the past null cone of any event in $\mathscr{M}_{\rm I}$ or $\mathscr{M}_{\rm III}$ encounters the curvature singularity $r' = 0$ at the past boundary of $\mathscr{M}_{\rm IV}$. In other words, any observer in $\mathscr{M}_{\rm I}$ or $\mathscr{M}_{\rm III}$ can receive signals from this singularity. Contrary to the singularity $r = 0$ at the future boundary of $\mathscr{M}_{\rm II}$, it is not "clothed" by a black hole horizon, but by a white hole horizon, which does not prevent null geodesics from moving from the singularity to the observer.

In a given spacetime, a naked singularity constitutes a limitation to predictability, since one cannot compute what may come out of the singularity at any instant. That no "reasonable" physical process can generate a naked singularity is known as the (weak) *cosmic censorship conjecture*, first formulated by Penrose in 1969 [210] (see also Sec. 10.1 of Wald's textbook [256]).

9.6.2 Astrophysical relevance

We shall see in Chaps. 14 and 15 that, in spherical symmetry, the formation of a black hole by the astrophysical process of gravitational collapse yields a spacetime that contains some parts of the regions $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm II}$ of Schwarzschild spacetime, but that do not contain any part of $\mathscr{M}_{\rm III}$ nor $\mathscr{M}_{\rm IV}$. In particular, it does not contain any white hole nor any naked singularity. In other words, the maximal extension of Schwarzschild spacetime cannot be formed by gravitational collapse of a star or a cloud of matter. This, of course, does not exclude by itself the existence of (approximate) maximally extended Schwarzschild regions in our universe. Such regions could exist because the universe is born with them built in. For this reason, the maximal extension of the Schwarzschild Mspacetime is sometimes called the *eternal Schwarzschild black hole*. Moreover, there exist some solutions of Einstein equation describing a spherically symmetric ball of matter that is expanding, reaches some maximal extension and then collapses and whose exterior contains parts of regions $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm IV}$, and even $\mathscr{M}_{\rm III}$ (in which case, it is called a *semiclosed world*) [196] (see also Sec. 2.7.2 of Frolov & Novikov's textbook [110]).

9.6.3 Use in theoretical physics

Even if the astrophysical motivation for the Schwarzschild maximal extension is pretty weak, this spacetime plays some role in theoretical physics. For instance, it has been advanced that quantum entanglement between two black holes can be realized by an Einstein-Rosen bridge [181], a conjecture that is known as $ER = EPR$, where ER stands for *Einstein-Rosen* and EPR for *Einstein-Podolsky-Rosen*, from the famous EPR paradox involving entangled particles (see e.g. Ref. [175]).

Chapter 10

Kerr black hole

Contents

10.1 Introduction

Having studied the Schwarzschild black hole in the preceding chapters, we turn now to its rotating generalization: the Kerr black hole. The Kerr metric is arguably the most important solution of general relativity, largely because of the no-hair theorem, according to which all stationary black holes in the Universe are Kerr black holes (cf. Sec. 5.5).

In this chapter, the Kerr solution is first presented in terms of the standard Boyer-Lindquist coordinates and its basic properties are discussed (Sec. 10.2). Then Kerr coordinates are introduced in Sec. 10.3; contrary to Boyer-Lindquist coordinates, they are regular on the two Killing horizons of Kerr spacetime. The Kerr coordinates are also tied to one of the two congruences of null geodesics related to the spacetime conformal structure: the principal null geodesics, which are introduced in Sec. 10.4. The second congruence, that of the so-called *outgoing* principal null geodesics, provides the generators of the black hole event horizon, which is studied in Sec. 10.5. Then Sec. 10.6 focuses

on global quantities characterizing Kerr spacetime: mass, angular momentum and horizon area. Section 10.7 presents various standard families of observers in Kerr spacetime. Finally, Sec. 10.8 concludes this chapter by discussing the maximal analytical extension of Kerr spacetime and the concept of Cauchy horizon.

10.2 The Kerr solution

10.2.1 Expression in Boyer-Lindquist coordinates

The Kerr solution depends on two constant non-negative real parameters:

- the ***mass parameter*** $m > 0$, to be interpreted in Sec. 10.6.1 as the spacetime total mass;
- the ***spin parameter*** $a \geq 0$, to be interpreted in Sec. 10.6.2 as the reduced angular momentum $a = J/m$, J being the spacetime total angular momentum.

In this chapter, we focus on Kerr solutions for which

$$0 < a < m, \tag{10.1}$$

postponing the case $a = m$ to Chap. 13. The Kerr solution is usually presented in the so-called ***Boyer-Lindquist coordinates*** (t, r, θ, φ). Except for the standard singularities of the spherical coordinates (θ, φ) on $\mathbb{S}^2$ at $\theta \in \{0, \pi\}$, we may consider that the Boyer-Lindquist coordinates cover the manifold $\mathbb{R}^2 \times \mathbb{S}^2$, with t spanning $\mathbb{R}$, r spanning[1] $\mathbb{R}$, θ spanning $(0, \pi)$ and φ spanning $(0, 2\pi)$. Hence (t, r) is a Cartesian chart covering $\mathbb{R}^2$ and (θ, φ) is the standard spherical chart of $\mathbb{S}^2$.

In this section, we choose the spacetime manifold to be the open subset $\mathscr{M}_{\rm BL}$ of $\mathbb{R}^2 \times \mathbb{S}^2$ formed by the disjoint union of the following three components (cf. Fig. 10.1):

$$\mathscr{M}_{\rm BL} := \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm II} \cup \mathscr{M}_{\rm III}, \tag{10.2a}$$
$$\mathscr{M}_{\rm I} := \mathbb{R} \times (r_+, +\infty) \times \mathbb{S}^2 \tag{10.2b}$$
$$\mathscr{M}_{\rm II} := \mathbb{R} \times (r_-, r_+) \times \mathbb{S}^2 \tag{10.2c}$$
$$\mathscr{M}_{\rm III} := \mathbb{R} \times (-\infty, r_-) \times \mathbb{S}^2 \setminus \mathscr{R}, \tag{10.2d}$$

where

$$\boxed{r_+ := m + \sqrt{m^2 - a^2}} \quad \text{and} \quad \boxed{r_- := m - \sqrt{m^2 - a^2}} \tag{10.3}$$

and $\mathscr{R}$ is the subset of $\mathbb{R}^2 \times \mathbb{S}^2$ defined in terms of the Boyer-Lindquist coordinates (t, r, θ, φ) by

$$\mathscr{R} := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = 0 \text{ and } \theta(p) = \frac{\pi}{2} \right\}. \tag{10.4}$$

[1] This contrasts with r spanning only $(0, +\infty)$ for the standard spherical coordinates (r, θ, φ) on $\mathbb{R}^3$.

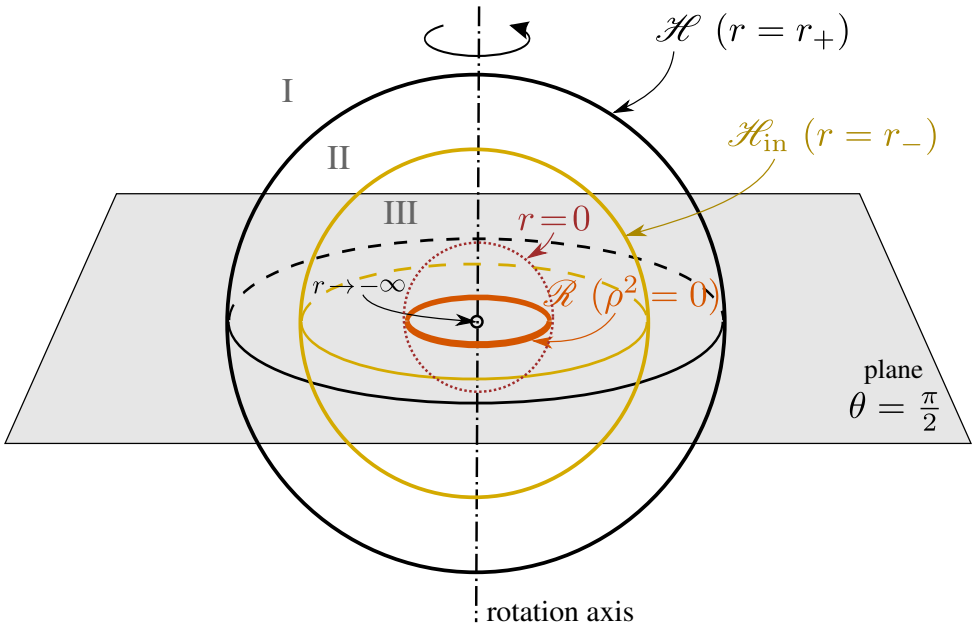


Figure 10.1: View of a section $t = \text{const}$ of $\mathbb{R}^2 \times \mathbb{S}^2$ in terms of the coordinates (R, θ, φ), with $R := \mathrm{e}^r$, so that the region $r \to -\infty$ is reduced to a single point at the centre of all the pictured spheres. Such coordinates have been introduced for pictorial purposes by O'Neill [200].

Remark 1: By construction, the points of $\mathbb{R}^2 \times \mathbb{S}^2$ obeying $r = r_-$ and $r = r_+$ are excluded from the spacetime manifold $\mathscr{M}_{\rm BL}$. In a latter stage (Sec. 10.3), we shall extend the spacetime manifold to include these points, so that the spacetime manifold will be $\mathscr{M} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathscr{R}$. Even latter on, after having noticed that $(\mathscr{M}, \boldsymbol{g})$ is not geodesically complete, we shall extend the spacetime manifold further and discuss the maximal analytic extension (Sec. 10.8).

Remark 2: One shall stress that the points having $r = 0$ are *not* special points in $\mathbb{R}^2 \times \mathbb{S}^2$ spanned by the coordinates (t, r, θ, φ). In particular, $(t, r) = (t_0, 0)$, where t_0 is a constant, defines a regular sphere $\mathscr{S}_0$ diffeomorphic to $\mathbb{S}^2$. On the spacetime manifold $\mathscr{M}_{\rm BL}$, the points $r = 0$ are part of $\mathscr{M}_{\rm III}$ and because of the exclusion of $\mathscr{R}$ in the definition (10.2d), $(t, r) = (t_0, 0)$ defines a sphere minus its equator (cf. Fig. 10.1, where the equator is the tick orange circle).

Note that thanks to the constraint (10.1), r_+ and r_- are well-defined and obey

$$0 < r_- < m < r_+ < 2m. \tag{10.5}$$

Note also that $\mathscr{R}$ is spanned by the coordinates (t, φ) and is diffeomorphic to the 2-dimensional cylinder $\mathbb{R} \times \mathbb{S}^1$:

$$\mathscr{R} \simeq \mathbb{R} \times \mathbb{S}^1. \tag{10.6}$$

This is so because $r = 0$ is *not* a peculiar value of r in $\mathbb{R}^2 \times \mathbb{S}^2$ (cf. remark 2 above). In view of Eqs. (10.2b)-(10.2d) and (10.4), it is clear that the various connected components

of $\mathscr{M}_{\rm BL}$ are defined in terms of the Boyer-Lindquist coordinates (t, r, θ, φ) by

$$\forall p \in \mathscr{M}_{\rm BL}, \quad p \in \mathscr{M}_{\rm I} \iff r(p) > r_+ \tag{10.7a}$$

$$p \in \mathscr{M}_{\rm II} \iff r_- < r(p) < r_+ \tag{10.7b}$$

$$p \in \mathscr{M}_{\rm III} \iff r(p) < r_- \text{ and } \left(r(p) \neq 0 \text{ or } \theta(p) \neq \frac{\pi}{2}\right). \tag{10.7c}$$

The ***Kerr metric*** is defined by the following components in terms of the Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$:

$$\boxed{\begin{aligned} g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu = & -\left(1 - \frac{2mr}{\rho^2}\right)\mathrm{d}t^2 - \frac{4amr\sin^2\theta}{\rho^2}\,\mathrm{d}t\,\mathrm{d}\varphi + \frac{\rho^2}{\Delta}\,\mathrm{d}r^2 \\ & + \rho^2 \mathrm{d}\theta^2 + \left(r^2 + a^2 + \frac{2a^2 mr\sin^2\theta}{\rho^2}\right)\sin^2\theta\,\mathrm{d}\varphi^2, \end{aligned}} \tag{10.8}$$

with

$$\boxed{\rho^2 := r^2 + a^2\cos^2\theta} \tag{10.9}$$

and

$$\boxed{\Delta := r^2 - 2mr + a^2 = (r - r_-)(r - r_+)}. \tag{10.10}$$

Note that on $\mathscr{M}_{\rm BL}$, $\rho \neq 0$ and $\Delta \neq 0$ (by construction of $\mathscr{M}_{\rm BL}$!), so that the metric components (10.8) are regular in $\mathscr{M}_{\rm BL}$, except for the standard singularities of the spherical coordinates (θ, φ).

By means of a computer algebra system (cf. the notebook D.4.1), it is easy to check that

$(\mathscr{M}_{\rm BL}, \boldsymbol{g})$ with $\boldsymbol{g}$ given by (10.8), is a solution of Einstein equation (1.38) in vacuum ($\boldsymbol{T} = 0$) and with a vanishing cosmological constant ($\Lambda = 0$).

Historical note: The Kerr solution has been found by the New Zealand mathematician Roy P. Kerr (then at the University of Texas at Austin) in the spring of 1963 [164]. Kerr was searching for *algebraically special* metrics, i.e. metrics whose Weyl conformal curvature tensor admits a *doubly degenerate principal null direction* (to be defined in Sec. 10.4 below), in the case where the principal null congruence has a non-vanishing twist (or "rotation"). The special case of vanishing twist (i.e. hypersurface-orthogonal congruence) had been treated by Ivor Robinson and Andrzej Trautman in 1962 [225]. Kerr used Cartan's structure equations in a null tetrad to manipulate the Einstein equation; he obtained the solution in coordinates different from the Boyer-Lindquist ones, namely *Kerr coordinates*, to be discussed in Sec. 10.3.1 (cf. the historical note page 268). For more details about this fantastic discovery, see the account by Kerr himself in Ref. [165]. The Boyer-Lindquist coordinates have been introduced in 1966 by Robert H. Boyer (see the historical note on page 117) and Richard W. Lindquist [30] (compare Eq. (2.13) of Ref. [30] with Eq. (10.8) above, keeping in mind that Boyer and Lindquist used $-a$ instead of a, following Kerr's convention in the discovery article [164], cf. the historical note on p. 268).

10.2.2 Basic properties

Various properties of the Kerr metric are immediate:

- For $r \to +\infty$ or $r \to -\infty$, one has $\rho^2 \sim r^2$ and $\rho^2/\Delta \sim (1-2m/r)^{-1}$, and $4amr/\rho^2\, \mathrm{d}t\, \mathrm{d}\varphi \sim 4am/r^2\, \mathrm{d}t\, r\mathrm{d}\varphi$, so that the metric (10.8) becomes

$$g_{\mu\nu}\, \mathrm{d}x^\mu \mathrm{d}x^\nu \simeq -\left(1-\frac{2m}{r}\right) \mathrm{d}t^2 + \left(1-\frac{2m}{r}\right)^{-1} \mathrm{d}r^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right) + O\left(\frac{1}{r^2}\right) \tag{10.11}$$

  For $r > 0$, we recognize the Schwarzschild metric expressed in Schwarzschild-Droste coordinates [cf. Eq. (6.15)]. For $r < 0$, the change of coordinate $r' = -r$ leads also to the Schwarzschild metric but with a negative mass parameter $m' = -m$. Hence, the Kerr metric has (at least) two asymptotically flat ends: one in $\mathscr{M}_{\mathrm{I}}$ for $r \to +\infty$ and one in $\mathscr{M}_{\mathrm{III}}$ for $r \to -\infty$.

- Since in (10.8) all the metric components $g_{\alpha\beta}$ are independent from t and φ, the spacetime $(\mathscr{M}_{\mathrm{BL}}, \boldsymbol{g})$ admits two isometries, generated by the Killing vectors

$$\boxed{\boldsymbol{\xi} := \boldsymbol{\partial}_t} \quad \text{and} \quad \boxed{\boldsymbol{\eta} := \boldsymbol{\partial}_\varphi}. \tag{10.12}$$

  Since t spans $\mathbb{R}$, the isometry group generated by $\boldsymbol{\xi}$ is clearly the translation group $(\mathbb{R}, +)$. Moreover, in view of (10.11), we have $\boldsymbol{\xi}\cdot\boldsymbol{\xi} = g_{tt} < 0$ as $r \to +\infty$, which means that the Killing vector $\boldsymbol{\xi}$ is asymptotically timelike. Given the definition of stationarity stated in Sec. 5.2.1, we conclude that the Kerr spacetime is stationary. On the other side, since φ is an azimuthal coordinate on $\mathbb{S}^2$, the isometry group generated by $\boldsymbol{\eta}$ is the rotation group $\mathrm{SO}(2) = \mathrm{U}(1)$. Hence, the Kerr spacetime is axisymmetric.

- When $a \neq 0$, as we have assumed in (10.1), the Kerr spacetime is not static, since the stationary Killing vector $\boldsymbol{\xi}$ is not orthogonal to the hypersurfaces $t = \text{const}$. Indeed from (10.8),

$$a \neq 0 \implies \boldsymbol{\xi}\cdot\boldsymbol{\eta} = g_{t\varphi} \neq 0;$$

  since $\boldsymbol{\eta}$ is tangent to the hypersurfaces $t = \text{const}$, this implies that $\boldsymbol{\xi}$ is not normal to these hypersurfaces.

- When $a \to 0$, we have $r_+ \to 2m$, $r_- \to 0$, $\rho^2 \sim r^2$, and $\rho^2/\Delta \sim (1-2m/r)^{-1}$, and we see on (10.8) that the Kerr metric reduces to the Schwarzschild metric.

- On the hypersurface $r = 0$, we have $\rho^2 = a^2\cos^2\theta$ and $\Delta = a^2$, so that the Kerr metric (10.11) induces the metric:

$$h_{ij}\, \mathrm{d}x^i \mathrm{d}x^j = -\mathrm{d}t^2 + a^2\left(\cos^2\theta\, \mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right), \tag{10.13}$$

  where $(x^i) = (t, \theta, \varphi)$. According to the assumption (10.1), $a \neq 0$ and the change of coordinates $x := a\sin\theta\cos\varphi$, $y := a\sin\theta\sin\varphi$ turns the right-hand side of (10.13)

into $-\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2$. We recognize a *flat* Minkowskian metric. In particular, for a fixed value of t, the $r = 0$ set

$$\mathcal{S}_{0,t} := \{p \in \mathscr{M}_{\mathrm{III}}, r(p) = 0, t(p) = t\} \tag{10.14}$$

is made of two connected components, which are two *flat open disks* of radius a, corresponding respectively to $\theta < \pi/2$ and $\theta > \pi/2$, since the equator $\theta = \pi/2$ is excluded by the very definition of $\mathscr{M}_{\mathrm{III}}$ (cf. remark 2 above). The set $\mathcal{S}_{0,t}$ is depicted by the dotted red line in Fig. 10.1. It is also depicted in terms of the so-called Kerr-Schild coordinates in Figs. C.1 - C.4 of Appendix C.

10.2.3 Determinant and inverse metric

The determinant of the metric $\boldsymbol{g}$ with respect to Boyer-Lindquist coordinates is deduced from (10.8); it takes a relatively simple form (see the notebook D.4.1 for the computation):

$$\det(g_{\alpha\beta}) = -\rho^4 \sin^2\theta. \tag{10.15}$$

The inverse metric is (see the notebook D.4.1 for the computation)

$$g^{\alpha\beta} = \begin{pmatrix} -\frac{1}{\Delta}\left(r^2 + a^2 + \frac{2a^2 mr \sin^2\theta}{\rho^2}\right) & 0 & 0 & -\frac{2amr}{\rho^2\Delta} \\ 0 & \frac{\Delta}{\rho^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\rho^2} & 0 \\ -\frac{2amr}{\rho^2\Delta} & 0 & 0 & \frac{1}{\Delta\sin^2\theta}\left(1 - \frac{2mr}{\rho^2}\right) \end{pmatrix}. \tag{10.16}$$

10.2.4 Ergoregion

Let us investigate the causal character of the stationary Killing vector $\boldsymbol{\xi}$. We have, according to (10.8) and (10.9),

$$\boldsymbol{\xi}\cdot\boldsymbol{\xi} = g_{tt} = -1 + \frac{2mr}{r^2 + a^2\cos^2\theta}.$$

Thus

$$\boldsymbol{\xi} \text{ timelike} \iff r^2 - 2mr + a^2\cos^2\theta > 0 \iff r < r_{\mathscr{E}^-}(\theta) \quad \text{or} \quad r > r_{\mathscr{E}^+}(\theta),$$

with

$$r_{\mathscr{E}^\pm}(\theta) := m \pm \sqrt{m^2 - a^2\cos^2\theta}. \tag{10.17}$$

Comparing with (10.3), we note that

$$0 \leq r_{\mathscr{E}^-}(\theta) \leq r_- \leq m \leq r_+ \leq r_{\mathscr{E}^+}(\theta) \leq 2m, \tag{10.18}$$

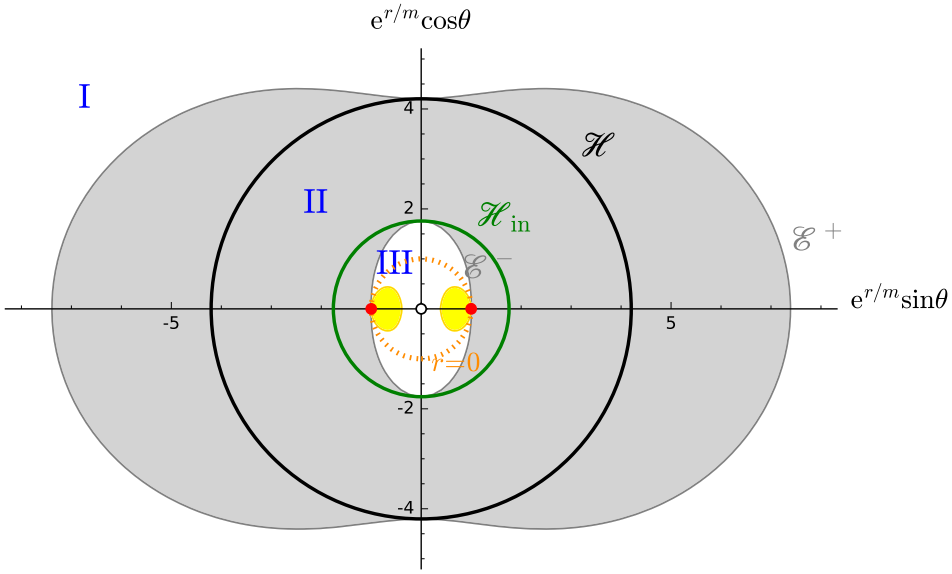


Figure 10.2: Meridional view of a section $t = \text{const}$ of Kerr spacetime with $a/m = 0.90$ in O'Neill exponential coordinates $x = \mathrm{e}^{r/m}\sin\theta$ and $z = \mathrm{e}^{r/m}\cos\theta$ (cf. Fig. 10.1). The right (resp. left) half of the figure corresponds to $\varphi = 0$ (resp. $\varphi = \pi$). The Roman numbers I, II, III denote the components $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$ of the manifold $\mathscr{M}_{\rm BL}$. The dotted orange circle marks the location of $r = 0$, while the small black circle at the center of the figure corresponds to $r \to -\infty$. The two red dots marks the curvature singularity $\mathscr{R}$. The ergoregion (cf. Sec. 10.2.4) is shown in grey, while the yellow part is Carter time machine (cf. Sec. 10.2.5).

with

$$r_{\mathscr{E}^-}(\pi/2) = 0 \tag{10.19a}$$
$$r_{\mathscr{E}^-}(0) = r_{\mathscr{E}^-}(\pi) = r_- \tag{10.19b}$$
$$r_{\mathscr{E}^+}(0) = r_{\mathscr{E}^+}(\pi) = r_+ \tag{10.19c}$$
$$r_{\mathscr{E}^+}(\pi/2) = 2m. \tag{10.19d}$$

Given the definition of $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$, we conclude that

- $\boldsymbol{\xi}$ is timelike in the region of $\mathscr{M}_{\rm I}$ defined by $r > r_{\mathscr{E}^+}(\theta)$ and in the region of $\mathscr{M}_{\rm III}$ defined by $r < r_{\mathscr{E}^-}(\theta)$;
- $\boldsymbol{\xi}$ is null on the hypersurface $\mathscr{E}^+$ of $\mathscr{M}_{\rm I}$ defined by $r = r_{\mathscr{E}^+}(\theta)$ and on the hypersurface $\mathscr{E}^-$ of $\mathscr{M}_{\rm III}$ defined by $r = r_{\mathscr{E}^-}(\theta)$;
- $\boldsymbol{\xi}$ is spacelike in all $\mathscr{M}_{\rm II}$ and in the region $\mathscr{G}^+$ of $\mathscr{M}_{\rm I}$ defined by $r < r_{\mathscr{E}^+}(\theta)$, as well as in the region $\mathscr{G}^-$ of $\mathscr{M}_{\rm III}$ defined by $r > r_{\mathscr{E}^-}(\theta)$.

According to the nomenclature introduced in Sec. 5.3.2, one calls $\mathscr{E}^+$ (resp. $\mathscr{E}^-$) the ***outer ergosphere*** (resp. ***inner ergosphere***) and $\mathscr{G}^+$ (resp. $\mathscr{G}^-$) the ***outer ergoregion*** (resp. ***inner ergoregion***). The part of $\mathscr{M}_{\rm BL}$ where $\boldsymbol{\xi}$ is spacelike, i.e. $\mathscr{G} = \mathscr{G}^+ \cup \mathscr{M}_{\rm II} \cup \mathscr{G}^-$, is

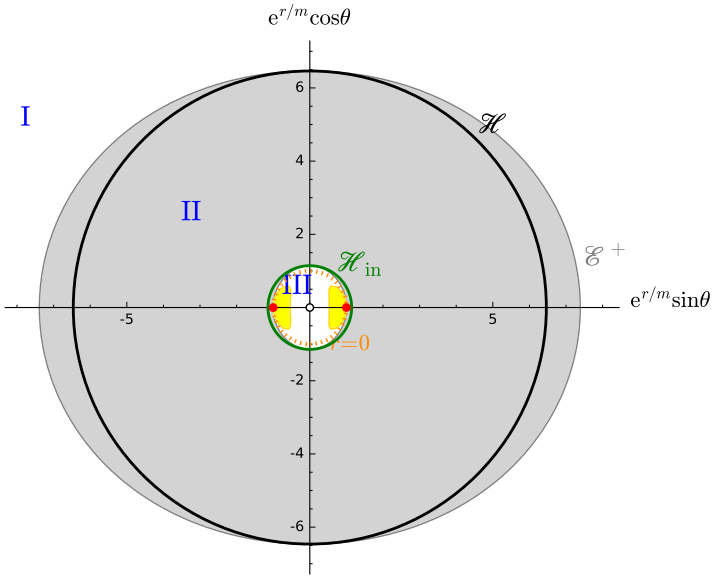


Figure 10.3: Same as Fig. 10.2 but for $a/m = 0.50$.

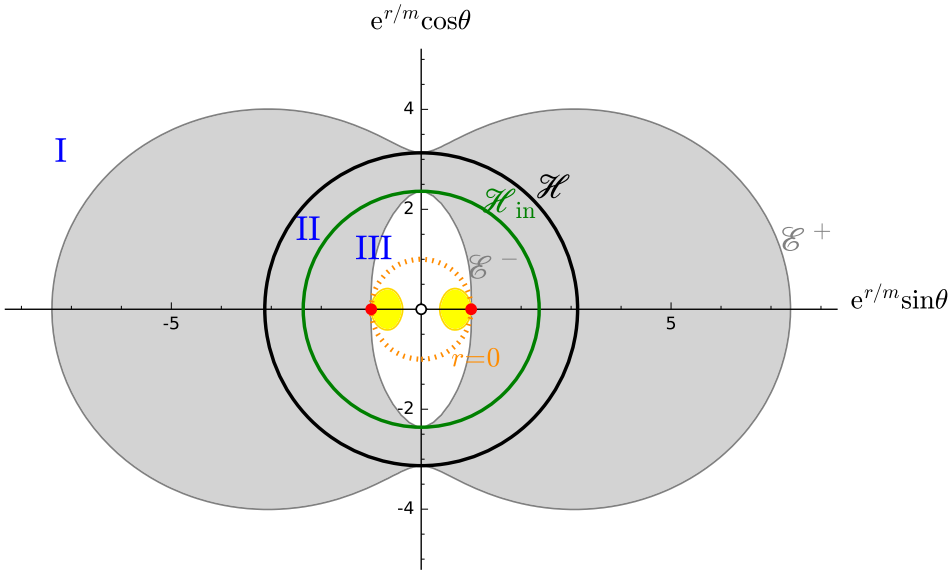


Figure 10.4: Same as Fig. 10.2 but for $a/m = 0.99$.

called the ***ergoregion***. Following the standard notation in topology, we shall denote by $\overline{\mathscr{G}}$ the closure of $\mathscr{G}$, i.e. the union of the ergoregion and the two ergospheres:

$$\overline{\mathscr{G}} := \mathscr{G} \cup \mathscr{E}^- \cup \mathscr{E}^+. \tag{10.20}$$

The locus where the Killing vector $\boldsymbol{\xi}$ is timelike is then $\mathscr{M}_{\rm BL} \setminus \overline{\mathscr{G}}$.

The ergoregion is depicted in Figs. 10.2–10.4. It is worth noticing that the location of the inner and outer ergospheres in the equatorial plane ($\theta = \pi/2$) do not depend on the spin parameter a: $r_{\mathscr{E}^-}(\pi/2) = 0$ [Eq. (10.19a)] and $r_{\mathscr{E}^+}(\pi/2) = 2m$ [Eq. (10.19d)], a feature that appears clearly in Figs. 10.2–10.4, given that $\mathrm{e}^2 \simeq 7.39$.

Remark 3: Sometimes the word ***ergosurface*** is used instead of *ergosphere*. One may encounter as well the name ***static limit*** for *ergosphere*, especially in old texts (cf. Sec. 10.7.2 below).

We shall see in Sec. 10.5.5 that the outer ergoregion plays a key role in an energy extraction mechanism known as the *Penrose process*.

10.2.5 Carter time machine

Let us now focus on the second Killing vector, $\boldsymbol{\eta}$. From (10.8) and (10.9), we have

$$\boldsymbol{\eta}\cdot\boldsymbol{\eta} = g_{\varphi\varphi} = \left(r^2 + a^2 + \frac{2a^2mr\sin^2\theta}{r^2 + a^2\cos^2\theta}\right)\sin^2\theta. \tag{10.21}$$

Hence

$$\boldsymbol{\eta} \text{ spacelike} \iff (r^2+a^2)(r^2+a^2\cos^2\theta) + 2a^2mr\sin^2\theta > 0.$$

For $\theta \to 0$ or $\theta \to \pi$, the left-hand side of the above inequality is always positive, but for $\theta = \pi/2$ and r negative with $|r|$ small enough so that $2a^2m|r| > r^2(r^2+a^2)$, it is negative. This feature is apparent on Fig. 10.5: for θ close to $\pi/2$, there is a region $\mathscr{T}$ defined by $r_{\mathscr{T}}(\theta) < r < 0$ for some negative function $r_{\mathscr{T}}(\theta)$, such that $g_{\varphi\varphi} < 0$. Since $\mathscr{T}$ corresponds to negative values of r, we have $\mathscr{T} \subset \mathscr{M}_{\rm III}$. Hence we conclude:

- $\boldsymbol{\eta}$ is spacelike in all $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm II}$, as well as outside the region $\mathscr{T}$ in $\mathscr{M}_{\rm III}$;
- $\boldsymbol{\eta}$ is timelike in the subset $\mathscr{T}$ of $\mathscr{M}_{\rm III}$;
- $\boldsymbol{\eta}$ is null at the boundary of $\mathscr{T}$.

The region $\mathscr{T}$ is called ***Carter time machine***. This name stems from the fact that thanks to $\mathscr{T}$, there is a future-directed timelike curve connecting any two points of $\mathscr{M}_{\rm III}$ (see e.g. Proposition 2.4.7 of O'Neill's textbook [200] for a demonstration, or Carter's original article [41]). The region $\mathscr{T}$ is depicted in yellow in the meridional diagrams of Figs. 10.2-10.4.

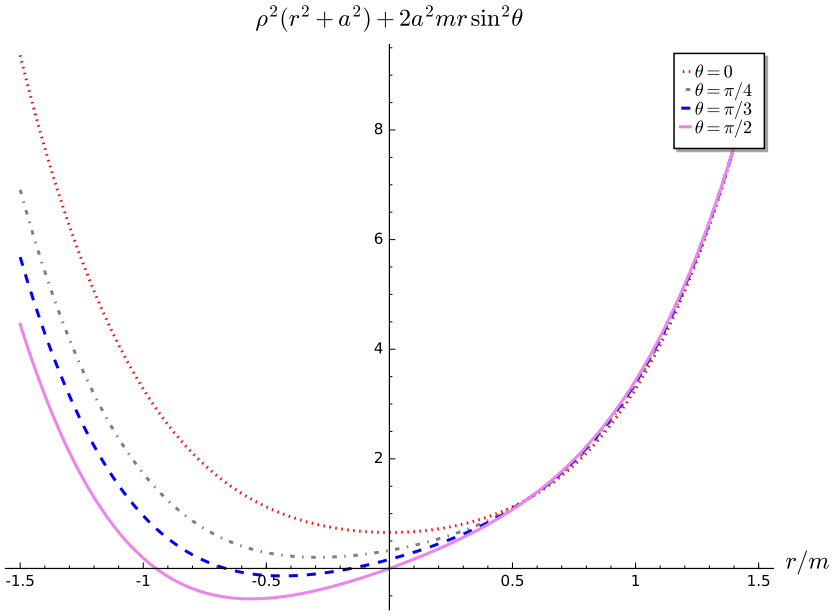


Figure 10.5: Graph of the function giving the sign of $g_{\varphi\varphi}$ for $a = 0.9m$ and various values of θ.

10.2.6 Singularities

The components $g_{\alpha\beta}$ of the Kerr metric as given by (10.8) are diverging at various locations:

- when $\rho^2 \to 0$, which, given (10.9) and assuming $a \neq 0$, is equivalent to approaching the cylinder $\mathscr{R}$ defined by (10.4);
- when $\Delta \to 0$, which, given (10.10), is equivalent to either $r \to r_-$ or $r \to r_+$; the first case corresponds to the boundary (within $\mathbb{R}^2 \times \mathbb{S}^2$) between $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$ and the second case to the boundary between $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm II}$.

The divergence when $\rho^2 \to 0$ corresponds to a *curvature singularity*. Indeed, the Kretschmann scalar of Kerr metric is (cf. Eq. (6.46) and the notebook D.4.1 for the computation)

$$K := R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} = 48\frac{m^2}{\rho^{12}}\left(r^6 - 15r^4a^2\cos^2\theta + 15r^2a^4\cos^4\theta - a^6\cos^6\theta\right). \qquad (10.22)$$

The value for $\theta = \pi/2$ is thus $K = 48m^2/r^6$, which clearly diverges for $r \to 0$ (i.e. $\rho^2 \to 0$). Hence $\mathscr{R}$ is called the ***ring singularity*** of Kerr spacetime, the word *ring* reflecting the fact that $t =$ const sections of $\mathscr{R}$ are circles [cf. Eq. (10.6)]. See Ref. [61] for an extended discussion of this singularity.

On the contrary, we shall see in the next section that the divergence of the metric components when $\Delta \to 0$ corresponds to a mere *coordinate singularity*, i.e. to a pathology of Boyer-Lindquist coordinates, which can be cured by switching to other coordinates.

10.3 Kerr coordinates and extension of the spacetime manifold through $\Delta = 0$

10.3.1 Null Kerr coordinates

The ***null Kerr coordinates*** are coordinates $(x^{\hat{\alpha}}) = (v, r, \theta, \tilde{\varphi})$ defined on $\mathbb{R}^2 \times \mathbb{S}^2$ and related to the Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$ introduced in Sec. 10.2.1 by

$$\boxed{\mathrm{d}v = \mathrm{d}t + \frac{r^2 + a^2}{\Delta}\,\mathrm{d}r} \tag{10.23a}$$

$$\boxed{\mathrm{d}\tilde{\varphi} = \mathrm{d}\varphi + \frac{a}{\Delta}\,\mathrm{d}r}. \tag{10.23b}$$

If $a = 0$, we note that the null Kerr coordinates are nothing but the null ingoing Eddington-Finkelstein coordinates on Schwarzschild spacetime (cf. Sec. 6.3.2 and compare (10.23a) with (6.29)).

Given that $\Delta = (r - r_-)(r - r_+) = r^2 + a^2 - 2mr$ [Eq. (10.10)], we have the identities

$$\frac{r^2 + a^2}{\Delta} = 1 + \frac{2m}{r_+ - r_-}\left(\frac{r_+}{r - r_+} - \frac{r_-}{r - r_-}\right) \quad \text{and} \quad \frac{a}{\Delta} = \frac{a}{r_+ - r_-}\left(\frac{1}{r - r_+} - \frac{1}{r - r_-}\right),$$

with $r_+ - r_- = 2\sqrt{m^2 - a^2}$, so that Eqs. (10.23) can be readily integrated to

$$\boxed{v = t + r + \frac{m}{\sqrt{m^2 - a^2}}\left(r_+ \ln\left|\frac{r - r_+}{2m}\right| - r_- \ln\left|\frac{r - r_-}{2m}\right|\right)} \tag{10.24a}$$

$$\boxed{\tilde{\varphi} = \varphi + \frac{a}{2\sqrt{m^2 - a^2}} \ln\left|\frac{r - r_+}{r - r_-}\right|}, \tag{10.24b}$$

up to some additive constant. When $a \to 0$, we have $r_+ \to 2m$ and $r_- \to 0$ and by comparing Eq. (10.24a) with Eq. (6.28), we recover the fact that the null Kerr coordinates reduces to the ingoing null Eddington-Finkelstein ones in this limit.

The components $g_{\hat{\alpha}\hat{\beta}}$ of the metric tensor $\boldsymbol{g}$ with respect to the null Kerr coordinates are computed from those with respect to the Boyer-Lindquist ones, as given by Eq. (10.8). One gets (cf. Appendix D or Eq. (5.31) of Ref. [146], or Lemma 2.5.2 of [200]):

$$\begin{aligned} g_{\hat{\mu}\hat{\nu}}\,\mathrm{d}x^{\hat{\mu}}\mathrm{d}x^{\hat{\nu}} = & -\left(1 - \frac{2mr}{\rho^2}\right)\mathrm{d}v^2 + 2\mathrm{d}v\,\mathrm{d}r - \frac{4amr\sin^2\theta}{\rho^2}\,\mathrm{d}v\,\mathrm{d}\tilde{\varphi} \\ & -2a\sin^2\theta\,\mathrm{d}r\,\mathrm{d}\tilde{\varphi} + \rho^2\mathrm{d}\theta^2 + \left(r^2 + a^2 + \frac{2a^2mr\sin^2\theta}{\rho^2}\right)\sin^2\theta\,\mathrm{d}\tilde{\varphi}^2. \end{aligned} \tag{10.25}$$

We note that these metric components do not have any divergence when $\Delta \to 0$, contrary to the Boyer-Lindquist ones. Hence, we may extend the Kerr metric to the points of $\mathbb{R}^2 \times \mathbb{S}^2$ where $\Delta = 0$, i.e. to the hypersurfaces (cf. Fig. 10.1)

$$\mathscr{H} := \left\{p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = r_+\right\} \tag{10.26}$$

and

$$\mathscr{H}_{\text{in}} := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = r_- \right\}. \tag{10.27}$$

The hypersurface $\mathscr{H}$ is actually the boundary between the regions $\mathscr{M}_{\text{I}}$ and $\mathscr{M}_{\text{II}}$, while $\mathscr{H}_{\text{in}}$ is the boundary between $\mathscr{M}_{\text{II}}$ and $\mathscr{M}_{\text{III}}$ (cf. Eq. (10.7) and Fig. 10.2). We thus consider

$$\boxed{\mathscr{M} := \mathscr{M}_{\text{BL}} \cup \mathscr{H} \cup \mathscr{H}_{\text{in}} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathscr{R}} \tag{10.28}$$

as the spacetime manifold. In order for $\boldsymbol{g}$ defined by (10.25) to be a well-defined metric on $\mathscr{M}$, it does not suffice that the components $g_{\hat{\alpha}\hat{\beta}}$ do not diverge at $\mathscr{H}$ and $\mathscr{H}_{\text{in}}$: one shall check as well that the bilinear form $\boldsymbol{g}$ is non-degenerate there. This is easily proven by considering the determinant of the metric components, which turns out to have a simple form (cf. the notebook D.4.2):

$$\det(g_{\hat{\alpha}\hat{\beta}}) = -\rho^4 \sin^2\theta. \tag{10.29}$$

Except at $\theta = 0$ and $\theta = \pi$ (the usual singularity of spherical coordinates), we have $\det(g_{\hat{\alpha}\hat{\beta}}) \neq 0$ everywhere on $\mathscr{M}$, since ρ vanishes only on $\mathscr{R}$, which is excluded from $\mathscr{M}$. Hence we conclude that

The symmetric bilinear form $\boldsymbol{g}$ is non-degenerate on the manifold $\mathscr{M}$ defined by Eq. (10.28) and thus $(\mathscr{M}, \boldsymbol{g})$ is a well-behaved spacetime — our ***Kerr spacetime*** from now on. We note that, contrary to $\mathscr{M}_{\text{BL}}$, $\mathscr{M}$ is a connected manifold.

We deduce from (10.23) that

$$\left.\frac{\partial v}{\partial t}\right|_{r,\theta,\varphi} = 1, \qquad \left.\frac{\partial v}{\partial r}\right|_{t,\theta,\varphi} = \frac{r^2+a^2}{\Delta}, \qquad \left.\frac{\partial \tilde{\varphi}}{\partial r}\right|_{t,\theta,\varphi} = \frac{a}{\Delta}, \qquad \left.\frac{\partial \tilde{\varphi}}{\partial \varphi}\right|_{t,r,\theta} = 1.$$

It follows then from the chain rule that the null Kerr coordinate frame is related to the Boyer-Lindquist coordinate frame by

$$\boldsymbol{\partial}_v = \boldsymbol{\partial}_t \tag{10.30a}$$

$$\boldsymbol{\partial}_{\hat{r}} = \boldsymbol{\partial}_r - \frac{a^2+r^2}{\Delta}\boldsymbol{\partial}_t - \frac{a}{\Delta}\boldsymbol{\partial}_\varphi \tag{10.30b}$$

$$\boldsymbol{\partial}_\theta = \boldsymbol{\partial}_\theta \tag{10.30c}$$

$$\boldsymbol{\partial}_{\tilde{\varphi}} = \boldsymbol{\partial}_\varphi. \tag{10.30d}$$

Note that we are using the notation $\boldsymbol{\partial}_{\hat{r}}$ for the $\partial/\partial r$ vector of the null Kerr coordinate system $(x^{\hat{\alpha}}) = (v, r, \theta, \tilde{\varphi})$, to distinguish it from the $\partial/\partial r$ vector of Boyer-Lindquist coordinates.

Historical note: The null Kerr coordinates are those in which Roy P. Kerr originally presented his solution in 1963 [164]. As noted by Kerr himself later [165], he used $-a$ instead of a, because he was "rather hurried in performing this calculation (angular momentum) and got the sign wrong" (footnote in Sec. 2.5 of Ref. [165]). Actually, it was shown in 1964 by Robert H. Boyer and T.G. Price [31] that the angular momentum about the rotation axis of the Kerr solution is $J = -am$, where a is Kerr's a. Taking this into account, the correspondence between our notations and those of Kerr's article [164] is $v \leftrightarrow u$ and $a \leftrightarrow -a$. Then we can check that the line element (10.25) coincides with the one given by an unnumbered (!) equation in Ref. [164].

10.3.2 Time orientation of Kerr spacetime

We read on (10.25) that $\boldsymbol{g}(\boldsymbol{\partial}_{\hat{r}}, \boldsymbol{\partial}_{\hat{r}}) = g_{rr} = 0$, which implies that $\boldsymbol{\partial}_{\hat{r}}$ is a global null vector field on $\mathscr{M}$. We may then use it to set the time orientation of $(\mathscr{M}, \boldsymbol{g})$ (cf. Sec. 1.2.2):

The Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ is time-orientable, and we choose its time orientation such that

$$\boldsymbol{k} := -\boldsymbol{\partial}_{\hat{r}} \tag{10.31}$$

is a future-directed null vector field in all $\mathscr{M}$.

Remark 1: The minus sign in the above definition, along with Eq. (10.30b), ensures that

$$\boldsymbol{k} \sim \boldsymbol{\partial}_t - \boldsymbol{\partial}_r \quad \text{when } r \to +\infty,$$

which shows that the time orientation set by $\boldsymbol{k}$ agrees asymptotically with that of $\boldsymbol{\xi} = \boldsymbol{\partial}_t$. The latter vector field could not have been chosen to set the time orientation of $(\mathscr{M}, \boldsymbol{g})$ since it is not causal everywhere, being spacelike in the ergoregion.

The field lines of $\boldsymbol{k}$ are future-directed null curves, which may be qualified of *ingoing* since, by definition, $-\boldsymbol{\partial}_{\hat{r}}$ points towards decreasing values of r. Note that, by the very definition of the coordinate vector $\boldsymbol{\partial}_{\hat{r}}$, the values of the coordinates $(v, \theta, \tilde{\varphi})$ are fixed along each of these null curves. We therefore denote them by $\mathscr{L}^{\text{in}}_{(v,\theta,\tilde{\varphi})}$. We shall see in Sec. 10.4 that each $\mathscr{L}^{\text{in}}_{(v,\theta,\tilde{\varphi})}$ is actually a null geodesic.

Decaying of r towards the future in $\mathscr{M}_{\text{II}}$

The scalar square of the vector $\boldsymbol{\partial}_r$ of Boyer-Lindquist coordinates is read from the metric components (10.8): $\boldsymbol{\partial}_r \cdot \boldsymbol{\partial}_r = \boldsymbol{g}(\boldsymbol{\partial}_r, \boldsymbol{\partial}_r) = g_{rr} = \rho^2/\Delta$. Since Δ is positive in $\mathscr{M}_{\text{I}}$ and $\mathscr{M}_{\text{III}}$ and negative in $\mathscr{M}_{\text{II}}$, we conclude that $\boldsymbol{\partial}_r$ is spacelike in $\mathscr{M}_{\text{I}} \cup \mathscr{M}_{\text{III}}$ and timelike in $\mathscr{M}_{\text{II}}$. Moreover, it follows from the above choice of time orientation that

In region $\mathscr{M}_{\text{II}}$, the vector $\boldsymbol{\partial}_r$ of Boyer-Lindquist coordinates is a past-directed timelike vector.

Proof. Applying Lemma 2 of Sec. 1.2.2 with $\boldsymbol{u} = \boldsymbol{k}$ and $\boldsymbol{v} = \boldsymbol{\partial}_r$, we get that $\boldsymbol{\partial}_r$ is past-directed iff $\boldsymbol{g}(\boldsymbol{k}, \boldsymbol{\partial}_r) > 0$. Now, in terms of the Boyer-Lindquist components (10.8), $\boldsymbol{g}(\boldsymbol{k}, \boldsymbol{\partial}_r) = g_{\mu r} k^\mu = g_{rr} k^r = (\rho^2/\Delta) k^r$. The Boyer-Lindquist component k^r is given by Eq. (10.30b) where $\boldsymbol{\partial}_{\hat{r}} = -\boldsymbol{k}$; we get $k^r = -1$. Hence $\boldsymbol{g}(\boldsymbol{k}, \boldsymbol{\partial}_r) = -\rho^2/\Delta > 0$ in $\mathscr{M}_{\text{II}}$, for $\Delta < 0$ there. □

An important consequence of the above property is

In region $\mathscr{M}_{\text{II}}$, the coordinate r must decrease towards the future along any causal (i.e. timelike or null) worldline.

Proof. Let $\mathscr{L}$ be a causal curve in region $\mathscr{M}_{\rm II}$ and λ a parameter along $\mathscr{L}$ increasing towards the future. The associated tangent vector $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ is then future-directed. According to the above result, $-\boldsymbol{\partial}_r$ is a future-directed timelike vector in $\mathscr{M}_{\rm II}$, so that we can apply Lemma 1 of Sec. 1.2.2 with $\boldsymbol{u} = -\boldsymbol{\partial}_r$ and get $\boldsymbol{g}(-\boldsymbol{\partial}_r, \boldsymbol{v}) < 0$. Now, using Boyer-Lindquist components, we have

$$\boldsymbol{g}(-\boldsymbol{\partial}_r, \boldsymbol{v}) = -g_{r\mu}v^\mu = -g_{rr}v^r = -g_{rr}\frac{\mathrm{d}r}{\mathrm{d}\lambda} = -\frac{\rho^2}{\Delta}\frac{\mathrm{d}r}{\mathrm{d}\lambda}.$$

Since $-\rho^2/\Delta > 0$ in $\mathscr{M}_{\rm II}$, $\boldsymbol{g}(-\boldsymbol{\partial}_r, \boldsymbol{v}) < 0$ is thus equivalent to $\mathrm{d}r/\mathrm{d}\lambda < 0$, which proves that r is decreasing along $\mathscr{L}$ as λ increases. □

10.3.3 Kerr coordinates

As in Sec. 6.3.2, we shall move from the null coordinate v to a (asymptotically) timelike one by setting

$$\boxed{\tilde{t} = v - r} \iff \boxed{v = \tilde{t} + r} \tag{10.32}$$

so that v appears as the advanced time $\tilde{t}+r$ (compare with Eq. (6.31)). We thus consider the coordinates

$$(x^{\tilde{\alpha}}) = (\tilde{t}, r, \theta, \tilde{\varphi}), \tag{10.33}$$

which we shall call ***Kerr coordinates*** (cf. the historical note below). It is worth to relate them to Boyer-Lindquist coordinates (t, r, θ, φ). This is easily achieved by combining (10.23) with $\mathrm{d}\tilde{t} = \mathrm{d}v - \mathrm{d}r$:

$$\boxed{\mathrm{d}\tilde{t} = \mathrm{d}t + \frac{2mr}{\Delta}\,\mathrm{d}r} \tag{10.34a}$$

$$\boxed{\mathrm{d}\tilde{\varphi} = \mathrm{d}\varphi + \frac{a}{\Delta}\,\mathrm{d}r}. \tag{10.34b}$$

The integrated version is deduced obtained by substituting (10.32) in Eq. (10.24):

$$\boxed{\tilde{t} = t + \frac{m}{\sqrt{m^2-a^2}}\left(r_+ \ln\left|\frac{r-r_+}{2m}\right| - r_- \ln\left|\frac{r-r_-}{2m}\right|\right)} \tag{10.35a}$$

$$\boxed{\tilde{\varphi} = \varphi + \frac{a}{2\sqrt{m^2-a^2}} \ln\left|\frac{r-r_+}{r-r_-}\right|}. \tag{10.35b}$$

Since the transform (10.32) leads to $\mathrm{d}v = \mathrm{d}\tilde{t} + \mathrm{d}r$, the metric components $g_{\tilde{\alpha}\tilde{\beta}}$ with respect to the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ are easily deduced from (10.25):

$$\begin{aligned}\boxed{g_{\tilde{\mu}\tilde{\nu}}\,\mathrm{d}x^{\tilde{\mu}}\,\mathrm{d}x^{\tilde{\nu}} = {} & -\left(1-\frac{2mr}{\rho^2}\right)\mathrm{d}\tilde{t}^2 + \frac{4mr}{\rho^2}\mathrm{d}\tilde{t}\,\mathrm{d}r - \frac{4amr\sin^2\theta}{\rho^2}\,\mathrm{d}\tilde{t}\,\mathrm{d}\tilde{\varphi} \\ & + \left(1+\frac{2mr}{\rho^2}\right)\mathrm{d}r^2 - 2a\left(1+\frac{2mr}{\rho^2}\right)\sin^2\theta\,\mathrm{d}r\,\mathrm{d}\tilde{\varphi} \\ & + \rho^2\mathrm{d}\theta^2 + \left(r^2+a^2+\frac{2a^2mr\sin^2\theta}{\rho^2}\right)\sin^2\theta\,\mathrm{d}\tilde{\varphi}^2.}\end{aligned} \tag{10.36}$$

Since we kept r, θ and $\tilde{\varphi}$ and simply changed v to $\tilde{t}$ via (10.32) when moving from the null Kerr coordinates to the Kerr coordinates, we easily get the relation between the two coordinate frames:

$$\boldsymbol{\partial}_{\tilde{t}} = \boldsymbol{\partial}_v \tag{10.37a}$$
$$\boldsymbol{\partial}_{\tilde{r}} = \boldsymbol{\partial}_v + \boldsymbol{\partial}_{\hat{r}} \tag{10.37b}$$
$$\boldsymbol{\partial}_\theta = \boldsymbol{\partial}_\theta \tag{10.37c}$$
$$\boldsymbol{\partial}_{\tilde{\varphi}} = \boldsymbol{\partial}_\varphi. \tag{10.37d}$$

Note that we have denoted by $\boldsymbol{\partial}_{\tilde{r}}$ the second vector of the coordinate frame associated to the Kerr coordinates $(x^{\tilde{\alpha}}) = (\tilde{t}, r, \theta, \tilde{\varphi})$, in order to distinguish it from the coordinate vector $\boldsymbol{\partial}_{\hat{r}}$ of the null Kerr coordinates $(x^{\hat{\alpha}}) = (v, r, \theta, \tilde{\varphi})$, as well as from the coordinate vector $\boldsymbol{\partial}_r$ of the Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$.

By combining (10.30) and (10.37), we get the relation between the Kerr coordinate frame and the Boyer-Lindquist coordinate frame:

$$\boldsymbol{\partial}_{\tilde{t}} = \boldsymbol{\partial}_t \tag{10.38a}$$
$$\boldsymbol{\partial}_{\tilde{r}} = \boldsymbol{\partial}_r - \frac{2mr}{\Delta}\boldsymbol{\partial}_t - \frac{a}{\Delta}\boldsymbol{\partial}_\varphi \tag{10.38b}$$
$$\boldsymbol{\partial}_\theta = \boldsymbol{\partial}_\theta \tag{10.38c}$$
$$\boldsymbol{\partial}_{\tilde{\varphi}} = \boldsymbol{\partial}_\varphi. \tag{10.38d}$$

We notice on (10.38a) and (10.38d) that the coordinate frame vectors $\boldsymbol{\partial}_{\tilde{t}}$ and $\boldsymbol{\partial}_{\tilde{\varphi}}$ coincide with the Killing vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$:

$$\boxed{\boldsymbol{\partial}_{\tilde{t}} = \boldsymbol{\xi}} \quad \text{and} \quad \boxed{\boldsymbol{\partial}_{\tilde{\varphi}} = \boldsymbol{\eta}}. \tag{10.39}$$

That $\boldsymbol{\partial}_{\tilde{t}}$ and $\boldsymbol{\partial}_{\tilde{\varphi}}$ are Killing vectors is not surprising since the metric components (10.36) do not depend on $\tilde{t}$ nor on $\tilde{\varphi}$.

The determinant of the metric components (10.36) takes a very simple form (see the notebook D.4.2 for the computation):

$$\det\left(g_{\tilde{\alpha}\tilde{\beta}}\right) = -\rho^4 \sin^2\theta. \tag{10.40}$$

The inverse metric takes also a rather simple form in terms of the Kerr coordinates (see the notebook D.4.2 for the computation):

$$g^{\tilde{\alpha}\tilde{\beta}} = \begin{pmatrix} -\left(1 + \frac{2mr}{\rho^2}\right) & \frac{2mr}{\rho^2} & 0 & 0 \\ \frac{2mr}{\rho^2} & \frac{\Delta}{\rho^2} & 0 & \frac{a}{\rho^2} \\ 0 & 0 & \frac{1}{\rho^2} & 0 \\ 0 & \frac{a}{\rho^2} & 0 & \frac{1}{\rho^2 \sin^2\theta} \end{pmatrix}. \tag{10.41}$$

Comparing (10.38) with (10.8), we note that the metric components in Kerr coordinates are slightly more complicated than those in Boyer-Lindquist coordinates, for they

contain extra off-diagonal terms: $g_{\tilde{t}r}$ and $g_{r\tilde{\varphi}}$. However the determinant (10.40) and the inverse metric (10.41) are pretty simple. Morever the Kerr coordinates are as well adapted to the spacetime symmetries as the Boyer-Lindquist ones, as (10.39) shows, and they have the great advantage to be regular on the boundary hypersurfaces $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, contrary to Boyer-Lindquist coordinates. The last feature is all the more important that $\mathscr{H}$ is the future event horizon of Kerr spacetime, as we are going to see. Therefore, we shall continue our study of Kerr spacetime, and especially the black hole aspect, by means of the Kerr coordinates.

Remark 2: The Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ are not related everywhere to a 3+1 slicing of spacetime. By ***3+1 slicing***, it is meant a foliation of $\mathscr{M}$ by spacelike hypersurfaces (see e.g. [122]). Now, the hypersurfaces $\tilde{t} =$ const are spacelike iff their normal $\vec{\nabla}\tilde{t}$ is timelike, which is equivalent to $\boldsymbol{g}(\vec{\nabla}\tilde{t}, \vec{\nabla}\tilde{t}) < 0$, with

$$\boldsymbol{g}(\vec{\nabla}\tilde{t}, \vec{\nabla}\tilde{t}) = g_{\tilde{\mu}\tilde{\nu}}\, \nabla^{\tilde{\mu}}\tilde{t} \nabla^{\tilde{\nu}}\tilde{t} = g^{\tilde{\mu}\tilde{\nu}}\, \partial_{\tilde{\mu}}\tilde{t}\, \partial_{\tilde{\nu}}\tilde{t} = g^{\tilde{t}\tilde{t}}.$$

Given the value of $g^{\tilde{t}\tilde{t}}$ read on (10.41), the hypersurface $\tilde{t} =$ const is spacelike iff $\rho^2 + 2mr > 0$, or equivalently $r^2 + 2mr + a^2\cos^2\theta > 0$. Now this second-order polynomial in r is positive everywhere except in the region of $\mathscr{M}_{\rm III}$ defined by

$$-m - \sqrt{m^2 - a^2\cos^2\theta} \leq r \leq -m + \sqrt{m^2 - a^2\cos^2\theta}. \tag{10.42}$$

Note that this region is contained in the negative-r part of $\mathscr{M}_{\rm III}$. We conclude that the coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ define a 3+1 slicing of $\mathscr{M}$ in $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and in the part of $\mathscr{M}_{\rm III}$ outside the region defined by (10.42).

Historical note: The coordinate $\tilde{t}$ has been introduced by Roy P. Kerr in the 1963 discovery article [164], by exactly the same transformation as (10.32) ($\tilde{t}$ is denoted by t and v by u in Ref. [164]). Kerr considered $\tilde{t}$ along with Cartesian-type coordinates (x, y, z) deduced from $(r, \theta, \tilde{\varphi})$ by spheroidal transformations, to form the coordinate system $(\tilde{t}, x, y, z)$, which is known today as ***Kerr-Schild coordinates*** (cf. Appendix C), despite they have been introduced first in Kerr's article [164] and not in the subsequent article by Kerr and Schild [166] (1965). Accordingly, the *Kerr coordinates* $(\tilde{t}, r, \theta, \tilde{\varphi})$ defined above are a mix of the coordinates $(v, r, \theta, \tilde{\varphi})$ in which Kerr exhibited his solution (cf. the historical note on p. 268), called here *null Kerr coordinates*, and the Kerr-Schild coordinates $(\tilde{t}, x, y, z)$. The Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ have been first explicitely considered by Robert H. Boyer and Richard W. Lindquist in 1966 [30], being called by them the "(E) frame" — "(E)" standing for Eddington, since these coordinates generalize Eddington-Finkelstein coordinates to the rotating case. We can check that the line element (10.36) coincides with Eq. (2.7) of Boyer and Lindquist's article [30], keeping in mind that these authors were using $-a$ for a, as Kerr in 1963 (cf. the discussion about the sign of a in the historical note on p. 268).

10.4 Principal null geodesics

10.4.1 Ingoing principal null geodesics

We have seen that the null Kerr coordinates $(v, r, \theta, \tilde{\varphi})$ introduced in Sec. 10.3.1 are such that the curves $\mathscr{L}^{\rm in}_{(v,\theta,\tilde{\varphi})}$ defined by $(v, \theta, \tilde{\varphi}) =$ const are null curves (cf. Sec. 10.3.2). Their

future-directed tangent vector field is $\boldsymbol{k} = -\boldsymbol{\partial}_{\hat{r}}$ [Eq. (10.31)], which can expressed in terms of the Kerr basis via (10.37):

$$\boxed{\boldsymbol{k} = \boldsymbol{\partial}_{\tilde{t}} - \boldsymbol{\partial}_{\tilde{r}}}. \tag{10.43}$$

The 1-form $\underline{\boldsymbol{k}}$ associated to $\boldsymbol{k}$ by $\boldsymbol{g}$-duality is easily computed from $k_{\tilde{\alpha}} = g_{\tilde{\alpha}\tilde{\mu}} k^{\tilde{\mu}}$, with $g_{\tilde{\alpha}\tilde{\mu}}$ given by Eq. (10.36). We get $k_{\tilde{\alpha}} = (-1, -1, 0, a\sin^2\theta)$, i.e.

$$\boxed{\underline{\boldsymbol{k}} = -\mathrm{d}\tilde{t} - \mathrm{d}r + a\sin^2\theta\, \mathrm{d}\tilde{\varphi}}. \tag{10.44}$$

A direct computation (cf. the notebook D.4.2) shows that

$$\boldsymbol{\nabla}_{\boldsymbol{k}}\, \boldsymbol{k} = 0. \tag{10.45}$$

It follows that each curve $\mathscr{L}^{\mathrm{in}}_{(v,\theta,\tilde{\varphi})}$ is a geodesic and that the parameter λ associated with $\boldsymbol{k}$ is an affine parameter of this geodesic (cf. Sec. B.2.1 in Appendix B). Since Eq. (10.43) implies

$$k^r = \frac{\mathrm{d}r}{\mathrm{d}\lambda} = -1,$$

we have, up to some additive constant,

$$\lambda = -r. \tag{10.46}$$

The geodesics $\mathscr{L}^{\mathrm{in}}_{(v,\theta,\tilde{\varphi})}$ are called the ***ingoing principal null geodesics***. The qualifier *ingoing* stems from the fact that r is decreasing along $\mathscr{L}^{\mathrm{in}}_{(v,\theta,\tilde{\varphi})}$ (in the future direction), which is an immediate consequence of $\lambda = -r$ being a future-directed parameter along $\mathscr{L}^{\mathrm{in}}_{(v,\theta,\tilde{\varphi})}$. The qualifier *principal* arises from a relation between $\boldsymbol{k}$ and the Weyl conformal curvature tensor $\boldsymbol{C}$ (cf. Sec. A.5.4) of Kerr spacetime, namely:

$$C^{\alpha}{}_{\mu\nu[\beta} k_{\gamma]} k^{\mu} k^{\nu} = 0 \qquad \text{and} \qquad {}^*C^{\alpha}{}_{\mu\nu[\beta} k_{\gamma]} k^{\mu} k^{\nu} = 0, \tag{10.47}$$

where ${}^*\boldsymbol{C}$ stands for the ***dual of the Weyl tensor***:

$${}^*C^{\alpha}{}_{\beta\gamma\delta} := \frac{1}{2} C^{\alpha}{}_{\beta\mu\nu}\, \epsilon^{\mu\nu}{}_{\gamma\delta}, \tag{10.48}$$

$\boldsymbol{\epsilon}$ being the Levi-Civita tensor (cf. Sec. A.3.4). In view of (10.47), one says that the vector field $\boldsymbol{k}$ constitutes a ***doubly degenerate, principal null direction*** of $\boldsymbol{C}$ (see e.g. Chap. 5 of O'Neill textbook [200] for details). We note that the ingoing principal null geodesics form a *congruence*: through each point of $\mathscr{M}$, there is one, and only one, curve $\mathscr{L}^{\mathrm{in}}_{(v,\theta,\tilde{\varphi})}$.

Remark 1: For the Kerr spacetime, as for any solution of the vacuum Einstein equation ($\boldsymbol{R} = 0$), the Weyl conformal curvature tensor $\boldsymbol{C}$ is equal to the Riemann curvature tensor **Riem**. Indeed, setting $\boldsymbol{R} = 0$ in Eq. (A.114) yields immediately $\mathbf{Riem} = \boldsymbol{C}$.

The components of $\boldsymbol{k}$ with respect to the Boyer-Lindquist coordinate frame are immediately deduced from Eq. (10.30b) and $\boldsymbol{k} = -\boldsymbol{\partial}_{\hat{r}}$:

$$\boldsymbol{k} = \frac{r^2 + a^2}{\Delta}\, \boldsymbol{\partial}_t - \boldsymbol{\partial}_r + \frac{a}{\Delta}\, \boldsymbol{\partial}_{\varphi}. \tag{10.49}$$

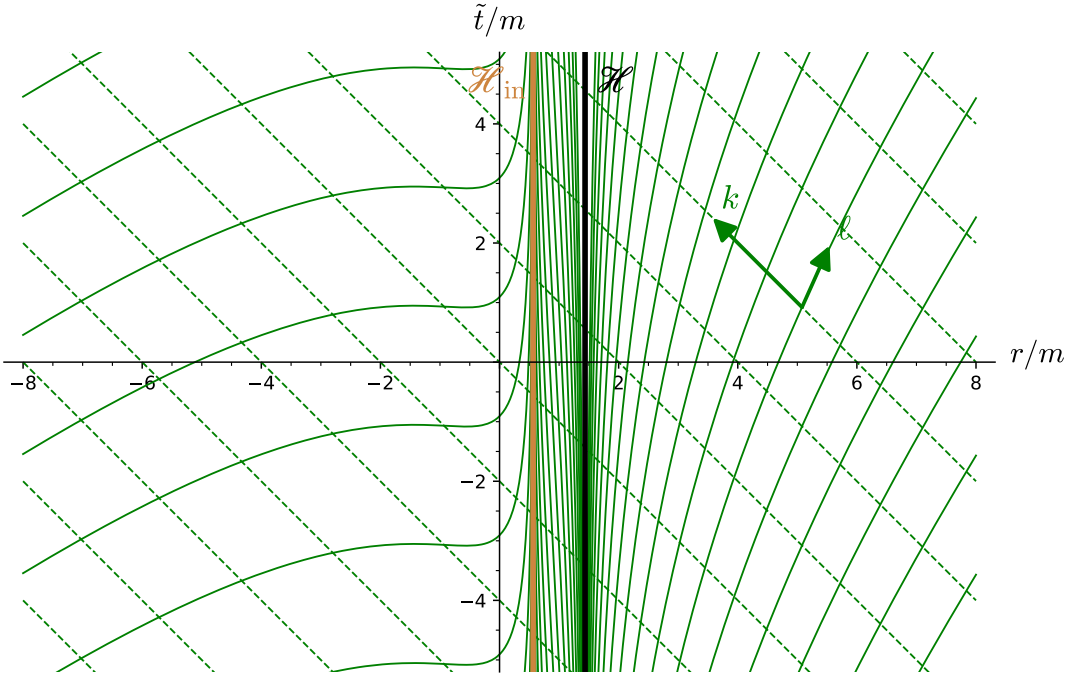


Figure 10.6: Principal null geodesics of Kerr spacetime viewed in terms of the Kerr coordinates $(\tilde{t}, r)$ for $a/m = 0.9$. The solid (resp. dashed) curves correspond to outgoing geodesics $\mathscr{L}^{\rm out}_{(u,\theta,\tilde{\varphi})}$ (resp. ingoing geodesics $\mathscr{L}^{\rm in}_{(v,\theta,\tilde{\varphi})}$), as given by Eq. (10.53) with $u = \text{const}$ (resp. Eq. (10.32) with $v = \text{const}$). The increment in u between two depicted outgoing geodesics is $\delta u = 2m$; similarly, two depicted ingoing geodesics differ in their values of v by $\delta v = 2m$. Note that the outgoing principal null geodesics tend to become tangent to the horizon $\mathscr{H}$ (resp. $\mathscr{H}_{\rm in}$) when $r \to r_+$ (resp. $r \to r_-$). Actually, we shall see in Sec. 10.5.3 and 10.8.3 that the generators of these two horizons belong to the outgoing principal null congruence. *[Figure generated by the notebook D.4.3]*

Substituting Eqs. (10.34) for $\mathrm{d}\tilde{t}$ and $\mathrm{d}\tilde{\varphi}$ in Eq. (10.44), we get the expression of the associated 1-form $\underline{\boldsymbol{k}}$ with respect to the Boyer-Lindquist coordinate coframe:

$$\underline{\boldsymbol{k}} = -\mathrm{d}t - \frac{\rho^2}{\Delta}\,\mathrm{d}r + a\sin^2\theta\,\mathrm{d}\varphi. \tag{10.50}$$

Remark 2: Expressions (10.49) and (10.50) are singular on the two horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, where $\Delta = 0$. This reflects of course not a pathology of the vector field $\boldsymbol{k}$ but merely the singularity of the Boyer-Lindquist coordinates on $\mathscr{H}$ and $\mathscr{H}_{\rm in}$. The expressions of $\boldsymbol{k}$ and $\underline{\boldsymbol{k}}$ in terms of the Kerr coordinates [Eqs. (10.43) and (10.44)] are perfectly regular in all $\mathscr{M}$.

10.4.2 Outgoing principal null geodesics

One can construct a second congruence of principal null geodesics from the *outgoing* Kerr coordinates instead of the ingoing ones considered in Sec. 10.3.1. The ***outgoing null Kerr coordinates*** $(u, r, \theta, \tilde{\varphi})$ are defined by relations to Boyer-Lindquist coordinates

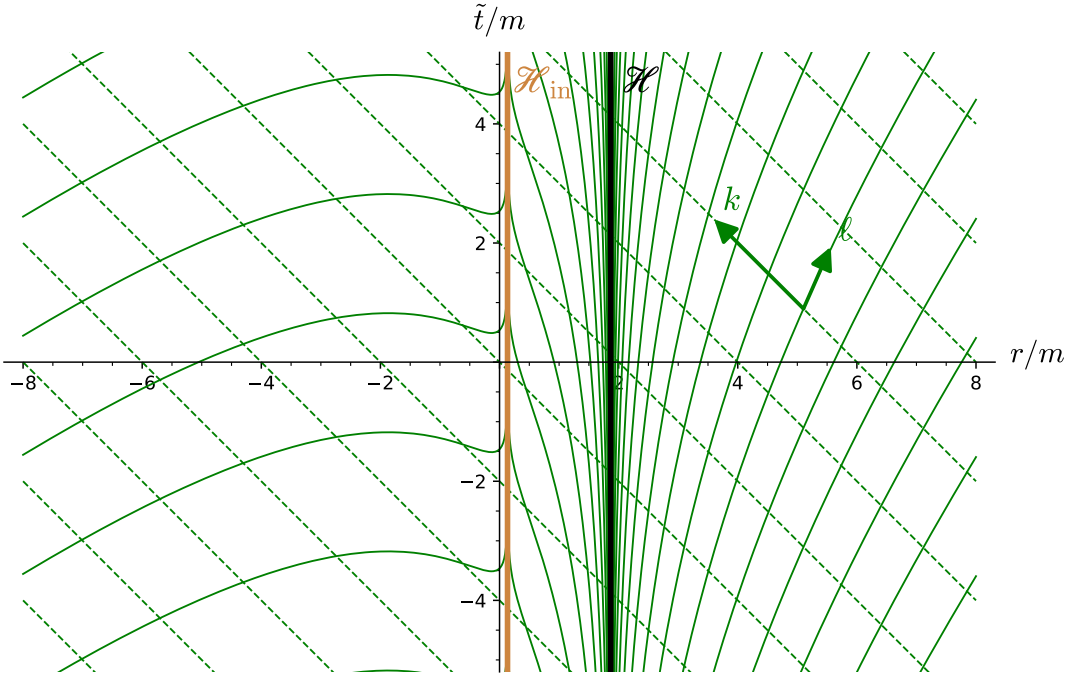


Figure 10.7: Same as Fig. 10.6, but for $a/m = 0.5$.

that are similar to (10.23), up to a change of sign:

$$\mathrm{d}u = \mathrm{d}t - \frac{r^2+a^2}{\Delta}\,\mathrm{d}r \tag{10.51a}$$

$$\mathrm{d}\tilde{\tilde{\varphi}} = \mathrm{d}\varphi - \frac{a}{\Delta}\,\mathrm{d}r. \tag{10.51b}$$

The coordinates $(u, r, \theta, \tilde{\tilde{\varphi}})$ generalize the null outgoing Eddington-Finkelstein coordinates introduced in Sec. 6.4 to the case $a \neq 0$. Thanks to the symmetry $(t, \varphi) \mapsto (-t, -\varphi)$ of the Kerr metric (10.8), which turns $(u, \tilde{\tilde{\varphi}})$ into $(-v, -\tilde{\varphi})$, it is clear that the curves $\mathscr{L}^{\rm out}_{(u,\theta,\tilde{\tilde{\varphi}})}$ defined by $(u, \theta, \tilde{\tilde{\varphi}}) = \text{const}$ constitute a second congruence of principal null geodesics, called the ***outgoing principal null geodesics***. A priori, this congruence is defined only in $\mathscr{M}_{\rm BL}$, i.e. where $\Delta \neq 0$, but we shall below that we can extend it to all $\mathscr{M}$. As $-r$ was a affine parameter along the ingoing principal null geodesics, r is an affine parameter along the outgoing principal null geodesics in $\mathscr{M}_{\rm BL}$. A difference with respect to the ingoing family is that, while $-r$ was always increasing towards the future along all ingoing geodesics, r is increasing towards the future along outgoing principal null geodesics only in regions $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$; in region $\mathscr{M}_{\rm II}$, r is decreasing towards the future, in agreement with the general rule for causal curves established in Sec. 10.3.2.

The fact that the Weyl tensor $\boldsymbol{C}$ admits two, and exactly two, congruences of principal null geodesics means that the Kerr metric is an *algebraically special* solution of Einstein equation: it belongs to the so-called *Petrov type D* [200].

Let us find the expression of the outgoing principal null geodesics in terms of the Kerr coordinates (which have been constructed on the *ingoing* principal null congruence).

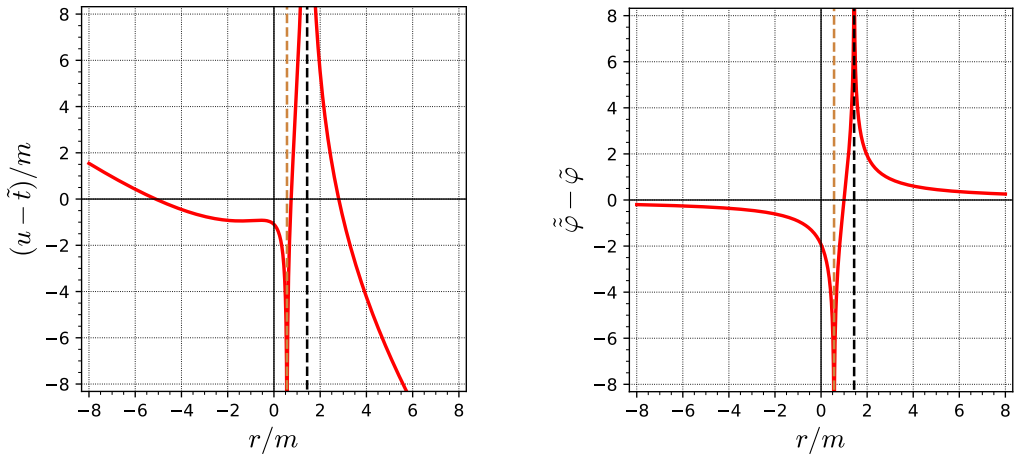


Figure 10.8: $u-\tilde{t}$ (left panel) and $\tilde{\tilde{\varphi}}-\tilde{\varphi}$ (right panel) as functions of r given by Eqs. (10.53) for $a/m = 0.9$. The dashed vertical lines correspond to $r = r_-$ ($\mathscr{H}_{\rm in}$) and $r = r_+$ ($\mathscr{H}$) and delimitate the regions $\mathscr{M}_{\rm III}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm I}$ (from the left to the right). *[Figure generated by the notebook D.4.3]*

Combining (10.51) with (10.34), we get

$$\mathrm{d}u = \mathrm{d}\tilde{t} - \frac{r^2+2mr+a^2}{\Delta}\,\mathrm{d}r \tag{10.52a}$$

$$\mathrm{d}\tilde{\tilde{\varphi}} = \mathrm{d}\tilde{\varphi} - \frac{2a}{\Delta}\,\mathrm{d}r. \tag{10.52b}$$

These equations can be integrated (cf. the computation leading to Eq. (10.24)), yielding

$$u = \tilde{t} - r - \frac{2m}{\sqrt{m^2-a^2}}\left(r_+ \ln\left|\frac{r-r_+}{2m}\right| - r_- \ln\left|\frac{r-r_-}{2m}\right|\right) \tag{10.53a}$$

$$\tilde{\tilde{\varphi}} = \tilde{\varphi} - \frac{a}{\sqrt{m^2-a^2}} \ln\left|\frac{r-r_+}{r-r_-}\right|, \tag{10.53b}$$

The quantities $u-\tilde{t}$ and $\tilde{\tilde{\varphi}}-\tilde{\varphi}$ are plotted in terms of r in Fig. 10.8. We see there that u takes all values in the range $(-\infty,+\infty)$ in each of the regions $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$. Accordingly, the 3-tuple $(u,\theta,\tilde{\tilde{\varphi}})$ defines uniquely an outgoing principal null geodesic only in each of these three regions. In other words, we have three families $\mathscr{L}^{\rm out,I}_{(u,\theta,\tilde{\tilde{\varphi}})}$, $\mathscr{L}^{\rm out,II}_{(u,\theta,\tilde{\tilde{\varphi}})}$ and $\mathscr{L}^{\rm out,III}_{(u,\theta,\tilde{\tilde{\varphi}})}$ of outgoing principal null geodesics, each family being labelled by $(u,\theta,\tilde{\tilde{\varphi}})$. In what follows, we shall denote by $\mathscr{L}^{\rm out}_{(u,\theta,\tilde{\tilde{\varphi}})}$ any member of one of these families.

The outgoing principal null geodesics are depicted, along with the ingoing ones, in Figs. 10.6 and 10.7. Note that the $\tilde{\varphi}$-motion of the outgoing geodesics, as expressed by Eq. (10.53b) with $\tilde{\tilde{\varphi}}$ held fixed, is not shown in these figures, which represent only traces in the $(\tilde{t},r)$ plane. Actually the geodesics $\mathscr{L}^{\rm out}_{(u,\theta,\tilde{\tilde{\varphi}})}$ are winding at the coordinate speed

$$\left.\frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}\tilde{t}}\right|_{\mathscr{L}^{\rm out}_{(u,\theta,\tilde{\tilde{\varphi}})}} = \frac{2a}{r^2+2mr+a^2}. \tag{10.54}$$

Proof. Along a null geodesic $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$, we have $\mathrm{d}u = 0$ and $\mathrm{d}\tilde\varphi = 0$, so that (10.52) yields

$$\left.\frac{\mathrm{d}\tilde t}{\mathrm{d}r}\right|_{\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}} = \frac{r^2+2mr+a^2}{\Delta} \qquad \text{and} \qquad \left.\frac{\mathrm{d}\tilde\varphi}{\mathrm{d}r}\right|_{\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}} = \frac{2a}{\Delta}. \tag{10.55}$$

Dividing the second expression by the first one yields (10.54). □

Note that in both asymptotically flat regions, when $r \to +\infty$ and $r \to -\infty$, the winding speed (10.54) goes to zero and the two congruences of geodesics are $\pm 45°$ lines in Figs. 10.6 and 10.7, as expected. Note also that, despite their name, the outgoing principal null geodesics are actually *ingoing* in $\mathscr{M}_{\rm II}$ (between $\mathscr{H}^-_{\rm in}$ and $\mathscr{H}^-$), i.e. have r decaying towards the future, in agreement with the general property stated in Sec. 10.3.2.

Remark 3: In Fig. 10.7, the outgoing geodesics seem to go "backward in time" for $-2m \lesssim r \lesssim 0$. This is an artefact due to the hypersurfaces $\tilde t = \text{const}$ being not spacelike there, as discussed in Remark 2 p. 272. Consequently it is possible to move to the future with decaying values of $\tilde t$ in this region. The same effect exists, but is less pronounced, for $a/m = 0.9$ (Fig. 10.6).

Another view of ingoing principal null geodesics is provided by Figs. C.1, C.2 and C.4 of Appendix C, which depict them in terms of Kerr-Schild coordinates.

10.4.3 Regular null tangent vector to the outgoing congruence

The presence of Δ in the denominators of expressions (10.55) shows that r can no longer be considered as a parameter along $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$ when $\Delta = 0$, i.e. on $\mathscr{H}^-$ and $\mathscr{H}^-_{\rm in}$. Actually, the family $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$ is not even defined there, since we read on (10.53) and see on Fig. 10.8 that $u \to \pm\infty$ and $\tilde\varphi \to \pm\infty$ when $r \to r_\pm$ (see also Fig. 6.5 for a pictorial view of $u \to +\infty$ when $r \to r_+$ in the Schwarzschild case). In order to extend the outgoing principal null family to $\mathscr{H}^-$ and $\mathscr{H}^-_{\rm in}$, let us introduce along $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$ in $\mathscr{M}_{\rm BL}$ the parameter λ that is related to the affine parameter r by

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda} = \frac{\Delta}{2(r^2+a^2)}. \tag{10.56}$$

The tangent vector to $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$ associated to λ, $\boldsymbol{\ell}$ say, has the following components w.r.t. Kerr coordinates, obtained by combining Eqs. (10.55) and (10.56):

$$\begin{aligned}
\ell^{\tilde t} &= \frac{\mathrm{d}\tilde t}{\mathrm{d}\lambda} = \frac{\mathrm{d}\tilde t}{\mathrm{d}r} \times \frac{\mathrm{d}r}{\mathrm{d}\lambda} = \frac{r^2+2mr+a^2}{2(r^2+a^2)} = \frac{1}{2} + \frac{mr}{r^2+a^2} \\
\ell^{r} &= \frac{\mathrm{d}r}{\mathrm{d}\lambda} = \frac{\Delta}{2(r^2+a^2)} = \frac{1}{2} - \frac{mr}{r^2+a^2} \\
\ell^{\theta} &= \frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = 0 \\
\ell^{\tilde\varphi} &= \frac{\mathrm{d}\tilde\varphi}{\mathrm{d}\lambda} = \frac{\mathrm{d}\tilde\varphi}{\mathrm{d}r} \times \frac{\mathrm{d}r}{\mathrm{d}\lambda} = \frac{a}{r^2+a^2}.
\end{aligned}$$

In other words, we have

$$\boxed{\boldsymbol{\ell} = \left(\frac{1}{2} + \frac{mr}{r^2+a^2}\right) \boldsymbol{\partial}_{\tilde{t}} + \left(\frac{1}{2} - \frac{mr}{r^2+a^2}\right) \boldsymbol{\partial}_{\tilde{r}} + \frac{a}{r^2+a^2} \boldsymbol{\partial}_{\tilde{\varphi}}}. \tag{10.57}$$

It is clear that this vector field is regular everywhere in $\mathscr{M}$. Given the metric components (10.36), it is easy to check that $\boldsymbol{g}(\boldsymbol{\ell}, \boldsymbol{\ell}) = 0$, i.e. that $\boldsymbol{\ell}$ is a null vector. Moreover, an explicit computation (cf. the notebook D.4.2) reveals that

$$\boldsymbol{\nabla}_{\boldsymbol{\ell}} \boldsymbol{\ell} = \kappa_{\boldsymbol{\ell}} \boldsymbol{\ell}, \tag{10.58}$$

with

$$\kappa_{\boldsymbol{\ell}} := \frac{m(r^2-a^2)}{(r^2+a^2)^2}. \tag{10.59}$$

Equation (10.58) shows that the integral curves of the vector field $\boldsymbol{\ell}$ are geodesics (cf. Secs. 2.3.3 and B.2.2). In $\mathscr{M}_{\mathrm{BL}}$, they are nothing but the null geodesics $\mathscr{L}^{\mathrm{out}}_{(u,\theta,\tilde{\varphi})}$. On $\mathscr{H}$ or $\mathscr{H}_{\mathrm{in}}$, they are null geodesics evolving at constant r ($= r_+$ or r_-) (since $\ell^r = 0$ there), constant θ (since $\ell^\theta = 0$) and constant ψ, where

$$\psi := \tilde{\varphi} - \frac{a}{2mr_\pm} \tilde{t}. \tag{10.60}$$

Indeed, in view of the components (10.57), we have along these geodesics

$$\frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}\tilde{t}} = \frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}\lambda} \frac{\mathrm{d}\lambda}{\mathrm{d}\tilde{t}} = \frac{\ell^{\tilde{\varphi}}}{\ell^{\tilde{t}}} = \frac{2a}{r_\pm^2 + 2mr_\pm + a^2} = \frac{a}{2mr_\pm}.$$

Since $a/(2mr_\pm)$ is constant, we get $\tilde{\varphi} = a/(2mr_\pm)\, \tilde{t} + \psi$, where ψ is some integration constant.

Given that $\boldsymbol{\ell}$ is a smooth vector field, the congruence of ***outgoing principal null geodesics*** can be smoothly extended to include the above integral curves of $\boldsymbol{\ell}$ on $\mathscr{H}$ and $\mathscr{H}_{\mathrm{in}}$. We shall denote them by $\mathscr{L}^{\mathrm{out},\mathscr{H}}_{(\theta,\psi)}$ and $\mathscr{L}^{\mathrm{out},\mathscr{H}_{\mathrm{in}}}_{(\theta,\psi)}$ and shall discuss them in more detail in Sec. 10.5.3.

The 1-form $\underline{\boldsymbol{\ell}}$ associated to $\boldsymbol{\ell}$ by $\boldsymbol{g}$-duality is obtained from $\ell_{\tilde{\alpha}} = g_{\tilde{\alpha}\tilde{\mu}} \ell^{\tilde{\mu}}$, with $g_{\tilde{\alpha}\tilde{\mu}}$ given by Eq. (10.36):

$$\boxed{\underline{\boldsymbol{\ell}} = -\frac{\Delta}{2(r^2+a^2)} \mathbf{d}\tilde{t} + \frac{2\rho^2 - \Delta}{2(r^2+a^2)} \mathbf{d}r + \frac{a\Delta \sin^2\theta}{2(r^2+a^2)} \mathbf{d}\tilde{\varphi}}. \tag{10.61}$$

The components of $\boldsymbol{\ell}$ and $\underline{\boldsymbol{\ell}}$ with respect to the Boyer-Lindquist coordinate frame/coframe are obtained by combining Eq. (10.57) with Eq. (10.38) and Eq. (10.61) with Eq. (10.34):

$$\boldsymbol{\ell} = \frac{1}{2} \boldsymbol{\partial}_t + \frac{\Delta}{2(r^2+a^2)} \boldsymbol{\partial}_r + \frac{a}{2(r^2+a^2)} \boldsymbol{\partial}_\varphi. \tag{10.62}$$

$$\underline{\boldsymbol{\ell}} = -\frac{\Delta}{2(r^2+a^2)}\,\mathrm{d}t + \frac{\rho^2}{2(r^2+a^2)}\,\mathrm{d}r + \frac{a\Delta\sin^2\theta}{2(r^2+a^2)}\,\mathrm{d}\varphi. \tag{10.63}$$

Remark 4: The reader may have noticed a certain dissymmetry between the chosen tangent vector $\boldsymbol{k}$ of ingoing principal null geodesics, which obeys $\boldsymbol{\nabla}_{\boldsymbol{k}}\,\boldsymbol{k} = 0$ [Eq. (10.45)] and the tangent vector $\boldsymbol{\ell}$ of the outgoing ones, which obeys $\boldsymbol{\nabla}_{\boldsymbol{\ell}}\,\boldsymbol{\ell} \neq 0$ [Eq. (10.58)]. The last property implies that the parameter λ associated to $\boldsymbol{\ell}$ is not affine, while the parameter $-r$ associated to $\boldsymbol{k}$ is (cf. Sec. B.2.2). The non-affine choice is the price to pay to have a parametrization of the outgoing family well-defined everywhere in $\mathscr{M}$, even where $\Delta = 0$. We shall see in Sec. 10.8 that in the maximal extension of the Kerr spacetime, there are other regions where these features are reversed, thereby restoring the symmetry between ingoing and outgoing principal null geodesics on the extended spacetime.

Using either the Kerr components (10.43) and (10.57) or the Boyer-Lindquist components (10.49) and (10.62), one can easily compute the scalar product of $\boldsymbol{k}$ and $\boldsymbol{\ell}$:

$$\boldsymbol{g}(\boldsymbol{k},\boldsymbol{\ell}) = -\frac{\rho^2}{r^2+a^2}. \tag{10.64}$$

This implies that $\boldsymbol{g}(\boldsymbol{k},\boldsymbol{\ell}) < 0$ everywhere (note that $\rho^2 = 0$ would correspond to the ring singularity $\mathscr{R}$, which has been excluded from the spacetime manifold $\mathscr{M}$). Using Lemma 2 of Sec. 1.2.2 with $\boldsymbol{u} = \boldsymbol{k}$ and $\boldsymbol{v} = \boldsymbol{\ell}$, we conclude that

> The tangent vector $\boldsymbol{\ell}$ to the outgoing principal null geodesics is future-directed in all the Kerr spacetime $\mathscr{M}$.

Remark 5: Instead of $\boldsymbol{\ell}$, another natural choise for the tangent vector to the outgoing principal null geodesics $\mathscr{L}^{\mathrm{out}}_{(u,\theta,\tilde{\varphi})}$, $\mathscr{L}^{\mathrm{out},\mathscr{H}}_{(\theta,\psi)}$ and $\mathscr{L}^{\mathrm{out},\mathscr{H}_{\mathrm{in}}}_{(\theta,\psi)}$ would have been the tangent vector $\hat{\boldsymbol{\ell}}$ such that

$$\boldsymbol{g}(\boldsymbol{k},\hat{\boldsymbol{\ell}}) = -1. \tag{10.65}$$

In view of (10.64), the two tangent vectors are related by

$$\hat{\boldsymbol{\ell}} = \frac{r^2+a^2}{\rho^2}\boldsymbol{\ell}. \tag{10.66}$$

As for $\boldsymbol{\ell}$, the tangent vector $\hat{\boldsymbol{\ell}}$ does not correspond to an affine parameter of the outgoing principal null geodesics, i.e. its non-affinity coefficient $\kappa_{\hat{\boldsymbol{\ell}}}$ is nonzero. While $\hat{\boldsymbol{\ell}}$ is used in many studies, we prefer $\boldsymbol{\ell}$ here because it coincides with the Killing vector $\boldsymbol{\xi} + \Omega_H\boldsymbol{\eta}$ on the event horizon [Eq. (10.77) below]. In particular, its non-affinity coefficient $\kappa_{\boldsymbol{\ell}}$ gives there the so-called black hole's surface gravity (Sec. 10.5.4 below).

10.5 Event horizon

10.5.1 The two Killing horizons

Let us consider the hypersurfaces of $\mathscr{M}$ defined by a fixed value of the coordinate r. $\mathscr{H}$ and $\mathscr{H}_{\mathrm{in}}$ are two particular cases, corresponding to $r = r_+$ and $r = r_-$ respectively. The

normal 1-form to these hypersurfaces is $\mathbf{d}r$; the corresponding gradient vector field is $\overrightarrow{\boldsymbol{\nabla}} r$, the components of which are $\nabla^{\tilde{\alpha}} r = g^{\tilde{\alpha}\tilde{\mu}} \partial_{\tilde{\mu}} r = g^{\tilde{\alpha} r}$. Given (10.41), this yields the following components with respect to the Kerr coordinates:

$$\nabla^{\tilde{\alpha}} r = \left(\frac{2mr}{\rho^2}, \frac{\Delta}{\rho^2}, 0, \frac{a}{\rho^2} \right). \tag{10.67}$$

It is then quite natural to consider the rescaled vector field

$$\boldsymbol{n} := \rho^2 \overrightarrow{\boldsymbol{\nabla}} r \tag{10.68}$$

as the normal to the hypersurfaces $r = \text{const}$, instead of $\overrightarrow{\boldsymbol{\nabla}} r$, for it has simpler components in the Kerr frame:

$$\boldsymbol{n} = 2mr\, \boldsymbol{\partial}_{\tilde{t}} + \Delta\, \boldsymbol{\partial}_{\tilde{r}} + a\, \boldsymbol{\partial}_{\tilde{\varphi}}. \tag{10.69}$$

The scalar square of $\boldsymbol{n}$ is

$$\boldsymbol{n} \cdot \boldsymbol{n} = \boldsymbol{g}(\boldsymbol{n}, \boldsymbol{n}) = n_\mu n^\mu = \rho^2 (\nabla_\mu r) n^\mu = \rho^2 n^r.$$

Hence, in view of (10.69),

$$\boldsymbol{n} \cdot \boldsymbol{n} = \rho^2 \Delta. \tag{10.70}$$

Since $\rho^2 > 0$ everywhere on $\mathscr{M}$ and $\Delta = (r - r_+)(r - r_-)$ [cf. Eq. (10.10)], we conclude that

- The hypersurfaces $r = \text{const}$ are timelike in regions $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$;
- The hypersurfaces $r = \text{const}$ are spacelike in region $\mathscr{M}_{\rm II}$;
- $\mathscr{H}$ (where $r = r_+$) and $\mathscr{H}_{\rm in}$ (where $r = r_-$) are null hypersurfaces.

On $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, $\Delta = 0$, so that Eqs. (10.67)-(10.68) yield

$$\boldsymbol{n} = 2mr_\pm\, \boldsymbol{\partial}_{\tilde{t}} + a\, \boldsymbol{\partial}_{\tilde{\varphi}} = 2mr_\pm\, \boldsymbol{\xi} + a\, \boldsymbol{\eta}, \tag{10.71}$$

where we have used (10.39) and $r_\pm$ stands for r_+ on $\mathscr{H}$ and r_- on $\mathscr{H}_{\rm in}$. On $\mathscr{H}$, we may rewrite this expression as

$$\boldsymbol{n} = 2mr_+\, \boldsymbol{\chi}, \tag{10.72}$$

with

$$\boxed{\boldsymbol{\chi} := \boldsymbol{\xi} + \Omega_H\, \boldsymbol{\eta}} \tag{10.73}$$

and

$$\boxed{\Omega_H := \frac{a}{2mr_+} = \frac{a}{r_+^2 + a^2} = \frac{a}{2m\left(m + \sqrt{m^2 - a^2}\right)}}. \tag{10.74}$$

Ω_H being a constant, the vector field $\boldsymbol{\chi}$ defined by (10.73) is a Killing vector field. Moreover, (10.72) shows that this Killing vector is normal to the null hypersurface $\mathscr{H}$. In view of the definition given in Sec. 3.3.2, we conclude that

$\mathscr{H}$ is a Killing horizon.

Similarly, on $\mathscr{H}_{\rm in}$, we may rewrite (10.71) as $\boldsymbol{n} = 2mr_- \boldsymbol{\chi}_{\rm in}$, with

$$\boldsymbol{\chi}_{\rm in} := \boldsymbol{\xi} + \Omega_{\rm in}\, \boldsymbol{\eta} \tag{10.75}$$

and

$$\Omega_{\rm in} := \frac{a}{2mr_-} = \frac{a}{r_-^2 + a^2} = \frac{a}{2m\left(m - \sqrt{m^2 - a^2}\right)}, \tag{10.76}$$

thereby arriving at the same conclusion:

$\mathscr{H}_{\rm in}$ is a Killing horizon.

We shall call $\mathscr{H}_{\rm in}$ the ***inner horizon***. We shall see in Sec. 10.8.3 that $\mathscr{H}_{\rm in}$ is actually (part of) a so-called *Cauchy horizon*.

Historical note: The identification of the hypersurfaces $\mathscr{H}$ and $\mathscr{H}_{\rm in}$ as the only two null hypersurfaces of Kerr spacetime that are stationary (i.e. Killing horizons in the modern language) has been first performed in 1964 by Robert H. Boyer and T.G. Price [31], who claimed: "These are 'horizons' in the sense that there can be a flow of matter or radiation across them in only one dimension. They are the analogues of the Schwarzschild null sphere or 'singularity'."

10.5.2 Black hole character

As a null hypersurface, $\mathscr{H}$ is a one-way membrane (cf. Sec. 2.2.2), therefore any (massive or null) particle that crossed it from $\mathscr{M}_{\rm I}$ to $\mathscr{M}_{\rm II}$ can never be back in $\mathscr{M}_{\rm I}$. Let us show that $\mathscr{H}$ is actually a black hole event horizon, as defined in Sec. 4.4.2.

We have seen in Sec. 10.2.2 that the asymptotics of region $\mathscr{M}_{\rm I}$ is the same as that of Schwarzschild spacetime. Hence one can perform a conformal completion of $(\mathscr{M}_{\rm I}, \boldsymbol{g})$ endowed with a future null infinity $\mathscr{I}^+$ and a past null infinity $\mathscr{I}^-$ (an explicit construction of $\mathscr{I}^+$ and $\mathscr{I}^-$ for Schwarzschild spacetime has been performed in Sec. 9.4.3). A conformal diagram representing $\mathscr{M}_{\rm I}$ along with $\mathscr{I}^+$ and $\mathscr{I}^-$ is given in Fig. 10.10 below.

Let us show that the causal past of the future null infinity coincide with $\mathscr{M}_{\rm I}$: $J^-(\mathscr{I}^+) = \mathscr{M}_{\rm I}$. Since, as stressed above, no future-directed causal curve can move from $\mathscr{M}_{\rm II}$ to $\mathscr{M}_{\rm I}$ and $\mathscr{I}^+$ is a boundary of $\mathscr{M}_{\rm I}$, we have $\mathscr{M}_{\rm II} \cap J^-(\mathscr{I}^+) = \varnothing$. A fortiori $\mathscr{M}_{\rm III} \cap J^-(\mathscr{I}^+) = \varnothing$. We have thus $J^-(\mathscr{I}^+) \subset \mathscr{M}_{\rm I}$. To show the equality between the two sets there remains to show that any point $p \in \mathscr{M}_{\rm I}$ can emit a signal reaching $\mathscr{I}^+$. Let $\mathscr{L}$ be the null geodesic through p of the outgoing principal null congruence $\mathscr{L}^{\rm out}_{(u,\theta,\tilde{\varphi})}$ introduced in Sec. 10.4, i.e. $\mathscr{L}$ is the unique geodesic departing from p with the tangent vector $\boldsymbol{\ell}$ given by (10.57). Along $\mathscr{L}$, one has

$$\frac{{\rm d}r}{{\rm d}\lambda} = \ell^r = \frac{1}{2} - \frac{mr}{r^2 + a^2},$$

where λ is the parameter associated with $\boldsymbol{\ell}$. In particular, at p, if we denote by r_0 the r-coordinate of p in the Kerr system,

$$\left.\frac{{\rm d}r}{{\rm d}\lambda}\right|_p = \frac{1}{2} - \frac{mr_0}{r_0^2 + a^2} > 0.$$

The above inequality simply translates the fact that $r_0 > r_+$ wherever p lies in $\mathscr{M}_{\rm I}$. Hence, initially r is increasing along $\mathscr{L}$ and we get, since $-mr/(r^2+a^2)$ is an increasing function of r,

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda} \geq \frac{1}{2} - \frac{mr_0}{r_0^2+a^2} =: \alpha > 0.$$

Since α is a constant, we deduce that

$$r \geq r_0 + \alpha(\lambda - \lambda_0),$$

where λ_0 is the value of $\mathscr{L}$'s parameter at p. When $\lambda \to +\infty$, we get $r \to +\infty$, which proves that the null curve $\mathscr{L}$ reaches $\mathscr{I}^+$. Hence we conclude:

$\mathscr{B} = \mathscr{M} \setminus \mathscr{M}_{\rm I}$ is the black hole region, the event horizon of which is $\mathscr{H}$.

Incidentally, since we have already shown that $\mathscr{H}$ is a Killing horizon (cf. Sec. 10.5.1), this illustrates Hawking's strong rigidity theorem discussed in Sec. 5.3.2: the event horizon of Kerr spacetime is a Killing horizon.

According to the discussion in Sec. 5.3.2, we may then call the quantity Ω_H introduced in Eqs. (10.73)-(10.74) the ***black hole rotation velocity***.

10.5.3 Null generators of the event horizon

The null vector field $\boldsymbol{\ell}$ defined by Eq. (10.57) coincides with the Killing vector $\boldsymbol{\chi}$ on $\mathscr{H}$, since $mr/(r^2+a^2) \overset{\mathscr{H}}{=} 1/2$ and $a/(r^2+a^2) \overset{\mathscr{H}}{=} \Omega_H$:

$$\boldsymbol{\ell} \overset{\mathscr{H}}{=} \boldsymbol{\chi}. \tag{10.77}$$

Since (i) $\boldsymbol{\chi}$ is tangent to the null geodesic generators of $\mathscr{H}$, being a null normal to it (cf. Sec. 2.3.3) and (ii) $\boldsymbol{\ell}$ is the tangent vector to the outgoing principal null geodesics $\mathscr{L}^{{\rm out},\mathscr{H}}_{(\theta,\psi)}$ (Sec. 10.4), we conclude that

The null generators of the event horizon $\mathscr{H}$ are the geodesics $\mathscr{L}^{{\rm out},\mathscr{H}}_{(\theta,\psi)}$ of the outgoing principal null congruence.

Similarly, we have

The null generators of the inner horizon $\mathscr{H}_{\rm in}$ are the geodesics $\mathscr{L}^{{\rm out},\mathscr{H}_{\rm in}}_{(\theta,\psi)}$ of the outgoing principal null congruence.

The reader is referred to Fig. 5.1 for a pictorial view of the event horizon $\mathscr{H}$ spanned by the rotating null generators, as well as to Figs. 10.6 and 10.7, where it appears clearly that at $r = r_+$ (resp. $r = r_-$) the outgoing principal null geodesics are tangent to $\mathscr{H}$ (resp. $\mathscr{H}_{\rm in}$).

Remark 1: Since $\Delta = 0$ on $\mathscr{H}$, we read immediately from Eq. (10.61) that

$$\underline{\boldsymbol{\ell}} \overset{\mathscr{H}}{=} \frac{\rho^2}{r^2+a^2}\,\mathbf{d}r,$$

3 which shows that, at any point of $\mathscr{H}$, the vector $\boldsymbol{\ell}$ is normal to the hypersurface $r = \text{const}$ through this point. This hypersurface being nothing but $\mathscr{H}$ itself, we recover the fact that $\boldsymbol{\ell}$ is normal to $\mathscr{H}$.

In view of expression (10.74) for Ω_H, Eq. (10.60) for the parameter ψ labelling the outgoing principal null geodesics $\mathscr{L}^{\text{out},\mathscr{H}}_{(\theta,\psi)}$ can be rewritten as

$$\psi = \tilde{\varphi} - \Omega_H \tilde{t}. \tag{10.78}$$

Hence ψ appears as a corotating azimuthal coordinate on $\mathscr{H}$. Moreover, we verify that the winding speed of the outgoing principal null geodesics given by Eq. (10.54) tends toward Ω_H when approaching the event horizon:

$$\lim_{r\to r_+} \left.\frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}\tilde{t}}\right|_{\mathscr{L}^{\text{out}}_{(u,\theta,\tilde{\varphi})}} = \Omega_H. \tag{10.79}$$

This follows immediately from $\boldsymbol{\ell} \overset{\mathscr{H}}{=} \boldsymbol{\chi} = \boldsymbol{\xi} + \Omega_H \boldsymbol{\eta} = \boldsymbol{\partial}_{\tilde{t}} + \Omega_H \boldsymbol{\partial}_{\tilde{\varphi}}$ and the identity $r_+^2 + a^2 = 2mr_+$ (compare Eq. (10.54) with $r \to r_+$ and Eq. (10.74)).

10.5.4 Surface gravity

In view of the pre-geodesic equation (10.58) satisfied by $\boldsymbol{\ell}$ and the identity (10.77), we deduce that

$$\boxed{\boldsymbol{\nabla}_{\boldsymbol{\chi}} \boldsymbol{\chi} \overset{\mathscr{H}}{=} \kappa \boldsymbol{\chi}}, \tag{10.80}$$

with the non-affinity coefficient given by Eq. (10.59):

$$\kappa = \kappa_{\boldsymbol{\ell}}|_{r=r_+} = \frac{m(r_+^2 - a^2)}{(r_+^2 + a^2)^2}.$$

r_+ being a zero of Δ, we have $r_+^2 + a^2 = 2mr_+$, so that we may rewrite the above expression in terms of r_+ and a only:

$$\kappa = \frac{r_+^2 - a^2}{2r_+(r_+^2 + a^2)}. \tag{10.81}$$

Substituting (10.3) for r_+, we get an expression involving the two basic Kerr parameters:

$$\boxed{\kappa = \frac{\sqrt{m^2 - a^2}}{2m(m + \sqrt{m^2 - a^2})}}. \tag{10.82}$$

Given the strict inequality $a < m$ assumed in this chapter [Eq. (10.1)], we have $\kappa \neq 0$, which, according to the classification introduced in Sec. 3.3.6, means that

> As long as $a < m$, the event horizon $\mathscr{H}$ is a non-degenerate Killing horizon.

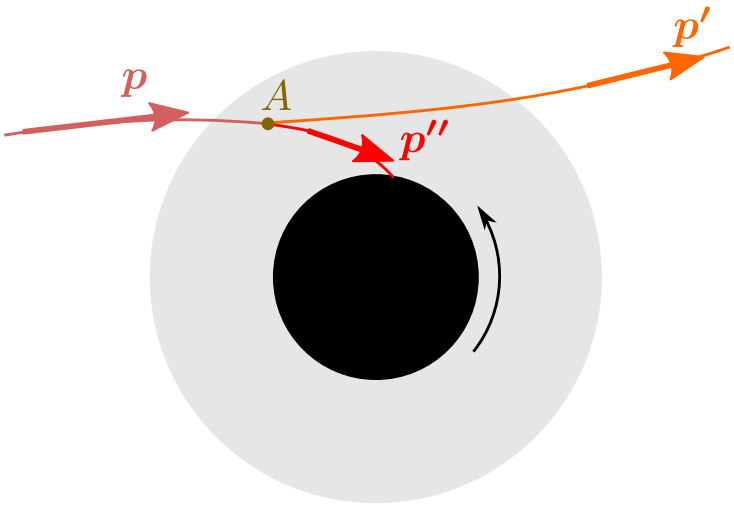


Figure 10.9: Projection in the equatorial plane $t = \text{const}$ and $\theta = \pi/2$ of the worldline of a particle and its 4-momentum $\boldsymbol{p}$, which decay at event A in two particles: one with 4-momentum $\boldsymbol{p'}$, which leaves to infinity, and one with 4-momentum $\boldsymbol{p''}$, which falls into the black hole (black region). The grey zone is the outer ergoregion.

In Sec. 3.3.7, we have seen that the non-affinity coefficient κ can be interpreted as a "rescaled" surface gravity. Hence κ is called the ***black hole surface gravity***. The fact that κ is a constant (i.e. does not depend on θ) is an illustration of the *zeroth law of black hole mechanics* established in Sec. 3.3.5 (cf. in particular Example 14 in that section).

Remark 2: As a check, if we let $a \to 0$ in Eq. (10.82), we get $\kappa = 1/(4m)$, i.e. we recover the Schwarzschild horizon value computed in Example 10 of Chap. 2 [cf. Eq. (2.31)].

10.5.5 The Penrose process

Having established that the black hole event horizon is $\mathscr{H}$, i.e. the boundary between $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm II}$, implies that one can escape from the outer ergoregion $\mathscr{G}^+$ introduced in Sec. 10.2.4, since $\mathscr{G}^+ \subset \mathscr{M}_{\rm I}$. Moreover, we are going to see that $\mathscr{G}^+$ is a region where energy can be extracted "from" the black hole. This results from the stationarity Killing vector $\boldsymbol{\xi}$ being spacelike there — the very definition of an ergoregion. Indeed, let us consider a particle $\mathscr{P}$, of 4-momentum $\boldsymbol{p}$, in free fall from infinity into the outer ergoregion. At some point $A \in \mathscr{G}^+$, the particle $\mathscr{P}$ splits (or decays) into two particles: $\mathscr{P}'$, of 4-momentum $\boldsymbol{p'}$, which leaves to infinity, and $\mathscr{P}''$, of 4-momentum $\boldsymbol{p''}$, which falls into the black hole (cf. Fig. 10.9).

We shall define the *energy gain* in the above scenario by

$$\Delta E = E_{\rm out} - E_{\rm in}, \tag{10.83}$$

where $E_{\rm in}$ (resp. $E_{\rm out}$) is the energy of $\mathscr{P}$ (resp. $\mathscr{P}'$) as measured by an inertial observer at rest with respect to the black hole and far from it (cf. Sec. 1.4):

$$E_{\rm in} = -\,\boldsymbol{\xi}\cdot\boldsymbol{p}|_\infty \qquad \text{and} \qquad E_{\rm out} = -\,\boldsymbol{\xi}\cdot\boldsymbol{p'}|_\infty\,. \tag{10.84}$$

The above formulas hold because the 4-velocity of the inertial observer is precisely the time-translation Killing vector $\boldsymbol{\xi} = \boldsymbol{\partial}_t$ in the far region (since $\boldsymbol{\xi}\cdot\boldsymbol{\xi} = g_{tt} \to -1$ as $r \to +\infty$) (compare with Eq. (1.21)). Assuming that particles $\mathscr{P}$ and $\mathscr{P}'$ are subject only to gravitation, i.e. are in free fall, their worldlines are timelike geodesics. $\boldsymbol{\xi}$ being a Killing vector, it follows then that the scalar products $\boldsymbol{\xi}\cdot\boldsymbol{p}$ and $\boldsymbol{\xi}\cdot\boldsymbol{p'}$ are constant along the worldlines of $\mathscr{P}$ and $\mathscr{P}'$ respectively (cf. Sec. B.5). In particular, we have $\boldsymbol{\xi}\cdot\boldsymbol{p}|_\infty = \boldsymbol{\xi}\cdot\boldsymbol{p}|_A$ and $\boldsymbol{\xi}\cdot\boldsymbol{p'}|_\infty = \boldsymbol{\xi}\cdot\boldsymbol{p'}|_A$. Accordingly, Eqs. (10.83) and (10.84) leads to

$$\Delta E = \boldsymbol{\xi}\cdot\boldsymbol{p}|_A - \boldsymbol{\xi}\cdot\boldsymbol{p'}|_A = \boldsymbol{\xi}|_A\cdot(\boldsymbol{p}|_A - \boldsymbol{p'}|_A).$$

Now, the conservation of energy-momentum at event A writes $\boldsymbol{p}|_A = \boldsymbol{p'}|_A + \boldsymbol{p''}|_A$, hence

$$\Delta E = \boldsymbol{\xi}\cdot\boldsymbol{p''}|_A\,. \tag{10.85}$$

As the 4-momentum of a (non-tachyonic) particle, $\boldsymbol{p''}$ is always a future-directed timelike ($\mathscr{P}''$ massive) or null ($\mathscr{P}''$ massless) vector. If A were located outside the ergoregion, then $\boldsymbol{\xi}$ would be a future-directed timelike vector and the scalar product $\boldsymbol{\xi}\cdot\boldsymbol{p''}$ would necessarily be negative, leading to $\Delta E < 0$, i.e. a net loss of energy. However, in the ergoregion, $\boldsymbol{\xi}$ is a spacelike vector and it is possible to have $\boldsymbol{\xi}\cdot\boldsymbol{p''} > 0$ with $\boldsymbol{p''}$ future-directed timelike or null. If this occurs, then $\Delta E > 0$: the outgoing particle has more energy than the ingoing one. The above process is then called a ***Penrose process***. It is particulary relevant to astrophysics, notably via the so-called *Blanford-Znajek mechanism* or via some magnetohydrodynamical processes (see Ref. [173] for an extended discussion).

Remark 3: Assuming that the particle $\mathscr{P}''$ follows a geodesic, the scalar product $\boldsymbol{\xi}\cdot\boldsymbol{p''}$ in the right-hand side of Eq. (10.85) is actually constant along the worldline of $\mathscr{P}''$, so that the mention of the point A can be omitted in Eq. (10.85).

Remark 4: If the Penrose process does occur, i.e. if $\Delta E > 0$, the particle $\mathscr{P}''$ is called a ***negative-energy particle***. This stems from the fact that the quantity $E''_{\mathscr{P}} := -\boldsymbol{\xi}\cdot\boldsymbol{p''}$ obeys then $E''_{\mathscr{P}} < 0$. However, it should be stressed that $E''_{\mathscr{P}}$ does not correspond to a locally measured energy by some physical observer, as given by formula (1.21), since $\boldsymbol{\xi}$ cannot be identified with any observer 4-velocity, being spacelike in the ergoregion, where $\mathscr{P}''$ lives.

Historical note: The Penrose process has first been suggested by Roger Penrose in 1969, in the review article [210] (cf. the footnote 7 in this article); the detailed calculation has been presented subsequently in an article written with Roger Floyd [211].

10.6 Global quantities

10.6.1 Mass

We have seen in Sec. 10.2.2 that when $r \to +\infty$, the Kerr metric tends towards the Schwarzschild metric of parameter m (cf. Eq. (10.11)); we conclude that m is nothing but the gravitational mass M (cf. Sec. 6.2.4). However, for any asymptotically flat spacetime endowed with an (asymptotically) timelike Killing vector $\boldsymbol{\xi}$, as the Kerr spacetime, there

is a generic definition of the mass, given by an invariant integral: the *Komar mass*. We are going to see explicitly that for the Kerr spacetime, the Komar mass coincides with m.

Let $(\mathscr{M}, \boldsymbol{g})$ be an asymptotically flat spacetime endowed with a Killing vector $\boldsymbol{\xi}$ which is timelike at least near infinity. The ***Komar mass*** is defined as

$$\boxed{M := -\frac{1}{8\pi} \int_{\mathscr{S}} \star(\mathrm{d}\underline{\boldsymbol{\xi}})}, \tag{10.86}$$

where

- $\mathscr{S}$ is a closed spacelike 2-surface;
- $\underline{\boldsymbol{\xi}}$ is the 1-form associated to the Killing vector $\boldsymbol{\xi}$ by metric duality (cf. Sec. A.3.3), i.e. the 1-form of components $\xi_\alpha = g_{\alpha\mu}\xi^\mu$, $\mathrm{d}\underline{\boldsymbol{\xi}}$ is the exterior derivative of $\underline{\boldsymbol{\xi}}$ (cf. Sec. A.4.3 and Eqs. (A.90) and (A.93)):

$$(\mathrm{d}\underline{\boldsymbol{\xi}})_{\alpha\beta} = \partial_\alpha \xi_\beta - \partial_\beta \xi_\alpha = \nabla_\alpha \xi_\beta - \nabla_\beta \xi_\alpha \tag{10.87}$$

  and $\star(\mathrm{d}\underline{\boldsymbol{\xi}})$ is the ***Hodge dual*** of the 2-form $\mathrm{d}\underline{\boldsymbol{\xi}}$, i.e. the 2-form defined by[2]

$$\star\,(\mathrm{d}\underline{\boldsymbol{\xi}})_{\alpha\beta} := \frac{1}{2}(\mathrm{d}\underline{\boldsymbol{\xi}})_{\mu\nu}\,\epsilon^{\mu\nu}{}_{\alpha\beta}, \tag{10.88}$$

  $\boldsymbol{\epsilon}$ being the Levi-Civita tensor associated with the metric $\boldsymbol{g}$ (cf. Sec. A.3.4).

An important property of the Komar mass is that its value does not depend on the choice of the 2-surface $\mathscr{S}$, as long as the latter is located in a vacuum region of spacetime and $\boldsymbol{g}$ fulfils the Einstein equation (see e.g. Sec. 8.6.1 of Ref. [122] for a demonstration).

Remark 1: As the integral of a 2-form over a 2-dimensional manifold, formula (10.86) is well posed.

Thanks to the Killing equation (3.26), one may rewrite Eq. (10.87) as

$$(\mathrm{d}\underline{\boldsymbol{\xi}})_{\alpha\beta} = 2\nabla_\alpha \xi_\beta. \tag{10.89}$$

Taking into account Eq. (10.88), the Komar mass formula (10.86) becomes then

$$M = -\frac{1}{8\pi} \int_{\mathscr{S}} \nabla^\mu \xi^\nu\, \epsilon_{\mu\nu\alpha\beta}, \tag{10.90}$$

where $\omega_{\alpha\beta} := \nabla^\mu \xi^\nu\, \epsilon_{\mu\nu\alpha\beta}$ stands for the 2-form defined by $\boldsymbol{\omega}(\boldsymbol{u}, \boldsymbol{v}) = \nabla^\mu \xi^\nu\, \epsilon_{\mu\nu\rho\sigma} u^\rho v^\sigma$ for any pair of vector fields $(\boldsymbol{u}, \boldsymbol{v})$.

Instead of integrals of 2-forms *along* the 2-surface $\mathscr{S}$, as in (10.86) and (10.90), one may express the Komar mass as a *flux integral*, i.e. the integral of a 2-form contracted with some "surface element", which is *normal* to $\mathscr{S}$. More precisely, let us introduce the ***area element bivector*** by

$$\mathrm{d}S^{\alpha\beta} := (s^\alpha n^\beta - n^\alpha s^\beta)\sqrt{q}\,\mathrm{d}x^2\,\mathrm{d}x^3, \tag{10.91}$$

where

[2] See e.g. Sec. 14.5 of Ref. [123] for an introduction to Hodge duality.

- $\boldsymbol{s}$ is a unit spacelike vector normal to $\mathscr{S}$ and oriented towards the exterior of $\mathscr{S}$, $\boldsymbol{n}$ is a unit timelike vector normal to $\mathscr{S}$ and oriented towards the future[3], such that, at each point $p \in \mathscr{S}$, $(\boldsymbol{n}, \boldsymbol{s})$ is an orthonormal basis of the timelike plane $T_p^\perp\mathscr{S}$ normal to $\mathscr{S}$.
- $(x^a) = (x^2, x^3)$ is a coordinate system on $\mathscr{S}$
- $q = \det(q_{ab})$ is the determinant w.r.t. (x^a) of the metric induced on $\mathscr{S}$ by the spacetime metric $\boldsymbol{g}$, so that $\sqrt{q}\,\mathrm{d}x^2\,\mathrm{d}x^3$ is the area element on $\mathscr{S}$.

To reexpress the Komar mass, we shall use the following identity:
Lemma: for any 2-form $\boldsymbol{A}$ defined in the vicinity of $\mathscr{S}$, one has

$$\int_{\mathscr{S}} \star\boldsymbol{A} = \frac{1}{2}\int_{\mathscr{S}} A_{\mu\nu}\,\mathrm{d}S^{\mu\nu}. \tag{10.92}$$

Proof. Using the definition of the Hodge dual, we have

$$\int_{\mathscr{S}} \star\boldsymbol{A} = \frac{1}{2}\int_{\mathscr{S}} A_{\mu\nu}\,\epsilon^{\mu\nu}{}_{\alpha\beta} = \frac{1}{2}\int_{\mathscr{S}} A^{\mu\nu}\,\epsilon_{\mu\nu\alpha\beta} = \frac{1}{2}\int_{\mathscr{S}} \boldsymbol{A}^\sharp(\boldsymbol{e}^{(\mu)}, \boldsymbol{e}^{(\nu)})\,\boldsymbol{\epsilon}(\boldsymbol{e}_{(\mu)}, \boldsymbol{e}_{(\nu)}, \mathrm{d}\boldsymbol{\ell}_2, \mathrm{d}\boldsymbol{\ell}_3),$$

where $(\boldsymbol{e}_{(\alpha)})$ is an orthonormal tetrad such that $\boldsymbol{e}_{(0)} = \boldsymbol{n}$ and $\boldsymbol{e}_{(1)} = \boldsymbol{s}$, $(\boldsymbol{e}^{(\alpha)})$ is its dual cobasis and $\mathrm{d}\boldsymbol{\ell}_2$ and $\mathrm{d}\boldsymbol{\ell}_3$ are displacement vectors forming elementary parallelograms on $\mathscr{S}$; for instance $\mathrm{d}\boldsymbol{\ell}_2 = \mathrm{d}x^2\,\boldsymbol{\partial}_2$ and $\mathrm{d}\boldsymbol{\ell}_3 = \mathrm{d}x^3\,\boldsymbol{\partial}_3$. Notice that the last equality in the expression above results from the very definition of the integral of a 2-form on a 2-surface. Given the definition of $(\boldsymbol{e}_{(\alpha)})$, $(\boldsymbol{e}_{(2)}, \boldsymbol{e}_{(3)})$ is necessarily a basis of the tangent space $T_p\mathscr{S}$; consequently $\mathrm{d}\boldsymbol{\ell}_2$ and $\mathrm{d}\boldsymbol{\ell}_3$ are linear combinations of $\boldsymbol{e}_{(2)}$ and $\boldsymbol{e}_{(3)}$). Given the alternate character of $\boldsymbol{\epsilon}$, the sum over the indices μ and ν can be then restricted to $\mu, \nu = 0$ or 1. Hence

$$\begin{aligned}
\int_{\mathscr{S}} \star\boldsymbol{A} &= \frac{1}{2}\int_{\mathscr{S}} \left[\boldsymbol{A}^\sharp(\boldsymbol{e}^{(0)}, \boldsymbol{e}^{(1)})\,\boldsymbol{\epsilon}(\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}, \mathrm{d}\boldsymbol{\ell}_2, \mathrm{d}\boldsymbol{\ell}_3) + \boldsymbol{A}^\sharp(\boldsymbol{e}^{(1)}, \boldsymbol{e}^{(0)})\,\boldsymbol{\epsilon}(\boldsymbol{e}_{(1)}, \boldsymbol{e}_{(0)}, \mathrm{d}\boldsymbol{\ell}_2, \mathrm{d}\boldsymbol{\ell}_3)\right] \\
&= \int_{\mathscr{S}} \boldsymbol{A}^\sharp(\boldsymbol{e}^{(0)}, \boldsymbol{e}^{(1)})\,\boldsymbol{\epsilon}(\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}, \mathrm{d}\boldsymbol{\ell}_2, \mathrm{d}\boldsymbol{\ell}_3) \\
&= \int_{\mathscr{S}} A^{(0)(1)}\,\boldsymbol{\epsilon}(\boldsymbol{n}, \boldsymbol{s}, \mathrm{d}\boldsymbol{\ell}_2, \mathrm{d}\boldsymbol{\ell}_3).
\end{aligned}$$

Now, since $(\boldsymbol{e}_{(\alpha)})$ is an orthonormal basis,

$$A^{(0)(1)} = g^{(0)(\mu)}g^{(1)(\nu)}A_{(\mu)(\nu)} = g^{(0)(0)}g^{(1)(1)}A_{(0)(1)} = (-1)\times 1 \times A_{(0)(1)} = -A_{(0)(1)},$$

with

$$A_{(0)(1)} = \boldsymbol{A}(\boldsymbol{e}_{(0)}, \boldsymbol{e}_{(1)}) = \boldsymbol{A}(\boldsymbol{n}, \boldsymbol{s}) = A_{\mu\nu}n^\mu s^\nu = -A_{\mu\nu}s^\mu n^\nu = -\frac{1}{2}A_{\mu\nu}(s^\mu n^\nu - n^\mu s^\nu).$$

On the other side, we recognize in ${}^{\mathscr{S}}\boldsymbol{\epsilon} := \boldsymbol{\epsilon}(\boldsymbol{n}, \boldsymbol{s}, ., .)$ the area element 2-form on $\mathscr{S}$, so that we may write, for $\mathrm{d}\boldsymbol{\ell}_2 = \mathrm{d}x^2\,\boldsymbol{\partial}_2$ and $\mathrm{d}\boldsymbol{\ell}_3 = \mathrm{d}x^3\,\boldsymbol{\partial}_3$,

$$\boldsymbol{\epsilon}(\boldsymbol{n}, \boldsymbol{s}, \mathrm{d}\boldsymbol{\ell}_2, \mathrm{d}\boldsymbol{\ell}_3) = {}^{\mathscr{S}}\epsilon_{ab}\,\mathrm{d}\ell_2^a\,\mathrm{d}\ell_3^b = \sqrt{q}\,\mathrm{d}x^2\,\mathrm{d}x^3.$$

Gathering the above results establishes (10.92). □

[3] The vector $\boldsymbol{n}$ considered here shall not be confused with the vector $\boldsymbol{n}$ introduced in Sec. 10.5.1.

Thanks to (10.92), we may reexpress the Komar mass (10.86) as

$$M = -\frac{1}{16\pi}\int_{\mathscr{S}} (\mathbf{d}\underline{\boldsymbol{\xi}})_{\mu\nu}\, \mathrm{d}S^{\mu\nu}. \tag{10.93}$$

Using the Killing equation as (10.89), this becomes

$$\boxed{M = -\frac{1}{8\pi}\int_{\mathscr{S}} \nabla_\mu \xi_\nu \, \mathrm{d}S^{\mu\nu}}. \tag{10.94}$$

Alternatively, we may express the exterior derivative in terms of partial derivatives and get

$$(\mathbf{d}\underline{\boldsymbol{\xi}})_{\mu\nu}\, \mathrm{d}S^{\mu\nu} = (\partial_\mu \xi_\nu - \partial_\nu \xi_\mu)(s^\mu n^\nu - n^\mu s^\nu)\sqrt{q}\, \mathrm{d}x^2\, \mathrm{d}x^3 = 2(s^\mu \partial_\mu \xi_\nu\, n^\nu - n^\mu \partial_\mu \xi_\nu\, s^\nu)\sqrt{q}\, \mathrm{d}x^2\, \mathrm{d}x^3,$$

so that (10.93) becomes

$$\boxed{M = -\frac{1}{8\pi}\int_{\mathscr{S}} (s^\mu \partial_\mu \xi_\nu\, n^\nu - n^\mu \partial_\mu \xi_\nu\, s^\nu)\sqrt{q}\, \mathrm{d}x^2\, \mathrm{d}x^3}. \tag{10.95}$$

Let us use this last expression to compute the Komar mass of the Kerr spacetime. For this purpose, we shall consider the Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$ and define $\mathscr{S}$ to be a sphere $t = \text{const}$ and $r = \text{const}$. Coordinates on $\mathscr{S}$ are then $(x^2, x^3) = (\theta, \varphi)$. Moreover, since the value of M does not depend on the choice of $\mathscr{S}$, we may set $r \to +\infty$ and use the asymptotic flatness of Kerr metric to get simple expressions. The unit normals to $\mathscr{S}$ are then $\boldsymbol{n} = \boldsymbol{\partial}_t$ and $\boldsymbol{s} = \boldsymbol{\partial}_r$. Moreover, when $r \to +\infty$, $\sqrt{q} = r^2 \sin\theta$. Hence (10.95) yields

$$M = -\frac{1}{8\pi}\lim_{r\to+\infty}\int_{\mathscr{S}} (\partial_r \xi_t - \underbrace{\partial_t \xi_r}_{0}) r^2 \sin\theta \, \mathrm{d}\theta\, \mathrm{d}\varphi,$$

with

$$\xi_t = g_{t\mu}\underbrace{\xi^\mu}_{\delta^\mu{}_t} = g_{tt} \simeq -1 + \frac{2m}{r},$$

the last expression resulting from the expansion (10.11). We have then $\partial_r \xi_t = -2m/r^2$, so that the above integral yields

$$\boxed{M = m}. \tag{10.96}$$

10.6.2 Angular momentum

As a timelike Killing vector gave birth to the Komar mass, an axisymmetry Killing vector gives birth an invariant integral quantity: the Komar angular momentum.

For a spacetime $(\mathscr{M}, \boldsymbol{g})$ that is axisymmetric, with $\boldsymbol{\eta}$ the corresponding Killing vector, one defines the ***Komar angular momentum*** by

$$\boxed{J := \frac{1}{16\pi}\int_{\mathscr{S}} \star(\mathbf{d}\underline{\boldsymbol{\eta}})}, \tag{10.97}$$

where $\mathscr{S}$ is any spacelike surface lying in the vacuum region and $\star(\mathbf{d}\underline{\boldsymbol{\eta}})$ is the Hodge dual of the 2-form $\mathbf{d}\underline{\boldsymbol{\eta}}$, given by

$$(\mathbf{d}\underline{\boldsymbol{\eta}})_{\alpha\beta} = \partial_\alpha \eta_\beta - \partial_\beta \eta_\alpha = \nabla_\alpha \eta_\beta - \nabla_\beta \eta_\alpha. \tag{10.98}$$

As the Komar mass M, J does not depend on the choice of $\mathscr{S}$, provided it lies in the vacuum region and $\boldsymbol{g}$ obeys Einstein equation.

Remark 2: Besides the sign and the change $\boldsymbol{\xi} \leftrightarrow \boldsymbol{\eta}$, we notice a difference by a factor 2 between the r.h.s. of (10.86) and (10.97). This is known as Komar's anomalous factor and is discussed further in Ref. [163].

Performing the same manipulations as for the Komar mass, we obtain

$$\boxed{J = \frac{1}{16\pi} \int_{\mathscr{S}} (s^\mu \partial_\mu \eta_\nu \, n^\nu - n^\mu \partial_\mu \eta_\nu \, s^\nu) \sqrt{q} \, \mathrm{d}x^2 \, \mathrm{d}x^3}. \tag{10.99}$$

As above, let us perform the computation in Boyer-Lindquist coordinates, choosing for $\mathscr{S}$ a 2-sphere $\{t = \text{const}, \ r = \text{const}\}$. In evaluating the terms $s^\mu \partial_\mu \eta_\nu \, n^\nu$ and $n^\mu \partial_\mu \eta_\nu \, s^\nu$ as $r \to +\infty$, we have to be a little more cautious than in Sec. 10.6.1, since one of the components η_α is diverging when $r \to +\infty$:

$$\eta_\alpha = g_{\alpha\mu} \, \eta^\mu = g_{\alpha\mu} \, \delta^\mu{}_\varphi = g_{\alpha\varphi} = (g_{t\varphi}, 0, 0, g_{\varphi\varphi})$$

with, according to (10.11), $g_{\varphi\varphi} \sim r^2 \sin^2\theta$ as $r \to +\infty$. Moreover, given the value of $g_{t\varphi}$ read on (10.8), we may write

$$\eta_\alpha \sim \left(-\frac{2am \sin^2\theta}{r}, 0, 0, r^2 \sin^2\theta \right) \quad \text{when} \quad r \to +\infty.$$

Let us choose for the timelike normal $\boldsymbol{n}$ to $\mathscr{S}$ the future-directed unit normal to the hypersurfaces $t = \text{const}$:

$$\underline{\boldsymbol{n}} = -N \mathbf{d}t, \tag{10.100}$$

where N is a normalization factor ensuring $\boldsymbol{n} \cdot \boldsymbol{n} = -1$. We do not need the precise value[4] of N, but simply the property $N \to 1$ as $r \to +\infty$. We have then $n_\alpha = (-N, 0, 0, 0)$, so that

$$n^\alpha = g^{\alpha\mu} n_\mu = g^{\alpha t}(-N) = (-N g^{tt}, 0, 0, -N g^{t\varphi}),$$

where the last equality follows from the expression (10.16) of $g^{\alpha\beta}$ in Boyer-Lindquist coordinates, with $g^{tt} \sim -1$ and $g^{t\varphi} \sim -2am/r^3$ when $r \to +\infty$; hence

$$n^\alpha \sim \left(1, 0, 0, \frac{2am}{r^3} \right) \quad \text{when} \quad r \to +\infty.$$

The choice of $\boldsymbol{n}$ completely determines that of $\boldsymbol{s}$:

$$s^\alpha = \left(0, \frac{\sqrt{\Delta}}{\rho}, 0, 0 \right) \sim (0, 1, 0, 0) \quad \text{when} \quad r \to +\infty.$$

[4] It is given by Eq. (10.113) below.

Indeed, given the metric components (10.8), we immediately check that $\boldsymbol{n}\cdot\boldsymbol{s} = 0$, $\boldsymbol{s}\cdot\boldsymbol{s} = 1$ and $\underline{\boldsymbol{s}} = (\rho/\sqrt{\Delta})\,\mathbf{d}r$, which does imply that $\boldsymbol{s}$ is normal to $\mathscr{S}$.

Given the above expressions for η_α, n^α and s^α, we get, for $r \to +\infty$,

$$s^\mu \partial_\mu \eta_\nu \, n^\nu \sim \partial_r \left(-\frac{2am\sin^2\theta}{r} \right) \times 1 + \partial_r \left(r^2 \sin^2\theta \right) \times \frac{2am}{r^3} \sim \frac{6am\sin^2\theta}{r^2}$$

and

$$n^\mu \partial_\mu \eta_\nu \, s^\nu \sim \partial_t \underbrace{\eta_r}_{0} + \frac{2am}{r^3} \partial_\varphi \underbrace{\eta_r}_{0} = 0.$$

Hence Eq. (10.99) leads to

$$J = \frac{1}{16\pi} \lim_{r\to+\infty} \int_{\mathscr{S}} \frac{6am\sin^2\theta}{r^2} \times r^2 \sin\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi = \frac{3am}{8\pi} \int_{\mathscr{S}} \sin^3\theta \, \mathrm{d}\theta \, \mathrm{d}\varphi = \frac{3am}{4} \underbrace{\int_0^\pi \sin^3\theta \, \mathrm{d}\theta}_{4/3},$$

i.e.

$$\boxed{J = am}. \tag{10.101}$$

We conclude that the parameter a is nothing but the total angular momentum divided by the total mass.

10.6.3 Black hole area

Since the event horizon $\mathscr{H}$ is a Killing horizon (cf. Sec. 10.5.1), it is a non-expanding horizon. As such, it has a well-defined area A, which is the common area of any of its cross-sections, as we have seen in Sec. 3.2.2. To compute A, we shall not use the Boyer-Lindquist coordinates as for M and J, because they are singular on $\mathscr{H}$; we shall use rather the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$, which are regular on $\mathscr{H}$. $\mathscr{H}$ is defined by $r = r_+$ and it is natural to consider a cross-section of it, $\mathscr{S}$ say, defined by $\{\tilde{t} = \text{const}, r = r_+\}$. Then $\mathscr{S}$ is spanned by the coordinates $(x^a) = (\theta, \tilde{\varphi})$ and the metric $\boldsymbol{q}$ induced on it by the spacetime metric is obtained by setting $r = r_+$, $\mathrm{d}\tilde{t} = 0$ and $\mathrm{d}r = 0$ in (10.36):

$$q_{ab}\,\mathrm{d}x^a \mathrm{d}x^b = (r_+^2 + a^2\cos^2\theta)\,\mathrm{d}\theta^2 + \left(r_+^2 + a^2 + \frac{2a^2 m r_+ \sin^2\theta}{r_+^2 + a^2\cos^2\theta} \right) \sin^2\theta \, \mathrm{d}\tilde{\varphi}^2.$$

Now, since r_+ is a zero of Δ [cf. Eq. (10.10)], we have $2mr_+ = r_+^2 + a^2$. Hence

$$q_{ab}\,\mathrm{d}x^a \mathrm{d}x^b = (r_+^2 + a^2\cos^2\theta)\,\mathrm{d}\theta^2 + (r_+^2 + a^2) \left(1 + \frac{a^2\sin^2\theta}{r_+^2 + a^2\cos^2\theta} \right) \sin^2\theta \, \mathrm{d}\tilde{\varphi}^2,$$

or, equivalently,

$$q_{ab}\,\mathrm{d}x^a \mathrm{d}x^b = (r_+^2 + a^2\cos^2\theta)\,\mathrm{d}\theta^2 + \frac{(r_+^2 + a^2)^2}{r_+^2 + a^2\cos^2\theta} \sin^2\theta \, \mathrm{d}\tilde{\varphi}^2. \tag{10.102}$$

The area of $\mathscr{S}$ is

$$A = \int_{\mathscr{S}} \sqrt{q}\, \mathrm{d}\theta\, \mathrm{d}\tilde{\varphi},$$

with, according to (10.102), $q = \det(q_{ab}) = (r_+^2 + a^2)^2 \sin^2\theta$. Hence

$$A = (r_+^2 + a^2) \underbrace{\int_{\mathscr{S}} \sin\theta\, \mathrm{d}\theta\, \mathrm{d}\tilde{\varphi}}_{4\pi}.$$

We have thus

$$\boxed{A = 4\pi(r_+^2 + a^2) = 8\pi m r_+}. \tag{10.103}$$

Via (10.3), one may recast this result to let appear only m and a:

$$\boxed{A = 8\pi m(m + \sqrt{m^2 - a^2})}. \tag{10.104}$$

10.7 Observers in Kerr spacetime

The concept of *observer* in a relativistic spacetime has been recalled in Sec. 1.4. We discuss here some families of observers well adapted to Kerr spacetime: the ZAMOs (Sec. 10.7.3) and the Carter observers (Sec. 10.7.4). These two families are actually particular cases of a more general concept, that of a *stationary observer*, which we introduce first.

10.7.1 Stationary observers

A ***stationary observer*** is an observer $\mathscr{O}$ in Kerr spacetime whose 4-velocity $\boldsymbol{u}$ is a linear combination of the Killing vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with constant coefficients[5]:

$$\boldsymbol{u} = \alpha\, \boldsymbol{\xi} + \beta\, \boldsymbol{\eta}, \qquad \alpha = \text{const}, \ \beta = \text{const}. \tag{10.105}$$

It follows that the worldline $\mathscr{L}$ of $\mathscr{O}$ is an orbit of the isometry group $\mathbb{R} \times \mathrm{SO}(3)$ of Kerr spacetime. In physical terms, this means that the spacetime geometry as perceived by observer $\mathscr{O}$ does not evolve, hence the name *stationary observer*.

A stationary observer moves necessarily at fixed values of the non-ignorable coordinates in a coordinate system adapted to the spacetime symmetries, like Boyer-Lindquist coordinates or Kerr ones. For instance, considering the Boyer-Lindquist coordinates (t, r, θ, φ), we have $\boldsymbol{\xi} = \boldsymbol{\partial}_t$ and $\boldsymbol{\eta} = \boldsymbol{\partial}_\varphi$ and we deduce from (10.105) and the definition (1.14) of the 4-velocity that along the worldline $\mathscr{L}$:

$$\frac{\mathrm{d}t}{\mathrm{d}\tau} = \alpha, \quad \frac{\mathrm{d}r}{\mathrm{d}\tau} = 0, \quad \frac{\mathrm{d}\theta}{\mathrm{d}\tau} = 0, \quad \frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = \beta,$$

where τ is the observer's proper time. It follows that the stationary observer evolves at fixed values of the coordinates r and θ.

[5] In the definition (10.105), the coefficients α and β are required to be constant along a given observer's worldline. When considering a family of observers, they may vary from one worldline to the other.

Outside the Carter time machine $\mathscr{T}$ (cf. Sec. 10.2.5), we have necessarily $\alpha \neq 0$, because $\boldsymbol{\eta}$ is spacelike in $\mathscr{M} \setminus \mathscr{T}$ and the 4-velocity $\boldsymbol{u}$ is necessarily timelike. Hence, by introducing $\omega := \beta/\alpha$, we may rewrite (10.105) as

$$\boxed{\boldsymbol{u} = \alpha\,(\boldsymbol{\xi} + \omega\,\boldsymbol{\eta})}, \qquad \alpha = \text{const}, \ \omega = \text{const}. \tag{10.106}$$

The coefficient ω gives the rate of variation of the azimuthal coordinate along the observer's worldline $\mathscr{L}$ in any coordinate system adapted to the spacetime symmetries, like the Boyer-Lindquist ones (t, r, θ, φ), the Kerr ones $(\tilde{t}, r, \theta, \tilde{\varphi})$, or the null Kerr ones $(v, r, \theta, \tilde{\varphi})$, according to

$$\boxed{\omega = \left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{\mathscr{L}} = \left.\frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}v}\right|_{\mathscr{L}} = \left.\frac{\mathrm{d}\tilde{\varphi}}{\mathrm{d}\tilde{t}}\right|_{\mathscr{L}}}. \tag{10.107}$$

Proof. Let us consider Boyer-Lindquist coordinates (t, r, θ, φ). Denoting by τ the proper time of $\mathscr{O}$, we have $\mathrm{d}\varphi/\mathrm{d}t|_{\mathscr{L}} = \mathrm{d}\varphi/\mathrm{d}\tau \times \mathrm{d}\tau/\mathrm{d}t$. Now by the definition (1.14) of the 4-velocity, $\mathrm{d}\varphi/\mathrm{d}\tau = u^\varphi$ and $\mathrm{d}\tau/\mathrm{d}t = (u^t)^{-1}$. We have thus $\mathrm{d}\varphi/\mathrm{d}t|_{\mathscr{L}} = u^\varphi/u^t$. From Eq. (10.106) along with $\boldsymbol{\xi} = \boldsymbol{\partial}_t$ and $\boldsymbol{\eta} = \boldsymbol{\partial}_\varphi$, we read $u^t = \alpha$ and $u^\varphi = \alpha\omega$. Hence $\mathrm{d}\varphi/\mathrm{d}t|_{\mathscr{L}} = \omega$. The same demonstration applies to Kerr and null Kerr coordinates because $\boldsymbol{\xi} = \boldsymbol{\partial}_v = \boldsymbol{\partial}_{\tilde{t}}$ and $\boldsymbol{\eta} = \boldsymbol{\partial}_{\tilde{\varphi}}$. □

Remark 1: That the ratios $\mathrm{d}\varphi/\mathrm{d}t$, $\mathrm{d}\tilde{\varphi}/\mathrm{d}v$ and $\mathrm{d}\tilde{\varphi}/\mathrm{d}\tilde{t}$ along $\mathscr{L}$ are all equal, as expressed in (10.107), is not surprising if one considers the links between the various coordinates expressed by Eqs. (10.23a), (10.23b) and (10.34). Setting $\mathrm{d}r = 0$ in these equations, since r is constant along $\mathscr{L}$, we get

$$\mathrm{d}t|_{\mathscr{L}} = \mathrm{d}v|_{\mathscr{L}} = \mathrm{d}\tilde{t}|_{\mathscr{L}} \qquad \text{and} \qquad \mathrm{d}\varphi|_{\mathscr{L}} = \mathrm{d}\tilde{\varphi}|_{\mathscr{L}}.$$

Equation (10.107) allows for a nice physical interpretation of ω. Indeed, we have seen in Sec. 7.3.3 [cf. Eq. (7.58)] that $\mathrm{d}\varphi/\mathrm{d}t|_{\mathscr{L}}$ is nothing but the angular velocity around the symmetry axis as measured by an asymptotically distant inertial observer. The demonstration was performed in the Schwarzschild case and for a circular orbit in the equatorial plane, but it used only the stationarity of Schwarzschild spacetime and the fact that $\mathrm{d}\varphi/\mathrm{d}t$ was constant along $\mathscr{L}$, so it applies to the present case as well.

10.7.2 Static observers

A ***static observer*** is a stationary observer $\mathscr{O}$ having $\omega = 0$. The denomination stems from the fact that the three coordinates (r, θ, φ) (or $(r, \theta, \tilde{\varphi})$) remain constant along $\mathscr{O}$'s worldline, since $\omega = 0$ in Eq. (10.107) implies $\varphi = \text{const}$ and $\tilde{\varphi} = \text{const}$. Moreover, such an observer appears not moving to an asymptotic inertial observer.

According to (10.106) with $\omega = 0$, the 4-velocity of $\mathscr{O}$ is collinear to the Killing vector $\boldsymbol{\xi}$: $\boldsymbol{u} = \alpha\boldsymbol{\xi}$. We conclude that a static observer can exist only where $\boldsymbol{\xi}$ is timelike, i.e. outside the ergoregion (cf. Sec. 10.2.4). This explains why the outer boundary of the ergoregion (the ergosphere) is sometimes called the *static limit* (cf. Remark 3 on p. 265).

Since static observers are not very useful for describing physical processes in the vicinity of a Kerr black hole (in particular, they cannot exist close to the event horizon, which is located in the ergoregion), we shall not discuss them further.

10.7.3 Zero-angular-momentum observers (ZAMO)

Let us consider an observer $\mathscr{O}$ whose worldline is normal to the hypersurfaces of constant Boyer-Lindquist coordinate t, Σ_t say. Of course, such an observer exists only where the hypersurface Σ_t is spacelike, so that the latter has a timelike normal (cf. Sec. 2.2.2). A normal 1-form to Σ_t is of course $\mathbf{d}t$. The associated normal vector, $\boldsymbol{N}$ say, is obtained by metric duality [cf. Eq. (A.42)-(A.43)]:

$$\boldsymbol{N} = -\overrightarrow{\mathbf{d}t} = -\overrightarrow{\boldsymbol{\nabla}}t. \tag{10.108}$$

In components:

$$N^\alpha = -g^{\alpha\mu}(\mathrm{d}t)_\mu = -g^{\alpha\mu}\delta^t{}_\mu = -g^{t\alpha}.$$

The minus sign has been chosen to have $\boldsymbol{N}$ future-directed, as we shall see below. In view of the Boyer-Lindquist components (10.16) of the inverse metric, we get

$$\boldsymbol{N} = \frac{1}{\Delta}\left(r^2 + a^2 + \frac{2a^2mr\sin^2\theta}{\rho^2}\right)\boldsymbol{\partial}_t + \frac{2amr}{\rho^2\Delta}\boldsymbol{\partial}_\varphi. \tag{10.109}$$

The scalar square of $\boldsymbol{N}$ is

$$\boldsymbol{N}\cdot\boldsymbol{N} = \boldsymbol{g}(\boldsymbol{N},\boldsymbol{N}) = N_\mu N^\mu = \delta^t{}_\mu N^\mu = N^t = -\frac{1}{\Delta}\left(r^2 + a^2 + \frac{2a^2mr\sin^2\theta}{\rho^2}\right). \tag{10.110}$$

Now, the quantity inside the parentheses is positive everywhere except in the Carter time machine $\mathscr{T} \subset \mathscr{M}_{\mathrm{III}}$ discussed in Sec. 10.2.5. Indeed, up to the factor $\sin^2\theta \geq 0$, it coincides with expression (10.21) of $\boldsymbol{\eta}\cdot\boldsymbol{\eta}$. Since $\Delta = (r-r_+)(r-r_-)$ is positive on $\mathscr{M}_{\mathrm{I}}$ and $\mathscr{M}_{\mathrm{III}}$, and negative on $\mathscr{M}_{\mathrm{II}}$, we conclude that the locus where $\boldsymbol{N}$ is timelike is

$$\mathscr{M}_{\mathrm{ZAMO}} := \mathscr{M}_{\mathrm{I}} \cup (\mathscr{M}_{\mathrm{III}} \setminus \mathscr{T}). \tag{10.111}$$

This is thus the region where the observer $\mathscr{O}$ is defined. Note that it does not contain the horizons $\mathscr{H}$ and $\mathscr{H}_{\mathrm{in}}$, which is not surprising since the Boyer-Lindquist coordinates are singular there, notably in terms of the spacetime slicing by the hypersurfaces Σ_t, as illustrated in Fig. 6.2 for the case $a = 0$.

In all $\mathscr{M}_{\mathrm{ZAMO}}$, the timelike vector $\boldsymbol{N}$ is future-directed with respect to the time orientation chosen in Sec. 10.3.2. Indeed, the latter is set by the global null vector field $\boldsymbol{k}$ and we have, using the Boyer-Lindquist components (10.49) of $\boldsymbol{k}$:

$$\boldsymbol{k}\cdot\boldsymbol{N} = N_\mu k^\mu = -k^t = -\frac{r^2+a^2}{\Delta} < 0 \text{ on } \mathscr{M}_{\mathrm{ZAMO}},$$

so that Lemma 2 in Sec. 1.2.2 [Eq. (1.4a)] let us conclude that $\boldsymbol{N}$ is future-directed.

Choosing the observer $\mathscr{O}$ as having his worldline orthogonal to Σ_t means that $\mathscr{O}$'s 4-velocity $\boldsymbol{n}$ is the unit timelike vector introduced by Eq (10.100). Equivalently, $\boldsymbol{n}$ is $\boldsymbol{N}$ rescaled to form a unit vector:

$$\boldsymbol{n} = N\boldsymbol{N} = -N\overrightarrow{\boldsymbol{\nabla}}t, \tag{10.112}$$

with[6] $N := (-\boldsymbol{N}\cdot\boldsymbol{N})^{-1/2}$. In view of Eq. (10.110), we get

$$N = \sqrt{\Delta}\left(r^2 + a^2 + \frac{2a^2mr\sin^2\theta}{\rho^2}\right)^{-1/2}. \tag{10.113}$$

N is called the ***lapse function***, for it relates the increment $\mathrm{d}\tau$ in the proper time of $\mathscr{O}$ to the change $\mathrm{d}t$ of the coordinate t when moving from Σ_t to $\Sigma_{t+\mathrm{d}t}$ via

$$\boxed{\mathrm{d}\tau = N\mathrm{d}t}. \tag{10.114}$$

Proof. The infinitesimal vector that connects the point of proper time τ to that of proper time $\tau+\mathrm{d}\tau$ along $\mathscr{O}$'s worldline is $\mathrm{d}\boldsymbol{x} = \mathrm{d}\tau\,\boldsymbol{n}$ (by the very definition of the 4-velocity $\boldsymbol{n}$, compare Eq. (1.14)). The corresponding increment in the coordinate t is given by formula (A.20): $\mathrm{d}t = \langle \mathbf{d}t, \mathrm{d}\boldsymbol{x}\rangle$; we have then, using Eq. (10.112) to express $\vec{\boldsymbol{\nabla}}t$,

$$\mathrm{d}t = \langle \mathbf{d}t, \mathrm{d}\boldsymbol{x}\rangle = \vec{\boldsymbol{\nabla}}t\cdot\mathrm{d}\boldsymbol{x} = \mathrm{d}\tau\,\vec{\boldsymbol{\nabla}}t\cdot\boldsymbol{n} = -\frac{\mathrm{d}\tau}{N}\underbrace{\boldsymbol{n}\cdot\boldsymbol{n}}_{-1} = \frac{\mathrm{d}\tau}{N},$$

hence formula (10.114). □

A ***zero-angular-momentum observer (ZAMO)*** is an observer $\mathscr{O}$ of the above type, i.e. whose worldline $\mathscr{L}$ is normal to the hypersurfaces Σ_t of constant Boyer-Lindquist time t, and whose orthonormal frame $(\boldsymbol{e}_{(\alpha)})$ is related to the Boyer-Lindquist coordinate frame $(\boldsymbol{\partial}_\alpha)$ by

$$\boldsymbol{e}_{(0)} = \boldsymbol{n} = \frac{\sqrt{\rho^2(r^2+a^2)+2a^2mr\sin^2\theta}}{\rho\sqrt{\Delta}}\,\boldsymbol{\partial}_t + \frac{2amr}{\rho\sqrt{\Delta[\rho^2(r^2+a^2)+2a^2mr\sin^2\theta]}}\,\boldsymbol{\partial}_\varphi \tag{10.115a}$$

$$\boldsymbol{e}_{(r)} = \frac{\sqrt{\Delta}}{\rho}\,\boldsymbol{\partial}_r \tag{10.115b}$$

$$\boldsymbol{e}_{(\theta)} = \frac{1}{\rho}\,\boldsymbol{\partial}_\theta \tag{10.115c}$$

$$\boldsymbol{e}_{(\varphi)} = \frac{\rho}{\sin\theta\sqrt{\rho^2(r^2+a^2)+2a^2mr\sin^2\theta}}\,\boldsymbol{\partial}_\varphi, \tag{10.115d}$$

where $\rho := \sqrt{r^2+a^2\cos^2\theta}$ (the positive square root of ρ^2, which is defined by Eq. (10.9)). Expression (10.115a) for $\boldsymbol{n}$ has been obtained by combining Eqs. (10.112), (10.113) and (10.109). Given the Boyer-Lindquist components (10.8) of $\boldsymbol{g}$, one readily check that $(\boldsymbol{e}_{(r)}, \boldsymbol{e}_{(\theta)}, \boldsymbol{e}_{(\varphi)})$ is an orthonormal basis of spacelike vectors (see the notebook D.4.5 for a SageMath computation). Moreover, these vectors are all tangent to Σ_t, since $(\boldsymbol{\partial}_r, \boldsymbol{\partial}_\theta, \boldsymbol{\partial}_\varphi)$ are. They are thus orthogonal to $\boldsymbol{n} = \boldsymbol{e}_{(0)}$, which is a unit timelike vector. This completes the proof that $(\boldsymbol{e}_{(\alpha)})$ is an orthonormal frame.

[6] Note that N is not defined as the (pseudo)norm of $\boldsymbol{N}$, as the notation might suggest, but rather as the inverse of it.

The ***ZAMO coframe*** is the 4-tuple of 1-forms $(\boldsymbol{e}^{(\alpha)})_{0\leq\alpha\leq 3}$ that constitutes, at each point $p \in \mathscr{L}$, a dual basis of $(\boldsymbol{e}_{(\alpha)}(p))$, namely $(\boldsymbol{e}^{(\alpha)})$ obeys $\langle \boldsymbol{e}^{(\alpha)}, \boldsymbol{e}_{(\beta)}\rangle = \delta^{\alpha}{}_{\beta}$ (cf. Eq. (A.22) in Appendix A). Its expression in terms of the Boyer-Lindquist coordinate coframe $(\mathbf{d}x^{\alpha})$ is (cf. the notebook D.4.5)

$$\boldsymbol{e}^{(0)} = \frac{\rho\sqrt{\Delta}}{\sqrt{\rho^2(r^2+a^2)+2a^2mr\sin^2\theta}}\,\mathbf{d}t \tag{10.116a}$$

$$\boldsymbol{e}^{(r)} = \frac{\rho}{\sqrt{\Delta}}\,\mathbf{d}r \tag{10.116b}$$

$$\boldsymbol{e}^{(\theta)} = \rho\,\mathbf{d}\theta \tag{10.116c}$$

$$\boldsymbol{e}^{(\varphi)} = -\frac{2amr\sin\theta}{\rho\sqrt{\rho^2(r^2+a^2)+2a^2mr\sin^2\theta}}\,\mathbf{d}t + \frac{\sin\theta}{\rho}\sqrt{\rho^2(r^2+a^2)+2a^2mr\sin^2\theta}\,\mathbf{d}\varphi. \tag{10.116d}$$

Each ZAMO can be characterized by its coordinates $(r_0, \theta_0, \varphi_0)$ at some fixed value t_0 of the Boyer-Lindquist time t. The set of ZAMOs is thus a 3-parameter family of observers filling $\mathscr{M}_{\rm ZAMO}$. The coordinates (r, θ, φ) span each hypersurface Σ_t. Contrary to the ZAMO worldlines, the curves of fixed (r, θ, φ), the tangent vector of which is $\boldsymbol{\partial}_t = \boldsymbol{\xi}$, are not orthogonal to Σ_t, except for $a = 0$. The orthogonal decomposition of $\boldsymbol{\xi}$ into a part along the normal $\boldsymbol{n}$ and a part tangent to Σ_t defines the ***shift vector*** $\boldsymbol{\beta}$:

$$\boxed{\boldsymbol{\xi} = N\,\boldsymbol{n} + \boldsymbol{\beta}}, \qquad \boldsymbol{n}\cdot\boldsymbol{\beta} = 0. \tag{10.117}$$

From Eq. (10.115a), we get

$$\boldsymbol{\beta} = -\frac{2amr}{\rho^2(r^2+a^2)+2a^2mr\sin^2\theta}\,\boldsymbol{\partial}_{\varphi}. \tag{10.118}$$

Remark 2: The terms *lapse* and *shift vector* are those used in the so-called *3+1 formalism of general relativity* (see e.g. Refs. [8, 20, 122, 232]). In this context, the ZAMO is called the ***Eulerian observer***, which is the generic denomination of the observer whose worldline is normal to the hypersurfaces Σ_t that constitute the 3+1 foliation of spacetime.

The ZAMO rotation velocity seen from infinity ω is obtained by comparing Eqs. (10.106) and Eq. (10.117) rewritten as $\boldsymbol{n} = N^{-1}(\boldsymbol{\xi} - \beta^{\varphi}\,\boldsymbol{\eta})$. We thus get immediately $\omega = -\beta^{\varphi}$. Hence

$$\boxed{\omega := \left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{\mathscr{L}} = \frac{2amr}{\rho^2(r^2+a^2)+2a^2mr\sin^2\theta}}. \tag{10.119}$$

Note that $\omega = 0$ for $a = 0$ (Schwarzschild black hole) and that ω decays quite rapidly with r:

$$\omega \underset{r\to\pm\infty}{\sim} \frac{2am}{r^3} = \frac{2J}{r^3}, \tag{10.120}$$

where Eq. (10.101) has been used to let appear the black hole angular momentum J.

A ZAMO is not an inertial observer: it is not in free-fall, since its 4-acceleration $\boldsymbol{a} := \boldsymbol{\nabla}_{\boldsymbol{n}} \boldsymbol{n}$ is nonzero. Indeed, for any family of observers orthogonal to a spacelike foliation $(\Sigma_t)_{t\in\mathbb{R}}$, it can be shown that $\boldsymbol{a}$ is the orthogonal projection onto Σ_t of the gradient of the logarithm of the lapse function (see e.g. Eq. (4.19) in Ref. [122]): $\boldsymbol{a} = \vec{\boldsymbol{\nabla}} \ln N + (\boldsymbol{\nabla}_{\boldsymbol{n}} \ln N)\, \boldsymbol{n}$. In the present case, this expression simplifies since $\boldsymbol{\nabla}_{\boldsymbol{n}} \ln N = n^\mu \partial_\mu \ln N = 0$, as a result of $\boldsymbol{n} = n^t \boldsymbol{\partial}_t + n^\varphi \boldsymbol{\partial}_\varphi$. We then get $\boldsymbol{a} = \vec{\boldsymbol{\nabla}} \ln N$, i.e. $a^\alpha = g^{\alpha\mu} \partial_\mu \ln N$, so that

$$\boldsymbol{a} = \frac{\Delta}{\rho^2 N} \frac{\partial N}{\partial r} \boldsymbol{\partial}_r + \frac{1}{\rho^2 N} \frac{\partial N}{\partial \theta} \boldsymbol{\partial}_\theta .$$

Using expression (10.113) for N, we get[7]

$$\boxed{\boldsymbol{a} = \frac{m[\rho^2(r^4 - a^4) + 2\Delta(ra\sin\theta)^2]}{\rho^3\sqrt{\Delta}[\rho^2(r^2+a^2) + 2a^2 mr \sin^2\theta]} \boldsymbol{e}_{(r)} - \frac{2a^2 mr(r^2+a^2)\sin\theta\cos\theta}{\rho^3[\rho^2(r^2+a^2) + 2a^2 mr\sin^2\theta]} \boldsymbol{e}_{(\theta)}} . \tag{10.121}$$

We have thus $\boldsymbol{a} \neq 0$ as soon as $m \neq 0$. In other words, the ZAMO's worldline is not a geodesic (cf. Eq. (B.1) in Appendix B) and the ZAMO feels some acceleration, the stronger, the closer to the black hole. Far from the black hole, we have $\sqrt{\Delta} \sim |r|$, $\rho \sim |r|$ so that $\rho^3\sqrt{\Delta} \sim r^4$ and we get

$$\boldsymbol{a} \underset{r\to\pm\infty}{\sim} \frac{m}{r^2} \boldsymbol{e}_{(r)} . \tag{10.122}$$

In the asymptotic regions, the non-relativistic gravitational field felt by the observer is $\boldsymbol{g} = -\boldsymbol{a}$, so that we recover the standard Newtonian expression for $r \to +\infty$ ($\mathscr{M}_{\rm I}$ side). In the asymptotic region of $\mathscr{M}_{\rm III}$, i.e. for $r \to -\infty$, $\boldsymbol{e}_{(r)} = (\sqrt{\Delta}/\rho)\, \boldsymbol{\partial}_r$ is oriented towards the black hole, so that $\boldsymbol{g}$ points outward, i.e. is a repelling field. This is of course in agreement with the negative mass aspect of the Kerr metric in region $\mathscr{M}_{\rm III}$ discussed in Sec. 10.2.2.

Despite ω as given by Eq. (10.119) is nonzero for $a \neq 0$, a ZAMO has a vanishing angular momentum about the rotation axis, hence the name *zero-angular momentum observer*. Indeed, the observer's specific "angular momentum", loosely defined by $\ell := \boldsymbol{\eta}\cdot\boldsymbol{n}$ [cf. Eq. (7.26)], is identically zero, the observer's 4-velocity $\boldsymbol{n}$ being orthogonal to $\boldsymbol{\eta} = \boldsymbol{\partial}_\varphi$. By *losely defined*, we mean that the above definition of ℓ fully makes sense for a *geodesic*, for which it leads to a conserved quantity along the worldline (cf. Sec. 7.2.1), and we are going to see that a ZAMO's worldline is not a geodesic. However, a ZAMO shares with a $\ell = 0$ geodesic crossing his worldline the same value ω of the angular velocity as seen from infinity, as we shall see in Sec. 11.3.3 [cf. Eq. (11.63)].

The name *locally nonrotating observer*, initially given to a ZAMO (cf. historical note below), conveys other specificities of such a observer $\mathscr{O}$:

- $\mathscr{O}$ does not measure any component along $\boldsymbol{e}_{(\varphi)}$ for the 3-momentum $\boldsymbol{P}$ of any null or timelike geodesic that has a zero conserved angular momentum ($L = 0$):

$$P^{(\varphi)} = \boldsymbol{e}_{(\varphi)} \cdot \boldsymbol{P} = \boldsymbol{e}_{(\varphi)} \cdot [\boldsymbol{p} + (\boldsymbol{n}\cdot\boldsymbol{p})\, \boldsymbol{n})] = \boldsymbol{e}_{(\varphi)} \cdot \boldsymbol{p} = e^\varphi_{(\varphi)}\, \boldsymbol{\eta}\cdot\boldsymbol{p} = e^\varphi_{(\varphi)}\, L = 0,$$

[7] See the notebook D.4.5; see also Eqs. (70)-(71) of Ref. [230] or Eqs. (A.9) and (A.10) of Ref. [25].

where $\boldsymbol{p}$ stands for the particle's 4-momentum and we have used expression (1.22) for the 3-momentum, along with $\boldsymbol{e}_{(\varphi)} \cdot \boldsymbol{n} = 0$, formula (10.115d) for $\boldsymbol{e}_{(\varphi)}$ and the definition $L := \boldsymbol{\eta} \cdot \boldsymbol{p}$ of the conserved angular momentum [cf. Eq. (7.1b), as well as Eq. (11.2b) below].

- for $\mathscr{O}$, the directions $\boldsymbol{e}_{(\varphi)}$ and $-\boldsymbol{e}_{(\varphi)}$ are equivalent, insofar as the proper time measured by $\mathscr{O}$ for a light signal to perform a full circuit on a circle at $r = \text{const}$, $\theta = \text{const}$ is the same for a forward motion (increasing φ) of the signal as for a backward one (decreasing φ) (see Ref. [16] for details).

However, a ZAMO is not totally "non rotating", for he has a nonzero *4-rotation* as soon as $a \neq 0$. Let us recall that the ***4-rotation vector*** of an observer $\mathscr{O}$ of 4-velocity $\boldsymbol{n}$ and 4-acceleration $\boldsymbol{a}$ is defined as the spacelike vector $\boldsymbol{\omega}_{\rm rot}$ orthogonal to $\boldsymbol{n}$ such that the evolution of $\mathscr{O}$'s orthonormal frame $(\boldsymbol{e}_{(\alpha)})$ along $\mathscr{O}$'s worldline $\mathscr{L}$ takes the form (see e.g. Sec. 13.6 of Ref. [190] or Sec. 4.5 of Ref. [123])

$$\boldsymbol{\nabla}_{\boldsymbol{n}} \boldsymbol{e}_{(\alpha)} = \underbrace{(\boldsymbol{a} \cdot \boldsymbol{e}_{(\alpha)})\boldsymbol{n} - (\boldsymbol{n} \cdot \boldsymbol{e}_{(\alpha)})\boldsymbol{a}}_{\boldsymbol{\Omega}_{\rm FW}(\boldsymbol{e}_{(\alpha)})} + \boldsymbol{\omega}_{\rm rot} \times_{\boldsymbol{n}} \boldsymbol{e}_{(\alpha)}, \tag{10.123}$$

where the cross product in the hyperplane orthogonal to $\boldsymbol{n}$, $\boldsymbol{\omega}_{\rm rot} \times_{\boldsymbol{n}} \boldsymbol{e}_{(\alpha)}$, is the unique vector orthogonal to both $\boldsymbol{n}$ and $\boldsymbol{\omega}_{\rm rot}$ such that $(\boldsymbol{\omega}_{\rm rot} \times_{\boldsymbol{n}} \boldsymbol{e}_{(\alpha)}) \cdot \boldsymbol{v} = \boldsymbol{\epsilon}(\boldsymbol{n}, \boldsymbol{\omega}_{\rm rot}, \boldsymbol{e}_{(\alpha)}, \boldsymbol{v})$ for any vector $\boldsymbol{v}$, $\boldsymbol{\epsilon}$ being the Levi-Civita tensor (cf. Sec. A.3.4). The $\boldsymbol{\Omega}_{\rm FW}$ operator that appears in the right-hand side of Eq. (10.123) is called the ***Fermi-Walker operator***. It appears as soon as the observer is accelerating, even if he is nonrotating. It corrects the parallel transport of $(\boldsymbol{e}_{(\alpha)})$ along $\mathscr{L}$, which would be realized by $\boldsymbol{\nabla}_{\boldsymbol{n}} \boldsymbol{e}_{(\alpha)} = 0$, to ensure that $(\boldsymbol{e}_{(\alpha)})$ remains an orthonormal frame. A vector field $\boldsymbol{v}$ that obeys $\boldsymbol{\nabla}_{\boldsymbol{n}} \boldsymbol{e} = \boldsymbol{\Omega}_{\rm FW}(\boldsymbol{v})$ is said to be ***Fermi-Walker transported*** along $\mathscr{L}$. Physically, Fermi-Walker transport is realized by a free gyroscope: its spin vector with respect to $\mathscr{O}$ is Fermi-Walker transported along $\mathscr{L}$. Hence the 4-rotation $\boldsymbol{\omega}_{\rm rot}$ of an observer, which is an *absolute* quantity (like the 4-acceleration $\boldsymbol{a}$, it does not depend on any other observer), measures the motion of his spacelike triad $(\boldsymbol{e}_{(i)})$ with respect to an orthonormal triad whose vectors are aligned along gyroscopes axes. For the ZAMO, the 4-rotation turns out to be[8]

$$\boxed{\boldsymbol{\omega}_{\rm rot} = -\frac{\omega}{\rho^3} \left\{ a^2 \sqrt{\Delta} \cos\theta \sin^2\theta \, \boldsymbol{e}_{(r)} + \left[r(r^2 + a^2) + \frac{\rho^2}{2r}(r^2 - a^2) \right] \sin\theta \, \boldsymbol{e}_{(\theta)} \right\}}, \tag{10.124}$$

where ω is given by Eq. (10.119). Hence, as soon as $a \neq 0$, $\boldsymbol{\omega}_{\rm rot} \neq 0$. Far from the black hole, $\rho \sim |r|$ and we get

$$\boldsymbol{\omega}_{\rm rot} \underset{r \to \pm\infty}{\sim} -\frac{3\omega}{2} \frac{r}{|r|} \sin\theta \, \boldsymbol{e}_{(\theta)} = -\frac{3J}{|r|^3} \sin\theta \, \boldsymbol{e}_{(\theta)}. \tag{10.125}$$

In particular, in the equatorial plane ($\theta = \pi/2$), $\boldsymbol{\omega}_{\rm rot} \sim 3J/|r|^3 \, \boldsymbol{e}_z$, where $\boldsymbol{e}_z = -\boldsymbol{e}_{(\theta)}$ is parallel the symmetry axis, with the same direction as the black hole spin.

Remark 3: As a family of observers, the ZAMOs form a zero-vorticity congruence. The ***vorticity 2-form*** of any congruence of timelike worldlines of 4-velocity $\boldsymbol{u}$ is defined as the "mag-

[8] See the notebook D.4.5 for the computation; see also Eqs. (73)-(74) of Ref. [230].

netic" part $\boldsymbol{\Omega}$ in the "electric/magnetic" decomposition[9] of the 2-form $\mathbf{d}\underline{\boldsymbol{u}}$ with respect to $\boldsymbol{u}$ (cf. Sec. A.4.3):

$$\mathbf{d}\underline{\boldsymbol{u}} = \underline{\boldsymbol{a}} \wedge \underline{\boldsymbol{u}} + \boldsymbol{\Omega} \quad \text{with} \quad \boldsymbol{\Omega}(\boldsymbol{u}, \cdot) = 0, \tag{10.126}$$

where $\underline{\boldsymbol{a}}$ is the 1-form associated by metric duality to the 4-acceleration $\boldsymbol{a} := \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u}$ of the worldlines. That $\boldsymbol{\Omega} = 0$ for the ZAMO congruence follows from its orthogonality to a family of hypersurfaces, namely $(\Sigma_t)_{t\in\mathbb{R}}$. Indeed, we deduce from (10.112) that $\mathbf{d}\underline{\boldsymbol{n}} = -\mathbf{d}(N\mathbf{d}t) = -\mathbf{d}N \wedge \mathbf{d}t$, since $\mathbf{d}\mathbf{d}t = 0$ [cf. Eq. (A.95)]. Using (10.112) again, we get $\mathbf{d}\underline{\boldsymbol{n}} = \mathbf{d}\ln N \wedge \underline{\boldsymbol{n}}$. Comparing with (10.126) with $\boldsymbol{u} = \boldsymbol{n}$, we conclude that $\boldsymbol{\Omega} = 0$ for the ZAMO congruence[10]. Physically, this means that if each ZAMO sets up a orthonormal spatial frame $(\boldsymbol{e}'_i)$ with axes aligned along the spin of free gyroscopes, instead of the frame $(\boldsymbol{e}_{(i)})$ defined by (10.115), then this frame will not rotate with respect to the frame defined similarly by a nearby ZAMO.

Historical note: The concept of ZAMO has been introduced by James M. Bardeen in 1970 [16] under the name of ***locally nonrotating observers*** and for generic stationary and axisymmetric spacetimes. The application to the exterior part of Kerr spacetime has been performed by Bardeen, William H. Press and Saul A. Teukolsky in 1972 [19]. The name *ZAMO* has been coined by Kip S. Thorne and Douglas MacDonald in 1982 [247].

10.7.4 Carter observers

A ***Carter observer*** is a stationary observer $\mathscr{O}$ defined in the region $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$ of Kerr spacetime, whose worldline $\mathscr{L}$ has the following equation in Boyer-Lindquist coordinates:

$$t(\tau) = \frac{r^2+a^2}{\rho\sqrt{\Delta}}\tau + \text{const}, \quad r(\tau) = \text{const}, \quad \theta(\tau) = \text{const}, \quad \varphi(\tau) = \frac{a}{\rho\sqrt{\Delta}}\tau + \text{const}, \tag{10.127}$$

where τ is $\mathscr{O}$'s proper time, and which is equipped with the following orthonormal frame:

$$\boldsymbol{\varepsilon}_{(0)} = \frac{r^2+a^2}{\rho\sqrt{\Delta}}\,\boldsymbol{\partial}_t + \frac{a}{\rho\sqrt{\Delta}}\,\boldsymbol{\partial}_\varphi \tag{10.128a}$$

$$\boldsymbol{\varepsilon}_{(r)} = \frac{\sqrt{\Delta}}{\rho}\,\boldsymbol{\partial}_r \tag{10.128b}$$

$$\boldsymbol{\varepsilon}_{(\theta)} = \frac{1}{\rho}\,\boldsymbol{\partial}_\theta \tag{10.128c}$$

$$\boldsymbol{\varepsilon}_{(\varphi)} = \frac{a}{\rho}\sin\theta\,\boldsymbol{\partial}_t + \frac{1}{\rho\sin\theta}\,\boldsymbol{\partial}_\varphi. \tag{10.128d}$$

We note that the frame $(\boldsymbol{\varepsilon}_{(\alpha)})$ is well-defined because $\Delta > 0$ in $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$. It is an easy exercise to check that $(\boldsymbol{\varepsilon}_{(\alpha)})$ is an orthonormal frame (see the notebook D.4.6). In

[9] This name stems from the decomposition of the electromagnetic field 2-form $\boldsymbol{F}$ with respect to an observer of 4-velocity $\boldsymbol{u}$ into the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{B}$, both measured by that observer, according to $\boldsymbol{F} = \underline{\boldsymbol{u}} \wedge \underline{\boldsymbol{E}} + \star(\underline{\boldsymbol{u}} \wedge \underline{\boldsymbol{B}})$, where $\star$ stands for the Hodge dual, cf. Eq. (10.88).

[10] Incidentally, we also get $\underline{\boldsymbol{a}} = \mathbf{d}\ln N + \alpha\,\underline{\boldsymbol{n}}$ and Eq. (10.121) shows that α is actually zero.

particular, we have $\boldsymbol{\varepsilon}_{(0)} \cdot \boldsymbol{\varepsilon}_{(0)} = -1$ and, using the Boyer-Lindquist components (10.50) of $\underline{\boldsymbol{k}}$:

$$\boldsymbol{k} \cdot \boldsymbol{\varepsilon}_{(0)} = k_\mu \varepsilon^\mu_{(0)} = -\frac{r^2+a^2}{\rho\sqrt{\Delta}} + a\sin^2\theta \frac{a}{\rho\sqrt{\Delta}} = -\frac{\rho}{\sqrt{\Delta}} < 0,$$

which, by virtue of Lemma 2 in Sec. 1.2.2, shows that $\boldsymbol{\varepsilon}_{(0)}$ is future-directed. Moreover, we notice from (10.127) and (10.128a) that $\varepsilon^\alpha_{(0)} = \mathrm{d}x^\alpha/\mathrm{d}\tau$, which proves that the worldline $\mathscr{L}$ is timelike and $\boldsymbol{\varepsilon}_{(0)}$ is the 4-velocity of observer $\mathscr{O}$. Since r and θ are constant along $\mathscr{L}$, we note also that expression (10.128a) for $\boldsymbol{\varepsilon}_{(0)}$ agrees with the general form (10.105) of a stationary observer's 4-velocity.

The algebraic dual of the Carter frame, called the ***Carter coframe***, is the 4-tuple of 1-forms $(\boldsymbol{\varepsilon}^{(\alpha)})_{0\le\alpha\le3}$ such that $\langle \boldsymbol{\varepsilon}^{(\alpha)}, \boldsymbol{\varepsilon}_{(\beta)} \rangle = \delta^\alpha{}_\beta$ (cf. Sec. A.2.4 in Appendix A). It is related to the Boyer-Lindquist coordinate coframe $(\mathbf{d}x^\alpha)$ by[11]

$$\boldsymbol{\varepsilon}^{(0)} = \frac{\sqrt{\Delta}}{\rho}\left(\mathbf{d}t - a\sin^2\theta\,\mathbf{d}\varphi\right) \tag{10.129a}$$

$$\boldsymbol{\varepsilon}^{(r)} = \frac{\rho}{\sqrt{\Delta}}\mathbf{d}r \tag{10.129b}$$

$$\boldsymbol{\varepsilon}^{(\theta)} = \rho\,\mathbf{d}\theta \tag{10.129c}$$

$$\boldsymbol{\varepsilon}^{(\varphi)} = \frac{\sin\theta}{\rho}\left(-a\mathbf{d}t + (r^2+a^2)\mathbf{d}\varphi\right). \tag{10.129d}$$

Remark 4: By definition of an orthonormal frame, the spacetime metric can be written

$$\boldsymbol{g} = -\boldsymbol{\varepsilon}^{(0)}\otimes\boldsymbol{\varepsilon}^{(0)} + \boldsymbol{\varepsilon}^{(r)}\otimes\boldsymbol{\varepsilon}^{(r)} + \boldsymbol{\varepsilon}^{(\theta)}\otimes\boldsymbol{\varepsilon}^{(\theta)} + \boldsymbol{\varepsilon}^{(\varphi)}\otimes\boldsymbol{\varepsilon}^{(\varphi)}. \tag{10.130}$$

Substituting Eqs. (10.129) for the 1-forms $\boldsymbol{\varepsilon}^{(\alpha)}$, we get the following expression of the metric line element in Boyer-Lindquist coordinates:

$$g_{\mu\nu}\,\mathrm{d}x^\mu\mathrm{d}x^\nu = -\frac{\Delta}{\rho^2}\left(\mathrm{d}t - a\sin^2\theta\,\mathrm{d}\varphi\right)^2 + \frac{\rho^2}{\Delta}\mathrm{d}r^2 + \rho^2\mathrm{d}\theta^2 + \frac{\sin^2\theta}{\rho^2}\left((r^2+a^2)\mathrm{d}\varphi - a\mathrm{d}t\right)^2. \tag{10.131}$$

This expression, which is equivalent to (10.8), is often found in textbooks (cf. e.g. Eq. (33.2) in Ref. [190]). It is called the *canonical form* by Carter [45].

The specificity of the Carter observer is to be linked to the principal null geodesic congruences of Kerr spacetime, which have been introduced in Sec. 10.4. Indeed, by adding and subtracting the first two vectors of the Carter frame, as given by Eqs. (10.128a) and (10.128b) and comparing the result with Eqs. (10.49) and (10.62), we can express the tangent vectors $\boldsymbol{k}$ and $\boldsymbol{\ell}$ of respectively the ingoing and outgoing principal null geodesics as

$$\boxed{\boldsymbol{k} = \frac{\rho}{\sqrt{\Delta}}\left(\boldsymbol{\varepsilon}_{(0)} - \boldsymbol{\varepsilon}_{(r)}\right)} \quad \text{and} \quad \boxed{\boldsymbol{\ell} = \frac{\rho\sqrt{\Delta}}{2(r^2+a^2)}\left(\boldsymbol{\varepsilon}_{(0)} + \boldsymbol{\varepsilon}_{(r)}\right)}. \tag{10.132}$$

In other words, the principal null vectors $\boldsymbol{k}$ and $\boldsymbol{\ell}$ lie in $\mathrm{Span}(\boldsymbol{\varepsilon}_{(0)}, \boldsymbol{\varepsilon}_{(r)})$. Physically, this means that the Carter observer see the principal null geodesics having a pure radial motion

[11] cf. the notebook D.4.6.

(no component along $\boldsymbol{\varepsilon}_{(\theta)}$ or $\boldsymbol{\varepsilon}_{(\varphi)}$). Since the principal null geodesics are related to the Weyl tensor and to the algebraically special character of the Kerr metric (cf. Sec. 10.4), this makes the Carter frame well tailored to the computation of the Riemann's curvature tensor via Cartan's formula (see Refs. [45, 200]).

By comparing expression (10.128a) for a Carter observer's 4-velocity $\boldsymbol{\varepsilon}_{(0)}$ with Eq. (10.106), we get immediately the angular velocity around the symmetry axis of the Carter observer as measured by an asymptotically distant static observer:

$$\boxed{\omega_{\rm C} := \left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{\mathscr{L}} = \frac{a}{r^2+a^2}}. \tag{10.133}$$

Remark 5: Expression (10.133) is much simpler than that of the angular velocity ω of a ZAMO, as given by Eq. (10.119). In particular, $\omega_{\rm C}$ does not depend on θ, contrary to ω.

The 4-acceleration $\boldsymbol{a}_{\rm C} := \boldsymbol{\nabla}_{\boldsymbol{\varepsilon}_{(0)}} \boldsymbol{\varepsilon}_{(0)}$ of the Carter observer is[12]

$$\boxed{\boldsymbol{a}_{\rm C} = \frac{m(r^2-a^2)+a^2(m-r)\sin^2\theta}{\rho^3\sqrt{\Delta}}\,\boldsymbol{\varepsilon}_{(r)} - \frac{a^2}{\rho^3}\sin\theta\cos\theta\,\boldsymbol{\varepsilon}_{(\theta)}}. \tag{10.134}$$

Far from the black hole, we have $\sqrt{\Delta} \sim |r|$, $\rho \sim |r|$ so that $\rho^3\sqrt{\Delta} \sim r^4$ and we get

$$\boldsymbol{a}_{\rm C} \underset{r\to\pm\infty}{\sim} \frac{m}{r^2}\,\boldsymbol{\varepsilon}_{(r)}, \tag{10.135}$$

i.e. the same Newtonian behavior as for the ZAMO 4-acceleration [Eq. (10.122)], with the attractive character in $\mathscr{M}_{\rm I}$ and the repelling one in $\mathscr{M}_{\rm III}$.

The 4-rotation vector $\boldsymbol{\omega}^{\rm C}_{\rm rot}$ of the Carter observer is computed in a way similar to that of the ZAMO [Eq. (10.124)], i.e. by subtracting the Fermi-Walker part from $\boldsymbol{\nabla}_{\boldsymbol{\varepsilon}_{(0)}} \boldsymbol{\varepsilon}_{(i)}$ for $i \in \{1,2,3\}$ and writing it as $\boldsymbol{\omega}^{\rm C}_{\rm rot} \times_{\boldsymbol{\varepsilon}_{(0)}} \boldsymbol{\varepsilon}_{(i)}$. We get[13]

$$\boxed{\boldsymbol{\omega}^{\rm C}_{\rm rot} = \frac{a}{\rho^3}\left[\sqrt{\Delta}\cos\theta\,\boldsymbol{\varepsilon}_{(r)} - r\sin\theta\,\boldsymbol{\varepsilon}_{(\theta)}\right]}. \tag{10.136}$$

Due to $\sqrt{\Delta} \sim |r|$ and $\rho \sim |r|$ when $r \to \pm\infty$, the asymptotic behavior is

$$\boldsymbol{\omega}^{\rm C}_{\rm rot} \underset{r\to\pm\infty}{\sim} \frac{a}{r^2}\left[\cos\theta\,\boldsymbol{\varepsilon}_{(r)} - (\mathrm{sgn}\, r)\sin\theta\,\boldsymbol{\varepsilon}_{(\theta)}\right], \tag{10.137}$$

with $\mathrm{sgn}\, r = +1$ in $\mathscr{M}_{\rm I}$ and $\mathrm{sgn}\, r = -1$ in the asymptotic region of $\mathscr{M}_{\rm III}$. Introducing in both the region $\mathscr{M}_{\rm I}$ and in the asymptotic region of $\mathscr{M}_{\rm III}$ a "Cartesian" coordinate system (x,y,z) defined $x = |r|\sin\theta\cos\varphi$, $y = |r|\sin\theta\sin\varphi$ and $z = |r|\cos\theta$, the above formula turns out to be equivalent to

$$\boldsymbol{\omega}^{\rm C}_{\rm rot} \underset{r\to\pm\infty}{\sim} (\mathrm{sgn}\, r)\frac{a}{r^2}\,\boldsymbol{\partial}_z. \tag{10.138}$$

[12] See the notebook D.4.6 for the computation; see as well Eqs. (90)-(91) of Ref. [230].
[13] See the notebook D.4.6 for the computation; see as well Eq. (93) of Ref. [230].

We have thus $\boldsymbol{\omega}_{\rm rot}^{\rm C}$ always orthogonal to the equatorial plane in the asymptotic regions.

Remark 6: This is not the case of the ZAMO's 4-rotation vector $\boldsymbol{\omega}_{\rm rot}$: Eq. (10.125) shows that $\boldsymbol{\omega}_{\rm rot}$ is collinear to $\boldsymbol{\partial}_z$ only at $\theta = \pi/2$. Note also that $\boldsymbol{\omega}_{\rm rot}$ decays faster than $\boldsymbol{\omega}_{\rm rot}^{\rm C}$: $|r|^{-3}$ versus r^{-2}. Besides note that the full expression of $\boldsymbol{\omega}_{\rm rot}^{\rm C}$, Eq. (10.136), is much simpler than that of $\boldsymbol{\omega}_{\rm rot}$, Eq. (10.124). The same remark holds for the 4-acceleration $\boldsymbol{a}_{\rm C}$ [Eq. (10.134)] versus the ZAMO 4-acceleration $\boldsymbol{a}$ [Eq. (10.121)]. This reflects the fact that the Carter observer is tightly bound to the spacetime structure, being closely related to the principal null geodesics, as shown by Eq. (10.132).

Historical note: The Carter frame has been introduced by Brandon Carter in 1968 [42], in the form of its dual, namely the Carter coframe, for a quite general class of spacetimes including the Kerr one (Eqs. (81)-(82) in Ref. [42]). The explicit form for Kerr spacetime, i.e. the system (10.129), can be found in Carter's lecture at the famous Les Houches School (1972) [45] (Eqs. (5.19)-(5.24) and (5.31)-(5.34); cf. footnote 41 in the reprinted version [45]). The Carter frame has been popularized by Roman Znajek in 1977 [266] for studying electromagnetic processes around a black hole and has been heavily used in Barrett O'Neill's monograph about Kerr black holes [200], where it is called the *Boyer-Lindquist frame field*. In particular O'Neill's *canonical Kerr vector fields* V and W are $V = \rho\sqrt{\Delta}\,\boldsymbol{\varepsilon}_{(0)}$ and $W = \rho\sin\theta\,\boldsymbol{\varepsilon}_{(\varphi)}$.

10.7.5 Asymptotic inertial observers

For $r \to \pm\infty$, we see on Eqs. (10.122), (10.125), (10.135) and (10.137) that the 4-acceleration and 4-rotation of both the ZAMOs and the Carter observers tend to zero. Hence, these observers become inertial observers. Actually, at the same value of (r, θ, φ), the ZAMO and Carter observer reduce to the same inertial observer. The latter has $\omega = 0$ (take the limit $r \to \pm\infty$ in Eqs. (10.120) and (10.133)). It is therefore a static observer (cf. Sec. 10.7.2). We shall call it the ***asymptotic inertial observer***. His 4-velocity is nothing but the Killing vector $\boldsymbol{\xi}$.

10.8 Maximal analytic extension

10.8.1 Carter-Penrose diagrams

It is useful to have a "compactified view" of the whole Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$, as we had for Schwarzschild spacetime via the Carter-Penrose diagrams discussed in Sec. 9.4. However, because the Kerr spacetime with $a \neq 0$ is not spherically symmetric, such a 2-dimensional diagram can only offer a truncated view. With this in mind, we shall call a ***Carter-Penrose diagram***[14] of Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ the image $\Pi(\mathscr{M})$ of a differentiable map $\Pi : \mathscr{M} \to \mathbb{R}^2$, $(\tilde{t}, r, \theta, \tilde{\varphi}) \mapsto (T, X)$ such that

1. $T = T(\tilde{t}, r)$, $X = X(\tilde{t}, r)$;
2. $\Pi(\mathscr{M})$ is a bounded region of $\mathbb{R}^2$;

[14] See the historical notes on p. 238 and p. 305.

3. the ingoing principal null geodesics are mapped to lines $T = -X + \text{const}$ and the outgoing ones are mapped to lines $T = X + \text{const}$, with T increasing towards the future of $(\mathscr{M}, \boldsymbol{g})$ in both cases.

Property 1 implies that Π acts as a *projection*, leaving out any information in the non-ignorable coordinate θ. Property 2 provides the "compactified" view[15], and Property 3 makes $\Pi(\mathscr{M})$ look "conformal" to the 2-dimensional Minkowski spacetime generated by the metric $\boldsymbol{f} = -\mathbf{d}T \otimes \mathbf{d}T + \mathbf{d}X \otimes \mathbf{d}X$ on $\mathbb{R}^2$. However, let us stress that Π is not a priori connected to a proper conformal completion of $(\mathscr{M}, \boldsymbol{g})$, as defined in Sec. 4.3. Note that to fulfill Property 3, it suffices to choose the functions $T(\tilde{t}, r)$ and $X(\tilde{t}, r)$ defining Π such that in each of the regions $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$,

$$T(\tilde{t}, r) = U(u) + V(v) \qquad \text{and} \qquad X(\tilde{t}, r) = V(v) - U(u), \tag{10.139}$$

where (i) $U(u)$ is a function of the retarded Kerr time $u = u(\tilde{t}, r)$ given by Eq. (10.53a) that is monotonic increasing in $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$ and decreasing in $\mathscr{M}_{\rm II}$ (cf. Fig. 10.8) and (ii) $V(v)$ is a monotonic increasing function of the advanced Kerr time $v = \tilde{t} + r$ [Eq. (10.32)].

This follows from u (resp. v) being constant along the outgoing (resp. ingoing) principal null geodesics. Note also that Property 3 ensures that the two horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, which are generated by outgoing principal null geodesics, appear as lines inclined by $+45°$ with respect to the X-axis, as for the black hole event horizon in the Carter-Penrose diagrams of Schwarzschild spacetime (Figs. 9.10 and 9.11).

Remark 1: A related but distinct concept of *projection diagram* has been introduced in Ref. [63] (see also Chap. 7 of Ref. [57]); it differs from that considered here by demanding, instead of Properties 1 to 3, that Π maps any timelike curve of[16] $(\mathscr{M} \setminus \mathscr{T}, \boldsymbol{g})$ to a timelike curve of the Minkowski spacetime $(\mathbb{R}^2, \boldsymbol{f})$ and that any timelike curve in $\Pi(\mathscr{M} \setminus \mathscr{T})$ is the image of a timelike curve of $(\mathscr{M} \setminus \mathscr{T}, \boldsymbol{g})$.

A schematic Carter-Penrose diagram of the Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ is shown in Fig. 10.10. By *schematic* it is meant that we do not provide any explicit construction via a map Π. Let us mention however that defining Π by choosing $U(u) = \arctan(u/m) - k\pi$ and $V(v) = \arctan(v/m)$ in Eq. (10.139), with $k = 0$, 1 and 2 on respectively $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$, would lead to such a diagram. By virtue of Property 1, one may think of each point in the diagram of Fig. 10.10 as being a 2-sphere, spanned by $(\theta, \tilde{\varphi})$, except along the curve $r = 0$ (thick dotted line), where each point is the union $\mathcal{S}_{0,t}$ of two flat open disks (cf. Sec. 10.2.2, Eq. (10.14)).

Each of the regions $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$ of the Kerr spacetime $\mathscr{M}$ is mapped to the interior of a square tilted by 45° (a "diamond") in the diagram of Fig. 10.10. In each of these diamond blocks, v increases from $-\infty$ to $+\infty$ in the South-West to North-East direction, while u increases from $-\infty$ to $+\infty$ in the South-East to North-West direction in $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$ and in the opposite direction in $\mathscr{M}_{\rm II}$. These ranges and directions follow directly from $v = \tilde{t} + r$ and the graph of $u - \tilde{t}$ shown in Fig. 10.8.

[15] Property 2 refers to a "bounded region" instead of "compact region", because $\Pi(\mathscr{M})$ is in general an open subset of $\mathbb{R}^2$ and therefore not a compact subset.

[16] $\mathscr{T}$ is the Carter time machine discussed in Sec. 10.2.5.

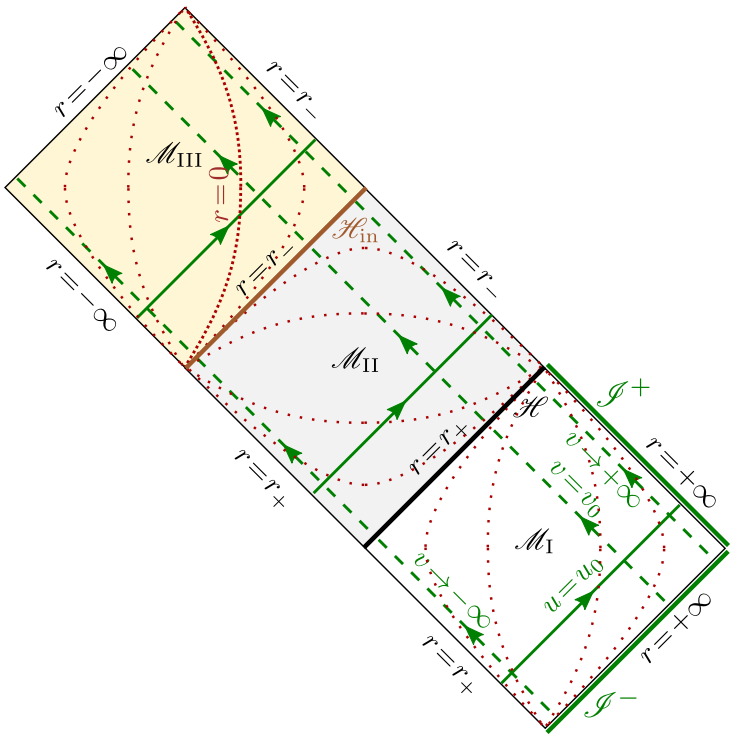


Figure 10.10: Carter-Penrose diagram of the Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$, with $\mathscr{M} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathscr{R}$ (cf. Eq. (10.28) in Sec. 10.3.1), spanned by the ingoing principal null geodesics (dashed green lines). The solid green lines are outgoing principal null geodesics, while the dotted curves mark some hypersurfaces $r = \text{const}$.

The dotted curves in Fig. 10.10 represent some hypersurfaces $r = \text{const}$. According to the results of Sec. 10.5.1, such hypersurfaces are timelike in $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$, spacelike in $\mathscr{M}_{\rm II}$ and null for $r = r_-$ or $r = r_+$. Since the Killing vector field $\boldsymbol{\xi}$ is tangent to the hypersurfaces $r = \text{const}$, the dotted curves in Fig. 10.10 can also be seen as the projections of the field lines of $\boldsymbol{\xi}$, i.e. of the orbits of the stationary group action.

As pointed out in Sec. 10.5.2, the asymptotic structure of region $\mathscr{M}_{\rm I}$ is identical to that of $\mathscr{M}_{\rm I}$ in Schwarzschild spacetime. It can thus be endowed with a conformal completion at null infinity, as constructed in Secs. 9.4.3 and 9.4.4. This allows us to add the future (resp. past) null infinity $\mathscr{I}^+$ (resp. $\mathscr{I}^-$) to the diagram of Fig. 10.10. Note that $\mathscr{I}^+$ corresponds to the limit $v \to +\infty$ in $\mathscr{M}_{\rm I}$, while $\mathscr{I}^-$ corresponds to the limit $u \to -\infty$ in $\mathscr{M}_{\rm I}$.

10.8.2 Constructing the maximal extension

The ingoing principal null geodesics $\mathscr{L}^{\rm in}_{(v,\theta,\tilde\varphi)}$ (dashed green lines in Fig. 10.10) are complete (cf. Sec. B.3.2), except for those lying in the equatorial plane, which hit the curvature singularity at $r = 0$. Indeed, the affine parameter $\lambda = -r$ of any ingoing principal null geodesic with $\theta \neq \pi/2$ ranges from $-\infty$ (lower right of the diagram) to $+\infty$ (upper left). On the contrary, the outgoing principal null geodesics $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$ (solid green lines in

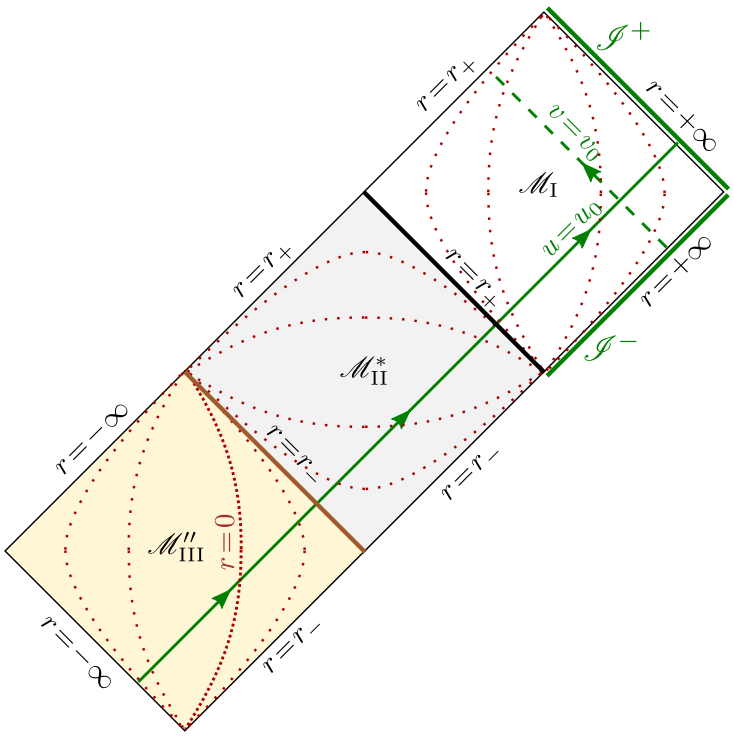


Figure 10.11: Carter-Penrose diagram of the minimal extension of $\mathscr{M}_{\rm I}$ to ensure complete outgoing principal null geodesics (one of them is drawn as a solid green line).

Fig. 10.10) are not complete: in $\mathscr{M}_{\rm I}$, their affine parameter $\lambda = r$ is bounded from below by r_+; in $\mathscr{M}_{\rm II}$, their affine parameter $\lambda = -r$ ranges in $(-r_+, -r_-)$ only and in $\mathscr{M}_{\rm III}$, their affine parameter $\lambda = r$ is bounded from above by r_-. Since these geodesics are not ending at any spacetime singularity (it can be shown that all curvature scalar invariants remain bounded along them), except those with $\theta = \pi/2$ in $\mathscr{M}_{\rm III}$, this indicates that the spacetime $(\mathscr{M}, \boldsymbol{g})$ can be extended. Moreover, for $0 \leq a < m$, the event horizon $\mathscr{H}$ is a *non-degenerate* Killing horizon (cf. Sec. 10.5.4) and we have seen in Sec. 5.4.2 that, generically, such horizons are part of a *bifurcate Killing horizon* in an extended spacetime.

A spacetime extending $(\mathscr{M}_{\rm I}, \boldsymbol{g})$ "to the past", so that the outgoing principal null geodesics are complete, is shown in Fig. 10.11. It is made by attaching to $(\mathscr{M}_{\rm I}, \boldsymbol{g})$ a time-reversed copy[17] of $(\mathscr{M}_{\rm II}, \boldsymbol{g})$, $(\mathscr{M}^*_{\rm II}, \boldsymbol{g})$ say, and then attaching to the latter a copy of $(\mathscr{M}_{\rm III}, \boldsymbol{g})$, $(\mathscr{M}''_{\rm III}, \boldsymbol{g})$ say[18]. By construction, the outgoing principal null geodesics are complete in this extension, but the ingoing ones are not: the affine parameter $\lambda = -r$ of that denoted by $v = v_0$ in Fig. 10.11 is bounded from above by $-r_+$. Actually the situation is completely symmetric to that of the original Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ (Fig. 10.10). In particular, the region $\mathscr{M}^*_{\rm II} \cup \mathscr{H}''_{\rm in} \cup \mathscr{M}''_{\rm III}$, where $\mathscr{H}''_{\rm in}$ is the (null) hypersurface $r = r_-$, is a *white hole*, since it is the complement of $J^+(\mathscr{I}^-)$ (cf. the definition (4.39) of a white hole). The hypersurface $r = r_+$ (black thick line in Fig. 10.11) is the corresponding *past*

[17] By *copy*, it is meant that $(\mathscr{M}^*_{\rm II}, \boldsymbol{g})$ is a spacetime isometric to $(\mathscr{M}_{\rm II}, \boldsymbol{g})$ and by *time-reversed* that r is increasing towards the future in $\mathscr{M}^*_{\rm II}$, while it is decreasing towards the future in $\mathscr{M}_{\rm II}$.

[18] The notation $\mathscr{M}''_{\rm III}$ instead of $\mathscr{M}'_{\rm III}$ is for later convenience.

event horizon.

By combining the diagrams of Figs. 10.10 and 10.11, one obtains a spacetime which still contains incomplete null geodesics: the outgoing ones in regions $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$, and the ingoing ones in regions $\mathscr{M}_{\rm II}^*$ and $\mathscr{M}_{\rm III}''$. To go further, one should add new regions isometric to one of the three blocks $(\mathscr{M}_{\rm I}, \boldsymbol{g})$, $(\mathscr{M}_{\rm II}, \boldsymbol{g})$ and $(\mathscr{M}_{\rm III}, \boldsymbol{g})$ and iterate indefinitely, leading to the Carter-Penrose diagram of Fig. 10.12. In this process, one shall make sure to have some analytic continuation of the metric between the various blocks. This is done by introducing Kruskal-type coordinates in the vicinity of the boundaries between the various blocks. We shall not do it here and refer the reader to the seminal articles by Boyer & Lindquist [30] and Carter [41], the famous Les Houches lectures by Carter [45] or the textbook by O'Neill [200].

In the diagram of Fig. 10.12, it is clear that the event horizon $\mathscr{H}$ and the inner horizon $\mathscr{H}_{\rm in}$ have been extended to bifurcate Killing horizons (cf. Sec. 5.4), the bifurcation surface of which being a 2-sphere depicted by a circular dot.

Historical note: In 1966, Brandon Carter [39] obtained the maximal analytic extension of the 2-dimensional manifold constituted by the rotation axis[19] $\mathscr{A}$ of the Kerr spacetime and drew a diagram similar to that of Fig. 10.12 (Fig. 1a of Ref. [39]). More precisely, Carter introduced on $\mathscr{A}$ a coordinate system (T, X) (denoted (ψ, ξ) by him) in which the metric induced by $\boldsymbol{g}$ on $\mathscr{A}$ is explicitely conformal the 2-dimensional Minkowski metric $\boldsymbol{f} = -\mathbf{d}T \otimes \mathbf{d}T + \mathbf{d}X \otimes \mathbf{d}X$. Thus Fig. 1a of Carter's article [39], which is reproduced as Fig. 28 of Hawking & Ellis' textbook [146], is a true conformal representation of $\mathscr{A}$, and not the mere "compactified projection" that we used to define a generic Carter-Penrose diagrams in Sec. 10.8.1. This was of course made possible because the rotation axis $\mathscr{A}$ is a 2-dimensional manifold. For the whole Kerr spacetime, the "projection" aspect is inevitable. In the same article [39], Carter suggested that the maximal analytic extension of the whole 4-dimensional manifold would be similar to that of $\mathscr{A}$. This was proven rigorously a year later by Robert H. Boyer and Richard W. Lindquist [30] and generalized to Kerr-Newmann spacetimes by Carter himself in 1968 [41].

10.8.3 Cauchy horizon

Let us consider a spacelike hypersurface Σ running from the asymptotically flat end of $\mathscr{M}_{\rm I}$ to the asymptotically flat end of $\mathscr{M}'_{\rm I}$, possibly through[20] $\mathscr{M}_{\rm II}$ or $\mathscr{M}_{\rm II}^*$ (cf. Fig. 10.13) that is ***acausal***, in the sense that no causal curve intersects it more than once. One says that Σ is a ***partial Cauchy surface***, the definition of the latter being an acausal hypersurface without edge [146]. The ***future Cauchy development*** (resp. ***past Cauchy development***) of Σ is the set $D^+(\Sigma)$ (resp. $D^-(\Sigma)$) of all spacetime points p such that each past-directed (resp. future directed) inextendible causal curve through p intersects Σ. The future Cauchy development of Σ is the hatched region in Fig. 10.13. The spacetime metric at every point in $D^+(\Sigma)$ is entirely determined by initial data on Σ through

[19] Let us recall that the *rotation axis* is the 2-dimensional Lorentzian submanifold $\mathscr{A}$ where the Killing vector $\boldsymbol{\eta}$ vanishes (cf. the definition given in Sec. 5.3.2); in terms of the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$, $\mathscr{A}$ is defined by by $\theta = 0$ or π and is naturally spanned by the coordinates $(\tilde{t}, r)$.

[20] In Fig. 10.13, this is not the case: Σ goes from $\mathscr{M}_{\rm I}$ to $\mathscr{M}'_{\rm I}$ via the bifurcation sphere.

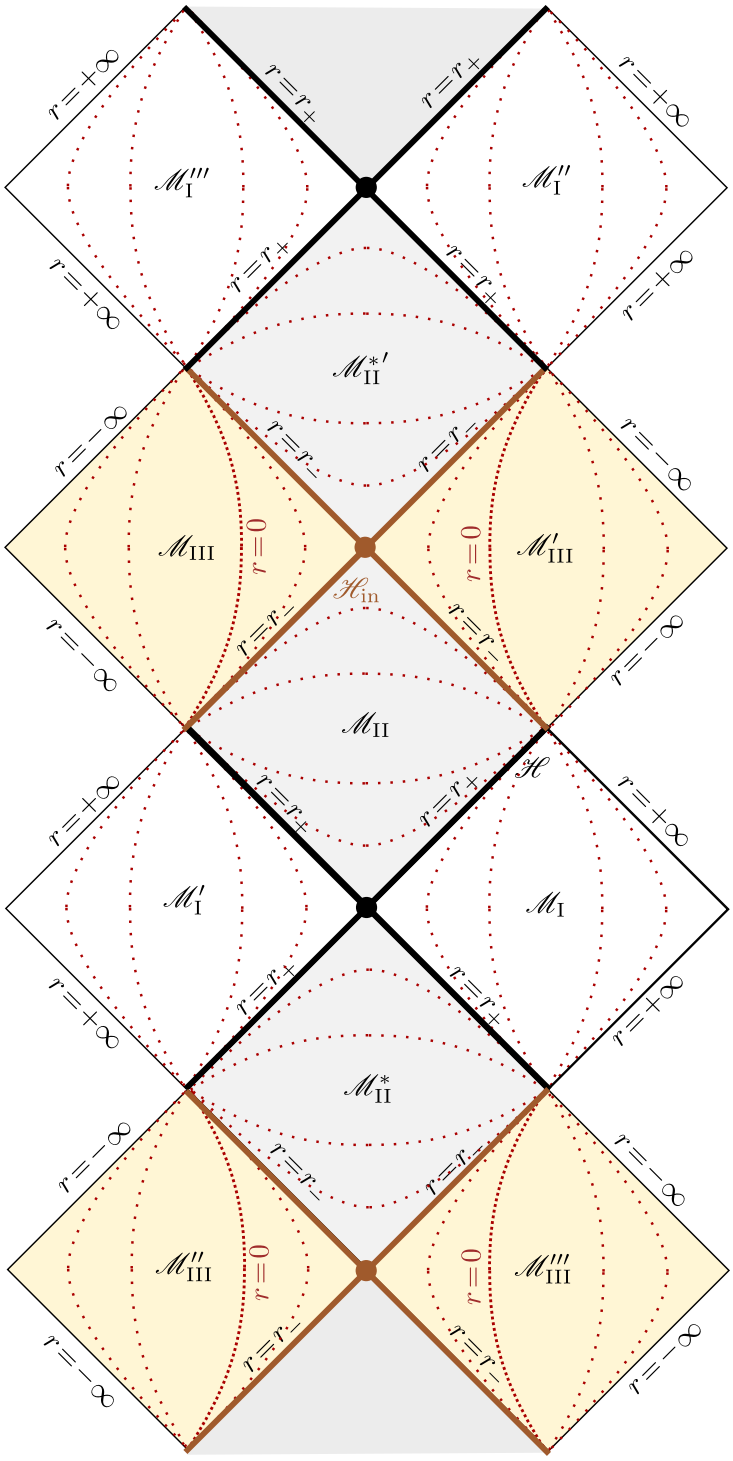


Figure 10.12: Carter-Penrose diagram of the maximal analytic extension of the Kerr spacetime. As in Figs. 10.10 and 10.11, the dotted curves mark some hypersurfaces $r = \text{const}$. The central black or light brown dots mark the bifurcation spheres of bifurcate Killing horizons.

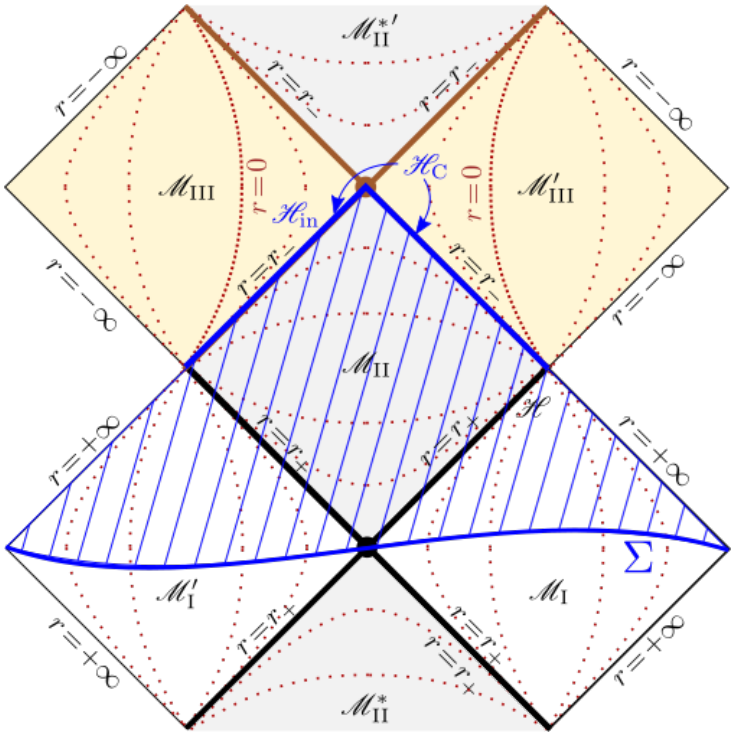


Figure 10.13: The partial Cauchy surface Σ, its future Cauchy development $D^+(\Sigma)$ (hatched) and the Cauchy horizon $\mathscr{H}_{\rm C}$. As in Figs. 10.10-10.12, the dotted curves marks some hypersurfaces $r = \text{const}$.

the Einstein equation, in its 3+1 form (see e.g. Chap. 5 of Ref. [122]); this reflects the well-posedness of general relativity as a *Cauchy problem*.

One says that Σ is a ***Cauchy surface*** iff $D^+(\Sigma) \cup D^-(\Sigma)$ is the whole spacetime, i.e. iff every inextendible causal curve intersects Σ. It is clear on Fig. 10.13 that Σ is a Cauchy surface for $(\mathscr{M}_1, \boldsymbol{g})$, with $\mathscr{M}_1 := \mathscr{M}_{\rm I} \cup \mathscr{M}'_{\rm I} \cup \mathscr{M}_{\rm II} \cup \mathscr{M}^*_{\rm II}$, but not for the whole extended Kerr spacetime. The future boundary of $D^+(\Sigma)$ is called the ***future Cauchy horizon*** of Σ and denoted by $H^+(\Sigma)$. In the present case, the future Cauchy horizon does not depend upon the choice of Σ, being the same for any partial Cauchy surface through $\mathscr{M}_{\rm I} \cup \mathscr{M}'_{\rm I}$. We shall therefore denote it by $\mathscr{H}_{\rm C}$. It is depicted as the blue thick line in Fig. 10.13. $\mathscr{H}_{\rm C}$ is the union of what we called the *inner horizon* in Sec. 10.5.1, i.e. the Killing horizon $\mathscr{H}_{\rm in}$, and the null boundary between $\mathscr{M}_{\rm II}$ and $\mathscr{M}'_{\rm III}$, $\mathscr{H}'_{\rm in}$ say, which is a Killing horizon as well:

$$\mathscr{H}_{\rm C} = \mathscr{H}_{\rm in} \cup \mathscr{H}'_{\rm in}. \tag{10.140}$$

Note that $\mathscr{H}_{\rm C}$ corresponds to a fixed value of the r coordinate of Kerr-type coordinate systems, namely $r = r_-$, in agreement with the primary definition of $\mathscr{H}_{\rm in}$ given in Eq. (10.27).

Remark 2: There is no such Cauchy horizon in the Schwarzschild spacetime, even in its maximally extended version. For instance the hypersurface defined in terms of the Kruskal-Szekeres coordinates by $T = 0$ and whose equatorial section is depicted in Fig. 9.14 (Flamm paraboloid), is a Cauchy surface for the whole maximally extended Schwarzschild spacetime (by looking at Fig. 9.12, you may convince yourself that any inextendible causal curve in Schwarzschild spacetime must go through $T = 0$).

Let us evaluate the non-affinity coefficient $\kappa_{\rm in}$ of the Killing generator $\boldsymbol{\chi}_{\rm in}$ of the part $\mathscr{H}_{\rm in}$ of the Cauchy horizon [cf. Eq. (10.75)]. Since $mr_-/(r_-^2+a^2) = 1/2$, we notice, from Eqs. (10.57), (10.75) and (10.76), that the Killing vector $\boldsymbol{\chi}_{\rm in}$ coincides with the vector $\boldsymbol{\ell}$ tangent to the outgoing principal null geodesics on $\mathscr{H}_{\rm in}$:

$$\boldsymbol{\chi}_{\rm in} \overset{\mathscr{H}_{\rm in}}{=} \boldsymbol{\ell}. \tag{10.141}$$

Hence, as the event horizon $\mathscr{H}$, the inner horizon $\mathscr{H}_{\rm in}$ is generated by outgoing principal null geodesics [compare Eq. (10.77) and see Figs. 10.6 and 10.7]. Equation (10.58) implies then

$$\boldsymbol{\nabla}_{\boldsymbol{\chi}_{\rm in}} \boldsymbol{\chi}_{\rm in} \overset{\mathscr{H}_{\rm in}}{=} \kappa_{\rm in}\, \boldsymbol{\chi}_{\rm in}, \tag{10.142}$$

with the non-affinity coefficient $\kappa_{\rm in}$ obtained by specializing Eq. (10.59) to $r = r_-$:

$$\kappa_{\rm in} = \kappa_{\boldsymbol{\ell}}\big|_{r=r_-} = \frac{m(r_-^2 - a^2)}{(r_-^2+a^2)^2}.$$

Using the expression of r_- in terms of m and a [Eq. (10.3)], we get

$$\boxed{\kappa_{\rm in} = -\frac{\sqrt{m^2-a^2}}{2m(m-\sqrt{m^2-a^2})}}. \tag{10.143}$$

Given the assumption $0 < a < m$, we have $\kappa_{\rm in} \neq 0$, which implies that (cf. Sec. 3.3.6)

> As long as $a < m$, the part $\mathscr{H}_{\rm in}$ of the Cauchy horizon $\mathscr{H}_{\rm C}$ is a non-degenerate Killing horizon.

We note that

$$\kappa_{\rm in} < 0. \tag{10.144}$$

According to the results of Sec. 5.4.2, it follows that

> $\mathscr{H}_{\rm in}$ is contained in a *bifurcate Killing horizon*, the bifurcation surface of which being the future boundary of $\mathscr{H}_{\rm in}$ (light brown small disk in Fig. 10.13). Actually, the whole Cauchy horizon $\mathscr{H}_{\rm C}$ is the past part of the bifurcate Killing horizon, the future part being formed by the future boundaries $r = r_-$ of $\mathscr{M}_{\rm III}$ and $\mathscr{M}'_{\rm III}$ (light brown thick lines in Fig. 10.13).

10.8.4 Physical relevance of the maximal extension

As for the maximal extension of Schwarzschild spacetime (Chap. 9), the maximal extension of the Kerr spacetime discussed above corresponds to an "eternal" black hole, not to any "astrophysical" black hole formed by gravitational collapse. Moreover, contrary to the non-rotating case, where the Schwarzschild geometry is exact outside the collapsing star (by virtue of Birkhoff's theorem), the spacetime outside a collapsing rotating star is *not* a part of Kerr spacetime. In particular, it contains gravitational waves and is not stationary.

Only at "late times", when all the "hairs" have been radiated away, does it settle to the Kerr spacetime.

Another physical issue regards the Cauchy horizon, which has been shown to be unstable: it suffers the so-called *mass inflation instability* discovered by Poisson & Israel [215] (see Ref. [33] for a recent study of this instability).

10.9 Further reading

For more material about the Kerr black hole, we refer the reader to O'Neill's very nice monograph [200], as well to the review articles by Heinicke & Hehl [149], Teukolsky [245] and Visser [255].

Chapter 11

Geodesics in Kerr spacetime: generic and timelike cases

Contents

11.1 Introduction

In various occasions during our study of Kerr spacetime in Chap. 10, we have already encountered some peculiar geodesics, namely the principal null geodesics. In this chapter, we begin the systematic study of causal (timelike or null) geodesics in Kerr spacetime. First, taking advantage of an algebraic particularity of this spacetime, giving birth to a non-trivial Killing tensor and the associated Carter constant, we shall see in Sec. 11.2 that Kerr geodesic motion is fully governed by a system of first order equations, the solution of which can be obtained by quadrature. The basic properties of geodesics that one can infer from this system are derived in Sec. 11.3. Then we focus on timelike geodesics (Sec. 11.4) and in particular on those that are bound, which represent orbits of massive particles or bodies around a Kerr black hole. In Sec. 11.5, we shall investigate the particular case of circular orbits in the equatorial plane and discuss their stability. The detailed study of null geodesics is deferred to the next chapter.

Of course, when the Kerr spin parameter a tends to zero, all results in this chapter reduce to those obtained for Schwarzschild geodesics in Chap. 7. As already suggested in the introduction of Chap. 7, before moving on, the reader might want to have a look at Appendix B, which recaps the properties of geodesics in pseudo-Riemannian manifolds.

11.2 Equations of geodesic motion

11.2.1 Introduction

In all this chapter, we are concerned with the motion of a particle $\mathscr{P}$ in the Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$, under the hypothesis that $\mathscr{P}$ feels only gravity, as described by $\boldsymbol{g}$ (freely falling particle). The worldline $\mathscr{L}$ of $\mathscr{P}$ is then necessarily a geodesic[1] of $(\mathscr{M}, \boldsymbol{g})$. It is a timelike geodesic if $\mathscr{P}$ is a massive particle and a null geodesic if $\mathscr{P}$ is massless (e.g. a photon).

In a given coordinate system (x^α), the geodesic $\mathscr{L}$ is described by a system of the form[2] $x^\alpha = x^\alpha(\lambda)$, where λ is an affine parameter[3] along $\mathscr{L}$. We choose λ to be the affine parameter associated with $\mathscr{P}$'s 4-momentum $\boldsymbol{p}$, i.e. such that $\mathrm{d}x^\alpha/\mathrm{d}\lambda = p^\alpha$ [Eq. (1.19)]. In particular, λ is dimensionless and is increasing towards the future[4] along $\mathscr{L}$. The curve $\mathscr{L}$ is a geodesic iff the functions $x^\alpha(\lambda)$ obey the geodesic equation [Eq. (B.10) in Appendix B]:

$$\frac{\mathrm{d}^2 x^\alpha}{\mathrm{d}\lambda^2} + \Gamma^\alpha{}_{\mu\nu} \frac{\mathrm{d}x^\mu}{\mathrm{d}\lambda} \frac{\mathrm{d}x^\nu}{\mathrm{d}\lambda} = 0, \qquad 0 \leq \alpha \leq 3, \tag{11.1}$$

where the $\Gamma^\alpha{}_{\mu\nu}$'s are the Christoffel symbols of the metric $\boldsymbol{g}$ with respect to the coordinates (x^α), as given by Eq. (A.71). Equation (11.1) is the component expression of $\boldsymbol{\nabla}_{\boldsymbol{p}} \boldsymbol{p} = 0$ [Eq. (1.9)]. It is a system of four coupled second-order differential equations, which are non-linear. We are going to see that it is actually not necessary to solve this system to compute the geodesics of Kerr spacetime. Indeed, as in the Schwarzschild case studied in Chap. 7, there exists enough first integrals of Eq. (11.1) to reduce the problem to four first-order equations of the type $\mathrm{d}x^\alpha/\mathrm{d}\lambda = F^\alpha(x^0, x^1, x^2, x^3)$.

Three integrals of motion are similar to those of Schwarzschild geodesics: one is the particle's mass μ (Sec. 11.2.3 below) and the two other ones are the conserved energy E and the conserved angular momentum L (Sec. 11.2.2), which arise from the common symmetries of Kerr and Schwarzschild spacetimes: stationarity (outside the ergoregion) and axisymmetry. In the Schwarzschild case, the fourth integral of motion was provided by spherical symmetry, which constrained all geodesics to be planar, so that a suitable choice of coordinates (t, r, θ, φ) makes a given geodesic confined to the hyperplane $\theta = \pi/2$, yielding the first integral $p^\theta = 0$ [Eq. (7.7)]. The Kerr spacetime with $a \neq 0$ being not spherically symmetric, we loose this property here. Fortunately there exists another integral of motion, known as the *Carter constant*; it arises from a remarkable property of Kerr spacetime: the existence of a non-trivial Killing tensor (Sec. 11.2.4). This makes a total of four integral of motions, which makes the problem integrable (Secs. 11.2.5 to 11.2.8).

[1] The definition and basic properties of geodesics are recalled in Appendix B; see also Sec. 1.3.2.

[2] In Appendices A and B, a different symbol is used for the function $X^\alpha(\lambda)$ defining $\mathscr{L}$ and the coordinate x^α (cf. Secs. A.2.3 and B.3.1). Following standard usage in the physics literature, we shall not do this in this chapter.

[3] See Sec. B.2.1 for the definition of an affine parameter along a geodesic.

[4] Let us recall that the Kerr spacetime is time-oriented, cf. Sec. 10.3.2.

11.2.2 Integrals of motion from spacetime symmetries

As for Schwarzschild spacetime (cf. Sec. 7.2.1), we have the following property.

The Killing vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ of Kerr spacetime, associated respectively with stationarity and axisymmetry [cf. Eq. (10.12)], give birth to two conserved quantities along any geodesic $\mathscr{L}$:

$$\boxed{E := -\boldsymbol{\xi}\cdot\boldsymbol{p} = -\boldsymbol{g}(\boldsymbol{\xi},\boldsymbol{p})} \tag{11.2a}$$

$$\boxed{L := \boldsymbol{\eta}\cdot\boldsymbol{p} = \boldsymbol{g}(\boldsymbol{\eta},\boldsymbol{p})}, \tag{11.2b}$$

where $\boldsymbol{p}$ is the 4-momentum of the particle $\mathscr{P}$ having $\mathscr{L}$ as worldline (cf. Sec. 1.3). For the same reasons as in Sec. 7.2.1, E is called the ***conserved energy*** or ***energy at infinity*** of $\mathscr{P}$, while L is called the ***conserved angular momentum*** or ***angular momentum at infinity*** of $\mathscr{P}$.

In particular, if $\mathscr{L}$ reaches the asymptotic region $|r| \to +\infty$, E is nothing but $\mathscr{P}$'s energy as measured by the asymptotic inertial observer introduced in Sec. 10.7.5, given that the 4-velocity of the latter coincides with the Killing vector $\boldsymbol{\xi}$ [compare Eq. (1.21) with Eq. (11.2a)]. Similarly, L is the component along the rotation axis of $\mathscr{P}$'s (total) angular momentum vector $\boldsymbol{L}_{\rm tot}$ as measured by the asymptotic inertial observer [cf. Eq. (7.4)].

Remark 1: Because it represents only a component of the total angular momentum, L is sometimes denoted by L_z.

In coordinates (t, r, θ, φ) adapted to spacetime symmetries, i.e. coordinates such that $\boldsymbol{\xi} = \boldsymbol{\partial}_t$ and $\boldsymbol{\eta} = \boldsymbol{\partial}_\varphi$, for instance Boyer-Lindquist coordinates (Sec. 10.2.1), null Kerr coordinates (Sec. 10.3.1) or Kerr coordinates (Sec. 10.3.3), one can rewrite (11.2) in terms of the components $p_t = g_{t\mu}\, p^\mu$ and $p_\varphi = g_{\varphi\mu}\, p^\mu$ of the 1-form $\underline{\boldsymbol{p}}$ associated to $\boldsymbol{p}$ by metric duality:

$$E = -p_t \tag{11.3a}$$

$$L = p_\varphi \tag{11.3b}$$

Indeed, in such a coordinate system, $\xi^\mu = \delta^\mu{}_t$ and $\eta^\mu = \delta^\mu{}_\varphi$, so that $E = -g_{\mu\nu}\,\xi^\mu p^\nu = -g_{t\nu}\, p^\nu = -p_t$ and $L = g_{\mu\nu}\,\eta^\mu p^\nu = g_{\varphi\nu}\, p^\nu = p_\varphi$.

Example 1 (Generators of the event and inner horizons): Let us choose for $\mathscr{L}$ a null geodesic generator of the event horizon $\mathscr{H}$ (cf. Sec. 10.5.3). $\mathscr{L}$ is then an outgoing principal null geodesic $\mathscr{L}^{{\rm out},\mathscr{H}}_{(\theta,\psi)}$ and the 4-momentum vector $\boldsymbol{p}$ is proportional to the null vector $\boldsymbol{\ell} \overset{\mathscr{H}}{=} \boldsymbol{\chi} = \boldsymbol{\xi} + \Omega_H \boldsymbol{\eta}$ [Eq. (10.77) and (10.73)]. By definition of a generator of a null hypersurface (cf. Sec. 2.3.3), $\boldsymbol{p}$ is normal to $\mathscr{H}$. Since the Killing vector fields $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ are tangent to $\mathscr{H}$ (for $\mathscr{H}$ is globally preserved by the spacetime symmetries), we get immediately from the definitions (11.2) $E = 0$ and $L = 0$. Similarly, if $\mathscr{L}$ is a null geodesic generator of the inner horizon $\mathscr{H}_{\rm in}$, $\mathscr{L}$ is an outgoing principal null geodesic $\mathscr{L}^{{\rm out},\mathscr{H}_{\rm in}}_{(\theta,\psi)}$, with $\boldsymbol{p}$ proportional to the null normal

$\boldsymbol{\ell} \overset{\mathscr{H}_{\rm in}}{=} \boldsymbol{\chi}_{\rm in} = \boldsymbol{\xi} + \Omega_{\rm in}\boldsymbol{\eta}$ [Eq. (10.141) and (10.75)], so that we have $E = 0$ and $L = 0$ as well. Hence we conclude:

$$\mathscr{L} \text{ null geodesic generator of } \mathscr{H} \text{ or } \mathscr{H}_{\rm in} \implies E = 0 \quad \text{and} \quad L = 0. \tag{11.4}$$

Example 2 (Ingoing principal null geodesics): For the ingoing principal null geodesics $\mathscr{L}^{\rm in}_{(v,\theta,\tilde\varphi)}$ introduced in Sec. 10.4, the 4-momentum vector is

$$\boldsymbol{p} = \alpha \boldsymbol{k}, \tag{11.5}$$

where α is a (positive) constant, since $\boldsymbol{k}$ is a geodesic vector [Eq. (10.45)], as $\boldsymbol{p}$ [Eq. (1.9)]. Equations (11.3) with the components relative to Kerr coordinates $(\tilde t, r, \theta, \tilde\varphi)$ lead then to

$$E = -\alpha k_{\tilde t} = -\alpha(-1) = \alpha \qquad \text{and} \qquad L = \alpha k_{\tilde\varphi} = \alpha a \sin^2\theta,$$

where the components $k_{\tilde t}$ and $k_{\tilde\varphi}$ have been read on Eq. (10.44). Hence for any ingoing principal null geodesic $\mathscr{L}^{\rm in}_{(v,\theta,\tilde\varphi)}$,

$$E > 0 \qquad \text{and} \qquad L = aE\sin^2\theta. \tag{11.6}$$

Recall that θ is constant along $\mathscr{L}^{\rm in}_{(v,\theta,\tilde\varphi)}$, so that the above formula does yield a constant value for L. Moreover, it fulfills $L \geq 0$ with $L = 0$ only for $a = 0$ or $\theta \in \{0, \pi\}$ (rotation axis).

Example 3 (Outgoing principal null geodesics): For the outgoing principal null geodesics $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$, $\mathscr{L}^{{\rm out},\mathscr{H}}_{(\theta,\psi)}$ and $\mathscr{L}^{{\rm out},\mathscr{H}_{\rm in}}_{(\theta,\psi)}$ introduced in Sec. 10.4, the 4-momentum vector is

$$\boldsymbol{p} = \beta(\lambda)\boldsymbol{\ell}. \tag{11.7}$$

where $\beta(\lambda)$ is a function of the affine parameter λ associated to $\boldsymbol{p}$ that obeys $\beta(\lambda) > 0$, since both $\boldsymbol{p}$ and $\boldsymbol{\ell}$ are future-directed (cf. Sec. 10.4). Contrary to the coefficient α in Eq. (11.5), $\beta(\lambda)$ is not constant because $\boldsymbol{\ell}$ is not a geodesic vector, but only a pregeodesic one: it fulfills $\boldsymbol{\nabla}_{\boldsymbol{\ell}}\boldsymbol{\ell} = \kappa_{\boldsymbol{\ell}}\boldsymbol{\ell}$ with $\kappa_{\boldsymbol{\ell}} \neq 0$ [Eqs. (10.58) and (10.59)] (cf. Remark 4 on p. 279). The geodesic equation $\boldsymbol{\nabla}_{\boldsymbol{p}}\boldsymbol{p} = 0$ implies that β obeys the differential equation $\boldsymbol{\nabla}_{\boldsymbol{p}}\beta + \kappa_{\boldsymbol{\ell}}\beta^2 = 0$, or equivalently

$$\frac{{\rm d}\beta}{{\rm d}\lambda} + \kappa_{\boldsymbol{\ell}}\beta^2 = 0. \tag{11.8}$$

The outgoing principal null geodesics either (i) are confined to one of the horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, generating them (case of $\mathscr{L}^{{\rm out},\mathscr{H}}_{(\theta,\psi)}$ and $\mathscr{L}^{{\rm out},\mathscr{H}_{\rm in}}_{(\theta,\psi)}$), or (ii) never intersect them (case of $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$, cf. the solid curves in Figs. 10.6 – 10.7). In the first case, $\kappa_{\boldsymbol{\ell}}$ is a constant on $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, given by Eq. (10.59) with $r = r_+$ for $\mathscr{H}$ and $r = r_-$ for $\mathscr{H}_{\rm in}$. The solution of Eq. (11.8) is then

$$\beta(\lambda) = \frac{1}{\kappa_{\boldsymbol{\ell}}(\lambda - \lambda_0)},$$

where λ_0 is a constant. Outside the horizons, $\kappa_{\boldsymbol{\ell}}$ is the function of r given by Eq. (10.59). We then search for a solution of Eq. (11.8) in the form $\beta(\lambda) = B(r(\lambda))$, where $r(\lambda)$ is the function giving the coordinate r along the geodesic. Since the latter obeys ${\rm d}r/{\rm d}\lambda = p^r = \beta\ell^r$ with ℓ^r read on Eq. (10.57), we get the following linear differential equation for B:

$$(r^2 - 2mr + a^2)\frac{{\rm d}B}{{\rm d}r} + 2m\frac{r^2 - a^2}{r^2 + a^2}B = 0.$$

The solution is

$$B(r) = 2\frac{r^2+a^2}{r^2-2mr+a^2}B_0 = \beta(\lambda),$$

where B_0 is a positive constant in $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$ and a negative constant in $\mathscr{M}_{\rm II}$, to ensure that $\beta(\lambda) > 0$.

Plugging (11.7) into Eqs. (11.3) with the components relative to Kerr coordinates yields

$$E = -\beta(\lambda)\ell_{\tilde{t}} = \beta(\lambda)\frac{\Delta}{2(r^2+a^2)} \qquad \text{and} \qquad L = \beta(\lambda)\ell_{\tilde{\varphi}} = \beta(\lambda)\frac{a\Delta\sin^2\theta}{2(r^2+a^2)},$$

where the components $\ell_{\tilde{t}}$ and $\ell_{\tilde{\varphi}}$ have been read on Eq. (10.61). Given the above results for $\beta(\lambda)$, we get $E = 0$ and $L = 0$ on $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, since $\Delta = 0$ there, and $E = B_0$ and $L = aB_0\sin^2\theta$ elsewhere. We conclude that for the outgoing principal null geodesics,

$$\text{in } \mathscr{H}\cup\mathscr{H}_{\rm in}, \quad E = 0 \quad \text{and} \quad L = 0 \tag{11.9a}$$

$$\text{in } \mathscr{M}_{\rm I}\cup\mathscr{M}_{\rm III}, \quad E > 0 \quad \text{and} \quad L = aE\sin^2\theta \tag{11.9b}$$

$$\text{in } \mathscr{M}_{\rm II}, \quad E < 0 \quad \text{and} \quad L = aE\sin^2\theta. \tag{11.9c}$$

In particular, (11.9a) is nothing but the result (11.4) already obtained in Example 1. We note also that the relation between L and E is identical to that obtained in Example 2 for the ingoing principal null geodesics [Eq. (11.6)].

In what follows, we will use Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$ as introduced in Sec. 10.2.1. Given the components (10.8) of the metric tensor $\boldsymbol{g}$ in these coordinates, evaluating E and L via $E = -g_{t\mu}\,p^\mu$ and $L = g_{\varphi\mu}\,p^\mu$ yields

$$E = \left(1-\frac{2mr}{\rho^2}\right)p^t + \frac{2amr\sin^2\theta}{\rho^2}\,p^\varphi. \tag{11.10}$$

$$L = -\frac{2amr\sin^2\theta}{\rho^2}\,p^t + \left(r^2+a^2+\frac{2a^2mr\sin^2\theta}{\rho^2}\right)\sin^2\theta\,p^\varphi, \tag{11.11}$$

where $\rho^2 := r^2 + a^2\cos^2\theta$ [Eq. (10.9)].

Let us recall that the components (p^α) of the 4-momentum are related to the parametric equation $x^\alpha = x^\alpha(\lambda)$ of the geodesic $\mathscr{L}$ in terms of the affine parameter λ by $p^\alpha = \mathrm{d}x^\alpha/\mathrm{d}\lambda$ [cf. Eq. (1.19)], i.e.

$$p^t = \frac{\mathrm{d}t}{\mathrm{d}\lambda}, \quad p^r = \frac{\mathrm{d}r}{\mathrm{d}\lambda}, \quad p^\theta = \frac{\mathrm{d}\theta}{\mathrm{d}\lambda}, \quad p^\varphi = \frac{\mathrm{d}\varphi}{\mathrm{d}\lambda}. \tag{11.12}$$

11.2.3 Mass as an integral of motion

The mass μ of particle $\mathscr{P}$ is related to the scalar square of the 4-momentum vector $\boldsymbol{p}$ via Eq. (1.8):

$$\mu^2 = -\boldsymbol{g}(\boldsymbol{p},\boldsymbol{p}). \tag{11.13}$$

The fact that $\mathscr{L}$ is a geodesic implies that μ is constant along $\mathscr{L}$ (cf. Eq. (B.6) in Appendix B). It therefore provides a third integral of motion, after E and L.

It is convenient to express (11.13) in terms of the inverse metric in order to let appear $p_t = -E$ and $p_\varphi = L$:

$$\mu^2 = -g^{\mu\nu} p_\mu p_\nu.$$

Given the components (10.16) of the inverse metric in Boyer-Lindquist coordinates, we get

$$\begin{aligned} \mu^2 &= \frac{1}{\Delta}\left(r^2 + a^2 + \frac{2a^2 mr \sin^2\theta}{\rho^2}\right) E^2 - \frac{4amr}{\rho^2\Delta} EL - \frac{1}{\Delta}\left(1 - \frac{2mr}{\rho^2}\right)\frac{L^2}{\sin^2\theta} \\ &\quad - \frac{\rho^2}{\Delta}\left(p^r\right)^2 - \rho^2\left(p^\theta\right)^2, \end{aligned} \tag{11.14}$$

where $\Delta := r^2 - 2mr + a^2$ [Eq. (10.10)]. Note that we have expressed p_r and p_θ in terms of p^r and p^θ thanks to the relations $p_r = g_{r\mu} p^\mu$ and $p_\theta = g_{\theta\mu} p^\mu$, which are very simple for the Boyer-Lindquist components (10.8) of $\boldsymbol{g}$:

$$p_r = \frac{\rho^2}{\Delta} p^r \qquad \text{and} \qquad p_\theta = \rho^2 p^\theta. \tag{11.15}$$

11.2.4 The fourth integral of motion: Carter constant

It turns out that the Kerr spacetime is endowed with a non-trivial Killing tensor of valence 2: the ***Walker-Penrose Killing tensor*** $\boldsymbol{K}$, which is the symmetric tensor of type $(0,2)$ defined by

$$\boxed{\boldsymbol{K} := (r^2 + a^2)\left(\underline{\boldsymbol{k}} \otimes \underline{\boldsymbol{\ell}} + \underline{\boldsymbol{\ell}} \otimes \underline{\boldsymbol{k}}\right) + r^2 \boldsymbol{g}}, \tag{11.16}$$

where $\underline{\boldsymbol{\ell}}$ and $\underline{\boldsymbol{k}}$ are the 1-forms associated by metric duality to the null vector fields $\boldsymbol{k}$ and $\boldsymbol{\ell}$ tangent to the principal null geodesics introduced in Sec. 10.4. In index notation, Eq. (11.16) writes

$$K_{\alpha\beta} = (r^2 + a^2)\left(k_\alpha \ell_\beta + \ell_\alpha k_\beta\right) + r^2 g_{\alpha\beta}. \tag{11.17}$$

$\boldsymbol{K}$ is called a ***Killing tensor*** because its symmetrized covariant derivative vanishes identically:

$$\boxed{\nabla_{(\alpha} K_{\beta\gamma)} = 0}. \tag{11.18}$$

This property can be seen as a generalization of the Killing equation (3.26) to tensors of valence 2. That the tensor $\boldsymbol{K}$ defined by (11.16) obeys the Killing identity (11.18) is established in the notebook D.4.7.

Killing tensors are discussed in Sec. B.5.2 of Appendix B. It is shown there that the Killing identity (11.18) implies that the following quantity is constant along any geodesic $\mathscr{L}$ [cf. Eq. (B.46)]:

$$\boxed{\mathscr{K} := \boldsymbol{K}(\boldsymbol{p}, \boldsymbol{p}) = K_{\mu\nu} p^\mu p^\nu}. \tag{11.19}$$

$\mathscr{K}$ is named ***Carter constant*** (cf. the historical note on p. 326). From the definition (11.16) of $\boldsymbol{K}$, we have

$$\mathscr{K} = 2(r^2 + a^2)\langle \underline{\boldsymbol{k}}, \boldsymbol{p}\rangle \langle \underline{\boldsymbol{\ell}}, \boldsymbol{p}\rangle + r^2 \boldsymbol{g}(\boldsymbol{p}, \boldsymbol{p}). \tag{11.20}$$

Now by Eq. (11.13), $\boldsymbol{g}(\boldsymbol{p},\boldsymbol{p}) = -\mu^2$. Besides, we have $\langle \underline{\boldsymbol{k}}, \boldsymbol{p}\rangle = k_\mu p^\mu = k^\mu p_\mu$. The last form lets appear the constants of motion $p_t = -E$ and $p_\varphi = L$ [Eq. (11.3)]. Using it with the components of $\boldsymbol{k}$ as given by Eq. (10.49), we get

$$\langle \underline{\boldsymbol{k}}, \boldsymbol{p}\rangle = -\frac{r^2+a^2}{\Delta}E - p_r + \frac{a}{\Delta}L.$$

Similarly, from the components (10.62) of $\boldsymbol{\ell}$, we obtain

$$\langle \underline{\boldsymbol{\ell}}, \boldsymbol{p}\rangle = -\frac{1}{2}E + \frac{\Delta}{2(r^2+a^2)}p_r + \frac{a}{2(r^2+a^2)}L.$$

Accordingly, Eq. (11.20) becomes

$$\mathscr{K} = \frac{1}{\Delta}\left[(r^2+a^2)E + \Delta p_r - aL\right]\left[(r^2+a^2)E - \Delta p_r - aL\right] - r^2\mu^2,$$

which can be rewritten as

$$\boxed{\mathscr{K} = \frac{1}{\Delta}\left[\left((r^2+a^2)E - aL\right)^2 - \rho^4 (p^r)^2\right] - r^2\mu^2}. \tag{11.21}$$

Note that we have expressed p_r in terms of p^r via Eq. (11.15).

Some physical interpretation of the Carter constant can be inferred from the above expression, in the case where the particle $\mathscr{P}$ following the geodesic $\mathscr{L}$ visits the asymptotic region $r \to +\infty$. Indeed, given that $\Delta := r^2 - 2mr + a^2 \sim r^2$ and $\rho^4 := (r^2 + a^2\cos^2\theta)^2 \sim r^4$ as $r \to +\infty$, we deduce from Eq. (11.21) that

$$\mathscr{K} \underset{r\to+\infty}{\sim} r^2\left[E^2 - \mu^2 - (p^r)^2\right]. \tag{11.22}$$

As discussed above, E is $\mathscr{P}$'s energy as measured by the asymptotic inertial observer $\mathscr{O}$. Then, according to Einstein's formula (1.24), $E^2 - \mu^2 = \boldsymbol{P}\cdot\boldsymbol{P}$, where $\boldsymbol{P}$ is $\mathscr{P}$'s linear momentum as measured by $\mathscr{O}$. Given that asymptotically, $p^r \sim P^r$ [cf. Eq. (1.20)], Eq. (11.22) becomes

$$\mathscr{K} \underset{r\to+\infty}{\sim} r^2\left[\boldsymbol{P}\cdot\boldsymbol{P} - (P^r)^2\right] = r^2\left[(P^{(\theta)})^2 + (P^{(\varphi)})^2\right],$$

where $P^{(\theta)} = rP^\theta \sim rp^\theta$ and $P^{(\varphi)} = r\sin\theta\, P^\varphi \sim r\sin\theta\, p^\varphi$ are the angular components of $\boldsymbol{P}$ in the orthonormal basis $(\boldsymbol{e}_{(r)}, \boldsymbol{e}_{(\theta)}, \boldsymbol{e}_{(\varphi)}) := (\boldsymbol{\partial}_r, r^{-1}\boldsymbol{\partial}_\theta, (r\sin\theta)^{-1}\boldsymbol{\partial}_\varphi)$. Now the total angular momentum of $\mathscr{P}$ measured by $\mathscr{O}$ is

$$\boldsymbol{L}_{\rm tot} := \boldsymbol{r}\times\boldsymbol{P} = -rP^{(\varphi)}\boldsymbol{e}_{(\theta)} + rP^{(\theta)}\boldsymbol{e}_{(\varphi)}.$$

Hence we may conclude that asymptotically, the Carter constant coincides with the square of $\mathscr{P}$'s angular momentum as measured by the inertial observer $\mathscr{O}$:

$$\mathscr{K} \underset{r\to+\infty}{\sim} \boldsymbol{L}_{\rm tot}\cdot\boldsymbol{L}_{\rm tot}. \tag{11.23}$$

We shall see later that one has always $\mathscr{K} \geq 0$. For now, let us establish the following characterization of null geodesics with $\mathscr{K} = 0$:

A null geodesic $\mathscr{L}$ has a vanishing Carter constant $\mathscr{K}$ if, and only if, $\mathscr{L}$ is a principal null geodesic:

$$\mathscr{K} = 0 \iff \mathscr{L} = \mathscr{L}^{\rm in}_{(v,\theta,\tilde\varphi)} \quad \text{or} \quad \mathscr{L} = \mathscr{L}^{\rm out}, \tag{11.24}$$

where $\mathscr{L}^{\rm in}_{(v,\theta,\tilde\varphi)}$ (resp. $\mathscr{L}^{\rm out}$) is one of the ingoing (resp. outgoing) principal null geodesic introduced in Sec. 10.4, with $\mathscr{L}^{\rm out}$ standing for $\mathscr{L}^{\rm out}_{(u,\theta,\tilde\varphi)}$, $\mathscr{L}^{{\rm out},\mathscr{H}}_{(\theta,\psi)}$ or $\mathscr{L}^{{\rm out},\mathscr{H}_{\rm in}}_{(\theta,\psi)}$.

Proof. Consider expression (11.20) for $\mathscr{K}$. If $\mathscr{L}$ is a null geodesic, the term $\boldsymbol{g}(\boldsymbol{p},\boldsymbol{p})$ vanishes identically, so that one is left with

$$\mathscr{K} = 2(r^2+a^2)(\boldsymbol{k}\cdot\boldsymbol{p})(\boldsymbol{\ell}\cdot\boldsymbol{p}).$$

Given that $r^2+a^2 \neq 0$ (this is clear for $a\neq 0$, while if $a=0$ (Schwarzschild case), $r=0$ is excluded from the spacetime manifold), we have then

$$\mathscr{K} = 0 \iff \boldsymbol{k}\cdot\boldsymbol{p} = 0 \quad \text{or} \quad \boldsymbol{\ell}\cdot\boldsymbol{p} = 0.$$

The vectors $\boldsymbol{k}$, $\boldsymbol{\ell}$ and $\boldsymbol{p}$ are all null. Now, according to the Corollary 2 in Sec. 1.2.2, two null vectors are orthogonal iff they are collinear. Hence

$$\mathscr{K} = 0 \iff \boldsymbol{p} = \alpha\boldsymbol{k} \quad \text{or} \quad \boldsymbol{p} = \alpha\boldsymbol{\ell}.$$

Since $\boldsymbol{k}$ (resp. $\boldsymbol{\ell}$) is tangent to $\mathscr{L}^{\rm in}_{(v,\theta,\tilde\varphi)}$ (resp. $\mathscr{L}^{\rm out}$), this completes the proof. □

Remark 2: The result (11.24) is consistent with the interpretation (11.23) of $\mathscr{K}$, since asymptotically the principal null geodesics are purely radial[5], and hence have $\boldsymbol{L}_{\rm tot} = 0$.

Example 4 (Generators of the event and inner horizons): We have seen in Example 1 in Sec. 11.2.2 that the null geodesic generators of the event horizon $\mathscr{H}$ and the inner horizon $\mathscr{H}_{\rm in}$ have $E=0$ and $L=0$. Since these generators are principal null geodesics (cf. Secs. 10.5.3 and 10.8.3), the above results shows that in addition $\mathscr{K}=0$. As $\mu=0$ by definition of a null geodesic, we see that the four integrals of motion μ, E, L and $\mathscr{K}$ all vanish for these geodesics.

The converse is true:

The null geodesic generators of the event horizon $\mathscr{H}$ and of the inner horizon $\mathscr{H}_{\rm in}$ are the only geodesics of Kerr spacetime having their four integrals of motion vanishing:

$$\mathscr{L} \text{ null geodesic generator of } \mathscr{H} \text{ or } \mathscr{H}_{\rm in} \iff (\mu, E, L, \mathscr{K}) = (0,0,0,0). \tag{11.25}$$

Proof. If a geodesic $\mathscr{L}$ has $(\mu,\mathscr{K})=(0,0)$, $\mathscr{L}$ is necessary null and (11.24) shows that $\mathscr{L}$ is a principal null geodesic. Moreover Eq. (11.21) with $(\mu,E,L,\mathscr{K})=(0,0,0,0)$ implies $p^r = 0$, i.e. $\mathscr{L}$ lies at a constant value of r. If $\mathscr{L}$ is ingoing, then $\boldsymbol{p}\propto\boldsymbol{k}$, with the Boyer-Lindquist components of $\boldsymbol{k}$ given by Eq. (10.49). Since $k^r=-1$, this precludes $p^r=0$. Hence $\mathscr{L}$ is an outgoing principal null geodesic and one has $\boldsymbol{p}\propto\boldsymbol{\ell}$, with the

[5] Indeed, we deduce from Eqs. (10.49) and (10.62) that, for $r\to+\infty$, $\boldsymbol{k}\sim\boldsymbol{\partial}_t-\boldsymbol{\partial}_r$ and $\boldsymbol{\ell}\sim(\boldsymbol{\partial}_t+\boldsymbol{\partial}_r)/2$.

Boyer-Lindquist components of $\boldsymbol{\ell}$ given by Eq. (10.62). We read $\ell^r = \Delta/(2(r^2+a^2))$ with $\Delta \neq 0$, except precisely on $\mathscr{H}$ and $\mathscr{H}_{\rm in}$[6]. Hences $p^r = 0$ is possible only on $\mathscr{H}$ and $\mathscr{H}_{\rm in}$. □

Historical note: That $\mathscr{K} = 0$ for principal null geodesics has been pointed out by Jiří Bičák and Zdeněk Stuchlík in 1976 [24].

11.2.5 First order equations of motion

We have thus four first integrals of the geodesic equation (11.1) at disposal: E [Eq. (11.10)], L [Eq. (11.11)], μ^2 [Eq. (11.14)] and $\mathscr{K}$ [Eq. (11.21)]. In the expressions of each of these integrals, p^α has to be thought of as the first order derivative $\mathrm{d}x^\alpha/\mathrm{d}\lambda$ [Eq. (11.12)]. Two first integrals, namely E and L, are linear in the p^α's, while the two others, namely μ^2 and $\mathscr{K}$, are quadratic. Furthermore, Eqs. (11.10) and (11.11) constitute a decoupled subsystem for (p^t, p^φ), which can easily be solved[7], yielding

$$\rho^2 p^t = \frac{1}{\Delta}[(r^2+a^2)^2 E - 2amrL] - a^2 E \sin^2\theta \tag{11.26}$$

$$\rho^2 p^\varphi = \frac{L}{\sin^2\theta} + \frac{a}{\Delta}(2mrE - aL). \tag{11.27}$$

Besides, Eq. (11.21) involves only p^r and can be recast as

$$\rho^4 (p^r)^2 = R(r), \tag{11.28}$$

with

$$\boxed{R(r) := \left[(r^2+a^2)E - aL\right]^2 - \Delta(r^2\mu^2 + \mathscr{K})}. \tag{11.29}$$

In the above expression, let us recall that Δ is the function of r given by Eq. (10.10): $\Delta := r^2 - 2mr + a^2$. All other quantities are constant. Accordingly, $R(r)$ is a 4th order polynomial in r. Equation (11.28) implies that, along the geodesic $\mathscr{L}$, it has to fulfill

$$\boxed{R(r) \geq 0}. \tag{11.30}$$

Finally, if we substitute p^r by the value given by Eqs. (11.28)-(11.29) in the mass first integral (11.14), we get, after simplification,

$$\rho^4 (p^\theta)^2 = \Theta(\theta), \tag{11.31}$$

with

$$\boxed{\Theta(\theta) := \mathscr{K} - \left(\frac{L}{\sin\theta} - aE\sin\theta\right)^2 - \mu^2 a^2 \cos^2\theta}. \tag{11.32}$$

[6] This is graphically confirmed by Figs. 10.6 and 10.7, which show that $\mathscr{H}$ and $\mathscr{H}_{\rm in}$ are the only locations where a principal null geodesic can have $r = \text{const}$.

[7] An intermediate step is combining Eqs. (11.10) and (11.11) to get $aE - L/\sin^2\theta = ap^t - (r^2+a^2)p^\varphi$ and $(r^2+a^2)E - aL = \Delta(p^t - a\sin^2\theta p^\varphi)$.

Equation. (11.31) imposes that $\Theta(\theta)$ is non-negative along the geodesic $\mathscr{L}$:

$$\boxed{\Theta(\theta) \geq 0}. \tag{11.33}$$

The following constant is often used instead of $\mathscr{K}$:

$$\boxed{Q := \mathscr{K} - (L - aE)^2} \tag{11.34}$$

Thanks to it, we may rewrite (11.32) as

$$\boxed{\Theta(\theta) = Q + \cos^2\theta \left[a^2(E^2 - \mu^2) - \frac{L^2}{\sin^2\theta}\right]}. \tag{11.35}$$

Following the standard usage, we call Q the ***Carter constant*** as well. To distinguish between the two Carter constants, we shall specify *Carter constant* Q or *Carter constant* $\mathscr{K}$. As we shall see in Sec. 11.3.6, Q is well adapted to the description of the θ-motion of geodesics. On the other hand, a nice property of $\mathscr{K}$, which is not shared by Q, is to be always non-negative, as Eqs. (11.32) and (11.33) show:

$$\boxed{\mathscr{K} \geq 0}. \tag{11.36}$$

If the particle $\mathscr{P}$ reaches the asymptotic region $r \gg m$, we deduce from Eqs. (11.34) and (11.23) the following behavior of Q:

$$Q \underset{r\to+\infty}{\sim} \boldsymbol{L}_{\rm tot} \cdot \boldsymbol{L}_{\rm tot} - L^2 + aE(2L - aE). \tag{11.37}$$

Hence, if $a = 0$, Q can be interpreted as the square of the part of $\mathscr{P}$'s angular momentum (measured by the asymptotic inertial observer) that is not in L.

Example 5 (Carter constant Q of the principal null geodesics): As (11.24) shows, for a principal null geodesic, be it ingoing or outgoing, the Carter constant $\mathscr{K}$ vanishes identically. According to Eq. (11.34), the Carter constant Q is then $Q = -(L - aE)^2$. In view of the relation $L = aE\sin^2\theta$ for these geodesics [Eqs. (11.6) and (11.9)], we get

$$Q = -a^2E^2\cos^4\theta, \tag{11.38}$$

where θ is the constant value of the θ-coordinate along the principal null geodesic. Note that the above relation holds in all Kerr spacetime, including on the horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, where it reduces to $Q = 0$, for $E = 0$ there [Eq. (11.9a)]. Equation (11.38) implies

$$Q \leq 0, \tag{11.39}$$

with $Q = 0$ only for principal null geodesics lying in the equatorial plane or for the outgoing principal null geodesics $\mathscr{L}^{\rm out,\mathscr{H}}_{(\theta,\psi)}$ and $\mathscr{L}^{\rm out,\mathscr{H}_{\rm in}}_{(\theta,\psi)}$ generating the horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$.

Remark 3: As for $\mathscr{K}$, one may derive the Carter constant Q from a Killing tensor. Indeed, from the Walker-Penrose Killing tensor $\boldsymbol{K}$ [Eq. (11.16)], let us form the tensor field

$$\tilde{\boldsymbol{K}} := \boldsymbol{K} - \underline{\tilde{\boldsymbol{\eta}}} \otimes \underline{\tilde{\boldsymbol{\eta}}}, \quad \text{where} \quad \tilde{\boldsymbol{\eta}} := \boldsymbol{\eta} + a\boldsymbol{\xi}. \tag{11.40}$$

Being a linear combination with constant coefficients of the Killing vectors $\boldsymbol{\eta}$ and $\boldsymbol{\xi}$, $\tilde{\boldsymbol{\eta}}$ is itself a Killing vector. It follows that $\underline{\tilde{\boldsymbol{\eta}}} \otimes \underline{\tilde{\boldsymbol{\eta}}}$ is a Killing tensor (cf. Example 3 in Sec. B.5.2), so that $\tilde{\boldsymbol{K}}$ a Killing tensor, the Killing equation (11.18) being linear. Applying $\tilde{\boldsymbol{K}}$ to the 4-momentum $\boldsymbol{p}$, we get

$$\tilde{\boldsymbol{K}}(\boldsymbol{p},\boldsymbol{p}) = \underbrace{\boldsymbol{K}(\boldsymbol{p},\boldsymbol{p})}_{\mathscr{K}} - (\underbrace{\langle \underline{\boldsymbol{\eta}}, \boldsymbol{p}\rangle}_{L} + a \underbrace{\langle \underline{\boldsymbol{\xi}}, \boldsymbol{p}\rangle}_{-E})^2 .$$

Comparing with the definition (11.34), we conclude that

$$Q = \tilde{\boldsymbol{K}}(\boldsymbol{p},\boldsymbol{p}). \tag{11.41}$$

The Boyer-Lindquist components of the contravariant tensor associated to $\tilde{\boldsymbol{K}}$ by metric duality have an expression particularly simple in terms of those of the inverse metric (cf. the notebook D.4.7):

$$\tilde{K}^{\alpha\beta} = -a^2 \cos^2\theta \, g^{\alpha\beta} + \mathrm{diag}(-a^2\cos^2\theta, 0, 1, \tan^{-2}\theta)^{\alpha\beta}. \tag{11.42}$$

In view of the relation (11.12) between the p^α's and the derivatives of the functions $x^\alpha(\lambda)$, we may collect Eqs. (11.26), (11.27), (11.28) and (11.32) as the first-order system

$$\boxed{\rho^2 \frac{\mathrm{d}t}{\mathrm{d}\lambda} = \frac{1}{\Delta}[(r^2+a^2)^2 E - 2amrL] - a^2 E \sin^2\theta} \tag{11.43a}$$

$$\boxed{\rho^2 \frac{\mathrm{d}r}{\mathrm{d}\lambda} = \epsilon_r \sqrt{R(r)}} \tag{11.43b}$$

$$\boxed{\rho^2 \frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \epsilon_\theta \sqrt{\Theta(\theta)}} \tag{11.43c}$$

$$\boxed{\rho^2 \frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{L}{\sin^2\theta} + \frac{a}{\Delta}(2mrE - aL)}, \tag{11.43d}$$

where $\epsilon_r := \mathrm{sgn}\, p^r = \pm 1$, $\epsilon_\theta := \mathrm{sgn}\, p^\theta = \pm 1$ and the functions $R(r)$ and $\Theta(\theta)$ are defined by Eq. (11.29) and Eq. (11.32) or (11.35). Since $p^r = \mathrm{d}r/\mathrm{d}\lambda$, ϵ_r is $+1$ (resp. -1) in the parts of the geodesic $\mathscr{L}$ where r increases (resp. decreases) with λ. Similarly, ϵ_θ is $+1$ (resp. -1) in the parts of $\mathscr{L}$ where θ increases (resp. decreases) with λ.

We may rewrite Eq. (11.29) for $R(r)$ in terms of Q instead of $\mathscr{K}$, via Eq. (11.34):

$$\boxed{R(r) := \left[(r^2+a^2)E - aL\right]^2 - \Delta\left[r^2\mu^2 + Q + (L-aE)^2)\right]}. \tag{11.44}$$

Note that, beside the constants of motion E, L, μ and Q, $R(r)$ depends on both Kerr parameters a and m (via $\Delta = r^2 - 2mr + a^2$), while $\Theta(\theta)$ depends on a only [cf. Eq. (11.35)]. Along $\mathscr{L}$, these functions must obey $R(r) \geq 0$ [Eq. (11.30)] and $\Theta(\theta) \geq 0$ [Eq. (11.33)].

11.2.6 Turning points

Let $\mathscr{L}$ be a geodesic that is not stuck at some constant value of the coordinate r. We define a ***r-turning point*** of $\mathscr{L}$ as a point $p_0 \in \mathscr{L}$, the r-coordinate r_0 of which obeys

$$R(r_0) = 0 \quad \text{and} \quad R'(r_0) \neq 0, \tag{11.45}$$

i.e. r_0 is a simple root of the polynomial R.

We have then

$$\left.\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right|_{\lambda_0} = 0 \quad \text{and} \quad \left.\frac{\mathrm{d}^2 r}{\mathrm{d}\lambda^2}\right|_{\lambda_0} = \frac{R'(r_0)}{2\rho_0^4} \neq 0, \tag{11.46}$$

where λ_0 is the value of the affine parameter λ at p_0 and $\rho_0 := \rho(p_0)$.

Proof. The vanishing of $\mathrm{d}r/\mathrm{d}\lambda$ at λ_0 follows immediately from Eq. (11.43b) with $R(r_0) = 0$, since ρ^2 never vanishes on $\mathscr{M}$. Besides, by taking the derivative of Eq. (11.43b) with respect to λ, we get

$$2\left(r\frac{\mathrm{d}r}{\mathrm{d}\lambda} + a^2\cos\theta\sin\theta\frac{\mathrm{d}\theta}{\mathrm{d}\lambda}\right)\frac{\mathrm{d}r}{\mathrm{d}\lambda} + \rho^2\frac{\mathrm{d}^2 r}{\mathrm{d}\lambda^2} = \epsilon_r\frac{R'(r)}{2\sqrt{R(r)}}\frac{\mathrm{d}r}{\mathrm{d}\lambda} = \frac{R'(r)}{2\rho^2}.$$

At $\lambda = \lambda_0$, the first term in left-hand side vanishes identically, due to the $\mathrm{d}r/\mathrm{d}\lambda$ factor, and we get the second part of (11.46). □

We deduce from the result (11.46) that at $\lambda = \lambda_0$, $\mathrm{d}r/\mathrm{d}\lambda$ moves from positive to negative values or vice-versa (depending on the sign of $R'(r_0)$), which means that the function $r(\lambda)$ switches from increasing to decreasing or vice-versa, hence the name *r-turning point*. The factor $\epsilon_r = \pm 1$ in Eq. (11.43b) necessarily changes sign at $\lambda = \lambda_0$.

Remark 4: For a generic smooth function $r(\lambda)$ with $\mathrm{d}r/\mathrm{d}\lambda = 0$ at λ_0, the condition $\mathrm{d}^2 r/\mathrm{d}\lambda^2 \neq 0$ at λ_0 is sufficient but not necessary for r to change its direction of variation there. Indeed, the same property holds with $\mathrm{d}^2 r/\mathrm{d}\lambda^2 = 0$ and higher order derivatives vanishing up to some even order k for which $\mathrm{d}^k r/\mathrm{d}\lambda^k \neq 0$. However, in the present case, $\mathrm{d}^2 r/\mathrm{d}\lambda^2 = 0$ would imply $R'(r_0) = 0$ and we shall see Sec. 11.3.5 that a geodesic can reach such a point only asymptotically, i.e. for $\lambda \to +\infty$. Hence it cannot be a turning point.

Similarly, if $\mathscr{L}$ is a geodesic that is not stuck at some constant value of the coordinate θ, we define a ***θ-turning point*** of $\mathscr{L}$ as a point $p_0 \in \mathscr{L}$, the θ-coordinate θ_0 of which obeys

$$\Theta(\theta_0) = 0 \quad \text{and} \quad \Theta'(\theta_0) \neq 0. \tag{11.47}$$

We deduce then from the equation of motion (11.43c):

$$\left.\frac{\mathrm{d}\theta}{\mathrm{d}\lambda}\right|_{\lambda_0} = 0 \quad \text{and} \quad \left.\frac{\mathrm{d}^2\theta}{\mathrm{d}\lambda^2}\right|_{\lambda_0} = \frac{\Theta'(\theta_0)}{2\rho_0^4} \neq 0. \tag{11.48}$$

This implies that at a θ-turning point, the function $\theta(\lambda)$ switches from increasing to decreasing or vice-versa. The factor $\epsilon_\theta = \pm 1$ in Eq. (11.43c) necessarily changes sign at such a point.

Remark 5: A comment similar to Remark 4 can be made: a geodesic with varying θ that has $\mathrm{d}\theta/\mathrm{d}\lambda = 0$ for some finite value of λ cannot have $\mathrm{d}^2\theta/\mathrm{d}\lambda^2 = 0$ at the same point, since we shall see in Sec. 11.3.5 that a value of θ with both $\Theta(\theta) = 0$ and $\Theta'(\theta) = 0$ can only be reached asymptotically along a geodesic. Hence (11.47) is a necessary and sufficient condition for a θ-turning point.

11.2.7 Equations of motion in terms of Mino parameter

In view of the right-hand sides of the system (11.43), it is quite natural to introduce a new parameter λ' along the geodesic $\mathscr{L}$ such that

$$\mathrm{d}\lambda' = \frac{\mathrm{d}\lambda}{\rho^2} = \frac{\mathrm{d}\lambda}{r(\lambda)^2 + a^2\cos^2\theta(\lambda)}. \tag{11.49}$$

Since ρ^2 never vanishes on the spacetime manifold $\mathscr{M}$ (by construction of the latter, cf. Eq. (10.28)), the above relation leads to a well-defined parameter[8] along $\mathscr{L}$. Moreover, since $\rho^2 > 0$, λ' increases towards the future, as λ. A difference between the two parametrizations is that λ' is not an affine parameter of $\mathscr{L}$ in general[9], contrary to λ. The parameter λ' is called ***Mino parameter*** [188].

In terms of Mino parameter, the system (11.43) becomes

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\lambda'} = \frac{1}{\Delta}[(r^2+a^2)^2E - 2amrL] - a^2E\sin^2\theta} \tag{11.50a}$$

$$\boxed{\frac{\mathrm{d}r}{\mathrm{d}\lambda'} = \epsilon_r\sqrt{R(r)}} \tag{11.50b}$$

$$\boxed{\frac{\mathrm{d}\theta}{\mathrm{d}\lambda'} = \epsilon_\theta\sqrt{\Theta(\theta)}} \tag{11.50c}$$

$$\boxed{\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda'} = \frac{L}{\sin^2\theta} + \frac{a}{\Delta}(2mrE - aL)}, \tag{11.50d}$$

where R is the quartic polynomial defined by Eq. (11.44) and Θ is the function defined by Eq. (11.35). It is remarkable that Eqs. (11.50b) and (11.50c) are two fully decoupled equations: Eq. (11.50b) is an ordinary differential equation (ODE) for the function $r(\lambda')$, while Eq. (11.50c) is an ordinary differential equation for the function $\theta(\lambda')$. This was not the case for Eqs. (11.43b) and (11.43c) since ρ^2 involves both r and θ.

11.2.8 Integration of the geodesic equations

The ODE (11.50b) can be integrated by the method of separation of variables. On a part of $\mathscr{L}$ where $R(r) \neq 0$, this yields

$$\lambda' - \lambda'_0 = \int_{r_0}^{r} \frac{\epsilon_r\,\mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}}, \tag{11.51}$$

with $r_0 := r(\lambda'_0)$. The hypothesis $R(r) \neq 0$ excludes any r-turning point between λ'_0 and λ', so that $\epsilon_r = \pm 1$ is constant along the considered part of $\mathscr{L}$.

Actually, the solution (11.51) can be extended to include a turning point at one or two of its boundaries, despite $R(r) = 0$ there. Indeed, let us assume that $r = r_1$ corresponds

[8] Note however that λ' may blow up if $\mathscr{L}$ comes arbitrarily close to the ring singularity, i.e. if $\rho \to 0$.

[9] The only exception is for a circular orbit at $\theta = \pi/2$, since then ρ^2 is constant and Eq. (11.49) reduces to an affine relation between λ and λ'.

to a r-turning point of $\mathscr{L}$. Due to $R'(r_1) \neq 0$ [Eq. (11.45)], the integral in the right-hand side of (11.51) with $r = r_1$ is finite. Indeed, the Taylor expansion $R(\bar{r}) = R'(r_1)(\bar{r} - r_1) + O((\bar{r} - r_1)^2)$ makes the integral behave near r_1 as[10]

$$\frac{1}{\sqrt{-R'(r_1)}} \int_{r_0}^{r_1} \frac{\mathrm{d}\bar{r}}{\sqrt{r_1 - \bar{r}}},$$

which is a convergent improper integral.

Remark 6: This is the same reasoning as in Sec. 8.3.6.

Let us assume that there are $M \geq 1$ r-turning points $p_1, \ldots, p_M$ between λ'_0 and λ'. Their radial coordinates take at most two distinct values, r_1 and r_2, such that $r(p_1) = r_1$, $r(p_2) = r_2$, $r(p_3) = r_1$, $r(p_4) = r_2$, etc. From the above convergence property, the solution of Eq. (11.50b) is then

$$\lambda' - \lambda'_0 = \int_{r_0}^{r_1} \frac{\epsilon_r \, \mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}} + (M-1) \int_{r_1}^{r_2} \frac{\epsilon_r \, \mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}} + \int_{r_{1,2}}^{r} \frac{\epsilon_r \, \mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}}, \tag{11.52}$$

where $r_{1,2} = r_1$ for M odd and $r_{1,2} = r_2$ for M even. Note that each term in the above sum is positive, the sign of ϵ_r compensating the order of the integral boundaries.

The right-hand side of Eq. (11.52) is actually a path integral and is often abriged by means of a slash notation:

$$\lambda' - \lambda'_0 = \not\int_{r_0}^{r} \frac{\epsilon_r \, \mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}}. \tag{11.53}$$

Similarly, if $\Theta(\theta) \neq 0$ along $\mathscr{L}$, except possibly at some θ-turning points, Eq. (11.50c) can be integrated as

$$\begin{aligned} \lambda' - \lambda'_0 &= \not\int_{\theta_0}^{\theta} \frac{\epsilon_\theta \, \mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} \\ &= \begin{cases} \displaystyle\int_{\theta_0}^{\theta} \frac{\epsilon_\theta \, \mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} & \text{if } N = 0 \\ \displaystyle\int_{\theta_0}^{\theta_1} \frac{\epsilon_\theta \, \mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} + (N-1) \int_{\theta_1}^{\theta_2} \frac{\epsilon_\theta \, \mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} + \int_{\theta_{1,2}}^{\theta} \frac{\epsilon_\theta \, \mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} & \text{if } N \geq 1, \end{cases} \end{aligned} \tag{11.54}$$

where $\theta_0 := \theta(\lambda'_0)$, N is the number of θ-turning points between λ'_0 and λ', θ_1 (resp. θ_2) is the value of θ at the first (resp. second) turning point, if any, $\theta_{1,2} = \theta_1$ for N odd and $\theta_{1,2} = \theta_2$ for N even.

We are now in position to state the full expression of the general solution for geodesic motion:

[10] We assume here $r_0 < r_1$, so that the constraint $R(\bar{r}) \geq 0$ on the interval $[r_0, r_1]$ implies $R'(r_1) < 0$.

Let $\mathscr{L}$ be a null or timelike geodesic in Kerr spacetime, with conserved energy E, conserved angular momentum L, mass μ and Carter constant Q. We assume that the associated quartic polynomial $R(r)$, as defined by Eq. (11.44) (see also Eq. (11.86) below), and the associated function $\Theta(\theta)$, as defined by Eq. (11.35), do not vanish along $\mathscr{L}$ except possibly at some r-turning points or θ-turning points. Let λ be the affine parameter of $\mathscr{L}$ associated with the 4-momentum $\boldsymbol{p}$ and λ' the Mino parameter. If for $\lambda = \lambda_0$ $\mathscr{L}$ lies at the point of Boyer-Lindquist coordinates $(t_0, r_0, \theta_0, \varphi_0)$, then at any value of λ, the Boyer-Lindquist coordinates (t, r, θ, φ) along $\mathscr{L}$ obey

$$\lambda' - \lambda'_0 = \int_{r_0}^{r} \frac{\epsilon_r \,\mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}} = \int_{\theta_0}^{\theta} \frac{\epsilon_\theta \,\mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} \tag{11.55a}$$

$$\lambda - \lambda_0 = \int_{r_0}^{r} \frac{\epsilon_r \bar{r}^2 \mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}} + a^2 \int_{\theta_0}^{\theta} \frac{\epsilon_\theta \cos^2\bar{\theta} \,\mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} \tag{11.55b}$$

$$t - t_0 = \int_{r_0}^{r} \frac{(\bar{r}^2+a^2)^2 E - 2am\bar{r}L}{\bar{r}^2 - 2m\bar{r} + a^2} \frac{\epsilon_r \mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}} - a^2 E \int_{\theta_0}^{\theta} \sin^2\bar{\theta} \frac{\epsilon_\theta \,\mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} \tag{11.55c}$$

$$\varphi - \varphi_0 = a \int_{r_0}^{r} \frac{2m\bar{r}E - aL}{\bar{r}^2 - 2m\bar{r} + a^2} \frac{\epsilon_r \,\mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}} + L \int_{\theta_0}^{\theta} \frac{1}{\sin^2\bar{\theta}} \frac{\epsilon_\theta \,\mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}}. \tag{11.55d}$$

Proof. Equation (11.55a) is nothing but the gathering of Eqs. (11.53) and (11.54). For Eq. (11.55b), it suffices to rewrite Eq. (11.49) as

$$\mathrm{d}\lambda = r^2 \,\mathrm{d}\lambda' + a^2 \cos^2\theta \,\mathrm{d}\lambda'$$

and to substitute $\mathrm{d}\lambda'$ by $\epsilon_r \mathrm{d}r/\sqrt{R(r)}$ in the first term [cf. Eq. (11.50b)] and by $\epsilon_\theta \mathrm{d}\theta/\sqrt{\Theta(\theta)}$ in the second term [cf. Eq. (11.50c)]. Similarly, by rewriting Eq. (11.50a) as

$$\mathrm{d}t = \frac{(r^2+a^2)^2 E - 2amrL}{r^2 - 2mr + a^2} \,\mathrm{d}\lambda' - a^2 E \sin^2\theta \,\mathrm{d}\lambda'$$

and performing the same substitutions for $\mathrm{d}\lambda'$ as above, we get Eq. (11.55c). Finally, Eq. (11.55d) is deduced in the same fashion from Eq. (11.50d). □

The system (11.55) shows that the geodesic motion can be fully solved in terms of the Mino parameter λ'. Indeed, the integral on r in Eq. (11.55a) can be evaluated by means of elliptic integrals since R is a polynomial of degree 4. This provides $\lambda' = \lambda'(r)$. Inverting this relation via Jacobi elliptic functions (cf. Remark 5 on p. 186) yields $r = r(\lambda')$. The integral on θ in Eq. (11.55a) can be evaluated by means of elliptic integrals as well since the change of variable $\zeta := \cos\theta$ along with expression (11.35) for $\Theta(\theta)$ results in

$$\int_{\theta_0}^{\theta} \frac{\epsilon_\theta \,\mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} = -\int_{\zeta_0}^{\zeta} \frac{\epsilon_\theta \,\mathrm{d}\bar{\zeta}}{\sqrt{Q(1-\bar{\zeta}^2) + \bar{\zeta}^2[a^2(E^2-\mu^2)(1-\bar{\zeta}^2) - L^2]}},$$

with the term under the square root in the right-hand side being a polynomial of degree 4 in $\bar{\zeta}$. The use of Jacobi elliptic functions leads then to $\theta = \theta(\lambda')$. From $r(\lambda')$ and $\theta(\lambda')$ one

can get the functions $\lambda(\lambda')$, $t(\lambda')$ and $\varphi(\lambda')$ by evaluating the integrals in the right-hand sides of Eqs. (11.55b)–(11.55d). Again, these integrals are reducible to elliptic integrals. We shall not give the details of all the elliptic integrals computations, referring the reader to Ref. [113] for bound timelike geodesics and to Ref. [131] for null geodesics.

Historical note: The constant of motion $\mathscr{K}$ has been discovered by Brandon Carter in 1968 [41], in a study about Kerr-Newman spacetimes, which generalize the Kerr ones to nonzero global electric charge. Carter actually did not get $\mathscr{K}$ from the Killing tensor $\boldsymbol{K}$, which was discovered only two years later by Martin Walker and Roger Penrose [258]; he started instead from the Lagrangian (B.39) governing both timelike and null geodesics, derived the corresponding Hamiltonian as $H = g^{\mu\nu} p_\mu p_\nu$ (for the uncharged (i.e. Kerr) case) and discovered that the Hamilton-Jacobi equation is separable, i.e. can be solved by separation of variables. $\mathscr{K}$ appeared then as a separation constant. Note that Carter used the Kerr coordinates[11] $(v, r, \theta, \tilde{\varphi})$ described in Sec. 10.3.1 and not the Boyer-Lindquist ones. Carter also introduced the constant Q via Eq. (11.34). He obtained the equivalent of the first-order system (11.43) for the Kerr coordinates. Actually two of his equations, those for $\mathrm{d}r/\mathrm{d}\lambda$ and $\mathrm{d}\theta/\mathrm{d}\lambda$, are identical to Eqs. (11.43b) and (11.43c) of the Boyer-Lindquist system. This is not surprising since the coordinates r and θ are the same in both systems. Carter's equations for the Kerr coordinates v and $\tilde{\varphi}$ are slightly more complicated than Eqs. (11.43a) and (11.43d) for the Boyer-Lindquist coordinates t and φ. It seems that the Boyer-Lindquist first-order system (11.43) has been first derived by Daniel Wilkins in 1972 [264], starting from Carter's system and performing the transformation to Boyer-Lindquist coordinates. The integrated geodesic equations (11.55) have been obtained by Carter in 1968 [41]: Eq. (11.55a) (without the $\lambda' - \lambda'_0$ part) is Eq. (58) of Ref. [41] and Eq. (11.55b) is Eq. (59) of Ref. [41]. Regarding Eqs. (11.55c) and (11.55d), Carter obtained equivalent ones for the Kerr coordinates v and $\tilde{\varphi}$, which he was using [his Eqs. (60) and (61)].

11.3 General properties of geodesics

11.3.1 Sign of E

We have:

If a null or timelike geodesic $\mathscr{L}$ has some part lying in the exterior of the ergoregion $\mathscr{G}$ (cf. Sec. 10.2.4), then the conserved energy E defined by Eq. (11.2a) is necessarily positive:

$$\mathscr{L} \not\subset \overline{\mathscr{G}} \quad \Longrightarrow \quad E > 0. \tag{11.56}$$

Proof. By the very definition of the ergoregion $\mathscr{G}$ (cf. Sec. 10.2.4), the Killing vector $\boldsymbol{\xi}$ is timelike in the exterior of $\mathscr{G}$, i.e. in $\mathscr{M} \setminus \overline{\mathscr{G}}$. Moreover, it is future-directed there, given the time orientation defined in Sec. 10.3.2. The 4-momentum $\boldsymbol{p}$ is either timelike or null and always future-directed. By Eq. (1.3a) in Lemma 1 of Sec. 1.2.2, one has then necessarily $\boldsymbol{\xi} \cdot \boldsymbol{p} < 0$; hence $E := -\boldsymbol{\xi} \cdot \boldsymbol{p} > 0$ in $\mathscr{M} \setminus \mathscr{G}$. Since E is constant along $\mathscr{L}$, it follows that $E > 0$ everywhere. □

[11] Carter's u is our v.

In particular, any timelike or null geodesic that reaches one of the asymptotic regions $r \to \pm\infty$ has $E > 0$.

Remark 1: The property (11.56) generalizes that obtained for Schwarzschild spacetime [cf. Eq. (7.5)] to the case $a \neq 0$, since for Schwarzschild spacetime, the exterior of the ergoregion is nothing but the exterior of the black hole region.

Inside the ergoregion, the Killing vector $\boldsymbol{\xi}$ is spacelike and E can be either positive, zero or negative. A particle with $E < 0$ (resp. $E = 0$) is called a ***negative-energy particle*** (resp. a ***zero-energy particle***). We have already encountered zero-energy particles in Example 1 of Sec. 11.2.2: the photons whose worldlines are the null generators of the horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$. Negative-energy particles are those involved in the Penrose process discussed in Sec. 10.5.5. Note that according to (11.56), neither a negative-energy particle nor a zero-energy one can exist in the exterior of the ergoregion.

Example 6 (Outgoing principal null geodesics in $\mathscr{M}_{\rm II}$): We have seen in Example 3 (p. 314) that in the region $\mathscr{M}_{\rm II}$, the outgoing principal null geodesics $\mathscr{L}^{\rm out}_{(u,\theta,\tilde{\varphi})}$ have $E < 0$ [Eq. (11.9c)]. This is consistent with (11.56) because $\mathscr{M}_{\rm II}$ is entirely contained in the ergoregion $\mathscr{G}$ (cf. Fig. 10.2).

Remark 2: As already stressed in Remark 4 on p. 285, any local observer measuring the energy of a negative-energy particle or a zero-energy one, according to the process described in Sec. 1.4, will find a positive value. Hence there is nothing mysterious about these particles.

11.3.2 Future-directed condition

Not all values of the constant of motions (μ, E, L, Q) lead to a solution $x^\alpha(\lambda)$ of the first-order system (11.43) that is future-directed, i.e. such that $\boldsymbol{p} = {\rm d}\boldsymbol{x}/{\rm d}\lambda$ is oriented everywhere towards the future. The latter condition is equivalent to have the affine parameter λ increase towards the future along the geodesic $\mathscr{L}$, as we are demanding in all our study (cf. Sec. 11.2.1). In particular, this allows one to identify λ with the proper time when $\mathscr{L}$ is timelike.

As discussed in Sec. 10.3.2, the time orientation of the Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ is set by the global null vector field $\boldsymbol{k}$ generating the ingoing principal null geodesics, cf. Eq. (10.31). By virtue of Lemma 2 of Sec. 1.2.2 [cf. Eq. (1.4a)], we have then, for $\boldsymbol{p}$ not collinear with $\boldsymbol{k}$ (in particular for $\mathscr{L}$ timelike),

$$\boldsymbol{p} \text{ future-directed} \iff \boldsymbol{g}(\boldsymbol{k},\boldsymbol{p}) < 0 \iff k^\mu p_\mu < 0 \iff \frac{r^2+a^2}{\Delta} p_t - p_r + \frac{a}{\Delta} p_\varphi < 0,$$

where the last inequality follows from the Boyer-Lindquist components (10.49) of $\boldsymbol{k}$. Given the identities $p_t = -E$, $p_\varphi = L$ [Eq. (11.3)] and $p_r = \epsilon_r \sqrt{R(r)}/\Delta$ [Eqs. (11.15) and (11.43b)], we conclude that

$$\begin{array}{c}\mathscr{L} \text{ timelike or} \\ \boldsymbol{k} \text{ not tangent to } \mathscr{L}\end{array} \implies \boxed{\frac{1}{\Delta}\left[(r^2+a^2)E - aL + \epsilon_r\sqrt{R(r)}\right] > 0}. \qquad (11.57)$$

Given expression (11.44) for $R(r)$ and $\Delta := r^2 - 2mr + a^2$ [Eq. (10.10)], we note that this constraint involves (μ, E, L, Q) and r, but not θ.

Another global future-directed null vector field on the Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ is the null tangent $\boldsymbol{\ell}$ to the outgoing principal null geodesics, introduced in Sec. 10.4 [Eq. (10.57)]. If we apply Lemma 2 of Sec. 1.2.2 with $\boldsymbol{\ell}$ instead of $\boldsymbol{k}$, we get, using the Boyer-Lindquist components (10.62) of $\boldsymbol{\ell}$,

$$\begin{array}{c}\mathscr{L} \text{ timelike or} \\ \boldsymbol{\ell} \text{ not tangent to } \mathscr{L}\end{array} \quad \Longrightarrow \quad \boxed{(r^2+a^2)E - aL - \epsilon_r \sqrt{R(r)} > 0}. \tag{11.58}$$

In $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$, $\Delta > 0$ and we can combine (11.57) and (11.58) to $(r^2+a^2)E - aL > |\epsilon_r \sqrt{R(r)}| = \sqrt{R(r)}$, thereby getting a constraint that is independent from ϵ_r:

$$\mathscr{L} \text{ not principal null geod.} \quad \Longrightarrow \quad \boxed{(r^2+a^2)E - aL > \sqrt{R(r)}} \quad \text{in} \quad \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}. \tag{11.59}$$

A related constraint in $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$ is obtained from the vector $\boldsymbol{\varepsilon}_{(0)}$ of the Carter frame introduced in Sec. 10.7.4. As shown there, $\boldsymbol{\varepsilon}_{(0)}$ is a linear combination of $\boldsymbol{k}$ and $\boldsymbol{\ell}$ [cf. Eq. (10.132)] and is a future-directed timelike vector in all $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$. We can then apply Lemma 1 of Sec. 1.2.2, using expression (10.128a) of $\boldsymbol{\varepsilon}_{(0)}$ in terms of $\boldsymbol{\xi} = \boldsymbol{\partial}_t$ and $\boldsymbol{\eta} = \boldsymbol{\partial}_\varphi$:

$$\boldsymbol{p} \text{ future-directed in } \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III} \iff \boldsymbol{g}(\boldsymbol{\varepsilon}_{(0)}, \boldsymbol{p}) < 0 \iff (r^2+a^2) \underbrace{\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{p})}_{-E} + a \underbrace{\boldsymbol{g}(\boldsymbol{\eta}, \boldsymbol{p})}_{L} < 0.$$

Hence the constraint

$$\boxed{(r^2+a^2)E - aL > 0} \quad \text{in } \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}. \tag{11.60}$$

Note that if $\mathscr{L}$ is not a principal null geodesic, (11.60) is a mere consequence of (11.59), given that $\sqrt{R(r)} \geq 0$.

An immediate corollary of (11.60) is:

In the outer ergoregion $\mathscr{G}^+ = \mathscr{G} \cap \mathscr{M}_{\rm I}$, a geodesic that has $E \leq 0$ must have $L < 0$.

Another constraint is obtained by considering the vector field $\boldsymbol{N}$ normal to the hypersurfaces of constant Boyer-Lindquist coordinate t, which has been introduced in Sec. 10.7.3: $\boldsymbol{N} = -\overrightarrow{\mathrm{d}t}$ [Eq. (10.108)]. As shown in Sec. 10.7.3, $\boldsymbol{N}$ is future-directed timelike in the subpart $\mathscr{M}_{\rm ZAMO} := \mathscr{M}_{\rm I} \cup (\mathscr{M}_{\rm III} \setminus \mathscr{T})$ of Kerr spacetime. Lemma 1 of Sec. 1.2.2 [cf. Eq. (1.3a)] then yields

$$\boldsymbol{p} \text{ future-directed in } \mathscr{M}_{\rm ZAMO} \iff \boldsymbol{g}(\boldsymbol{N}, \boldsymbol{p}) < 0 \iff -\underbrace{\langle \mathrm{d}t, \boldsymbol{p} \rangle}_{p^t} < 0 \iff p^t > 0.$$

Hence the Boyer-Lindquist component p^t of the 4-momentum $\boldsymbol{p}$ must be positive in $\mathscr{M}_{\rm ZAMO}$:

$$\boxed{p^t > 0} \quad \text{in } \mathscr{M}_{\rm I} \cup (\mathscr{M}_{\rm III} \setminus \mathscr{T}). \tag{11.61}$$

Let us recall that $\mathscr{T}$ is the Carter time machine (cf. Sec. 10.2.5), which occupies a very limited portion of $\mathscr{M}_{\rm III}$. Since $p^t = {\rm d}t/{\rm d}\lambda$ and λ increases towards the future (cf. Sec. 11.2.1), (11.61) implies

In the region $\mathscr{M}_{\rm I} \cup (\mathscr{M}_{\rm III} \setminus \mathscr{T})$ of Kerr spacetime, the Boyer-Lindquist coordinate t increases towards the future along any geodesic $\mathscr{L}$.

Using Eq. (11.43a) with ${\rm d}t/{\rm d}\lambda = p^t$ and the fact that $\Delta > 0$ in $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$, we can rewrite (11.61) as

$$\boxed{\rho^2(r^2+a^2)E - 2amr(L - aE\sin^2\theta) > 0} \quad \text{in } \mathscr{M}_{\rm I} \cup (\mathscr{M}_{\rm III} \setminus \mathscr{T}). \tag{11.62}$$

Remark 3: In the outer ergoregion $\mathscr{G}^+ = \mathscr{G} \cap \mathscr{M}_{\rm I}$ (i.e. the part of the ergoregion outside of the black hole), $E \leq 0$ is allowed, but the result (11.61) shows that $p^t \leq 0$ is not allowed.

Historical note: That $E \leq 0$ in the outer ergoregion implies $L < 0$ has been shown for equatorial geodesics by George Contopoulos in 1984 [69].

11.3.3 Lense-Thirring effect

Let us consider a null or timelike geodesic $\mathscr{L}$ with a vanishing angular momentum: $L = 0$. The equations of motion (11.43a) and (11.43d) reduce to

$$\rho^2 \frac{{\rm d}t}{{\rm d}\lambda} = E\left(\frac{(r^2+a^2)^2}{\Delta} - a^2\sin^2\theta\right) = \frac{E}{\Delta}\left[\rho^2(r^2+a^2) + 2a^2mr\sin^2\theta\right]$$
$$\rho^2 \frac{{\rm d}\varphi}{{\rm d}\lambda} = \frac{2amrE}{\Delta},$$

so that we get

$$\boxed{\left.\frac{{\rm d}\varphi}{{\rm d}t}\right|_{\mathscr{L}} = \frac{2amr}{\rho^2(r^2+a^2) + 2a^2mr\sin^2\theta}}_{L=0}. \tag{11.63}$$

Hence, for $a \neq 0$, a geodesic with zero angular momentum about the symmetry axis has nevertheless a rotational motion around that axis. This is a manifestation of the ***Lense-Thirring effect***, also called ***dragging of inertial frames***, or, in short, ***frame dragging***.

The Lense-Thirring effect is illustrated in Fig. 11.1 which shows the trajectory a timelike particle with $L = 0$ which is asymptotically at rest (marginally bound particle, having $E = \mu$, to be discussed in Sec. 11.3.7). The trajectory is initialy radial, but as r decreases, ${\rm d}\varphi/{\rm d}t$ increases according to formula (11.63). In Fig. 11.1 and in all figures of this chapter, we are using the ***Cartesian Boyer-Lindquist coordinates*** (t, x, y, z), which are defined

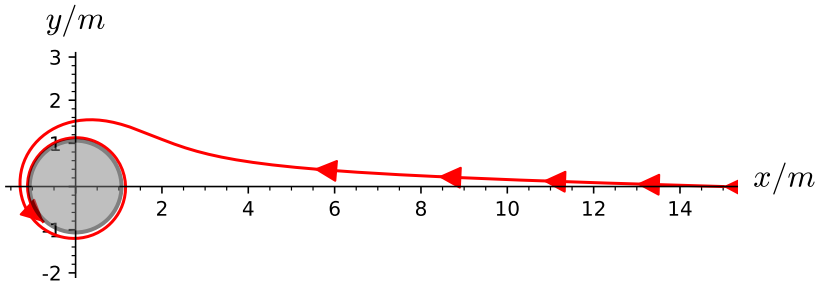


Figure 11.1: Trajectory in the equatorial plane of an incoming timelike geodesic with $E = \mu$, $L = 0$ and $Q = 0$, plunging into a Kerr black hole with $a = 0.998\,m$. The figure is drawn in terms of the Cartesian Boyer-Lindquist coordinates (x, y) defined by Eq. (11.64) with $\theta = \pi/2$ (equatorial plane) and the grey disk marks the black hole region. *[Figure generated by the notebook D.4.8]*

in the $r > 0$ region of Kerr spacetime and are related to the Boyer-Lindquist coordinates (t, r, θ, φ) by the standard transformation from spherical to Cartesian coordinates:

$$x := r \sin\theta \cos\varphi, \qquad y := r \sin\theta \sin\varphi, \qquad z := r \cos\theta. \tag{11.64}$$

We note that, at a given point (r, θ), the angular velocity (11.63) coincides with that of the zero-angular momentum observer (ZAMO) at that point [compare Eq. (10.119)].

11.3.4 Winding near the event horizon and the inner horizon

Let us consider a null or timelike geodesic $\mathscr{L}$ in the vicinity of the black hole event horizon $\mathscr{H}$. On $\mathscr{H}$, $r = r_+$ and $\Delta = 0$. Then, the term Δ^{-1} in Eq. (11.43a) makes $\mathrm{d}t/\mathrm{d}\lambda$ diverge as $r \to r_+$, except in the very special where $(r_+^2 + a^2)E - aL = 0$, which is equivalent to $E = \Omega_H L$ according to Eq. (10.74). Similarly, $\mathrm{d}\varphi/\mathrm{d}\lambda$, as given by Eq. (11.43d), diverges as $r \to r_+$, except for $E = \Omega_H L$. These two divergences are not a pathology of $\mathscr{L}$ per se; they reflect merely the singularity of Boyer-Lindquist coordinates (t, r, θ, φ) at $\mathscr{H}$ (cf. Sec. 10.2.6). However, we read on Eqs. (11.43a) and (11.43d) that the ratio $\mathrm{d}\varphi/\mathrm{d}t|_{\mathscr{L}} := \mathrm{d}\varphi/\mathrm{d}\lambda \times (\mathrm{d}t/\mathrm{d}\lambda)^{-1}$ converges to a finite value:

$$\lim_{r \to r_+} \left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{\mathscr{L}} = \frac{a}{r_+^2 + a^2} = \Omega_H, \tag{11.65}$$

where the second equality follows from Eq. (10.74), letting the black hole rotation velocity Ω_H appear. Hence we conclude that

> Any null or timelike geodesic that approaches the event horizon $\mathscr{H}$ is winding around it in terms of the Boyer-Lindquist coordinates at exactly the black hole rotation velocity Ω_H.

Remark 4: For a timelike geodesic on a circular orbit, we have argued in Sec. 7.3.3 that $\mathrm{d}\varphi/\mathrm{d}t|_{\mathscr{L}}$ is the angular velocity of as seen by an asymptotic inertial observer. The reasoning was given

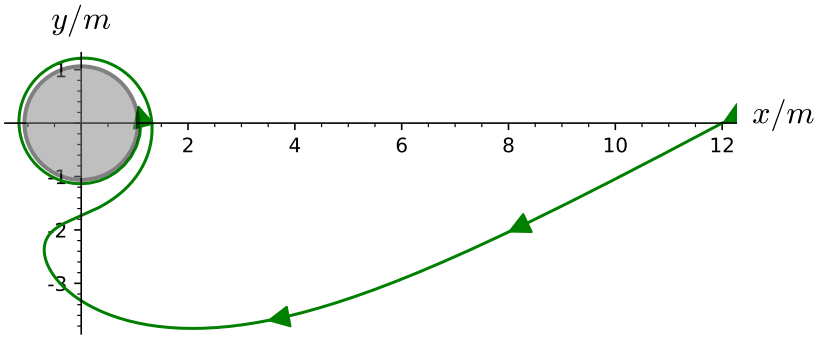


Figure 11.2: Trajectory in the equatorial plane of an incoming null geodesic with $L = -6E < 0$ and $Q = 0$, plunging into a Kerr black hole with $a = 0.998\,m$. Note the turning point in φ and the final winding in the direction of the black rotation (counterclockwise in the figure). The figure is drawn in terms of the Cartesian Boyer-Lindquist coordinates (x, y) defined by Eq. (11.64) with $\theta = \pi/2$ (equatorial plane) and the grey disk marks the black hole region. *[Figure generated by the notebook D.4.8]*

in the Schwarzschild spacetime context but it actually involved only the spacetime symmetry by translation in t, so it is applicable here.

Remark 5: When $a \neq 0$, a geodesic that starts far from the black hole with $\mathrm{d}\varphi/\mathrm{d}t|_{\mathscr{L}} < 0$ [according to Eq. (11.43d) with $r \gg m$, this occurs for $L < 0$] must necessarily have a turning point in φ if it reaches the event horizon, in order to fulfill (11.65), where Ω_H is positive. This is illustrated in Fig. 11.2 and is in sharp contrast with the Schwarzschild case, where φ is always a monotonous function of λ, as shown in Sec. 7.2.2.

Remark 6: The winding property does not hold for the Kerr or null Kerr coordinates. Indeed, we have seen in Sec. 10.4 that the ingoing principal null geodesics $\mathscr{L}^{\mathrm{in}}_{(v,\theta,\tilde{\varphi})}$ are geodesics along which $\tilde{\varphi}$ is constant. They are therefore not winding in terms of neither the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ nor the null Kerr ones $(v, r, \theta, \tilde{\varphi})$. This difference of (coordinate) behavior is understandable if one considers the diverging behavior in Δ^{-1} of the relation $\mathrm{d}\tilde{\varphi} = \mathrm{d}\varphi + a/\Delta\,\mathrm{d}r$ [Eq. (10.23b)] between the angular coordinates φ and $\tilde{\varphi}$.

Regarding the Boyer-Lindquist coordinate behavior in the vicinity of the inner horizon $\mathscr{H}_{\mathrm{in}}$ (corresponding to the second root r_- of Δ), we deduce from Eqs. (11.43a) and (11.43d) that

$$\lim_{r\to r_-} \left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{\mathscr{L}} = \frac{a}{r_-^2 + a^2} = \Omega_{\mathrm{in}}, \tag{11.66}$$

where the second equality follows from Eq. (10.76); it involves the rotation velocity Ω_{in} of the inner horizon $\mathscr{H}_{\mathrm{in}}$. Hence

> Any null or timelike geodesic that approaches the inner horizon $\mathscr{H}_{\mathrm{in}}$ is winding around it in terms of the Boyer-Lindquist coordinates at exactly the rotation velocity Ω_{in} of $\mathscr{H}_{\mathrm{in}}$.

11.3.5 Asymptotic r-values and θ-values

Let $\mathscr{L}$ be a geodesic and p_0 a point of $\mathscr{L}$ at which r varies, i.e. such that $R(r_0) \neq 0$, where $r_0 = r(p_0)$. Let us assume that $\mathscr{L}$ comes close to $r = r_1$, where r_1 is a double root of R, i.e. fulfils $R(r_1) = 0$, $R'(r_1) = 0$ and $R''(r_1) \neq 0$, such that $R(r) > 0$ on the interval $[r_0, r_1)$. Let us reconsider the argument in Sec. 11.2.7 that lead to extend the integral (11.51) to $r = r_1$, where r_1 corresponded to a simple root of R ($R(r_1) = 0$ and $R'(r_1) \neq 0$). For a double root, the integral in the right-hand side of Eq. (11.51) with $r = r_1$ behaves near r_1 as

$$\frac{\sqrt{2}}{\sqrt{R''(r_1)}} \int_{r_0}^{r_1} \frac{\mathrm{d}\bar{r}}{r_1 - \bar{r}},$$

which is a divergent improper integral. If r_1 is a higher order root of R, the divergence is even worse, being triggered by a higher power of $1/(r_1 - \bar{r})$. Consequently, Eq. (11.51) with $r = r_1$ implies that the Mino parameter diverges: $\lambda' \to +\infty$. Since $\rho^2 > 0$, provided that ρ^2 does not tend to 0 (the ring singularity) as $\lambda' \to +\infty$, this implies that the affine parameter λ tends to $+\infty$ as well, given the relation (11.49) between the two parameters. Hence we conclude

Let $\mathscr{L}$ be a geodesic of Kerr spacetime that does not lie at a fixed value of r and let $R(r)$ be the associated polynomial (11.44). A point away from the ring singularity $\mathscr{R}$ and of r-coordinate r_1 such that both $R(r_1) = 0$ and $R'(r_1) = 0$ can only be reached asymptotically by $\mathscr{L}$, i.e. in the limit $\lambda \to \pm\infty$ of the affine parameter λ. We call r_1 an ***asymptotic r-value*** of the geodesic $\mathscr{L}$.

This property explains why $R(r_0) = 0$ and $R'(r_0) \neq 0$ is a necessary condition to have a r-turning point (cf. Remark 4 on p. 322).

By the same reasoning, we have

Let $\mathscr{L}$ be a geodesic of Kerr spacetime that does not lie at a fixed value of θ and let $\Theta(\theta)$ be the associated function (11.32). A point away from the ring singularity $\mathscr{R}$ and of θ-coordinate θ_1 such that both $\Theta(\theta_1) = 0$ and $\Theta'(\theta_1) = 0$ can only be reached asymptotically by $\mathscr{L}$, i.e. in the limit $\lambda \to \pm\infty$ of the affine parameter λ. We call θ_1 an ***asymptotic θ-value*** of the geodesic $\mathscr{L}$.

11.3.6 Latitudinal motion

We start by analysing the variation of the θ coordinate along a geodesic, as governed by the decoupled equation (11.50c), since this provides some constraints to discuss later the r-motion.

Given expression (11.35) for $\Theta(\theta)$, the first-order equation of motion (11.50c) can be rewritten as

$$\boxed{\left(\frac{\mathrm{d}\theta}{\mathrm{d}\lambda'}\right)^2 + V(\theta) = Q}, \tag{11.67}$$

with

$$\boxed{V(\theta) := \cos^2\theta \left[a^2(\mu^2 - E^2) + \frac{L^2}{\sin^2\theta} \right]}. \tag{11.68}$$

Note that $V(\theta)$ is related to $\Theta(\theta)$ by

$$\Theta(\theta) = Q - V(\theta). \tag{11.69}$$

In particular, the θ-turning points (cf. Sec. 11.2.6) are characterized by $V(\theta) = Q$ and $V'(\theta) \neq 0$, while the asymptotic θ-values (cf. Sec. 11.3.5) correspond to $V(\theta) = Q$ and $V'(\theta) = 0$.

We recognize in Eq. (11.67) the first integral of a (non-relativistic) 1-dimensional motion in the potential $V(\theta)$, which we shall call the ***effective θ-potential***. The discussion of the geodesic θ-motion is then based on the properties of that potential, Q in the right-hand side of Eq. (11.67) playing the role of the constant "total energy". Since $(\mathrm{d}\theta/\mathrm{d}\lambda')^2 \geq 0$, Eq. (11.67) implies

$$V(\theta) \leq Q. \tag{11.70}$$

Accordingly, given a plot of $V(\theta)$, as in Figs. 11.3 and 11.4, the allowed range of θ is determined by the part of the graph of $V(\theta)$ that lies below the horizontal line of ordinate equal to Q.

We shall distinguish the case $L = 0$ from the case $L \neq 0$, since they lead to different shapes of the potential $V(\theta)$.

Geodesics with $L = 0$

If $L = 0$, the effective θ-potential reduces to $V(\theta) = a^2(\mu^2 - E^2)\cos^2\theta$. We have then three subcases:

- **Case** $a^2(E^2 - \mu^2) < 0 \iff a \neq 0$ and $|E| < \mu$: the corresponding graph of $V(\theta)$ is shown in Fig. 11.3 (left part); we deduce immediately from it that $Q \geq 0$, with
  - $Q = 0$: the motion is confined to the equatorial plane $\theta = \pi/2$; since it corresponds to a minimum of the effective potential, this is a stable configuration.
  - $0 < Q < a^2(\mu^2 - E^2)$: θ oscillates between two θ-turning points, which are symmetric about the equatorial plane[12]: $\theta_\mathrm{m} := \arccos\sqrt{Q/(a^2(\mu^2 - E^2))}$ and $\pi - \theta_\mathrm{m}$ (cf. the trajectory $Q = Q_1$ in Fig. 11.3, left).
  - $Q = a^2(\mu^2 - E^2)$: $\theta = 0$ and $\theta = \pi$ are either unstable positions or asymptotic θ-values (cf. Sec. 11.3.5), i.e. the geodesic reaches the rotation axis for $\lambda \to \pm\infty$.
  - $Q > a^2(\mu^2 - E^2)$: the range of θ is not limited (cf. the trajectory $Q = Q_2$ in Fig. 11.3, left) and each time the geodesic reaches the rotation axis ($\theta = 0$ or $\theta = \pi$), it crosses it, since the velocity $\mathrm{d}\theta/\mathrm{d}\lambda'$ does not vanish there. This leads to $\theta < 0$ or $\theta > \pi$; to keep θ within the interval $[0, \pi]$, one shall use the identification of the points (θ, φ), $(-\theta, \varphi + \pi)$ and $(\theta - \pi, \varphi + \pi)$, which holds on the sphere $\mathbb{S}^2$.

[12] The index m in θ_m stands for *minimal*.

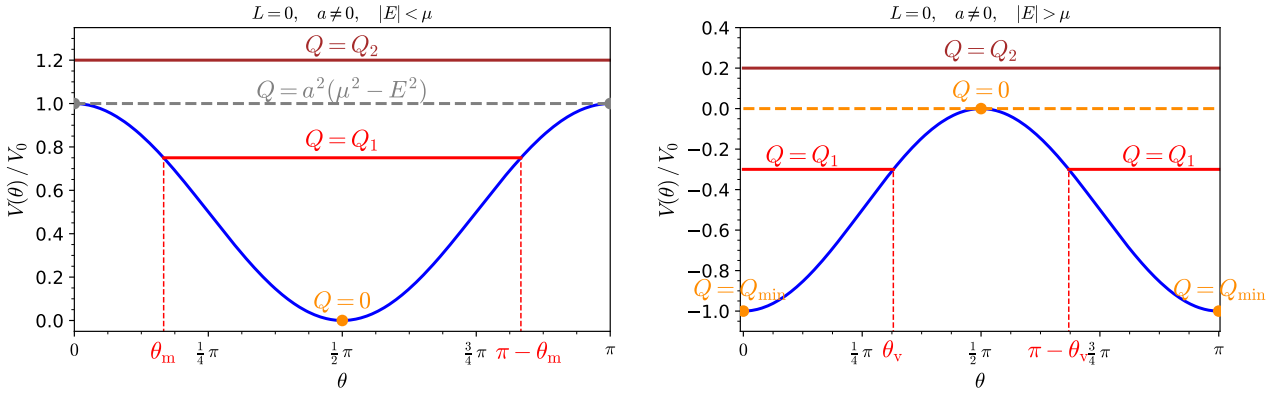


Figure 11.3: Effective θ-potential $V(\theta)$ in the case $L = 0$ and $a \neq 0$. $V(\theta)$ is plotted in units of $V_0 := a^2|\mu^2 - E^2|$. The left figure is for $|E| < \mu$, with colored dots or horizontal lines corresponding to geodesic trajectories for four values of the Carter constant Q: 0, $Q_1 = 0.75V_0$, $a^2(\mu^2-E^2)$ and $Q_2 = 1.2V_0$. The right figure is for $|E| > \mu$, with trajectories corresponding to four values of Q: $Q_{\rm min} = -a^2(E^2-\mu^2)$, $Q_1 = -0.3V_0$, 0 and $Q_2 = 0.2V_0$. Dashed lines indicate trajectories with asympotic θ-values.

- **Case** $a^2(E^2-\mu^2) = 0 \iff a = 0$ or $|E| = \mu$: $V(\theta) = 0$ and Eq.(11.67) reduces to $(\mathrm{d}\theta/\mathrm{d}\lambda')^2 = Q$. This implies $Q \geq 0$ and the solution is $\theta(\lambda') = \pm\sqrt{Q}\,\lambda' + \theta_0$, so that
  - for $Q = 0$, the geodesic lies at a constant value of θ, within the range $[0, \pi]$;
  - for $Q > 0$, θ varies monotonically along the geodesic, which therefore crosses the rotation axis an infinite number of times.
- **Case** $a^2(E^2-\mu^2) > 0 \iff a \neq 0$ and $|E| > \mu$: the corresponding graph of $V(\theta)$ is shown in Fig. 11.3 (right part); the Carter constant must obey $Q \geq -a^2(E^2-\mu^2)$, with
  - $Q = -a^2(E^2-\mu^2)$: only $\theta = 0$ and $\theta = \pi$ are possible and they correspond to minima of $V(\theta)$; the geodesic is then stably located on the rotation axis.
  - $-a^2(E^2-\mu^2) < Q < 0$: the geodesic oscillates about the rotation axis, without reaching the equator; one has either $\theta \in [0, \theta_{\rm v}]$ or $\theta \in [\pi-\theta_{\rm v}, \pi]$, with the turning point value[13] $\theta_{\rm v} := \arccos\sqrt{|Q|/(a^2(E^2-\mu^2))} < \pi/2$ (cf. the trajectories $Q = Q_1$ in Fig. 11.3, right).
  - $Q = 0$: $\theta = \pi/2$ is either an unstable position or an asymptotic θ-value, i.e. the geodesic approaches the equatorial plane when $\lambda \to \pm\infty$.
  - $Q > 0$: θ varies in all the range $[0, \pi]$; when the geodesic reaches the rotation axis, it crosses it (cf. the trajectory $Q = Q_2$ in Fig. 11.3, right).

[13] The index v in $\theta_{\rm v}$ stands for *vortical*, the definition of which is given at the end of this section.

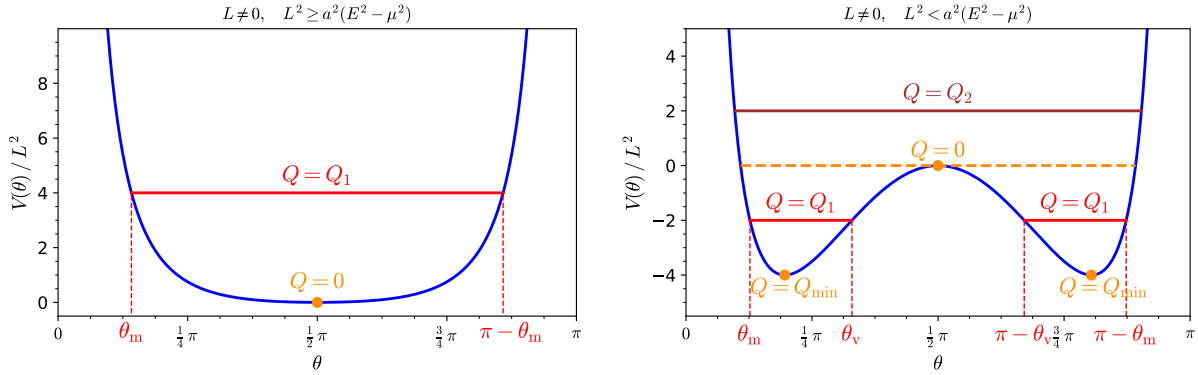


Figure 11.4: Effective θ-potential $V(\theta)$ in the case $L \neq 0$. The left figure is for $L^2 \geq a^2(E^2-\mu^2)$, with with colored dots or horizontal lines corresponding to geodesic trajectories for two values of the Carter constant Q: 0 and $Q_1 = 4L^2$. The right figure is for $L^2 < a^2(E^2-\mu^2)$, with trajectories corresponding to four values of Q: $Q_{\rm min}$, $Q_1 = -2L^2$, 0 and $Q_2 = 2L^2$. The dashed line ($Q = 0$) indicates the trajectory with an asympotic θ-value, which is $\pi/2$.

Geodesics with $L \neq 0$

When $L \neq 0$, we see from expression (11.68) that $V(\theta) \to +\infty$ when $\theta \to 0$ or π. We conclude immediately that the geodesic cannot reach the rotation axis for $L \neq 0$. Moreover, we have

$$\frac{\mathrm{d}V}{\mathrm{d}\theta} = -2\cos\theta\sin\theta\left[a^2(\mu^2 - E^2) + \frac{L^2}{\sin^4\theta}\right].$$

The extrema of V in $(0,\pi)$ are then given by

$$\frac{\mathrm{d}V}{\mathrm{d}\theta} = 0 \iff \theta = \frac{\pi}{2} \quad \text{or} \quad \sin^4\theta = \frac{L^2}{a^2(E^2-\mu^2)}. \tag{11.71}$$

We are thus led to distinguish two cases, depending whether or not the equation involving $\sin^4\theta$ has solutions distinct from $\pi/2$:

- **Case** $L^2 \geq a^2(E^2-\mu^2) \iff a = 0$ or $|E| \leq \sqrt{\mu^2 + \frac{L^2}{a^2}}$: the only extremum solution is $\theta = \pi/2$. Its value is $V(\pi/2) = 0$ and it is necessarily a minimum, since $V(\theta) \to +\infty$ at the boundaries of the interval $[0,\pi]$ (cf. left part of Fig. 11.4). Hence one must have $Q \geq 0$, with
  - $Q = 0$: the geodesic stays stably in the equatorial plane.
  - $Q > 0$: the geodesic oscillates about the equatorial plane, between two turning points at $\theta = \theta_{\rm m}$ and $\pi - \theta_{\rm m}$, $\theta_{\rm m}$ being the solution of $V(\theta_{\rm m}) = Q$ in $(0,\pi/2)$ (cf. the trajectory $Q = Q_1$ in Fig. 11.4, left).
- **Case** $L^2 < a^2(E^2-\mu^2) \iff a \neq 0$ and $|E| > \sqrt{\mu^2 + \frac{L^2}{a^2}}$: $V(\theta)$ has three extrema

(cf. right part of Fig. 11.4), which are located at $\theta = \theta_*$, $\pi/2$ and $\pi - \theta_*$ with

$$\theta_* := \arcsin \sqrt{\frac{|L|}{a\sqrt{E^2 - \mu^2}}} \quad \text{and} \quad 0 < \theta_* < \frac{\pi}{2}. \tag{11.72}$$

θ_* and $\pi - \theta_*$ correspond to the minimum of $V(\theta)$, while $\pi/2$ corresponds to a local maximum. The minimum is $V(\theta_*) = (1 - \sin^2\theta_*)(a^2(\mu^2 - E^2) + L^2/\sin^2\theta_*) = -(a\sqrt{E^2 - \mu^2} - |L|)^2$. This is necessary the minimal value of Q [cf. Eq. (11.70)]:

$$Q_{\rm min} = -\left(a\sqrt{E^2 - \mu^2} - |L|\right)^2. \tag{11.73}$$

We have then (cf. right panel of Fig. 11.4)

- $Q = Q_{\rm min}$: the geodesic stays stably at a fixed value of θ, either θ_* or $\pi - \theta_*$.
- $Q_{\rm min} < Q < 0$: the geodesic oscillates between two θ-turning points either in the Northern hemisphere ($\theta < \pi/2$) or the Southern one ($\theta > \pi/2$), without reaching the equator nor the rotation axis (cf. the trajectories $Q = Q_1$ in Fig. 11.4, right).
- $Q = 0$: the geodesic lies unstably in the equatorial plane or moves asymptotically towards it (possibly after a turning point at $\theta_{\rm m} \neq \pi/2$), $\pi/2$ being an asymptotic θ-value, since $\Theta'(\pi/2) = -V'(\pi/2) = 0$ (cf. Sec. 11.3.5).
- $Q > 0$: the geodesic oscillates about the equatorial plane, between two turning points at $\theta = \theta_{\rm m}$ and $\pi - \theta_{\rm m}$, $\theta_{\rm m}$ being the solution of $V(\theta_{\rm m}) = Q$ in $(0, \pi/2)$ (cf. the trajectory $Q = Q_2$ in Fig. 11.4, right).

Expression of the θ-turning points

The θ-turning points $\theta_{\rm m}$ and $\theta_{\rm v}$ mentioned above are solutions of $\Theta(\theta) = 0$ and $\Theta'(\theta) \neq 0$ [Eq. (11.47)]. We search only for θ-turning points with $0 < \theta < \pi/2$, since 0 and $\pi/2$ cannot be θ-turning points (for $\Theta'(0) = 0$ and $\Theta'(\pi/2) = 0$) and θ-turnings points with $\pi/2 < \theta < \pi$ are deduced from those with $0 < \theta < \pi/2$ by $\theta \mapsto \pi - \theta$ (symmetry with respect to the equatorial plane). Using expression (11.35) for Θ and introducing

$$x := \cos^2\theta, \tag{11.74}$$

the equation $\Theta(\theta) = 0$ is equivalent to

$$a^2(E^2 - \mu^2)x^2 + \left[L^2 + Q - a^2(E^2 - \mu^2)\right] x - Q = 0. \tag{11.75}$$

If $a^2(E^2 - \mu^2) = 0$, the solution is $x = Q/(L^2 + Q)$. Moreover, in this case one has necessarily $Q \geq 0$, so that $0 \leq x \leq 1$ and the unique solution in $(0, \pi/2)$ is

$$\boxed{\theta_{\rm m} = \arccos \sqrt{\frac{Q}{L^2 + Q}}}_{a^2(E^2-\mu^2)=0}. \tag{11.76}$$

In the rest of this section, we assume that $a^2(E^2-\mu^2)\neq 0$. Equation (11.75) is then a quadratic equation in x, the discriminant of which is

$$\begin{aligned}\Delta &= \left[L^2+Q-a^2(E^2-\mu^2)\right]^2+4Qa^2(E^2-\mu^2)\\ &= \left[L^2+Q+a^2(E^2-\mu^2)\right]^2-4L^2a^2(E^2-\mu^2).\end{aligned}$$

It turns out that Δ is always non-negative. Indeed, from the second equality above, we have clearly $\Delta\geq 0$ if $a^2(E^2-\mu^2)\leq 0$. In the complementary case, i.e. when $a^2(E^2-\mu^2)>0$, then $\Delta\geq 0$ as soon as $Q\geq Q_2$, where Q_2 is the larger of the two roots of Δ considered as a quadratic polynomial in Q. Again, considering the second equality in the expression of Δ, we have $Q_2=-L^2-a^2(E^2-\mu^2)+2|L|a\sqrt{E^2-\mu^2}=-(|L|-a\sqrt{E^2-\mu^2})^2$. In view of Eq. (11.73), we realize that $Q_2=Q_{\rm min}$, so that $Q\geq Q_2$ always holds and $\Delta\geq 0$. Accordingly the two roots of Eq. (11.75) are real and are given by

$$x_\pm=\frac{1}{2}\left(1-\frac{L^2+Q}{a^2(E^2-\mu^2)}\pm\sqrt{\left[1-\frac{L^2+Q}{a^2(E^2-\mu^2)}\right]^2+\frac{4Q}{a^2(E^2-\mu^2)}}\right). \tag{11.77}$$

Since $x:=\cos^2\theta$ and $\theta=0$ and $\theta=\pi/2$ cannot correspond to a θ-turning point, acceptable solutions must obey

$$0<x_\pm<1. \tag{11.78}$$

For $L=0$, Eq. (11.77) simplifies to[14]

$$x_+=1 \qquad \text{and} \qquad x_-=-\frac{Q}{a^2(E^2-\mu^2)}$$

Now, $x_+=1$ is excluded by (11.78). There remains only x_-, leading to the turning point value found in the $L=0$ cases discussed above, which we can combine into a single formula:

$$\boxed{\theta_{\rm m}=\theta_{\rm v}=\arccos\sqrt{\frac{Q}{a^2(\mu^2-E^2)}}}_{L=0}. \tag{11.79}$$

This formula assumes that $0<Q/(a^2(\mu^2-E^2))<1$ (as for the trajectories $Q=Q_1$ in both panels of Fig. 11.3), otherwise there is no θ-turning point (as for the trajectories $Q=Q_2$ in both panels of Fig. 11.3). More precisely, for $Q>0$ and $\mu^2-E^2>0$, it leads to $\theta_{\rm m}$, while for $Q<0$ and $\mu^2-E^2<0$, it leads to $\theta_{\rm v}$.

For $L\neq 0$, one shall distinguish between two cases, since the case $|E|=\mu$ has been covered above [Eq. (11.76)]:

- if $|E|<\mu$, then only x_- fulfils the criterion (11.78), leading to the turning point value:

$$\boxed{\theta_{\rm m}=\arccos\sqrt{\frac{1}{2}\left[1+\frac{L^2+Q}{a^2(\mu^2-E^2)}-\sqrt{\left[1+\frac{L^2+Q}{a^2(\mu^2-E^2)}\right]^2-\frac{4Q}{a^2(\mu^2-E^2)}}\right]}} \tag{11.80}$$

[14] A switch between x_+ and x_- is performed if $1+Q/(a^2(E^2-\mu^2))<0$.

- if $|E| > \mu$, then x_+ always fulfils the criterion (11.78), leading to the turning point value:

$$\theta_{\rm m} = \arccos \sqrt{\frac{1}{2}\left[1 - \frac{L^2+Q}{a^2(E^2-\mu^2)} + \sqrt{\left[1 - \frac{L^2+Q}{a^2(E^2-\mu^2)}\right]^2 + \frac{4Q}{a^2(E^2-\mu^2)}}\right]}. \tag{11.81}$$

  If, in addition $Q < 0$, then x_- also fulfils (11.78) and we get a second turning point value (cf. the case $Q = Q_1$ in the right panel of Fig. 11.4):

$$\theta_{\rm v} = \arccos \sqrt{\frac{1}{2}\left[1 - \frac{L^2+Q}{a^2(E^2-\mu^2)} - \sqrt{\left[1 - \frac{L^2+Q}{a^2(E^2-\mu^2)}\right]^2 + \frac{4Q}{a^2(E^2-\mu^2)}}\right]}. \tag{11.82}$$

Remark 7: At the limit $L \to 0$, formulas (11.80) and (11.82) reduce both to (11.79).

Summary

We can summarize the above results by

- A geodesic $\mathscr{L}$ of Kerr spacetime cannot encounter the rotation axis unless it has $L = 0$.
- If $L^2 \geq a^2(E^2 - \mu^2)$, the Carter constant Q is necessarily non-negative:

$$Q \geq 0. \tag{11.83}$$

- The Carter constant Q can take negative values only if $L^2 < a^2(E^2 - \mu^2)$, which implies $a \neq 0$ and $|E| > \mu$; the range of Q is then limited from below:

$$Q \geq Q_{\rm min} = -\left(a\sqrt{E^2-\mu^2} - |L|\right)^2. \tag{11.84}$$

  A geodesic with $Q < 0$ is called ***vortical***; it never encounters the equatorial plane.

- If $Q > 0$, $\mathscr{L}$ oscillates symmetrically about the equatorial plane, between two θ-turning points, at $\theta = \theta_{\rm m}$ and $\theta = \pi - \theta_{\rm m}$, where $\theta_{\rm m} \in (0, \pi/2)$ is given by Eq. (11.76) for $a^2(E^2 - \mu^2) = 0$, Eq. (11.79) for $L = 0$ and $Q < a^2(\mu^2 - E^2)$, Eq. (11.80) for $|E| < \mu$ and Eq. (11.81) for $|E| > \mu$, except for two subcases with $L = 0$: (i) $Q = a^2(\mu^2 - E^2)$: $\mathscr{L}$ lies unstably along the rotation axis or approaches it asymptotically and (ii) $Q > a^2(\mu^2 - E^2)$: $\mathscr{L}$ crosses repeatedly the rotation axis, with θ taking all values in the range $[0, \pi]$.

- If $Q = 0$, $\mathscr{L}$ is stably confined to the equatorial plane for $L^2 > a^2(E^2 - \mu^2)$ or $L^2 = a^2(E^2 - \mu^2) \neq 0$; for $L^2 < a^2(E^2 - \mu^2)$, $\mathscr{L}$ either lies unstably in the equatorial plane or approaches it asymptotically from one side, while for $L^2 = a^2(E^2 - \mu^2) = 0$, $\mathscr{L}$ lies at a constant value $\theta = \theta_0 \in [0, \pi]$.

- If $Q_{\rm min} < Q < 0$, $\mathscr{L}$ never encounters the equatorial plane, having a θ-motion entirely confined either to the Northern hemisphere ($0 < \theta < \pi/2$) or to the Southern one ($\pi/2 < \theta < \pi$); if $L \neq 0$, $\mathscr{L}$ oscillates between two θ-turning points, at $\theta = \theta_{\rm m}$ and $\theta = \theta_{\rm v}$ (Northern hemisphere) or at $\theta = \pi - \theta_{\rm v}$ and $\theta = \pi - \theta_{\rm m}$ (Southern hemisphere), where $\theta_{\rm m}$ and $\theta_{\rm v}$ are given by Eqs. (11.81) and (11.82) respectively; if $L = 0$, $\mathscr{L}$ oscillates about the rotation axis, with a θ-turning point at $\theta = \theta_{\rm v}$ or $\theta = \pi - \theta_{\rm v}$, where $\theta_{\rm v}$ is given by Eq. (11.79).

- If $Q = Q_{\rm min}$, $\mathscr{L}$ lies stably at a constant value $\theta = \theta_*$ or $\theta = \pi - \theta_*$, with $\theta_* \in [0, \pi/2)$ given by[a]

$$\theta_* := \arcsin \sqrt{\frac{|L|}{a\sqrt{E^2 - \mu^2}}}. \tag{11.85}$$

[a] This is Eq. (11.72) generalized to encompass the case $L = 0$.

Remark 8: We could have deduced that $L = 0$ is a necessary condition for a geodesic to encounter the rotation axis without studying the potential $V(\theta)$. Indeed, by definition, $L = \boldsymbol{\eta} \cdot \boldsymbol{p}$ [Eq. (11.2b)], with the Killing vector $\boldsymbol{\eta}$ being zero on the rotation axis, since the latter is pointwise invariant under the action of the rotation group SO(2). Hence $L = 0$ on the rotation axis. Since L is constant along the geodesic, the result follows immediately.

Remark 9: That a vortical geodesic never intersects the equatorial plane has been obtained by examining the various cases $Q < 0$. This result can be derived directly from the constraint $\Theta(\theta) \geq 0$ [Eq. (11.33)], by noticing the identity $\Theta(\pi/2) = Q$, which follows immediately from Eq. (11.35).

Example 7 (Schwarzschild geodesics): It is instructive to apply the above results to $a = 0$ and recover the geodesics in Schwarzschild spacetime studied in Chaps. 7 and 8. There, spherical symmetry was used to select the coordinates (t, r, θ, φ) so that the geodesic was confined to the hyperplane $\theta = \pi/2$. Consequently there was no θ-motion. Here, we keep the coordinates (t, r, θ, φ) fixed and do not assume that they are adapted to the geodesic $\mathscr{L}$ under consideration. So θ may vary along $\mathscr{L}$. In Schwarzschild spacetime, the Carter constant Q is always non-negative, since the inequality $L^2 \geq a^2(E^2 - \mu^2)$ is trivially satisfied for $a = 0$ [cf. Eq. (11.83)].
If $Q = 0$, then for $L = 0$, $\mathscr{L}$ lies at a constant value of θ: this actually corresponds to a purely radial geodesic. Indeed, the total angular momentum measured in the asymptotic region is $\boldsymbol{L}_{\rm tot} = 0$ in that case [set $Q = 0$, $L = 0$ and $a = 0$ in Eq. (11.37)]. If $L \neq 0$, still with $Q = 0$, $\mathscr{L}$ lies stably in the equatorial plane (this is the only possibility for $a = 0$ among the $Q = 0$ cases listed above).
If $Q > 0$, then $L = 0$ is necessarily the subcase (ii) listed above: $\mathscr{L}$ crosses repeatedly the z-axis ($\theta \in \{0, \pi\}$); this corresponds to the case where the orbital plane contains the z-axis. All the angular momentum measured asymptotically is then contained in Q [cf. Eq. (11.37)]. For

$Q > 0$ and $L \neq 0$, $\mathscr{L}$ oscillates symmetrically about the equatorial plane: this is the case where the orbital plane is inclined by an angle $\iota \in (0, \pi/2)$ with respect to the equatorial plane; the θ-turning point of $\mathscr{L}$ is then $\theta_{\rm m} = \pi/2 - \iota$.

11.3.7 Radial motion

The r-part of geodesic motion is constrained by Eq. (11.30): $R(r) \geq 0$. Let us expand expression (11.44) for the quartic polynomial $R(r)$ in powers of r:

$$\boxed{R(r) = (E^2 - \mu^2) r^4 + 2m\mu^2 r^3 + \left[a^2(E^2 - \mu^2) - Q - L^2\right] r^2 + 2m\left[Q + (L - aE)^2\right] r - a^2 Q}. \tag{11.86}$$

Geodesics with $|E| < \mu$

For $r \to \pm\infty$ and $|E| \neq \mu$, we have $R(r) \sim (E^2 - \mu^2) r^4$ and in particular $\lim_{r\to\pm\infty} R(r) = +\infty$ for $|E| > \mu$ and $\lim_{r\to\pm\infty} R(r) = -\infty$ for $|E| < \mu$. In the latter case, the constraint $R(r) \geq 0$ cannot be satisfied for large values of $|r|$. A geodesic with $|E| < \mu$ is therefore located within a bounded region in terms of the radial coordinate r. Moreover, $|E| < \mu$ implies necessarily $L^2 > a^2(E^2 - \mu^2)$, so that the results of Sec. 11.3.6 lead to $Q \geq 0$ [cf. Eq. (11.83)]. Therefore, $-a^2 Q \leq 0$. It follows then that r cannot be negative. Indeed, we see on expression (11.86) that for $|E| < \mu$ (which implies $\mu > 0$ and $Q \geq 0$) and $r < 0$,

$$R(r) = \underbrace{(E^2 - \mu^2) r^4}_{<0} + \underbrace{2m\mu^2 r^3}_{<0} + \underbrace{\left[a^2(E^2 - \mu^2) - Q - L^2\right] r^2}_{\leq 0} + \underbrace{2m\left[Q + (L - aE)^2\right] r}_{\leq 0} \underbrace{- a^2 Q}_{\leq 0},$$

which contradicts $R(r) \geq 0$. Hence we conclude:

> Any geodesic $\mathscr{L}$ with $|E| < \mu$ is necessarily timelike ($\mu > 0$), has a non-negative Carter constant Q, is confined to the region $r \geq 0$ of Kerr spacetime and cannot reach arbitrary large values of r. $\mathscr{L}$ is called a ***geodesic with a bound orbit***, or in short, a ***bound geodesic***.

Remark 10: That $\mathscr{L}$ cannot reach the asymptotic region $r \to +\infty$ could have been found directly from the definition of E as the "energy at infinity" of the particle $\mathscr{P}$ having $\mathscr{L}$ as worldline [Eq. (11.2a)]. Indeed, since $\mathscr{L}$ is timelike, the 4-momentum of $\mathscr{P}$ is $\boldsymbol{p} = \mu \boldsymbol{u}$ [Eq. (1.16)], where $\boldsymbol{u}$ is the 4-velocity of $\mathscr{P}$, so that Eq. (11.2a) yields $E = -\mu\, \boldsymbol{\xi} \cdot \boldsymbol{u}$. Now, when $r \to +\infty$, $\boldsymbol{\xi}$ tends to the 4-velocity of an inertial observer (cf. Sec. 10.7.5). If $\mathscr{L}$ could reach the asymptotic region, the scalar product of the two 4-velocities $\boldsymbol{\xi}$ and $\boldsymbol{u}$ would be necessarily lower or equal to -1, or equivalently $\Gamma := -\boldsymbol{\xi} \cdot \boldsymbol{u} \geq 1$ (the proof lies in expression (1.33) of Γ with $0 \leq \boldsymbol{V} \cdot \boldsymbol{V} < 1$), so that we would have $E \geq \mu$, which contradicts $|E| < \mu$.

Geodesics with $|E| = \mu$

Contrary to the case $|E| < \mu$, a geodesic with $|E| = \mu$ can be null, provided it has $E = 0$ (cf. Example 1 on p. 313). We have

Any geodesic with $|E| = \mu$ has $Q \geq 0$ and is necessarily confined to the region $r \geq 0$ of Kerr spacetime.

Proof. $Q \geq 0$ follows directly from the criteria $|E| \leq \sqrt{\mu^2 + L^2/a^2}$ [cf. Eq. (11.83)], which is evidently fulfilled for $|E| = \mu$. Furthermore, for $|E| = \mu$, expression (11.86) for $R(r)$ simplifies to

$$R(r) = 2m\mu^2 r^3 - (Q + L^2)r^2 + 2m\left[Q + (L - aE)^2\right] r - a^2 Q.$$

For $r < 0$, all the four terms in the above sum are ≤ 0. If $\mathscr{L}$ is timelike, then $\mu \neq 0$ and the first term is < 0, so that $r < 0 \Longrightarrow R(r) < 0$, which violates the constraint $R(r) \geq 0$. Let us now assume that $\mathscr{L}$ is null. We have then $\mu = 0$ and $E = 0$, so that

$$R(r) = -(Q + L^2)r^2 + 2m(Q + L^2)r - a^2 Q = (Q + L^2)r(2m - r) - a^2 Q.$$

If $Q + L^2 \neq 0$, $R(r)$ is a quadratic polynomial that is either negative everywhere or negative outside the interval $[r_1, r_2]$, when the two roots r_1 and r_2 of $R(r)$ are reals. In the latter case, we deduce from the signs of the coefficients of $R(r)$ that $r_1 + r_2 > 0$ and $r_1 r_2 \geq 0$, so that $r_1 \geq 0$ and $r_2 \geq 0$. This implies that $R(r) < 0$ for $r < 0$, which is not permitted. If $Q + L^2 = 0$, the property $Q \geq 0$ implies $Q = 0$ and $L = 0$. The Carter constant $\mathscr{K}$ is then zero as well, since $\mathscr{K} = Q + (L - aE)^2$ [Eq. (11.34)]. Then by the result (11.25), $\mathscr{L}$ is a null geodesic generator of either the event horizon $\mathscr{H}$, which is located at $r = r_+$, or the inner horizon $\mathscr{H}_{\rm in}$, which is located at $r = r_-$. Since both r_+ and r_- are positive [Eq. (10.5)], we cannot have $r < 0$ in this case either. □

A timelike geodesic $\mathscr{L}$ with $E = \mu$ can reach the asymptotic region $r \to +\infty$. It has then a Lorentz factor with respect to the asymptotic static observer of 4-velocity $\boldsymbol{\xi}$ equal to one (cf. Remark 10 above), which implies that $\boldsymbol{p}$ is collinear to $\boldsymbol{\xi}$. In that sense, $\mathscr{L}$ is asymptotically "at rest". Such a geodesic is called ***marginally bound***.

On the opposite, a null geodesic with $E = \mu (= 0)$ cannot reach the asymptotic region $r \to +\infty$. Actually, it cannot even exist outside the ergoregion, by virtue of the property (11.56).

Example 8: The null geodesics generating the horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$ considered in Example 1 on p. 313 have $E = 0$ [Eq. (11.4)] and are indeed fully located in the ergoregion $\mathscr{G}$, since $\mathscr{H} \subset \mathscr{G}$ and $\mathscr{H}_{\rm in} \subset \mathscr{G}$ (cf. Fig. 10.2).

Geodesics with $|E| > \mu$

An immediate corollary of the properties obtained for $|E| < \mu$ and $|E| = \mu$ is

Only geodesics with $|E| > \mu$ may have some part in the region $r < 0$ of Kerr spacetime, $\mathscr{M}_-$:

$$\mathscr{L} \cap \mathscr{M}_- \neq \varnothing \Longrightarrow |E| > \mu. \tag{11.87}$$

Decay of r towards the future in $\mathscr{M}_{\rm II}$

As a particular case of the result established in Sec. 10.3.2 for any causal worldline (not necessarily a geodesic), we have

In region $\mathscr{M}_{\rm II}$, the coordinate r must decrease towards the future along any timelike or null geodesic:

$$\boxed{\frac{\mathrm{d}r}{\mathrm{d}\lambda} < 0}_{\mathscr{M}_{\rm II}} . \tag{11.88}$$

Example 9 (Principal null geodesics): For the ingoing principal null geodesics $\mathscr{L}^{\rm in}_{(v,\theta,\tilde{\varphi})}$, a future-directed tangent vector is $\boldsymbol{k}$ and we have $k^r = \mathrm{d}r/\mathrm{d}\lambda = -1$, since the affine parameter associated with $\boldsymbol{k}$ is $\lambda = -r$ [Eq. (10.46)]. Hence for these geodesics, r is decreasing towards the future everywhere in $\mathscr{M}$, and in particular in $\mathscr{M}_{\rm II}$. For the outgoing principal null geodesics $\mathscr{L}^{\rm out}_{(u,\theta,\tilde{\varphi})}$, a future-directed tangent vector is $\boldsymbol{\ell}$ and the associated (non-affine) parameter obeys Eq. (10.56):

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda} = \frac{\Delta}{2(r^2+a^2)}.$$

Thus $\mathrm{d}r/\mathrm{d}\lambda < 0$ in $\mathscr{M}_{\rm II}$, since $\Delta < 0$ there.

11.3.8 Geodesics reaching or emanating from the ring singularity

In the Schwarzschild spacetime, any causal geodesic that enters into the black hole region inevitably terminates at the curvature singularity $r = 0$, as it is clear on the Kruskal or Carter-Penrose diagrams constructed in Chap. 9. For the Kerr spacetime with $a \neq 0$, we are going to see that, on the contrary, most causal geodesics in the black hole region *avoid* the curvature singularity. In all this section, we assume $a \neq 0$, so that the curvature singularity is the ring singularity $\mathscr{R}$ discussed in Sec. 10.2.6.

Formally, $\mathscr{R}$ is not part of the Kerr spacetime $\mathscr{M}$ but of the larger manifold $\mathbb{R}^2 \times \mathbb{S}^2$ [cf. the construction (10.28)]. It is located at $\rho^2 = 0$, i.e. at $r = 0$ and $\theta = \pi/2$. We shall say that a geodesic $\mathscr{L}$ of affine parameter λ (oriented towards the future) ***approaches the ring singularity*** if $\mathscr{L}$ has both $r(\lambda) \to 0$ and $\theta(\lambda) \to \pi/2$ as $\lambda \to \lambda_*$, with λ_* being finite or infinite. If λ_* is finite, we shall say that $\mathscr{L}$ ***hits the ring singularity*** for $\lambda \to \lambda_*^-$ and ***emanates from the ring singularity*** for $\lambda \to \lambda_*^+$. If $\lambda_* = \pm\infty$, we shall say that $\mathscr{L}$ ***asymptotically approaches the ring singularity***.

A first key result is

A null or timelike geodesic $\mathscr{L}$ that approaches the ring singularity has a vanishing Carter constant Q.

Proof. From Eqs. (11.86) and (11.35), we have

$$\lim_{r\to 0} R(r) = -a^2 Q \quad \text{and} \quad \lim_{\theta\to\pi/2} \Theta(\theta) = Q.$$

The constraints $R(r) \geq 0$ [Eq. (11.30)] and $\Theta(\theta) \geq 0$ [Eq. (11.33)] imply then respectively $Q \leq 0$ and $Q \geq 0$, from which we get $Q = 0$. □

The above property can be refined[15]:

A null or timelike geodesic cannot approach the ring singularity unless it lies entirely in the equatorial plane.

Proof. Let $\mathscr{L}$ be a causal geodesic that approaches $\mathscr{R}$. From the previous result, $\mathscr{L}$ has $Q = 0$. Reviewing the subcases $Q = 0$ among all the cases considered in Sec. 11.3.6, we see that $\mathscr{L}$ can approach $\theta = \pi/2$ iff (i) $\mathscr{L}$ has $|E| \leq \sqrt{\mu^2 + L^2/a^2}$ and lies stably in the equatorial plane or (ii) $\mathscr{L}$ has $|E| > \sqrt{\mu^2 + L^2/a^2}$ and approaches asymptotically $\theta = \pi/2$ as the Mino parameter λ' tends to $\pm\infty$. Let us show that (ii) is not compatible with $r \to 0$. For $Q = 0$, expression (11.86) for $R(r)$ reduces to

$$R(r) = r\left[(E^2 - \mu^2)r^3 + 2m\mu^2 r^2 + \left(a^2(E^2 - \mu^2) - L^2\right) r + 2m(L - aE)^2\right],$$

with the constant term inside the square brackets $2m(L - aE)^2 \neq 0$, since $L = aE$ is not compatible with $|E| > \sqrt{\mu^2 + L^2/a^2}$. It follows that $r = 0$ is a simple root of $R(r)$. If $\mathscr{L}$ would reach both $r = 0$ and $\theta = \pi/2$ when λ tends to some value λ_*, then Eq. (11.55a) would yield

$$\lambda'_* - \lambda'_0 = \int_{r_0}^{0} \frac{\epsilon_r \, \mathrm{d}r}{\sqrt{R(r)}} = \int_{\theta_0}^{\frac{\pi}{2}} \frac{\epsilon_\theta \, \mathrm{d}\theta}{\sqrt{\Theta(\theta)}}, \tag{11.89}$$

where λ'_* is the Mino parameter corresponding to λ_*. Note that λ'_* may be infinite even if λ_* is finite, due to the relation (11.49) with $\rho^2 \to 0$ when $\lambda \to \lambda_*$. Since $r = 0$ is a simple root of $R(r)$, the integral on r has a finite value. Now, for $Q = 0$, expression (11.35) for Θ reduces to

$$\Theta(\theta) = \cos^2\theta \left[a^2(E^2 - \mu^2) - \frac{L^2}{\sin^2\theta}\right].$$

The $\cos^2\theta$ term, which behaves as $(\theta - \pi/2)^2$ for θ near $\pi/2$, makes the integral on θ in Eq. (11.89) divergent, which is incompatible with the finite value of the integral on r in the left-hand side. Hence only (i) is possible, which completes the proof. □

The above result has been obtained for a generic approach to the ring singularity, i.e. for a geodesic $\mathscr{L}$ that hits $\mathscr{R}$, emanates from $\mathscr{R}$ or asymptotically approach $\mathscr{L}$ in the future or the past. If we apply it to null geodesics (light rays) emanating from $\mathscr{R}$, we conclude that an intrepid observer diving into the black hole would not see the ring singularity at all, except when he crosses the equatorial plane. At this instant, the singularity would appear to him as a 1-dimensional segment, and the image would disappear as soon as the observer leaves the equatorial plane. In particular, the observer will never see a ring-like image.

[15] This is a refinement because not all geodesics with $Q = 0$ lie in the equatorial plane. For instance the null geodesic generators of the event horizon $\mathscr{H}$ have $Q = 0$ [Eq. (11.25)] but those with $\theta \neq \pi/2$ lie outside the equatorial plane.

11.3.9 Moving to the negative-r side

If it maintains $\theta \neq \pi/2$ in the vicinity of $r = 0$, a geodesic $\mathscr{L}$ can a priori move from the region $r > 0$ of $\mathscr{M}$ to the region $r < 0$, or vice-versa, through one the two open disks $r = 0$ (either the disk $\theta < \pi/2$ or the disk $\theta > \pi/2$) delimited by the ring singularity (cf. Sec. 10.2.2). However, such a motion is possible only if $R(0) \geq 0$ (condition (11.30) at the boundary $r = 0$). The case $R(0) = 0$ is excluded for it would correspond either to a r-turning point (Sec. 11.2.6) or to an asymptotic r-value (Sec. 11.3.5). In both cases, $\mathscr{L}$ would remain on a single side of the hypersurface $r = 0$. The necessary condition for $r = 0$ crossing is thus $R(0) > 0$. Now, in view of expression (11.86) of R, we have $R(0) = -a^2 Q$, so that $R(0) > 0 \iff Q < 0$. We thus conclude

> Only a vortical geodesic ($Q < 0$) can cross the hypersurface $r = 0$ and thus move from the positive-r region of spacetime ($\mathscr{M}_+$) to the negative-r one ($\mathscr{M}_-$), or vice-versa. In particular, such a geodesic must have a high energy, i.e. it must obey $|E| > \sqrt{\mu^2 + L^2/a^2}$.

The energy condition is simply the necessary condition for $Q < 0$ stated in Sec. 11.3.6. We note that it implies $|E| > \mu$, which is consistent with the property (11.87) required to travel in $\mathscr{M}_-$.

Example 10: The ingoing principal null geodesics with $\theta \neq \pi/2$ cross the hypersurface $r = 0$ (cf. the dashed green lines in Figs. 10.6 and 10.10, as well as the green lines for $\theta = \pi/6$ in Fig. C.2) and are indeed vortical, since their Carter constants (11.38) obey $Q < 0$ for $\theta \neq \pi/2$. Moreover, they have $\mu = 0$ and $L = aE \sin^2\theta$ [Eq. (11.6)], with $\sin^2\theta < 1$ for $\theta \neq \pi/2$, so that they fulfil $|E| > \sqrt{\mu^2 + L^2/a^2}$.

11.4 Timelike geodesics

11.4.1 Parametrization

Whenever the geodesic $\mathscr{L}$ is timelike, it is relevant to rescale everything by the mass μ of the particle $\mathscr{P}$ whose worldline is $\mathscr{L}$. In particular, as in the Schwarzschild case treated in Sec. 11.4, let us parameterize $\mathscr{L}$ by the proper time τ, which is the affine parameter τ related to the affine parameter λ associated to the 4-momentum $\boldsymbol{p}$ by

$$\tau = \mu\lambda \tag{11.90}$$

and let us introduce the ***specific conserved energy*** ε, ***specific conserved angular momentum*** ℓ , and ***reduced Carter constant*** q:

$$\boxed{\varepsilon := \frac{E}{\mu}}, \qquad \boxed{\ell := \frac{L}{\mu}} \qquad \text{and} \qquad \boxed{q := \frac{Q}{\mu^2}}. \tag{11.91}$$

We shall refer to ε, ℓ and q as the ***reduced integrals of motion*** of the geodesic $\mathscr{L}$. Note that in geometrized units ($c = G = 1$), ε is dimensionless, ℓ has the dimension of a length and q that of a squared length.

The first-order equations of motion (11.50) can be rewritten in terms of the above quantities:

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\tau'} = T_1(r) + T_2(\theta)} \tag{11.92a}$$

$$\boxed{\left(\frac{\mathrm{d}r}{\mathrm{d}\tau'}\right)^2 - \mathcal{R}(r) = 0} \tag{11.92b}$$

$$\boxed{\left(\frac{\mathrm{d}\theta}{\mathrm{d}\tau'}\right)^2 - \tilde{\Theta}(\theta) = 0} \tag{11.92c}$$

$$\boxed{\frac{\mathrm{d}\varphi}{\mathrm{d}\tau'} = \Phi_1(r) + \Phi_2(\theta)} \tag{11.92d}$$

where τ' is related to the Mino parameter λ' by

$$\tau' = \mu\lambda' \tag{11.93}$$

and

$$T_1(r) := \frac{\varepsilon(r^2+a^2)^2 - 2am\ell r}{r^2 - 2mr + a^2}, \qquad T_2(\theta) := -a^2\varepsilon\sin^2\theta, \tag{11.94}$$

$$\Phi_1(r) := \frac{a(2m\varepsilon r - a\ell)}{r^2 - 2mr + a^2}, \qquad \Phi_2(\theta) := \frac{\ell}{\sin^2\theta}, \tag{11.95}$$

$$\boxed{\mathcal{R}(r) := (\varepsilon^2-1)r^4 + 2mr^3 + \left[a^2(\varepsilon^2-1) - q - \ell^2\right] r^2 + 2m\left[q + (\ell - a\varepsilon)^2\right] r - a^2 q}, \tag{11.96}$$

$$\boxed{\tilde{\Theta}(\theta) := q + \cos^2\theta\left[a^2(\varepsilon^2-1) - \frac{\ell^2}{\sin^2\theta}\right]}. \tag{11.97}$$

Note that $\mathcal{R}(r) = R(r)/\mu^2$, $\tilde{\Theta}(\theta) := \Theta(\theta)/\mu^2$ and that expressions (11.96) and (11.97) follow respectively from Eqs. (11.86) and (11.35).

Using Eqs. (11.90) and (11.49), we can relate τ' to the proper time τ:

$$\boxed{\mathrm{d}\tau' = \frac{\mathrm{d}\tau}{\rho^2} = \frac{\mathrm{d}\tau}{r(\tau)^2 + a^2\cos^2\theta(\tau)}}. \tag{11.98}$$

We shall call τ' ***Mino time*** along the geodesic $\mathscr{L}$.

Remark 1: Despite its name, the Mino time has not the dimension of a time, but rather that of a time inverse or length inverse (in the units $G = c = 1$ that we are using).

11.4.2 Bound orbits

We consider here a timelike geodesic $\mathscr{L}$ with $|E| < \mu$, or equivalently

$$|\varepsilon| < 1. \tag{11.99}$$

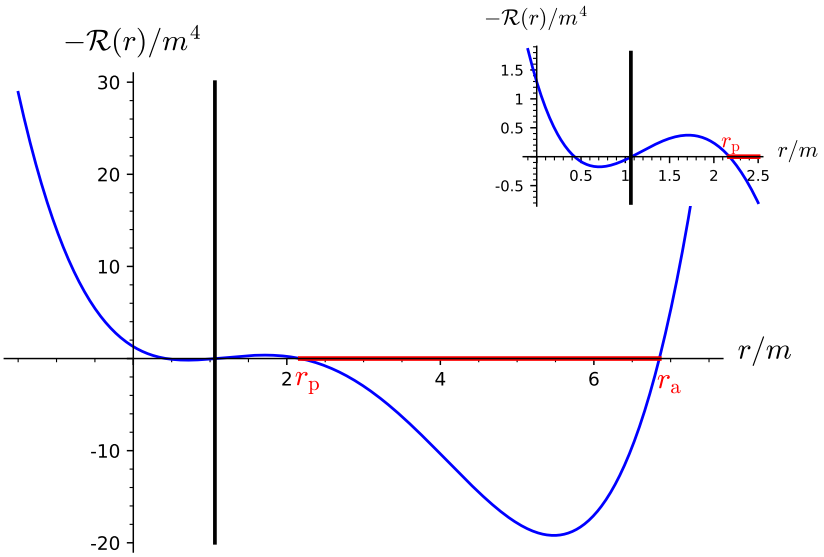


Figure 11.5: Effective potential $-\mathcal{R}(r)$ corresponding to $a = 0.998\,m$, $\varepsilon = 0.9$, $\ell = 2m$ and $q = 1.3\,m^2$. The black vertical line marks the black hole horizon at $r = r_+ \simeq 1.063\,m$. Note that the polynomial $-\mathcal{R}(r)$ has four real roots, one of them being very close to, but distinct from, r_+. The two largest roots give the periastron and apoastron of the bound orbit with the above values of (ε, ℓ, q); they are respectively $r_p \simeq 2.175\,m$ and $r_a \simeq 6.853\,m$. *[Figure generated by the notebook D.4.8]*

As shown in Sec. 11.3.7, such a geodesic has necessarily a non-negative Carter constant:

$$q \geq 0 \tag{11.100}$$

and is located in a radially bounded part of the $r \geq 0$ region of Kerr spacetime.

Equation (11.92b) can be viewed as the first-integral of a 1-dimensional motion in the effective potential $U(r) := -\mathcal{R}(r)$, with a vanishing "total energy" — the right-hand side of Eq. (11.92b). The motion is thus possible wherever $U(r) \leq 0$ or, equivalently, wherever $\mathcal{R}(r) \geq 0$, which is nothing but the constraint (11.30). Property (11.99) implies that the coefficient of r^4 in formula (11.96) for $\mathcal{R}(r)$ is negative. It follows then that $\lim_{r\to\pm\infty} U(r) = +\infty$ and the segments with $U(r) \leq 0$ are located between two roots of $\mathcal{R}(r)$ (cf. Fig. 11.5). Let us denote the lower of these two roots by $r_{\rm p}$, for ***periastron***, and the larger one by $r_{\rm a}$, for ***apoastron***. The periastron and apoastron are of course *r-turning points* of the geodesic, as defined in Sec. 11.2.6.

It is clear that the motion in the potential well $U(r) = -\mathcal{R}(r)$ as governed by Eq. (11.92b) (cf. Fig. 11.5) is periodic:

$$\forall n \in \mathbb{Z}, \quad r(\tau' + n\Lambda_r) = r(\tau'), \tag{11.101}$$

the period Λ_r being the Mino time $\Delta\tau'$ spent to perform a round-trip between $r_{\rm p}$ and $r_{\rm a}$. By rewriting Eq. (11.92b) as $\mathrm{d}\tau' = \pm \mathrm{d}r/\sqrt{\mathcal{R}(r)}$, we get

$$\boxed{\Lambda_r = 2\int_{r_{\rm p}}^{r_{\rm a}} \frac{\mathrm{d}r}{\sqrt{\mathcal{R}(r)}}}. \tag{11.102}$$

Since $\mathcal{R}(r)$ is a polynomial of degree 4 [cf. Eq. (11.96)], the integral in the right-hand side of Eq. (11.102) can be evaluated by means of elliptic integrals. We shall not give the detail here, referring the interested reader to the article [113].

Regarding the θ-motion, we are in the case $L^2 > a^2(E^2-\mu^2)$ considered in Sec. 11.3.6, since for bound geodesics $E^2 - \mu^2 < 0$. The effective θ-potential $V(\theta) = Q - \Theta(\theta)$ has then the shape shown in the left panel of Fig. 11.3 for $\ell = 0$ and in the left panel of Fig. 11.4 for $\ell \neq 0$. We can then distinguish three types of bound orbits:

- ***polar orbit***: $\mathscr{L}$ crosses the rotational axis an infinite number of times; this occurs iff $\ell = 0$ and $q > a^2(1-\varepsilon^2)$;
- ***equatorial orbit***: $\mathscr{L}$ is entirely contained in the equatorial plane $\theta = \pi/2$; for a bound orbit, this occurs iff $q = 0$;
- ***non-polar and non-equatorial orbit***: $\mathscr{L}$ never crosses the rotational axis and oscillates symmetrically about the equatorial plane between two θ-turning points at $\theta = \theta_{\rm m} \in (0, \pi/2)$ and $\theta = \pi - \theta_{\rm m}$; for a bound orbit, this occurs iff $q > 0$ and ($\ell \neq 0$ or $q < a^2(1-\varepsilon^2)$).

Strictly speaking, there is also the exceptional case $\ell = 0$ and $q = a^2(1-\varepsilon^2)$, for which the rotation axis is reached asymptotically (cf. Sec. 11.3.5 and the grey dashed curve in the left panel of Fig. 11.3).

Circular equatorial orbits will be discussed in Sec. 11.5. In the following, we focus on non-polar and non-equatorial orbits, which constitute the generic category of bound timelike orbits. One has then $\theta_{\rm m} \leq \theta \leq \pi - \theta_{\rm m}$, with $\theta_{\rm m}$ given by Eq. (11.80):

$$\boxed{\theta_{\rm m} = \arccos \sqrt{\frac{1}{2}\left[1 + \frac{\ell^2+q}{a^2(1-\varepsilon^2)} - \sqrt{\left[1 + \frac{\ell^2+q}{a^2(1-\varepsilon^2)}\right]^2 - \frac{4q}{a^2(1-\varepsilon^2)}}\right]}}, \tag{11.103}$$

which reduces to Eq. (11.79) for $\ell = 0$:

$$\boxed{\theta_{\rm m} = \arccos \sqrt{\frac{q}{a^2(1-\varepsilon^2)}}.}_{\ell=0} \tag{11.104}$$

From the viewpoint of Eq. (11.92c) above, non-polar orbits have a periodic θ-motion in the potential well $-\tilde{\Theta}(\theta)$:

$$\forall n \in \mathbb{Z}, \quad \theta(\tau' + n\Lambda_\theta) = \theta(\tau'). \tag{11.105}$$

The period Λ_θ is the Mino time $\Delta\tau'$ spent in a round-trip between $\theta_{\rm m}$ and $\pi - \theta_{\rm m}$. By rewriting Eq. (11.92c) as $\mathrm{d}\tau' = \pm \mathrm{d}\theta / \sqrt{\tilde{\Theta}(\theta)}$, we get

$$\boxed{\Lambda_\theta = 2\int_{\theta_{\rm m}}^{\pi-\theta_{\rm m}} \frac{\mathrm{d}\theta}{\sqrt{\tilde{\Theta}(\theta)}} = 4\int_{\theta_{\rm m}}^{\pi/2} \frac{\mathrm{d}\theta}{\sqrt{\tilde{\Theta}(\theta)}}}, \tag{11.106}$$

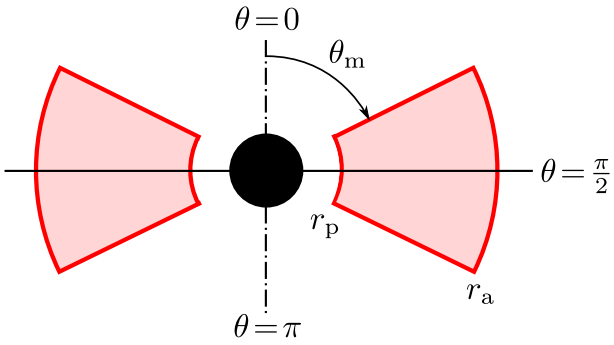


Figure 11.6: Meridional section of the torus occupied by a bound timelike geodesic. The dot-dashed line is the rotation axis and the solid line marks the equatorial plane.

where the last equality results from the symmetry of $\tilde{\Theta}(\theta)$ with respect to $\pi/2$. One naturally associates to the periods Λ_r and Λ_θ the ***Mino angular frequencies***:

$$\boxed{\Upsilon_r := \frac{2\pi}{\Lambda_r}} \quad \text{and} \quad \boxed{\Upsilon_\theta := \frac{2\pi}{\Lambda_\theta}}. \tag{11.107}$$

For non-polar orbits, the motion is restricted by

$$\boxed{r_{\rm p} \leq r \leq r_{\rm a}} \quad \text{and} \quad \boxed{\theta_{\rm m} \leq \theta \leq \pi - \theta_{\rm m}}. \tag{11.108}$$

This means that in terms of the Cartesian Boyer-Lindquist coordinates (11.64), the geodesic is confined inside a torus (cf. Fig. 11.6), which is symmetric about the equatorial plane. Moreover, if the two frequencies Υ_r and Υ_θ are not commensurable, i.e. if $\Upsilon_\theta/\Upsilon_r \notin \mathbb{Q}$, the geodesic fills the torus. Examples of orbits with $\Upsilon_\theta/\Upsilon_r \in \mathbb{Q}$, i.e. orbits with a closed trajectory in the (r,θ) plane, can be found in Ref. [135].

As an illustration, Figs. 11.7–11.9 show a bound timelike geodesic $\mathscr{L}$ of reduced integrals of motion $\varepsilon = 0.9$, $\ell = 2m$ and $q = 1.3\,m^2$ orbiting around a Kerr black hole with $a = 0.998\,m$. It has $r_p \simeq 2.175\,m$, $r_a \simeq 6.853\,m$ and $\theta_{\rm m} \simeq 1.060\,\text{rad} \simeq 60.75°$. The effective radial potential $-\mathcal{R}(r)$ of this geodesic is the one depicted in Fig. 11.5. Figures 11.7–11.9 show actually the segment $0 \leq \tau \leq 600\,m$ of $\mathscr{L}$, starting at the point of Boyer-Lindquist coordinates $(0, (r_{\rm p}+r_{\rm a})/2, \pi/2, 0)$. That $\theta(\tau)$ lies in the interval $[\theta_{\rm m}, \pi-\theta_{\rm m}] \sim [60°, 120°]$ appears clearly in the right panel of Fig. 11.8. On Figs. 11.8 (left panel) and 11.9, we notice a generic feature of eccentric orbits in the Kerr metric, known as ***zoom-whirl***[16]: the particle falls from its apoastron to the central region ("zoom in") and performs some quasi-circular revolutions at close distance to the black hole (it "whirls"), reaching the periastron, and finally goes back to the apoastron. The zoom-whirl behavior exists as well for eccentric orbits in Schwarzschild spacetime (cf. right panel of Fig. 7.12), but is less pronounced, i.e. the number of whirls is smaller (only one in Fig. 7.12).

Let us now consider the evolution of the coordinates t and φ along the geodesic, as governed by Eqs. (11.92a) and (11.92d). Since r and θ are periodic functions of τ', so are

[16] According to Ref. [119], the name "zoom-whirl" has been forged by Curt Cutler and Eric Poisson and may have been suggested by Kip Thorne.

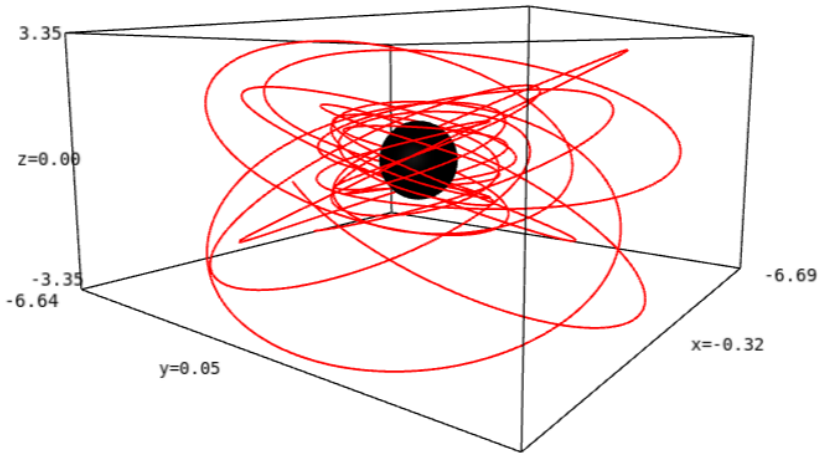


Figure 11.7: Trajectory with respect to the Cartesian Boyer-Lindquist coordinates (x, y, z) [Eq. (11.64)] of a bound timelike geodesic of reduced integrals of motion $\varepsilon = 0.9$, $\ell = 2m$ and $q = 1.3\,m^2$, orbiting around a Kerr black hole with $a = 0.998\,m$. These parameters are the same as for the $-\mathcal{R}(r)$ plot in Fig. 11.5. *[Figure generated by the notebook D.4.8]*

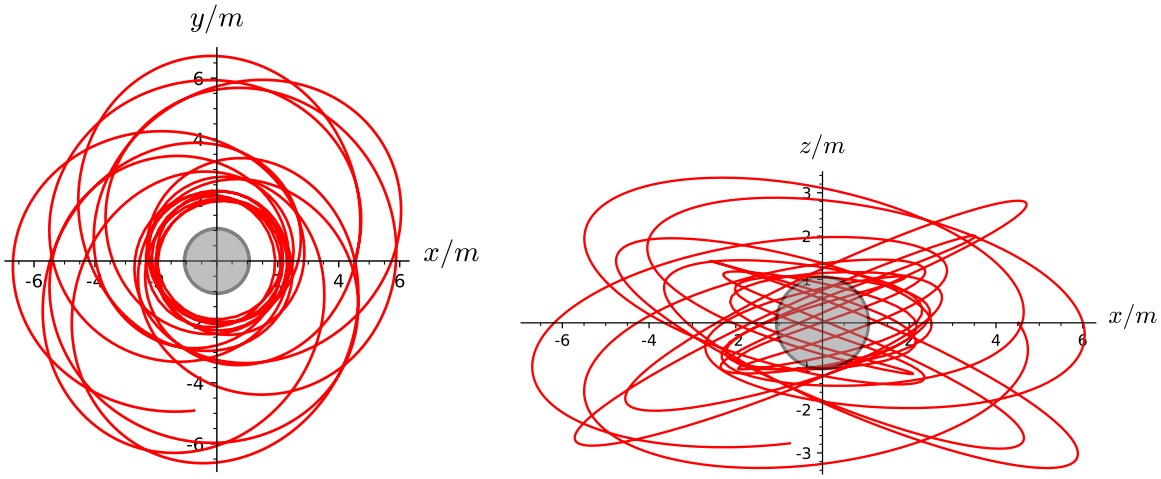


Figure 11.8: Projection in the (x, y) plane (equatorial plane; left panel) and the (x, z) plane (meridional plane; right panel) of the bound timelike geodesic considered in Fig. 11.7. The grey disk depicts the black hole region. On the left panel, there appears clearly that the minimal value of r is $r_p \simeq 2.2\,m$, while its maximal value is $r_a \simeq 6.9\,m$. On the right panel, we check that $\theta_{\rm m} \leq \theta \leq \pi - \theta_{\rm m}$ with $\theta_{\rm m} \sim 60°$, in agreement with the generic behavior illustrated in Fig. 11.6. *[Figure generated by the notebook D.4.8]*

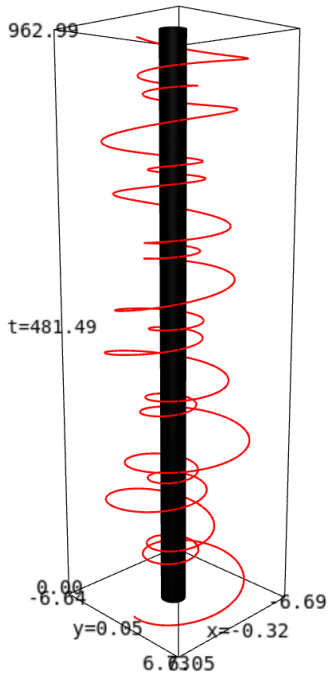


Figure 11.9: Spacetime diagram based on the Cartesian Boyer-Lindquist coordinates (t, x, y) [Eq. (11.64)] depicting the bound timelike geodesic considered in Figs. 11.7 and 11.8. The black cylinder is the black hole event horizon. *[Figure generated by the notebook D.4.8]*

the terms $T_1(r)$, $T_2(\theta)$, $\Phi_1(r)$ and $\Phi_2(\theta)$ that appear in the right-hand side of Eqs. (11.92a) and (11.92d). We can thus expand them in Fourier series:

$$T_1(r) = \sum_{k=-\infty}^{+\infty} \hat{T}_{1k}\, \mathrm{e}^{\mathrm{i}k\Upsilon_r \tau'} = \langle T_1(r)\rangle + \sum_{k=1}^{+\infty} \left(\hat{T}_{1k}\, \mathrm{e}^{\mathrm{i}k\Upsilon_r \tau'} + \text{c.c.}\right) \tag{11.109a}$$

$$T_2(\theta) = \sum_{k=-\infty}^{+\infty} \hat{T}_{2k}\, \mathrm{e}^{\mathrm{i}k\Upsilon_\theta \tau'} = \langle T_2(\theta)\rangle + \sum_{k=1}^{+\infty} \left(\hat{T}_{2k}\, \mathrm{e}^{\mathrm{i}k\Upsilon_\theta \tau'} + \text{c.c.}\right) \tag{11.109b}$$

$$\Phi_1(r) = \sum_{k=-\infty}^{+\infty} \hat{\Phi}_{1k}\, \mathrm{e}^{\mathrm{i}k\Upsilon_r \tau'} = \langle \Phi_1(r)\rangle + \sum_{k=1}^{+\infty} \left(\hat{\Phi}_{1k}\, \mathrm{e}^{\mathrm{i}k\Upsilon_r \tau'} + \text{c.c.}\right) \tag{11.109c}$$

$$\Phi_2(\theta) = \sum_{k=-\infty}^{+\infty} \hat{\Phi}_{2k}\, \mathrm{e}^{\mathrm{i}k\Upsilon_\theta \tau'} = \langle \Phi_2(\theta)\rangle + \sum_{k=1}^{+\infty} \left(\hat{\Phi}_{2k}\, \mathrm{e}^{\mathrm{i}k\Upsilon_\theta \tau'} + \text{c.c.}\right), \tag{11.109d}$$

where "c.c." stands for *complex conjugate* and $\langle T_1(r)\rangle = \hat{T}_{10}$, $\langle T_2(\theta)\rangle = \hat{T}_{20}$, $\langle \Phi_1(r)\rangle = \hat{\Phi}_{10}$ and $\langle \Phi_2(\theta)\rangle = \hat{\Phi}_{20}$ are the mean values of the functions $T_1(r(\tau'))$, $T_2(\theta(\tau'))$, $\Phi_1(r(\tau'))$ and

$\Phi_2(\theta(\tau'))$ over one period:

$$\langle T_1(r)\rangle = \frac{1}{\Lambda_r}\int_0^{\Lambda_r} T_1(r(\tau'))\,\mathrm{d}\tau' = \frac{2}{\Lambda_r}\int_{r_\mathrm{p}}^{r_\mathrm{a}} \frac{T_1(r)}{\sqrt{R(r)}}\,\mathrm{d}r \tag{11.110a}$$

$$\langle T_2(\theta)\rangle = \frac{1}{\Lambda_\theta}\int_0^{\Lambda_\theta} T_2(\theta(\tau'))\,\mathrm{d}\tau' = \frac{4}{\Lambda_\theta}\int_{\theta_\mathrm{m}}^{\pi/2} \frac{T_2(\theta)}{\sqrt{\tilde{\Theta}(\theta)}}\,\mathrm{d}\theta, \tag{11.110b}$$

with similar formulas for $\langle\Phi_1(r)\rangle$ and $\langle\Phi_2(\theta)\rangle$.

We may then rewrite Eqs. (11.92a) and (11.92d) as

$$\frac{\mathrm{d}t}{\mathrm{d}\tau'} = \Gamma + \sum_{k=1}^{+\infty}\left(\hat{T}_{1k}\,\mathrm{e}^{\mathrm{i}k\Upsilon_r\tau'} + \text{c.c.}\right) + \sum_{k=1}^{+\infty}\left(\hat{T}_{2k}\,\mathrm{e}^{\mathrm{i}k\Upsilon_\theta\tau'} + \text{c.c.}\right) \tag{11.111a}$$

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\tau'} = \Upsilon_\varphi + \sum_{k=1}^{+\infty}\left(\hat{\Phi}_{1k}\,\mathrm{e}^{\mathrm{i}k\Upsilon_r\tau'} + \text{c.c.}\right) + \sum_{k=1}^{+\infty}\left(\hat{\Phi}_{2k}\,\mathrm{e}^{\mathrm{i}k\Upsilon_\theta\tau'} + \text{c.c.}\right), \tag{11.111b}$$

with

$$\Gamma := \langle T_1(r)\rangle + \langle T_2(\theta)\rangle, \tag{11.112}$$

and

$$\Upsilon_\varphi := \langle\Phi_1(r)\rangle + \langle\Phi_2(\theta)\rangle. \tag{11.113}$$

Γ and Υ_φ represent the constant (or average) parts of respectively $\mathrm{d}t/\mathrm{d}\tau'$ and $\mathrm{d}\varphi/\mathrm{d}\tau'$, while the oscillatory parts are the terms involved in the sums over $k \geq 1$. Since Γ is the average value of $\mathrm{d}t/\mathrm{d}\tau'$, we can define the ***average angular frequencies with respect to Boyer-Lindquist time*** by [94, 113]

$$\Omega_r := \frac{\Upsilon_r}{\Gamma}, \qquad \Omega_\theta := \frac{\Upsilon_\theta}{\Gamma}, \qquad \Omega_\varphi := \frac{\Upsilon_\varphi}{\Gamma}. \tag{11.114}$$

We may summarize the above results by

> The motion of a generic bound timelike geodesic in Kerr spacetime is periodic in r and θ only in terms of the Mino time τ', the corresponding angular frequencies being Υ_r and Υ_θ. In terms of the Boyer-Lindquist time t, the motion in r or θ is not periodic. In this respect, the "average frequencies" Ω_r and Ω_θ defined above are not true frequencies like Υ_r or Υ_θ. Regarding the motion in φ, it is not periodic, neither in terms of the Mino time, nor in terms of the Boyer-Lindquist time, so that both Υ_φ and Ω_φ are average frequencies.

Quasi-Keplerian parametrization

Given a bound timelike geodesic $\mathscr{L}$ of Kerr spacetime, one may define its ***eccentricity*** e and (dimensionless) ***semilatus rectum*** p by means of Keplerian-like formulas [228, 95, 235]:

$$r_\mathrm{p} =: \frac{pm}{1+e} \qquad \text{and} \qquad r_\mathrm{a} =: \frac{pm}{1-e}, \tag{11.115}$$

or equivalently

$$p = \frac{2r_\mathrm{a} r_\mathrm{p}}{m(r_\mathrm{a}+r_\mathrm{p})} \qquad \text{and} \qquad e = \frac{r_\mathrm{a}-r_\mathrm{p}}{r_\mathrm{a}+r_\mathrm{p}}. \tag{11.116}$$

One may also introduce the ***inclination angle*** θ_inc by

$$\theta_\mathrm{inc} := \frac{\pi}{2} - (\mathrm{sgn}\,\ell)\theta_\mathrm{m}. \tag{11.117}$$

The sign of ℓ appears in this formula to enforce $\theta_\mathrm{inc} \in [0, \pi/2]$ for prograde orbits ($\ell > 0$) and $\theta_\mathrm{inc} \in [\pi/2, \pi]$ for retrograde orbits ($\ell < 0$).

Instead of (ε, ℓ, q), a bound timelike geodesic can be parameterized[17] by $(p, e, \theta_\mathrm{inc})$. Actually, there is a one-to-one correspondence between (ε, ℓ, q) and $(p, e, \theta_\mathrm{inc})$: it is clear from formulas (11.116) and (11.116) that $(p, e, \theta_\mathrm{inc})$ are functions of (ε, ℓ, q), because (i) r_a and r_p are some roots of the quartic polynomial $\mathcal{R}(r)$, whose coefficients are functions of (ε, ℓ, q) only (at fixed Kerr parameters (m, a)) [cf. Eq. (11.96)] and (ii) θ_m in the right-hand side of expression (11.117) for θ_inc is the function of (ε, ℓ, q) given by Eq. (11.103). Conversely, to express (ε, ℓ, q) in terms of $(p, e, \theta_\mathrm{inc})$, one shall write

$$\begin{cases} \mathcal{R}(r_\mathrm{p}) = 0 \\ \mathcal{R}(r_\mathrm{a}) = 0 \\ \tilde{\Theta}(\theta_\mathrm{m}) = 0, \end{cases}$$

where r_p and r_a are the functions of (e, p) given by Eq. (11.115) and θ_m is the function of θ_inc given by Eq. (11.117). The above system is then a system of 3 equations for the 3 unknowns (ε, ℓ, q); see Appendix B of Ref. [228] or Appendix A of Ref. [95] for details.

Example 11: The geodesic $\mathscr{L}$ depicted in Figs. 11.7–11.9 has $\varepsilon = 0.9$, $\ell = 2m$ and $q = 1.3\,m^2$, from which we get $r_p = 2.175\,m$, $r_a = 6.853\,m$ and $\theta_\mathrm{m} = 1.060\,\mathrm{rad} = 60.75°$. Then Eqs. (11.116) and (11.117) yield $p = 3.302$, $e = 0.518$ and $\theta_\mathrm{inc} = 29.26°$.

The parametrization $(p, e, \theta_\mathrm{inc})$ allows one to introduce along the geodesic $\mathscr{L}$ the radial phase angle $\psi(\tau)$ and the meridional phase angle $\chi(\tau)$ as monotonically increasing functions of τ such that

$$r(\tau) =: \frac{pm}{1 + e\cos\psi(\tau)} \qquad \text{and} \qquad \theta(\tau) =: \arccos\left(\cos\theta_\mathrm{m}\,\cos\chi(\tau)\right). \tag{11.118}$$

For numerical integration, $\psi(\tau)$ and $\chi(\tau)$ are preferred to $r(\tau)$ and $\theta(\tau)$ since the latter are tricky at the turning points, where $\mathrm{d}r/\mathrm{d}\tau$ and $\mathrm{d}\theta/\mathrm{d}\tau$ vanish and change sign [94].

11.5 Circular timelike orbits in the equatorial plane

11.5.1 Equations of motion in the equatorial plane

Let us focus on timelike geodesics confined to the equatorial plane: $\theta = \pi/2$. They thus have $\mathrm{d}\theta/\mathrm{d}\tau' = 0$ and Eq. (11.92c) leads to $\tilde{\Theta}(\theta) = 0$. Given expression (11.97) for $\tilde{\Theta}(\theta)$,

[17] Some authors use $x := \cos\theta_\mathrm{inc}$ instead of θ_inc [235].

this implies

$$\boxed{q = 0}. \tag{11.119}$$

Remark 1: The above result follows as well from the analysis of the θ-motion performed in Sec. 11.3.6 and according to which the only possible value of the Cartan constant Q for a motion confined to the equatorial plane is $Q = 0$, be the geodesic timelike or null. The converse is not true: there exist (timelike or null) geodesics with $Q = 0$ outside the equatorial plane; they are asymptotically approaching the equatorial from one side, except for the exceptional case $L = 0$ and $a^2E^2 = a^2\mu^2$, for which they are moving with a constant value of θ (cf. Sec. 11.3.6).

When $\theta = \pi/2$, we have $\sin^2\theta = 1$, $\rho^2 = r^2$ and $\mathrm{d}\tau' = r^{-2}\mathrm{d}\tau$ [cf. Eq. (11.98)], so that the system (11.92) reduces to

$$\boxed{\frac{\mathrm{d}t}{\mathrm{d}\tau} = \frac{1}{r^2 - 2mr + a^2}\left[\varepsilon(r^2 + a^2) + \frac{2am}{r}(a\varepsilon - \ell)\right]} \tag{11.120a}$$

$$\boxed{\left(\frac{\mathrm{d}r}{\mathrm{d}\tau}\right)^2 + \mathcal{V}(r) = 0} \tag{11.120b}$$

$$\boxed{\frac{\mathrm{d}\varphi}{\mathrm{d}\tau} = \frac{1}{r^2 - 2mr + a^2}\left[\ell\left(1 - \frac{2m}{r}\right) + \frac{2am\varepsilon}{r}\right]}, \tag{11.120c}$$

with

$$\boxed{\mathcal{V}(r) := 1 - \varepsilon^2 - \frac{2m}{r} + \frac{\ell^2 + a^2(1 - \varepsilon^2)}{r^2} - \frac{2m(\ell - a\varepsilon)^2}{r^3}}. \tag{11.121}$$

Note that $\mathcal{V}(r) = -\mathcal{R}(r)/r^4 = -R(r)/(\mu^2 r^4)$ [cf. Eqs. (11.96) and (11.86)].

Remark 2: At the Schwarzschild limit ($a = 0$) we recover the differential equations of Chap. 7, namely Eq. (11.120a) reduces to Eq. (7.34), Eq. (11.120b) to Eq. (7.24) and Eq. (11.120c) to Eq. (7.35). In particular we have $\mathcal{V}(r) = 1 - \varepsilon^2 + 2V_\ell(r)$, where $V_\ell(r)$ is the effective potential (7.25). Note however that when $a \neq 0$, $\mathcal{V}(r)$ cannot be considered as an effective potential for the radial motion, because it depends on ε, in addition to ℓ, even if we subtract from it the constant term $1 - \varepsilon^2$. This is why we did not bother to add a 1/2 factor in Eq. (11.120b), as we did for Eq. (7.24) to make it match the first integral of a 1-dimensional motion in a potential well.

11.5.2 Equatorial circular orbits: definition and existence

The simplest equatorial geodesics are of course the circular ones: a ***circular orbit in the equatorial plane*** is a geodesic that obeys

$$\forall \tau \in \mathbb{R}, \quad \theta(\tau) = \pi/2 \qquad \text{and} \qquad r(\tau) = r_0 = \text{const.} \tag{11.122}$$

The constant r_0 is called the ***radius*** of the circular orbit. One has then $\mathrm{d}r/\mathrm{d}\tau = 0$, so that Eq. (11.120b) implies

$$\boxed{\mathcal{V}(r_0) = 0}. \tag{11.123}$$

The above condition is not sufficient to single out a geodesic worldline: there exist worldlines with constant $r = r_0$ that obey the whole system (11.120) (which includes (11.123)) but that are not solution of the geodesic equation (11.1) (see Ref. [251] for concrete examples). One must add a second condition, which is obtained as follows. For any timelike equatorial geodesic (not necessarily circular), the following relation must hold:

$$\frac{\mathrm{d}r}{\mathrm{d}\tau}\left[2\frac{\mathrm{d}^2r}{\mathrm{d}\tau^2} + \mathcal{V}'(r)\right] = 0.$$

It is obtained by differentiating Eq. (11.120b) with respect to τ. If the geodesic is not circular, at any point where r is not stationary (i.e. excluding r-turning points), we have $\mathrm{d}r/\mathrm{d}\tau \neq 0$ and the above relation implies

$$\mathcal{V}'(r) = -2\frac{\mathrm{d}^2r}{\mathrm{d}\tau^2}.$$

Since in the vicinity of any circular geodesic there exist non-circular ones (those with small eccentricity e, as defined by Eq. (11.116)), by continuity, we shall ask that this relation holds for circular geodesics as well. For the latter ones, $r = r_0$ and $\mathrm{d}^2r/\mathrm{d}\tau^2 = 0$, so that it simplifies to

$$\boxed{\mathcal{V}'(r_0) = 0}. \tag{11.124}$$

One can verify[18] that this relation, in conjunction with Eq. (11.123), is sufficient to eliminate the non-geodesic circular worldlines.

To summarize, circular timelike geodesics in the equatorial plane are obtained by solving the system (11.123)-(11.124). Given expression (11.121) for $\mathcal{V}$, this system takes the form

$$\begin{cases} (1-\varepsilon^2)r_0^3 - 2mr_0^2 + \left[\ell^2 + a^2(1-\varepsilon^2)\right] r_0 - 2m(\ell - a\varepsilon)^2 = 0 & \text{(11.125a)} \\ mr_0^2 - \left[\ell^2 + a^2(1-\varepsilon^2)\right] r_0 + 3m(\ell - a\varepsilon)^2 = 0. & \text{(11.125b)} \end{cases}$$

This is a system of 2 nonlinear equations for 3 unknowns: r_0, ε and ℓ. We thus expect a 1-parameter family of solutions. It is convenient to choose r_0 as the parameter. We have thus to solve (11.125) for (ε, ℓ). The change of variable

$$\tilde{\ell} := \ell - a\varepsilon \tag{11.126}$$

turns (11.125) into the system

$$\begin{cases} r_0^3\varepsilon^2 = m\tilde{\ell}^2 + r_0^2(r_0 - m) & \text{(11.127a)} \\ 2ar_0\,\varepsilon\tilde{\ell} = (3m - r_0)\tilde{\ell}^2 + r_0(mr_0 - a^2). & \text{(11.127b)} \end{cases}$$

Let us consider the square of Eq. (11.127b):

$$4a^2r_0^2\,\varepsilon^2\tilde{\ell}^2 = \left[(3m - r_0)\tilde{\ell}^2 + r_0(mr_0 - a^2)\right]^2, \tag{11.128}$$

[18] See Ref. [251] for details.

and substitute (11.127a) for ε^2 in it; we get

$$\left[r_0(r_0-3m)^2-4a^2m\right]\tilde{\ell}^4+2r_0^2\left[(3m^2-a^2)r_0-m(r_0^2+a^2)\right]\tilde{\ell}^2+r_0^3(mr_0-a^2)^2=0.$$

This is a quadratic equation for $X:=\tilde{\ell}^2$. Its discriminant turns out to have a simple form:

$$\Delta=16a^2mr_0^3(r_0^2-2mr_0+a^2)^2.$$

It follows immediately that $\Delta\geq 0 \iff r_0\geq 0$. Hence, there is no solution for $r_0<0$:

There does not exist any equatorial circular timelike orbit in the region $r<0$ of Kerr spacetime.

Remark 3: For $r_0<0$ with $|r_0|\gg m$, the above result is not surprising since we have seen in Sec. 10.2.2 that under these conditions, the Kerr metric appears as a Schwarzschild metric with a negative mass. The asymptotic gravitational field is then repulsive and certainly does not admit any circular orbit. In the region $r<0$ and $|r|$ small, the above argument does not apply. However, the results obtained in Sec. 11.3.7 show that if a circular orbit would exist there, it would necessarily have $|E|>\mu$ [cf. Eq. (11.87)], i.e. would be unbound.

The case $r_0=0$ is excluded as well, since in the equatorial plane, it would correspond to the curvature singularity. In what follows, we therefore assume

$$\boxed{r_0>0}. \tag{11.129}$$

We have then $\sqrt{\Delta}=4ar_0\sqrt{mr_0}|r_0^2-2mr_0+a^2|$ and the solutions of the quadratic equation are[19]

$$\tilde{\ell}_\pm^2=r_0\frac{(a^2-3m^2)r_0^2+mr_0(r_0^2+a^2)\mp 2a\sqrt{mr_0}(r_0^2-2mr_0+a^2)}{r_0(r_0-3m)^2-4a^2m}. \tag{11.130}$$

The denominator of the right-hand side can be written as

$$r_0(r_0-3m)^2-4a^2m=\frac{1}{r_0}\left[(r_0(r_0-3m))^2-(2a\sqrt{mr_0})^2\right]=\frac{1}{r_0}A_+A_-,$$

with

$$A_\pm:=r_0(r_0-3m)\pm 2a\sqrt{mr_0}.$$

In the numerator of (11.130), we may use the identity $\mp 2a\sqrt{mr_0}=A_\mp-r_0(r_0-3m)$ and get, after simplification,

$$\tilde{\ell}_\pm^2=\frac{r_0^2}{A_+A_-}\left[A_\mp(r_0^2-2mr_0+a^2)-A_+A_-\right]=\frac{r_0^2}{A_\pm}\left(r_0^2-2mr_0+a^2-A_\pm\right).$$

[19] For future convenience, we have chosen the sign $\mp$ in the numerator to be the opposite of the label $\pm$ in $\tilde{\ell}_\pm^2$.

Using $r_0^2 - 2mr_0 + a^2 - A_\pm = a^2 \mp 2a\sqrt{mr_0} + mr_0 = (a \mp \sqrt{mr_0})^2$, we obtain

$$\tilde{\ell}_\pm^2 = \frac{r_0^2(a \mp \sqrt{mr_0})^2}{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}. \tag{11.131}$$

Since obviously $\tilde{\ell}_\pm^2 \geq 0$, this expression gives birth to the constraint $r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0} > 0$, which is equivalent to

$$r_0^{3/2} - 3mr_0^{1/2} \pm 2a\sqrt{m} > 0. \tag{11.132}$$

The left hand-side of this inequality is a cubic polynomial in $x := r_0^{1/2}$ and determining its sign amounts to compute its roots. Fortunately, this corresponds to a depressed cubic equation $x^3 + px + q = 0$, with $p := -3m$ and $q := \pm 2a\sqrt{m}$, the discriminant of which is $-(4p^3 + 27q^2) = 108m(m^2 - a^2) \geq 0$. The roots $(x_k)_{k\in\{0,1,2\}}$ are then all real and are given by Viète's formula (8.22). They obey $x_0 + x_1 + x_2 = 0$ and $x_0x_1x_2 = -q = \mp 2a\sqrt{m}$, from which we deduce that for $\pm = +$ in Eq. (11.132), i.e. for $\tilde{\ell}_+^2$, there are two positive roots, which are x_0 and x_2 as given by Eq. (8.22), while for $\pm = -$, i.e. for $\tilde{\ell}_-^2$, there is a single positive root, which is x_0 as given by Eq. (8.22). Going back to $r_0 = x^2$, with the roots x_0 and x_2 given by Viète's formula (8.22), we get the following constraints for circular orbits in the equatorial plane:

$$\text{for } \tilde{\ell}_+^2 : \qquad 0 < r_0 < r_{\rm ph}^* \qquad \text{or} \qquad r_0 > r_{\rm ph}^+ \tag{11.133a}$$

$$\text{for } \tilde{\ell}_-^2 : \qquad r_0 > r_{\rm ph}^-, \tag{11.133b}$$

with

$$\boxed{r_{\rm ph}^\pm := 4m\cos^2\left[\frac{1}{3}\arccos\left(\mp\frac{a}{m}\right)\right]} \tag{11.134}$$

and

$$\boxed{r_{\rm ph}^* := 4m\cos^2\left[\frac{1}{3}\arccos\left(-\frac{a}{m}\right) + \frac{4\pi}{3}\right]}. \tag{11.135}$$

The index "ph" stands for *photon*, because we shall see in Sec. 11.5.4 and in Sec. 12.3 that these radii actually correspond to photon orbits (circular null geodesics). $r_{\rm ph}^\pm$ and $r_{\rm ph}^*$ are plotted as functions of a in Fig. 11.10 (green curves). Note that $m \leq r_{\rm ph}^+ \leq 3m$ and $3m \leq r_{\rm ph}^- \leq 4m$, with

$$\lim_{a\to 0} r_{\rm ph}^\pm = 3m, \quad \lim_{a\to m} r_{\rm ph}^+ = m, \quad \lim_{a\to m} r_{\rm ph}^- = 4m. \tag{11.136}$$

Besides, $0 \leq r_{\rm ph}^* \leq m$, with

$$\lim_{a\to 0} r_{\rm ph}^* = 0, \quad \lim_{a\to m} r_{\rm ph}^* = m. \tag{11.137}$$

Furthermore, one can check that

$$r_{\rm ph}^\pm \geq r_+ \tag{11.138}$$

and

$$0 \leq r_{\rm ph}^* \leq r_-, \tag{11.139}$$

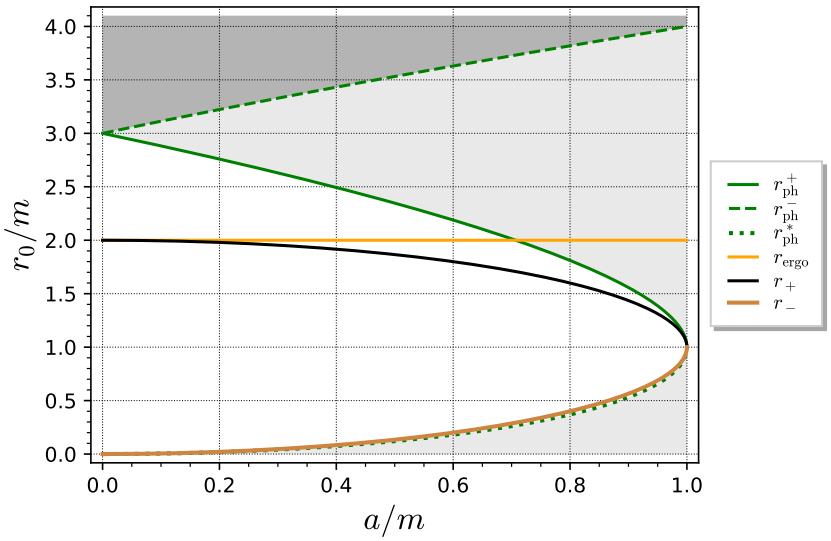


Figure 11.10: Domain of existence of circular equatorial timelike orbits in the (a, r_0) plane: orbits of the (ε_+, ℓ_+) family exist in both the light and dark grey regions, while orbits of the (ε_-, ℓ_-) family exist only in the dark grey region. The orange line marks the location of the outer ergosphere in the equatorial plane, which is $r_{\rm ergo} = r_{\mathscr{E}^+}(\pi/2) = 2m$ [cf. Eq. (10.19d)], the black curve corresponds to the black hole event horizon $\mathscr{H}$ ($r = r_+$) and the maroon one to the inner horizon $\mathscr{H}_{\rm in}$ ($r = r_-$). *[Figure generated by the notebook D.4.9]*

where $r_+ := m^2 - \sqrt{m^2 - a^2}$ and $r_- := m^2 - \sqrt{m^2 - a^2}$ [Eq. (10.3)] are the radii of respectively the event horizon $\mathscr{H}$ and the inner horizon $\mathscr{H}_{\rm in}$. Regarding the upper bound in Eq. (11.139), one can see from Fig. 11.10 that $r_{\rm ph}^*$ is always very close to r_-. The maximum discrepancy is $r_- - r_{\rm ph}^* \simeq 0.032\, m$ and is achieved for $a \simeq 0.9\, m$ (cf. the notebook D.4.9). We conclude that the first permitted range for circular orbits in (11.133a) lies entirely in the region $\mathscr{M}_{\rm III}$ of Kerr spacetime, while the second range in (11.133a) lies in entirely in $\mathscr{M}_{\rm I}$. Regarding the orbits for $\tilde{\ell}_-^2$, the range (11.133b) lies in $\mathscr{M}_{\rm I}$ as well.

Substituting expression (11.131) for $\tilde{\ell}^2$ in Eq. (11.127a), we get, after simplification,

$$\varepsilon_\pm^2 = \frac{\left(r_0^2 - 2mr_0 \pm a\sqrt{mr_0}\right)^2}{r_0^2\left(r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}\right)}. \tag{11.140}$$

The $\pm$'s in Eqs. (11.131) and (11.140) have to be consistent. In other words, we have two pairs of solutions for $(\varepsilon^2, \tilde{\ell}^2)$, namely $(\varepsilon_+^2, \tilde{\ell}_+^2)$ and $(\varepsilon_-^2, \tilde{\ell}_-^2)$. A priori, ε_+^2 leads to two values for ε, namely $\pm\sqrt{\varepsilon_+^2}$, and $\tilde{\ell}_+^2$ leads to two values for $\tilde{\ell}$, namely $\pm\sqrt{\tilde{\ell}_+^2}$, so that the pair $(\varepsilon_+^2, \tilde{\ell}_+^2)$ would generate four solutions for $(\varepsilon, \tilde{\ell})$, and similarly the pair $(\varepsilon_-^2, \tilde{\ell}_-^2)$ would generate four extra solutions for $(\varepsilon, \tilde{\ell})$, leading to a total of eight solutions. However, by construction these solutions satisfy the "squared" equation (11.128) but not all of them satisfy the original equation (11.127b). To see this, let us consider the following square

roots of Eqs. (11.140) and (11.131):

$$\boxed{\varepsilon_\pm = \frac{r_0^2 - 2mr_0 \pm a\sqrt{mr_0}}{r_0\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}}}, \tag{11.141}$$

$$\tilde{\ell}_\pm = -\frac{r_0(a \mp \sqrt{mr_0})}{\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}} = \frac{r_0(\pm\sqrt{mr_0} - a)}{\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}} \tag{11.142}$$

and write the eight solutions of (11.128) as

$$(\varepsilon, \tilde{\ell}) = \left(\epsilon_1\varepsilon_\pm, \epsilon_2\tilde{\ell}_\pm\right), \qquad \text{with} \quad \epsilon_1 = \pm 1 \qquad \text{and} \qquad \epsilon_2 = \pm 1.$$

By a direct calculation, we get

$$2ar_0\,\varepsilon\tilde{\ell} - (3m - r_0)\tilde{\ell}^2 - r_0(mr_0 - a^2) = \frac{2(1 - \epsilon_1\epsilon_2)ar_0(a \mp \sqrt{mr_0})(r_0^2 - 2mr_0 \pm a\sqrt{mr_0})}{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}.$$

Hence Eq. (11.127b) is fulfilled iff $1 - \epsilon_1\epsilon_2 = 0$, i.e. iff $\epsilon_1\epsilon_2 = 1$. This reduces the number of possible solutions from eight to four:

$$(\varepsilon, \tilde{\ell}) = (\varepsilon_+, \tilde{\ell}_+) \quad \text{or} \quad (-\varepsilon_+, -\tilde{\ell}_+) \quad \text{or} \quad (\varepsilon_-, \tilde{\ell}_-) \quad \text{or} \quad (-\varepsilon_-, -\tilde{\ell}_-). \tag{11.143}$$

A further reduction of the number of solutions is provided by the future-directed condition (11.57). Since $R(r_0) = -\mu^2 r_0^4 \mathcal{V}(r_0) = 0$ by virtue of Eq. (11.123), for circular equatorial orbits, Eq. (11.57) reduces to

$$\frac{1}{\Delta_0}\left(\varepsilon - \frac{a}{r_0^2 + a^2}\ell\right) > 0,$$

where $\Delta_0 := r_0^2 - 2mr_0 + a^2 = (r_0 - r_+)(r_0 - r_-)$. Now, we have observed above that circular orbits lie either in $\mathscr{M}_{\rm I}$ or $\mathscr{M}_{\rm III}$, where $\Delta_0 > 0$. Therefore, we can further simplify the future-directed condition to

$$\varepsilon - \frac{a}{r_0^2 + a^2}\ell > 0.$$

Once reexpressed in terms of $\tilde{\ell} = \ell - a\varepsilon$, it is equivalent to

$$r_0^2\varepsilon - a\tilde{\ell} > 0. \tag{11.144}$$

Let us check each of the four solutions (11.143):

$$\begin{aligned}
r_0^2\varepsilon_+ - a\tilde{\ell}_+ &= \frac{r_0\Delta_0}{\sqrt{r_0^2 - 3mr_0 + 2a\sqrt{mr_0}}} > 0\\
r_0^2(-\varepsilon_+) - a(-\tilde{\ell}_+) &= -(r_0^2\varepsilon_+ - a\tilde{\ell}_+) < 0\\
r_0^2\varepsilon_- - a\tilde{\ell}_- &= \frac{r_0\Delta_0}{\sqrt{r_0^2 - 3mr_0 - 2a\sqrt{mr_0}}} > 0\\
r_0^2(-\varepsilon_-) - a(-\tilde{\ell}_-) &= -(r_0^2\varepsilon_- - a\tilde{\ell}_-) < 0.
\end{aligned}$$

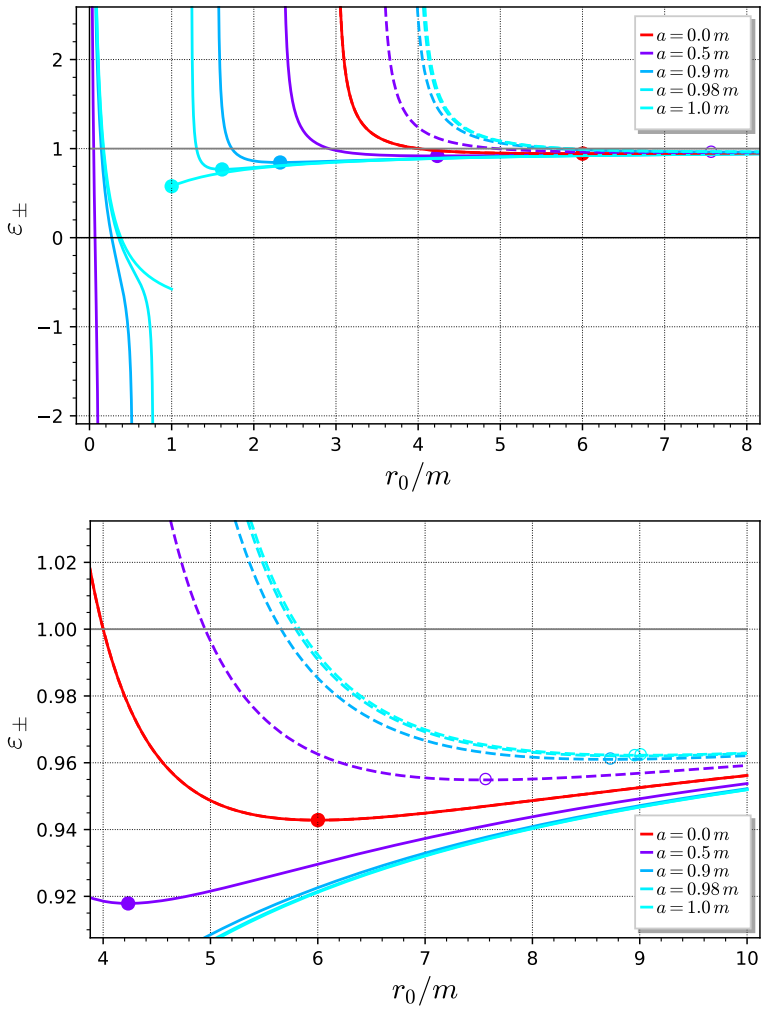


Figure 11.11: Specific conserved energy $\varepsilon = \varepsilon_+$ (solid curves) or $\varepsilon = \varepsilon_-$ (dashed curves) along circular timelike orbits in the equatorial plane, as a function of the orbital radius r_0 [Eq. (11.141)], for selected values of the Kerr spin parameter a. Curves in the inner region ($r_0 < m$) terminate by a vertical asymptote at $r = r^*_{\rm ph}(a)$ given by Eq. (11.135), while the other curves start along a vertical asymptote at $r = r^\pm_{\rm ph}(a)$ given by Eq. (11.134). Dots on ε_+ curves and open circles on ε_- ones mark the ISCO: all configurations on the left of these points are unstable. The bottom panel is a zoom on the region $4\,m \leq r_0 \leq 10\,m$. Note that the red curve ($a = 0$) is the same as in Fig. 7.8. *[Figure generated by the notebook D.4.9]*

We conclude that only two solutions remain:

$$(\varepsilon, \tilde{\ell}) = (\varepsilon_+, \tilde{\ell}_+) \quad \text{or} \quad (\varepsilon_-, \tilde{\ell}_-). \tag{11.145}$$

Given that $\ell = \tilde{\ell} + a\varepsilon$ [Eq. (11.126)] and expressions (11.141) and (11.142) for respectively $\varepsilon_\pm$ and $\tilde{\ell}_\pm$, these two solutions can be reexpressed in terms of (ε, ℓ) as

$$(\varepsilon, \ell) = \boxed{(\varepsilon_+, \ell_+)} \quad \text{or} \quad \boxed{(\varepsilon_-, \ell_-)}, \tag{11.146}$$

where

$$\boxed{\ell_\pm = \pm\sqrt{\frac{m}{r_0}}\frac{r_0^2 + a^2 \mp 2a\sqrt{mr_0}}{\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}}}. \tag{11.147}$$

The quantities $\varepsilon_\pm$ and $\ell_\pm$ are plotted in terms of r_0 in Figs. 11.11 and 11.12. Orbits with $\ell > 0$ (resp. $\ell < 0$) are called ***prograde*** (resp. ***retrograde***). We have immediately $\ell_- < 0$, while the sign of ℓ_+ is that of of $r_0^2 + a^2 - 2a\sqrt{mr_0}$. Outside the event horizon, i.e. in $\mathscr{M}_{\rm I}$, we have $r_0 > m$, which implies $r_0^2 + a^2 - 2a\sqrt{mr_0} > r_0^2 + a^2 - 2ar_0 = (r_0 - a)^2 > 0$. Hence

$$r_0 > r_{\rm ph}^+ \quad \Longrightarrow \quad \ell_+ > 0. \tag{11.148}$$

But for $r_0 < r_{\rm ph}^*$, i.e. in region $\mathscr{M}_{\rm III}$, we may have $\ell_+ < 0$ (cf. Fig. 11.12). In view of this and the result (11.133), we introduce the following nomenclature:

- orbits of the (ε_+, ℓ_+) family with $r_0 > r_{\rm ph}^+$ are called ***prograde outer circular orbits***;
- orbits of the (ε_+, ℓ_+) family with $0 < r_0 < r_{\rm ph}^*$ are called ***inner circular orbits***;
- all orbits of the (ε_-, ℓ_-) family are called ***retrograde outer circular orbits***.

We note from Figs. 11.11 and 11.12, or from formulas (11.141) and (11.147), that both ε and ℓ are diverging at the boundaries of the various domains of existence of circular orbits. We shall comment further on this behavior at the end of Sec. 11.5.4.

Figure 11.13 shows ε in terms of ℓ for the three families of circular orbits, with the indication of the stability of the various branches, as determined in the next section.

11.5.3 Stability of circular timelike orbits

Latitudinal stability

A natural question regarding the stability of equatorial circular orbits is whether these orbits are stably confined into the equatorial plane $\theta = \pi/2$. The answer is very simple:

All circular timelike orbits are stable with respect to any perturbation away from the equatorial plane.

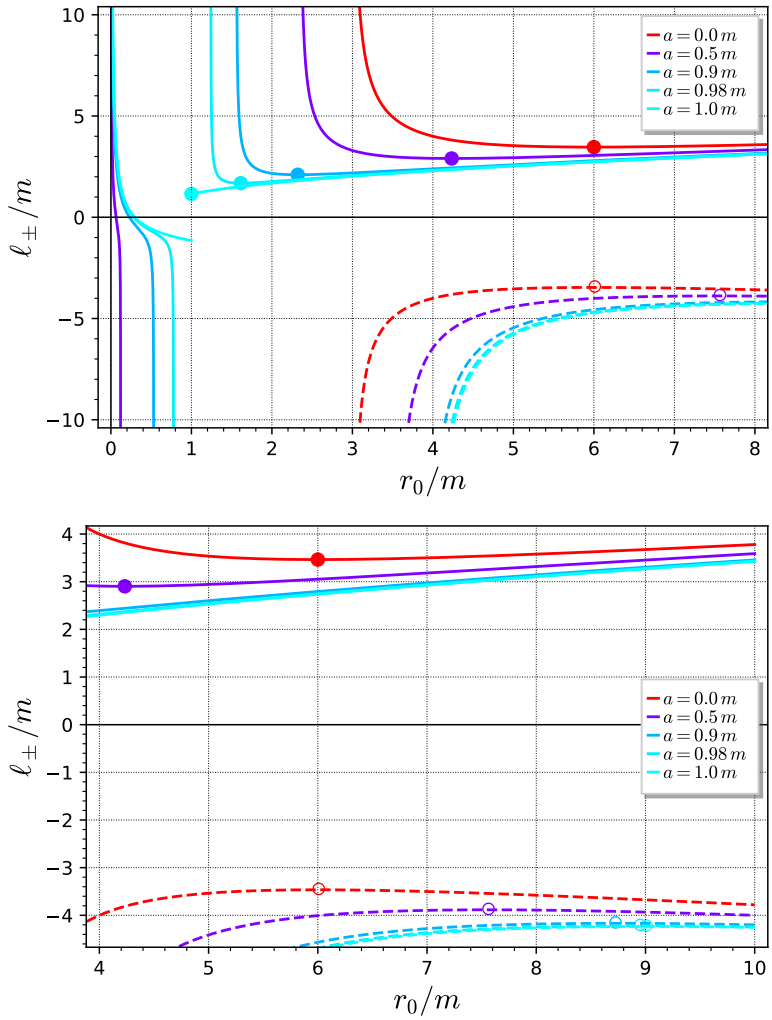


Figure 11.12: Specific conserved angular momentum $\ell = \ell_+$ (solid curves) or $\ell = \ell_-$ (dashed curves) along circular timelike orbits in the equatorial plane, as a function of the orbital radius r_0 [Eq. (11.147)], for selected values of the Kerr spin parameter a. Curves in the inner region ($r_0 < m$) terminate by a vertical asymptote at $r = r^*_{\rm ph}(a)$ given by Eq. (11.135), while the other curves start along a vertical asymptote at $r = r^\pm_{\rm ph}(a)$ given by Eq. (11.134). Dots on ℓ_+ curves and open circles on ℓ_- ones mark the ISCO: all configurations on the left of these points are unstable. The bottom panel is a zoom on the region $4\,m \leq r_0 \leq 10\,m$. *[Figure generated by the notebook D.4.9]*

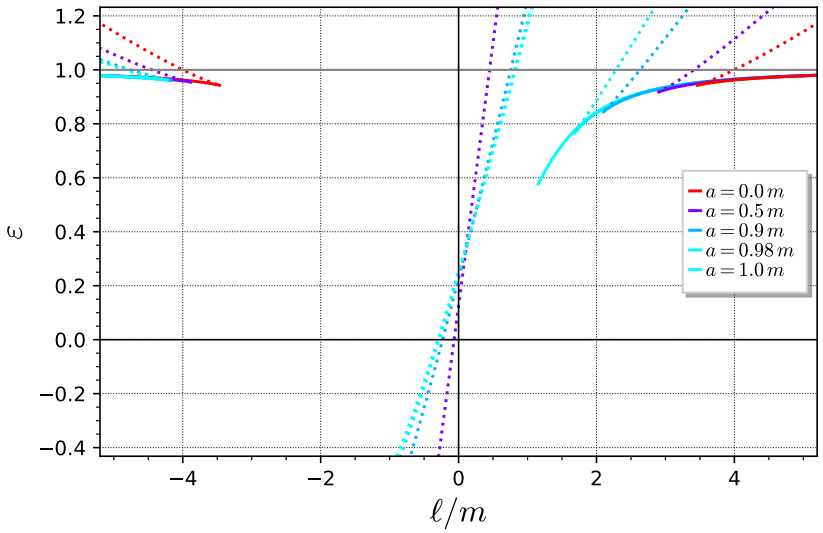


Figure 11.13: Circular timelike orbits in the (ℓ, ε) plane, ℓ being the specific conserved angular momentum and ε the specific conserved energy. Solid (resp. dotted) curves correspond to stable (unstable) orbits. Retrograde outer circular orbits are on the left side, inner circular orbits in the middle and prograde outer circular orbits on the right side. The cusps in the left and right curves correspond to ISCOs. *[Figure generated by the notebook D.4.9]*

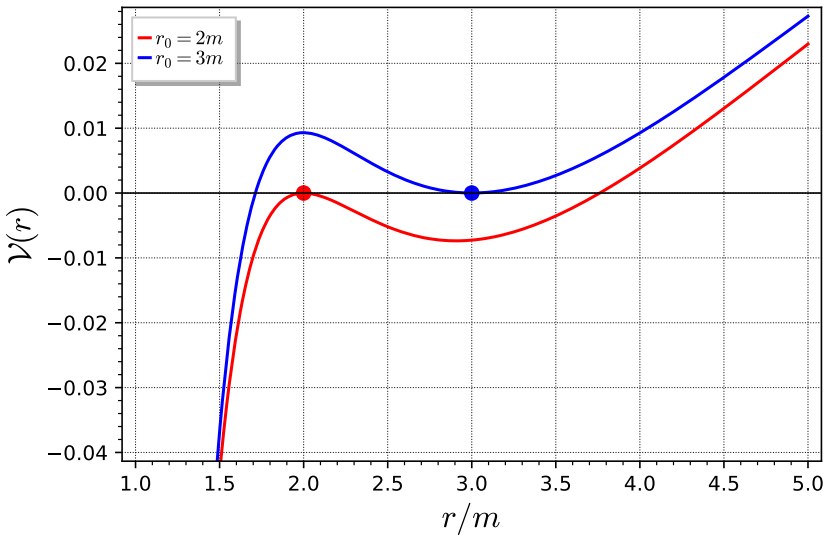


Figure 11.14: Function $\mathcal{V}(r)$, as defined by Eq. (11.121), with $a = 0.9\,m$, $\varepsilon = \varepsilon_+(r_0)$ [Eq. (11.141)], $\ell = \ell_+(r_0)$ [Eq. (11.147)] for $r_0 = 2\,m$ (red curve) and $r_0 = 3\,m$ (blue curve). By definition of ε_+ and ℓ_+, one has both $\mathcal{V}(r_0) = 0$ and $\mathcal{V}'(r_0) = 0$. *[Figure generated by the notebook D.4.9]*

Proof. All equatorial orbits have a vanishing Carter constant Q [cf. Eq. (11.119)] and we have seen in Sec. 11.3.6 that a $Q = 0$ geodesic stays stably at $\theta = \pi/2$ iff $a^2(E^2-\mu^2) \leq L^2$, i.e. iff

$$a^2(\varepsilon^2 - 1) \leq \ell^2 \tag{11.149}$$

for a timelike geodesic. Now, for a circular orbit of radius r_0, Eq. (11.125b) implies

$$\ell^2 - a^2(\varepsilon^2 - 1) = mr_0 + 3m(\ell - a\varepsilon)^2/r_0 > 0,$$

where the inequality follows from $r_0 > 0$ [Eq. (11.129)]. Hence the stability condition (11.149) is always fulfilled. □

Radial stability

Let us now investigate the stability with respect to a radial perturbation. A circular orbit at $r = r_0$ obeys both $\mathcal{V}(r_0) = 0$ and $\mathcal{V}'(r_0) = 0$ [Eqs. (11.123) and (11.124)]. The latter property means that it corresponds to a local extremum of the function $\mathcal{V}$. This extremum can be either a local maximum (red curve in Fig. 11.14) or a local minimum (blue curve in Fig. 11.14). Now, in view of Eq. (11.120b) rewritten as $\mathcal{V}(r) = -(\mathrm{d}r/\mathrm{d}\tau)^2$, any motion in the equatorial plane must obey $\mathcal{V}(r) \leq 0$. If the circular orbit corresponds to a local minimum, then for r close to r_0, but distinct from it, $\mathcal{V}(r) > 0$, since $\mathcal{V}(r_0) = 0$ (cf. the blue curve in Fig. 11.14). This means that no geodesic motion with the same values of the conserved quantities ε and ℓ is possible in the vicinity of r_0 except for precisely $r = r_0$. We conclude that the circular orbit at r_0 is stable in that case. On the contrary, when r_0 is a local maximum of $\mathcal{V}$, one has $\mathcal{V}(r) < 0$ for r close to r_0, but distinct from it (cf. the red curve in Fig. 11.14). Motion away from r_0 is then possible for the same values of ε and ℓ; we conclude that the circular orbit is unstable in that case. Assuming that $\mathcal{V}''(r_0) \neq 0$, a local minimum (resp. maximum) is equivalent to $\mathcal{V}''(r_0) > 0$ (resp. $\mathcal{V}''(r_0) < 0$). We have then

$$\text{The circular orbit of radius } r_0 \text{ is stable} \iff \mathcal{V}''(r_0) > 0. \tag{11.150}$$

From the definition (11.121) of $\mathcal{V}(r)$ and the relation $\ell = \tilde{\ell} + a\varepsilon$ [Eq. (11.126)], we get

$$\mathcal{V}''(r_0) = -\frac{4m}{r_0^3} + \frac{6(\tilde{\ell}^2 + 2a\varepsilon\tilde{\ell} + a^2)}{r_0^4} - \frac{24m\tilde{\ell}^2}{r_0^5}.$$

Substituting Eq. (11.127b) for $2a\varepsilon\tilde{\ell}$, we get a simple expression, involving only $\tilde{\ell}$:

$$\mathcal{V}''(r_0) = \frac{2m}{r_0^3}\left(1 - 3\frac{\tilde{\ell}^2}{r_0^2}\right).$$

For a circular orbit, $\tilde{\ell}$ is the function (11.142) of r_0. Using it, we get

$$\mathcal{V}''(r_0) = \frac{2m(r_0^2 - 6mr_0 \pm 8a\sqrt{mr_0} - 3a^2)}{r_0^3(r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0})}.$$

Since the denominator of the right-hand-side expression is always positive, by virtue of Eqs. (11.129) and (11.132), the sign of $\mathcal{V}''(r_0)$ is entirely determined by the numerator, so that we may rewrite (11.150) as

$$\begin{aligned} \text{The circular orbit of radius } r_0 \text{ is stable} \quad &\iff \quad r_0^2 - 6mr_0 \pm 8a\sqrt{mr_0} - 3a^2 > 0 \\ &\iff \quad x^4 - 6x^2 \pm 8\bar{a}x - 3\bar{a}^2 > 0, \end{aligned} \tag{11.151}$$

where we have introduced the dimensionless variables

$$x := \sqrt{\frac{r_0}{m}} \qquad \text{and} \qquad \bar{a} := \frac{a}{m}. \tag{11.152}$$

The problem amounts to finding the range of x where the quartic polynomial $P(x) := x^4 - 6x^2 \pm 8\bar{a}x - 3\bar{a}^2$ is positive. This requires computing the roots of P. We shall do it via Ferrari's method. The first step is to introduce a parameter Z_1 and rewrite $P(x)$ as

$$P(x) = (x^2 - Z_1)^2 - T(x), \quad \text{with} \quad T(x) := 2(3 - Z_1)x^2 \mp 8\bar{a}x + Z_1^2 + 3\bar{a}^2. \tag{11.153}$$

The above expression is an identity, which holds for any value of Z_1; the core of Ferrari's method it to find Z_1 so that the quadratic polynomial $T(x)$ has a double root, x_0 say. We will have then $T(x) = S(x)^2$, with $S(x) := \sqrt{2(3 - Z_1)}(x - x_0)$, and $P(x) = (x^2 - Z_1)^2 - S(x)^2 = (x^2 - Z_1 - S(x))(x^2 - Z_1 + S(x))$, so that

$$P(x) = 0 \quad \iff \quad x^2 - Z_1 - S(x) = 0 \quad \text{or} \quad x^2 - Z_1 + S(x) = 0. \tag{11.154}$$

In other words, the four (possibly complex) solutions of the quartic equation $P(x) = 0$ are the two solutions of the quadratic equation $x^2 - Z_1 - S(x) = 0$ plus the two solutions of $x^2 - Z_1 + S(x) = 0$. A necessary and sufficient condition for $T(x)$ to have a double root is that its discriminant vanishes, which is equivalent to

$$Z_1^3 - 3Z_1^2 + 3\bar{a}^2 Z_1 - \bar{a}^2 = 0. \tag{11.155}$$

We have thus to solve a cubic equation in Z_1. Let us reduce it to a depressed cubic equation (i.e. an equation free of any square term) via the change of variable $Z_1 =: Z + 1$:

$$Z^3 + 3(\bar{a}^2 - 1)Z + 2(\bar{a}^2 - 1) = 0.$$

The discriminant of this cubic equation is $\Delta = -(4p^3 + 27q^2)$, where $p := 3(\bar{a}^2 - 1)$ and $q := 2(\bar{a}^2 - 1)$. We get $\Delta = -108\bar{a}^2(1 - \bar{a}^2)^2$. Hence $\Delta < 0$: there exist only one real solution. It is given by Cardano's formula:

$$Z = \sqrt[3]{\frac{1}{2}\left(-q + \sqrt{\frac{-\Delta}{27}}\right)} + \sqrt[3]{\frac{1}{2}\left(-q - \sqrt{\frac{-\Delta}{27}}\right)} = \sqrt[3]{1 - \bar{a}^2}\left(\sqrt[3]{1 + \bar{a}} + \sqrt[3]{1 - \bar{a}}\right).$$

Hence

$$Z_1 = 1 + \sqrt[3]{1 - \bar{a}^2}\left(\sqrt[3]{1 + \bar{a}} + \sqrt[3]{1 - \bar{a}}\right). \tag{11.156}$$

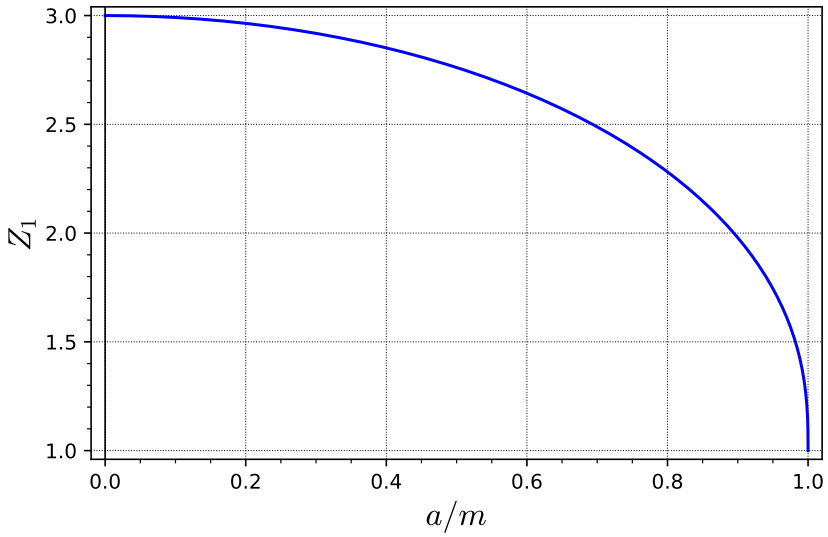


Figure 11.15: Function Z_1, as defined by Eq. (11.156). *[Figure generated by the notebook D.4.9]*

Z_1 is plotted in terms of $\bar{a} = a/m$ in Fig. 11.15. In particular, we notice that $1 \leq Z_1 \leq 3$. When Z_1 takes the value (11.156), the double root of $T(x)$ is $x_0 = \pm 2a/(3 - Z_1)$, where the $\pm$ is the opposite of the $\mp$ in the definition (11.153) of $T(x)$, and therefore indicates which family of circular orbits, among (ε_+, ℓ_+) or (ε_-, ℓ_-), is considered. We have then

$$S(x) = \sqrt{2(3 - Z_1)}\, x \mp \frac{2\sqrt{2}\bar{a}}{\sqrt{3 - Z_1}}.$$

When $\bar{a} \to 0$, the ratio $\bar{a}/\sqrt{3 - Z_1}$ is of the undetermined type "0/0". We can rearrange it by noticing the identity $8\bar{a}^2/(3 - Z_1) = Z_1^2 + 3\bar{a}^2$, which is a direct consequence of Eq. (11.155). We then write

$$S(x) = \sqrt{2(3 - Z_1)}\, x \mp Z_2, \quad \text{with} \quad Z_2 := \sqrt{Z_1^2 + 3\bar{a}^2}.$$

The solutions of $P(x) = 0$ are obtained by solving the two quadratic equations [cf. Eq. (11.154)]

$$x^2 - \sqrt{2(3 - Z_1)}\, x - Z_1 \pm Z_2 = 0 \tag{11.157a}$$
$$x^2 + \sqrt{2(3 - Z_1)}\, x - Z_1 \mp Z_2 = 0. \tag{11.157b}$$

Moreover, physically acceptable solutions must obey $x > 0$ (recall that $x := \sqrt{r_0/m}$). The discriminant of (11.157a) is $\Delta = 2(3 + Z_1 \mp Z_2)$. It is non-negative only when $\mp = +$, i.e. for orbits in the (ϵ_-, ℓ_-) family. The positive solution of (11.157a) is then

$$x_- = \frac{1}{\sqrt{2}} \left(\sqrt{3 + Z_1 + 2Z_2} + \sqrt{3 - Z_1} \right).$$

On the other side, the discriminant of (11.157b) is $\Delta = 2(3+Z_1 \pm 2Z_2)$. It is non-negative only when $\pm = +$, i.e. for orbits in the (ϵ_+, ℓ_+) family. The positive solution of (11.157b) is then

$$x_+ = \frac{1}{\sqrt{2}}\left(\sqrt{3+Z_1+2Z_2} - \sqrt{3-Z_1}\right).$$

We conclude that the quartic polynomial $P(x)$ has a single positive root, which is

$$x_\pm = \frac{1}{\sqrt{2}}\left(\sqrt{3+Z_1+2Z_2} \mp \sqrt{3-Z_1}\right).$$

Going back to $r_0 = mx^2$, we get the single solution of $\mathcal{V}''(r_0) = 0$ on $(0, +\infty)$:

$$\boxed{r_{\rm ISCO}^\pm = m\left[3 + Z_2 \mp \sqrt{(3-Z_1)(3+Z_1+2Z_2)}\right]}, \tag{11.158}$$

$$Z_1 := 1 + \sqrt[3]{1-\bar{a}^2}\left(\sqrt[3]{1+\bar{a}} + \sqrt[3]{1-\bar{a}}\right); \qquad Z_2 := \sqrt{Z_1^2 + 3\bar{a}^2}; \qquad \bar{a} := a/m.$$

where $\pm$ indicates which family among (ε_+, ℓ_+) and (ε_-, ℓ_-) is considered and *ISCO* stands for ***innermost stable circular orbit***. Indeed, $r_{\rm ISCO}^\pm$ being the unique zero of $r_0^2 - 6mr_0 \pm 8a\sqrt{mr_0} - 3a^2$, (11.151) is equivalent to

$$\text{A circular orbit of radius } r_0 \text{ is radially stable} \iff r_0 > r_{\rm ISCO}^\pm. \tag{11.159}$$

Note that $r_{\rm ISCO}^\pm/m$ is a function of $\bar{a} := a/m$ only; this function is depicted in Fig. 11.16. We notice immediately from it that

$$r_{\rm ph}^* < r_{\rm ISCO}^+ \tag{11.160}$$

and conclude that all the inner circular orbits are unstable. We also see on Fig. 11.16 that in the limit $a = 0$, $r_{\rm ISCO}^+ = r_{\rm ISCO}^- = 6m$, i.e. we recover the Schwarzschild ISCO discussed in Sec. 7.3.3.

Remark 4: The ISCO is also called ***marginally stable circular orbit*** by some authors (e.g. [19]), $r_{\rm ISCO}$ is then denoted by $r_{\rm ms}$.

Example 12: For $a = 0.9\,m$, Eq. (11.158) yields $r_{\rm ISCO}^+ = 2.32088\,m$. Accordingly, the prograde circular orbit at $r_0 = 3\,m$ is stable, while that at $r_0 = 2m$ is unstable, in agreement with the plot of $\mathcal{V}(r)$ in Fig. 11.14.

An interesting property of the ISCO is that it corresponds to extrema of $\varepsilon_\pm(r_0)$ and $\ell_\pm(r_0)$:

Among all prograde outer circular orbits, the ISCO is that for which the functions $\varepsilon_+(r_0)$ and $\ell_+(r_0)$ are minimal. Similarly, among all retrograde outer circular orbits, the ISCO is that for which the function $\varepsilon_-(r_0)$ is minimal and the function $\ell_-(r_0)$ is maximal.

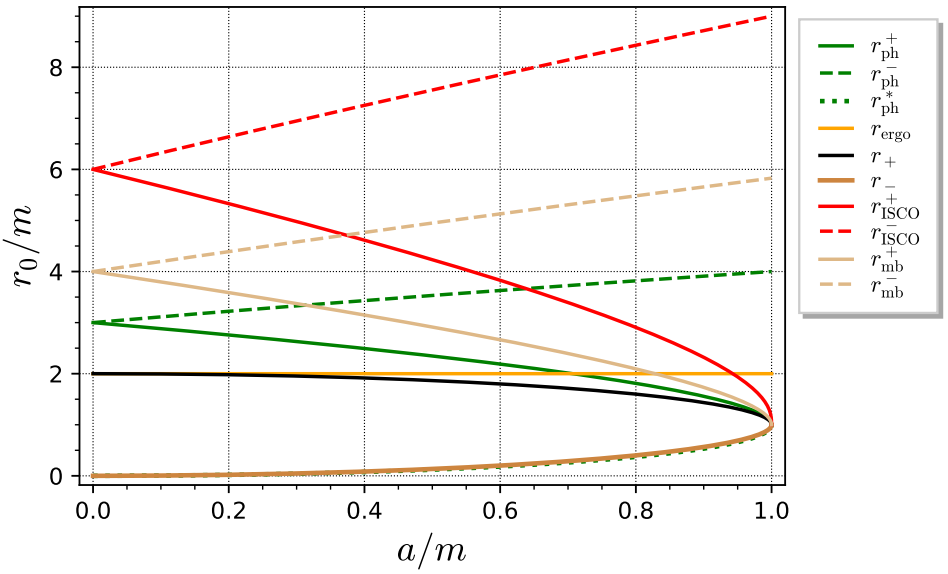


Figure 11.16: Various critical radii for circular orbits as functions of the Kerr spin parameter a. The red curves correspond to the innermost stable circular orbit $r^\pm_{\rm ISCO}$ [Eq. (11.158)], the light brown ones to the marginally bound circular orbit $r^\pm_{\rm mb}$ [Eq. (11.172)], while the other curves are the same as in Fig. 11.10. *[Figure generated by the notebook D.4.9]*

Proof. In what precedes, we have considered $\mathcal{V}$, defined by Eq. (11.121), as a function of r only. Let us consider it instead as a function of (r, ε, ℓ): $\mathcal{V} = \mathcal{V}(r, \varepsilon, \ell)$. What we have denoted by $\mathcal{V}'(r)$ is then $\partial\mathcal{V}/\partial r$ and Eqs. (11.123) and (11.124) can be rewritten as

$$\mathcal{V}\left(r_0,\ \varepsilon_\pm(r_0),\ \ell_\pm(r_0)\right) = 0 \tag{11.161a}$$

$$\frac{\partial \mathcal{V}}{\partial r}\left(r_0,\ \varepsilon_\pm(r_0),\ \ell_\pm(r_0)\right) = 0. \tag{11.161b}$$

These equations are valid for any circular orbit, i.e. any value of r_0. Let us take the derivative of Eq. (11.161a) with respect to r_0. Using the chain rule, we get

$$\underbrace{\left.\frac{\partial \mathcal{V}}{\partial r}\right|_0}_{=0} + \left.\frac{\partial \mathcal{V}}{\partial \epsilon}\right|_0 \varepsilon'_\pm(r_0) + \left.\frac{\partial \mathcal{V}}{\partial \ell}\right|_0 \ell'_\pm(r_0) = 0,$$

where $|_0$ means that the quantity is evaluated at $(r, \varepsilon, \ell) = (r_0,\ \varepsilon_\pm(r_0),\ \ell_\pm(r_0))$ and the vanishing of the first term results from Eq. (11.161b). Similarly, deriving Eq. (11.161b) with respect to r_0 yields

$$\left.\frac{\partial^2 \mathcal{V}}{\partial r^2}\right|_0 + \left.\frac{\partial^2 \mathcal{V}}{\partial \epsilon \partial r}\right|_0 \varepsilon'_\pm(r_0) + \left.\frac{\partial^2 \mathcal{V}}{\partial \ell \partial r}\right|_0 \ell'_\pm(r_0) = 0.$$

Now, at the ISCO, one has precisely $\partial^2\mathcal{V}/\partial r^2|_0 = 0$ ($\mathcal{V}''(r_0) = 0$ in the preceding notations). Hence, at the ISCO, the following two equations must hold:

$$\begin{cases} \left.\frac{\partial \mathcal{V}}{\partial \epsilon}\right|_0 \varepsilon'_\pm(r_0) + \left.\frac{\partial \mathcal{V}}{\partial \ell}\right|_0 \ell'_\pm(r_0) = 0 \\ \left.\frac{\partial^2 \mathcal{V}}{\partial \epsilon \partial r}\right|_0 \varepsilon'_\pm(r_0) + \left.\frac{\partial^2 \mathcal{V}}{\partial \ell \partial r}\right|_0 \ell'_\pm(r_0) = 0 \end{cases} \tag{11.162}$$

This constitutes a linear homogeneous system for the two unknowns $\big(\varepsilon'_\pm(r_0),\ \ell'_\pm(r_0)\big)$. Since there are no obvious reason for its determinant to vanish, we deduce that the only solution is $\big(\varepsilon'_\pm(r_0),\ \ell'_\pm(r_0)\big) = (0,0)$. Hence the ISCO realizes an extremum of both $\varepsilon_\pm(r_0)$ and $\ell_\pm(r_0)$. A quick look at Figs. 11.11 and 11.12, especially their bottom panels, enables us to conclude that the extremum is a minimum for $\varepsilon_\pm(r_0)$ and $\ell_+(r_0)$ and a maximum for $\ell_-(r_0)$. □

Because it is an extremum of both $\varepsilon(r_0)$ and $\ell(r_0)$, the ISCO is located at the cusp in the (ℓ,ε) curves shown in Fig. 11.13.

Summary

Equatorial circular timelike orbits exist only in the region $r > 0$ of Kerr spacetime. They are all stable with respect to perturbations away from the equatorial plane. They are of three kinds:

1. The *prograde outer circular orbits*: they are located outside the black hole event horizon (i.e. in region $\mathscr{M}_{\rm I}$), having a Boyer-Lindquist radius r_0 ranging from $r^+_{\rm ph} = 4m\cos^2[\arccos(-a/m)/3]$ [Eq. (11.134)] to $+\infty$; the minimal radius $r^+_{\rm ph}$ decreases with a monotonically from $3m$ ($a=0$) to m ($a=m$). These orbits are radially unstable for $r^+_{\rm ph} < r_0 \leq r^+_{\rm ISCO}$ and stable for $r_0 > r^+_{\rm ISCO}$, where $r^+_{\rm ISCO}$ is given by Eq. (11.158), decreasing monotonically from $6m$ ($a=0$) to m ($a=m$). Their specific conserved energy and angular momentum are $\varepsilon = \varepsilon_+(r_0) > 0$ and $\ell = \ell_+(r_0) > 0$, with $\varepsilon_+(r_0)$ given by Eq. (11.141) and $\ell_+(r_0)$ by Eq. (11.147), both being plotted as the solid curves in the region $r_0 \geq m$ of Figs. 11.11 and 11.12. The functions $\varepsilon_+(r_0)$ and $\ell_+(r_0)$ are minimal at the ISCO.

2. The *retrograde outer circular orbits*: they are located outside the black hole event horizon (i.e. in region $\mathscr{M}_{\rm I}$), having a Boyer-Lindquist radius r_0 ranging from $r^-_{\rm ph} = 4m\cos^2[\arccos(a/m)/3]$ [Eq. (11.134)] to $+\infty$; the minimal radius $r^-_{\rm ph}$ increases with a monotonically from $3m$ ($a=0$) to $4m$ ($a=m$). These orbits are radially unstable for $r^-_{\rm ph} < r_0 \leq r^-_{\rm ISCO}$ and stable for $r_0 > r^-_{\rm ISCO}$, where $r^-_{\rm ISCO}$ is given by Eq. (11.158), increasing monotonically from $6m$ ($a=0$) to $9m$ ($a=m$). Their specific conserved energy and angular momentum are $\varepsilon = \varepsilon_-(r_0) > 0$ and $\ell = \ell_-(r_0) < 0$, with $\varepsilon_-(r_0)$ given by Eq. (11.141) and $\ell_-(r_0)$ by Eq. (11.147), both being plotted as the dashed curves in Figs. 11.11 and 11.12. The functions $\varepsilon_-(r_0)$ and $|\ell_-(r_0)|$ are minimal at the ISCO.

3. The *inner circular orbits*: they are located inside the inner horizon (in the part $r > 0$ of region $\mathscr{M}_{\rm III}$), having a Boyer-Lindquist radius r_0 ranging from 0 (the ring singularity) to $r^*_{\rm ph} = 4m\cos^2[\arccos(-a/m)/3 + 4\pi/3]$ [Eq. (11.135)]; the maximal radius $r^*_{\rm ph}$ increases with a monotonically from 0 ($a=0$) to m ($a=m$). These orbits are all radially unstable. Their specific conserved energy and angular momentum are $\varepsilon = \varepsilon_+(r_0)$ and $\ell = \ell_+(r_0)$, with $\varepsilon_+(r_0)$ given by

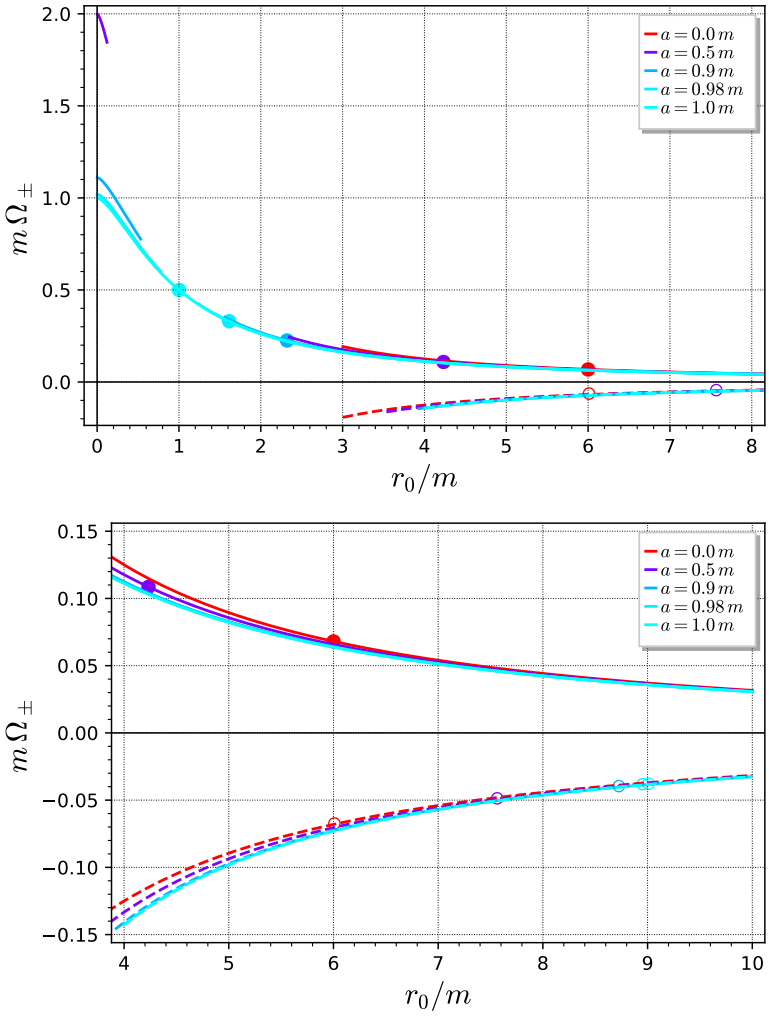


Figure 11.17: Angular velocity $\Omega := \mathrm{d}\varphi/\mathrm{d}t = \Omega_+$ (solid curves) or $\Omega = \Omega_-$ (dashed curves) along circular timelike orbits in the equatorial plane, as a function of the orbital radius r_0 [Eq. (11.163)], for selected values of the Kerr spin parameter a. Dots on Ω_+ curves and open circles on Ω_- ones mark the ISCO: all configurations on the left of these points are unstable. The bottom panel is a zoom on the region $4\,m \leq r_0 \leq 10\,m$. *[Figure generated by the notebook D.4.9]*

Eq. (11.141) and $\ell_+(r_0)$ by Eq. (11.147), both being plotted as the solid curves in the region $r_0 \leq m$ of Figs. 11.11 and 11.12.

In particular, there are no equatorial circular orbits in region $\mathscr{M}_{\mathrm{II}}$ of Kerr spacetime.

11.5.4 4-velocity and angular velocities

The orbiting angular velocity as seen by an asymptotic inertial observer is $\Omega := \mathrm{d}\varphi/\mathrm{d}t|_{\mathscr{L}}$ (cf. Sec. 7.3.3 for the Schwarzschild case and Sec. 10.7.1 for the extension to Kerr space-

time). We evaluate it by combining Eqs. (11.120a) and (11.120c):

$$\Omega = \frac{\mathrm{d}\varphi}{\mathrm{d}\tau} \times \frac{\mathrm{d}\tau}{\mathrm{d}t} = \frac{\left(1 - \frac{2m}{r_0}\right)\ell + \frac{2am\varepsilon}{r_0}}{(r_0^2 + a^2)\varepsilon + \frac{2am}{r_0}(a\varepsilon - \ell)}.$$

Using $\tilde{\ell} = \ell - a\varepsilon$ [Eq. (11.126)] instead of ℓ, we get

$$\Omega = \frac{(r_0 - 2m)\tilde{\ell} + ar_0\varepsilon}{r_0(r_0^2 + a^2)\varepsilon - 2am\tilde{\ell}}.$$

Substituting ε and $\tilde{\ell}$ by the actual values $\varepsilon_\pm$ and $\tilde{\ell}_\pm$ taken on a circular orbit [Eqs. (11.141) and (11.142)], we obtain, after simplification,

$$\boxed{\Omega_\pm = \pm\frac{\sqrt{m}}{r_0^{3/2} \pm a\sqrt{m}}}, \tag{11.163}$$

where Ω_+ is the value of Ω for prograde outer circular orbits and inner circular orbits and Ω_- is the value for retrograde outer circular orbits. Ω_+ and Ω_- are drawn in term of r_0 in Fig. 11.17. Note that for $r_0 \gtrsim 7\,m$, the effect of spin parameter a on $\Omega_\pm$ is hardly perceptible.

Remark 5: For $a \to 0$, Eq. (11.163) reduces to Eq. (7.60), as it should.

From the very definition (11.122) of a circular orbit in the equatorial plane, the 4-velocity $\boldsymbol{u}$ along such an orbit obeys $u^r = \mathrm{d}r/\mathrm{d}\tau = 0$ and $u^\theta = \mathrm{d}\theta/\mathrm{d}\tau = 0$. Since moreover $\Omega = \mathrm{d}\varphi/\mathrm{d}t = u^\varphi/u^t$, $\boldsymbol{\partial}_t = \boldsymbol{\xi}$ and $\boldsymbol{\partial}_\varphi = \boldsymbol{\eta}$, we can write the 4-velocity as

$$\boxed{\boldsymbol{u} = u^t\,(\boldsymbol{\xi} + \Omega_\pm \boldsymbol{\eta})}. \tag{11.164}$$

The component $u^t = \mathrm{d}t/\mathrm{d}\tau$ is given by Eq. (11.120a). Substituting in it Eq. (11.141) for ε and Eq. (11.142) for $\tilde{\ell} := \ell - a\varepsilon$, we get, after simplification:

$$u^t = \frac{r_0 \pm a\sqrt{m/r_0}}{\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}}. \tag{11.165}$$

Equation (11.164) shows that the 4-velocity is a linear combination of the two Killing vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ with coefficients that are constant along the worldline $\mathscr{L}$. An observer on a circular orbit in the equatorial plane is thus a *stationary observer* as defined in Sec. 10.7.1 [compare Eq. (10.106)]. He does not notice any change in the spacetime geometry. We shall call him a ***circular geodesic observer***. Contrary to the families of stationary observers considered in Sec. 10.7, a circular geodesic observer is in free fall: by construction, his worldline is a geodesic, so that his 4-acceleration is zero.

The orbital velocity measured by the circular geodesic observer himself is $\Omega_{\mathscr{P}} = \mathrm{d}\varphi/\mathrm{d}\tau = \mathrm{d}\varphi/\mathrm{d}t \times \mathrm{d}t/\mathrm{d}\tau = u^t\Omega$. Using the values (11.163) and (11.165), we get

$$\Omega_{\mathscr{P}}^\pm = \pm\frac{\sqrt{m} \pm amr_0^{-3/2}}{(\sqrt{r_0} \pm a\sqrt{m}/r_0)\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}}. \tag{11.166}$$

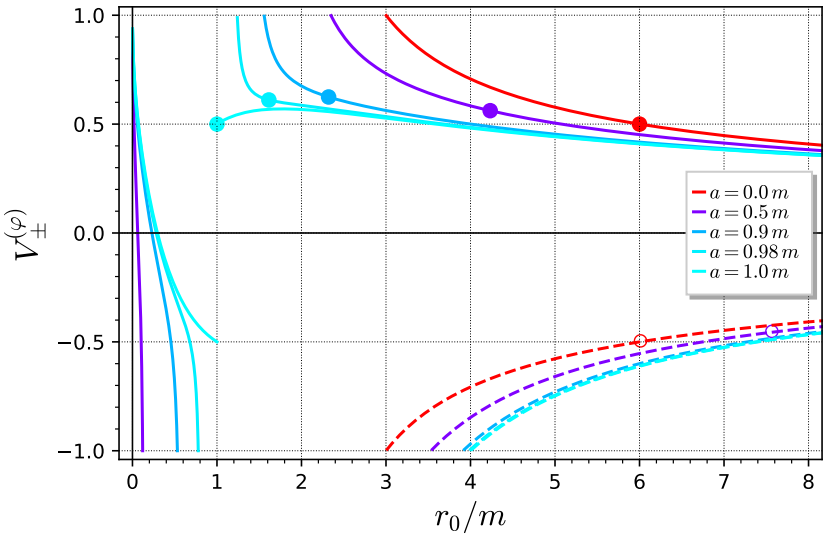


Figure 11.18: Component $V_+^{(\varphi)}$ (solid curves) or $V_-^{(\varphi)}$ (dashed curves) of the velocity $\boldsymbol{V} = V_\pm^{(\varphi)}\,\boldsymbol{e}_{(\varphi)}$ of a particle moving along a circular orbit as measured by the ZAMO [Eq. (11.169)] for selected values of the Kerr spin parameter a. Dots on $V_+^{(\varphi)}$ curves and open circles on $V_-^{(\varphi)}$ ones mark the ISCO: all configurations on the left of these points are unstable. *[Figure generated by the notebook D.4.9]*

The orbital period measured by the circular geodesic observer is $T_{\mathscr{P}}^{\pm} = 2\pi/|\Omega_{\mathscr{P}}^{\pm}|$.

It is instructive to evaluate the velocity $\boldsymbol{V}$ of a particle $\mathscr{P}$ on a circular orbit as measured by a zero-angular momentum observer (ZAMO) (cf. Sec. 10.7.3). Let us first note that all circular orbits are within the domain of ZAMOs, $\mathscr{M}_{\rm ZAMO} = \mathscr{M}_{\rm I} \cup (\mathscr{M}_{\rm III} \setminus \mathscr{T})$ [Eq. (10.111)]. Indeed, the outer circular orbits are in $\mathscr{M}_{\rm I}$ and the inner ones are in the part $r > 0$ of $\mathscr{M}_{\rm III}$, while the Carter time machine $\mathscr{T}$ is located in the part $r < 0$ (cf. Sec. 10.2.5). Let us then rewrite formula (11.164) for $\mathscr{P}$'s 4-velocity $\boldsymbol{u}$ by expressing $\boldsymbol{\xi}$ in terms of the ZAMO's 4-velocity $\boldsymbol{n}$, the lapse N and the shift vector $\boldsymbol{\beta}$ via Eq. (10.117):

$$\boldsymbol{u} = u^t\,(N\boldsymbol{n} + \boldsymbol{\beta} + \Omega_\pm \boldsymbol{\eta}) = u^t\,[N\boldsymbol{n} + (\Omega_\pm - \omega)\,\boldsymbol{\eta}] = N u^t \left[\boldsymbol{n} + \frac{1}{N}(\Omega_\pm - \omega)\,\boldsymbol{\eta}\right],$$

where the second equality results from $\boldsymbol{\beta} = \beta^\varphi \boldsymbol{\partial}_\varphi = \beta^\varphi \boldsymbol{\eta} = -\omega\boldsymbol{\eta}$, ω being the rotation angular velocity of the ZAMO seen from infinity [Eq. (10.119)]. Since $\boldsymbol{n}\cdot\boldsymbol{\eta} = 0$, the above formula constitute the orthogonal decomposition of $\boldsymbol{u}$ with respect to the ZAMO's 4-velocity $\boldsymbol{n}$, which we can compare to the generic formula (1.34) and thereby conclude that the velocity of $\mathscr{P}$ with respect to the ZAMO is

$$\boldsymbol{V} = \frac{1}{N}(\Omega_\pm - \omega)\,\boldsymbol{\eta} \tag{11.167}$$

and the Lorentz factor of $\mathscr{P}$ with respect to the ZAMO is

$$\Gamma = (1 - \boldsymbol{V}\cdot\boldsymbol{V})^{-1/2} = N u^t. \tag{11.168}$$

Let us expand $\boldsymbol{V}$ on the ZAMO's orthonormal fame $(\boldsymbol{e}_{(\alpha)})$ [Eq. (10.115)]. Given relation (10.115d) between $\boldsymbol{\eta} = \boldsymbol{\partial}_\varphi$ and $\boldsymbol{e}_{(\varphi)}$ we get, after substituting Eq. (10.113) (with $\theta = \pi/2$)

for N:

$$\boldsymbol{V} = \frac{r_0^2 + a^2 + 2a^2 m/r_0}{\sqrt{r_0^2 - 2mr_0 + a^2}} (\Omega_\pm - \omega)\, \boldsymbol{e}_{(\varphi)}.$$

Finally, let us substitute Eq. (11.163) for $\Omega_\pm$ and Eq. (10.119) (with $\theta = \pi/2$) for ω. We obtain, after simplification:

$$\boxed{\boldsymbol{V} = V_\pm^{(\varphi)}\, \boldsymbol{e}_{(\varphi)}} \quad \text{with} \quad \boxed{V_\pm^{(\varphi)} = \pm \frac{\sqrt{m}\left(r_0^2 \mp 2a\sqrt{mr_0} + a^2\right)}{(r_0^{3/2} \pm a\sqrt{m})\sqrt{r_0^2 - 2mr_0 + a^2}}}. \tag{11.169}$$

As in all formulas of this section, the sign $\pm$ is $+$ for a prograde outer circular orbit or an inner circular orbit and $-$ for a retrograde outer circular orbit. The velocity $V_\pm^{(\varphi)}$ is depicted in Fig. 11.18. We note that, except for $a = m$,

$$\lim_{r_0 \to 0} V_+^{(\varphi)} = 1, \quad \lim_{r_0 \to r_{\rm ph}^*} V_+^{(\varphi)} = -1, \quad \lim_{r_0 \to r_{\rm ph}^+} V_+^{(\varphi)} = 1, \quad \lim_{r_0 \to r_{\rm ph}^-} V_-^{(\varphi)} = -1. \tag{11.170}$$

Since $\boldsymbol{e}_{(\varphi)}$ is a unit vector, this means that

- the inner orbits rotate close to the speed of light with respect to the ZAMO near the ring singularity ($r_0 \to 0$) and near the outer boundary of their domain of existence ($r_0 \to r_{\rm ph}^*$), this motion being in the prograde (resp. retrograde) direction in the first (resp. second) case;
- the outer orbits rotate at the speed of light with respect to the ZAMO near the minimal radius of their domain of existence ($r_0 \to r_{\rm ph}^\pm$).

Given that the velocity of a massive particle with respect to any observer cannot be larger than the speed of light, this provides a physical explanation for the boundaries of the domains of existence of circular orbits obtained in Sec. 11.5.2. We shall see in Sec. 12.3 that these boundaries correspond to photon circular orbits. This also provides a physical explanation why the specific conserved energy ε and the specific conserved angular momentum ℓ diverge at these boundaries, as observed in Figs. 11.11 and 11.12: $\varepsilon := E/\mu$ and $\ell := L/\mu$ must tend to $\pm\infty$ when the particle's mass μ tends to 0 (the photon limit).

The particular case $a = m$ will be discussed in Sec. ??.

11.5.5 Marginally bound circular orbit

We see on Fig. 11.11 that for $a \neq m$ and r_0 lower than a critical value, $r_{\rm mb}^+$ say, the prograde outer circular orbits have $\varepsilon > 1$ (cf. the grey horizontal line in Fig. 11.11), which is equivalent to $E > \mu$; they are thus unbound (cf. Sec. 11.3.7). Conversely, all orbits with $r_0 > r_{\rm mb}^+$ are bound. Similarly, but this time of all values of $a \leq m$, the retrograde outer circular orbits are bound iff r_0 is larger than a critical value, $r_{\rm mb}^-$ say. The orbit at $r_0 = r_{\rm mb}^\pm$ is called the ***marginally bound circular orbit***.

We note from Fig. 11.11 that

$$r_{\rm ph}^\pm < r_{\rm mb}^\pm < r_{\rm ISCO}^\pm. \tag{11.171}$$

Accordingly, all unbound orbits are unstable.

To evaluate $r_{\rm mb}^{\pm}$, let us solve the equation $\varepsilon_\pm(r_0) = 1$ for r_0. Using expression (11.141) for $\varepsilon_\pm(r_0)$, taking the square and simplifying, we get

$$r_0^2 - 4mr_0 \pm 4a\sqrt{mr_0} - a^2 = 0,$$

which we can write

$$r_0^2 = (2\sqrt{mr_0} \mp a)^2.$$

Since we are looking for solutions $r_0 \geq r_+ = m + \sqrt{m^2 - a^2}$, we have $2\sqrt{mr_0} \mp a > 0$, so that we can reduce the above equation to $r_0 = 2\sqrt{mr_0} \mp a$. Introducing $x := \sqrt{r_0/m}$, we end up solving the quadratic equation $x^2 - 2x \pm a/m = 0$. The only solution $x \geq 1$, which is implied by $r_0 \geq r_+$, is $x = 1 + \sqrt{1 \mp a/m}$. This yields

$$\boxed{r_{\rm mb}^{\pm} = 2m \mp a + 2\sqrt{m(m \mp a)}}. \tag{11.172}$$

$r_{\rm mb}^{\pm}$ is plotted as a function of a in Fig. 11.16.

11.5.6 Circular orbits in the ergoregion

We see in Fig. 11.16 that for a sufficiently large, prograde outer circular orbits can exist in the outer ergoregion $\mathscr{G}^+$ (cf. Sec. 10.2.4), while no retrograde outer circular orbits can exist there. In the equatorial plane, the coordinate r of the external boundary of $\mathscr{G}^+$ (the outer ergosphere) is simply $r_{\mathscr{E}^+}(\pi/2) = 2m$ [Eq. (10.19d)] — the orange horizontal line in Fig. 11.16. Thus, prograde outer circular orbits exist in the ergoregion if $r_{\rm ph}^+ < 2m$ and they can be stable if $r_{\rm ISCO}^+ < 2m$. The limiting value of a for the first case is obtained by setting $r_0 = 2m$ in the equation governing the existence of prograde circular orbits, namely Eq. (11.132) with the + sign: $(2m)^{3/2} - 3m\sqrt{2m} + 2a\sqrt{m} > 0$, from which we get immediately

$$\left(\begin{array}{c}\text{timelike circular orbits}\\ \text{exist in the outer ergoregion}\end{array}\right) \iff a > \frac{m}{\sqrt{2}} \simeq 0.707\, m. \tag{11.173}$$

Similarly, the limiting value of a for the stability of circular orbits in the ergoregion is obtained by setting $r_0 = 2m$ in Eq. (11.151) with the + sign: $(2m)^2 - 12m^2 + 8am\sqrt{2} - 3a^2 > 0$. The right-hand side being a quadratic polynomial in a, we get easily, given the constraint $a \leq m$,

$$\left(\begin{array}{c}\text{timelike circular orbits}\\ \text{exist stably in the ergoregion}\end{array}\right) \iff a > \frac{2\sqrt{2}}{3} m \simeq 0.943\, m. \tag{11.174}$$

Both in (11.173) and (11.174), the orbits referred to belong to the prograde outer family.

Remark 6: Regarding the limit (11.173), the existence is specified to be in the *outer* ergoregion, because timelike circular orbits always exist, as soon as $a > 0$, in the inner ergoregion: they are

the (unstable) inner circular orbits found in Sec. 11.5.2. Indeed, these orbits exist in the range $0 < r_0 < r^*_{\rm ph}$, which is entirely contained in the inner ergoregion: the boundaries of the latter in the equatorial plane are $r = 0$ [Eq. (10.19a)] and $r = r_-$, with $r_- > r^*_{\rm ph}$. On the contrary, in the limit (11.174), we have dropped the qualifier *outer* for the ergoregion, since there is no other place where stable circular orbits can be found, the inner circular orbits being all unstable.

Remark 7: As discussed in Sec. 11.3.1, negative-energy or zero-energy particles can exist in the outer ergoregion. However, Fig. 11.11 shows us that none of them can follow a circular orbit.

Historical note: The solutions for the circular geodesic motion in the equatorial plane for outer prograde and outer retrograde orbits have been published for the first time by James M. Bardeen, William H. Press and Saul A. Teukolsky in 1972 [19]. In his lecture notes at the famous 1972 Les Houches School [17], Bardeen says that they have been derived by Teukolsky. In the article [19], it is mentioned that they have been derived by means of computer algebra techniques.

11.6 Going further

For an extended discussion of geodesics of Kerr spacetime, including those that cross the various blocks of the maximal analytic extension presented in Sec. 10.8, see Chap. 4 of O'Neill textbook [200].

Chapter 12

Null geodesics and images in Kerr spacetime

Contents

12.1 Introduction

Having investigated the properties of generic causal geodesics in Kerr spacetime in Chap. 11, we focus here on null geodesics, with application to images of a Kerr black hole. First we discuss the main properties of null geodesics in Sec. 12.2, in great part by taking the $\mu = 0$ limit of results obtained for generic causal geodesics in Chap. 11. Then, in Sec. 12.3, we focus on null geodesics evolving at a fixed value of the coordinate r — the so-called *spherical photon orbits*. These geodesics play a crucial role in the formation of the images perceived by an observer. In particular, they are related to the key concepts of *critical curve* and *shadow* in the observer's screen, which are investigated in Sec. 12.4. Finally, we discuss the images themselves in Sec. 12.5, first by considering computed images from a simplified model of accretion disk and then by analyzing the actual image of the surroundings of the black hole M87*, as obtained recently by the Event Horizon Telescope [5].

12.2 Main properties of null geodesics

We shall distinguish the null geodesics with $E = 0$ (the so-called *zero-energy* geodesics, cf. Sec. 11.3.1) from those having $E \neq 0$. Indeed, in the latter case, we will rescale the angular momentum L and the Carter constant Q by E, so that only two constants of motion become pertinent for the study: L/E and Q/E^2. We thus treat first the particular case $E = 0$.

12.2.1 Zero-energy null geodesics

First, we note that a geodesic $\mathscr{L}$ with $E = 0$ cannot exist outside the ergoregion $\mathscr{G}$, by virtue of the result (11.56). In particular, it cannot exist far from the black hole.

Another property of $\mathscr{L}$ is to have a non-negative Carter constant:

$$\boxed{Q \geq 0}_{E=0}. \tag{12.1}$$

This follows immediately from the result of Sec. 11.3.6 stating that a necessary condition for $Q < 0$ is $a \neq 0$ and $|E| > \sqrt{\mu^2 + L^2/a^2}$. Specializing this last inequality to $\mu = 0$ and $E = 0$, we get $0 > |L|$, which is impossible.

Besides, if $\mathscr{L}$ has some part in $\mathscr{M}_{\rm I}$ (necessarily in the outer ergoregion) or in $\mathscr{M}_{\rm III}$ (necessarily in the inner ergoregion), the constraint (11.60) reduces to $L < 0$:

$$\mathscr{L} \cap (\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}) \neq \varnothing \quad \Longrightarrow \quad L < 0. \tag{12.2}$$

We shall see below that actually $L \leq 0$ for all zero-energy null geodesics, as soon as $a \neq 0$.

The equations of geodesic motion expressed in terms of the Mino parameter λ' [system (11.50)] simplify considerably for a geodesic $\mathscr{L}$ with $\mu = 0$ and $E = 0$:

$$\frac{\mathrm{d}t}{\mathrm{d}\lambda'} = -\frac{2amLr}{\Delta} \tag{12.3a}$$

$$\frac{\mathrm{d}r}{\mathrm{d}\lambda'} = \epsilon_r \sqrt{R(r)} \tag{12.3b}$$

$$\frac{\mathrm{d}\theta}{\mathrm{d}\lambda'} = \epsilon_\theta \sqrt{\Theta(\theta)} \tag{12.3c}$$

$$\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda'} = \frac{L}{\Delta \sin^2\theta}\left(r^2 - 2mr + a^2\cos^2\theta\right), \tag{12.3d}$$

with [cf. Eqs. (11.86) and (11.35)]:

$$R(r) = -(Q + L^2)r^2 + 2m(Q + L^2)r - a^2 Q \tag{12.4}$$

$$\Theta(\theta) = Q - \frac{L^2}{\tan^2\theta}. \tag{12.5}$$

By combining (12.3a) and (12.3d), we get

$$\boxed{\left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{\mathscr{L}} = \frac{2mr - r^2 - a^2\cos^2\theta}{2amr\sin^2\theta}}_{E=0}. \tag{12.6}$$

It is remarkable that this expression does not depend on L or Q; it is therefore the same for all zero-energy null geodesics. Moreover, we note that the numerator of the right-hand side is always positive or zero in the closure $\overline{\mathscr{G}}$ of the ergoregion, which is precisely defined by $2mr - r^2 - a^2\cos^2\theta \geq 0$ (cf. Sec. 10.2.4) and where $\mathscr{L}$ is necessarily confined. Since moreover $r > 0$ in $\overline{\mathscr{G}}$, we conclude that

$$\left.\frac{\mathrm{d}\varphi}{\mathrm{d}t}\right|_{\mathscr{L}} \geq 0. \tag{12.7}$$

To proceed, we shall distinguish the subcases $Q \neq 0$ and $Q = 0$.

Case $Q \neq 0$

This case actually corresponds to $Q > 0$, since $Q < 0$ is forbidden by (12.1). We set

$$\bar{L} := \frac{L}{\sqrt{Q}} \tag{12.8}$$

and rewrite expression (12.4) for $R(r)$ as

$$R(r)/Q = -(1+\bar{L}^2)r^2 + 2m(1+\bar{L}^2)r - a^2. \tag{12.9}$$

Since $1 + \bar{L}^2 \neq 0$, this is a second-order polynomial in r, the two roots of which are

$$r_{\min} = m - \sqrt{m^2 - \frac{a^2}{1+\bar{L}^2}} \qquad \text{and} \qquad r_{\max} = m + \sqrt{m^2 - \frac{a^2}{1+\bar{L}^2}}. \tag{12.10}$$

Since $m^2 \geq a^2$, the two roots are real. They are distinct except for $a = m$ and $L = 0$. The range of radial motion being determined by $R(r) \geq 0$ [Eq. (11.30)], we get

$$r_{\min} \leq r \leq r_{\max}, \tag{12.11}$$

with a turning point at $r_{\min}$ and at $r_{\max}$. Given that $r_- = m - \sqrt{m^2 - a^2}$ and $r_+ = m + \sqrt{m^2 - a^2}$ [Eq. (10.3)], we note that

$$0 \leq r_{\min} \leq r_- \leq m \leq r_+ \leq r_{\max} \leq 2m, \tag{12.12}$$

with $r_{\min} = 0$ for $a = 0$ or $\bar{L}^2 \to +\infty$, $r_{\min} = r_-$ for $L = 0$, $r_{\max} = 2m$ for $a = 0$ or $\bar{L}^2 \to +\infty$ and $r_{\max} = r_+$ for $L = 0$. If $L \neq 0$ and $a \neq 0$, then $r_{\max} > r_+$, so that $\mathscr{L}$ has a part in the outer ergoregion and (12.2) implies that $L < 0$. Hence

$$a \neq 0 \implies L \leq 0. \tag{12.13}$$

Let us consider a zero-energy null geodesic $\mathscr{L}$ emitted outward (i.e. with $\epsilon_r = +1$) from a point A in the outer ergoregion $\mathscr{G}^+$. The coordinate r increases along $\mathscr{L}$ from r_A to $r_{\max}$, which corresponds to a r-turning point. Then r decreases to r_+, which means that $\mathscr{L}$ crosses the black hole event horizon $\mathscr{H}$ and enters the region $\mathscr{M}_{\mathrm{II}}$. In all $\mathscr{M}_{\mathrm{II}}$, r keeps decreasing and reaches r_-. There $\mathscr{L}$ crosses the inner horizon $\mathscr{H}_{\mathrm{in}}$ and enters the

region $\mathscr{M}_{\rm III}$, where r continues to decrease until it reaches $r_{\rm min}$. The latter corresponding to a r-turning point, r starts to increase and reaches r_- again. There one might think that $\mathscr{L}$ crosses the inner horizon $\mathscr{H}_{\rm in}$ and enters into $\mathscr{M}_{\rm II}$. But this is impossible since $\mathscr{H}_{\rm in}$ is a 1-way membrane: it can be crossed by a causal curve from $\mathscr{M}_{\rm II}$ to $\mathscr{M}_{\rm III}$ but not in the reverse way. Moreover, r could not continue to increase into $\mathscr{M}_{\rm II}$ since r must be decreasing towards the future in all this region (this follows from the hypersurfaces $r=$ const being spacelike in $\mathscr{M}_{\rm II}$, cf. Sec. 10.5.1). The solution to this apparent puzzle is immediate as soon as one realizes that the boundary $r=r_-$ of $\mathscr{M}_{\rm III}$ is not entirely constituted by $\mathscr{H}_{\rm in}$: it also comprises a null hypersurface that separates $\mathscr{M}_{\rm III}$ from a region distinct from $\mathscr{M}_{\rm II}$ in the maximally extended Kerr spacetime, cf. Fig. 10.12. This region is a "time-reversed" copy of $\mathscr{M}_{\rm II}$ and is denoted by $\mathscr{M}_{\rm II}^{*\prime}$ in Fig. 10.12 (cf. Sec. 10.8 for details). So actually, when it reaches $r=r_-$, the null geodesic $\mathscr{L}$ enters $\mathscr{M}_{\rm II}^{*\prime}$. There r necessarily increases towards the future, at the opposite of $\mathscr{M}_{\rm II}$. It reaches then $r=r_+$, where $\mathscr{L}$ crosses a white hole horizon and emerges into the asymptotically flat region $\mathscr{M}_{\rm I}''$, as illustrated in Fig. 12.1. The region $\mathscr{M}_{\rm I}''$ is similar to $\mathscr{M}_{\rm I}$. In particular, $\mathscr{L}$ is confined into the outer ergoregion of $\mathscr{M}_{\rm I}''$, having a r-turning point at $r=r_{\rm max}$ (same value (12.10) as in $\mathscr{M}_{\rm I}$). Then a new cycle begins, with $\mathscr{L}$ entering the future event horizon of $\mathscr{M}_{\rm I}''$.

The θ-motion of $\mathscr{L}$ is constrained by $\Theta(\theta)\geq 0$ [Eq. (11.33)], which, given expression (12.5) for Θ, is equivalent to

$$\theta_{\rm m}\leq\theta\leq\pi-\theta_{\rm m}\quad\text{with}\quad\theta_{\rm m}:=\arctan(-\bar{L}). \tag{12.14}$$

Remark 1: The general formula for $\theta_{\rm m}$ in the case $a^2(E^2-\mu^2)=0$, Eq. (11.76), which holds here since $\mu=0$ and $E=0$, yields $\theta_{\rm m}=\arccos\sqrt{1/(1+\bar{L}^2)}=\arctan|\bar{L}|$. Hence we recover the above formula.

For $L=0$, one has $\theta_{\rm m}=0$, so that θ takes all values in the range $[0,\pi]$, which means that $\mathscr{L}$ crosses repeatedly the rotation axis. For $L<0$, one has $0<\theta_{\rm m}<\pi/2$ and $\mathscr{L}$ oscillates symmetrically about the equatorial plane, having two θ-turning points, at $\theta_{\rm m}$ and $\pi-\theta_{\rm m}$. Of course, we recover the general results for $Q>0$ of Sec. 11.3.6.

We can obtain r as a function of θ along $\mathscr{L}$ by evaluating the integrals in the identity (11.55a):

$$\int_{r_0}^{r}\frac{\epsilon_r\,\mathrm{d}\bar{r}}{\sqrt{R(\bar{r})}}=\int_{\theta_0}^{\theta}\frac{\epsilon_\theta\,\mathrm{d}\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}}$$

Using (12.9) and (12.5), we get on any portion of $\mathscr{L}$ where ϵ_r and ϵ_θ are constant,

$$\epsilon_r\int_{r_0}^{r}\frac{\mathrm{d}\bar{r}}{\sqrt{-(1+\bar{L}^2)\bar{r}^2+2m(1+\bar{L}^2)\bar{r}-a^2}}=\epsilon_\theta\int_{\theta_0}^{\theta}\frac{\mathrm{d}\bar{\theta}}{\sqrt{1-\bar{L}^2/\tan^2\bar{\theta}}}.$$

The changes of variables

$$x=\frac{r/m-1}{\sqrt{1-\frac{a^2}{m^2(1+\bar{L}^2)}}}\qquad\text{and}\qquad\mu=\cos\theta$$

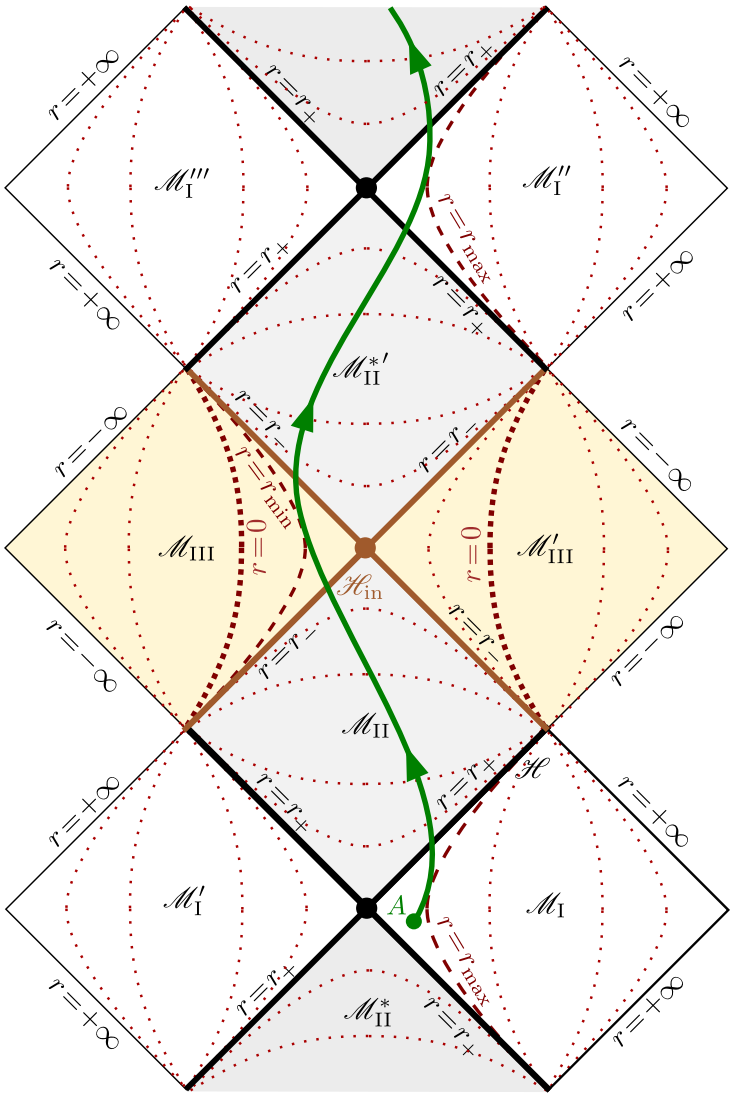


Figure 12.1: Trajectory in the extended Kerr spacetime of a null geodesic with $E = 0$, $Q > 0$ and $L < 0$, emitted from a point A in the outer ergoregion.

lead to

$$\frac{\epsilon_r}{\sqrt{1+\bar{L}^2}} \int_{x_0}^{x} \frac{\mathrm{d}\bar{x}}{\sqrt{1-\bar{x}^2}} = -\epsilon_\theta \int_{\cos\theta_0}^{\cos\theta} \frac{\mathrm{d}\mu}{\sqrt{1-(1+\bar{L}^2)\mu^2}}.$$

The integration is then immediate: $\arcsin x = -\epsilon_r\epsilon_\theta \arcsin(\sqrt{1+\bar{L}^2}\cos\theta) + K$, where K is a constant, from which we get

$$\boxed{r = m + m\sqrt{1 - \frac{a^2}{m^2(1+\bar{L}^2)}}\ \sin\left[K - \epsilon_r\epsilon_\theta \arcsin\left(\sqrt{1+\bar{L}^2}\cos\theta\right)\right]}. \tag{12.15}$$

Since $\sqrt{1+\bar{L}^2}\cos\theta_{\rm m} = 1$, we see that the constant K is related to the value of r at $\theta = \theta_{\rm m}$

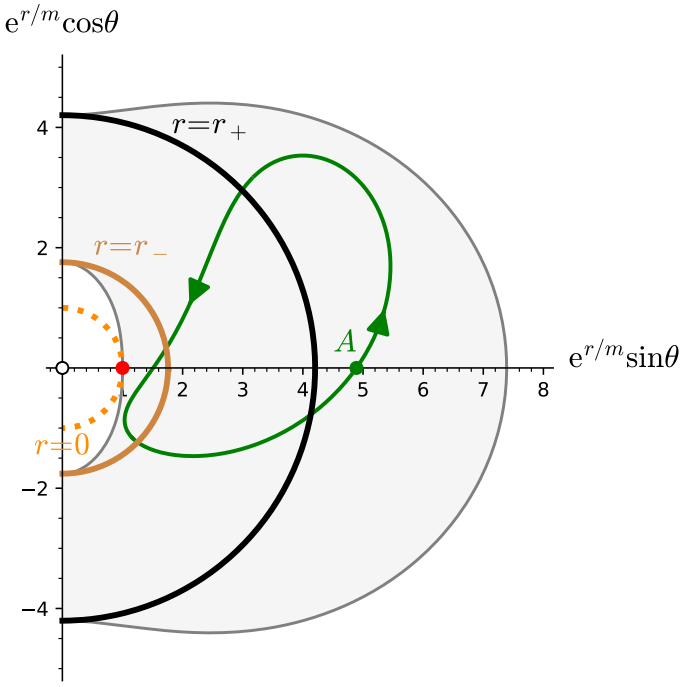


Figure 12.2: Trajectory in the meridional plane, as given by Eq. (12.15), of a null geodesic (green curve) with $E = 0$, $Q > 0$, $L = -\sqrt{Q}$ and $r(\theta_{\rm m}) = 1.5\,m$ in the Kerr spacetime with $a/m = 0.9$. The meridional plane is described in terms of O'Neill exponential coordinates $x = e^{r/m}\sin\theta$ and $z = e^{r/m}\cos\theta$, as in Figs. 10.2 – 10.4. The ergoregion is shown in grey. The black (resp. light brown) half-circle at $r = r_+$ (resp. $r = r_-$) is the trace of the outer (resp. inner) Killing horizon. The dotted orange half-circle marks the locus of $r = 0$, with the red dot indicating the curvature singularity at $r = 0$ and $\theta = \pi/2$. The area $r > r_+$ corresponds to the regions $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm I}''$ in Fig. 12.1, the area $r_- < r < r_+$ corresponds to the regions $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm II}^{*\prime}$ in Fig. 12.1 and the area $r < r_-$ corresponds to the region $\mathscr{M}_{\rm III}$ in Fig. 12.1. *[Figure generated by the notebook D.4.10]*

by

$$K = \arcsin\left(\frac{r(\theta_{\rm m})/m - 1}{\sqrt{1 - \frac{a^2}{m^2(1+\bar{L}^2)}}}\right) + \epsilon_r\epsilon_\theta\frac{\pi}{2}. \tag{12.16}$$

Note that K is not a constant all along $\mathscr{L}$, but only on portions where ϵ_r and ϵ_θ are constant. Expression 12.15 gives the trace of the zero-energy null geodesic $\mathscr{L}$ in a meridional plane. It depends on Q and L only through the ratio $\bar{L} := L/\sqrt{Q}$. It depends as well on the value of r at $\theta_{\rm m}$ via K, as it appears in Eq. (12.16). An example is shown in Fig. 12.2 for $a/m = 0.9$, $L/\sqrt{Q} = -1$ and $r(\theta_{\rm m}) = 1.5\,m$. It has $\theta_{\rm m} = \pi/4$, $r_{\rm min} \simeq 0.229\,m$ and $r_{\rm max} \simeq 1.771\,m$, while for $a/m = 0.9$, one has $r_- \simeq 0.564\,m$ and $r_+ \simeq 1.436\,m$. For concreteness, the arrows indicate some direction of motion, but depending upon some initial conditions, the opposite direction is possible. In particular, one may consider that the geodesic is the same as that shown in Fig. 12.1, being emitted outward in the outer ergoregion from a point A in the equatorial plane ($\theta = \pi/2$).

Case $Q = 0$

If the zero-energy null geodesic $\mathscr{L}$ has a vanishing Carter constant Q, Eq. (12.5) reduces to $\Theta(\theta) = -L^2/\tan^2\theta$, so that the constraint $\Theta(\theta) \geq 0$ [Eq. (11.33)] implies $L = 0$ or $\theta = \pi/2$.

In the first case, the four constants of motion μ, E, L and Q are zero. By virtue of the result (11.25), $\mathscr{L}$ is nothing but a null geodesic generator of the event horizon $\mathscr{H}$ or of the inner horizon $\mathscr{H}_{\rm in}$.

In the second case ($\theta = \pi/2$), $\mathscr{L}$ is confined to the equatorial plane. If $L = 0$, we are back to the first case: $\mathscr{L}$ is null geodesic generator of $\mathscr{H}$ or $\mathscr{H}_{\rm in}$ lying in the equatorial plane. If $L \neq 0$, the radial motion of $\mathscr{L}$ is governed by Eq. (12.3b) with the expression (12.4) of $R(r)$ reduced to

$$R(r) = L^2 r(2m - r). \tag{12.17}$$

The constraint $R(r) \geq 0$ [Eq. (11.30)] implies then that the motion is within the range $0 \leq r \leq 2m$, with $r = 2m$ being a r-turning point, since it is a simple root of $R(r)$ (cf. Sec. 11.2.6). It corresponds to the outer edge of the ergoregion in the equatorial plane, cf. Eq. (10.19d). Hence we have necessarily $\mathscr{L} \cap \mathscr{M}_{\rm I} \neq \varnothing$ and (12.2) applies: $L < 0$. The inner boundary of the radial motion, $r = 0$, is the ring singularity. Accordingly, in the maximally extended Kerr spacetime, $\mathscr{L}$ starts at the ring singularity in a $\mathscr{M}_{\rm III}$-type region (cf. Fig. 12.3), has r increasing, enters a $\mathscr{M}^*_{\rm II}$-type region (time reversed copy of $\mathscr{M}_{\rm II}$), emerges in $\mathscr{M}_{\rm I}$ via the white hole horizon at $r = r_+$ and reaches a r-turning point at $r = 2m$, then r decreases continuously until $\mathscr{L}$ terminates at the ring singularity of $\mathscr{M}_{\rm III}$, after having crossed the black hole horizon $\mathscr{H}$ and the inner horizon $\mathscr{H}_{\rm in}$. This trajectory, depicted in Fig. 12.3, is similar to that for $Q \neq 0$ shown in Fig. 12.1, except that it is "blocked" by two ring singularities and cannot oscillate forever between distinct $\mathscr{M}_{\rm I}$-type regions.

Remark 2: For $Q \neq 0$ and $L \neq 0$, the limit $Q \to 0$ corresponds to $\bar{L}^2 \to +\infty$ [cf. Eq. (12.8)], so that Eq. (12.10) yields $r_{\rm min} \to 0$ and $r_{\rm max} \to 2m$. We recover then the range $[0, 2m]$ for r obtained here for $Q = 0$.

We conclude that

> Any null geodesic with $E = 0$ and $Q = 0$ is either a null generator of one of the two Killing horizons $\mathscr{H}$ or $\mathscr{H}_{\rm in}$ (in which case, it has $L = 0$) or it has $L < 0$, lies in the equatorial plane, emanates from a ring singularity, reaches the outer ergosphere ($r = 2m$), where it has a r-turning point, and terminates at a ring singularity.

Regarding the sign of L, we can combine the above result with that obtained for $Q \neq 0$ [Eq. (12.13)] to conclude:

> For $a \neq 0$, any null geodesic with $E = 0$ has necessarily
>
> $$\boxed{L \leq 0}_{\substack{a\neq 0 \\ E=0}} . \tag{12.18}$$

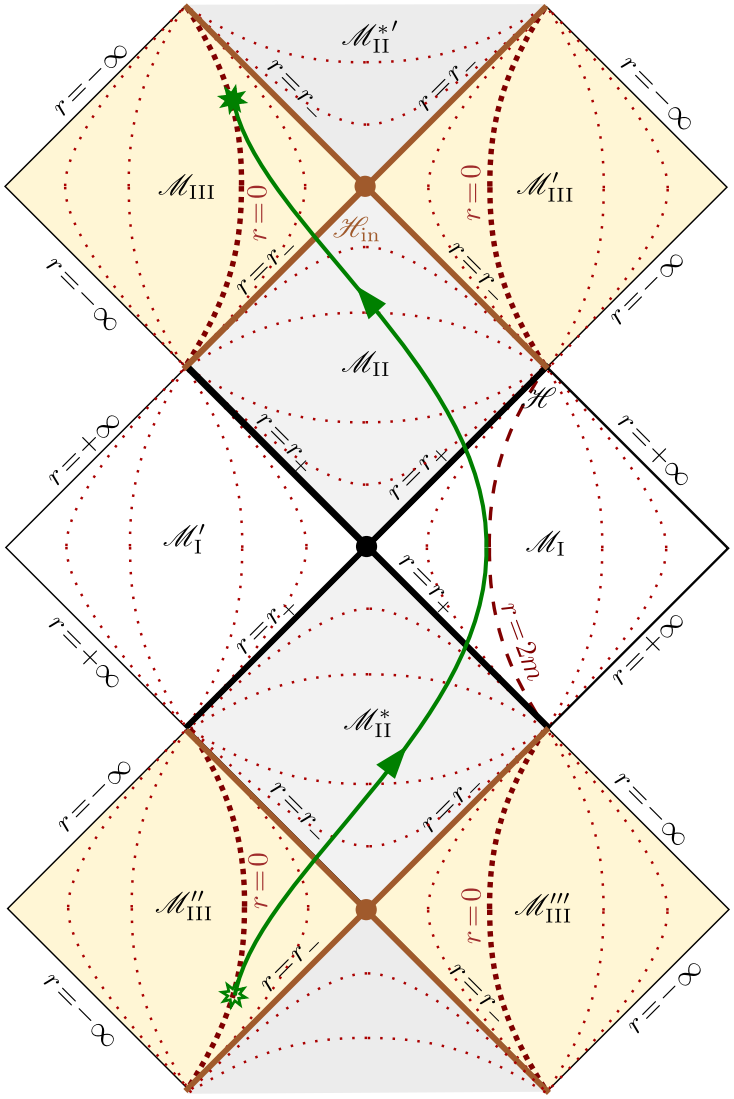


Figure 12.3: Trajectory in the extended Kerr spacetime of a null geodesic with $E = 0$, $Q = 0$ and $L < 0$.

Historical note: The zero-energy null geodesics in Kerr spacetime appear to have been first studied by Zdeněk Stuchlík in 1981, in the appendix of the article [238]; some corrections and refinement of his results have been performed by George Contopoulos in 1984 [69], who studied zero-energy timelike geodesics as well.

12.2.2 Equations of geodesic motion for $E \neq 0$

For any null geodesic $\mathscr{L}$ with $E \neq 0$, we introduce the reduced constants of motion

$$\boxed{\ell := \frac{L}{E}} \qquad \text{and} \qquad \boxed{q := \frac{Q}{E^2}}. \tag{12.19}$$

Note that, in geometrized units ($G = 1$ and $c = 1$), ℓ has the dimension of a length and q that of a squared length.

Remark 3: In the literature, ℓ is sometimes denoted by λ (e.g. Refs. [17, 131, 92]) or by ξ (e.g. Ref. [50]), while q is sometimes denoted by η (e.g. Refs. [17, 50, 131]). Moreover, in studies restricted to $q \geq 0$, it may happen that the notation q stands for the square root of the quantity q defined by (12.19) (e.g. Refs. [88, 133, 92]).

Remark 4: Contrary to L and Q, the quantities ℓ and q are independent from the affine parametrization of the geodesic $\mathscr{L}$. Indeed, if instead of the affine parameter λ associated with the particle's 4-momentum $\boldsymbol{p}$, one considers the affine parameter $\lambda' = \alpha\lambda$, where α is a positive constant, the tangent vector field becomes $\boldsymbol{p}' = \alpha^{-1}\boldsymbol{p}$, so that the associated conserved quantities are $E' = -\boldsymbol{\xi}\cdot\boldsymbol{p}' = \alpha^{-1}E$ [cf. Eq. (11.2a)], $L' = \boldsymbol{\eta}\cdot\boldsymbol{p}' = \alpha^{-1}L$ [cf. Eq. (11.2b)] and $Q' = \tilde{\boldsymbol{K}}(\boldsymbol{p}', \boldsymbol{p}') = \alpha^{-2}Q$ [cf. Eq. (11.41)], so that $\ell' := L'/E' = \ell$ and $q' := Q'/E'^2 = q$.

The non-negative property of Carter constant $\mathscr{K}$ [Eq. (11.36)], along with the identity (11.34) leads to the following constraint on the parameters ℓ and q:

$$\boxed{q + (\ell - a)^2 \geq 0}. \tag{12.20}$$

Example 1 (Principal null geodesic): A principal null geodesic has $E \neq 0$ if, and only if, it does not belong to the outgoing family generating the horizon $\mathscr{H}$ or $\mathscr{H}_{\rm in}$. This follows immediately from Eqs. (11.6) and (11.9). One has then $L = aE\sin^2\theta_0$ [Eqs. (11.6) and (11.9)] and $Q = -a^2E^2\cos^2\theta_0$ [Eq. (11.38)], where θ_0 is the constant value of θ along the geodesic. We have thus

$$\ell = a\sin^2\theta_0 \qquad \text{and} \qquad q = -a^2\cos^4\theta_0. \tag{12.21}$$

This yields

$$q + (\ell - a)^2 = 0, \tag{12.22}$$

so that the inequality (12.20) is saturated for principal null geodesics.

The equations of motion in term of the Mino parameter λ', Eqs. (11.50) specialized to $\mu = 0$, can be rewritten as

$$\boxed{\frac{1}{E}\frac{\mathrm{d}t}{\mathrm{d}\lambda'} = \frac{1}{\Delta}\left[(r^2+a^2)^2 - 2am\ell r\right] - a^2\sin^2\theta} \tag{12.23a}$$

$$\boxed{\frac{1}{|E|}\frac{\mathrm{d}r}{\mathrm{d}\lambda'} = \epsilon_r\sqrt{\mathcal{R}(r)}} \tag{12.23b}$$

$$\boxed{\frac{1}{|E|}\frac{\mathrm{d}\theta}{\mathrm{d}\lambda'} = \epsilon_\theta\sqrt{\tilde{\Theta}(\theta)}} \tag{12.23c}$$

$$\boxed{\frac{1}{E}\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda'} = \frac{\ell}{\sin^2\theta} + \frac{a}{\Delta}(2mr - a\ell)}, \tag{12.23d}$$

with

$$\mathcal{R}(r) := \frac{R(r)}{E^2} \qquad \text{and} \qquad \tilde{\Theta}(\theta) := \frac{\Theta(\theta)}{E^2}. \tag{12.24}$$

From the general expressions (11.44), (11.86) and (11.35) specialized to $\mu = 0$, we get

$$\boxed{\mathcal{R}(r) = (r^2 + a^2 - a\ell)^2 - \Delta\left[q + (\ell - a)^2\right]} \tag{12.25a}$$

$$\boxed{\mathcal{R}(r) = r^4 + (a^2 - \ell^2 - q)r^2 + 2m\left[q + (\ell - a)^2\right] r - a^2 q} \tag{12.25b}$$

and

$$\boxed{\tilde{\Theta}(\theta) = q + \cos^2\theta\left(a^2 - \frac{\ell^2}{\sin^2\theta}\right)}. \tag{12.26}$$

It suffices to use the parameter $\lambda'' := |E|\lambda'$ to make E disappear from the system (12.23). We therefore conclude that

In Kerr spacetime, a null geodesic with $E \neq 0$ is entirely determined by the two constants (ℓ, q), by the sign of E and by the values of the two signs $\epsilon_r = \pm 1$ and $\epsilon_\theta = \pm 1$ at a given point.

Example 2 (Principal null geodesic): Given the values (12.21) of ℓ and q for a principal null geodesic, Eq. (12.25b) reduces to a simple expression for the quartic polynomial $\mathcal{R}$:

$$\mathcal{R}(r) = \left(r^2 + a^2\cos^2\theta_0\right)^2. \tag{12.27}$$

We note that $\mathcal{R}(r) = \rho^4$, which makes sense because $\theta = \theta_0$ is constant along such a geodesic. For any principal null geodesic lying in the equatorial plane, the polynomial simplifies even further:

$$\mathcal{R}(r) = r^4 \qquad \left(\theta_0 = \frac{\pi}{2}\right). \tag{12.28}$$

12.2.3 Position on a remote observer's screen

The constants (ℓ, q) are closely related to the impact coordinates (α, β) on the screen of an asymptotic inertial observer (cf. Sec. 10.7.5) in case the null geodesic $\mathscr{L}$ reaches the asymptotic region $r \to +\infty$. Indeed, let us consider an asymptotic inertial observer $\mathscr{O}$ located at (fixed) Boyer-Lindquist coordinates $(r_\mathscr{O}, \theta_\mathscr{O}, \varphi_\mathscr{O})$, with $r_\mathscr{O} \gg m$ (cf. Fig. 12.4). In order to reach $\mathscr{O}$, the geodesic $\mathscr{L}$ must be such that the constraints constraints $\mathcal{R}(r_\mathscr{O}) \geq 0$ [Eq. (12.46)] and $\tilde{\Theta}(\theta_\mathscr{O}) \geq 0$ [Eq. (11.33)] are fulfilled. The first one is always satisfied due to the assumption that $\mathscr{O}$ is an asymptotic observer, since $\mathcal{R}(r) \sim r^4$ for $r \to +\infty$ [cf. Eq. (12.25b)]. In view of expression (12.26) for $\tilde{\Theta}$, the second constraint is equivalent to

$$\left(q + a^2\cos^2\theta_\mathscr{O}\right)\sin^2\theta_\mathscr{O} - \ell^2\cos^2\theta_\mathscr{O} \geq 0. \tag{12.29}$$

If $\sin\theta_\mathscr{O}$ is small, this constraint limits significantly the amplitude of ℓ.

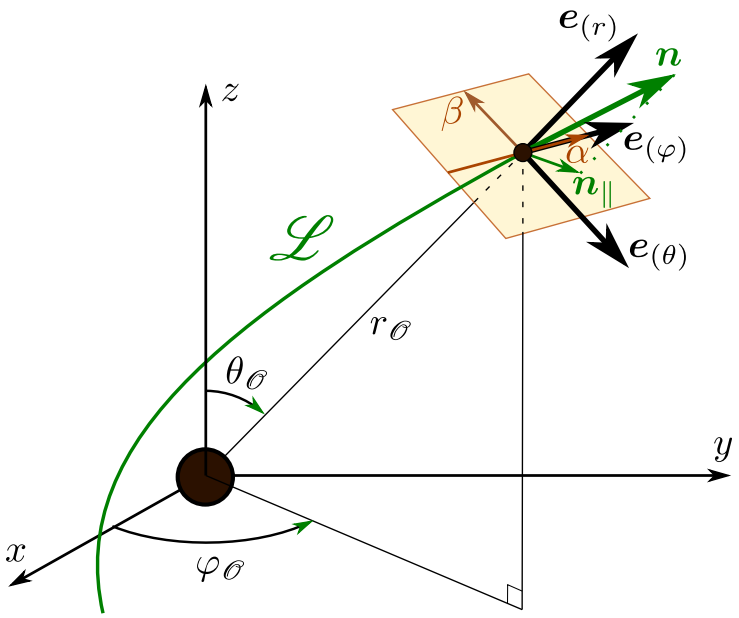


Figure 12.4: Impact of a null geodesic $\mathscr{L}$ onto the screen of a remote observer. (x, y, z) are the Cartesian Boyer-Lindquist coordinates defined by Eq. (11.64).

Observer not located on the rotation axis

Here we treat the generic case of an observer $\mathscr{O}$ that is not located on the rotation axis, i.e. we assume $\theta_{\mathscr{O}} \notin \{0, \pi\}$. The orthonormal frame of $\mathscr{O}$ is $\boldsymbol{e}_{(0)} = \boldsymbol{\xi}$, $\boldsymbol{e}_{(r)} = \boldsymbol{\partial}_r$, $\boldsymbol{e}_{(\theta)} = r_{\mathscr{O}}^{-1}\boldsymbol{\partial}_\theta$ and $\boldsymbol{e}_{(\varphi)} = (r_{\mathscr{O}} \sin\theta_{\mathscr{O}})^{-1}\boldsymbol{\partial}_\varphi$ [Eq. (10.115) with $r = r_{\mathscr{O}} \to +\infty$]. Let us assume that the observer set up a screen (via a telescope) centered on the direction to the black hole, i.e. such that $\boldsymbol{e}_{(r)}$ is the normal to the screen. The 4-momentum of the photon having $\mathscr{L}$ as worldline at the location of $\mathscr{O}$ writes

$$\boldsymbol{p} = p^t\, \boldsymbol{\xi} + p^r\, \boldsymbol{e}_{(r)} + p^\theta r_{\mathscr{O}}\, \boldsymbol{e}_{(\theta)} + p^\varphi r_{\mathscr{O}} \sin\theta_{\mathscr{O}}\, \boldsymbol{e}_{(\varphi)} = E(\boldsymbol{\xi} + \boldsymbol{n}),$$

where the second equality follows from the orthogonal decomposition (1.29) with respect to $\mathscr{O}$, $\boldsymbol{\xi}$ being the 4-velocity of $\mathscr{O}$ and the unit spacelike vector $\boldsymbol{n}$ being the photon's velocity[1] relative to $\mathscr{O}$; it obeys $\boldsymbol{\xi} \cdot \boldsymbol{n} = 0$ and $\boldsymbol{n} \cdot \boldsymbol{n} = 1$ [Eq. (1.35)]. That the conserved energy E appears in the above equation is a direct consequence of its definition as $E := -\boldsymbol{\xi} \cdot \boldsymbol{p}$ [Eq. (11.2a)] with $\boldsymbol{\xi} \cdot \boldsymbol{\xi} = -1$ for $r_{\mathscr{O}} \to +\infty$. The above identity implies $p^t = E$ and

$$\boldsymbol{n} = \frac{p^r}{E}\, \boldsymbol{e}_{(r)} + \frac{p^\theta}{E} r_{\mathscr{O}}\, \boldsymbol{e}_{(\theta)} + \frac{p^\varphi}{E} r_{\mathscr{O}} \sin\theta_{\mathscr{O}}\, \boldsymbol{e}_{(\varphi)}.$$

With respect to $\mathscr{O}$, the incoming direction of the photon is given by the vector $-\boldsymbol{n}$ and the trace on the screen is indicated by the component of that vector that is tangent to the screen, namely

$$\boldsymbol{m} = -\boldsymbol{n}_\parallel = -\frac{p^\theta}{E} r_{\mathscr{O}}\, \boldsymbol{e}_{(\theta)} - \frac{p^\varphi}{E} r_{\mathscr{O}} \sin\theta_{\mathscr{O}}\, \boldsymbol{e}_{(\varphi)} \tag{12.30}$$

[1] $\boldsymbol{n}$ is denoted by $\boldsymbol{V}$ in Eq. (1.29).

since the screen plane is spanned by $\boldsymbol{e}_{(\theta)}$ and $\boldsymbol{e}_{(\varphi)}$ (cf. Fig. 12.4). Let us choose the screen dimensionless Cartesian coordinates (α, β) so that the black hole's rotation axis appears as the β-axis (cf. Fig. 12.4), then

$$\boldsymbol{m} = \alpha \boldsymbol{e}_{(\alpha)} + \beta \boldsymbol{e}_{(\beta)}, \quad \text{with} \quad \boldsymbol{e}_{(\alpha)} = \boldsymbol{e}_{(\varphi)} \quad \text{and} \quad \boldsymbol{e}_{(\beta)} = -\boldsymbol{e}_{(\theta)}. \tag{12.31}$$

By comparing with (12.30), we get

$$\alpha = -\frac{p^\varphi}{E} r_{\mathscr{O}} \sin\theta_{\mathscr{O}} \qquad \text{and} \qquad \beta = \frac{p^\theta}{E} r_{\mathscr{O}}.$$

Now, for $r_{\mathscr{O}} \to +\infty$,

$$\frac{p^\theta}{E} = \frac{1}{E}\frac{\mathrm{d}\theta}{\mathrm{d}\lambda} = \frac{1}{E r_{\mathscr{O}}^2}\frac{\mathrm{d}\theta}{\mathrm{d}\lambda'} = \frac{\epsilon_\theta}{r_{\mathscr{O}}^2}\sqrt{\tilde{\Theta}(\theta_{\mathscr{O}})}$$

$$\frac{p^\varphi}{E} = \frac{1}{E}\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda} = \frac{1}{E r_{\mathscr{O}}^2}\frac{\mathrm{d}\varphi}{\mathrm{d}\lambda'} = \frac{\ell}{r_{\mathscr{O}}^2 \sin^2\theta_{\mathscr{O}}},$$

where we have used Eqs. (12.23c) and (12.23d), with the term involving a/Δ neglected since $\Delta \sim r^2$ for $r \to +\infty$. By inserting these formula into the above expressions of α and β, and using Eq. (12.26) for $\tilde{\Theta}(\theta_{\mathscr{O}})$, we get the sought relation between the constants of motion (ℓ, q) and the screen coordinates:

$$\boxed{\alpha = -\frac{\ell}{r_{\mathscr{O}} \sin\theta_{\mathscr{O}}}} \tag{12.32a}$$

$$\boxed{\beta = \frac{\epsilon_\theta}{r_{\mathscr{O}}}\sqrt{q + \cos^2\theta_{\mathscr{O}}\left(a^2 - \frac{\ell^2}{\sin^2\theta_{\mathscr{O}}}\right)}}. \tag{12.32b}$$

We have defined (α, β) as dimensionless Cartesian coordinates on the screen, cf. Eq. (12.31), where $\boldsymbol{m}$ is dimensionless and $(\boldsymbol{e}_{(\alpha)}, \boldsymbol{e}_{(\beta)})$ is the screen's orthonormal basis. In practice, their values are tiny, being exactly zero at the limit $r_{\mathscr{O}} \to +\infty$. (α, β) can thus be interpreted as *angular* coordinates, measuring the departure from the direction of the black hole "center" on the celestial sphere of observer $\mathscr{O}$. We shall then call (α, β) the ***screen angular coordinates***.

Remark 5: When studying null geodesics in Schwarzschild spacetime in Chap. 8, we introduced the impact parameter b as $b := |L|/E$ [Eq. (8.12)], hence b is related to ℓ by

$$b = |\ell|. \tag{12.33}$$

Moreover, thanks to spherical symmetry, we could reduce the study to the case where both the observer and the geodesic lie in the equatorial plane, which imply $\theta_{\mathscr{O}} = \pi/2$ and $q = 0$. Equations (12.32) yield then

$$|\alpha| = \frac{b}{r_{\mathscr{O}}} = \hat{b} \qquad \text{and} \qquad \beta = 0, \tag{12.34}$$

where $\hat{b}$ is the angle introduced by formula (8.118).

Remark 6: The angular impact parameters (α, β) depend on the geodesic $\mathscr{L}$ as a curve in spacetime and not on any affine parametrization of $\mathscr{L}$. Their direct connection with (ℓ, q) expressed by (12.32) is thus in perfect agreement with the invariance of (ℓ, q) in any affine reparametrization of $\mathscr{L}$, as noticed in Remark 4 p. 383.

We deduce from Eqs. (12.32) a simple relation between the squared angular distance to the screen center, $\alpha^2 + \beta^2$, and the constants of motion (ℓ, q) of the incoming geodesic:

$$\alpha^2 + \beta^2 = \frac{1}{r_{\mathscr{O}}^2} \left(\ell^2 + q + a^2 \cos^2 \theta_{\mathscr{O}} \right) . \tag{12.35}$$

Observer on the rotation axis

If the asymptotic inertial observer $\mathscr{O}$ is located on the rotation axis, i.e. if $\theta_{\mathscr{O}} = 0$ or $\theta_{\mathscr{O}} = \pi$, the only value of ℓ compatible with the constraint (12.29) is

$$\ell = 0 \qquad (\theta_{\mathscr{O}} = 0 \text{ or } \theta_{\mathscr{O}} = \pi). \tag{12.36}$$

Given that $\ell = L/E$, we recover one of the properties listed in Sec. 11.3.6: a geodesic cannot encounter the rotation axis unless it has $L = 0$.

Moreover, on the rotation axis, the vectors $\boldsymbol{e}_{(\theta)}$ and $\boldsymbol{e}_{(\varphi)}$ are not defined, due to the singularity of spherical coordinates there. Consequently, the screen coordinates (α, β) cannot be defined by (12.31). In particular, the rotation axis appears as single point on the screen, which forbids to use it to define the β-axis. One has then to pick an arbitrary orthonormal frame $(\boldsymbol{e}_{(\alpha)}, \boldsymbol{e}_{(\beta)})$ in the screen plane to define (α, β). Formulas (12.88) do no longer hold, but formula (12.35) is still valid, since the distance from the screen's center is a quantity independent from the orientation of the frame $(\boldsymbol{e}_{(\alpha)}, \boldsymbol{e}_{(\beta)})$. Another way to see that (12.35) is still valid is to notice that it admits a well-defined limit for $\theta_{\mathscr{O}} \to 0$ or π. Taking into account $\ell = 0$, we obtain

$$\alpha^2 + \beta^2 = \frac{1}{r_{\mathscr{O}}^2} \left(q + a^2 \right) \qquad (\theta_{\mathscr{O}} = 0 \text{ or } \theta_{\mathscr{O}} = \pi). \tag{12.37}$$

12.2.4 Latitudinal motion

Specializing the general results of Sec. 11.3.6 to $\mu = 0$, we get

- A null geodesic $\mathscr{L}$ of Kerr spacetime cannot encounter the rotation axis unless it has $\ell = 0$.
- If $|\ell| \geq a$, the reduced Carter constant q is necessarily non-negative:

$$q \geq 0. \tag{12.38}$$

- The reduced Carter constant q can take negative values only if $|\ell| < a$ (which implies $a \neq 0$); its range is then limited from below:

$$q \geq q_{\min} = -\left(a - |\ell|\right)^2 . \tag{12.39}$$

If $q < 0$, $\mathscr{L}$ is called a ***vortical null geodesic***; it never encounters the equatorial plane.

- If $q > 0$ and $\ell \neq 0$, $\mathscr{L}$ oscillates symmetrically about the equatorial plane, between two θ-turning points, at $\theta = \theta_{\rm m}$ and $\theta = \pi - \theta_{\rm m}$, where $\theta_{\rm m} \in (0, \pi/2)$ is given by Eqs. (11.76) and (11.81):

$$\theta_{\rm m} = \arccos \sqrt{\frac{q}{\ell^2 + q}} \quad \text{for} \quad a = 0 \tag{12.40}$$

$$\theta_{\rm m} = \arccos \sqrt{\frac{1}{2}\left[1 - \frac{\ell^2 + q}{a^2} + \sqrt{\left(1 - \frac{\ell^2 + q}{a^2}\right)^2 + \frac{4q}{a^2}}\right]} \quad \text{for} \quad a \neq 0. \tag{12.41}$$

If $q > 0$ and $\ell = 0$, $\mathscr{L}$ crosses repeatedly the rotation axis, with θ taking all values in the range $[0, \pi]$.

- If $q = 0$, $\mathscr{L}$ is stably confined to the equatorial plane for $|\ell| > a$ or $|\ell| = a \neq 0$; for $|\ell| < a$, $\mathscr{L}$ either lies unstably in the equatorial plane or approaches it asymptotically from one side, while for $\ell = 0$ and $a = 0$, $\mathscr{L}$ lies at a constant value $\theta = \theta_0 \in [0, \pi]$.

- If $q_{\rm min} < q < 0$, $\mathscr{L}$ never encounters the equatorial plane, having a θ-motion entirely confined either to the Northern hemisphere ($0 < \theta < \pi/2$) or to the Southern one ($\pi/2 < \theta < \pi$); if $\ell \neq 0$, $\mathscr{L}$ oscillates between two θ-turning points, at $\theta = \theta_{\rm m}$ and $\theta = \theta_{\rm v}$ (Northern hemisphere) or at $\theta = \pi - \theta_{\rm v}$ and $\theta = \pi - \theta_{\rm m}$ (Southern hemisphere), where $\theta_{\rm m}$ is given by Eq. (12.41) above and $\theta_{\rm v}$ is given by Eq. (11.82):

$$\theta_{\rm v} = \arccos \sqrt{\frac{1}{2}\left[1 - \frac{\ell^2 + q}{a^2} - \sqrt{\left(1 - \frac{\ell^2 + q}{a^2}\right)^2 + \frac{4q}{a^2}}\right]}; \tag{12.42}$$

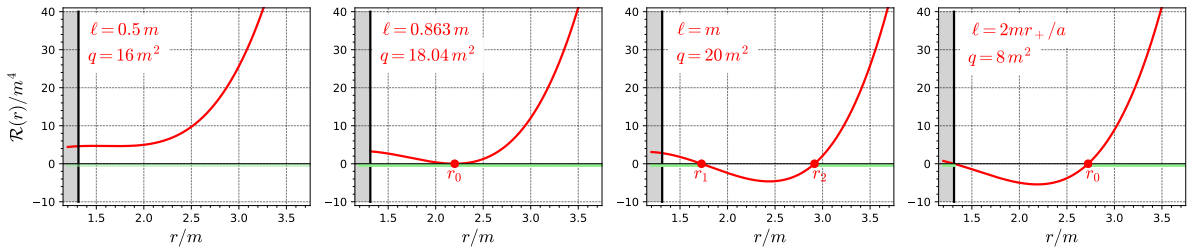
if $\ell = 0$, $\mathscr{L}$ oscillates about the rotation axis, with a θ-turning point at $\theta = \theta_{\rm v}$ or $\theta = \pi - \theta_{\rm v}$, where $\theta_{\rm v}$ is given by Eq. (11.79), or equivalently by the $\ell \to 0$ limit of Eq. (12.42):

$$\theta_{\rm v} = \arccos\left(\frac{\sqrt{-q}}{a}\right) \qquad (\ell = 0). \tag{12.43}$$

- If $q = q_{\rm min}$, $\mathscr{L}$ lies stably at a constant value $\theta = \theta_*$ or $\theta = \pi - \theta_*$, with $\theta_* \in [0, \pi/2)$ given by

$$\theta_* := \arcsin \sqrt{\frac{|\ell|}{a}}. \tag{12.44}$$

Remark 7: For $\ell < 0$, the constraint (12.39) is tighter than (12.20).



(a) (b) (c) (d)

Figure 12.5: Quartic polynomial $\mathcal{R}(r)$ in the region $\mathscr{M}_{\rm I}$ for four values of (ℓ, q) and for $a = 0.95\,m$. The grey area marks the black hole region, with the vertical black line at the event horizon, at $r = r_+ \simeq 1.312\,m$. The green part of the r axis corresponds to $\mathcal{R}(r) \geq 0$, i.e. to regions where the geodesic motion is allowed. *[Figure generated by the notebook D.4.13]*

Example 3 (Principal null geodesic): A principal null geodesic moves at a constant angle $\theta = \theta_0$ and has $\ell = a \sin^2\theta_0$ [Eq. (12.21)]. For $\theta_0 \neq \pi/2$, we have $|\ell| < a$ and Eq. (12.39) yields $q_{\rm min} = -a^2 \cos^4\theta_0$. Comparing with the value of q given by Eq. (12.21), we note that

$$q = q_{\rm min}, \tag{12.45}$$

which agrees with the last case listed above, i.e. motion at constant θ, with $\theta_0 = \theta_*$ or $\theta_0 = \pi - \theta_*$ according to Eq. (12.44), since $\sqrt{|\ell|/a} = \sin\theta_0$.

12.2.5 Radial motion

As for any geodesic, the radial motion of a null geodesic $\mathscr{L}$ is ruled by $R(r) \geq 0$ [Eq. (11.30)], which, in terms of $\mathcal{R}(r)$ [Eq. (12.25)], writes

$$\boxed{\mathcal{R}(r) \geq 0}. \tag{12.46}$$

Example 4 (Principal null geodesic): Given the value (12.27) of $\mathcal{R}(r)$ for a principal null geodesic, we note that $\mathcal{R}(r) > 0$ in all Kerr spacetime. This is consistent with the fact that, for $E \neq 0$ and $\theta_0 \neq \pi/2$, ingoing principal null geodesics travel from $r = +\infty$ to $r = -\infty$ (cf. the dashed green curve in Fig. 10.10) and outgoing principal null geodesics travel from $r = -\infty$ to $r = +\infty$ (cf. the solid green curve in Fig. 10.11).

Since $\mathcal{R}(r)$ is a polynomial of degree 4 in r, its behavior can be relatively complicated. However, it turns out that in the black hole exterior, its behavior is quite simple, according to the following lemma:

Lemma ($\mathcal{R}(r)$ in $\mathscr{M}_{\rm I}$)

In region $\mathscr{M}_{\rm I}$ of Kerr spacetime, i.e. for $r > r_+$, the quartic polynomial $\mathcal{R}(r)$ associated to a given geodesic [Eq. (12.25)] has one of the following behaviors, depending on the value of (ℓ, q):

1. $\mathcal{R}(r)$ has no root in $(r_+, +\infty)$ and $\mathcal{R}(r) > 0$ there (Fig. 12.5a);

2. $\mathcal{R}(r)$ has a double root in $(r_+, +\infty)$, at $r = r_0$ say, and $\mathcal{R}(r) > 0$ iff $r \in (r_+, r_0)$ or $r \in (r_0, +\infty)$ (Fig. 12.5b);

3. $\mathcal{R}(r)$ has two simple roots in $(r_+, +\infty)$, at $r = r_1$ and $r = r_2$ say, and $\mathcal{R}(r) > 0$ iff $r \in (r_+, r_1)$ or $r \in (r_2, +\infty)$ (Fig. 12.5c);

4. $\mathcal{R}(r)$ has a unique simple root in $(r_+, +\infty)$, at $r = r_0$ say; then necessarily $\ell = 2mr_+/a$, $\mathcal{R}(r_+) = 0$ and $\mathcal{R}(r) > 0$ iff $r \in (r_0, +\infty)$ (Fig. 12.5d).

Proof. First, we observe that $\mathcal{R}(r)$ is positive or zero at both ends of $\mathscr{M}_\mathrm{I}$. This is clear for the asymptotic flat end since, according to expression (12.25b), $\mathcal{R}(r) \sim r^4 > 0$ when $r \to +\infty$. At the inner end, namely for $r \to r_+$, we have $\Delta \to 0$ and expression (12.25a) yields the limit value $\mathcal{R}(r_+) = (r_+^2 + a^2 - a\ell)^2 \geq 0$. Now, as a polynomial of degree four, $\mathcal{R}(r)$ has at most four real roots. Given the above boundary conditions, $\mathcal{R}(r)$ can have a unique simple root in $(r_+, +\infty)$ only if $\mathcal{R}(r_+) = 0$ (cf. Fig. 12.5d), which occurs for a very specific value of ℓ, namely $\ell = (r_+^2 + a^2)/a = 2mr_+/a$. This is case 4 above. Similarly, the case of three roots in $(r_+, +\infty)$, either three simple roots or one double and one simple root, is compatible with the boundary conditions only if $\mathcal{R}(r_+) = 0$, i.e. if r_+ is a fourth root of $\mathcal{R}(r)$. But then the four roots of $\mathcal{R}(r)$ would be positive (since $r_+ > 0$). Now, since there is no r^3 term in the expression (12.25b) of $\mathcal{R}(r)$, the sum of the roots of $\mathcal{R}(r)$ is zero, which is impossible with all the roots being positive. Hence there cannot be three roots in $(r_+, +\infty)$. The same argument about the zero sum of the roots excludes as well the case of four roots of $\mathcal{R}(r)$ in $(r_+, +\infty)$. There remains then the cases of no root at all in $(r_+, +\infty)$ (case 1 of the lemma) and that of two roots (cases 2 and 3). These three cases are compatible with the boundary conditions and the vanishing of the sum of the roots. □

Remark 8: Case 4 of the lemma can be seen as the limit $r_1 \to r_+$ of case 3.

Remark 9: In case 4, $\mathcal{R}(r_+) = 0$ occurs for $\Omega_H \ell = 1$, where Ω_H is the black hole rotation velocity, as given by Eq. (10.74). Similarly, $\mathcal{R}(r_-) = 0$ for $\Omega_\mathrm{in} \ell = 1$, where Ω_in is the rotation velocity of the inner horizon, cf. Eq. (10.76).

Remark 10: The region $\mathscr{M}_\mathrm{III}$ shares with $\mathscr{M}_\mathrm{I}$ that $\mathcal{R}(r)$ is non-negative at each of its ends: $\mathcal{R}(r) \sim r^4 > 0$ when $r \to -\infty$ and $\mathcal{R}(r_-) = (r_-^2 + a^2 - a\ell)^2 \geq 0$. However, the argument used to limit the number of roots in $\mathscr{M}_\mathrm{I}$ cannot be applied to $\mathscr{M}_\mathrm{III}$ because the latter can accommodate for both positive and negative values of r and thus four roots of $\mathcal{R}(r)$ can be located in $\mathscr{M}_\mathrm{III}$.

According to the definition given in Sec. 11.2.6, a r-turning of a null geodesic $\mathscr{L}$ is a point $p_0 \in \mathscr{L}$ such that $r_0 := r(p)$ is a simple root of $\mathcal{R}(r)$. The above lemma leads then to:

A null geodesic in Kerr spacetime has

- at most one r-turning point in region $\mathscr{M}_\mathrm{I}$;

- no r-turning point in region $\mathscr{M}_{\rm II}$.

Proof. If a geodesic would have two turning points in $\mathscr{M}_{\rm I}$, this would mean that there exist two simple roots in $\mathscr{M}_{\rm I}$ and that $\mathcal{R}(r)$ is positive between them (so that the motion is possible there). But this is excluded by the lemma. In region $\mathscr{M}_{\rm II}$, we have $\Delta < 0$ and Eqs. (12.25a) and (12.20) show that $\mathcal{R}(r)$ is the sum of two non-negative terms. The only possibility to have $\mathcal{R}(r) = 0$ is then that each term vanishes separately: $r^2 + a^2 - a\ell = 0$ and $q + (\ell - a)^2 = 0$, i.e.

$$r^2 = a(\ell - a) \qquad \text{and} \qquad q = -(\ell - a)^2.$$

The second equation implies $q \leq 0$. The case $q = 0$ would lead to $\ell = a$ and $r^2 = 0$, which is impossible in $\mathscr{M}_{\rm II}$. There remains $q < 0$, but according to the results of Sec. 12.2.4, this implies $|\ell| < a$, so that $\ell - a < 0$ and the first equation above would yield $r^2 < 0$, which is impossible. There is thus no r-turning point in $\mathscr{M}_{\rm II}$. □

Remark 11: That no r-turning point can exist in $\mathscr{M}_{\rm II}$ has been established above by some considerations on $\mathcal{R}(r)$. This property can also be deduced as an immediate consequence of the result (11.88), namely that r must be a strictly decreasing function of λ at any point in $\mathscr{M}_{\rm II}$.

Let us consider case 2 of the lemma (double root of $\mathcal{R}(r)$ in $\mathscr{M}_{\rm I}$); it corresponds to two distinct situations regarding the null geodesic $\mathscr{L}$ having $\mathcal{R}(r)$ as radial polynomial. First, $\mathscr{L}$ can lie at a constant value of r, which is necessarily the double root r_0 of $\mathcal{R}(r)$ by virtue of Eq. (12.23b); $\mathscr{L}$ belongs then to the category of the *spherical photon orbits*, which will be studied in Sec. 12.3. If r is not constant along $\mathscr{L}$, then according to the definition in Sec. 11.3.5, $\mathscr{L}$ has as an asymptotic r-value, which is the double root r_0. Given that r_0 is the only root of $\mathcal{R}(r)$ in $(r_+, +\infty)$, Eq. (12.23b) implies $\mathrm{d}r/\mathrm{d}\lambda \neq 0$ all along $\mathscr{L}$, so that $r(\lambda) \to r_0$ for $\lambda \to +\infty$ (future asymptotic value) or $\lambda \to -\infty$ (past asymptotic value). Such a geodesic belong to the category of the *critical null geodesics*, which will be studied in Sec. 12.4.1.

In view of the above results, we can state:

In the region $\mathscr{M}_{\rm I}$ of Kerr spacetime, any null geodesic $\mathscr{L}$ has one of the following behaviors. For *generic* values of the constant of motions (ℓ, q), the possibilities are:

1. $\mathscr{L}$ arises from the past null infinity of $\mathscr{M}_{\rm I}$, $\mathscr{I}^-$, has a r-turning point, which we may call the ***periastron***, and terminates at the future null infinity of $\mathscr{M}_{\rm I}$, $\mathscr{I}^+$;

2. $\mathscr{L}$ arises from the past null infinity $\mathscr{I}^-$, has r decreasing monotonically and crosses the black hole event horizon $\mathscr{H}$;

3. $\mathscr{L}$ arises from the past event horizon $\mathscr{H}^-$ separating $\mathscr{M}_{\rm I}$ from the white hole region $\mathscr{M}_{\rm II}^*$, has r increasing monotonically and terminates at the future null infinity $\mathscr{I}^+$;

4. $\mathscr{L}$ arises from the past event horizon $\mathscr{H}^-$, has a r-turning point, which we may call the ***apoastron***, and crosses the black hole event horizon $\mathscr{H}$;

For some *specific* values of the constant of motions (ℓ, q), forming a 1-dimensional (hence zero-measure) subset of the set of all possible values[a], the possibilities are:

5. $\mathscr{L}$ evolves at a fixed value of r;
6. $\mathscr{L}$ arises from the past null infinity $\mathscr{I}^-$ and has r decreasing monotonically to a future asymptotic r-value at $r_0 > r_+$;
7. $\mathscr{L}$ arises from a past asymptotic r-value at $r_0 > r_+$, has r increasing monotonically and terminates at the future null infinity $\mathscr{I}^+$;
8. $\mathscr{L}$ arises from a past asymptotic r-value at $r_0 > r_+$, has r decreasing monotonically and crosses the black hole event horizon $\mathscr{H}$;
9. $\mathscr{L}$ arises from the past event horizon $\mathscr{H}^-$, has r increasing monotonically to a future asymptotic r-value at $r_0 > r_+$.

[a] This subset is given in parametric form by Eqs. (12.56)-(12.57) below.

Case 1 corresponds to a scattering trajectory, leading to the standard phenomenon of deflection of light. The polynomial $\mathcal{R}(r)$ belongs then to case 3 or 4 of the lemma. Ingoing principal null geodesics belong to case 2, while the outgoing ones with $E \neq 0$ belong to case 3 (cf. Example 5 below). Both cases 2 and 3 correspond to the lemma's case 1 (no root of $\mathcal{R}(r)$ in $\mathscr{M}_{\rm I}$). Cases 5–9 correspond to the lemma's case 2 (double root of $\mathcal{R}(r)$ in $\mathscr{M}_{\rm I}$). Case 5 is that of *spherical photon orbits* and will be discussed in Sec. 12.3, while cases 6–9 are those of *critical null geodesics*, to be discussed in Sec. 12.4.1.

Remark 12: The terminology *(periastron, apoastron)* employed here agrees with that introduced for the Schwarzschild case in Sec. 8.2.4.

Example 5 (principal null geodesics): That the principal null geodesics with $E \neq 0$ belong to cases 2 and 3 above is clear from their value (12.27) for $\mathcal{R}(r)$: $\mathcal{R}(r) = \rho^4 > 0$ in all $\mathscr{M}$, which precludes the existence of any r-turning point nor any r-asymptotic value along these geodesics.

Remark 13: As a sequel of Remark 10 above, a null geodesic can have two turnings points in region $\mathscr{M}_{\rm III}$. There can thus exist null geodesics that are trapped between two distinct values of r in $\mathscr{M}_{\rm III}$.

12.3 Spherical photon orbits

In the Schwarzschild case studied in Chap. 8, circular photon orbits at $r = 3m$ played a central role in the computation of the black hole shadow and the image of an accretion structure. For the Kerr black hole, a similar role is played by spherical photon orbits. They are also null geodesics evolving at a fixed value of r but, contrary to circular photon orbits of Schwarzschild spacetime, they are not planar in general.

12.3.1 Existence of spherical null geodesics

We shall say that a null geodesic $\mathscr{L}$ is ***spherical*** or is a ***spherical photon orbit*** iff $\mathscr{L}$ lies at a constant value of the coordinate r, r_0 say. We have already encountered such geodesics, namely the null generators $\mathscr{L}^{\mathrm{out},\mathscr{H}}_{(\theta,\psi)}$ and $\mathscr{L}^{\mathrm{out},\mathscr{H}_{\mathrm{in}}}_{(\theta,\psi)}$ of the two Killing horizons $\mathscr{H}$ and $\mathscr{H}_{\mathrm{in}}$ (cf. Secs. 10.5.3 and 10.8.3). Indeed, they lie at a constant value of r: $r_0 = r_+$ for $\mathscr{H}$ and $r_0 = r_-$ for $\mathscr{H}_{\mathrm{in}}$. These spherical orbits have $E = 0$ [cf. Eq. (11.25)]. From the results of Sec. 12.2.1, there is no other null geodesic with $E = 0$ at constant r. Hence we conclude

> In Kerr spacetime, the only spherical photon orbits with $E = 0$ are the null generators of the two Killing horizons $\mathscr{H}$ and $\mathscr{H}_{\mathrm{in}}$.

Remark 1: A spherical photon orbit of Kerr spacetime is an example of what has been called a ***fundamental photon orbit*** in a generic stationary and axisymmetric spacetime [71, 72], namely a null geodesic $\mathscr{L}$ of affine parameter λ such that (i) $\mathscr{L}$ is restricted a spatially compact region and (ii) there exists $\Lambda \in \mathbb{R}^+$ such that

$$\forall \lambda \in \mathbb{R}, \quad \exists \Phi \in \mathbb{R} \times \mathrm{SO}(2), \quad p(\lambda + \Lambda) = \Phi(p(\lambda)),$$

where $p(\lambda)$ stands for the point of affine parameter λ along $\mathscr{L}$ and $\mathbb{R} \times \mathrm{SO}(2)$ is the symmetry group implementing stationarity and axisymmetry. Basically (i) says that $\mathscr{L}$ is a bound orbit and (ii) that $\mathscr{L}$ is periodic, of period Λ in terms of λ, up to some isometry. Note that the definition of a fundamental photon orbit is coordinate-independent. Introducing coordinates (t, x^1, x^2, φ) adapted to the spacetime symmetries, the requirement (ii) is equivalent to demanding that x^1 and x^2 are periodic functions of λ. In the present case $x^1 = r$ and $x^2 = \theta$. For a spherical photon orbit, r is obviously a periodic function of λ, being a constant function, while we shall see in Sec. 12.3.2 that θ is indeed a periodic function of λ.

In the remainder of this section, we focus on spherical photon orbits with $E \neq 0$. We shall then describe these geodesics by the reduced constants of motion $\ell := L/E$ and $q := Q/E^2$ introduced in Sec. 12.2.2.

For a spherical photon orbit $\mathscr{L}$, the quartic polynomial $\mathcal{R}(r)$ associated to (ℓ, q) by Eq. (12.25) must obey

$$\mathcal{R}(r_0) = 0 \qquad \text{and} \qquad \mathcal{R}'(r_0) = 0. \tag{12.47}$$

Proof. Since $r = r_0$ with r_0 constant, we have $\mathrm{d}r/\mathrm{d}\lambda' = 0$ where λ' is the Mino parameter along $\mathscr{L}$, so that Eq. (12.23b) implies $\mathcal{R}(r_0) = 0$. If $\mathcal{R}'(r_0) \neq 0$, then r_0 would correspond to a r-turning point of $\mathscr{L}$ (cf. Eq. (11.46)), which would contradict r being constant. □

In other words, r_0 is a double root the polynomial $\mathcal{R}(r)$. A spherical photon orbit in $\mathscr{M}_{\mathrm{I}}$ correspond then to case 2 of the lemma in Sec. 12.2.5 and an example of $\mathcal{R}(r)$ is shown for $a = 0.95\,m$ and $r_0 = 2.2\,m$ in Fig. 12.5b.

In view of expression (12.25a) for $\mathcal{R}(r)$, the system (12.47) is equivalent to

$$(r_0^2 + a^2 - a\ell)^2 - \tilde{q}\Delta_0 = 0 \tag{12.48a}$$

$$2r_0(r_0^2 + a^2 - a\ell) - \tilde{q}(r_0 - m) = 0, \tag{12.48b}$$

where $\Delta_0 := r_0^2 - 2mr_0 + a^2$ and

$$\tilde{q} := q + (\ell - a)^2. \tag{12.49}$$

This is a system of 2 equations for 3 unknowns: (r_0, ℓ, q). We thus expect a one-parameter family of solutions. It is convenient to consider r_0 as the parameter and to solve the system (12.48) for (ℓ, q). We shall distinguish the case $r_0 = m$ from $r_0 \neq m$. If $r_0 = m$, the system (12.48) reduces to

$$\left\{ \begin{array}{l} (m^2 + a^2 - a\ell)^2 - \tilde{q}(a^2 - m^2) = 0 \\ m^2 + a^2 - a\ell = 0 \end{array} \right. \iff \left\{ \begin{array}{l} \ell = a + \frac{m^2}{a} \\ \tilde{q} = 0 \quad \text{or} \quad a = m. \end{array} \right.$$

Now, $\tilde{q} = 0$ is equivalent to $q = -(\ell - a)^2 = -m^4/a^2$, which implies $q < 0$, which is impossible for $\ell = a + \frac{m^2}{a} > a$, due to the property $|\ell| > a \implies q \geq 0$ (cf. Sec. 12.2.4). There remains $a = m$, which implies $\ell = 2m$. Hence we conclude

$$\boxed{r_0 = m \iff \left\{ \begin{array}{l} a = m \\ \ell = 2m. \end{array} \right.} \tag{12.50}$$

We shall discuss this case further in Sec. 12.4.4, as well as in Chap. 13, which is devoted to the extreme Kerr black hole ($a = m$). In the remainder of this section, we assume $r_0 \neq m$. Equation (12.48b) is then equivalent to

$$\tilde{q} = \frac{2r_0(r_0^2 + a^2 - a\ell)}{r_0 - m}. \tag{12.51}$$

Substituting this relation into Eq. (12.48a), we get an equation involving ℓ only:

$$(r_0^2 + a^2 - a\ell)\left(r_0^2 + a^2 - a\ell - \frac{2r_0}{r_0 - m}\Delta_0\right) = 0.$$

The two solutions are immediate:

$$\ell = a + \frac{r_0^2}{a} \tag{12.52}$$

or

$$\ell = a + \frac{r_0}{a(r_0 - m)}\left[r_0(r_0 - m) - 2\Delta_0\right]. \tag{12.53}$$

The solution (12.52), once inserted in (12.51) leads to $\tilde{q} = 0$, i.e. to $q = -(\ell - a)^2 \leq 0$. Now, $q < 0$ is excluded since Eq. (12.52) implies $|\ell| \geq a$ (cf. Sec. 12.2.4). There remains $q = 0$, which yields $\ell = a$. Then Eq. (12.52) leads to $r_0 = 0$. However, according to the results stated in Sec. 12.2.4, $q = 0$ and $\ell = a$ imply that the geodesic is confined in the equatorial plane, where $r_0 = 0$ corresponds to the ring singularity. We thus conclude that the solution (12.52) is not permitted. The remains then the solution (12.53). Substituting it into Eq. (12.51), we get

$$\tilde{q} = \frac{4r_0^2\Delta_0}{(r_0 - m)^2}, \tag{12.54}$$

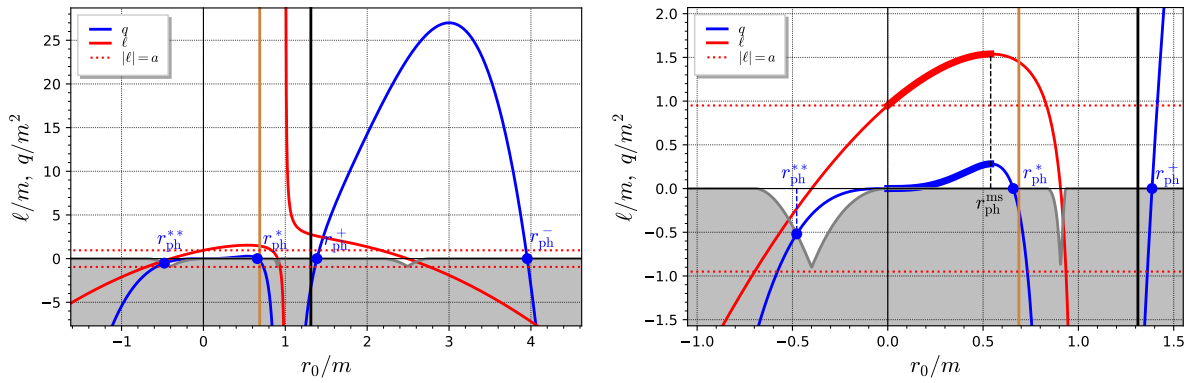


Figure 12.6: Functions $\ell_{\rm c}(r_0)$ (in red) and $q_{\rm c}(r_0)$ (in blue) giving the reduced angular momentum and reduced Carter constant of a spherical photon orbit of radius r_0, according to Eqs. (12.56) and (12.57) and for $a = 0.95\,m$. The right figure is a zoom on the part $-m \leq r_0 \leq 3m/2$. The thick vertical lines mark the two horizons: $\mathscr{H}$ (black) and $\mathscr{H}_{\rm in}$ (light brown). The horizontal dotted lines mark the boundary of the region $|\ell| < a$, where q can take negative values. Values of q in the grey zone are nonphysical, i.e. do not fulfill the conditions (12.58), so that only values of r_0 for which the blue curve lies above the grey zone correspond to spherical photon orbits. This occurs for $r_{\rm ph}^{**} \leq r_0 \leq r_{\rm ph}^{*}$ and $r_{\rm ph}^{+} \leq r_0 \leq r_{\rm ph}^{-}$. The thick segments of the ℓ and q curves correspond to stable orbits, for which $0 \leq r_0 \leq r_{\rm ph}^{\rm ms}$. *[Figure generated by the notebook D.4.11]*

so that Eqs. (12.49) and (12.53) yield, after simplifications,

$$q = \frac{r_0^3}{a^2(r_0-m)^2}\left[-r_0^3 + 6mr_0^2 - 9m^2 r_0 + 4a^2 m\right]. \tag{12.55}$$

Recasting expressions (12.53) and (12.55), we conclude that a spherical photon orbit at $r = r_0 \neq m$ has a reduced angular momentum ℓ and a reduced Carter constant q given by

$$\boxed{\ell = \ell_{\rm c}(r_0) := \frac{r_0^2(3m-r_0) - a^2(r_0+m)}{a(r_0-m)}}, \tag{12.56}$$

$$\boxed{q = q_{\rm c}(r_0) := \frac{r_0^3}{a^2(r_0-m)^2}\left[4a^2 m - r_0(r_0-3m)^2\right]}. \tag{12.57}$$

For a solution to exist, the constraint (12.20) must be obeyed; it writes $\tilde{q} \geq 0$. Given expression (12.54) for $\tilde{q}$, we see that it is equivalent $\Delta_0 \geq 0$. Hence spherical photon orbits do not exist in region $\mathscr{M}_{\rm II}$ of Kerr spacetime. This is in agreement with the general result (11.88).

Equations (12.56) and (12.57) provide the general solution to the system $\mathcal{R}(r_0) = 0$ and $\mathcal{R}'(r_0) = 0$ [Eq. (12.47)], but not all of them correspond to spherical photon orbits. Indeed, we do not expect spherical photon orbits to exist for any value of r_0, in particular for $|r_0| \gg m$. Actually, not all values of q given by Eq. (12.57) are permitted, but only those that fulfill the constraints established in Sec. 12.2.4, namely

$$q \geq 0 \quad \text{if } |\ell| \geq a \tag{12.58a}$$

$$q \geq -(a - |\ell|)^2 \quad \text{if } |\ell| < a. \tag{12.58b}$$

The solutions ℓ and q given by Eqs. (12.56)-(12.57) are plotted as functions of r_0 for $a = 0.95\,m$ in Fig. 12.6, where the region excluded by (12.58) is colored in grey. Consequently spherical photon orbits exist only for values of r_0 for which the q curve (in blue) lies above the grey region. We see that this occurs in three intervals:

$$\boxed{r_0 \in [r_{\rm ph}^{**}, 0)}, \quad \boxed{r_0 \in (0, r_{\rm ph}^{*}]} \quad \text{and} \quad \boxed{r_0 \in [r_{\rm ph}^{+}, r_{\rm ph}^{-}]}, \tag{12.59}$$

where $r_{\rm ph}^{*}$, $r_{\rm ph}^{+}$ and $r_{\rm ph}^{-}$ are the three (ordered) roots distinct from 0 of the equation $q_{\rm c}(r_0) = 0$ (boundary for condition (12.58a)) and $r_{\rm ph}^{**}$ is the unique root of the equation $q_{\rm c}(r_0) = -\left(a - |\ell_{\rm c}(r_0)|\right)^2$ when $|\ell_{\rm c}(r_0)| < a$ (boundary for condition (12.58b)). The above reasoning is based on Fig. 12.6, which has been drawn for $a = 0.95\,m$; however the conclusions (12.59) hold for any value of a (see the notebook D.4.11 for figures with $a = 0.5\,m$ or $a = 0.998\,m$). We have excluded $r_0 = 0$ in (12.59) because it would yield $\ell = a$ and $q = 0$ following Eqs. (12.56)-(12.57). But according to the results of Sec. 12.2.4, such an orbit would be confined to the equatorial plane, where $r = 0$ is not permitted (the ring singularity!).

Given expression (12.57) for q, we see that $r_{\rm ph}^{*}$, $r_{\rm ph}^{+}$ and $r_{\rm ph}^{-}$ are the three roots of the cubic equation

$$r_0(r_0 - 3m)^2 - 4a^2 m = 0. \tag{12.60}$$

We can solve this equation by bringing it to a depressed form in order to use Viète's formulas (8.22). However, we may rely on an already solved equation by noticing the following equivalences:

$$\begin{aligned} r_0(r_0 - 3m)^2 - 4a^2 m = 0 \quad &\iff \quad r_0 \geq 0 \quad \text{and} \quad \sqrt{r_0}|r_0 - 3m| = 2a\sqrt{m} \\ &\iff \quad r_0 \geq 0 \quad \text{and} \quad \sqrt{r_0}(r_0 - 3m) \pm 2a\sqrt{m} = 0 \\ &\iff \quad r_0 \geq 0 \quad \text{and} \quad r_0^{3/2} - 3m r_0^{1/2} \pm 2a\sqrt{m} = 0, \end{aligned}$$

where $\pm$ is $+$ for $r_0 \leq 3m$ and $-$ for $r_0 \geq 3m$. We recognize the function of r_0 which appears in the left-hand side of Eq. (11.132). As shown in Sec. 11.5.2, there are three real roots, $r_{\rm ph}^{*}$, $r_{\rm ph}^{+}$ and $r_{\rm ph}^{-}$, with $r_{\rm ph}^{*}, r_{\rm ph}^{+} \leq 3m$ $(\pm = +)$ and $r_{\rm ph}^{-} \geq 3m$ $(\pm = -)$. They are given by Eqs. (11.135) and (11.134):

$$\boxed{r_{\rm ph}^{*} := 4m\cos^2\left[\frac{1}{3}\arccos\left(-\frac{a}{m}\right) + \frac{4\pi}{3}\right]}. \tag{12.61}$$

$$\boxed{r_{\rm ph}^{\pm} := 4m\cos^2\left[\frac{1}{3}\arccos\left(\mp\frac{a}{m}\right)\right]} \tag{12.62}$$

As for $r_{\rm ph}^{**}$, since $q_{\rm c}(r_0) = -(a - |\ell_{\rm c}(r_0)|)^2$ with $|\ell_{\rm c}(r_0)| < a$ occurs in a region where $\ell_{\rm c}(r_0) < 0$ (cf. Fig. 12.6), we get that $r_{\rm ph}^{**}$ is a solution of the equation $q_{\rm c}(r_0) = -(a + \ell_{\rm c}(r_0))^2$. Given expressions (12.56) and (12.57) for respectively $\ell_{\rm c}(r_0)$ and $q_{\rm c}(r_0)$, we get that $r_{\rm ph}^{**}$ is a root of the cubic equation

$$2r_0^3 - 3mr_0^2 + a^2 m = 0. \tag{12.63}$$

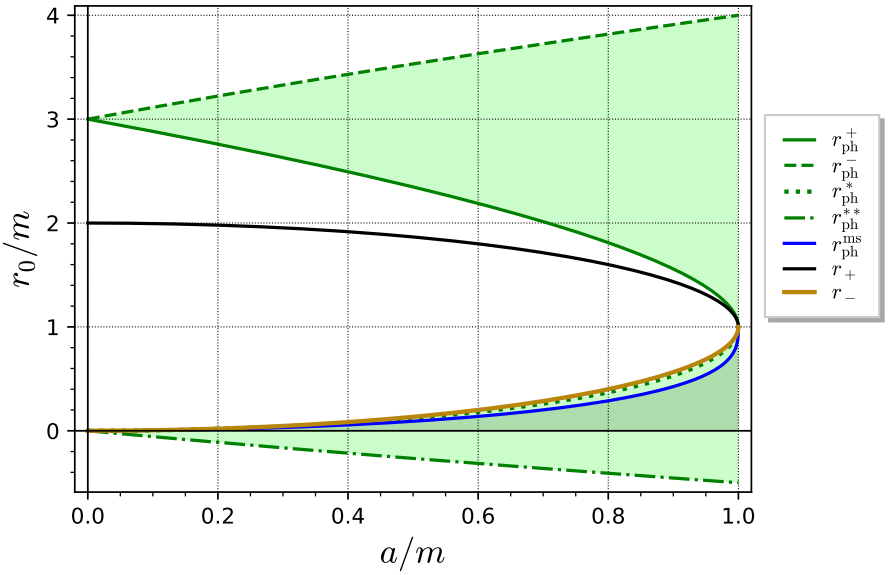


Figure 12.7: Domain of existence of spherical photon orbits in the (a, r_0) plane (in green). The boundaries of the domain are the radii $r_{\rm ph}^{**}$, $r_{\rm ph}^{*}$, $r_{\rm ph}^{+}$ and $r_{\rm ph}^{-}$ given by Eqs. (12.64), (12.61) and (12.62). The shaded area correspond to stable spherical orbits; its upper boundary (blue curve) is the radius $r_{\rm ph}^{\rm ms}$ given by Eq. (12.78). The black curve indicates the black hole horizon $\mathscr{H}$ and the light brown one the inner horizon $\mathscr{H}_{\rm in}$. *[Figure generated by the notebook D.4.11]*

By the change of variable $r_0 =: x + m/2$, we turn this equation into a depressed one:

$$x^3 - \frac{3}{4}m^2 x + \frac{m}{2}\left(a^2 - \frac{m^2}{2}\right) = 0,$$

i.e. $x^3 + px + q = 0$, with $p := -3m^2/4$ and $q := (m/2)(a^2 - m^2/2)$. The discriminant is $-(4p^3 + 27q^2) = 27a^2m^2(m^2 - a^2)/4 \geq 0$. The three roots $(x_k)_{k\in\{0,1,2\}}$ are then all real and are given by Viète's formula (8.22). Only the root x_1 leads to a negative value of r_0, which is the value we are looking for (cf. Fig. 12.6). Viète's formula (8.22) with $k = 1$ yields

$$x_1 = m\cos\left[\frac{1}{3}\arccos\left(1 - 2\frac{a^2}{m^2}\right) + \frac{2\pi}{3}\right] = m\cos\left[\frac{2}{3}\arcsin\left(\frac{a}{m}\right) + \frac{2\pi}{3}\right],$$

from which we obtain

$$\boxed{r_{\rm ph}^{**} = \frac{m}{2} + m\cos\left[\frac{2}{3}\arcsin\left(\frac{a}{m}\right) + \frac{2\pi}{3}\right]}. \tag{12.64}$$

The four critical radii $r_{\rm ph}^{**}$, $r_{\rm ph}^{*}$, $r_{\rm ph}^{+}$ and $r_{\rm ph}^{-}$ are plotted in terms of a in Fig. 12.7. We notice that

$$\boxed{-\frac{m}{2} \leq r_{\rm ph}^{**} \leq 0 \leq r_{\rm ph}^{*} \leq r_{-} \leq m \leq r_{+} \leq r_{\rm ph}^{+} \leq 3m \leq r_{\rm ph}^{-} \leq 4m}, \tag{12.65}$$

with the limits (11.136) and (11.137). In addition,

$$\lim_{a\to 0} r_{\rm ph}^{**} = 0 \qquad \text{and} \qquad \lim_{a\to m} r_{\rm ph}^{**} = -\frac{m}{2}. \tag{12.66}$$

As already stressed in Sec. 11.5.2, $r_{\rm ph}^{*}$ is lower than, but very close to, the inner horizon radius r_-, with $\max(r_- - r_{\rm ph}^{*}) \simeq 0.032\,m$, achieved for $a \simeq 0.9\,m$. In view of the above inequalities and the ranges (12.59), we conclude that

Spherical photon orbits with $E \neq 0$ exist in two regions of Kerr spacetime:

- orbits with $r_0 \in [r_{\rm ph}^{**}, 0) \cup (0, r_{\rm ph}^{*}]$ are located in $\mathscr{M}_{\rm III}$; we shall call them the ***inner spherical photon orbits***;
- orbits with $r_0 \in [r_{\rm ph}^{+}, r_{\rm ph}^{-}]$ are located in $\mathscr{M}_{\rm I}$; we shall call them the ***outer spherical photon orbits***.

Sign of E

All $E \neq 0$ spherical photon orbits lie in $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$. We may then apply the general result (11.60) to them. Since $L = E\ell$, we get $E(r_0^2 + a^2 - a\ell) > 0$. Hence, using formula (12.56) for ℓ,

$$\begin{aligned} E > 0 &\iff r_0^2 + a^2 - a\ell > 0 \\ &\iff \frac{r_0\Delta_0}{r_0 - m} > 0 \\ &\iff \frac{r_0}{r_0 - m} > 0 \\ &\iff r_0 > m \text{ or } r_0 < 0, \end{aligned}$$

where the third line follows from $\Delta_0 > 0$ in $\mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$. In view of (12.65), we conclude that

All outer spherical photon orbits have $E > 0$, as well as inner spherical photon orbits with $r_0 \in [r_{\rm ph}^{**}, 0)$, while inner spherical photon orbits with $r_0 \in (0, r_{\rm ph}^{*}]$ have $E < 0$.

12.3.2 Latitudinal motion

The latitudinal motion of spherical photon orbits is deduced from the general results of Sec. 12.2.4, where geodesics with $\ell = 0$ appear as a special case. We thus treat this case first.

Polar spherical photon orbits

We see on Fig. 12.6 that the reduced conserved angular momentum ℓ (red curve) vanishes at two places in the physically allowed range of r_0, i.e. where the blue curve lies above

the grey region: $r_0 = r_{\rm ph}^{\rm pol,in}$ and $r_0 = r_{\rm ph}^{\rm pol}$ with $r_{\rm ph}^{**} < r_{\rm ph}^{\rm pol,in} < 0$ and $r_{\rm ph}^{+} < r_{\rm ph}^{\rm pol} < r_{\rm ph}^{-}$. The value $r_{\rm ph}^{\rm pol,in}$ (resp. $r_{\rm ph}^{\rm pol}$) corresponds thus to inner (resp. outer) spherical orbits. The results of Sec. 12.2.4 show that these orbits are the only ones that encounter the rotation axis; moreover, they cross it repeatedly. We therefore call them ***polar spherical photon orbits***.

The values of $r_{\rm ph}^{\rm pol,in}$ and $r_{\rm ph}^{\rm pol}$ are obtained by solving $\ell_{\rm c}(r_0) = 0$, with $\ell_{\rm c}(r_0)$ given by Eq. (12.56). We get the cubic equation $r_0^3 - 3mr_0^2 + a^2 r_0 + a^2 m = 0$. Setting $r_0 =: x + m$ reduces it to the depressed cubic $x^3 + px + q = 0$, with $p := a^2 - 3m^2$ and $q := 2m(a^2 - m^2)$. The discriminant $\Delta = -4p^3 - 27q^2$ being positive, the solutions are provided by Viète's formulas, Eq. (8.22), with $k = 0$ and $k = 1$ ($k = 2$ leads to the third root of $\ell_{\rm c}(r_0)$ seen on Fig. 12.6, which lies in the nonphysical range of r_0 — blue curve in the grey region):

$$r_{\rm ph}^{\rm pol} = m + 2\sqrt{m^2 - \frac{a^2}{3}} \cos\left[\frac{1}{3}\arccos\left(\frac{m(m^2 - a^2)}{(m^2 - a^2/3)^{3/2}}\right)\right] \tag{12.67a}$$

$$r_{\rm ph}^{\rm pol,in} = m + 2\sqrt{m^2 - \frac{a^2}{3}} \cos\left[\frac{1}{3}\arccos\left(\frac{m(m^2 - a^2)}{(m^2 - a^2/3)^{3/2}}\right) + \frac{2\pi}{3}\right]. \tag{12.67b}$$

As a function of a, $r_{\rm ph}^{\rm pol}$ decays monotonically from $\lim_{a\to 0} r_{\rm ph}^{\rm pol} = 3m$ to $\lim_{a\to m} r_{\rm ph}^{\rm pol} = (1 + \sqrt{2})m \simeq 2.414214\, m$, while $r_{\rm ph}^{\rm pol,in}$ decays monotonically from $\lim_{a\to 0} r_{\rm ph}^{\rm pol,in} = 0$ to $\lim_{a\to m} r_{\rm ph}^{\rm pol,in} = (1 - \sqrt{2})m \simeq -0.414214\, m$.

Sign of ℓ and L

For spherical orbits with $\ell \neq 0$, the sign of ℓ is deduced from expression (12.56), whose numerator has a sign governed by the position of r_0 with respect to the roots $r_{\rm ph}^{\rm pol,in}$ and $r_{\rm ph}^{\rm pol}$ determined above and whose denominator is positive iff $r_0 > m$, which occurs only for outer spherical orbits. We get then

- Outer spherical orbits with $r_0 \in [r_{\rm ph}^{+}, r_{\rm ph}^{\rm pol})$ have $\ell > 0$;
- outer spherical orbits with $r_0 \in (r_{\rm ph}^{\rm pol}, r_{\rm ph}^{-}]$ have $\ell < 0$;
- inner spherical orbits with $r_0 \in (r_{\rm ph}^{\rm pol,in}, r_{\rm ph}^{*}]$ have $\ell > 0$;
- inner spherical orbits with $r_0 \in [r_{\rm ph}^{**}, r_{\rm ph}^{\rm pol,in})$ have $\ell < 0$.

The sign of L is deduced from that of $\ell = L/E$ by combining with the sign of E obtained in Sec. 12.3.1.

Latitudinal motion

By applying the results of Sec. 12.2.4, we get

- All outer spherical photon orbits with $r_0 \notin \left\{r_{\rm ph}^+, r_{\rm ph}^-\right\}$ and the inner ones with $r_0 \in (0, r_{\rm ph}^*)$ have $q > 0$ (cf. Fig. 12.6); they therefore cross the equatorial plane. Moreover,
  - those with $r_0 = r_{\rm ph}^{\rm pol}$ (outer polar spherical photon orbits) have $\ell = 0$ and cross repeatedly the rotation axis and the equatorial plane, with θ taking all values in the range $[0, \pi]$;
  - those with $r_0 \neq r_{\rm ph}^{\rm pol}$ oscillate about the equatorial plane, having two θ-turning points symmetrical about it, at $\theta = \theta_{\rm m}$ and $\theta = \pi - \theta_{\rm m}$, with $\theta_{\rm m}$ given by Eq. (12.41), in which ℓ and q are to be considered as the functions (12.56)-(12.57) of r_0.
- Inner spherical photon orbits with $r_0 \in (r_{\rm ph}^{**}, 0)$ have $q < 0$ (cf. right panel of Fig. 12.6); they are thus vortical and never encounter the equatorial plane. Moreover,
  - those with $r_0 = r_{\rm ph}^{\rm pol,in}$ (inner polar spherical photon orbits) have $\ell = 0$ and oscillate about the rotation axis, with a θ-turning point at $\theta = \theta_{\rm v}$ (Northern hemisphere) or $\theta = \pi - \theta_{\rm v}$ (Southern hemisphere), where $\theta_{\rm v}$ is given by Eq. (12.43) in which q is to be considered as the function (12.57) of $r_{\rm ph}^{\rm pol,in}$.
  - those with $r_0 \neq r_{\rm ph}^{\rm pol,in}$ neither encounter the rotation axis nor the equatorial plane, having either $\theta \in [\theta_{\rm m}, \theta_{\rm v}]$ (Northern hemisphere) or $\theta \in [\pi - \theta_{\rm v}, \pi - \theta_{\rm m}]$ (Southern hemisphere), with $\theta_{\rm m}$ and $\theta_{\rm v}$ given by Eqs. (12.41) and (12.42), in which ℓ and q are to be considered as the functions (12.56)-(12.57) of r_0.

Spherical photon orbits at $r_0 = r_{\rm ph}^{**}$, $r_{\rm ph}^*$, $r_{\rm ph}^+$ or $r_{\rm ph}^-$, which have been excluded from the above list, will be discussed in details in Sec. 12.3.3.

In addition, we have

For a spherical photon orbit, θ is either a constant or a periodic function of the affine parameter λ, the period being

$$\Lambda_\theta = \frac{2}{|E|} \int_{\theta_{\rm min}}^{\theta_{\rm max}} \frac{r_0^2 + a^2 \cos^2\theta}{\sqrt{\tilde{\Theta}(\theta)}} \, {\rm d}\theta, \tag{12.68}$$

with $(\theta_{\rm min}, \theta_{\rm max}) = (0, \pi)$ (outer polar orbit), $(0, \theta_{\rm v})$ (Northern inner polar orbit), $(\pi - \theta_{\rm v}, \pi)$ (Southern inner polar orbit), $(\theta_{\rm m}, \pi - \theta_{\rm m})$ (non-polar orbit with $r_0 \in (0, r_{\rm ph}^*) \cup (r_{\rm ph}^+, r_{\rm ph}^-)$), $(\theta_{\rm m}, \theta_{\rm v})$ (Northern non-polar vortical inner orbit) or $(\pi - \theta_{\rm v}, \pi - \theta_{\rm m})$ (Southern non-polar vortical inner orbit).

Proof. Switching from the Mino parameter λ' to the affine parameter λ via the relation

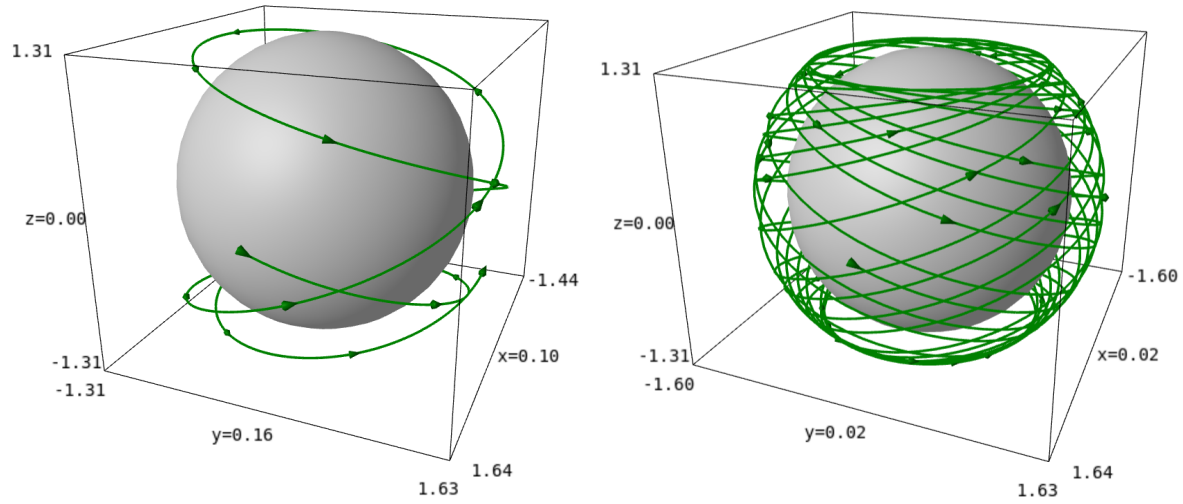


Figure 12.8: Spherical photon orbit at $r_0 = 1.6\,m$ around a Kerr black hole with $a = 0.95\,m$, depicted in terms of the Cartesian Boyer-Lindquist coordinates (x, y, z) defined by Eq. (11.64) and scaled in units of m. The geodesic starts at $\lambda = 0$ and $\varphi = 0$ in the equatorial plane, in the direction of the Southern hemisphere ($\mathrm{d}\theta/\mathrm{d}\lambda > 0$). The left panel corresponds to $0 \leq \lambda \leq 7.7\,m/E$, while the right one extends the range to $0 \leq \lambda \leq 70\,m/E$. The grey sphere is the black hole event horizon. *[Figure generated by the notebook D.4.12]*

(11.49) with $r(\lambda) = r_0$, we may rewrite the equation of motion (12.23c) as

$$\left(\frac{\mathrm{d}\theta}{\mathrm{d}\lambda}\right)^2 + \mathcal{V}(\theta) = 0, \quad \text{with } \mathcal{V}(\theta) := -\frac{E^2\tilde{\Theta}(\theta)}{(r_0^2 + a^2\cos^2\theta)^2}.$$

This is the 1-dimensional equation of motion in the potential well $\mathcal{V}$. It is then clear that $\theta(\lambda)$ is a periodic function. The period Λ_θ is evaluated by integrating $\mathrm{d}\lambda = \epsilon_\theta(r_0^2 + a^2\cos^2\theta)^2/(|E|\sqrt{\tilde{\Theta}(\theta)})$ over a "round-trip" to the same values of θ and $\mathrm{d}\theta/\mathrm{d}\lambda$. □

Remark 2: The azimuthal coordinate φ is not a periodic function of the affine parameter λ in general, nor of the Mino parameter λ'. An exception regards polar spherical geodesics, since setting $\ell = 0$ in the equation of motion (12.23d) results in $\mathrm{d}\varphi/\mathrm{d}\lambda' = \text{const}$, so that φ is a periodic function of λ' (but still not of λ).

Example 6 (Outer spherical photon orbits): Let us consider a Kerr black hole with $a = 0.95\,m$ (same value as in Fig. 12.6). The event and inner horizon radii are $r_+ = 1.312\,m$ and $r_- = 0.688\,m$ and one has $r_{\rm ph}^{**} = -0.478\,m$, $r_{\rm ph}^{*} = 0.658\,m$, $r_{\rm ph}^{+} = 1.386\,m$ and $r_{\rm ph}^{-} = 3.955\,m$. Some outer spherical photon orbits are plotted in Figs. 12.8–12.11 for various values of r_0:

- $r_0 = 1.6\,m$ (Fig. 12.8): this orbit has $\ell = 2.171\,m$, $q = 5.976\,m^2$, $\theta_{\rm m} = 0.706$ rad and $\mathrm{d}\varphi/\mathrm{d}\lambda > 0$;
- $r_0 = 2.8\,m$ (Fig. 12.9): this orbit has $\ell = -1.089\,m$, $q = 26.260\,m^2$ and $\theta_{\rm m} = 0.206$ rad; $\varphi(\lambda)$ is not a monotonous function (cf. the projection onto the xy-plane): $\mathrm{d}\varphi/\mathrm{d}\lambda < 0$ almost everywhere, except near the equator, where the Lense-Thirring effect (cf. Sec. 11.3.3) is the strongest[2] and enforces $\mathrm{d}\varphi/\mathrm{d}\lambda > 0$;

[2] This can be seen by considering the limit $\theta \to \pi/2$ in Eq. (12.23d).

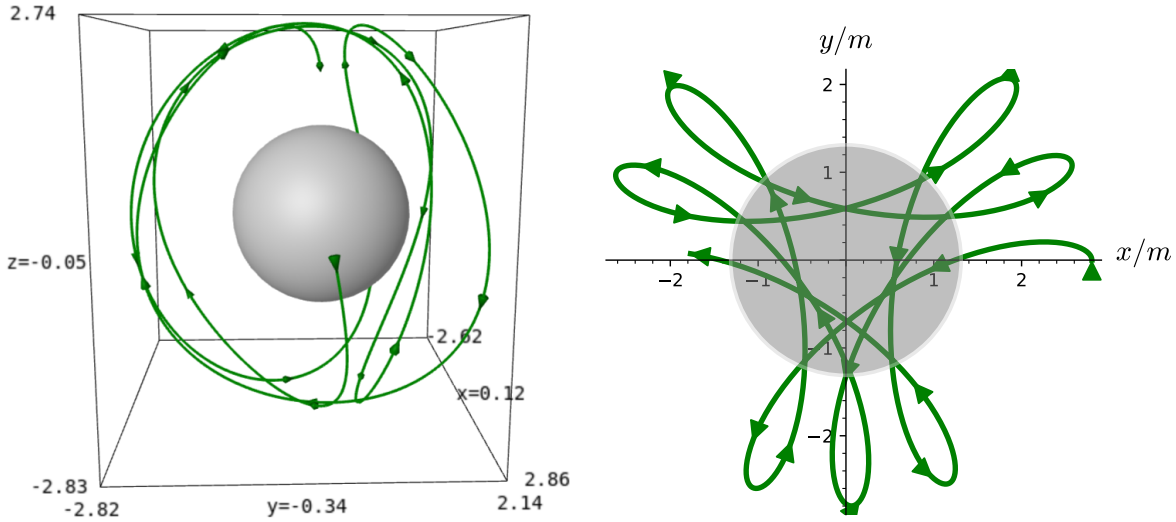


Figure 12.9: Spherical photon orbit at $r_0 = 2.8\,m$ around a Kerr black hole with $a = 0.95\,m$, depicted in terms of the Cartesian Boyer-Lindquist coordinates (x, y, z) defined by Eq. (11.64) and scaled in units of m, with the grey sphere representing the black hole event horizon. The right panel depicts the projection of the orbit onto the xy-plane (the overlap with the grey area is a mere projection effect, since of course the orbit never crosses the event horizon). The geodesic starts at $\lambda = 0$ and $\varphi = 0$ in the equatorial plane, in the direction of the Southern hemisphere. The plotted range is $0 \leq \lambda \leq 38\,m/E$. *[Figure generated by the notebook D.4.12]*

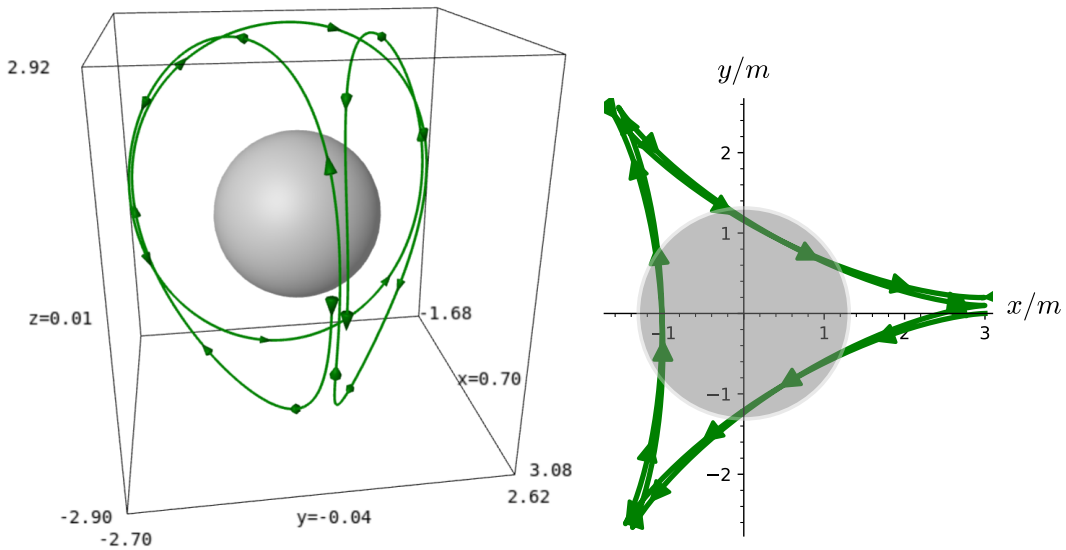


Figure 12.10: Same as in Fig. 12.9 but for a spherical photon orbit at $r_0 = 3m$, with $0 \leq \lambda \leq 32\,m/E$. *[Figure generated by the notebook D.4.12]*

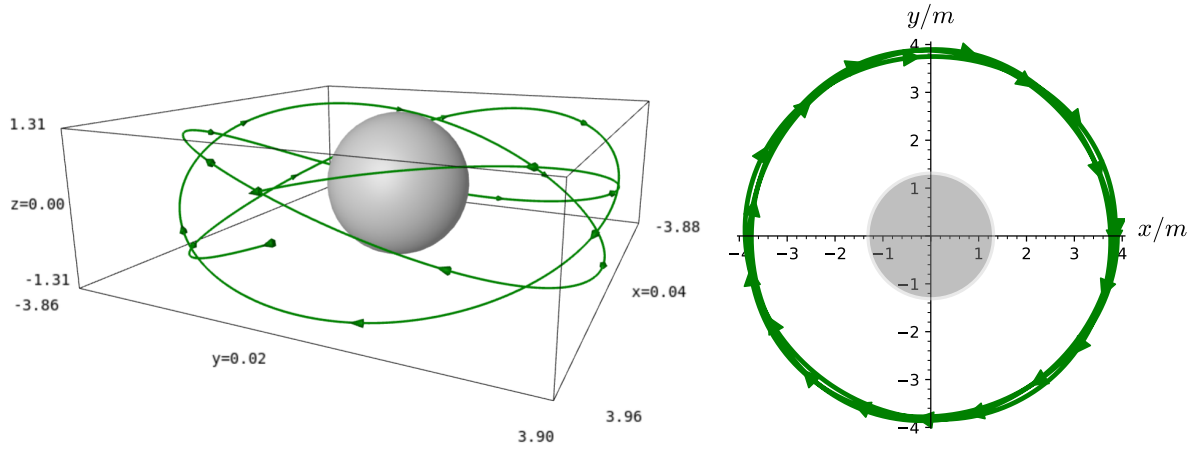


Figure 12.11: Same as in Fig. 12.9 but for a spherical photon orbit at $r_0 = 3.9\,m$, with $0 \leq \lambda \leq 55\,m/E$. *[Figure generated by the notebook D.4.12]*

- $r_0 = 3m$ (Fig. 12.10): this orbit has $\ell = -1.9\,m$, $q = 27\,m^2$, $\theta_{\rm m} = 0.346$ rad and $\mathrm{d}\varphi/\mathrm{d}\lambda \leq 0$, with $\mathrm{d}\varphi/\mathrm{d}\lambda = 0$ in the equatorial plane, as it can be easily checked by plugging $r = 3m$, $\theta = \pi/2$ as well as the above values of ℓ and q into Eq. (12.23d); this orbit maximizes the value of q among all spherical photon orbits (cf. Fig. 12.6 and Eq. (12.83) below);
- $r_0 = 3.9\,m$ (Fig. 12.11): this orbit has $\ell = -6.574\,m$, $q = 3.525\,m^2$, $\theta_{\rm m} = 1.290$ rad and $\mathrm{d}\varphi/\mathrm{d}\lambda < 0$.

The next example is devoted to inner spherical photon orbits. The Cartesian Boyer-Lindquist coordinates (t, x, y, z) are not well suited to depict these orbits, in particular those that have $r < 0$. We could use O'Neill exponential coordinates, as in Fig. 12.2. However, we are going to use instead the radial coordinate $\hat{r}$ and the associated Cartesian-type coordinates $(\hat{x}, \hat{y}, \hat{z})$ defined by

$$\hat{r} := \frac{1}{2}\left(r + \sqrt{r^2 + 4m^2}\right) \tag{12.69a}$$

$$\hat{x} := \hat{r}\sin\theta\cos\varphi, \quad \hat{y} := \hat{r}\sin\theta\sin\varphi, \quad \hat{z} := \hat{r}\cos\theta. \tag{12.69b}$$

As O'Neill coordinates, this brings the whole range $(-\infty, +\infty)$ for r to $(0, +\infty)$ for $\hat{r}$, with $\hat{r} \to 0$ corresponding to $r \to -\infty$. Contrary to O'Neill coordinate $\mathrm{e}^{r/m}$, the coordinate $\hat{r}$ does not enlarge too much the region exterior to the black hole, which makes it better suited for plots covering both the black hole interior and exterior, as in Fig. 12.17 below. Note that in $\mathscr{M}_{\rm I}$, $\hat{r}$ is asymptotically equivalent to r: $\hat{r} \sim r$ as $r \to +\infty$.

Example 7 (Inner spherical photon orbits): We consider the same Kerr spacetime as in Example 6, i.e. $a = 0.95\,m$, but this time, we focus on inner spherical photon orbits:

- $r_0 = 0.540\,m$ (Fig. 12.12): this orbit has $\ell = 1.539\,m$, $q = 0.282\,m^2$, $\theta_{\rm m} = 1.173$ rad $\mathrm{d}\varphi/\mathrm{d}\lambda > 0$, $E < 0$ and $L < 0$; it is actually a marginally stable photon orbit ($r_0 = r_{\rm ph}^{\rm ms}$), as we shall see in Sec. 12.3.4;

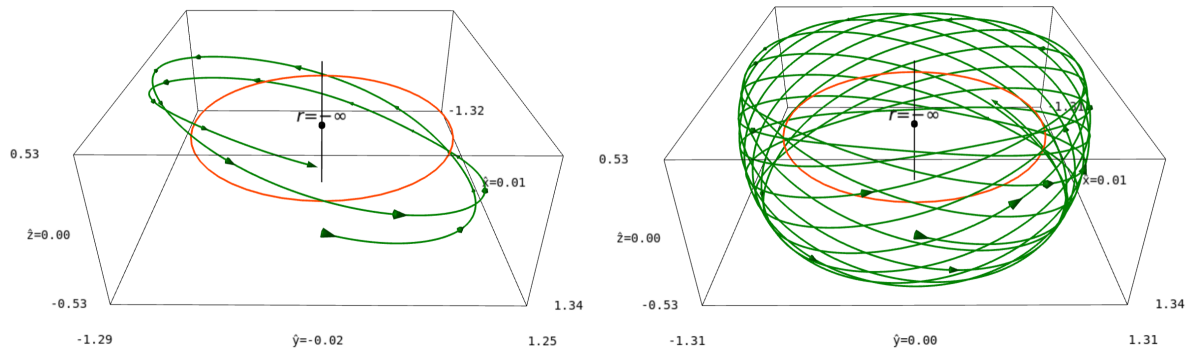


Figure 12.12: Marginally stable spherical photon orbit inside a Kerr black hole with $a = 0.95\,m$, depicted in terms of the coordinates $(\hat{x}, \hat{y}, \hat{z})$ defined by Eq. (12.69) and scaled in units of m. The geodesic is located at $r_0 = 0.540\,m$ and starts at $\lambda = 0$ and $\varphi = 0$ in the equatorial plane, in the direction of the Southern hemisphere ($\mathrm{d}\theta/\mathrm{d}\lambda > 0$). The left panel corresponds to $0 \leq \lambda \leq 3\,m/|E|$, while the right one extends the range to $0 \leq \lambda \leq 20, m/|E|$. The orange red circle is the ring singularity and the vertical black line is the rotation axis, with a black dot marking the asymptotic end $r \to -\infty$. *[Figure generated by the notebook D.4.12]*

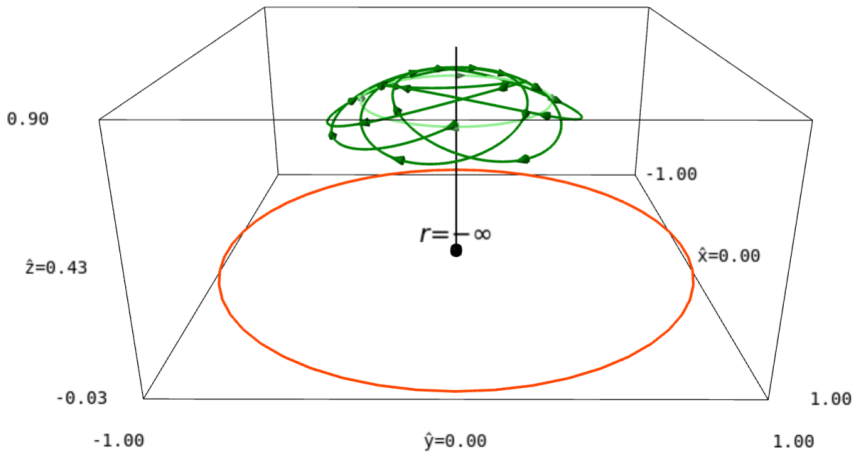


Figure 12.13: Same as in Fig. 12.12 but for a vortical spherical photon orbit at $r_0 = -0.46\,m$, with $0 \leq \lambda \leq 20\,m/E$. Also shown is the Northern vortical circular photon orbit at $r = r_{\mathrm{ph}}^{**}$ (in light green). *[Figure generated by the notebook D.4.12]*

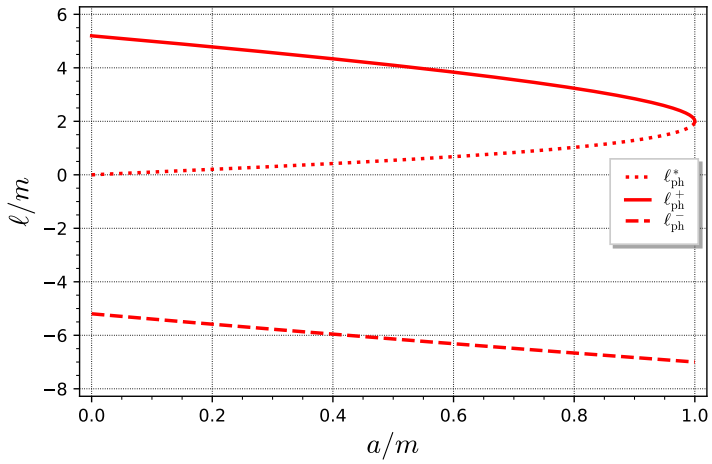


Figure 12.14: Reduced angular momentum ℓ of the three circular photon orbits in the equatorial plane, as a function of the Kerr spin parameter a. *[Figure generated by the notebook D.4.11]*

- $r_0 = -0.46\,m$ (Fig. 12.13): this orbit has $\ell = -0.176\,m$, $q = -0.461\,m^2$ (hence it is vortical), $\theta_{\rm m} = 0.282$ rad, $\theta_{\rm v} = 0.731$ rad, $\mathrm{d}\varphi/\mathrm{d}\lambda < 0$, $E > 0$ and $L < 0$;
- $r_0 = -0.478\,m = r_{\rm ph}^{**}$ (Fig. 12.13, light green curve): this orbit has $\ell = -0.229\,m$, $q = -0.519\,m^2$ (hence it is vortical), $\theta_{\rm m} = \theta_{\rm v} = \theta_{\rm ph}^{**} = 0.514$ rad, $\mathrm{d}\varphi/\mathrm{d}\lambda < 0$, $E > 0$ and $L < 0$.

12.3.3 Circular photon orbits

In any stationary and axisymmetric spacetime, such as the Kerr spacetime, one may define a ***circular photon orbit*** as a null geodesic $\mathscr{L}$ whose tangent vector field $\boldsymbol{p} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ is a linear combination of the two Killing vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ generating respectively the stationarity and the axisymmetry, with a non-vanishing component along $\boldsymbol{\eta}$:

$$\boldsymbol{p} = \alpha\boldsymbol{\xi} + \beta\boldsymbol{\eta}, \quad \text{with } \beta \neq 0. \tag{12.70}$$

Note that the above definition is independent from any coordinate system. It is worth to compare Eq. (12.70) with expression (11.164) for the 4-velocity $\boldsymbol{u} = \mu^{-1}\boldsymbol{p}$ of a circular timelike orbit in the equatorial plane.

Remark 3: Circular photon orbits are sometimes called ***light rings*** [71].

In Kerr spacetime, when using coordinates adapted to the spacetime symmetries, such as Boyer-Lindquist coordinates (t, r, θ, φ), one has $\boldsymbol{\xi} = \boldsymbol{\partial}_t$, $\boldsymbol{\eta} = \boldsymbol{\partial}_\varphi$, $\alpha = p^t$, $\beta = p^\varphi$ and the definition (12.70) is equivalent to $p^r = \mathrm{d}r/\mathrm{d}\lambda = 0$ and $p^\theta = \mathrm{d}\theta/\mathrm{d}\lambda = 0$. It follows immediately that a circular photon orbit is any null geodesic along which both r and θ are constant. It is thus a spherical photon orbit lying at a constant value of θ.

Remark 4: In Schwarzschild spacetime, the photon circular orbits forming the photon sphere at $r = 3m$ (cf. Sec. 8.2.3) do not have $\theta = \text{const}$, except for the orbit in the equatorial plane. However, for a given orbit non-equatorial orbit on the photon sphere, one may use the spherical

symmetry of Schwarzschild spacetime to perform a change of coordinates $(\theta,\varphi)\mapsto(\theta',\varphi')$ such that $\theta'=$ const is constant for that orbit. In Kerr spacetime with $a\neq 0$, such "oblique" circular orbits cannot exist due to Lense-Thirring precession.

Example 8: The spherical photon orbits with $E=0$ discussed in Sec. 12.3.1, namely the null generators of the horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$ are circular photon orbits, since they are the outgoing principal null geodesics $\mathscr{L}^{\rm out,\mathscr{H}}_{(\theta,\psi)}$ and $\mathscr{L}^{\rm out,\mathscr{H}_{\rm in}}_{(\theta,\psi)}$, which have $\theta=$ const.

The spherical photon orbits with $r_0=r_{\rm ph}^*$, $r_{\rm ph}^+$ and $r_{\rm ph}^-$ have $q=0$ and $|\ell|>a$. According to the results of Sec. 12.2.4, they necessary lie in the equatorial plane $\theta=\pi/2$. They are thus circular. According to expression (12.57) for q, they are the only circular orbits in the equatorial plane, because the only other root of $q_{\rm c}(r_0)=0$ is $r_0=0$ (see also Fig. 12.6), which would correspond to the ring singularity. Let us denote by $\ell_{\rm ph}^*$, $\ell_{\rm ph}^+$ and $\ell_{\rm ph}^-$ the reduced angular momentum ℓ of respectively the circular photon orbit at $r_{\rm ph}^*$, $r_{\rm ph}^+$ and $r_{\rm ph}^-$. These values of ℓ are deduced from Eq. (12.56) and are plotted in terms of a in Fig. 12.14. We have $\ell_{\rm ph}^*>0$, $\ell_{\rm ph}^+>0$ and $\ell_{\rm ph}^-<0$. Moreover,

$$\lim_{a\to 0}\ell_{\rm ph}^+=\lim_{a\to 0}\left|\ell_{\rm ph}^-\right|=3\sqrt{3}m\simeq 5.196\,m, \tag{12.71}$$

in agreement with the Schwarzschild result (8.21), while

$$\lim_{a\to m}\ell_{\rm ph}^+=\lim_{a\to m}\ell_{\rm ph}^*=2m, \tag{12.72}$$

in agreement with Eq. (12.50), since $\lim_{a\to m} r_{\rm ph}^+=\lim_{a\to m} r_{\rm ph}^*=m$.

The spherical orbits with $r_0=r_{\rm ph}^{**}$ have $q<0$, i.e. are vortical geodesics. Moreover, they fulfill $q=q_{\rm min}=-(a-|\ell|)^2$. According to the results of Sec. 12.2.4, this corresponds to two orbits at a fixed value of θ, which are symmetrical with respect to the equatorial plane. They lie at $\theta=\theta_{\rm ph}^{**}$ (Northern hemisphere) and $\theta=\pi-\theta_{\rm ph}^{**}$ (Southern hemisphere), where $\theta_{\rm ph}^{**}$ is given by Eq. (12.44), using for ℓ the value (12.56) with $r_0=r_{\rm ph}^{**}$. Since $3m(r_{\rm ph}^{**})^2=2(r_{\rm ph}^{**})^3+a^2m$ by virtue of Eq. (12.63), some simplification occurs and we get

$$\theta_{\rm ph}^{**}=\arcsin\sqrt{\frac{|r_{\rm ph}^{**}|}{m-r_{\rm ph}^{**}}\left(1-\frac{(r_{\rm ph}^{**})^2}{a^2}\right)}. \tag{12.73}$$

$\theta_{\rm ph}^{**}$ is an increasing function of a/m, plotted in Fig. 12.15. We have $\lim_{a\to 0}\theta_{\rm ph}^{**}=0$ (the rotation axis) and $\lim_{a\to m}\theta_{\rm ph}^{**}=\pi/6$. Since they occur at a fixed value of θ, the two vortical spherical photon orbits are actually circular. Such an orbit is shown in Fig. 12.13 (light green circle). The reduced angular momentum and Carter constant of these orbits, denoted by $\ell_{\rm ph}^{**}$ and $q_{\rm ph}^{**}$ respectively, are depicted in terms of a in Fig. 12.16. They tend to zero as $a\to 0$ and obey

$$\lim_{a\to m}\ell_{\rm ph}^{**}=-\frac{m}{4}\qquad\text{and}\qquad\lim_{a\to m}q_{\rm ph}^{**}=-\frac{9}{16}m^2. \tag{12.74}$$

We may summarize the above results by

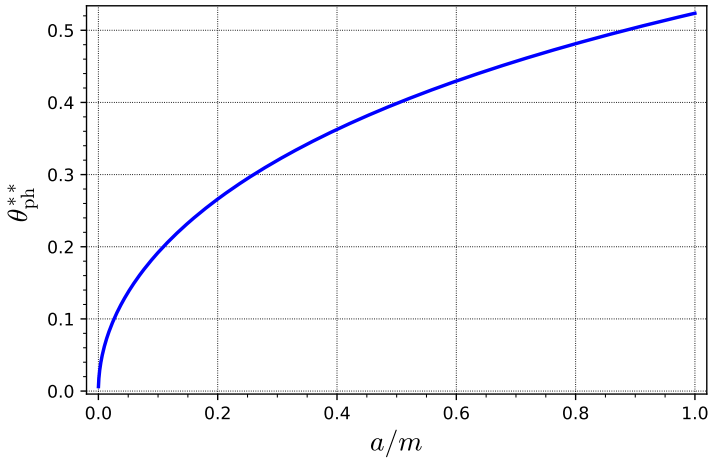


Figure 12.15: Angle θ of the Northern vortical circular photon orbit at $r = r^{**}_{\rm ph}$ as a function of Kerr spin parameter a. *[Figure generated by the notebook D.4.11]*

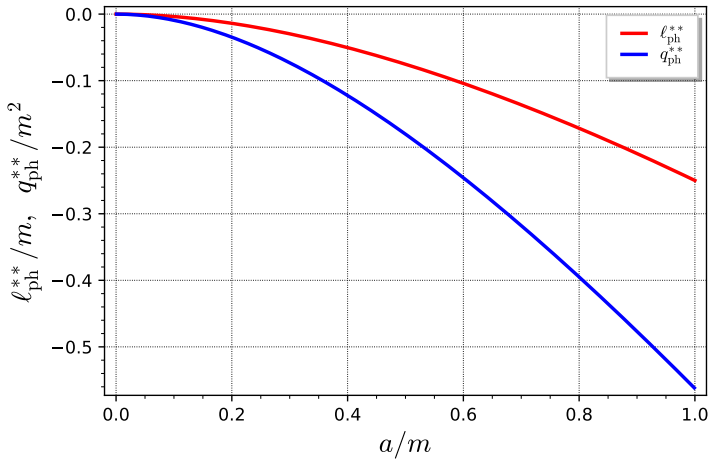


Figure 12.16: Reduced angular momentum $\ell^{**}_{\rm ph}$ and reduced Carter constant $q^{**}_{\rm ph}$ of the two vortical circular photon orbits at $r = r^{**}_{\rm ph}$, as functions of Kerr spin parameter a. *[Figure generated by the notebook D.4.11]*

The geodesics at the boundaries of the domain of existence of spherical photon orbits [Eq. (12.59)] are circular orbits:

- the orbit at $(r,\theta) = (r^{**}_{\rm ph}, \theta^{**}_{\rm ph})$ is called the ***Northern vortical circular photon orbit***;
- the orbit at $(r,\theta) = (r^{**}_{\rm ph}, \pi - \theta^{**}_{\rm ph})$ is called the ***Southern vortical circular photon orbit***;
- the orbit at $(r,\theta) = (r^{*}_{\rm ph}, \pi/2)$ is called the ***equatorial inner circular photon orbit***;
- the orbit at $(r,\theta) = (r^{+}_{\rm ph}, \pi/2)$ is called the ***prograde outer circular photon orbit***;
- the orbit at $(r,\theta) = (r^{-}_{\rm ph}, \pi/2)$ is called the ***retrograde outer circular photon orbit***;

Moreover, there is no other circular photon orbit with $E \neq 0$. The circular photon orbits with $E = 0$ are the null generators of the two Killing horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$.

That the orbits listed above are the only $E \neq 0$ circular orbits follows from the fact that all the other $E \neq 0$ spherical photon orbits have either $q > 0$ or $q_{\rm min} < q < 0$ (cf. Fig. 12.6), which imply that they have a varying θ (cf. Sec. 12.2.4).

12.3.4 Stability of spherical photon orbits

The radial stability of spherical photon orbits is derived by the same argument as that used in Sec. 11.5.3 for timelike circular orbits: a spherical photon orbit at $r = r_0$ is stable iff no geodesic motion with the same values of the conserved quantities ℓ and q is possible in the vicinity of r_0 except for $r = r_0$. Given the same value of (ℓ, q) imply the same polynomial $\mathcal{R}$ and that a geodesic motion is possible only if $\mathcal{R}(r) \geq 0$ [Eq. (12.46)], the above criteria is equivalent to $\mathcal{R}(r) < 0$ for r distinct from r_0 but close to it. Since $\mathcal{R}(r_0) = 0$ and $\mathcal{R}'(r_0) = 0$ [Eq. (12.47)], this is equivalent to $\mathcal{R}$ having a maximum at r_0. In other words:

$$\text{The spherical photon orbit of radius } r_0 \text{ is stable} \iff \mathcal{R}''(r_0) < 0. \tag{12.75}$$

Remark 5: The criterion (11.150) for timelike circular orbits involves $\mathcal{V}''(r_0) > 0$ simply because of the minus sign in the definition of $\mathcal{V}$ from R: $\mathcal{V}(r) := -R(r)/(\mu^2 r^4)$, whereas we have here $\mathcal{R}(r) := R(r)/E^2$.

An immediate consequence of the lemma in Sec. 12.2.5 is

All spherical photon orbits in the black hole exterior are unstable.

Proof. The quartic polynomial $\mathcal{R}(r)$ of spherical photon orbits in $\mathscr{M}_{\rm I}$ belong to case 2 of the lemma, which states that $\mathcal{R}(r) > 0$ for $r \in (r_+, r_0) \cup (r_0, +\infty)$, i.e. r_0 corresponds to a minimum of $\mathcal{R}$ (cf. Fig. 12.5b). □

We are however going to see that there exist stable spherical photon orbits in region $\mathscr{M}_{\rm III}$. From expression (12.25b) for $\mathcal{R}$, we get $\mathcal{R}''(r_0) = 12r_0^2 + 2(a^2 - \ell^2 - q)$. Substituting the values (12.56) and (12.57) for respectively ℓ and q yields

$$\mathcal{R}''(r_0) = \frac{8r_0}{(r_0-m)^2}\left[(r_0-m)^3 + m^3\left(1-\frac{a^2}{m^2}\right)\right]. \tag{12.76}$$

Hence the stability criterion (12.75) becomes

$$\mathcal{R}''(r_0) < 0 \iff \begin{cases} r_0 < 0 \quad \text{and} \quad (r_0-m)^3 + m^3\left(1-\frac{a^2}{m^2}\right) > 0 \\ \text{or} \\ r_0 > 0 \quad \text{and} \quad (r_0-m)^3 + m^3\left(1-\frac{a^2}{m^2}\right) < 0. \end{cases} \tag{12.77}$$

Now

$$(r_0-m)^3 + m^3\left(1-\frac{a^2}{m^2}\right) > 0 \iff r_0 - m > -m\left(1-\frac{a^2}{m^2}\right)^{1/3} \iff r_0 > r_{\rm ph}^{\rm ms},$$

where

$$\boxed{r_{\rm ph}^{\rm ms} := m\left[1-\left(1-\frac{a^2}{m^2}\right)^{1/3}\right]}. \tag{12.78}$$

Since obviously $r_{\rm ph}^{\rm ms} \geq 0$, we conclude that the first case in (12.77) is excluded and that the second case holds for $0 < r_0 < r_{\rm ph}^{\rm ms}$:

$$\text{A spherical photon orbit of radius } r_0 \text{ is stable} \iff 0 < r_0 < r_{\rm ph}^{\rm ms}. \tag{12.79}$$

The index 'ms' in $r_{\rm ph}^{\rm ms}$ stands for ***marginally stable***. $r_{\rm ph}^{\rm ms}$ is plotted as a function of a in Fig. 12.7. By comparing the blue solid curve with the green dotted one in that figure, we note that

$$r_{\rm ph}^{\rm ms} \leq r_{\rm ph}^{*}, \tag{12.80}$$

with equality iff $a = m$. We conclude

All spherical photon orbits are unstable with respect to radial perturbations, except for a subclass of inner orbits with negative energy ($E < 0$): those that have a radius $r_0 \in (0, r_{\rm ph}^{\rm ms})$, where $r_{\rm ph}^{\rm ms}$ is given by Eq. (12.78). In particular, all spherical photon orbits in $\mathscr{M}_{\rm I}$ (the black hole exterior) are unstable, as well as all circular photon orbits discussed in Sec. 12.3.3.

That all stable orbits have $E < 0$ follows from the results on the sign of E obtained in Sec. 12.3.1.

The stable spherical photon orbits have $q > 0$ and $\ell > 0$ (cf. right panel of Fig. 12.6). According to the results of Sec. 12.2.4, they thus oscillate symmetrically about the equatorial plane, between two θ-turning points, $\theta_{\rm m}$ and $\pi - \theta_{\rm m}$, such that $0 < \theta_{\rm m} < \pi/2$. In particular, $r_0 = r_{\rm ms}$ does not correspond to a unique orbit, but to a 1-parameter family of marginally stable orbits; the parameter can be chosen to be the azimuthal coordinate φ at the first value $\theta = \theta_{\rm m}$ (or $\theta = \pi/2$) after $t = 0$. A marginally stable spherical photon orbit is shown in Fig. 12.12.

We have the following property, which appears clearly on Fig. 12.6:

Among all inner spherical photon orbits, the marginally stable orbits at $r_0 = r_{\rm ph}^{\rm ms}$ are those for which the reduced angular momentum ℓ and reduced Carter constant q are maximal.

Proof. From Eqs. (12.56) and (12.57), we get

$$\frac{{\rm d}q_{\rm c}}{{\rm d}r_0} = -\frac{4r_0^2(r_0-3m)\left[(r_0-m)^3 + m^3\left(1-\frac{a^2}{m^2}\right)\right]}{a^2(r_0-m)^3} \tag{12.81}$$

$$\frac{{\rm d}\ell_{\rm c}}{{\rm d}r_0} = -2\frac{(r_0-m)^3 + m^3\left(1-\frac{a^2}{m^2}\right)}{a(r_0-m)^2} \tag{12.82}$$

Since $r_{\rm ph}^{\rm ms}$ is the unique real root of $(r_0-m)^3 + m^3(1-a^2/m^2) = 0$ [cf. Eq. (12.76)], it is clear that the function $\ell_{\rm c}(r_0)$ has a unique extremum, which is achieved by the marginally stable orbits. From the graph of $\ell_{\rm c}(r_0)$ shown in Fig. 12.6, we see that this extremum is a maximum. Regarding the function $q_{\rm c}(r_0)$, the above expression of ${\rm d}q_{\rm c}/{\rm d}r_0$ leads to two extrema: $r_0 = r_{\rm ph}^{\rm ms}$ and $r_0 = 3m$. The former regards the inner spherical orbits, while the latter regards the outer ones. Again, from the graph shown in Fig. 12.6, it is clear that $r_0 = r_{\rm ph}^{\rm ms}$ realizes a maximum of $q_{\rm c}$ among all inner spherical orbits. □

Remark 6: Another proof can be given by using the same general argument as that employed in Sec. 11.5.3 (p. 366) for showing that the ISCO realizes an extremum of the specific energy and specific angular momentum of timelike circular equatorial orbits. Indeed, considering $\mathcal{R}$ as a function of ℓ and q, in addition to r, i.e. writing $\mathcal{R} = \mathcal{R}(r,\ell,q)$, the marginal stable orbit obeys

$$\mathcal{R}(r_0,\ell_{\rm c}(r_0),q_{\rm c}(r_0)) = 0, \quad \frac{\partial\mathcal{R}}{\partial r}(r_0,\ell_{\rm c}(r_0),q_{\rm c}(r_0)) = 0, \quad \frac{\partial^2\mathcal{R}}{\partial r^2}(r_0,\ell_{\rm c}(r_0),q_{\rm c}(r_0)) = 0.$$

Deriving the first and second equations with respect to r_0 and using the third equation, we get a homogeneous linear system for $({\rm d}\ell_{\rm c}/{\rm d}r_0,\, {\rm d}q_{\rm c}/{\rm d}r_0)$, similar to the system (11.162). The unique solution is then $({\rm d}\ell_{\rm c}/{\rm d}r_0,\, {\rm d}q_{\rm c}/{\rm d}r_0) = (0,0)$, yielding the extremum in the functions $\ell_{\rm c}(r_0)$ and $q_{\rm c}(r_0)$ at $r_0 = r_{\rm ph}^{\rm ms}$.

Equation (12.81) shows that, in addition to that at $r_0 = r_{\rm ph}^{\rm ms}$, the function $q_{\rm c}(r_0)$ admits a second extremum at $r_0 = 3m$, i.e. for outer spherical orbits. This extremum is actually

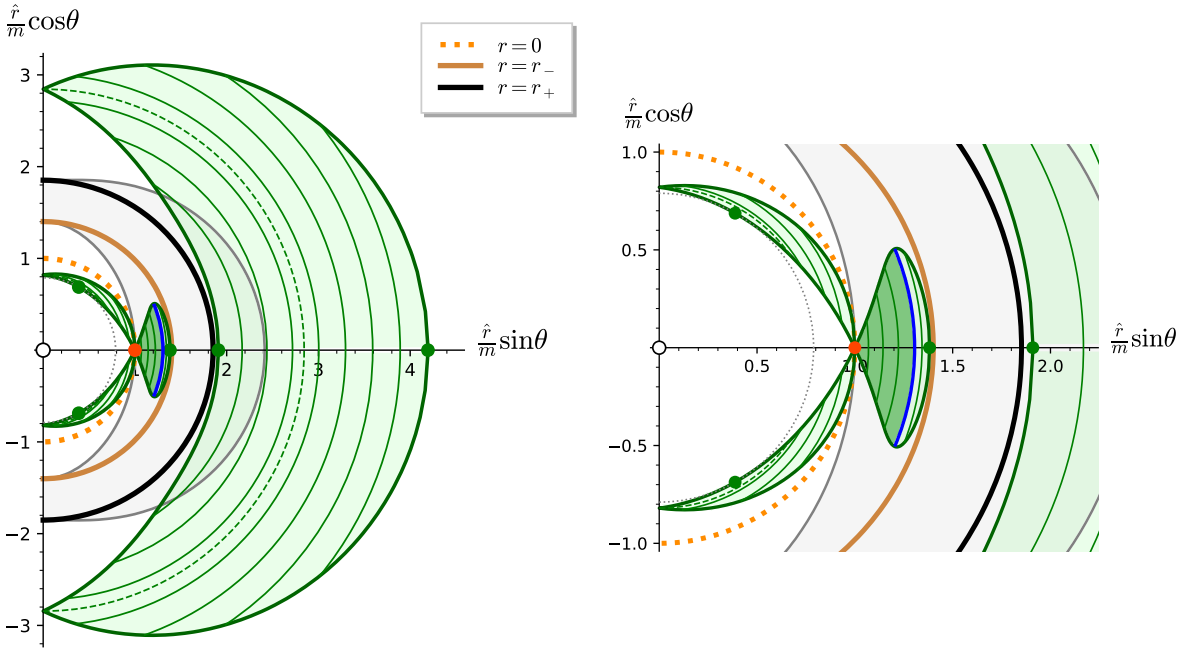


Figure 12.17: Trace of the photon region (pale and dark green areas) in a meridional plane $(t,\varphi)=$ const of a Kerr spacetime with $a=0.95\,m$ (as in Fig. 12.6). The right panel is a zoom on the inner part of the left one. The meridional plane is described in terms of the coordinates $(\hat{r},\theta)$ where $\hat{r}:=(r+\sqrt{r^2+4m^2})/2$ [cf. Eq. (12.69)]. Some spherical photon orbits are shown as green circular arcs, with the dashed ones representing the polar spherical orbits, i.e. $\ell=0$ orbits. The black (resp. light brown) half-circle at $r=r_+$ (resp. $r=r_-$) is the trace of the outer (resp. inner) Killing horizon. The dotted orange half-circle marks the locus of $r=0$, with the red dot indicating the curvature singularity at $r=0$ and $\theta=\pi/2$. The ergoregion is shown in grey. The region of stable spherical orbits is colored in dark green, with its boundary at $r=r_{\rm ph}^{\rm ms}$ drawn in blue. Green dots marks photon circular orbits: from the left to the right, they are the two vortical circular orbits at $r=r_{\rm ph}^{**}$, the equatorial inner circular orbit at $r=r_{\rm ph}^{*}$, the prograde outer circular orbit at $r=r_{\rm ph}^{+}$ and the retrograde outer circular orbit at $r=r_{\rm ph}^{-}$. The thin dotted grey half-circle marks $r=r_{\rm ph}^{**}$. *[Figure generated by the notebook D.4.11]*

a maximum and its value, obtained by setting $r_0=3m$ in Eq. (12.57), turns out to be independent from a:

$$\max q_{\rm c}(r_0)=27m^2. \tag{12.83}$$

This is the maximum of the reduced Carter constant over all the spherical photon orbits (cf. Fig. 12.6).

12.3.5 Photon region

A corollary of the results obtained in Sec. 12.2.5 is that outside the black hole, i.e. in region $\mathscr{M}_{\rm I}$, a photon cannot be trapped (i.e. move in a limited range of r) unless it moves on a (unstable) spherical orbit. The part of Kerr spacetime made of points through which a spherical photon orbit can pass is called the ***photon region*** or sometimes the ***photon shell*** [161]. In terms of the Boyer-Lindquist r-coordinate, the photon region lies in the three intervals (12.59). The range of the θ-coordinate has been discussed in Sec. 12.3.2:

spherical photon orbits with $r_0 > 0$ oscillate about the equatorial plane with the limiting angle $\theta_{\rm m}$ given by Eq. (12.41) ($\theta_{\rm m} = \pi/2$ for the three circular orbits at $r_0 = r_{\rm ph}^*$, $r_0 = r_{\rm ph}^+$ and $r_0 = r_{\rm ph}^-$, which lie in the equatorial plane). On the other side, the inner orbits with $r_0 < 0$ are all vortical, with the limiting angles $\theta_{\rm m}$ and $\theta_{\rm v}$ given by Eqs. (12.41) and (12.42). There is no constraint on the Boyer-Lindquist φ-coordinate: the photon region occupies all the range $[0, 2\pi)$.

The photon region is depicted in Fig. 12.17, which represents a meridional plane $(t, \varphi) = \text{const}$ of a Kerr spacetime with $a = 0.95\,m$ — the same value of a as in Fig. 12.6. Polar spherical photon orbits, at $r_0 = r_{\rm ph}^{\rm pol}$ and $r_0 = r_{\rm ph}^{\rm pol,in}$ (cf. Sec. 12.3.2), are plotted as dashed green curves. It is graphically clear that they are the only orbits that encounter the rotation axis. We also recover from this figure that, apart from those generating the two horizons $\mathscr{H}_{\rm in}$ and $\mathscr{H}$, the only circular photon orbits of Kerr spacetime are the five ones considered in Sec. 12.3.3 (the five green dots). We note also that a part of the outer spherical photon orbits lie in the ergoregion. These orbits have however $E > 0$, according to the results of the end of Sec. 12.3.1.

Historical note: The prograde and retrograde outer circular photon orbits in the equatorial plane of a Kerr black hole have been found by James M. Bardeen, William H. Press and Saul A. Teukolsky in 1972 [19]. The existence of stable spherical photon orbits under the inner horizon of a Kerr black hole has been shown by Zdeněk Stuchlík in 1981 [238]. The systematic study of spherical photon orbits in the black hole exterior has been performed by Edward Teo in 2003 [244].

12.4 Black hole shadow and critical curve

12.4.1 Critical null geodesics

As in the Schwarzschild case (Sec. 8.3.2), let us define a ***critical null geodesic*** as a null geodesic with $E \neq 0$ that has the same constants of motion (ℓ, q) as a spherical photon orbit, but that does not stay at a fixed value of r. Equivalently, a critical null geodesic is a null geodesic with varying r and for which the quartic polynomial $\mathcal{R}(r)$ [Eq. (12.25)] admits a double root [Eq. (12.47)].

In Schwarzschild spacetime (Chap. 8), a critical null geodesic was simply any null geodesic with varying r that has $b = |\ell| = 3\sqrt{3}\,m$. In the Kerr case with $0 < a < m$, we have instead a 1-parameter family of critical values: the family $(\ell_{\rm c}(r_0), q_{\rm c}(r_0))$ given by Eqs. (12.56)–(12.57), the parameter being r_0. By construction, the polynomial $\mathcal{R}(r)$ of a critical null geodesic $\mathscr{L}$ of parameter[3] r_0 has a double root at $r = r_0$ [cf. Eq. (12.47)]. More precisely, inserting formula (12.56) for ℓ and formula (12.54) for $\tilde{q} := q + (\ell - a)^2$ into expression (12.25a) for $\mathcal{R}(r)$ leads to

$$\mathcal{R}(r) = (r - r_0)^2 \left(r^2 + 2r_0 r - \frac{a^2 q_{\rm c}(r_0)}{r_0^2} \right), \tag{12.84}$$

[3]The word *parameter* is used here for the index among the family of all null critical geodesics and shall not be confused with the *affine parameter* along $\mathscr{L}$.

where $q_{\rm c}(r_0)$ is the function (12.57). This writing of $\mathcal{R}(r)$ clearly exhibits the double root at $r = r_0$. As discussed in Sec. 11.3.5 and 12.2.5, a consequence of the double-root behavior is that any critical null geodesic $\mathscr{L}$ of parameter r_0 has r_0 as an asymptotic r-value, i.e. $r(\lambda) \to r_0$ when $\lambda \to +\infty$ or $\lambda \to -\infty$, where λ is the affine parameter along $\mathscr{L}$.

Let us consider a critical null geodesic $\mathscr{L}$ of parameter r_0 that goes through a point A of Boyer-Lindquist coordinates $(t_A, r_A, \theta_A, \varphi_A)$. By plugging (12.84) into the integrated equation of motion (11.55a), we get, at any point of coordinates (t, r, θ, φ) along $\mathscr{L}$,

$$\int_{r_A}^{r} \frac{\epsilon_r \, \mathrm{d}\bar{r}}{|\bar{r} - r_0| \sqrt{\bar{r}^2 + 2r_0\bar{r} - a^2 q_{\rm c}(r_0)/r_0^2}} = \int_{\theta_A}^{\theta} \frac{\epsilon_\theta \, \mathrm{d}\bar{\theta}}{\sqrt{\tilde{\Theta}(\bar{\theta})}}. \tag{12.85}$$

When $r \to r_0$, it is clear that the integral in the left-hand side diverges logarithmically in terms of $|r - r_0|$. The path integral in the right-hand side must therefore diverge as well, which implies that $\mathscr{L}$ has an endless oscillatory θ-motion as $r \to r_0$. Similarly the integrated equation of motion (11.55d) with expression (12.84) for $\mathcal{R}(r)$ becomes

$$\varphi - \varphi_{\rm em} = a \int_{r_A}^{r} \frac{(2m\bar{r} - a\ell)\, \epsilon_r \, \mathrm{d}\bar{r}}{|\bar{r} - r_0|(\bar{r}^2 - 2m\bar{r} + a^2) \sqrt{\bar{r}^2 + 2r_0\bar{r} - a^2 q_{\rm c}(r_0)/r_0^2}} + \ell \int_{\theta_A}^{\theta} \frac{\epsilon_\theta \, \mathrm{d}\bar{\theta}}{\sin^2\bar{\theta} \sqrt{\tilde{\Theta}(\bar{\theta})}}. \tag{12.86}$$

Again, the first integral diverges when $r \to r_0$. The second one, which is a path integral on θ, diverges as well because the path integral in the right-hand side of Eq. (12.85) diverges. From expression (12.56) for ℓ it can be checked that $2mr_0 - a\ell > 0$ in the allowed range of r_0 (i.e. in the range (12.59), where spherical orbits exist). Given that $\epsilon_r \, \mathrm{d}\bar{r} > 0$ and $\epsilon_\theta \, \mathrm{d}\bar{\theta} > 0$, we conclude that the integral on r and the path integral on θ both tend to $+\infty$ when $r \to r_0$. If $\ell \geq 0$, we get then immediately $\varphi \to +\infty$ when $r \to r_0$. If $\ell < 0$, the two diverging terms in the right-hand side of Eq. (12.86) have opposite signs. Disregarding some unexpected subtle compensation, one term (in practice the second one) dominates over the other one, so that we conclude that, whatever the sign of ℓ,

$$\varphi \to \pm\infty \quad \text{when} \quad r \to r_0. \tag{12.87}$$

We may summarize the above results as follows.

A critical null geodesic $\mathscr{L}$ of parameter r_0 has r_0 as an asymptotic r-value, either in the future ($r(\lambda) \to r_0$ when $\lambda \to +\infty$) or in the past ($r(\lambda) \to r_0$ when $\lambda \to -\infty$), λ being the (future-directed) affine parameter of $\mathscr{L}$. In particular, $\mathscr{L}$ never crosses the sphere $r = r_0$. Moreover $\mathscr{L}$ is winding endlessly on the sphere $r = r_0$ when either $\lambda \to +\infty$ or $\lambda \to -\infty$, mimicking there the behavior of a spherical photon orbit.

Example 9: Figure 12.18 shows a critical null geodesic emitted from the equatorial plane at a large distance from the black hole (the emission point is not shown on the figure that is truncated at $r \sim 10\,m$). We note the winding around the sphere $r = r_0 = 2.2\,m$, in the same fashion as a spherical photon orbit (compare with Figs. 12.8 and 12.9).

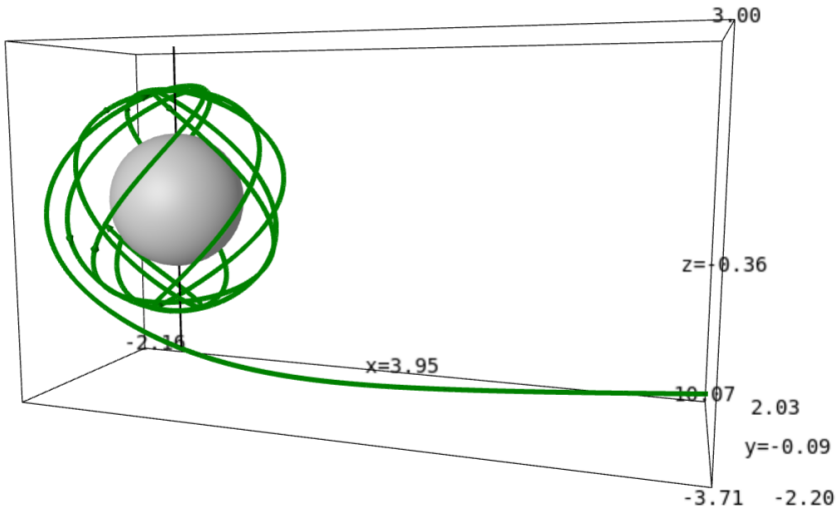


Figure 12.18: Critical null geodesic of parameter $r_0 = 2.2\,m$ in a Kerr spacetime with $a = 0.95\,m$, depicted in terms of the Cartesian Boyer-Lindquist coordinates (x, y, z) [Eq. (11.64)]. This geodesic has $\ell = 0.863\,m$, $q = 18.042\,m^2$ and is emitted at $\lambda = 0$ from the point of Boyer-Lindquist coordinates $(r, \theta, \varphi) = (40\,m, \pi/2, 0) \iff (x, y, z) = (40\,m, 0, 0)$, towards the black hole ($\epsilon_r = -1$). The drawing is interrupted at $\lambda = 80\,m/E$. The grey sphere is the black hole event horizon at $r = r_+ = 1.312\,m$. *[Figure generated by the notebook D.4.13]*

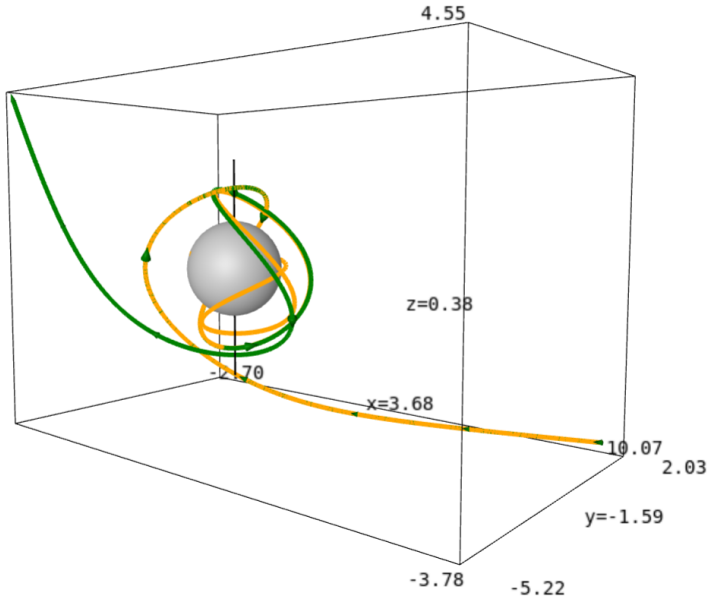


Figure 12.19: Two null geodesics with constants of motion (ℓ, q) close to those of a critical null geodesic, $(\ell_{\rm c}(r_0), q_{\rm c}(r_0))$ [Eqs. (12.56)–(12.57)]. Here, as in Fig. 12.18, $r_0 = 2.2\,m$ and $a = 0.95\,m$. The two geodesics depart from the same point at $(x, y, z) = (40\,m, 0, 0)$ (outside the scope of the figure) as the critical null geodesic shown in Fig. 12.18. Both geodesics have $q = q_{\rm c}(r_0)$, but the green one has $\ell = 1.0001\,\ell_{\rm c}(r_0)$ while the orange one has $\ell = 0.9999\,\ell_{\rm c}(r_0)$. *[Figure generated by the notebook D.4.13]*

12.4.2 Critical curve and black hole shadow

In this section, we focus on null geodesics emitted in the black hole exterior, i.e. in region $\mathscr{M}_{\rm I}$, having in mind the formation of images on the screen of a distant observer. As in the Schwarzschild case studied in Chap. 8, the family of critical null geodesics separates the null geodesics emitted far from the black hole between those that can escape to infinity and those that fall into the black hole. More precisely, according to the lemma established in Sec. 12.2.5, this family separates null geodesics whose quartic polynomial $\mathscr{R}$ has no root in $\mathscr{M}_{\rm I}$ from those for which $\mathscr{R}$ has a double root in $\mathscr{M}_{\rm I}$. It follows that the family of critical null geodesics separates null geodesics that have a r-turning point in $\mathscr{M}_{\rm I}$ from those that do not have any, given that a null geodesic has at most one r-turning point in $\mathscr{M}_{\rm I}$ (cf. Sec. 12.2.5). This is illustrated by the following example.

Example 10: Figure 12.19 depicts two null geodesics initially very close to a critical one. They are emitted at the same point as the critical geodesic considered in Example 9 and with parameters close to critical: the same reduced Carter constant q and values of the reduced angular momentum ℓ almost equal to the critical one, $\ell_{\rm c}(r_0)$, up to a relative difference of 10^{-4}. The geodesic $\mathscr{L}_1$ (green curve) has $\ell = 1.0001\,\ell_{\rm c}(r_0)$; it performs a few turns onto (actually very close to) the sphere $\mathscr{S}_0$ of coordinate radius $r_0 = 2.2\,m$ and eventually depart to infinity; this means that $\mathscr{L}_1$ has a r-turning point, somewhere close to $\mathscr{S}_0$. The geodesic $\mathscr{L}_2$ (orange curve) has $\ell = 0.9999\,\ell_{\rm c}(r_0)$; it is graphically indistinguishable from $\mathscr{L}_1$ until a full turn onto $\mathscr{S}_0$ (this is best seen in the interactive 3D view in the online notebook D.4.13), it then clearly departs from it and eventually terminates into the black hole. Hence $\mathscr{L}_2$ has no r-turning point.

Let us consider an asymptotic inertial observer $\mathscr{O}$ (cf. Sec. 10.7.5), i.e. a static observer located at Boyer-Lindquist coordinates $(t, r_{\mathscr{O}}, \theta_{\mathscr{O}}, \varphi_{\mathscr{O}})$, where $(r_{\mathscr{O}}, \theta_{\mathscr{O}}, \varphi_{\mathscr{O}})$ are constant and $r_{\mathscr{O}} \gg m$. As in the Schwarzschild case (Sec. 8.5.4), the concept of black hole shadow is defined by considering an emitting large sphere $\mathscr{S}$ of constant Boyer-Lindquist coordinate $r_{\mathscr{S}}$ that encompasses both the black hole and observer $\mathscr{O}$; this means that $r_{\mathscr{S}} > r_{\mathscr{O}}$. In a more astrophysical setting, one could image $\mathscr{S}$ as being made of many far-away light sources. We assume that $\mathscr{O}$ is equipped with a screen in the direction of the black hole, in the same set up as described in Sec. 12.2.3 (cf. Fig. 12.4). Given the relative position of $\mathscr{O}$, the black hole and the emitting sphere $\mathscr{S}$, any photon emitted from $\mathscr{S}$ that reaches $\mathscr{O}$'s screen has necessarily a point of minimal approach to the black hole, i.e. the associated null geodesic has necessarily a r-turning point between $\mathscr{S}$ and $\mathscr{O}$. Reciprocally, any null geodesic that impacts $\mathscr{O}$'s screen and has no r-turning point in its past cannot have emerged from $\mathscr{S}$ and therefore results in a black dot on the screen. According to the above discussion, the boundary of the black area on $\mathscr{O}$'s screen is made by the impact points of critical null geodesics and is called the ***critical curve*** [129, 130]; we shall denote it by $\mathscr{C}$. The black area itself is called the ***black hole shadow***. It is called so essentially because if the black hole were not present, it would not exist and $\mathscr{O}$'s screen would be uniformly bright.

Remark 1: The concept of black hole shadow is rather academic, since it requires the sources of light to be far from the black hole and to surround it, as well as the observer, from any direction. On the contrary, astrophysical black holes are illuminated by close sources (e.g. an accretion disk) and we shall see in Sec. 12.5 that the black area on astronomical images has little

resemblance with the shadow defined above. On the other hand, the critical curve has a true observational significance, as we shall discuss in Sec. 12.5, and is definitely worth to study.

In what follows, we shall assume that the observer $\mathscr{O}$ in not located on the black hole rotation axis, i.e. $\theta_{\mathscr{O}} \notin \{0, \pi\}$, leaving the special case $\theta_{\mathscr{O}} = 0$ or π to Sec. 12.4.3. Then $\sin\theta_{\mathscr{O}} \neq 0$ and we have seen in Sec. 12.2.3 that the impact point of a null geodesic $\mathscr{L}$ on $\mathscr{O}$'s screen, measured by the screen angular coordinates (α, β), is related to the reduced angular momentum ℓ and reduced Carter constant q of $\mathscr{L}$ by formulas (12.32). Given the above definition, the critical curve $\mathscr{C}$ is obtained by using for (ℓ, q) in (12.32) the critical values $(\ell_{\rm c}(r_0), q_{\rm c}(r_0))$ given by Eqs. (12.56)–(12.57):

$$\alpha = -\frac{\ell_{\rm c}(r_0)}{r_{\mathscr{O}} \sin\theta_{\mathscr{O}}} \tag{12.88a}$$

$$\beta = \frac{\epsilon_\theta}{r_{\mathscr{O}}} \sqrt{q_{\rm c}(r_0) + \cos^2\theta_{\mathscr{O}} \left(a^2 - \frac{\ell_{\rm c}(r_0)^2}{\sin^2\theta_{\mathscr{O}}} \right)}. \tag{12.88b}$$

This provides the equation of $\mathscr{C}$ in parametric form, the parameter being r_0 — the radius of the spherical orbits that have the same value of (ℓ, q) as the critical null geodesic that impacts $\mathscr{O}$'s screen at the point (α, β). Let us recall that (α, β) are angular coordinates and are therefore dimensionless (cf. Sec. 12.2.3). The range of r_0 for spherical orbits in the black hole exterior is $r_{\rm ph}^+ \leq r_0 \leq r_{\rm ph}^-$ [Eq. (12.59)], with $r_{\rm ph}^+$ and $r_{\rm ph}^-$ given by Eq. (12.62). The corresponding range of $\ell = \ell_{\rm c}(r_0)$ is then $\ell_{\rm c}(r_{\rm ph}^-) \leq \ell \leq \ell_{\rm c}(r_{\rm ph}^+)$ (cf. the red curve in Fig. 12.6) and the corresponding range of $q = q_{\rm c}(r_0)$ is $0 \leq q \leq 27\,m^2$ [cf. Eq. (12.83)]. However, not all these values of (ℓ, q) are allowed, since in order for the corresponding geodesic to reach $\mathscr{O}$, they have to obey the constraint (12.29):

$$\left(q_{\rm c}(r_0) + a^2 \cos^2\theta_{\mathscr{O}}\right) \sin^2\theta_{\mathscr{O}} - \ell_{\rm c}(r_0)^2 \cos^2\theta_{\mathscr{O}} \geq 0. \tag{12.89}$$

Except for $\theta_{\mathscr{O}} = \pi/2$ (observer in the equatorial plane), this constraint limits the range of r_0 to a subinterval $[r_0^{\rm min}, r_0^{\rm max}]$ of $[r_{\rm ph}^+, r_{\rm ph}^-]$. We note that the radius $r_{\rm ph}^{\rm pol}$ of polar spherical photon orbits, as given by Eq. (12.67a), lies necessarily in that subinterval; indeed by definition $\ell_{\rm c}(r_{\rm ph}^{\rm pol}) = 0$, which ensures that the constraint (12.89) is fulfilled for $r_0 = r_{\rm ph}^{\rm pol}$, whatever the value of $\theta_{\mathscr{O}}$.

The critical curve $\mathscr{C}$ in the observer's screen is obtained by first solving Eq. (12.89) with equality in the sign $\geq$ to determine $r_0^{\rm min}$ and $r_0^{\rm max}$. Then, the upper half of $\mathscr{C}$ is computed by means of formulas (12.88) with $\epsilon_\theta = +1$ and r_0 ranging in $[r_0^{\rm min}, r_0^{\rm max}]$. The lower half is obtained similarly, but with $\epsilon_\theta = -1$ in Eq. (12.88b), so that it is the symmetric of the upper half with respect to the α-axis.

The result of the computation is shown in Fig. 12.20 for a Kerr black hole with $a = 0.95\,m$ and various inclination angles of observer $\mathscr{O}$. For $\theta_{\mathscr{O}} = \pi/2$ the parameter r_0 of the critical null geodesics forming $\mathscr{C}$ spans the full interval $[r_{\rm ph}^+, r_{\rm ph}^-] \simeq [1.386\,m,\ 3.955\,m]$. For $\theta_{\mathscr{O}} = \pi/6$, the range is restricted to $[r_0^{\rm min}, r_0^{\rm max}] \simeq [1.768\,m,\ 3.237\,m]$, while for $\theta_{\mathscr{O}} = 0$ (to be discussed in Sec. 12.4.3), r_0 can take only one value: $r_0 = r_{\rm ph}^{\rm pol} \simeq 2.493\,m$. At a fixed value of a/m, the black hole shadow depends on m and $r_{\mathscr{O}}$ only through the dimensionless ratio $m/r_{\mathscr{O}}$, which sets the global scale of the shadow. Accordingly, in Fig. 12.20, the

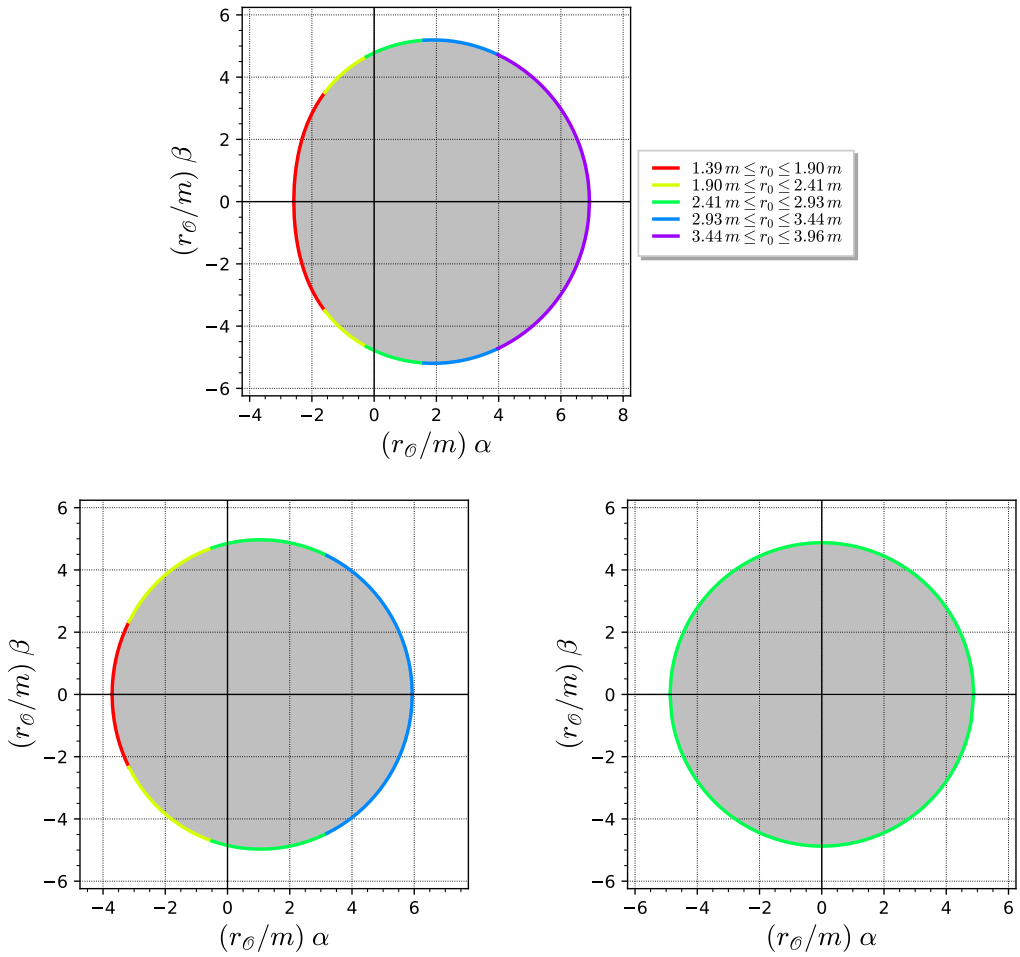


Figure 12.20: Shadow of a Kerr black hole of mass m and spin parameter $a = 0.95\,m$ on the screen of an asymptotic inertial observer $\mathscr{O}$ located at $r = r_\mathscr{O}$ and $\theta = \theta_\mathscr{O}$, for three values of $\theta_\mathscr{O}$: $\theta_\mathscr{O} = \pi/2$ ($\mathscr{O}$ in the equatorial plane; upper panel), $\theta_\mathscr{O} = \pi/6$ (lower left) and $\theta_\mathscr{O} = 0$ ($\mathscr{O}$ on the rotation axis; lower right). The screen is spanned by the angular coordinates (α, β), rescaled by the inverse of the factor $m/r_\mathscr{O}$, whose values for some astrophysical black holes can be found in Table 12.1. The shadow is bounded by the critical curve $\mathscr{C}$, which is depicted with colors indicating the range of the parameter r_0 of the critical null geodesics forming it. *[Figure generated by the notebook D.4.14]*

	Sgr A*	M87*	M31*	Cyg X-1
m $[M_\odot]$	$4.1\ 10^6$	$6.2\ 10^9$	$1.5\ 10^8$	15
$r_{\mathscr{O}}$ [kpc]	8.12	$1.67\ 10^4$	$7.6\ 10^2$	1.86
$m/r_{\mathscr{O}}$	$2.4\ 10^{-11}$	$1.8\ 10^{-11}$	$9.4\ 10^{-12}$	$3.9\ 10^{-16}$
$m/r_{\mathscr{O}}$ [μas]	5.0	3.7	1.9	$8.0\ 10^{-5}$

Table 12.1: Scale factor $m/r_{\mathscr{O}}$ for various astrophysical black holes: the supermassive black hole at the center of our galaxy, Sagittarius A* (data taken from Table A.1 of Ref. [3]), the supermassive black hole M87* in the nucleus of the galaxy Messier 87 (data from Refs. [114, 5]; see also Appendix I and Table 9 of Ref. [7]), the supermassive black hole M31* in the nucleus of the Andromeda Galaxy (Messier 31) (data from Ref. [22]) and the stellar black hole Cygnus X-1 (data from Refs. [201, 218, 121]). The last line gives $m/r_{\mathscr{O}}$ in microarcseconds (1 μas $= 4.848\ 10^{-12}$ rad).

screen angular coordinates (α, β) have been rescaled by $(m/r_{\mathscr{O}})^{-1}$. Values of $m/r_{\mathscr{O}}$ for some astrophysical black holes are provided in Table 12.1; the observer's radial coordinate $r_{\mathscr{O}}$ is then nothing but the distance of the black hole to the Earth. These values of $m/r_{\mathscr{O}}$ are tiny, being at most $2.4\ 10^{-11} = 5.0$ μas (microarcseconds; 1 μas $= 4.848\ 10^{-12}$ rad) for the (known) black hole of largest apparent size as seen from Earth, Sgr A*. Given that the diameter of the shadow is ~ 10 in the scale $m/r_{\mathscr{O}}$ used in Fig. 12.20, this means that the angular size of Sgr A* shadow as seen from Earth is only ~ 50 μas. Such a small value[4] has entered recently in the realm of observational astronomy, with the advent of the Event Horizon Telescope [5, 263], whose angular resolution is of order 20 μas. We also note from Table 12.1 that the size of the shadow of a stellar-mass black hole in our galaxy, such as Cygnus X-1, is even more tiny, by a factor 10^{-5}, which makes the shadow of this kind of black holes out of reach by the current technology.

The main feature that appears in Fig. 12.20 is that for $\theta_{\mathscr{O}} = \pi/6$ and even more for $\theta_{\mathscr{O}} = \pi/2$, the shadow is shifted to the right and its left edge is flattened, as compared to the shadow for $\theta_{\mathscr{O}} = 0$ (or π). This can be understood by noticing that the critical null geodesics forming the left edge arise form regions of smaller r_0 than those on the right edge (cf. the color code). For instance, the left critical null geodesic at $\beta = 0$ has $\alpha < 0$ and $\ell > 0$ (cf. the minus sign in Eq. (12.88a)) and arises from the prograde outer circular photon orbit lying at $r_0 = r_{\rm ph}^+$ in the equatorial plane (cf. Sec. 12.3.3), while the right one at $\beta = 0$ has $\alpha > 0$ and $\ell < 0$ and arises from the retrograde outer circular photon orbit at $r_0 = r_{\rm ph}^-$. That prograde (resp. retrograde) geodesics impact the screen on the left (resp. right) side is easily recovered by remembering that the β-axis ($\alpha = 0$) coincides with the orthogonal projection of the black hole's spin onto the observer's screen, with the spin being oriented upward.

[4] For comparison, the angular resolution of the Hubble Space Telescope is $\sim 0.1'' = 10^5$ μas!

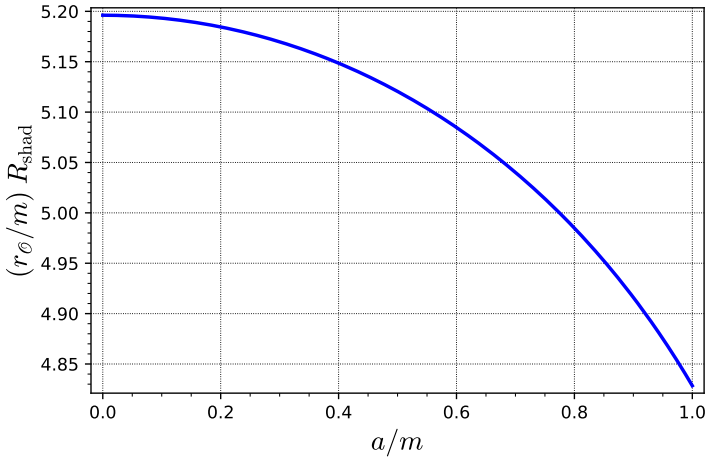


Figure 12.21: Angular radius $R_{\rm shad}$ of the (circular) critical curve of a Kerr black hole as seen by an asymptotic inertial observer located at $r = r_{\mathscr{O}}$ on the black hole's rotation axis, as a function of the Kerr spin parameter a [Eq. (12.90)]. $R_{\rm shad}$ is given in units of $m/r_{\mathscr{O}}$ (cf. Table 12.1). *[Figure generated by the notebook D.4.14]*

12.4.3 Shadow for an observer on the rotation axis

As stressed in Sec. 12.2.3, if $\theta_{\mathscr{O}} \in \{0, \pi\}$, the screen angular coordinates (α, β) can no longer be defined from the orthonormal frame $(\boldsymbol{e}_{(\theta)}, \boldsymbol{e}_{(\varphi)})$ associated with the Boyer-Lindquist coordinates (θ, φ). In this case, we pick an arbitrary orthonormal frame $(\boldsymbol{e}_{(\alpha)}, \boldsymbol{e}_{(\beta)})$ in the screen plane to define (α, β). Due to the axisymmetry of spacetime, when $\mathscr{O}$ lies on the rotation axis, the black hole shadow is symmetric by any rotation around the screen's center. It is therefore necessarily a disk. Its boundary, the critical curve $\mathscr{C}$, is then a circle. A critical null geodesic impacting the screen on $\mathscr{C}$ has necessarily $\ell = 0$ [Eq. (12.36)]. It follows that the parameter r_0 can take only a single value: $r_0 = r_{\rm ph}^{\rm pol}$, which is given by Eq. (12.67a). The radius of $\mathscr{C}$ is then given by Eq. (12.37), with $q = q_{\rm c}(r_{\rm ph}^{\rm pol})$:

$$\alpha^2 + \beta^2 = \frac{1}{r_{\mathscr{O}}^2} \left(q_{\rm c}(r_{\rm ph}^{\rm pol}) + a^2 \right) .$$

Given expression (12.57) of the function $q_{\rm c}$ and the fact that $r_{\rm ph}^{\rm pol}$ obeys $(r_{\rm ph}^{\rm pol})^4 (r_{\rm ph}^{\rm pol} - 3m)^2 = a^4 (r_{\rm ph}^{\rm pol} + m)^2$, as a solution of $\ell_{\rm c}(r_0) = 0$, we get

$$\boxed{R_{\rm shad} := \sqrt{\alpha^2 + \beta^2} = 2 \frac{m}{r_{\mathscr{O}}} \sqrt{\frac{r_{\rm ph}^{\rm pol}}{m}} \frac{\sqrt{(r_{\rm ph}^{\rm pol})^2 - a^2}}{r_{\rm ph}^{\rm pol} - m}} \qquad (\theta_{\mathscr{O}} = 0 \text{ or } \theta_{\mathscr{O}} = \pi) \qquad (12.90)$$

where $r_{\rm ph}^{\rm pol}$ is the function (12.67a) of (m, a) and the scale factor $m/r_{\mathscr{O}}$ is given for some astrophysical black holes in Table 12.1.

Example 11: An example of shadow seen from the rotation axis is shown in the lower right panel of Fig. 12.20. The Kerr spin parameter is $a = 0.95\, m$, for which formula (12.67a) yields

$r_{\rm ph}^{\rm pol} \simeq 2.493\,m$, so that Eq. (12.90) results in $R_{\rm shad} \simeq 4.875\,m/r_{\mathscr{O}}$. We note that the critical curve $\mathscr{C}$ is depicted in the same color (green) as the part of the critical curve in the other panels that crosses the β-axis, which is expected since the critical null geodesics that reach the screen along that axis have $\ell = 0$ and thus the same parameter $r_0 = r_{\rm ph}^{\rm pol}$ as all the critical null geodesics forming $\mathscr{C}$ for $\theta_{\mathscr{O}} = 0$ or π.

The shadow radius $R_{\rm shad}$ is a decreasing function of a, which is depicted in Fig. 12.21. Note that its variation range is pretty limited, since the limits $\lim_{a\to 0} r_{\rm ph}^{\rm pol} = 3m$ and $\lim_{a\to m} r_{\rm ph}^{\rm pol} = (\sqrt{2}+1)m$ (cf. Sec. 12.3.2) yield respectively

$$\lim_{a\to 0} R_{\rm shad} = 3\sqrt{3}\frac{m}{r_{\mathscr{O}}} \simeq 5.196\,\frac{m}{r_{\mathscr{O}}} \quad \text{and} \quad \lim_{a\to m} R_{\rm shad} = 2(\sqrt{2}+1)\frac{m}{r_{\mathscr{O}}} \simeq 4.828\,\frac{m}{r_{\mathscr{O}}}. \tag{12.91}$$

The value for $a = m$ is thus only 7% lower than the value for $a = 0$.

Remark 2: The limit $a \to 0$ is in agreement with the result for the shadow of a Schwarzschild black hole obtained in Sec. 8.5.4.

12.4.4 Shadow of an extremal Kerr black hole

The case $a = m$ corresponds to the extremal Kerr black hole, which will be studied in Chap. 13. However, we shall discuss its shadow and critical curve here, by taking some appropriate limits. For $a = m$, formulas (12.56)-(12.57) for $\ell_{\rm c}(r_0)$ and $q_{\rm c}(r_0)$ simplify significantly:

$$\ell_{\rm c}(r_0) = -\frac{r_0^2}{m} + 2r_0 + m \qquad (a = m) \tag{12.92}$$

$$q_{\rm c}(r_0) = \frac{r_0^3}{m^2}(4m - r_0) \qquad (a = m). \tag{12.93}$$

Substituting these values into the system (12.88) leads to the screen angular coordinates determining the critical curve $\mathscr{C}$ of an extremal Kerr black hole:

$$\alpha = \frac{(r_0 - m)^2 - 2m^2}{m r_{\mathscr{O}} \sin\theta_{\mathscr{O}}} \tag{12.94a}$$

$$\beta = \frac{\epsilon_\theta}{m r_{\mathscr{O}} \sin\theta_{\mathscr{O}}} \sqrt{r_0^3(4m - r_0) - m^2\cos^2\theta_{\mathscr{O}}\,[2r_0(r_0+2m) + m^2\cos^2\theta_{\mathscr{O}}]}. \tag{12.94b}$$

Besides, Eq. (12.62) yields

$$r_{\rm ph}^+ = m \qquad \text{and} \qquad r_{\rm ph}^- = 4m \qquad (a = m). \tag{12.95}$$

The range of r_0 is determined by $r_{\rm ph}^+ \le r_0 \le r_{\rm ph}^-$ and $\tilde\Theta(\theta_{\mathscr{O}}) \ge 0$. The last condition is equivalent to demanding that the quantity under the square root in expression (12.94b) for β is non-negative. This leads[5] to some interval $[r_{\rm min}, r_{\rm max}] \subset [r_{\rm ph}^+, r_{\rm ph}^-]$. As in the case

[5] The values of $r_{\rm min}$ and $r_{\rm max}$ can be computed exactly as $r_{\rm min} = \max(r_1, m)$ and $r_{\rm max} = r_2$, where r_1 and r_2 are two roots of the quartic polynomial in r_0 that appears under the square root in Eq. (12.94b). However, doing so would lead to complicated expressions, while a computation by numerical root finding, as in the notebook D.4.14, is sufficient in practice.

$a < m$ treated above, we have $r_{\rm max} \leq r_{\rm ph}^-$ with $r_{\rm max} = r_{\rm ph}^- \iff \theta_{\mathscr{O}} = \pi/2$. However, $r_{\rm min}$ behaves differently. As we going to see, $r_{\rm min} = r_{\rm ph}^+ (= m)$ for a finite-width interval of values of $\theta_{\mathscr{O}}$ around $\pi/2$ and not only for $\pi/2$. To see this, let us start by noticing that expression (12.93) results in $q_{\rm c}(r_{\rm ph}^-) = 0$, while $q_{\rm c}(r_{\rm ph}^+) = 3m^2 \neq 0$. This last result seems to contradict the fact that $r_{\rm ph}^+$ has been obtained in Sec. 12.3.1 by searching for the zeros of $q_{\rm c}$. However, the generic expression (12.57) for $q_{\rm c}$ has a factor $(r_0 - m)^2$ in its denominator, so that taking the limits $a \to m$ and $r_0 \to r_{\rm ph}^+ = m$ yields the indeterminate form "0/0". As a consequence of $q_{\rm c}(r_{\rm ph}^+) \neq 0$, one has $\beta \neq 0$ for $r_0 = r_{\rm ph}^+ = m$ and $|\theta_{\mathscr{O}} - \pi/2|$ sufficiently small. This implies that $r_{\rm min} = m$ for a finite range of values of $\theta_{\mathscr{O}}$ around $\pi/2$. More precisely, according to (12.94), the screen coordinates for $r_0 = r_{\rm ph}^+ = m$ are

$$\alpha|_{r_0=m} = -2\frac{m}{r_{\mathscr{O}} \sin\theta_{\mathscr{O}}} \tag{12.96a}$$

$$\beta|_{r_0=m} = \frac{\epsilon_\theta m}{r_{\mathscr{O}} \sin\theta_{\mathscr{O}}} \sqrt{3 - 6\cos^2\theta_{\mathscr{O}} - \cos^4\theta_{\mathscr{O}}}. \tag{12.96b}$$

In particular, for $\theta_{\mathscr{O}} = \pi/2$ (observer $\mathscr{O}$ in the equatorial plane), we get

$$\alpha|_{r_0=m} = -2\,\frac{m}{r_{\mathscr{O}}} \qquad \text{and} \qquad \beta|_{r_0=m} = \epsilon_\theta \sqrt{3}\,\frac{m}{r_{\mathscr{O}}} \qquad \left(\theta_{\mathscr{O}} = \frac{\pi}{2}\right). \tag{12.97}$$

Equation (12.96b) shows that, for $\epsilon_\theta = +1$, $\beta|_{r_0=m} > 0 \iff \theta_{\rm crit} < \theta_{\mathscr{O}} < \pi - \theta_{\rm crit}$, where $\theta_{\rm crit}$ is such that $\cos^2\theta_{\rm crit}$ is the positive root of the quadratic polynomial $-x^2 - 6x + 3 = 0$. We get $\cos^2\theta_{\rm crit} = 2\sqrt{3} - 3$, from which $\sin^2\theta_{\rm crit} = 4 - 2\sqrt{3} = (\sqrt{3} - 1)^2$, hence

$$\theta_{\rm crit} = \arcsin(\sqrt{3} - 1) \simeq 0.8213\,\text{rad} \simeq 47.06^\circ. \tag{12.98}$$

For $\theta_{\mathscr{O}} = \theta_{\rm crit}$ or $\pi - \theta_{\rm crit}$, we have exactly $\beta|_{r_0=m} = 0$, while for $\theta_{\mathscr{O}} < \theta_{\rm crit}$ or $\theta_{\mathscr{O}} > \pi - \theta_{\rm crit}$, $\beta|_{r_0=m}$ is imaginary. This means that for $\theta_{\rm crit} \leq \theta_{\mathscr{O}} \leq \pi - \theta_{\rm crit}$, the range of the parameter r_0 is $[m, r_{\rm max}]$, while for for $\theta_{\mathscr{O}} < \theta_{\rm crit}$ or $\theta_{\mathscr{O}} > \pi - \theta_{\rm crit}$, the range is $[r_{\rm min}, r_{\rm max}]$ with $r_{\rm min} > m$, as for the case $\theta_{\mathscr{O}} \neq \pi/2$ of the shadows with $a < m$ discussed in Sec. 12.4.2. In this last case, the critical curve is a smooth curve with $\beta|_{r_0=r_{\rm min}} = 0$ (in addition to $\beta|_{r_0=r_{\rm max}} = 0$, which always holds).

Let us focus on the first case, i.e. $\theta_{\rm crit} \leq \theta_{\mathscr{O}} \leq \pi - \theta_{\rm crit}$. The parametric curve defined by the system (12.94) terminates at $r_0 = m$ on the two points given by Eq. (12.96), or Eq. (12.97) in the particular case $\theta_{\mathscr{O}} = \pi/2$ (one point for $\epsilon_\theta = +1$, and the other one for $\epsilon_\theta = -1$). For $\theta_{\rm crit} < \theta_{\mathscr{O}} < \pi - \theta_{\rm crit}$, one has $\beta \neq 0$ at these two points, so that the curve is not closed, as one can see on Fig. 12.22 (drawn for $\theta_{\mathscr{O}} = \pi/2$). One may be puzzled by this feature: the shadow boundary has to be closed! We are thus missing some critical null geodesics to complete the boundary. It is easy to find the missing ones as soon as we remember that in the special case $a = m$, we had found an extra family of spherical photon orbits in Sec. 12.3.1: those given by Eq. (12.50), namely the spherical photon orbits at $r_0 = m$ that have $\ell = 2m$. Obviously, this family cannot be parameterized by r_0; on the other hand, the reduced Carter constant q is a valid parameter. Since q is not constrained by Eq. (12.50), it can take all values in the range $[0, +\infty)$. Note that $q < 0$ is not permitted here since $\ell = 2m > a$ [cf. property (12.38)]. However, not any

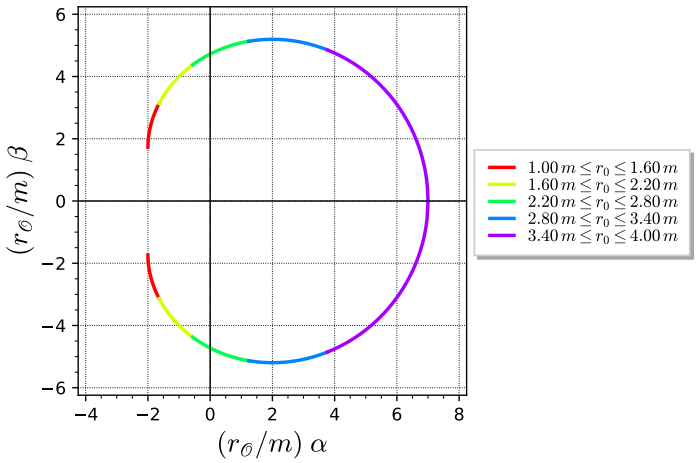


Figure 12.22: Part of the critical curve of an extremal Kerr black hole ($a = m$) in the screen of an asymptotic inertial observer $\mathscr{O}$ in the equatorial plane ($\theta_{\mathscr{O}} = \pi/2$), as defined by the parametric equations (12.94), with $r_0 \in [m, 4m]$. The two endpoints are achieved at $r_0 = m$ and are given by Eq. (12.97): $(\bar\alpha, \bar\beta) = (-2, \pm\sqrt{3}) \simeq (-2, \pm 1.732)$, where $\bar\alpha := \alpha r_{\mathscr{O}}/m$ and $\bar\beta := \beta r_{\mathscr{O}}/m$. *[Figure generated by the notebook D.4.14]*

value of q corresponds to a spherical photon orbit associated with a critical null geodesic that reaches the asymptotic inertial observer $\mathscr{O}$. Indeed, q must give birth to a radial polynomial $\mathcal{R}(r)$ positive for all $r > m$, so that the radial motion is possible between the spherical orbit at $r_0 = m$ and the observer [condition (12.46)]. Specializing expression (12.25) for $\mathcal{R}(r)$ to $a = m$ and $\ell = 2m$, we get

$$\mathcal{R}(r) = (r - m)^2(r^2 + 2mr - q). \tag{12.99}$$

$r = m$ appears as a double root of $\mathcal{R}$, which confirms that we are dealing with spherical photon orbits at $r_0 = m$. There are two other roots, which depend on q:

$$r_q^{\pm} = \pm\sqrt{m^2 + q} - m. \tag{12.100}$$

Since $q \ge 0$, we have $r_q^+ \ge 0$ and $r_q^- \le -2m$. It is then clear (cf. Fig. 12.23) that

$$(\forall r \in (m, +\infty), \quad \mathcal{R}(r) > 0) \iff r_q^+ \le m. \tag{12.101}$$

The critical value is $r_q^+ = m$, which correspond to $q = 3m^2$ (red curve in Fig. 12.23). $r = m$ is then a triple root of $\mathcal{R}$: $q = 3m^2 \implies \mathcal{R}(r) = (r - m)^3(r + 3m)$. Since r_q^+ is an increasing function of q, we conclude that

For $a = m$, a critical null geodesic of constants of motion (ℓ, q) can reach the asymptotic region $r \gg m$ from the vicinity of a spherical photon orbit at $r_0 = m$ iff

$$\ell = 2m \qquad \text{and} \qquad 0 \le q \le 3m^2. \tag{12.102}$$

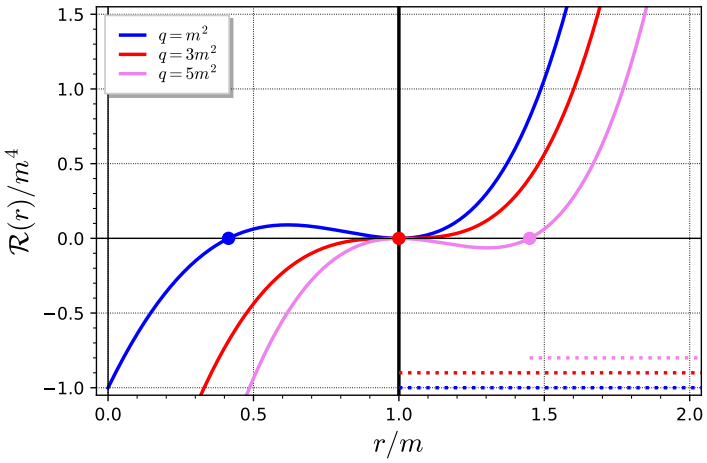


Figure 12.23: Radial polynomial $\mathcal{R}(r)$ for $a = m$, $\ell = 2m$ and three values of q. All the polynomials have a double root at $r = m$ and the dots mark the root r_q^+ [Eq. (12.100)], the fourth root $r_q^- \leq -2m$ being out of the figure's scope. The dotted horizontal lines indicate the range of r where a null geodesic motion to the asymptotic inertial observer is possible. *[Figure generated by the notebook D.4.14]*

Furthermore, such a critical null geodesic can reach the asymptotic inertial observer $\mathscr{O}$, who is located at $\theta = \theta_{\mathscr{O}}$, only if the $\tilde{\Theta}$ function associated with (ℓ, q) obeys $\tilde{\Theta}(\theta_{\mathscr{O}}) \geq 0$. Setting $a = m$ and $\ell = 2m$ into expression (12.26) for $\tilde{\Theta}$, we get the criterion

$$q \geq \frac{3 + \cos^2\theta_{\mathscr{O}}}{\tan^2\theta_{\mathscr{O}}}\, m^2. \tag{12.103}$$

The screen angular coordinates corresponding to the critical null geodesic are obtained by plugging the above values of ℓ and q, as well as $a = m$, into formulas (12.32):

$$\alpha = -\frac{2m}{r_{\mathscr{O}} \sin\theta_{\mathscr{O}}} \tag{12.104a}$$

$$\beta = \frac{\epsilon_\theta m}{r_{\mathscr{O}} \sin\theta_{\mathscr{O}}} \sqrt{\frac{q}{m^2}\sin^2\theta_{\mathscr{O}} - \cos^2\theta_{\mathscr{O}}\,(3 + \cos^2\theta_{\mathscr{O}})}, \quad \frac{3 + \cos^2\theta_{\mathscr{O}}}{\tan^2\theta_{\mathscr{O}}} \leq \frac{q}{m^2} \leq 3. \tag{12.104b}$$

The system (12.104) defines a curve parameterized by q, with the range of q obtained by combining (12.102) and (12.103). For $\mathscr{O}$ in the equatorial plane ($\theta_{\mathscr{O}} = \pi/2$), this parametric equation simplifies to

$$\alpha = -2\frac{m}{r_{\mathscr{O}}} \qquad \text{and} \qquad \beta = \epsilon_\theta \frac{m}{r_{\mathscr{O}}}\sqrt{\frac{q}{m^2}}, \quad 0 \leq \frac{q}{m^2} \leq 3 \qquad \left(\theta_{\mathscr{O}} = \frac{\pi}{2}\right). \tag{12.105}$$

In all cases, the curve (12.104) is actually a segment of a vertical straight line on $\mathscr{O}$'s screen, since it has $\alpha = \text{const}$. Following Ref. [133], we shall call this segment the ***NHEK line***, *NHEK* standing for *Near Horizon Extremal Kerr*, given that spherical photon orbits at $r_0 = m$ are close[6] to the event horizon of the extremal Kerr black hole. The key feature

[6]Despite they have the same radial coordinate as the horizon ($r_+ = m$ for $a = m$), the spherical photon orbits at $r_0 = m$ are not located on the horizon, for r fails to be a well-behaved coordinate at $r = m$ in the extremal Kerr geometry, as we shall discuss in Chap. 13.

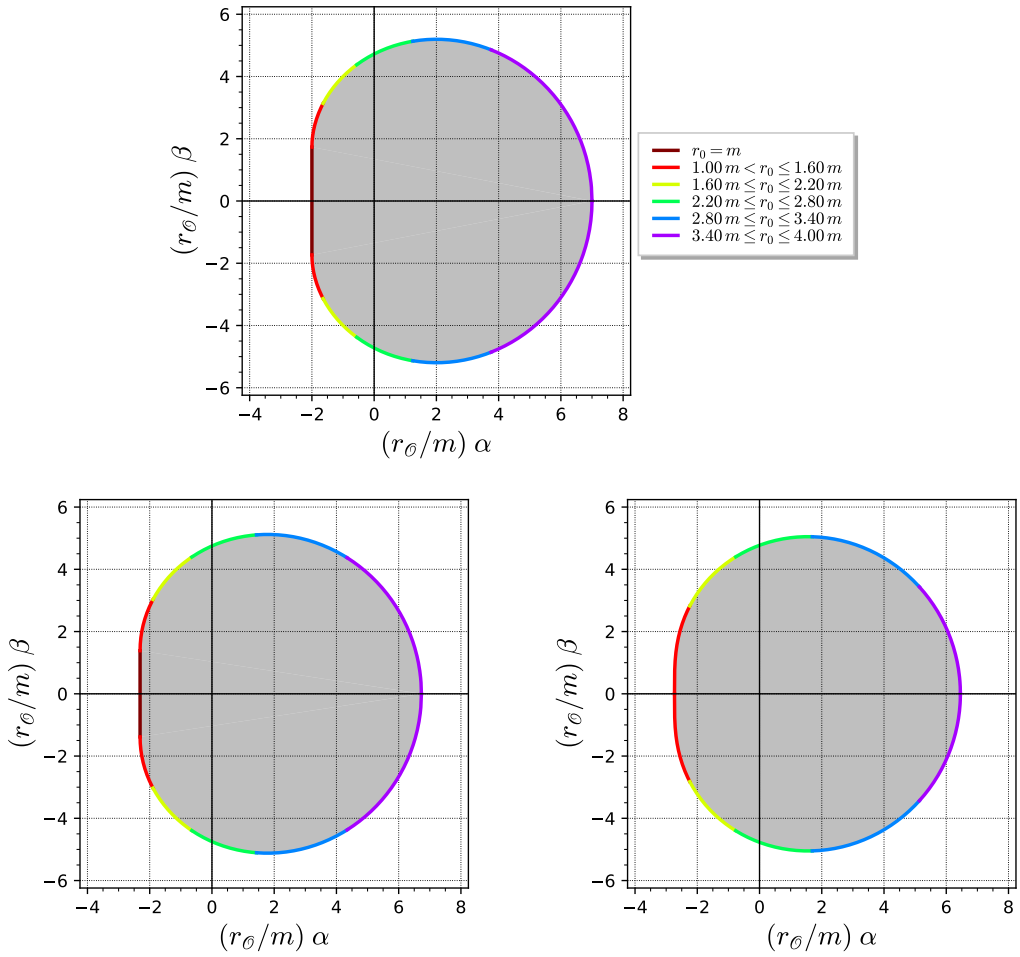


Figure 12.24: Critical curve $\mathscr{C}$ (colored curve) and shadow (grey area) of an extremal Kerr black hole on the screen of an asymptotic inertial observer $\mathscr{O}$ located at $r = r_\mathscr{O}$ and $\theta = \theta_\mathscr{O}$, for three values of $\theta_\mathscr{O}$: $\theta_\mathscr{O} = \pi/2$ ($\mathscr{O}$ in the equatorial plane; upper panel), $\theta_\mathscr{O} = \pi/3$ (lower left) and $\theta_\mathscr{O} = \theta_{\rm crit} = \arcsin(\sqrt{3}-1) \simeq 0.821$ [Eq. (12.98)]. The NHEK line (plotted in maroon) is present in the first two images, while it is vanishing in the third one (marginal case), since there is no NHEK line in $\mathscr{C}$ for $\theta < \theta_{\rm crit}$ or $\theta > \pi - \theta_{\rm crit}$. As in Fig. 12.20, the color code corresponds to some selected ranges for the parameter r_0 of the critical null geodesics forming $\mathscr{C}$. *[Figure generated by the notebook D.4.14]*

is that the end points of the NHEK line, which are obtained for $q/m^2 = 3$, are exactly the two points defined by Eq. (12.96), i.e. the end points of the curve parameterized by r_0 (cf. Fig. 12.22). By adding the NHEK line, we are thus closing the boundary of the black hole shadow! The result is shown in Fig. 12.24, where the NHEK line is drawn in maroon on the left edge of the plots for $\theta_\mathscr{O} = \pi/2$ and $\theta_\mathscr{O} = \pi/3 > \theta_{\rm crit}$.

Cardioid shape

As seen from Fig. 12.24, the departure of the critical curve $\mathscr{C}$ from a perfect circle is maximal when the observer lies in the equatorial plane, i.e. when $\theta_\mathscr{O} = \pi/2$. The critical

curve is often said to have a *D-shape* (from the letter D). It is interesting that this shape corresponds actually to a simple mathematical curve: the convex hull of a cardioid [105]. Indeed, for $\theta_{\mathscr{O}} = \pi/2$, equations (12.94), which govern the part of $\mathscr{C}$ parameterized by r_0, simplify to

$$\alpha = \frac{(r_0 - m)^2 - 2m^2}{m r_{\mathscr{O}}} \tag{12.106a}$$

$$\beta = \frac{\epsilon_\theta}{m r_{\mathscr{O}}} r_0 \sqrt{r_0(4m - r_0)}. \tag{12.106b}$$

Let us then extract r_0 from Eq. (12.106a):

$$r_0 = m\left(1 + \sqrt{2 + \frac{r_{\mathscr{O}}}{m}\alpha}\right)$$

and substitute it into the square of Eq. (12.106b); we get (using $\epsilon_\theta^2 = 1$)

$$\beta^2 = \frac{m^2}{r_{\mathscr{O}}^2}\left(1 + \sqrt{2 + \frac{r_{\mathscr{O}}}{m}\alpha}\right)^3 \left(3 - \sqrt{2 + \frac{r_{\mathscr{O}}}{m}\alpha}\right). \tag{12.107}$$

At this stage, it is worth to introduce explicitly the rescaled angular coordinates used in Figs. 12.22 and 12.24:

$$\bar\alpha := \frac{r_{\mathscr{O}}}{m}\alpha \qquad \text{and} \qquad \bar\beta := \frac{r_{\mathscr{O}}}{m}\beta. \tag{12.108}$$

Then, by expanding its right-hand side, we can rewrite Eq. (12.107) as

$$\bar\beta^2 = -\bar\alpha^2 + 2\bar\alpha + 8\sqrt{\bar\alpha + 2} + 11. \tag{12.109}$$

Let us introduce new screen coordinates $(\hat\alpha, \hat\beta)$ by shifting the origin to $(\bar\alpha, \bar\beta) = (-1, 0)$:

$$\hat\alpha := \bar\alpha + 1 \qquad \text{and} \qquad \hat\beta := \bar\beta. \tag{12.110}$$

We can then recast Eq. (12.109) as

$$\hat\alpha^2 + \hat\beta^2 - 4\hat\alpha = 8\left(\sqrt{\hat\alpha + 1} + 1\right). \tag{12.111}$$

The square of this expression is $(\hat\alpha^2 + \hat\beta^2 - 4\hat\alpha)^2 = 64(\hat\alpha + 2\sqrt{\hat\alpha + 1} + 2)$. Using Eq. (12.111) to get rid of the square root, we obtain

$$\left(\hat\alpha^2 + \hat\beta^2 - 4\hat\alpha\right)^2 = 16\left(\hat\alpha^2 + \hat\beta^2\right). \tag{12.112}$$

The reader may have recognized the Cartesian equation of a cardioid. To put it in a more familiar form, let us introduce the polar coordinates (ρ, ϕ) defined by $\rho^2 := \hat\alpha^2 + \hat\beta^2$, $\cos\phi := \hat\alpha/\rho$ and $\sin\phi := \hat\beta/\rho$. Then Eq. (12.112) becomes $(\rho^2 - 4\rho\cos\phi)^2 = 16\rho^2$, which is equivalent to $(\rho - 4\cos\phi)^2 = 16$, i.e. to $\rho = \pm 4 + 4\cos\phi$. Given that $\rho \geq 0$, the $\pm$ sign must be $+$, so that we end up with

$$\rho = 4(1 + \cos\phi). \tag{12.113}$$

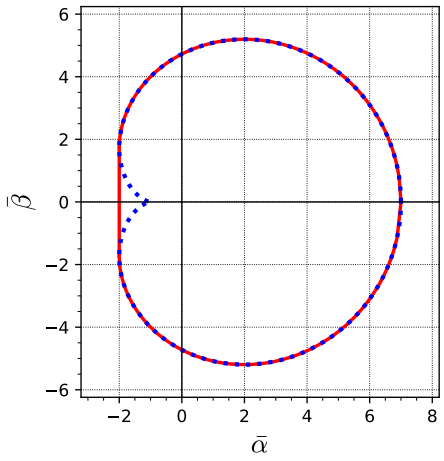


Figure 12.25: Cardioid defined by the polar equation $\rho = 4\,(1+\cos\phi)$ (dotted blue curve) and critical curve of an extremal Kerr black hole seen from the equatorial plane (red curve). The screen coordinates $(\bar{\alpha},\bar{\beta})$ are defined by Eq. (12.108), while (ρ,ϕ) are polar coordinates around the point $(\bar{\alpha},\bar{\beta}) = (-1,0)$, i.e. $\rho := \sqrt{(\bar{\alpha}+1)^2+\bar{\beta}^2}$ and $\sin\phi := \bar{\beta}/\rho$. *[Figure generated by the notebook D.4.14]*

We recognize the polar equation of a cardioid, generated by a circle of radius 2 rolling around a circle of the same radius and centered at $(\hat{\alpha},\hat{\beta}) = (2,0) \iff (\bar{\alpha},\bar{\beta}) = (1,0)$.

As it is clear from the starting point of the above calculation, the cardioid corresponds only to the part of the critical curve parameterized by r_0, i.e. the part depicted in Fig. 12.22. It does not reproduce the NHEK line. Actually, adding the NHEK line results in the convex hull of the cardioid, as shown in Fig. 12.25.

Remark 3: It is amusing to note that the cardioid is involved in two optical phenomena of very distinct origin: in classical optics, it appears as the caustic generated by reflection of light on a circular material, such as a bowl or a coffee cup, and at the same time, it delineates the shadow of some black holes of general relativity.

Remark 4: For $\theta_{\mathscr{O}} \neq \pi/2$, the critical curve $\mathscr{C}$ of the extremal Kerr black hole is no longer a cardioid. It is however still a quartic algebraic curve, i.e. $(\bar{\alpha},\bar{\beta})$ obey an algebraic equation of degree 4, as in Eq. (12.112). Moreover, $\mathscr{C}$ is a classical elementary curve, namely a *Cartesian oval*, also known as *oval of Descartes*, as shown in Ref. [132]. More precisely, $\mathscr{C}$ is exactly such an oval for $\theta_{\mathscr{O}} \leq \theta_{\rm crit}$ or $\theta_{\mathscr{O}} \geq \pi - \theta_{\rm crit}$. For $\theta_{\rm crit} < \theta_{\mathscr{O}} < \pi - \theta_{\rm crit}$, i.e. when the NHEK line is present, $\mathscr{C}$ is the convex hull of a Cartesian oval.

12.4.5 Comparing the critical curves at fixed inclination

From an observational point of view, it may happen that the inclination angle $\theta_{\mathscr{O}}$ is known, as we shall see for M87* in Sec. 12.5.3. Figure 12.26 compares then the critical curves at a fixed value of $\theta_{\mathscr{O}}$ for various values of the black hole spin parameter a. For $\theta_{\mathscr{O}} = 0$, we recover the feature found in Sec. 12.4.3, namely that the critical curve depends very weakly on the spin.

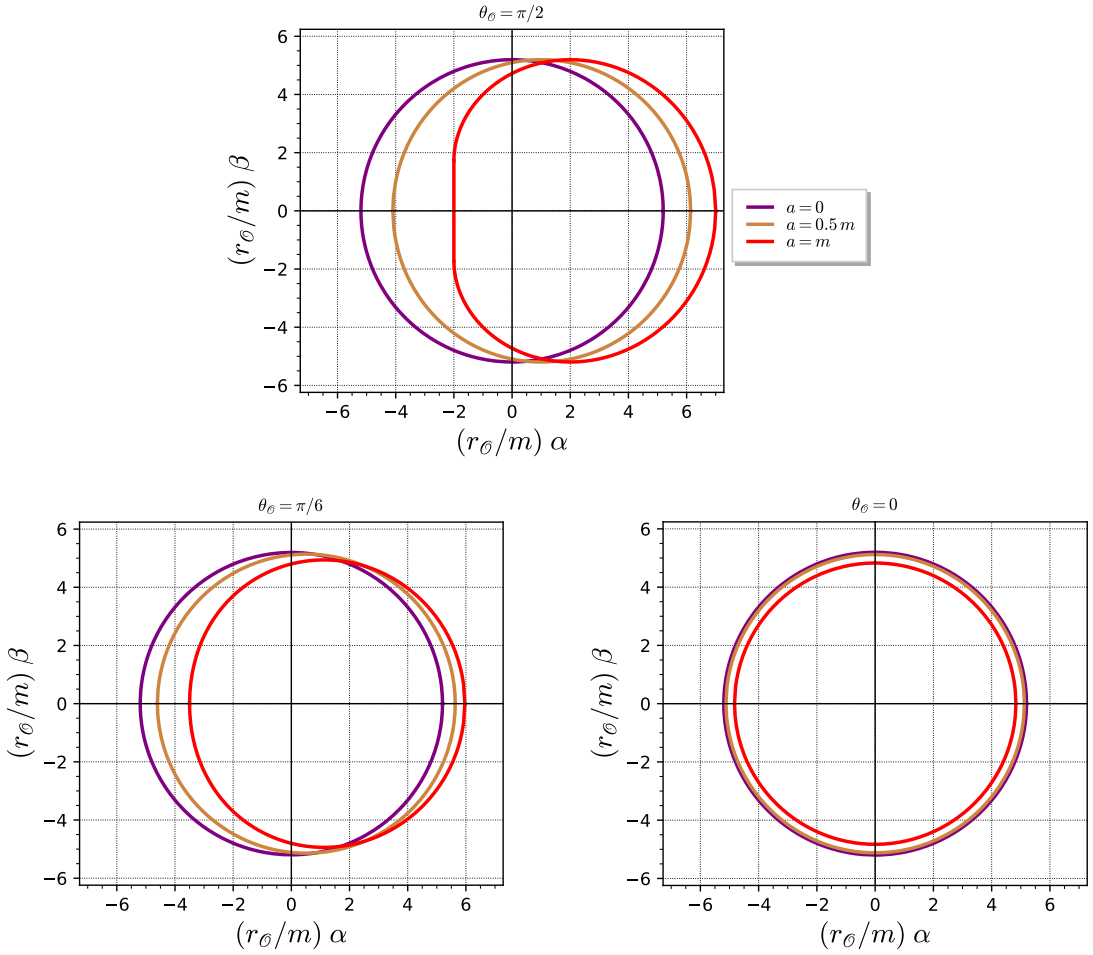


Figure 12.26: Critical curves for different values of a at fixed observer's inclination angle: $\theta_\mathscr{O} = \pi/2$ (top panel), $\theta_\mathscr{O} = \pi/6$ (lower left panel) and $\theta_\mathscr{O} = 0$ (lower right panel). *[Figure generated by the notebook D.4.14]*

Historical note: The critical curve of a Kerr black hole has been computed first by James M. Bardeen in 1972 [17] (cf. the historical note on p. 211), in the form of the parametric system (12.88). He wrote: *The rim of the "black hole"*[7] *corresponds to photon trajectories which are marginally trapped by the black hole; they spiral around many times before they reach the observer. It is conceptually interesting, if not astrophysically very important, to calculate the precise apparent shape of the black hole.* Fifty years later, with the first image from the Event Horizon Telescope [5, 263], this has become astrophysically important! For $a = m$ and $\theta_\mathscr{O} = \pi/2$, Bardeen derived the system (12.106) governing the part of the critical curve parameterized by r_0. For the NEHK line, he simply noted[8]: *The non-uniform nature of the limit $a \to m$ allows q to range between 0 and $3m^2$ at $r = m$.* Bardeen presented then a plot of the resulting critical curve (Fig. 6 in Ref. [17]) similar to that shown in the top panel of Fig. 12.24.

[7] For Bardeen, "black hole" (with quotes) stands for the black spot on the observer's screen.
[8] q is denoted by η by Bardeen.

The term *black hole shadow* for the interior of the critical curve has been coined by Heino Falcke, Fulvio Melia and Eric Agol in 2000 [104], cf. the historical note on p. 434.

12.5 Images

12.5.1 Multiples images of a single source

The main features of images of a single luminous source on the screen of a remote observer in Kerr spacetime are similar to those obtained for Schwarzschild spacetime in Sec. 8.5. In particular, a single point-like source, wherever localized in the black hole exterior (region $\mathscr{M}_{\rm I}$), gives birth to two infinite sequences of images on the screen of the asymptotic inertial observer, both sequences converging to the critical curve $\mathscr{C}$. Typically, one sequence is formed by null geodesics having a reduced angular momentum $\ell > 0$ and the other one by null geodesics with $\ell < 0$. Each image in a given sequence can be labeled by the number n of half-round trips around the black hole and is dimmer and dimmer as n increases.

12.5.2 Image of an accretion disk

Beside gravitational waves, the main way of observing black holes is through the electromagnetic radiation from material orbiting around them, either in the form of stars, as in the case of Sgr A* [3, 4], or in the form of an accretion disk [1]. In the latter case, most observations are spectra and time evolution of the global luminosity (the so-called *light curve*) of the unresolved accretion disk. However recently, the Event Horizon Telescope team has produced the first resolved image of an accretion disk around a black hole [5, 263]. In order to discuss this image in Sec. 12.5.3, let us take a look at the generic properties of (theoretical) images of accretion flows around a Kerr black hole.

Figures 12.27 and 12.28 present some computed images of an accretion disk around a Kerr black hole of spin parameter $a = 0.5\,m$ and $a = 0.95\,m$, respectively. The accretion disk is a simple model developed in Ref. [254]. It consists in a geometrically thick and optically thin accretion disk with an opening angle of 27° and an inner boundary located at the prograde ISCO (cf. Sec. 11.5.3): $r_{\rm in} = r_{\rm ISCO}^+$, with $r_{\rm ISCO}^+ \simeq 4.233\,m$ for $a = 0.5\,m$ (Fig. 12.27) and $r_{\rm ISCO}^+ \simeq 1.937\,m$ for $a = 0.95\,m$ (Fig. 12.28). The disc is in Keplerian rotation and the electromagnetic emission is due to thermal synchrotron radiation of electrons in the local magnetic field (see Ref. [254] for details). The images have been generated by the open-source ray-tracing code Gyoto [253] (cf. Appendix E).

Figure 12.27 provides images for moderate black hole spin: $a = 0.5\,m$. For a given inclination $\theta_{\mathscr{O}}$, the images are rather similar to those of the accretion disk around a Schwarzschild black hole displayed in Fig. 8.25. In particular, one notices three images of the disk, as in Fig. 8.25: a broad primary image, a secondary image with a narrow annular shape and a thin tertiary image appearing as a very thin ring, which is dotted by lack of resolution. The qualitative explanation of these images is basically the same as in the Schwarzschild case. In particular, for the disk face-on view ($\theta_{\mathscr{O}} = 0$), one could use a figure similar to Fig. 8.26, with the orange null geodesics generating the primary image,

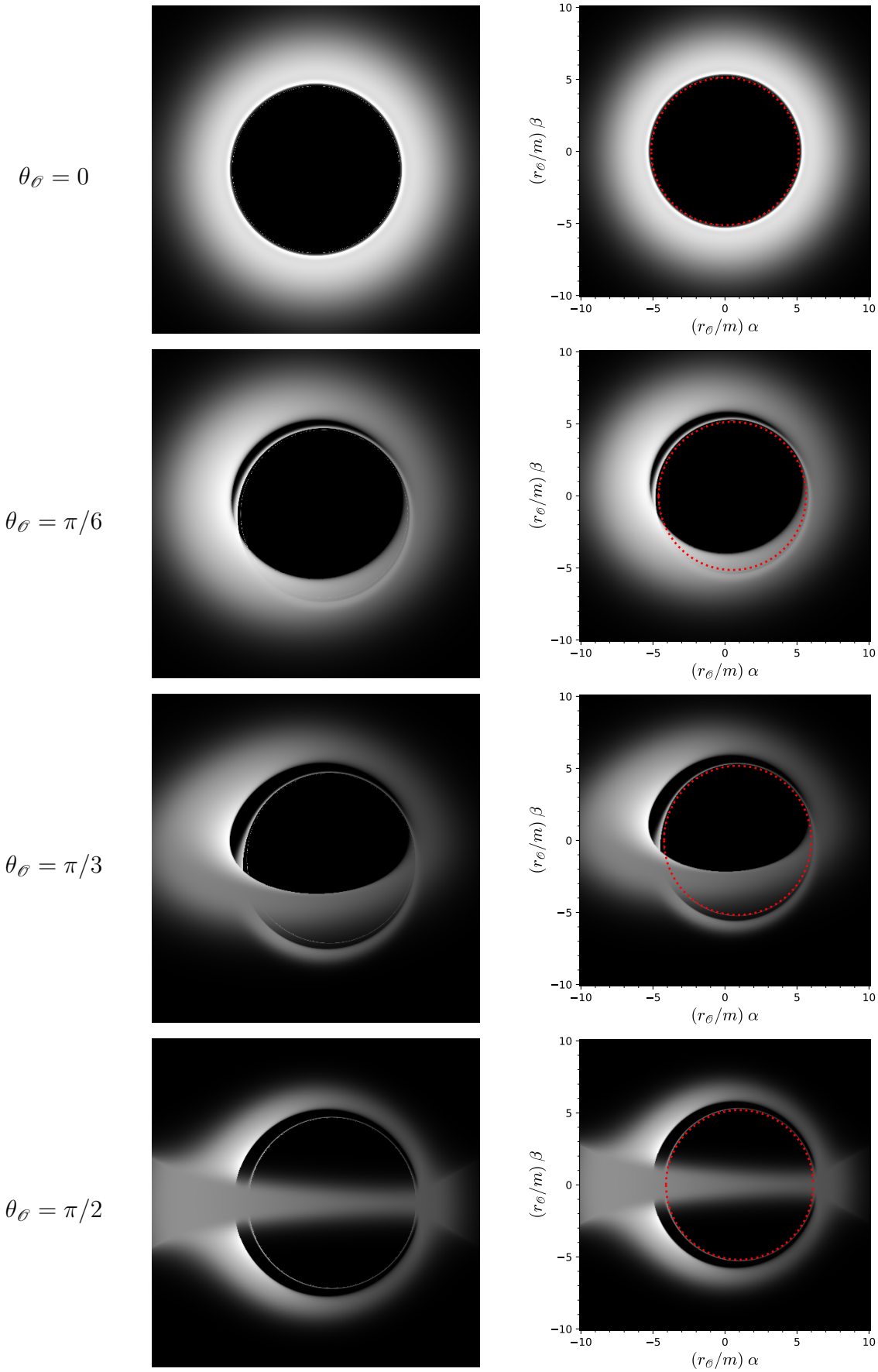


Figure 12.27: Images of a thick accretion disk around a Kerr black bole with $a = 0.5\,m$ for three observer's inclination angles $\theta_{\mathscr{O}}$. The right column shows the critical curve (red dotted line) superposed on the image. *[Images produced by Gyoto with the input files given in Sec. E.2.2; the addition of the critical curve in the right column has been performed by means of the SageMath notebook D.4.15]*

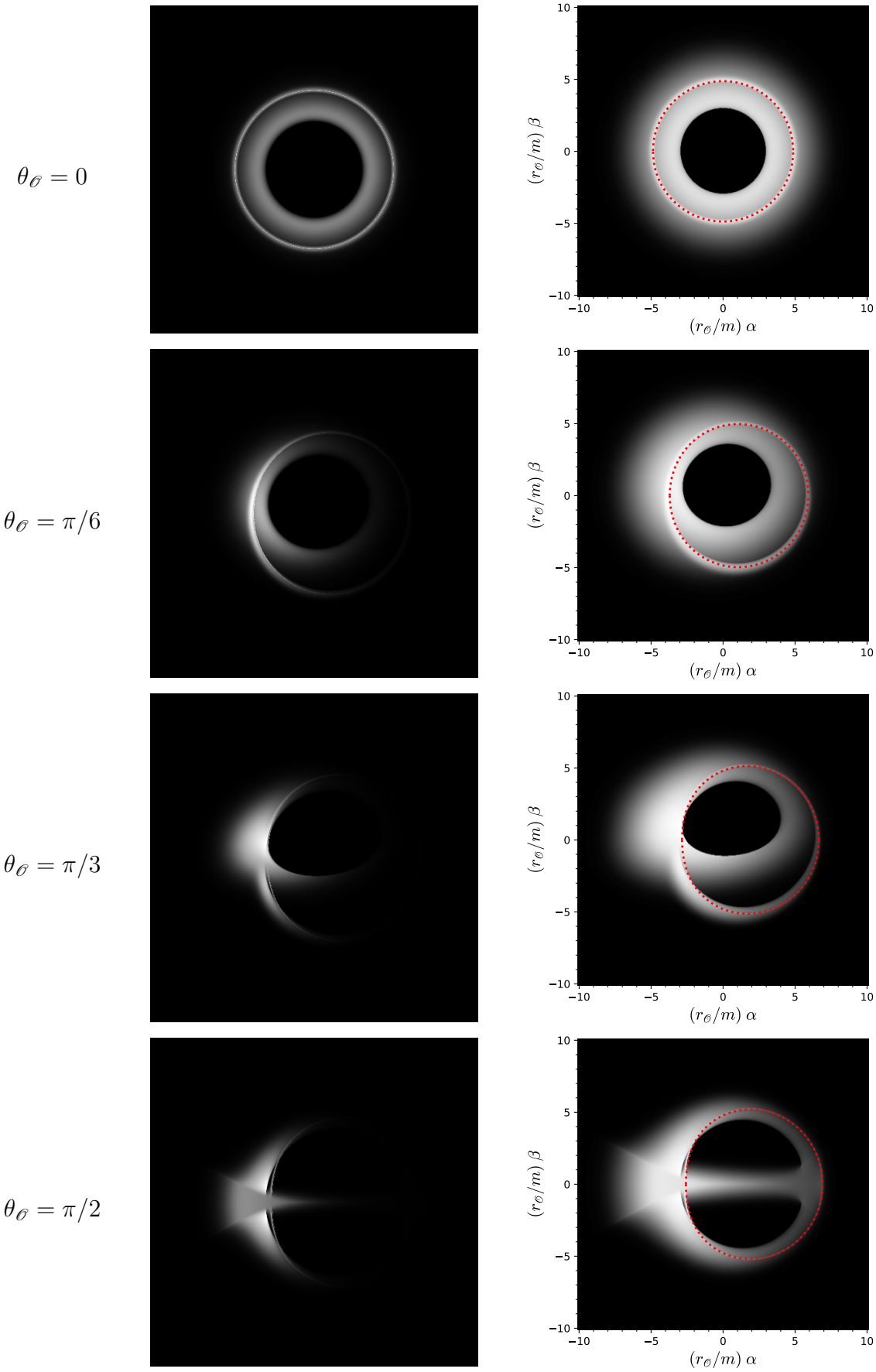


Figure 12.28: Same as Fig. 12.27 but for $a = 0.95\,m$ and the images in the right column using a log scale to reveal the faint parts. The critical curves are the same as in Fig. 12.20.

the brown ones the secondary image and the red ones the tertiary image. The green circle would then mark the location of polar spherical photon orbits, at $r_{\rm ph}^{\rm pol} \simeq 2.883\,m$ for $a = 0.5\,m$, since only the $\ell = 0$ critical null geodesics matter for $\theta_{\mathscr{O}} = 0$. There is actually an infinite sequence of images, the image of order n being generated by null geodesics that have performed n half-turns around the sphere $r = r_{\rm ph}^{\rm pol}$ before leaving toward the observer, with $n = 0$ corresponding to the primary image. A confirmation of this interpretation is provided by the superposition of the critical curve $\mathscr{C}$ onto the image, as the red dotted line in the right column of Fig. 12.27. The tertiary image ($n = 2$) is then almost indistinguishable from $\mathscr{C}$, in agreement with the exponential convergence of high order images to $\mathscr{C}$. This exponential convergence has been established for $a = 0$ by Eq. (8.131), but it holds for $a \neq 0$ as well [130].

Another common feature with the Schwarzschild images of Fig. 8.25 is the left part of the primary image being brighter than the right part as soon as $\theta_{\mathscr{O}} \neq 0$. As discussed in Sec. 8.5.5, this is due to the Doppler boosting resulting from the rotation of the accretion disk, the left part moving towards the observer, while the right part is receding.

A difference with the Schwarzschild images discussed in Sec. 8.5.5 is that in the current case, the disk is not confined to the equatorial plane, being geometrically thick, and moreover it is optically thin. This means that a given null geodesic touching the screen "carries" not only a photon from the disk's surface but also photons from the disk interior, actually all the photons emitted along the path of the geodesic inside the disk. This cumulative property results in a enhanced brightness of the secondary image and makes it appear as a relatively bright ring, as it is clearly seen on the $\theta_{\mathscr{O}} = 0$ and $\theta_{\mathscr{O}} = \pi/6$ images of Fig. 12.27. Another consequence of optical thinness is that the bottom part of the secondary and tertiary images are not blocked by the part of the accretion disk standing in the foreground for $\theta_{\mathscr{O}}$ close to $\pi/2$, as they are in the images of Fig. 8.25, which have been obtained for an optically thick disk model.

Another difference between the two sets of images is that in Fig. 12.27 the inner boundary of the top part of the primary image is closer to the secondary image than in Fig. 8.25 (for $\theta_{\mathscr{O}} = 0$, the primary and secondary images in Fig. 12.27 seem even to touch each other). This is a mere consequence of the inner radius of the accretion disk being closer to the black hole in Fig. 12.27, since $r_{\rm ISCO}^{+} \simeq 4.233\,m$ for $a = 0.5\,m$ and $r_{\rm ISCO}^{+} = 6m$ for $a = 0$.

In the images of Fig. 12.27, the interior of the critical curve, i.e. the black hole shadow according to the definition given in Sec. 12.4.2, appears black, except when crossed by the bottom of the primary image, which arises from the foreground part of the disk. This feature is similar to the Schwarzschild case presented in Fig. 8.25. However, it does not persist for highly spinning black holes, as we can see on Fig. 12.28, which shows images for the Kerr parameter $a = 0.95\,m$. There, the boundary of the central dark spot is clearly distinct from the critical curve and lies within the latter. This is due to the inner part of the accretion disk being located well within some spherical photon orbits, including the polar ones. Indeed, for $a = 0.95\,m$, we have $r_{\rm ISCO}^{+} \simeq 1.937\,m$, while the radius of the retrograde circular photon orbit is $r_{\rm ph}^{-} \simeq 3.995\,m$ and that of polar spherical orbits is $r_{\rm ph}^{\rm pol} \simeq 2.493\,m$. Given that $r_{\rm ph}^{+} \simeq 1.386\,m$, we have thus $r_{\rm ph}^{+} < r_{\rm ISCO}^{+} < r_{\rm ph}^{\rm pol} < r_{\rm ph}^{-}$. More

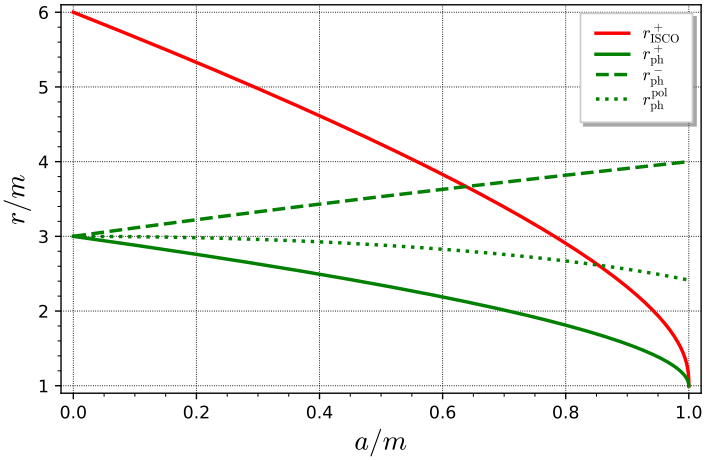


Figure 12.29: Radius of the timelike prograde ISCO, $r^+_{\rm ISCO}$ [Eq. (11.158)], compared with the radii of the prograde and retrograde outer circular photon orbits, $r^+_{\rm ph}$ and $r^-_{\rm ph}$ [Eq. (12.62)] and the radius of polar spherical photon orbits, $r^{\rm pol}_{\rm ph}$ [Eq. (12.67a)]. *[Figure generated by the notebook D.4.15]*

generally, using formulas (11.158), (12.62) and (12.67a), we have

$$r^+_{\rm ISCO} < r^-_{\rm ph} \iff a > 0.638\,m \qquad \text{and} \qquad r^+_{\rm ISCO} < r^{\rm pol}_{\rm ph} \iff a > 0.853\,m. \tag{12.114}$$

These properties are illustrated in Fig. 12.29. Hence, for large spins, if the inner boundary of the accretion disk is set by the prograde ISCO, a large part of the primary image lies strictly inside the critical curve $\mathscr{C}$. Let us for instance consider the image for $\theta_{\mathscr{O}} = 0$ in Fig. 12.28, which is the easiest to understand since for $\theta_{\mathscr{O}} = 0$, $\mathscr{C}$ involves only the $\ell = 0$ critical null geodesics (cf. Sec. 12.4.3), i.e. those that are rolling indefinitely around the sphere $r_0 = r^{\rm pol}_{\rm ph}$ in their asymptotic past. Since $r^+_{\rm ISCO} < r^{\rm pol}_{\rm ph}$ in the present case, this implies that the part of the disk located at $r \in [r^+_{\rm ISCO}, r^{\rm pol}_{\rm ph})$ is an emitting region below the polar spherical photon orbits at $r_0 = r^{\rm pol}_{\rm ph}$. The emitted photons encounter then the screen scritly inside $\mathscr{C}$. We shall not provide a full demonstration here, but simply recall that this was established rigorously for $a = 0$ in Chap. 8: formula (8.131) along with $A(r_{\rm em}) < 0$ for $r_{\rm em} < 3m = r^{\rm pol}_{\rm ph}(a = 0)$ (cf. Fig. 8.22) yields impact parameters $b^\pm_n$ that are lower than $b_{\rm c}$, the latter being nothing but the radius of $\mathscr{C}$ for $a = 0$. Hence we conclude that

> If the black hole is surrounded by some emitting material located within some of the spherical photon orbits, the central black spot in the images does not correspond to the shadow defined in Sec. 12.4.2 as the interior of the critical curve $\mathscr{C}$. The black spot is actually smaller than the shadow, lying strictly inside $\mathscr{C}$.

This holds for $a = 0$ as well, if the emitting matter is not limited inward by the ISCO as in Fig. 8.25 (see e.g. Fig. 5 in Ref. [129] or Fig. 2 in Ref. [254]). Note however that in the case of a purely radial inflow, the emitted photons suffer such a strong Doppler effect that the interior of the critical curve is very dark, so that the effective shadow coincides with the academic shadow of Sec. 12.4.2, as shown in Refs. [104, 193].

A striking feature of the images shown in Fig. 12.28 is that while the central black spot departs considerably from spherical symmetry at large inclination angle $\theta_{\mathscr{O}}$, the secondary and higher order images are pretty circular. This is because they accumulate onto the critical curve, which is quite close to a circle, as shown in Fig. 12.20. The only major effect of a high inclination angle is to move the high order images the right.

Besides, we note on Fig. 12.28 that the secondary image ($n = 1$) is brighter than the primary one ($n = 0$). Moreover, the tertiary one ($n = 3$), which is very thin and superposed onto the secondary image, is even brighter. This is because the higher the image order, the larger the number n of half-turns around the black hole of the involved geodesics, and thus the longer the path length of the geodesics inside the disk, resulting a larger number of photons "carried" by a given geodesic and accumulating on the same pixel of the screen. Note that this holds only because the disk is optically thin. Would it be optically thick, a null geodesic impacting the screen, even if very close to a critical geodesic, would carry only a single photon: the photon emitted at the first intersection of the geodesic with the disk's surface when traced backward in time from the screen. In Sec. 8.5.5 (Schwarzschild case), we considered such an optically thick disk (Page-Thorne model) and the secondary and tertiary images were not brighter than the primary one (cf. Fig. 8.25). However in that case, the main reason was that the emitting region was too far from the black hole for a close-to-critical geodesic to visit various parts of the disk. For instance, in the lower part of Fig. 8.26, consider the red geodesic that intersects the disk at the point of coordinates $(x, y) = (0, 6m)$; it contributes to the tertiary image and if one would extend it to the past beyond the point $(0, 6m)$, it would never encounter the disk again, leaving the figure at the upper left corner in a more or less straight line. We shall therefore state:

> The set constituted by the secondary image and higher order images, i.e. all images of order $n \geq 1$, is called the ***photon ring***. It appears as a bright feature on the observer's screen under two conditions: the emitting material must (i) be optically thin and (ii) be located in the vicinity of some spherical photon orbits.

As shown in Fig. 12.28, the photon ring generally appears as a single ring because the successive images are superposed on each other and converge exponentially to the the critical curve $\mathscr{C}$. Only the $n = 1$ and $n = 2$ parts of the photon ring are visible in Fig. 12.28 by lack of resolution, since the thickness of the n-th image decays exponentially with n [130]. We note on Figs. 12.27 and 12.28 that the $n = 2$ image (tertiary image) is already very close to $\mathscr{C}$. We conclude that

> In practice the photon ring, and more precisely the part $n \geq 2$ of it, materializes the critical curve.

The photon ring is therefore a potentially observable feature that can provide information on the spacetime metric independently of the emission model [161]. For instance, by analyzing the photon ring, one could check that the metric is the Kerr metric and measure the spin parameter a.

Remark 1: The reader is warned that the term *photon ring* is sometimes used in the literature

for circular photon orbits, as the orbit at $r = 3m$ in Schwarzschild spacetime and the orbits at $r = r_{\rm ph}^{\pm}$ in Kerr spacetime (see e.g. Ref. [37] for a recent example). We are using it here not for an orbit around the black hole but for a feature on the screen of a remote observer, in agreement with Refs. [21, 160, 161, 130, 132]. Recently the term *secondary ring*. has been introduced [254] to denote the subpart of the photon ring that contributes significantly to the flux in the image. Besides, some authors have distinguished the secondary image ($n = 1$) by calling it the *lensing ring* [129].

Historical note: In 1973, Christopher T. Cunningham and James M. Bardeen [73] computed the multiple images of a star orbiting an extremal Kerr black hole ($a = m$). In 1979, Jean-Pierre Luminet [179] predicted that a Schwarzschild black hole illuminated by a parallel light beam would appear to a remote observer as a black disk of radius $R_{\rm shad} = 3\sqrt{3}m/r_{\mathscr{O}}$ (cf. Eq. (12.91)) surrounded by a sequence of rings, called *ghost rings* by Luminet, converging exponentiallty to the rim of the black disk (which is the critical curve). These rings are fainter and fainter, the brightest ring being the outermost one. In the very same article [179], Luminet presented the first computed image of an accretion disk around a Schwarzschild black hole (cf. historical note on p. 215). The first images of an accretion disk around a Kerr black hole with $a \neq 0$ have been computed by S.U. Viergutz in 1993 [252], for $a = 0.998\,m$. In 1997, Michał Jaroszyński and Andrzej Kurpiewski [158] studied an optically thin accretion disk extending inwards down to the event horizon of a Kerr black hole with $a = 0$, $0.5\,m$ and $0.9\,m$, with some nonzero radial component of the accretion flow. They showed that the observed intensity just outside the critical curve is enhanced by the long path length of geodesics orbiting many times around the black hole, while the intensity inside the critical curve is low, even if there is some emitting material close to the black hole. In 2000, Heino Falcke, Fulvio Melia and Eric Agol [104] (see also Ref. [103]) have computed images of an optically thin accretion flow around a Kerr black hole with $a = 0.998\,m$. The accretion flow was assumed to extend down to the event horizon, infalling with constant angular momentum inside the ISCO. In agreement with Jaroszyński and Kupriewski's prediction, they obtained a large intensity just outside the critical curve and low intensity inside it. For this reason, they introduced the term *shadow of the black hole* for the interior of the critical curve, the latter being called by them the *apparent boundary of the black hole*. Moreover, in the same article [104], they advanced that, in the case of Galactic Center black hole Sgr A*, the shadow could be observed by means of very-long-baseline interferometry (VLBI), thereby proposing what would become the Event Horizon Telescope two decades later [5].

12.5.3 EHT image of M87*

The computed black hole images, as those shown in Figs. 12.27 and 12.28, can be contrasted with actual observations since the release of the very first observed image by the Event Horizon Telescope (EHT) collaboration in 2019 [5, 263]. This image, shown in Fig. 12.30, is that of the supermassive black hole M87* in the nucleus of the giant elliptic galaxy Messier 87 at the center of the Virgo cluster. It is actually a *reconstructed* image, after more than one year of data processing of an incomplete data set in the Fourier plane of the image obtained by very-long-baseline interferometry (VLBI) in April 2017.

In order to interpret the EHT image, it is natural to superpose the (theoretical) critical curve $\mathscr{C}$ of a Kerr black hole onto it. This is all the more meaningful that $\mathscr{C}$ is

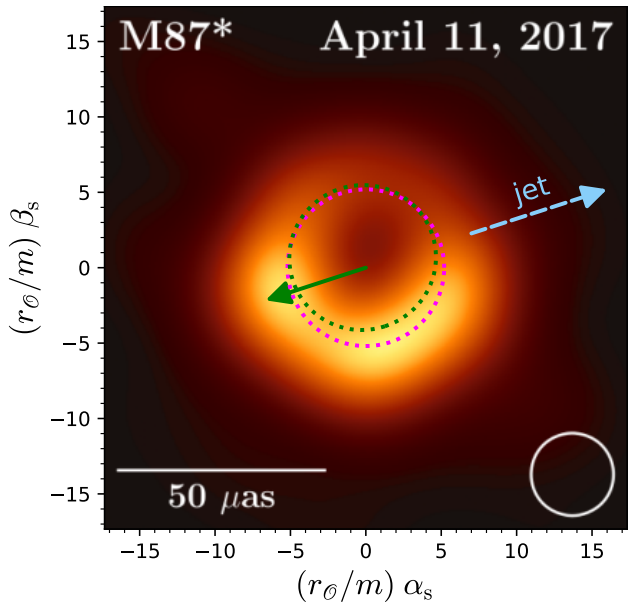


Figure 12.30: Image of the immediate vicinity of the supermassive black hole M87*, as released by the EHT collaboration in 2019 [5], with two critical curves superposed: that of a Schwarzschild black hole (magenta dotted circle) and that of an extremal Kerr black hole seen under the inclination $\theta_{\mathscr{O}} = 163°$ (green dotted curve), with the projection of the spin axis onto the screen indicated by the green arrow (position angle $\Theta = 108°$ with respect to the β_s-axis). The critical curves are scaled by assuming $m/r_{\mathscr{O}} = 3.7$ μas (cf. Table 12.1). (α_s, β_s) are screen angular coordinates, closely related to celestial equatorial coordinates: α_s is minus the (relative) right ascension and β_s is the (relative) declination. The white circle in the bottom right corner indicates the EHT resolution (approx. 20 μas). *[Figure generated by the notebook D.4.16; source of the EHT image: Fig. 3 of Ref. [5]]*

closely related to the photon ring, as discussed in Sec. 12.5.2. In order to perform the superposition for a given value of the Kerr spin parameter a, one must know three things: (i) the overall scale factor $m/r_{\mathscr{O}}$, (ii) the angle $\theta_{\mathscr{O}}$ between the black hole rotation axis and the line of sight and (iii) the orientation Θ of the projection of the rotation axis onto the plane of the sky. It turns out that for M87* these three quantities can be estimated. Indeed, the distance from the Earth to M87* has been measured to $r_{\mathscr{O}} \simeq 16.7$ Mpc and the mass of M87* has been deduced from stellar dynamics in the inner part of the galactic nucleus, resulting in[9] $m \simeq 6.2\ 10^9\, M_\odot$ (cf. Table 12.1 and references therein), hence the scale factor $m/r_{\mathscr{O}} \simeq 3.7$ μas. Regarding $\theta_{\mathscr{O}}$ and Θ, both angles can be determined via reasonable assumptions, thanks to the large relativistic jet emanating from the close vicinity of the black hole[10], since most (all?) theoretical models predict that the jet is aligned with the black hole's rotation axis. There are actually two jets, emitted from both

[9] Another method, based on the dynamics of the gas in the nucleus disk at $r < 40$ pc, yields a mass of only $m = 3.5\ 10^9\, M_\odot$ [260]; the discrepancy with the value from stellar dynamics has not been explained yet.

[10] This jet has been discovered in 1918 [74], hence well before one suspects a black hole to lie in the core of the galaxy M87! It makes M87 belong to the category of galaxies with an *active galactic nucleus (AGN)*. The jet engine is related to the black hole, possibly via the Penrose process discussed in Sec. 10.5.5, see e.g. Ref. [1].

sides of the black hole in opposite directions. Due to a strong Doppler beaming effect, the jet moving away from us, usually called the *counterjet*, is hardly visible, so that the large observed jet is the one moving towards us. The position angle of the jet in the plane of the sky, measured from the North axis, is $\Theta_{\rm jet} \sim -72°$ (see e.g. Fig. 1 in Ref. [259]); it is indicated by the light blue dashed arrow in Fig. 12.30 (the jet is not visible on the EHT image). The inclination ι of the jet with respect to the line of sight is estimated from observed motions of "nodes" within the jet, resulting in a value around $\iota \sim 17°$ [186, 259]. In particular, the detection of apparent superluminal motions, up to $v_{\rm app} \sim 6c$, implies both relativistic velocities and ι being a small angle (see e.g. Sec. 5.7.4 of Ref. [123]). Assuming that the jet axis is aligned with the black holes's rotation axis, we have either $\theta_{\mathscr{O}} = \iota$ or $\theta_{\mathscr{O}} = \pi - \iota$.

The main feature of the EHT image (Fig. 12.30) is a broad annular region, whose bottom part is much brighter than its top part. It is natural to interpret this brightness discrepancy as resulting from relativistic Doppler beaming: the bottom (resp. top) part corresponds to emitting material moving towards us (resp. receding from us) [6]. Assuming that, so close to the black hole, the emitting matter is rotating in the same sense as the black hole, due to the Lense-Thirring effect (cf. Sec. 11.3.3), this implies that the spin of black hole is pointing away from us, i.e. that $\theta_{\mathscr{O}} = \pi - \iota \sim 163°$ and $\Theta = \pi + \Theta_{\rm jet} \sim 108°$. The projection of the resulting black hole spin vector onto the plane of the sky is shown by the green arrow in Fig. 12.30. To superpose the critical curves computed in Sec. 12.4 onto the EHT image, it suffices then to rotate them by the angle Θ and to use the M87* scale factor $m/r_{\mathscr{O}} = 3.7$ μas determined above. In Fig. 12.30, we have done it for the two extreme values that can be taken by the spin parameter a: $a = 0$ (magenta dotted cirle) and $a = m$ (green dotted curve). The two critical curves are quite close to each other because the Earth observer is located almost on the axis of rotation, $\theta_{\mathscr{O}}$ being close to π. To have a stronger discrepancy among various possible values of a, it would have been better to have $\theta_{\mathscr{O}}$ close to $\pi/2$ instead, as illustrated in Fig. 12.26.

The main conclusion that one can draw from Fig. 12.30 is that the overall scale of the EHT image fits well with what could be expected from emitting matter in the close vicinity of a black hole of mass around $6\ 10^9\, M_\odot$. Unfortunately, the resolution of the EHT image ($\sim$ 20 μas $\sim$ 1/6 of the image width!) is not sufficient to draw sharper conclusions. In particular, one cannot assert whether the annular structure in the image is the photon ring discussed in Sec. 12.5.2 blurred to the EHT resolution or the primary image of an accretion disk or some more complicated accretion flow. It follows that it is not possible to infer the (currently unknown) spin parameter a from that image, not speaking about performing any strong test of general relativity (see Refs. [128, 120] for an extended discussion of this last point). However, this is the very first image of this type and the next images, which will result from some EHT upgrade or from space-based VLBI, will undoubtedly be sharper and will certainly revolutionize black hole astronomy.

12.5.4 Going further

For a detailed study of images of the surroundings of a Kerr black hole, see Refs. [90, 91, 92, 130]. In particular, Ref. [130] provides exact solutions of the null geodesic equation via

elliptic integrals, generalizing those derived in Chap. 8 for the Schwarzschild black hole. We have discussed only images as seen by a remote observer, which is the "astronomical" setting. For the images seen by an observer travelling close to, and even inside, a Kerr black hole (the "science-fiction" setting), see the recent study [222].

Chapter 13

Extremal Kerr black hole

Contents

13.1 Introduction

The Kerr solution of Einstein equation has been introduced in Sec. 10.2; it depends on two parameters: the mass $m > 0$ and the spin parameter $a \geq 0$. In Chaps. 10–12, we have considered the Kerr solution with $0 < a < m$, while the case $a = 0$ (Schwarzschild solution) was the subject of Chaps. 6–9. Here we focus on the case $a = m$. As we going to see, it has many properties that are not shared by the Kerr solution with $a < m$. In particular, the black hole event horizon is *degenerate*, in the sense defined in Sec. 3.3.6, i.e. it has a vanishing surface gravity κ. Another specific property is the geometry near the horizon admitting an enlarged symmetry group, which is generated by four independent Killing vectors, instead of two for the geometry of the global solution. Let us mention that $a = m$ is the highest value of a for which the Kerr solution describes a black hole. Indeed, for $a > m$, the Kerr metric is still an exact solution of the vacuum Einstein equation, but it describes a *naked singularity* (cf. Sec. 9.6.1): the ring curvature singularity is not hidden by any horizon to asymptotic observers.

13.2 Definition and basic properties

13.2.1 The extremal Kerr solution

Let us consider the manifold $\mathbb{R}^2 \times \mathbb{S}^2$ and describe it by coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ such that $(\tilde{t}, r)$ cover $\mathbb{R}^2$ and $(\theta, \tilde{\varphi})$ are standard spherical coordinates on $\mathbb{S}^2$. The ***extremal Kerr***

spacetime of mass $m > 0$ is defined as the pair $(\mathscr{M}, \boldsymbol{g})$ where the manifold $\mathscr{M}$ is the following open subset of $\mathbb{R}^2 \times \mathbb{S}^2$:

$$\mathscr{M} := \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathscr{R} \tag{13.1}$$

with

$$\mathscr{R} := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = 0 \text{ and } \theta(p) = \frac{\pi}{2} \right\}, \tag{13.2}$$

and the metric $\boldsymbol{g}$ has the following expression in terms of the coordinates $(x^{\tilde{\alpha}}) = (\tilde{t}, r, \theta, \tilde{\varphi})$:

$$\begin{aligned} g_{\tilde{\mu}\tilde{\nu}}\, \mathrm{d}x^{\tilde{\mu}}\, \mathrm{d}x^{\tilde{\nu}} = & -\left(1 - \frac{2mr}{\rho^2}\right) \mathrm{d}\tilde{t}^2 + \frac{4mr}{\rho^2} \mathrm{d}\tilde{t}\, \mathrm{d}r - \frac{4m^2 r \sin^2\theta}{\rho^2} \mathrm{d}\tilde{t}\, \mathrm{d}\tilde{\varphi} \\ & + \left(1 + \frac{2mr}{\rho^2}\right) \mathrm{d}r^2 - 2m \left(1 + \frac{2mr}{\rho^2}\right) \sin^2\theta\, \mathrm{d}r\, \mathrm{d}\tilde{\varphi} \\ & + \rho^2 \mathrm{d}\theta^2 + \left(r^2 + m^2 + \frac{2m^3 r \sin^2\theta}{\rho^2}\right) \sin^2\theta\, \mathrm{d}\tilde{\varphi}^2, \end{aligned} \tag{13.3}$$

with

$$\rho^2 := r^2 + m^2 \cos^2\theta. \tag{13.4}$$

In this context, the coordinates $(x^{\tilde{\alpha}}) = (\tilde{t}, r, \theta, \tilde{\varphi})$ are called ***Kerr coordinates*** and we recognize in (13.3) the limit $a \to m$ of expression (10.36) for the Kerr metric with $a < m$.

The metric (13.3) is regular in all $\mathscr{M}$, since the components $g_{\tilde{\alpha}\tilde{\beta}}$ are singular only for $\rho = 0$, i.e. for $r = 0$ and $\theta = \pi/2$, which defines the set $\mathscr{R}$ that has precisely been excluded from $\mathscr{M}$ by the definition (13.1). The Kretschmann curvature invariant $K := R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$ is given by Eq. (10.22) with $a = m$; it diverges for $\rho \to 0$. Therefore, as for the Kerr spacetime with $a < m$ (cf. Sec. 10.2.6), we shall call $\mathscr{R}$ the ***ring singularity*** of the extremal Kerr spacetime. Note that, formally, it is not part of the spacetime manifold $\mathscr{M}$ [cf. Eq. (13.1)].

Moreover, the Ricci tensor of the metric (13.3) is identically zero in all $\mathscr{M}$ (see the notebook D.4.17 for the computation). Hence, we have:

The metric $\boldsymbol{g}$ of the extremal Kerr spacetime is a solution of Einstein equation (1.38) in vacuum ($\boldsymbol{T} = 0$) and with a vanishing cosmological constant ($\Lambda = 0$).

The inverse metric is

$$g^{\tilde{\alpha}\tilde{\beta}} = \begin{pmatrix} -\left(1 + \frac{2mr}{\rho^2}\right) & \frac{2mr}{\rho^2} & 0 & 0 \\ \frac{2mr}{\rho^2} & \frac{(r-m)^2}{\rho^2} & 0 & \frac{m}{\rho^2} \\ 0 & 0 & \frac{1}{\rho^2} & 0 \\ 0 & \frac{m}{\rho^2} & 0 & \frac{1}{\rho^2 \sin^2\theta} \end{pmatrix}. \tag{13.5}$$

13.2.2 Boyer-Lindquist coordinates

For $a < m$, the Kerr manifold $\mathscr{M}$ has been split in three open regions, $\mathscr{M}_{\rm I}$, $\mathscr{M}_{\rm II}$ and $\mathscr{M}_{\rm III}$, separated by the two Killing horizons $\mathscr{H}$ and $\mathscr{H}_{\rm in}$ [cf. Eqs. (10.28) and (10.2a)]. Since $\mathscr{H}$ was defined by $r = r_+ := m+\sqrt{m^2-a^2}$ [Eq. (10.26)] and $\mathscr{H}_{\rm in}$ by $r = r_- := m-\sqrt{m^2-a^2}$ [Eq. 10.27)], we notice that $r_+ = r_- = m$ in the $a \to m$ limit. This implies that $\mathscr{H}$ and $\mathscr{H}_{\rm in}$ coincide when $a \to m$ and the region $\mathscr{M}_{\rm II}$, which is bounded by $\mathscr{H}$ and $\mathscr{H}_{\rm in}$, disappears. Accordingly, we shall split the extremal Kerr manifold $\mathscr{M}$ in two open regions only, $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$, separated by a single hypersurface $\mathscr{H}$:

$$\boxed{\mathscr{M} = \mathscr{M}_{\rm I} \cup \mathscr{H} \cup \mathscr{M}_{\rm III}}, \tag{13.6}$$

with

$$\mathscr{M}_{\rm I} := \{p \in \mathscr{M}\,, \quad r(p) > m\} \tag{13.7a}$$
$$\mathscr{H} := \{p \in \mathscr{M}\,, \quad r(p) = m\} \tag{13.7b}$$
$$\mathscr{M}_{\rm III} := \{p \in \mathscr{M}\,, \quad r(p) < m\}\,. \tag{13.7c}$$

Remark 1: We are using the notation $\mathscr{M}_{\rm III}$ for the "second" region, and not $\mathscr{M}_{\rm II}$, to be consistent with Chaps. 10–12, i.e. with the $a \to m$ limit of the results obtained in these chapters.

The quadratic polynomial in r introduced in Chap. 10, $\Delta = r^2 - 2mr + a^2 = (r - r_+)(r - r_-)$ reduces to $\Delta = (r-m)^2$ in the limit $a \to m$. Its double root, $r = m$, defines the hypersurface $\mathscr{H}$.

In the region $\mathscr{M}_{\rm BL} := \mathscr{M} \setminus \mathscr{H} = \mathscr{M}_{\rm I} \cup \mathscr{M}_{\rm III}$, one may introduce the ***Boyer-Lindquist coordinates*** (t, r, θ, φ) such that (r, θ) are the same coordinates as in Kerr coordinates, while t and φ are related to the Kerr coordinates $\tilde{t}$, r and $\tilde{\varphi}$ by

$$t = \tilde{t} + \frac{2m^2}{r-m} - 2m \ln \left| \frac{r-m}{m} \right| \tag{13.8a}$$
$$\varphi = \tilde{\varphi} + \frac{m}{r-m}. \tag{13.8b}$$

Differentiating these relations leads to

$$\mathrm{d}\tilde{t} = \mathrm{d}t + \frac{2mr}{(r-m)^2}\,\mathrm{d}r \qquad \text{and} \qquad \mathrm{d}\tilde{\varphi} = \mathrm{d}\varphi + \frac{m}{(r-m)^2}\,\mathrm{d}r. \tag{13.9}$$

Remark 2: The differential relations (13.9) can be obtained immediately by substituting a by m in relations (10.34). However, to get the integrated relations (13.8) from their $a < m$ analog (10.35), one must perform an expansion in $\varepsilon := \sqrt{m^2-a^2}/m$, taking into account that $r_\pm = m(1 \pm \varepsilon)$. One obtains then (13.8a) up to the additive constant $(2 \ln 2)m$, while (13.8b) is recovered in the same form.

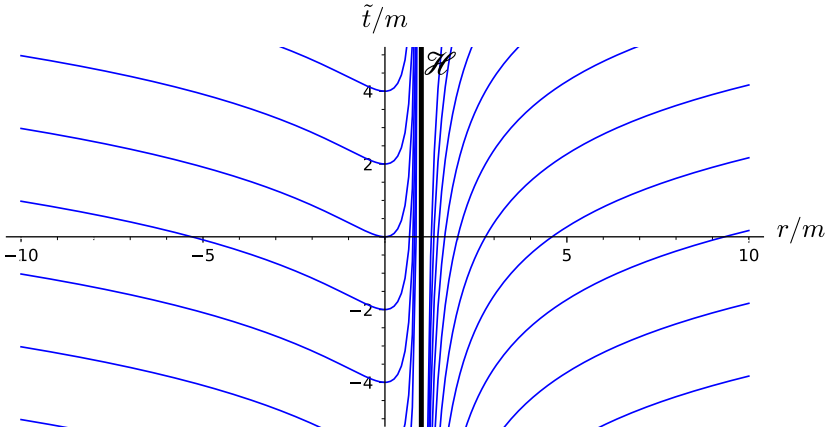


Figure 13.1: Trace of the hypersurfaces of constant Boyer-Lindquist time t in the plane $(\tilde{t}, r)$. *[Figure generated by the notebook D.4.17]*

It follows from the transformations (13.8) that the Boyer-Lindquist coordinate frame $(\boldsymbol{\partial}_\alpha)$ and the Kerr coordinate frame $(\boldsymbol{\partial}_{\tilde{\alpha}})$ are related by[1]

$$\boldsymbol{\partial}_t = \boldsymbol{\partial}_{\tilde{t}} \tag{13.10a}$$

$$\boldsymbol{\partial}_r = \boldsymbol{\partial}_{\tilde{r}} + \frac{2mr}{(r-m)^2}\boldsymbol{\partial}_{\tilde{t}} + \frac{m}{(r-m)^2}\boldsymbol{\partial}_{\tilde{\varphi}} \tag{13.10b}$$

$$\boldsymbol{\partial}_\theta = \boldsymbol{\partial}_\theta \tag{13.10c}$$

$$\boldsymbol{\partial}_\varphi = \boldsymbol{\partial}_{\tilde{\varphi}}. \tag{13.10d}$$

Note that, as in Chap. 10, we have denoted by $\boldsymbol{\partial}_{\tilde{r}}$ the second vector of the coordinate frame associated to the Kerr coordinates $(x^{\tilde{\alpha}}) = (\tilde{t}, r, \theta, \tilde{\varphi})$, in order to distinguish it from the coordinate vector $\boldsymbol{\partial}_r$ of the Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$.

The metric components $(g_{\alpha\beta})$ with respect to the Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$ are given by

$$\boxed{\begin{aligned} g_{\mu\nu}\,\mathrm{d}x^\mu \mathrm{d}x^\nu = & -\left(1 - \frac{2mr}{\rho^2}\right)\mathrm{d}t^2 - \frac{4m^2 r \sin^2\theta}{\rho^2}\,\mathrm{d}t\,\mathrm{d}\varphi + \frac{\rho^2}{(r-m)^2}\,\mathrm{d}r^2 \\ & + \rho^2 \mathrm{d}\theta^2 + \left(r^2 + m^2 + \frac{2m^3 r \sin^2\theta}{\rho^2}\right)\sin^2\theta\,\mathrm{d}\varphi^2. \end{aligned}} \tag{13.11}$$

This expression can be obtained either by taking the limit $a \to m$ of Eq. (10.8) or by using (13.9) to substitute $\mathrm{d}\tilde{t}$ and $\mathrm{d}\tilde{\varphi}$ in Eq. (13.3). We note that $g_{rr} \to +\infty$ when $r \to m$, which reflects the singularity of Boyer-Lindquist coordinates on $\mathscr{H}$. This singularity is clearly apparent in the coordinate transformations (13.8), as well as in the spacetime slicing by the hypersurfaces $t = \text{const}$ depicted in Fig. 13.1: the slices accumulate onto $\mathscr{H}$, without crossing it, so that the points on $\mathscr{H}$ do not belong to any hypersurface $t = \text{const}$. That

[1] See also the $a \to m$ limit of Eq. (10.38).

the $t = \text{const}$ hypersurfaces do not provide a regular slicing of the extremal Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ is also manifest on the Carter-Penrose diagram shown in Fig. 13.4.

The inverse metric $\boldsymbol{g}^{-1}$ has the following components in the Boyer-Lindquist coordinates (cf. the $a \to m$ limit of Eq. (10.16)):

$$g^{\alpha\beta} = \begin{pmatrix} -\frac{1}{(r-m)^2}\left(r^2+m^2+\frac{2m^3 r\sin^2\theta}{\rho^2}\right) & 0 & 0 & -\frac{2m^2 r}{\rho^2(r-m)^2} \\ 0 & \frac{(r-m)^2}{\rho^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\rho^2} & 0 \\ -\frac{2m^2 r}{\rho^2(r-m)^2} & 0 & 0 & \frac{1}{(r-m)^2\sin^2\theta}\left(1-\frac{2mr}{\rho^2}\right) \end{pmatrix}. \tag{13.12}$$

13.2.3 Symetries

The extremal Kerr metric (13.3) is stationary[2] and axisymmetric. The corresponding isometry group is $\mathbb{R} \times \mathrm{SO}(2)$ and is generated by two commuting Killing vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$. Both the Kerr coordinates and the Boyer-Lindquist ones are adapted to the spacetime symmetries, i.e. $(\tilde{t}, \tilde{\varphi})$ and (t, φ) are ignorable coordinates, as it is clear on the line elements (13.3) and (13.11). Accordingly, one can normalize the Killing vectors so that

$$\boxed{\boldsymbol{\xi} = \boldsymbol{\partial}_{\tilde{t}} = \boldsymbol{\partial}_t} \qquad \text{and} \qquad \boxed{\boldsymbol{\eta} = \boldsymbol{\partial}_{\tilde{\varphi}} = \boldsymbol{\partial}_\varphi}. \tag{13.13}$$

13.2.4 Principal null geodesics

As discussed in Sec. 10.4, the Kerr spacetime is endowed with two congruences of null geodesics that are tied to the spacetime structure, as described by the Weyl conformal curvature tensor. All the results of Sec. 10.4 remain valid at the limit $a \to m$. We can summarize them as follows:

- The ***ingoing principal null geodesics*** are the curves

$$\mathscr{L}^{\text{in}}_{(v,\theta,\tilde{\varphi})} : \quad (v, \theta, \tilde{\varphi}) = \text{const} \in \mathbb{R} \times [0, \pi] \times [0, 2\pi), \tag{13.14}$$

  where v is the Kerr advanced time:

$$v := \tilde{t} + r = t + r - \frac{2m^2}{r-m} + 2m \ln\left|\frac{r-m}{m}\right|. \tag{13.15}$$

  Along the geodesic $\mathscr{L}^{\text{in}}_{(v,\theta,\tilde{\varphi})}$, $-r$ is an affine parameter increasing towards the future; the corresponding tangent vector is

$$\boldsymbol{k} = \boldsymbol{\partial}_{\tilde{t}} - \boldsymbol{\partial}_{\tilde{r}} \tag{13.16a}$$

$$\boldsymbol{k} = \frac{r^2+m^2}{(r-m)^2}\boldsymbol{\partial}_t - \boldsymbol{\partial}_r + \frac{m}{(r-m)^2}\boldsymbol{\partial}_\varphi \quad \text{in } \mathscr{M} \setminus \mathscr{H}. \tag{13.16b}$$

[2] Cf. Sec. 5.2.1 for the definition of *stationary* and a discussion about the terminology.

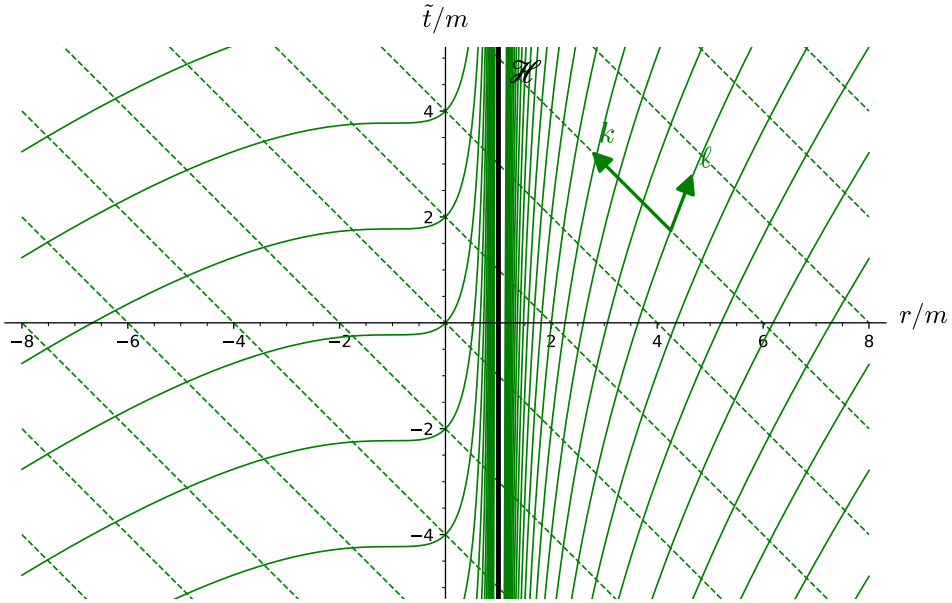


Figure 13.2: Trace of the principal null geodesics in the plane $(\tilde{t}, r)$. The dashed lines correspond to the ingoing principal null geodesics $\mathscr{L}^{\rm in}_{(v,\theta,\tilde{\varphi})}$ and the solid curves to the outgoing principal null geodesics $\mathscr{L}^{\rm out,I}_{(u,\theta,\tilde{\varphi})}$ for $r > m$ and $\mathscr{L}^{\rm out,III}_{(u,\theta,\tilde{\varphi})}$ for $r < m$. *[Figure generated by the notebook D.4.17]*

- The ***outgoing principal null geodesics*** are the curves

$$\text{in } \mathscr{M}_{\rm I}: \quad \mathscr{L}^{\rm out,I}_{(u,\theta,\tilde{\varphi})}: \quad (u,\theta,\tilde{\varphi}) = \text{const} \in \mathbb{R}\times[0,\pi]\times[0,2\pi), \tag{13.17a}$$

$$\text{in } \mathscr{M}_{\rm III}: \quad \mathscr{L}^{\rm out,III}_{(u,\theta,\tilde{\varphi})}: \quad (u,\theta,\tilde{\varphi}) = \text{const} \in \mathbb{R}\times[0,\pi]\times[0,2\pi), \tag{13.17b}$$

$$\text{on } \mathscr{H}: \quad \mathscr{L}^{\rm out,\mathscr{H}}_{(\theta,\psi)}: \quad (\theta,\psi) = \text{const} \in [0,\pi]\times[0,2\pi), \tag{13.17c}$$

where u is the Kerr retarded time:

$$u := \tilde{t} - r + \frac{4m^2}{r-m} - 4m\ln\left|\frac{r-m}{m}\right| = t - r + \frac{2m^2}{r-m} - 2m\ln\left|\frac{r-m}{m}\right|, \tag{13.18}$$

$$\tilde{\tilde{\varphi}} := \tilde{\varphi} + \frac{2m}{r-m} = \varphi + \frac{m}{r-m} \tag{13.19}$$

and

$$\psi := \tilde{\varphi} - \frac{\tilde{t}}{2m}. \tag{13.20}$$

Along the geodesics $\mathscr{L}^{\rm out,I}_{(u,\theta,\tilde{\varphi})}$ and $\mathscr{L}^{\rm out,III}_{(u,\theta,\tilde{\varphi})}$, r is an affine parameter increasing towards the future, while along $\mathscr{L}^{\rm out,\mathscr{H}}_{(\theta,\psi)}$, such an affine parameter is $\tilde{t}$. The null tangent vector $\boldsymbol{\ell}$ along the outgoing principal null geodesics that coincides with

the Killing vector $\boldsymbol{\xi} + (2m)^{-1}\boldsymbol{\eta}$ on $\mathscr{H}$ is

$$\boldsymbol{\ell} = \frac{(r+m)^2}{2(r^2+m^2)}\,\boldsymbol{\partial}_{\tilde{t}} + \frac{(r-m)^2}{2(r^2+m^2)}\,\boldsymbol{\partial}_{\tilde{r}} + \frac{m}{r^2+m^2}\,\boldsymbol{\partial}_{\tilde{\varphi}} \tag{13.21a}$$

$$\boldsymbol{\ell} = \frac{1}{2}\,\boldsymbol{\partial}_t + \frac{(r-m)^2}{2(r^2+m^2)}\,\boldsymbol{\partial}_r + \frac{m}{2(r^2+m^2)}\,\boldsymbol{\partial}_{\varphi} \quad \text{in } \mathscr{M}\setminus\mathscr{H}. \tag{13.21b}$$

Proof. The second equality in Eq. (13.15) follows from Eq. (13.8a). Eq. (13.16b) follow from Eq. (13.16a) via Eq. (13.10). Eqs. (13.18) and (13.19) are the integrated version of the system (10.52) with $a = m$. Equations (13.20), (13.21a) and (13.21b) are the $a = m$ versions of respectively Eqs. (10.60), (10.57) and (10.62). All the other statements follow from the $a \to m$ limit of results of Sec. 10.4, except for $\tilde{t}$ being an affine parameter along $\mathscr{L}^{\text{out},\mathscr{H}}_{(\theta,\psi)}$, which is peculiar to the extremal Kerr horizon and will be proven in Sec. 13.2.5. □

The principal null geodesic congruences are depicted in terms of the $(\tilde{t}, r)$ coordinates in Fig. 13.2. Note that the outgoing geodesics $\mathscr{L}^{\text{out,I}}_{(u,\theta,\tilde{\varphi})}$ and $\mathscr{L}^{\text{out,III}}_{(u,\theta,\tilde{\varphi})}$ tend to become tangent to $\mathscr{H}$ for $r \to m$; this agrees with $\mathscr{H}$ being generated by some members of the outgoing principal null congruence, namely the geodesics $\mathscr{L}^{\text{out},\mathscr{H}}_{(\theta,\psi)}$, as we shall see in details in the next subsection. Another view of the principal null geodesics is provided by the Carter-Penrose diagram of $(\mathscr{M}, \boldsymbol{g})$ shown in Fig. 13.3.

13.2.5 The degenerate horizon

$\mathscr{H}$ is the hypersurface of $\mathscr{M}$ defined by $r = m$ [Eq. (13.7b)]. Given that the component $g^{rr} = (r-m)^2/\rho^2$ of the inverse metric with respect to Kerr coordinates vanishes at $r = m$ [cf. Eq. (13.5)], we have $g^{\tilde{\mu}\tilde{\nu}}\partial_{\tilde{\mu}} r \partial_{\tilde{\nu}} r = 0$ on $\mathscr{H}$, which implies that the gradient $\vec{\boldsymbol{\nabla}} r$ is a null vector there and that $\mathscr{H}$ is a null hypersurface. Moreover, since the components of $\vec{\boldsymbol{\nabla}} r$ are $\nabla^{\tilde{\alpha}} = g^{\tilde{\alpha} r}$, we read on Eq. (13.5) that

$$\vec{\boldsymbol{\nabla}} r \overset{\mathscr{H}}{=} \frac{2m^2}{\rho^2}\boldsymbol{\partial}_{\tilde{t}} + \frac{m}{\rho^2}\boldsymbol{\partial}_{\tilde{\varphi}} \overset{\mathscr{H}}{=} \frac{2m^2}{\rho^2}\boldsymbol{\chi}, \tag{13.22}$$

where $\boldsymbol{\chi}$ is the Killing vector field

$$\boldsymbol{\chi} := \boldsymbol{\xi} + \Omega_H \boldsymbol{\eta}, \qquad \text{with} \quad \Omega_H := \frac{1}{2m}. \tag{13.23}$$

It follows immediately that $\mathscr{H}$ is a *Killing horizon*, i.e. a null hypersurface that admits a Killing vector as null normal (cf. the definition given in Sec. 3.3.2).

From expression (13.21a) for $\boldsymbol{\ell}$, we have immediately

$$\boldsymbol{\ell} \overset{\mathscr{H}}{=} \boldsymbol{\chi}. \tag{13.24}$$

This means that the null generators of $\mathscr{H}$ are the outgoing principal null geodesics $\mathscr{L}^{\text{out},\mathscr{H}}_{(\theta,\psi)}$.

The *surface gravity* κ of the Killing horizon $\mathscr{H}$ has been defined in Sec. 3.3.5 as the non-affinity coefficient of the Killing-vector normal $\boldsymbol{\chi}$ to $\mathscr{H}$: $\boldsymbol{\nabla}_{\boldsymbol{\chi}}\boldsymbol{\chi} \overset{\mathscr{H}}{=} \kappa\boldsymbol{\chi}$ [cf. Eq. (3.35)]. Given identity (13.24), κ coincides with the value on $\mathscr{H}$ of the non-affinity coefficient $\kappa_{\boldsymbol{\ell}}$ of the tangent $\boldsymbol{\ell}$ to the outgoing principal null geodesics: $\boldsymbol{\nabla}_{\boldsymbol{\ell}}\boldsymbol{\ell} = \kappa_{\boldsymbol{\ell}}\boldsymbol{\ell}$. A direct computation (cf. the notebook D.4.17) reveals that

$$\kappa_{\boldsymbol{\ell}} = m\frac{r^2 - m^2}{(r^2+m^2)^2}. \tag{13.25}$$

In particular, $\kappa_{\boldsymbol{\ell}}$ vanishes for $r = m$, i.e. on $\mathscr{H}$. Since $\kappa = \kappa_{\boldsymbol{\ell}}|_{\mathscr{H}}$, we get[3]

$$\boxed{\kappa = 0}. \tag{13.26}$$

According to the classification introduced in Sec. 3.3.6, one says that $\mathscr{H}$ is a *degenerate Killing horizon*. The vanishing of the non-affinity coefficient κ means that $\boldsymbol{\ell}$ is a geodesic vector on $\mathscr{H}$, and not only a pregeodesic one (cf. Remark 1 in Sec. 1.3.2). Equivalently, at any given point $p \in \mathscr{H}$, $\boldsymbol{\ell}$ is the tangent vector associated to an affine parameter λ of the null geodesic $\mathscr{L}^{\mathrm{out},\mathscr{H}}_{(\theta,\psi)}$ through p. Moreover, the affine parameter λ coincides with $\tilde{t}$, up to some additive constant. Indeed, Eq. (13.24) implies

$$\ell^{\tilde{t}} = \frac{\mathrm{d}\tilde{t}}{\mathrm{d}\lambda} = \chi^{\tilde{t}} = 1,$$

from which $\lambda = \tilde{t} + \text{const}$. Since the range of $\tilde{t}$ is $(-\infty, +\infty)$, we conclude that $\mathscr{L}^{\mathrm{out},\mathscr{H}}_{(\theta,\psi)}$ is a complete geodesic. This constrasts with the null generators of a non-degenerate Killing horizon, which are incomplete, as shown in Sec. 5.4.2.

Let us summarize the results obtained above:

> In the extremal Kerr spacetime, the hypersurface $\mathscr{H}$ defined by $r = m$ is a degenerate Killing horizon. Its generators are the outgoing principal null geodesics $\mathscr{L}^{\mathrm{out},\mathscr{H}}_{(\theta,\psi)}$, which are complete geodesics and which admit $\tilde{t}$ as an affine parameter. The tangent vector associated to this affine parameter is the Killing vector $\boldsymbol{\chi} = \boldsymbol{\xi} + \Omega_H\boldsymbol{\eta}$ [Eq. (13.23)], which coincides on $\mathscr{H}$ with the tangent vector $\boldsymbol{\ell}$ to the outgoing principal null congruence.

13.2.6 Black hole character

As a Killing horizon, $\mathscr{H}$ is a null hypersurface and thus a one-way membrane. Since the ingoing principal null geodesics $\mathscr{L}^{\mathrm{in}}_{(v,\theta,\tilde{\varphi})}$ cross it from $\mathscr{M}_{\mathrm{I}}$ to $\mathscr{M}_{\mathrm{III}}$ (cf. Fig. 13.2), we conclude that no (massive or null) particle can cross $\mathscr{H}$ from $\mathscr{M}_{\mathrm{III}}$ to $\mathscr{M}_{\mathrm{I}}$. In order to show that $\mathscr{H}$ is actually a black hole event horizon, it suffices to proceed as for the $a < m$ case treated in Sec. 10.5.2. We shall not repeat the argument here (which is based on the asymptotics of Kerr spacetime being that of Schwarzschild spacetime — a property that holds for the extremal Kerr spacetime as well) and jump directly to the conclusion:

[3] The vanishing of κ can also be obtained by taking the limit $a \to m$ of expression (10.82) derived for $a < m$.

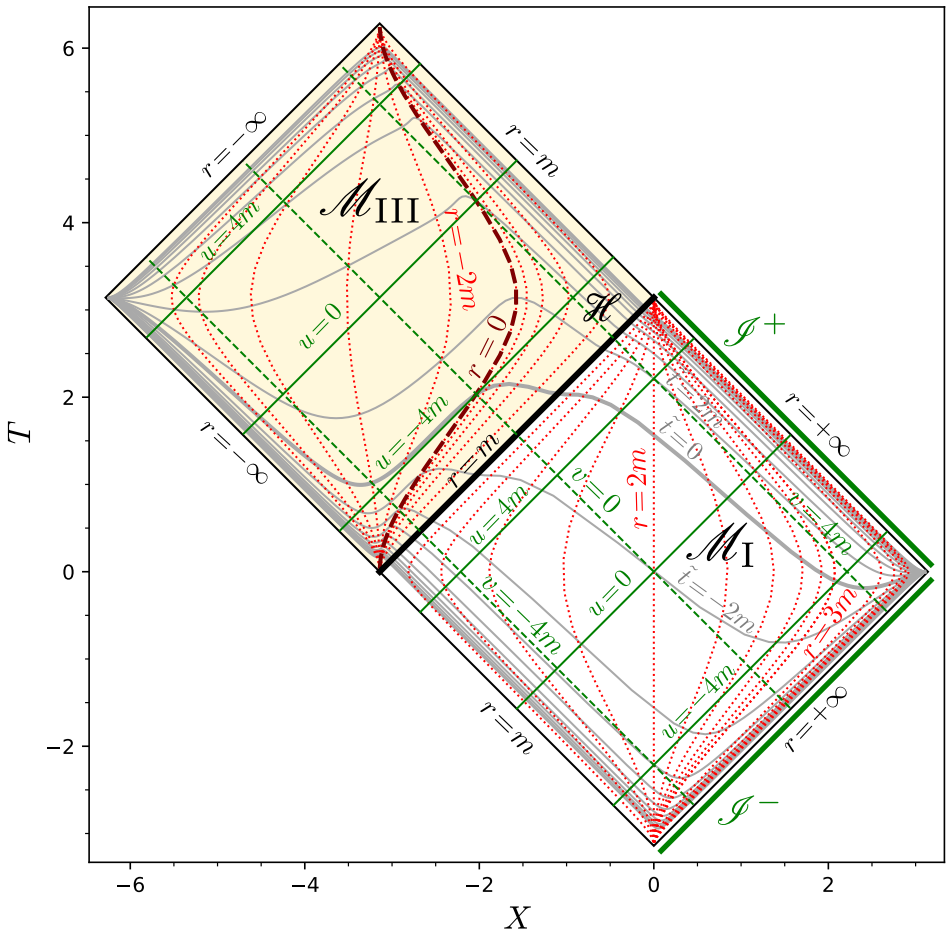


Figure 13.3: Carter-Penrose diagram of the extremal Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ constructed via the projection map $\Pi : \mathscr{M} \to \mathbb{R}^2$, $(\tilde{t}, r, \theta, \tilde{\varphi}) \mapsto (T, X)$ defined by Eqs. (13.27)-(13.28). The grey curves represent hypersurfaces $\tilde{t} = \text{const}$, with $\tilde{t} \in [-20m, 20m]$ and the increment $\delta\tilde{t} = 2m$ between two successive hypersurfaces. The hypersurface $\tilde{t} = 0$ is singled out by a larger thickness. The red dotted curves represent hypersurfaces $r = \text{const}$, with the increment δr between two successive hypersurfaces being $\delta r = 2m$ for $r < 0$ and $r > 3m$ and $\delta r = 0.2\,m$ for $0 \leq r \leq 3m$. The hypersurface $r = 0$ is marked by the brown dashed curve. The green straight lines depict some selected principal null geodesics (dashed = ingoing, solid = outgoing, as in Fig. 13.2). *[Figure generated by the notebook D.4.17]*

The extremal Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ can be endowed with a conformal completion at null infinity such that the future and past null infinities $\mathscr{I}^+$ and $\mathscr{I}^-$ are located at the boundary of $\mathscr{M}_{\rm I}$. The region $\mathscr{M}_{\rm III}$ is then the interior of a black hole, the event horizon of which is the Killing horizon $\mathscr{H}$.

The future null infinity $\mathscr{I}^+$ and the past null infinity $\mathscr{I}^-$ relative to $\mathscr{M}_{\rm I}$ are depicted in the Carter-Penrose diagram of Figs. 13.3 -13.4. In this diagram, it is clear that $\mathscr{M}_{\rm III}$ is a black hole region for $\mathscr{M}_{\rm I}$ and that $\mathscr{H}$ is the corresponding event horizon.

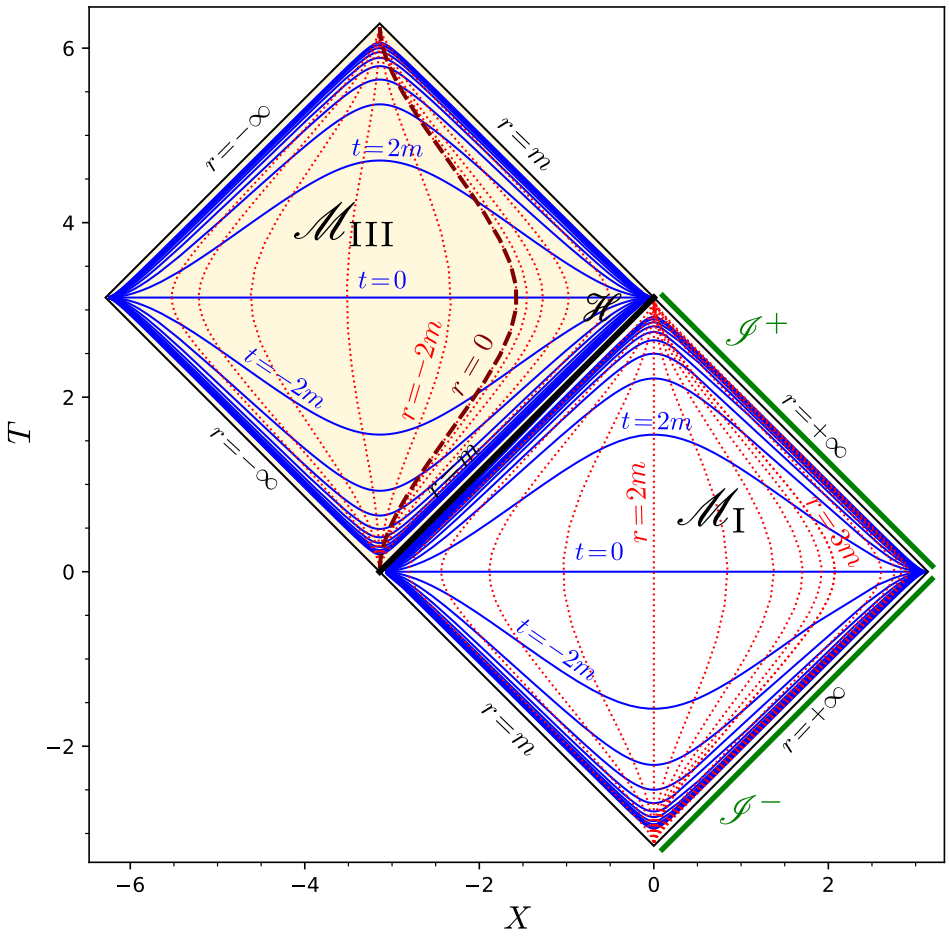


Figure 13.4: Same as Fig. 13.3 but for the time slicing associated to Boyer-Lindquist coordinates (t, r, θ, φ). The blue curves represent hypersurfaces $t = \text{const}$, with $t \in [-20m, 20m]$ and the increment $\delta t = 2m$ between two successive hypersurfaces. Note that the spacetime slicing by the hypersurfaces $t = \text{const}$ is singular at $\mathscr{H}$, contrary to the slicing by the hypersurfaces of constant Kerr time $\tilde{t}$ shown in Fig. 13.3. *[Figure generated by the notebook D.4.17]*

13.3 Maximal analytic extension

13.3.1 Extension of $\mathscr{M}_{\rm I}$ for complete outgoing principal null geodesics

Figure 13.3 depicts a Carter-Penrose diagram of the extremal Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$ built by means of the projection map[4] $\Pi : \mathscr{M} \to \mathbb{R}^2, (\tilde{t}, r, \theta, \tilde{\varphi}) \mapsto (T, X)$ defined by

$$T = T_0(u, v) \quad \text{and} \quad X = X_0(u, v) \quad \text{in } \mathscr{M}_{\rm I} \tag{13.27a}$$
$$T = \arctan(v/2) + \pi/2 \quad \text{and} \quad X = \arctan(v/2) - \pi/2 \quad \text{on } \mathscr{H} \tag{13.27b}$$
$$T = T_0(u, v) + \pi \quad \text{and} \quad X = X_0(u, v) - \pi \quad \text{in } \mathscr{M}_{\rm III} \tag{13.27c}$$

[4] cf. the definition of a Carter-Penrose diagram given in Sec. 10.8.1.

with

$$T_0(u,v) := \arctan\left(\frac{u}{2}\right) + \arctan\left(\frac{v}{2}\right) \tag{13.28a}$$

$$X_0(u,v) := \arctan\left(\frac{v}{2}\right) - \arctan\left(\frac{u}{2}\right), \tag{13.28b}$$

where u and v are the functions of $(\tilde{t}, r)$ given by Eqs. (13.18) and (13.15) respectively. Some outgoing principal null geodesics $\mathscr{L}^{\mathrm{out,I}}_{(u,\theta,\tilde{\tilde{\varphi}})}$ and $\mathscr{L}^{\mathrm{out,III}}_{(u,\theta,\tilde{\tilde{\varphi}})}$ are plotted for selected values of u (solid green lines). Let us consider the "external" region $\mathscr{M}_{\mathrm{I}}$. Since r is an affine parameter along the null geodesics $\mathscr{L}^{\mathrm{out,I}}_{(u,\theta,\tilde{\tilde{\varphi}})}$, it is clear that these geodesics are incomplete, for they all terminate in the past at the finite value $r = m$ (the South-West boundary of $\mathscr{M}_{\mathrm{I}}$ in Fig. 13.3), without any possible extension into $\mathscr{M}_{\mathrm{III}}$ from there. To extend $\mathscr{M}_{\mathrm{I}}$ so that all geodesics $\mathscr{L}^{\mathrm{out,I}}_{(u,\theta,\tilde{\tilde{\varphi}})}$ are complete, let us introduce a coordinate system on $\mathscr{M}_{\mathrm{I}}$ that is adapted to the outgoing principal null geodesics, as the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ were adapted to the ingoing ones. We thus define the ***outgoing Kerr coordinates*** $(x^{\tilde{\tilde{\alpha}}}) = (\tilde{\tilde{t}}, r, \theta, \tilde{\tilde{\varphi}})$ by

$$u = \tilde{\tilde{t}} - r \iff \tilde{\tilde{t}} = u + r, \tag{13.29}$$

where u is the retarded Kerr time (13.18) and $\tilde{\tilde{\varphi}}$ is related to the angle $\tilde{\varphi}$ of Kerr coordinates or to the angle φ of Boyer-Lindquist coordinates by Eq. (13.19). Substituting Eq. (13.18) for u into $\tilde{\tilde{t}} = u + r$ and using Eq. (13.19) linking $\tilde{\tilde{\varphi}}$ to $\tilde{\varphi}$, we get the transition map between the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ and the outgoing Kerr coordinates $(\tilde{\tilde{t}}, r, \theta, \tilde{\tilde{\varphi}})$:

$$\text{on } \mathscr{M}_{\mathrm{I}}, \quad \begin{cases} \tilde{\tilde{t}} = \tilde{t} + \dfrac{4m^2}{r-m} - 4m \ln\left(\dfrac{r-m}{m}\right) \\ \tilde{\tilde{\varphi}} = \tilde{\varphi} + \dfrac{2m}{r-m}. \end{cases} \tag{13.30}$$

By construction, the tangent vector to $\mathscr{L}^{\mathrm{out,I}}_{(u,\theta,\tilde{\tilde{\varphi}})}$ associated with the affine parameter r is then

$$\boldsymbol{\ell}' = \boldsymbol{\partial}_{\tilde{\tilde{t}}} + \boldsymbol{\partial}_{\tilde{\tilde{r}}}, \tag{13.31}$$

where $\boldsymbol{\partial}_{\tilde{\tilde{r}}}$ stands for the vector $\partial/\partial r$ of the coordinates $(\tilde{\tilde{t}}, r, \theta, \tilde{\tilde{\varphi}})$. Indeed $\ell'^{\tilde{\tilde{\alpha}}} = \mathrm{d}x^{\tilde{\tilde{\alpha}}}/\mathrm{d}r = (1, 1, 0, 0)$ since along $\mathscr{L}^{\mathrm{out,I}}_{(u,\theta,\tilde{\tilde{\varphi}})}$, $\tilde{\tilde{t}} = r + u$, with u constant, and both θ and $\tilde{\tilde{\varphi}}$ are constant. An explicit computation (cf. the notebook D.4.18) shows that $\boldsymbol{\ell}'$ is a geodesic vector:

$$\boldsymbol{\nabla}_{\boldsymbol{\ell}'} \boldsymbol{\ell}' = 0, \tag{13.32}$$

which confirms that r is an affine parameter along $\mathscr{L}^{\mathrm{out,I}}_{(u,\theta,\tilde{\tilde{\varphi}})}$. $\boldsymbol{\ell}'$ is thus similar to $\boldsymbol{k}$, which is the tangent vector to the ingoing principal null geodesics $\mathscr{L}^{\mathrm{in}}_{(v,\theta,\tilde{\varphi})}$ associated with the affine parameter $-r$ along them. In this respect, note the symmetry between the relations $\boldsymbol{\ell}' = \boldsymbol{\partial}_{\tilde{\tilde{t}}} + \boldsymbol{\partial}_{\tilde{\tilde{r}}}$ and $\boldsymbol{k} = \boldsymbol{\partial}_{\tilde{t}} - \boldsymbol{\partial}_{\tilde{r}}$ [Eq. (13.16a)]. The link between $\boldsymbol{\ell}'$ and the tangent vector $\boldsymbol{\ell}$ to $\mathscr{L}^{\mathrm{out,I}}_{(u,\theta,\tilde{\tilde{\varphi}})}$ introduced in Sec. 13.2.4 is easily obtained from the definition of a tangent vector to a curve:

$$\boldsymbol{\ell}' = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}r} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}r} = \left(\frac{\mathrm{d}r}{\mathrm{d}\lambda}\right)^{-1} \boldsymbol{\ell},$$

where λ is the (non-affine) parameter of $\mathscr{L}^{\mathrm{out,I}}_{(u,\theta,\tilde{\tilde{\varphi}})}$ associated with $\boldsymbol{\ell}$. We read on the Kerr components (13.21a) of $\boldsymbol{\ell}$, as well as on the Boyer-Lindquist ones (13.21b), that $\mathrm{d}r/\mathrm{d}\lambda = \ell^r = (r-m)^2/(2(r^2+m^2))$. Hence

$$\boldsymbol{\ell}' = 2\frac{r^2+m^2}{(r-m)^2}\,\boldsymbol{\ell}. \tag{13.33}$$

Differentiating Eq. (13.30) leads to

$$\mathrm{d}\tilde{\tilde{t}} = \mathrm{d}\tilde{t} - \frac{4mr}{(r-m)^2}\mathrm{d}r \qquad \text{and} \qquad \mathrm{d}\tilde{\tilde{\varphi}} = \mathrm{d}\tilde{\varphi} - \frac{2m}{(r-m)^2}\mathrm{d}r. \tag{13.34}$$

From these relations and the chain rule, we get immediately the link between the outgoing Kerr coordinate frame and the Kerr coordinate frame:

$$\boldsymbol{\partial}_{\tilde{\tilde{t}}} = \boldsymbol{\partial}_{\tilde{t}}, \qquad \boldsymbol{\partial}_{\tilde{\tilde{r}}} = \boldsymbol{\partial}_{\tilde{r}} + \frac{4mr}{(r-m)^2}\boldsymbol{\partial}_{\tilde{t}} + \frac{2m}{(r-m)^2}\boldsymbol{\partial}_{\tilde{\varphi}}, \qquad \boldsymbol{\partial}_{\theta} = \boldsymbol{\partial}_{\theta}, \qquad \boldsymbol{\partial}_{\tilde{\tilde{\varphi}}} = \boldsymbol{\partial}_{\tilde{\varphi}}. \tag{13.35}$$

In view of Eq. (13.13), we conclude that $\boldsymbol{\partial}_{\tilde{\tilde{t}}}$ and $\boldsymbol{\partial}_{\tilde{\tilde{\varphi}}}$ coincide with the Killing vectors $\boldsymbol{\xi}$ and $\boldsymbol{\eta}$ of Kerr metric:

$$\boldsymbol{\partial}_{\tilde{\tilde{t}}} = \boldsymbol{\xi} \qquad \text{and} \qquad \boldsymbol{\partial}_{\tilde{\tilde{\varphi}}} = \boldsymbol{\eta}. \tag{13.36}$$

The link between the outgoing Kerr coordinates and the Boyer-Lindquist ones is obtained by substituting Eq. (13.18) for u into $\tilde{\tilde{t}} = u + r$ [Eq. (13.29)]:

$$\tilde{\tilde{t}} = t + \frac{2m^2}{r-m} - 2m\ln\left|\frac{r-m}{m}\right|. \tag{13.37}$$

This relation is to be supplemented by Eq. (13.19) to fully specify the transformation from the Boyer-Lindquist coordinates (t, r, θ, φ) to the outgoing Kerr coordinates $(\tilde{\tilde{t}}, r, \theta, \tilde{\tilde{\varphi}})$. Differentiating Eqs. (13.37) and (13.19) leads to

$$\mathrm{d}\tilde{\tilde{t}} = \mathrm{d}t - \frac{2mr}{(r-m)^2}\mathrm{d}r \qquad \text{and} \qquad \mathrm{d}\tilde{\tilde{\varphi}} = \mathrm{d}\varphi - \frac{m}{(r-m)^2}\mathrm{d}r \tag{13.38}$$

Remark 1: Equation (13.38) differs from Eq. (13.9) only by the sign + changed to − in the right-hand side. This reflects the complete symmetry between the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ and the outgoing Kerr coordinates $(\tilde{\tilde{t}}, r, \theta, \tilde{\tilde{\varphi}})$ from the point of view of the Boyer-Lindquist coordinates (t, r, θ, φ) (cf. the discussion at the beginning of Sec. 10.4.2).

The components of the metric tensor $\boldsymbol{g}$ with respect to the outgoing Kerr coordinates are easily obtained by substituting $\mathrm{d}t$ and $\mathrm{d}\varphi$ from Eq. (13.38) into the Boyer-Lindquist expression (13.11) (see also the notebook D.4.18); we get

$$\begin{aligned} g_{\tilde{\tilde{\mu}}\tilde{\tilde{\nu}}}\,\mathrm{d}x^{\tilde{\tilde{\mu}}}\,\mathrm{d}x^{\tilde{\tilde{\nu}}} = & -\left(1-\frac{2mr}{\rho^2}\right)\mathrm{d}\tilde{\tilde{t}}^2 - \frac{4mr}{\rho^2}\mathrm{d}\tilde{\tilde{t}}\,\mathrm{d}r - \frac{4m^2 r\sin^2\theta}{\rho^2}\,\mathrm{d}\tilde{\tilde{t}}\,\mathrm{d}\tilde{\tilde{\varphi}} \\ & + \left(1+\frac{2mr}{\rho^2}\right)\mathrm{d}r^2 + 2m\left(1+\frac{2mr}{\rho^2}\right)\sin^2\theta\,\mathrm{d}r\,\mathrm{d}\tilde{\tilde{\varphi}} \\ & + \rho^2\mathrm{d}\theta^2 + \left(r^2+m^2+\frac{2m^3 r\sin^2\theta}{\rho^2}\right)\sin^2\theta\,\mathrm{d}\tilde{\tilde{\varphi}}^2. \end{aligned} \tag{13.39}$$

These metric components are very similar to those in Kerr coordinates, as given by Eq. (13.3): the only differences are $g_{\tilde{t}r}$ and $g_{r\tilde{\varphi}}$, which have a sign opposite to that of respectively $g_{\tilde{t}r}$ and $g_{r\tilde{\varphi}}$. Apart from the standard singularities of the spherical coordinates $(\theta, \tilde{\tilde{\varphi}})$ on the axis $\theta \in \{0, \pi\}$, the only singularity of the metric components (13.39) would occur at $\rho = 0$, which does not happen in $\mathscr{M}_{\rm I}$. In particular there is no divergence for $r \to m$. This can be used to extend smoothly the spacetime $(\mathscr{M}_{\rm I}, \boldsymbol{g})$ to $r \in (-\infty, m]$, so that the outgoing principal null geodesics $\mathscr{L}^{\rm out,I}_{(u,\theta,\tilde{\tilde{\varphi}})}$ with $\theta \neq \pi/2$ become complete. But the extension to $r < m$ cannot be $\mathscr{M}_{\rm III}$ as it appears clearly on Fig. 13.3 that the end point of $\mathscr{L}^{\rm out,I}_{(u,\theta,\tilde{\tilde{\varphi}})}$ for $r \to m^+$ is not located at the boundary between $\mathscr{M}_{\rm I}$ and $\mathscr{M}_{\rm III}$. We thus introduce a spacetime $(\mathscr{M}', \boldsymbol{g})$ from a new copy of $\mathbb{R}^2 \times \mathbb{S}^2$ with $\mathbb{R}^2$ spanned by the coordinates $(\tilde{\tilde{t}}, r)$ and $\mathbb{S}^2$ spanned by the coordinates $(\theta, \tilde{\tilde{\varphi}})$, such that (i) the manifold $\mathscr{M}'$ is

$$\mathscr{M}' := \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathscr{R}' \quad \text{where} \quad \mathscr{R}' := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = 0 \text{ and } \theta(p) = \frac{\pi}{2} \right\}, \tag{13.40}$$

(ii) $\mathscr{M}_{\rm I}$ is identified with the part $r > m$ of $\mathscr{M}'$ and (iii) in all $\mathscr{M}'$, $\boldsymbol{g}$ has the components given by expression (13.39). Furthermore, we define

$$\mathscr{H}' := \{ p \in \mathscr{M}', \quad r(p) = m \} \qquad \text{and} \qquad \mathscr{M}'_{\rm III} := \{ p \in \mathscr{M}', \quad r(p) < m \}. \tag{13.41}$$

We have then $\mathscr{M}_{\rm I} = \mathscr{M} \cap \mathscr{M}'$. In $\mathscr{M}'_{\rm III}$, one can define Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ from $(\tilde{\tilde{t}}, r, \varphi, \tilde{\tilde{\varphi}})$ via formulas (13.18) (with $u = \tilde{\tilde{t}} - r$) and (13.19). It appears then immediately that $(\mathscr{M}'_{\rm III}, \boldsymbol{g})$ is isometric to $(\mathscr{M}_{\rm III}, \boldsymbol{g})$. It follows that $\boldsymbol{g}$ obeys the vacuum Einstein equation in all $\mathscr{M}'$. The vector field $\boldsymbol{\ell}' := \boldsymbol{\partial}_{\tilde{\tilde{t}}} + \boldsymbol{\partial}_{\tilde{\tilde{r}}}$ [cf. Eq. (13.31)] is a smooth non-vanishing null vector field on $\mathscr{M}'$. Since it is future directed in $(\mathscr{M}_{\rm I}, \boldsymbol{g})$ considered as a part of $(\mathscr{M}, \boldsymbol{g})$, we use it to set the time orientation in all $\mathscr{M}'$.

The tangent vector $\boldsymbol{k}$ to ingoing principal null geodesics has the following components with respect to the outgoing Kerr coordinates:

$$\boldsymbol{k} = \frac{(r+m)^2}{(r-m)^2} \boldsymbol{\partial}_{\tilde{\tilde{t}}} - \boldsymbol{\partial}_{\tilde{\tilde{r}}} + \frac{2m}{(r-m)^2} \boldsymbol{\partial}_{\tilde{\tilde{\varphi}}}. \tag{13.42}$$

This follows immediately from $\boldsymbol{k} = \boldsymbol{\partial}_{\tilde{t}} - \boldsymbol{\partial}_{\tilde{r}}$ [Eq. (13.16a)] and using Eq. (13.35) to substitute $\boldsymbol{\partial}_{\tilde{t}}$ and $\boldsymbol{\partial}_{\tilde{r}}$. Extending Equation (13.42) to all $\mathscr{M}'$ leads to a vector field that is singular on $\mathscr{H}'$. To get a vector field everywhere regular on $\mathscr{M}'$, we rescale it by the inverse of the factor relating $\boldsymbol{\ell}$ to $\boldsymbol{\ell}'$ in Eq. (13.33), i.e. we define

$$\boldsymbol{k}' := \frac{(r-m)^2}{2(r^2+m^2)} \boldsymbol{k}. \tag{13.43}$$

Hence

$$\boldsymbol{k}' = \frac{(r+m)^2}{2(r^2+m^2)} \boldsymbol{\partial}_{\tilde{\tilde{t}}} - \frac{(r-m)^2}{2(r^2+m^2)} \boldsymbol{\partial}_{\tilde{\tilde{r}}} + \frac{m}{r^2+m^2} \boldsymbol{\partial}_{\tilde{\tilde{\varphi}}}. \tag{13.44}$$

This vector field is clearly regular in all $\mathscr{M}'$. Accordingly, it can be used to extend smoothly the family of ingoing principal null geodesics to $\mathscr{H}'$. The price to pay is that $\boldsymbol{k}'$

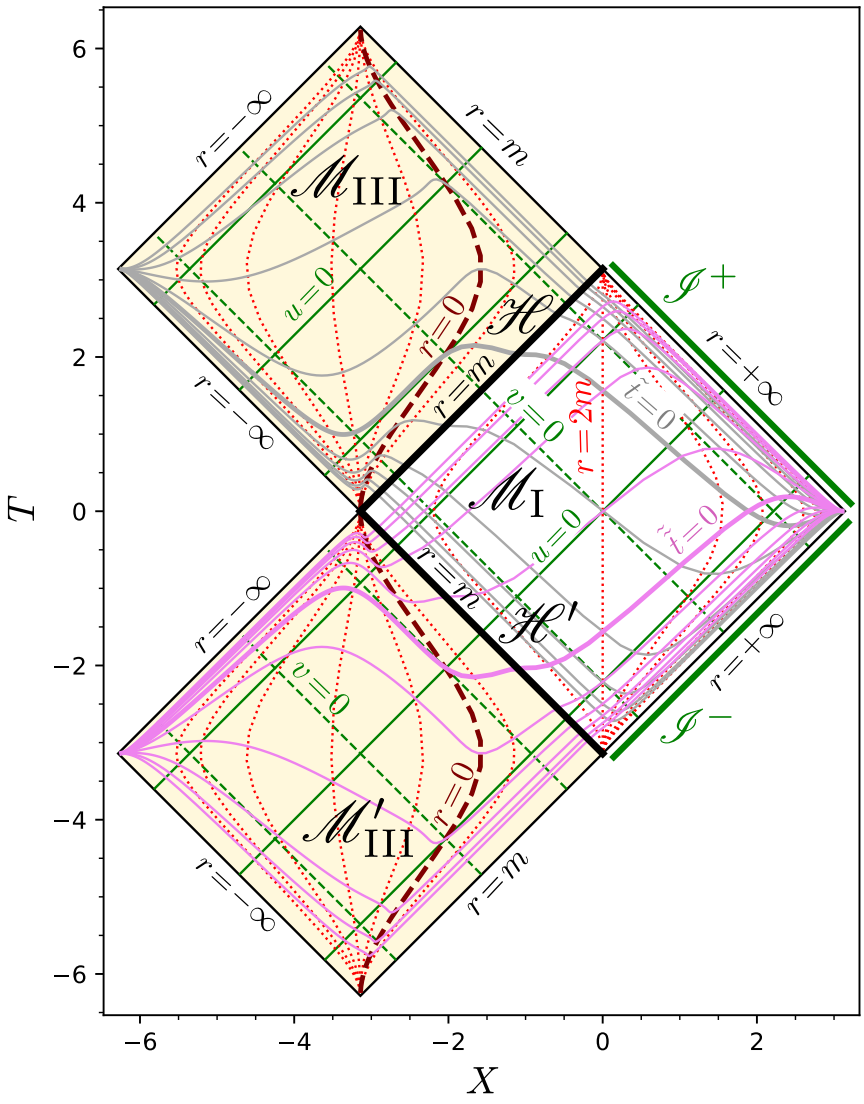


Figure 13.5: Carter-Penrose diagram of the (partially) extended extremal Kerr spacetime $(\mathscr{M}_0, \boldsymbol{g})$. The grey curves represent hypersurfaces $\tilde{t} = \text{const}$ in $\mathscr{M}$, with $\tilde{t} \in [-10m, 10m]$ and the increment $\delta\tilde{t} = 2m$ between two successive hypersurfaces. The hypersurface $\tilde{t} = 0$ is singled out by a larger thickness. The purple curves represent hypersurfaces $\tilde{\tilde{t}} = \text{const}$ in $\mathscr{M}'$, with $\tilde{\tilde{t}} \in [-10m, 10m]$ and the increment $\delta\tilde{\tilde{t}} = 2m$ between two successive hypersurfaces. The hypersurface $\tilde{\tilde{t}} = 0$ is singled out by a larger thickness. The red dotted curves represent hypersurfaces $r = \text{const}$, with the increment δr between two successive hypersurfaces being $\delta r = 2m$ for $r < 0$ and $r > 3m$ and $\delta r = 0.5\,m$ for $0 \leq r \leq 3m$. The hypersurfaces $r = 0$ are marked by brown dashed curves. The dashed (resp. solid) green straight lines depict some selected ingoing (resp. outgoing) principal null geodesics. *[Figure generated by the notebook D.4.18]*

is only a pregeodesic vector field, while $\boldsymbol{k}$ was geodesic, being associated with the affine parameter $-r$. The pair $(\boldsymbol{k}, \boldsymbol{k}')$ plays actually the same role as the pair $(\boldsymbol{\ell}', \boldsymbol{\ell})$ (note the order!) regarding the outgoing principal null geodesics.

As $\mathscr{H}$ in $(\mathscr{M}, \boldsymbol{g})$, $\mathscr{H}'$ is a degenerate Killing horizon of $(\mathscr{M}', \boldsymbol{g})$. Indeed, by the same reasoning as in Sec. 13.2.5, we get that the null normal to $\mathscr{H}'$ is the Killing vector $\boldsymbol{\chi} = \boldsymbol{\xi} + 1/(2m)\,\boldsymbol{\eta}$ [Eq. (13.23) extended to $\mathscr{M}'$]. This normal coincides with $\boldsymbol{k}'$ on $\mathscr{H}'$, as we can see by setting $r = m$ in Eq. (13.44). This implies that the null geodesic generators of $\mathscr{H}'$ belong to the ingoing principal null congruence. The non-affinity coefficient of $\boldsymbol{k}'$ is (cf. the notebook D.4.18):

$$\kappa_{\boldsymbol{k}'} = m\frac{m^2 - r^2}{(r^2 + m^2)^2}. \tag{13.45}$$

We have thus $\kappa_{\boldsymbol{k}'} = 0$ on $\mathscr{H}'$, so that $\mathscr{H}'$ is a degenerate Killing horizon.

With $\mathscr{M}'$, our extended extremal Kerr spacetime is thus $(\mathscr{M}_0, \boldsymbol{g})$ with

$$\boxed{\mathscr{M}_0 := \mathscr{M} \cup \mathscr{M}' = \mathscr{M}_{\mathrm{I}} \cup \mathscr{M}_{\mathrm{III}} \cup \mathscr{M}'_{\mathrm{III}} \cup \mathscr{H} \cup \mathscr{H}'}. \tag{13.46}$$

A Carter-Penrose diagram of $\mathscr{M}_0$ is shown in Fig. 13.5. The projection operator used to build this diagram (cf. Sec. 10.8.1) is $\Pi : \mathscr{M}_0 \to \mathbb{R}^2$, with $\mathbb{R}^2$ spanned by coordinates (T, X), such that Π is defined by Eqs. (13.27)-(13.28) on $\mathscr{M}$, while on $\mathscr{M}'$, Π is defined by

$$T = T_0(u, v) \quad \text{and} \quad X = X_0(u, v) \quad \text{in } \mathscr{M}_{\mathrm{I}} \tag{13.47a}$$

$$T = \arctan(u/2) - \pi/2 \quad \text{and} \quad X = -\arctan(u/2) - \pi/2 \quad \text{on } \mathscr{H}' \tag{13.47b}$$

$$T = T_0(u, v) - \pi \quad \text{and} \quad X = X_0(u, v) - \pi \quad \text{in } \mathscr{M}'_{\mathrm{III}} \tag{13.47c}$$

where $T_0(u, v)$ and $X_0(u, v)$ are given by Eq. (13.28) and u and v are functions of $(\tilde{\tilde{t}}, r)$ defined by respectively $u = \tilde{\tilde{t}} - r$ [Eq. (13.29)] and

$$v = \tilde{\tilde{t}} + r - \frac{4m^2}{r - m} + 4m \ln \left| \frac{r - m}{m} \right|. \tag{13.48}$$

The last equation is obtained by combining Eqs. (13.15) and (13.18).

Remark 2: As we have constructed it, the manifold $\mathscr{M}_0$ is covered by two coordinate charts: $(\tilde{t}, r, \theta, \tilde{\varphi})$ on $\mathscr{M}_{\mathrm{I}} \cup \mathscr{H} \cup \mathscr{M}_{\mathrm{III}}$ and $(\tilde{\tilde{t}}, r, \theta, \tilde{\tilde{\varphi}})$ on $\mathscr{M}_{\mathrm{I}} \cup \mathscr{H}' \cup \mathscr{M}'_{\mathrm{III}}$. These two charts overlap in $\mathscr{M}_{\mathrm{I}}$, where the transition between them is provided by Eq. (13.30). The Carter-Penrose diagram of Fig. 13.5 might give the impression that, by means of the coordinates (T, X), one could cover $\mathscr{M}_0$ by a single chart, in a fashion similar to the covering of the entire maximal extension of Schwarzschild manifold by the Kruskal-Szekeres coordinates (T, X, θ, φ) (cf. Sec. 9.3). However, this is not possible in a simple way, due to singularity issues with the azimuthal coordinates: $\tilde{\varphi}$ diverges on $\mathscr{H}'$, $\tilde{\tilde{\varphi}}$ diverges on $\mathscr{H}$ and the Boyer-Lindquist coordinate φ diverges on both $\mathscr{H}$ and $\mathscr{H}'$. Accordingly, none of $(T, X, \theta, \tilde{\varphi})$, $(T, X, \theta, \tilde{\tilde{\varphi}})$ and (T, X, θ, φ) would provide a regular chart of $\mathscr{M}_0$.

Remark 3: The spacetime $(\mathscr{M}_0, \boldsymbol{g})$ is analytic: (i) $\mathscr{M}_0$ is an analytic manifold, given that the change of coordinates (13.30) between the two charts covering $\mathscr{M}_0$ is analytic (cf. Remark 3

in Sec. A.2.1) and (ii) the components (13.3) and (13.39) of $\boldsymbol{g}$ are analytic functions of the coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ and $(\tilde{\tilde{t}}, r, \theta, \tilde{\tilde{\varphi}})$ respectively.

The Killing horizon $\mathscr{H}'$ is actually a white hole horizon from the point of view of $\mathscr{M}_{\rm I}$. More precisely, we can endow $(\mathscr{M}_0, \boldsymbol{g})$ with the same conformal completion at null infinity as that used for $(\mathscr{M}, \boldsymbol{g})$ in Sec. 13.2.6, i.e. such that the conformal boundary $\mathscr{I}$ is constituted by the same future and past null infinities $\mathscr{I}^+$ and $\mathscr{I}^-$ located at the boundary of $\mathscr{M}_{\rm I}$ (cf. Fig. 13.5). It appears then that

$$\mathscr{M}'_{\rm III} \cup \mathscr{H}' = \mathscr{M}_0 \setminus (J^+(\mathscr{I}^-) \cap \mathscr{M}_0) \tag{13.49}$$

In view of the definition (4.39), we conclude that

> $\mathscr{M}'_{\rm III}$ is the interior of a white hole region, the boundary of which is $\mathscr{H}'$.

Since $\mathscr{M}_{\rm III}$ was shown in Sec. 13.2.6 to be the black hole region for the same conformal completion at null infinity, we may state, according to the terminology introduced in Sec. 4.4.2:

> $\mathscr{M}_{\rm I}$ is the domain of outer communications of the spacetime $(\mathscr{M}_0, \boldsymbol{g})$.

13.3.2 Construction of the maximal analytic extension

With $(\mathscr{M}_0, \boldsymbol{g})$, we have acheived our first goal: all the outgoing principal null geodesics crossing $\mathscr{M}_{\rm I}$ and not lying in the equatorial plane (i.e. the geodesics extending the $\mathscr{L}^{\rm out,I}_{(u,\theta,\tilde{\varphi})}$ family with $\theta \neq \pi/2$ to the past) are complete. However, there remains incomplete geodesics in $\mathscr{M}_0$: the outgoing principal null geodesics crossing $\mathscr{M}_{\rm III}$ all stop at the value $r = m$ of their affine parameter, while the ingoing principal null geodesics crossing $\mathscr{M}'_{\rm III}$ all start at the value $-r = -m$ of their affine parameter (cf. Fig. 13.5). To construct a spacetime with complete geodesics, except for those that encouter the curvature singularity at $r = 0$ and $\theta = \pi/2$, one introduces an infinite number of copies of $\mathscr{M}_0$, $(\mathscr{M}_n)_{n\in\mathbb{Z}}$ say, and identify the region $\mathscr{M}'_{\rm III}$ of $\mathscr{M}_n$ with the region $\mathscr{M}_{\rm III}$ of $\mathscr{M}_{n-1}$ for all $n \in \mathbb{Z}$. The manifold hence obtained,

$$\mathscr{M}_* := \bigcup_{n\in\mathbb{Z}} \mathscr{M}_n, \tag{13.50}$$

is depicted via a Carter-Penrose diagram in Fig. 13.6. The region $\mathscr{M}_{\rm I}$ (resp. $\mathscr{M}_{\rm III}$) of $\mathscr{M}_n$ is denoted by $\mathscr{M}_{\rm I}^{(n)}$ (resp. $\mathscr{M}_{\rm III}^{(n)}$). Similarly, the Killing horizon $\mathscr{H}$ (resp. $\mathscr{H}'$) of $\mathscr{M}_n$ is denoted by $\mathscr{H}^+_{(n)}$ (resp. $\mathscr{H}^-_{(n)}$). Note that $\mathscr{H}^+_{(n)}$ is a future event horizon (black hole horizon), while $\mathscr{H}^-_{(n)}$ is a past event horizon (white hole horizon), for the conformal completion at null infinity with the future and past null infinities $\mathscr{I}^+_{(n)}$ and $\mathscr{I}^-_{(n)}$ as copies of $\mathscr{I}^+$ and $\mathscr{I}^-$ introduced for $\mathscr{M}_0$.

It is clear that $(\mathscr{M}_*, \boldsymbol{g})$ is an analytic spacetime, since $(\mathscr{M}_0, \boldsymbol{g})$ is (cf. Remark 3 on p. 453).

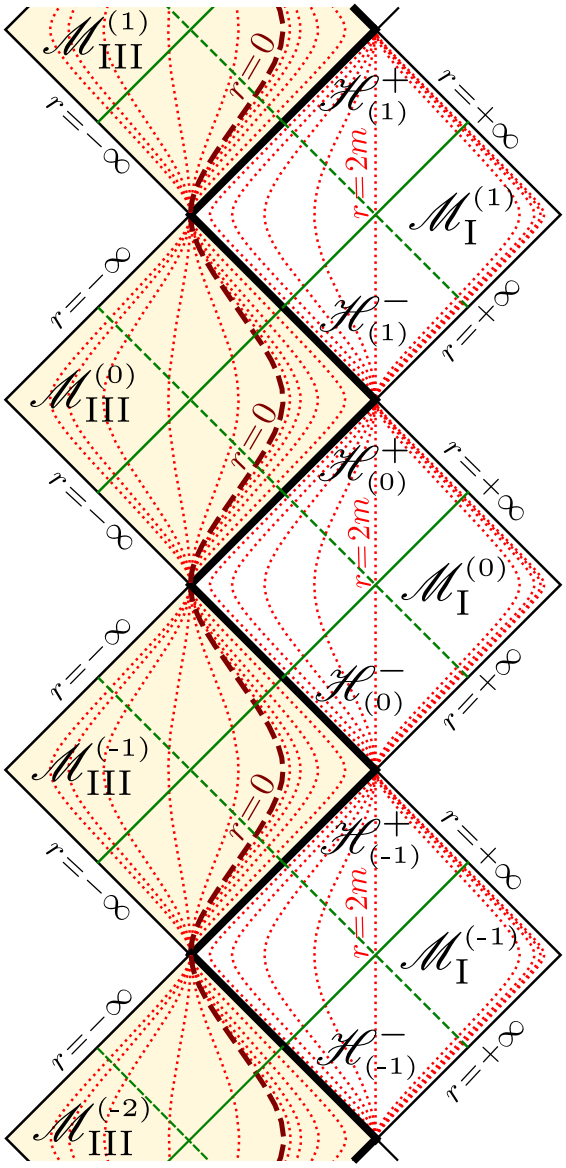


Figure 13.6: Carter-Penrose diagram of the maximal analytic extension $(\mathscr{M}_*, \boldsymbol{g})$ of the extremal Kerr spacetime. The red dotted curves represent hypersurfaces $r = \text{const}$, with the increment δr between two successive hypersurfaces being $\delta r = 2m$ for $r < 0$ and $r > 2m$ and $\delta r = 0.2\,m$ for $0 \leq r \leq 2m$. The hypersurfaces $r = 0$ are marked by brown dashed curves. The dashed (resp. solid) green straight lines depict ingoing (resp. outgoing) principal null geodesics with $v = 0$ (resp. $u = 0$). *[Figure generated by the notebook D.4.18]*

By construction, all principal null geodesics are complete in $\mathscr{M}_*$, except for those that encounter the curvature singularity at $r = 0$ and $\theta = \pi/2$. In particular, this holds for the outgoing principal null geodesics generating $\mathscr{H}^+_{(n)}$ and for the ingoing ones generating $\mathscr{H}^-_{(n)}$. Contemplating the Carter-Penrose diagram of Fig. 13.6, we might thus perceive each (solid or dashed) green straight line from an $r = -\infty$ end to an $r = +\infty$ end, as well as each thick black straight line, as representating a complete principal null geodesic.

It can be shown that actually *all* timelike or null geodesics of $(\mathscr{M}_*, \boldsymbol{g})$, and not only the principal null ones, are complete (see Carter's article [41] for the proof), so that we can conclude:

$(\mathscr{M}_*, \boldsymbol{g})$ is the maximal analytic extension of the extremal Kerr spacetime.

Remark 4: The construction of the maximal extension of the extremal Kerr spacetime is simpler than that of the Kerr spacetime with $a < m$, as presented in Sec. 10.8. Indeed, for the latter, if one uses only the ingoing and outgoing Kerr coordinate patches $(\tilde{t}, r, \theta, \tilde{\varphi})$ and $(\tilde{\tilde{t}}, r, \theta, \tilde{\tilde{\varphi}})$, one ends up with a manifold that is not maximal: the null geodesics generating the Killing horizons at $r = r_\pm$ are not complete, because the bifurcation spheres (the central dots in Fig. 10.12), accross which these geodesics can be extended, are missing. Indeed the bifurcation spheres are not covered by Kerr coordinates. To include them and thus get the full bifurcate Killing horizons (cf. Sec. 5.4), one has to introduce Kruskal-Szekeres-type coordinates in the vicinity of each bifurcation sphere, in a way similar to the Schwarzschild case treated in Secs. 9.2 and 9.3. In the present case, each Killing horizon $\mathscr{H}^+_{(n)}$ or $\mathscr{H}^-_{(n)}$ is degenerate and therefore made of complete null geodesics. In particular, $\mathscr{H}^+_{(n)}$ and $\mathscr{H}^-_{(n)}$ are not part of a bifurcate Killing horizon. In other words, there are no bifurcation spheres in the maximal extension of the extremal Kerr spacetime and the ingoing and outgoing Kerr coordinate patches are sufficient to cover it entirely.

Historical note: In 1966, Brandon Carter [39] obtained the maximal analytic extension of the rotation axis $\mathscr{A}$ of the extremal Kerr spacetime (cf. historical note on p. 305). In particular, he drew a diagram (Fig. 1b of Ref. [39]) similar to that of Fig. 13.6, the difference being that Carter's one is a true conformal representation, owning to the fact that $\mathscr{A}$ is 2-dimensional, while the diagram of Fig. 13.6 is a mere projection of the 4-dimensional manifold $\mathscr{M}_*$. In their 1967 classical study entitled *Maximal Analytic Extension of the Kerr metric* [30], Robert H. Boyer and Richard W. Lindquist focussed on the case $a < m$. For $a = m$, they refered to Carter's study [39] and stated simply that *although his* [Carter's] *work was confined to the symmetry axis ($\theta = 0, \pi$), it is clear that his conclusions apply with equal force to the full metric.* The detailed construction of the maximal analytic extension of the full extremal Kerr spacetime was actually presented by Carter himself in 1968 [41], as the special case of vanishing electric charge of the extremal Kerr-Newman spacetime. The construction was also performed in details in his famous lectures at Les Houches Summer School in 1972 [45].

13.4 Near-horizon extremal Kerr metric

13.4.1 The extremal Kerr throat

Chapter 14

Black hole formation 1: dust collapse

Contents

14.1 Introduction

After having investigated black holes in equilibrium, in the form of Schwarzschild and Kerr solutions, we turn to dynamical black hole, more specifically to the standard process of black hole formation: *gravitational collapse*. To deal with analytical solutions, we simplify the problem as much as possible. First we assume spherical symmetry, which is quite natural as a first approximation for modelling the gravitational collapse of a stellar core or a gas cloud. A drawback is that this forbids the study of gravitational waves, since by virtue of Birkhoff's theorem (to be proven in this Chapter) the exterior of any spherically symmetric collapsing object is a piece of Schwarzschild spacetime, i.e. does not contain any gravitational radiation. The second major approximation is to consider *pressureless matter*, commonly referred to as *dust*. An alternative, certainly more academic, is to consider the collapse of shell of pure electromagnetic radiation; this will be considered in Chap. 15.

14.2 Lemaître-Tolman equations

14.2.1 Hypotheses

As mentioned in the Introduction, we shall restrict ourselves to spherically symmetric[1] spacetimes, and for concretness, to 4-dimensional ones. The most general spherically

[1] See Sec. 6.2.2 for a precise definition of *spherically symmetric*.

symmetric 4-dimensional spacetime $(\mathscr{M}, \boldsymbol{g})$ can be described in terms of coordinates $x^\alpha = (\tau, \chi, \theta, \varphi)$ such that the components of the metric tensor are written

$$g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\mathrm{d}\tau^2 + a(\tau,\chi)^2\mathrm{d}\chi^2 + r(\tau,\chi)^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right), \tag{14.1}$$

where $a(\tau,\chi)$ and $r(\tau,\chi)$ are generic positive functions. These coordinates are called ***Lemaître synchronous coordinates***, the qualifier ***synchronous*** meaning that τ is the proper time of a observer staying at fixed value of the spatial coordinates (χ,θ,φ). Note that the function $r(\tau,\chi)$ gives the *areal radius* of the 2-spheres defined by $(\tau,\chi) = \text{const}$, which are the orbits of the SO(3) group action (cf. Sec. 6.2.2), i.e. the metric area of these 2-spheres is $4\pi r(\tau,\chi)^2$.

For simplification, we consider only a pressureless matter, in the form of a perfect fluid of 4-velocity $\boldsymbol{u}$ with zero pressure. The matter energy-momentum tensor is then

$$\boldsymbol{T} = \rho\boldsymbol{u}\otimes\boldsymbol{u}, \tag{14.2}$$

where the scalar field ρ can be interpreted as the fluid energy density measured in the fluid frame. Let us recall that the energy-mometum tensor of a generic perfect fluid is $\boldsymbol{T} = (\rho + p)\boldsymbol{u}\otimes\boldsymbol{u} + p\boldsymbol{g}$, where p is the fluid pressure. The expression (14.2) corresponds thus to the special case $p = 0$. Inside the matter, we link the coordinates $(\tau,\chi,\theta,\varphi)$ to the fluid by demanding that they are ***comoving*** with the fluid, i.e. that a fluid particle stays at fixed values of (χ,θ,φ). Because the 4-velocity obeys $u^\alpha = \mathrm{d}x^\alpha/\mathrm{d}\tau_{\mathrm{fl}}$, where τ_{fl} is the fluid proper time [cf. Eq. (1.14)], this amounts to set $u^\chi = u^\theta = u^\varphi = 0$, i.e. to have

$$\boldsymbol{u} = \boldsymbol{\partial}_\tau. \tag{14.3}$$

A priori, one should have only $\boldsymbol{u} = u^\tau\boldsymbol{\partial}_\tau$, but the synchronous coordinate condition $g_{\tau\tau} = -1$ along with the normalization $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{u}) = -1$ implies $u^\tau = 1$. Since $u^\tau = \mathrm{d}\tau/\mathrm{d}\tau_{\mathrm{fl}}$, we get $\tau = \tau_{\mathrm{fl}}$ (up to some additive constant), which provides the physical interpretation of Lemaître coordinate τ as the fluid proper time.

14.2.2 Geodesic matter flow

The equation of energy-momentum conservation $\boldsymbol{\nabla}\cdot\overrightarrow{\boldsymbol{T}} = 0$ [Eq. (1.40)], which follows from the Einstein equation (1.38) and the contracted Bianchi identity (A.110) (cf. Sec. 1.5), implies that

> The worldlines of the fluid particles obeying the pressureless matter model (14.2) are timelike geodesics of spacetime $(\mathscr{M}, \boldsymbol{g})$.

Proof. If we plug the energy-momentum tensor (14.2) in the energy-momentum conservation law (1.40), we obtain

$$\nabla_\mu(\rho u^\mu u^\alpha) = 0,$$

i.e.

$$\nabla_\mu(\rho u^\mu)u^\alpha + \rho u^\mu\nabla_\mu u^\alpha = 0. \tag{14.4}$$

Now the two terms in the left-hand side of this equation are orthogonal to each other, as an immediate consequence of the normalization of the 4-velocity $\boldsymbol{u}$ [Eq. (1.15)]: $\boldsymbol{u}\cdot\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0$. In particular, $\boldsymbol{u}$ is a timelike vector, while the 4-acceleration $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u}$ is a spacelike one. The only way for Eq. (14.4) to hold is thus that each term in the left-hand side vanishes separately:

$$\nabla_\mu(\rho u^\mu) = 0 \qquad \text{and} \qquad u^\mu \nabla_\mu u^\alpha = 0.$$

The second equation above is nothing but the geodesic equation [Eq. (B.1)] for the field lines of $\boldsymbol{u}$, i.e. the fluid worldlines. □

Each fluid particle is thus in free-fall and moves independently of its neighbours, which is not surprising since the pressure is zero. This justify the term ***dust*** given to the matter model (14.2).

14.2.3 From the Einstein equation to the Lemaître-Tolman system

Let us write the Einstein equation (1.38) in terms of Lemaître synchronous coordinates $(\tau, \chi, \theta, \varphi)$ and with the energy-momentum tensor (14.2)-(14.3) in its right-hand side. As detailed in Sec. D.5.1, if one disregards the peculiar case[2] $\partial r/\partial\chi = 0$, the $\tau\chi$ component yields

$$a(\tau,\chi) = \frac{1}{f(\chi)}\frac{\partial r}{\partial \chi}, \tag{14.5}$$

where $f(\chi)$ is an arbitrary function of χ. Accordingly, we may rewrite the metric components (14.1)

$$\boxed{g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\mathrm{d}\tau^2 + \frac{1}{f(\chi)^2}\left(\frac{\partial r}{\partial\chi}\right)^2 \mathrm{d}\chi^2 + r(\tau,\chi)^2\left(\mathrm{d}\theta^2 + \sin^2\theta\,\mathrm{d}\varphi^2\right)}. \tag{14.6}$$

Taking into account (14.5), the $\chi\chi$ and $\tau\tau$ components of the Einstein equation yield respectively to (cf. Sec. D.5.1)

$$\boxed{\left(\frac{\partial r}{\partial\tau}\right)^2 = f(\chi)^2 - 1 + \frac{2m(\chi)}{r(\tau,\chi)} + \frac{\Lambda}{3}r(\tau,\chi)^2} \tag{14.7a}$$

$$\boxed{\frac{\mathrm{d}m}{\mathrm{d}\chi} = 4\pi r(\tau,\chi)^2\rho(\tau,\chi)\frac{\partial r}{\partial\chi}}, \tag{14.7b}$$

where $m(\chi)$ is another arbitrary function of χ. There is no other independent component of Einstein equation. Equations (14.7) constitute the ***Lemaître-Tolman system***.

The function $m(\chi)$ is known in the literature as the ***Misner-Sharp mass*** or ***Misner-Sharp energy***, in reference of a study by Misner and Sharp in 1964 [189], despite it has

[2] For $\Lambda = 0$, this case leads to Datt solution [81].

been introduced more than 30 years earlier by Lemaître [178]. This quantity is invariantly defined for any spherically symmetric spacetime from the areal radius r:

$$m := \frac{r}{2}\left(1 - \nabla_\mu r \nabla^\mu r - \frac{\Lambda}{3} r^2\right). \tag{14.8}$$

It is easy to check that the above relation holds in the present case: we have, thanks to (14.1),

$$\nabla_\mu r \nabla^\mu r = g^{\mu\nu}\frac{\partial r}{\partial x^\mu}\frac{\partial r}{\partial x^\nu} = g^{\tau\tau}\left(\frac{\partial r}{\partial \tau}\right)^2 + g^{\chi\chi}\left(\frac{\partial r}{\partial \chi}\right)^2 = -\left(\frac{\partial r}{\partial \tau}\right)^2 + \frac{1}{a(\tau,\chi)^2}\left(\frac{\partial r}{\partial \chi}\right)^2$$

Using Eq. (14.5), this expression reduces to

$$\nabla_\mu r \nabla^\mu r = -\left(\frac{\partial r}{\partial \tau}\right)^2 + f(\chi)^2.$$

In view of the Lemaître-Tolman equation (14.7a), we conclude that (14.8) holds.

Historical note: The Lemaître-Tolman system (14.7) has been first derived in 1932 by Georges Lemaître [178]: Eqs. (14.6), (14.7a) and (14.7b) are respectively Eqs. (8.1), (8.2) and (8.3) of Ref. [178], up to some slight change of notations. It however became known as *Tolman model* or *Tolman-Bondi model*, in reference to posterior works by Richard Tolman (1934) [248] and by Hermann Bondi (1947) [28]. This happened despite Tolman fully acknowledged Lemaître's work [178] in his article [248] (Tolman actually met Lemaître in 1932-33 during the latter's trip to United States [101]) and Bondi [28] mentioned that *"Lemaître studies a problem very closely related to ours and many equations given in the appendix can be found in the* (Lemaître's) *paper"*. We refer to Eisenstaedt's article [101] for a detailed historical study of Lemaître paper [178] (see also Krasiński's note [169]).

We follow the suggestion of Plebański & Krasiński [213] to call the system (14.7) *Lemaître-Tolman*, and not merely *Lemaître*, in order to distiguish it from other Lemaître contributions to general relativity and cosmology.

14.2.4 Solutions for a vanishing cosmological constant

In the remaining of this chapter, we assume $\Lambda = 0$, since we are mainly interested in gravitational collapse in asymptotically flat spacetimes. The Lemaître-Tolman equation (14.7a) can be then rewritten as

$$\frac{1}{2}\dot{r}^2 - \frac{m(\chi)}{r} = E(\chi), \tag{14.9}$$

where $\dot{r} := \partial r/\partial\tau$ and

$$E(\chi) := \frac{f(\chi)^2 - 1}{2}. \tag{14.10}$$

For a fixed value of χ, we recognize in (14.9) the equation ruling the 1-dimensional non-relativistic motion of a particle in a Newtonian potential $V = -m/r$; $E(\chi)$ is then nothing but the total mechanical energy of the particle per unit mass. As it is well known, the solution of (14.9) depends on the sign of $E(\chi)$:

- if $E(\chi) > 0$, the solution is given in parameterized form (parameter η) by

$$\begin{cases} \tau = \dfrac{m(\chi)}{(2E(\chi))^{3/2}} (\sinh\eta - \eta) + \tau_0(\chi) \\ r(\tau,\chi) = \dfrac{m(\chi)}{2E(\chi)} (\cosh\eta - 1) \end{cases} \tag{14.11}$$

- if $E(\chi) = 0$, the solution is

$$r(\tau,\chi) = \left(\frac{9m(\chi)}{2}(\tau - \tau_0(\chi))^2\right)^{1/3} \tag{14.12}$$

- if $E(\chi) < 0$, the solution is given in parameterized form (parameter η) by

$$\begin{cases} \tau = \dfrac{m(\chi)}{|2E(\chi)|^{3/2}} (\eta + \sin\eta) + \tau_0(\chi) \\ r(\tau,\chi) = \dfrac{m(\chi)}{|2E(\chi)|} (1 + \cos\eta) \end{cases} \tag{14.13}$$

In the above formulas, $\tau_0(\chi)$ is an arbitrary function of χ. For $E > 0$ and $E = 0$, it sets the value of τ for which $r = 0$, while for $E < 0$, it sets the value of τ for which r takes its maximal value ($m/|E|$).
Exercise: prove that each of formulas (14.11)-(14.13) provides a solution of Eq. (14.9).

The procedure to get a full solution is (i) choose the functions $f(\chi)$, $m(\chi)$ and $\tau_0(\chi)$; (ii) evaluate $E(\chi)$ via (14.10); (iii) depending of on the value of $E(\chi)$, use (14.11), (14.12) or (14.13) to get the solution for $r(\tau,\chi)$; (iv) plug this solution into the remaining Lemaître-Tolman equation, Eq. (14.7b), to get $\rho(\tau,\chi)$ and into (14.6) to get the metric tensor.

14.2.5 Schwarzschild solution in Lemaître coordinates

One can recover Schwarzschild solution from the above setting by considering the vacuum case, i.e. $\rho = 0$. Then Eq. (14.7b) imposes $m(\chi)$ to be a constant, which we shall denote simply by m. Regarding the function $f(\chi)$, let us choose for simplicity $f(\chi) = 1$. Then $E(\chi) = 0$ and $r(\tau,\chi)$ is given by Eq. (14.12). Since m is constant, we cannot choose $\tau_0(\chi)$ to be a constant, otherwise $\partial r/\partial\chi$ would be zero and the metric (14.6) would be degenerate. The simplest non-constant choice is $\tau_0(\chi) = \chi$. To summarize, the three functions of χ determining the solution are set to

$$m(\chi) = m = \text{const}, \qquad f(\chi) = 1 \qquad \text{and} \qquad \tau_0(\chi) = \chi. \tag{14.14}$$

Equation (14.12), with the above values for $m(\chi)$ and $\tau_0(\chi)$, yields

$$r(\tau,\chi) = \left(\frac{9m}{2}\right)^{1/3} (\chi - \tau)^{2/3}. \tag{14.15}$$

In what follows, we assume $\chi \geq \tau$. Then

$$\chi - \tau = \frac{1}{3}\sqrt{\frac{2}{m}}\, r^{3/2} \tag{14.16}$$

and

$$\frac{\partial r}{\partial \chi} = \left(\frac{4m}{3}\right)^{1/3} (\chi - \tau)^{-1/3} = \sqrt{\frac{2m}{r}}. \tag{14.17}$$

Accordingly, Eq. (14.6) becomes

$$\boxed{g_{\mu\nu}\, \mathrm{d}x^\mu\, \mathrm{d}x^\nu = -\mathrm{d}\tau^2 + \frac{2m}{r}\mathrm{d}\chi^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)}. \tag{14.18}$$

In this expression, r is the function of (τ, χ) given by (14.15).

The metric (14.18) is actually the Schwarzschild metric of mass parameter m. To prove it, let us first promote r as a coordinate, instead of χ, i.e. we consider the coordinate system $(x'^\alpha) := (\tau, r, \theta, \varphi)$, which are called ***Painlevé-Gullstrand coordinates***. The relation to Lemaître coordinates $(x^\alpha) = (\tau, \chi, \theta, \varphi)$ is obtained by differentiating (14.15): we have clearly $\partial r/\partial\tau = -\partial r/\partial\chi$, so that, taking into account (14.17),

$$\mathrm{d}r = \sqrt{\frac{2m}{r}}(\mathrm{d}\chi - \mathrm{d}\tau).$$

Hence

$$\sqrt{\frac{2m}{r}}\mathrm{d}\chi = \sqrt{\frac{2m}{r}}\mathrm{d}\tau + \mathrm{d}r \quad \Longrightarrow \quad \frac{2m}{r}\mathrm{d}\chi^2 = \frac{2m}{r}\mathrm{d}\tau^2 + 2\sqrt{\frac{2m}{r}}\mathrm{d}\tau\, \mathrm{d}r + \mathrm{d}r^2.$$

Substituting this relation in Eq. (14.18) yields the expression of the metric tensor in terms of Painlevé-Gullstrand coordinates (x'^α):

$$\boxed{g'_{\mu\nu}\, \mathrm{d}x'^\mu\, \mathrm{d}x'^\nu = -\left(1 - \frac{2m}{r}\right)\mathrm{d}\tau^2 + 2\sqrt{\frac{2m}{r}}\mathrm{d}\tau\, \mathrm{d}r + \mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right)}. \tag{14.19}$$

We can rearrange it as

$$\begin{aligned} g'_{\mu\nu}\, \mathrm{d}x'^\mu\, \mathrm{d}x'^\nu &= -\left(1 - \frac{2m}{r}\right)\left(\mathrm{d}\tau^2 - 2\frac{\sqrt{\frac{2m}{r}}}{1 - \frac{2m}{r}}\, \mathrm{d}\tau\, \mathrm{d}r\right) + \mathrm{d}r^2 + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right) \\ &= -\left(1 - \frac{2m}{r}\right)\left(\mathrm{d}\tau - \frac{\sqrt{\frac{2m}{r}}}{1 - \frac{2m}{r}}\, \mathrm{d}r\right)^2 + \frac{\mathrm{d}r^2}{1 - \frac{2m}{r}} + r^2\left(\mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\right). \end{aligned} \tag{14.20}$$

If we introduce, instead of τ, a coordinate t such that

$$\mathrm{d}t = \mathrm{d}\tau - \frac{\sqrt{\frac{2m}{r}}}{1 - \frac{2m}{r}}\, \mathrm{d}r, \tag{14.21}$$

Eq. (14.20) yields immediately to the familiar expression of Schwarzschild metric in Schwarzschild-Droste coordinates (t, r, θ, φ) [Eq. (6.15)]. Hence this proves that the vacuum solution (14.18) is nothing but Schwarzschild metric. Incidently, since our starting point was the most general metric for a spherically symmetric spacetime [Eq. (14.1)], we have proven ***Birkhoff's theorem***:

> In vacuum, the unique spherically symmetric solution of the 4-dimensional Einstein equation with $\Lambda = 0$ is Schwarzschild metric.

In particular, outside any spherically symmetric body, the spacetime is a piece of Schwarzschild spacetime. Note that this implies that this part of spacetime is static, even if the central body is not (for instance oscillate radially, keeping its spherical symmetry). In other words, there are no gravitational waves in spherical symmetry.

Remark 1: Birkhoff's theorem can be viewed as a generalization of Gauss' theorem in Newtonian gravity: the gravitational field outside any spherical source is entirely determined by the mass m of the source, being identical to that generated by a point of mass m located at the symmetry center.

The relation between Lemaître coordinates and Schwarzschild-Droste ones can be made explicit by integrating Eq. (14.21); one gets

$$\tau = t + 4m\sqrt{\frac{r}{2m}} + 2m \ln\left(\frac{\sqrt{r/2m} - 1}{\sqrt{r/2m} + 1}\right) + \text{const.} \tag{14.22}$$

The expression of χ in terms of (t, r) is then deduced from Eq. (14.16):

$$\chi = t + 4m\sqrt{\frac{r}{2m}}\left(1 + \frac{r}{6m}\right) + 2m \ln\left(\frac{\sqrt{r/2m} - 1}{\sqrt{r/2m} + 1}\right) + \text{const.} \tag{14.23}$$

We deduce easily from these formulas the expression of the stationarity Killing vector $\boldsymbol{\xi}$ of Schwarzschild spacetime in terms of the Lemaître coordinates. Since $\boldsymbol{\xi} = \boldsymbol{\partial}_t$ [Eq. (6.7)], and the above formulas imply $\partial\tau/\partial t = 1$ and $\partial\chi/\partial t = 1$, we get, applying the chain rule $\partial/\partial t = \partial/\partial\tau \times \partial\tau/\partial t + \partial/\partial\chi \times \partial\chi/\partial t$,

$$\boxed{\boldsymbol{\xi} = \boldsymbol{\partial}_\tau + \boldsymbol{\partial}_\chi}. \tag{14.24}$$

Remark 2: Although very simple, this relation shows that Lemaître coordinates are *not* adapted to the spacetime symmetry generated by the Killing vector $\boldsymbol{\xi}$: the latter does not coincide with any Lemaître coordinate vector. This reflects the fact that the metric components (14.18) depend on τ (via the function $r(\tau, \chi)$), in addition to χ.

Despite the vacuum hypothesis means that we can no longer interpret the Lemaître coordinates as comoving with some free-falling dust (cf. Sec. 14.2.1), their geodesic character remains. Indeed, the vector $\boldsymbol{u} := \boldsymbol{\partial}_\tau$ is geodesic: $\boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u} = 0$, which implies that the

curves $(\chi, \theta, \varphi) = \text{const}$ are timelike geodesics. Moreover, the conserved energy per unit mass along these geodesics (cf. Sec. B.5) is

$$\varepsilon = -\boldsymbol{\xi} \cdot \boldsymbol{u} = -(\boldsymbol{\partial}_\tau + \boldsymbol{\partial}_\chi) \cdot \boldsymbol{\partial}_\tau = -\underbrace{g_{\tau\tau}}_{-1} - \underbrace{g_{\chi\tau}}_{0} = 1,$$

where use has been made of (14.24). $\varepsilon = 1$ means that the geodesics are marginally bound: they describe a free fall from rest at infinity.

As it is clear on the line element (14.18), a key feature of Lemaître coordinates is to be regular at $r = 2m$, i.e. accross the event horizon of Schwarzschild spacetime, contrary to the Schwarzschild-Droste coordinates.

Historical note: The Schwarzschild metric in the form (14.18) has been obtained in 1932 by Georges Lemaître [178], as a vacuum solution of the Lemaître-Tolman system: cf. Eq. (11.12) of Ref. [178]. Remarkably, Lemaître pointed out that the metric components (14.18) are regular at $r = 2m$ and was the first author to conclude that the singularity of Schwarzschild's solution at $r = 2m$ is a mere coordinate singularity. As pointed out in the historical note on p. 131, eight years before, Eddington exhibited a coordinate system that is regular at $r = 2m$ [97] but he did not mention this feature.

Remark 3: The Painlevé-Gullstrand coordinates, which have been introduced in our way from the Lemaître coordinates to the Schwarzschild-Droste ones, have a noticeable feature: the hypersurfaces $\tau = \text{const}$ are flat manifolds, i.e. the metric induced of them is the flat Euclidean metric. This is immediate if we set $\mathrm{d}\tau = 0$ in Eq. (14.19): the line element reduces to

$$\mathrm{d}s^2 = \mathrm{d}r^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\varphi^2\right),$$

which is nothing but the Euclidean line element expressed in spherical coordinates (r, θ, φ). This proves that the Schwarzschild spacetime can be sliced by a family of flat hypersurfaces. The associated 3+1 decomposition of the metric is revealed by rewritting the line element (14.19) as

$$\boxed{g'_{\mu\nu} \, \mathrm{d}x'^\mu \, \mathrm{d}x'^\nu = -\mathrm{d}\tau^2 + \left(\mathrm{d}r + \sqrt{\frac{2m}{r}}\mathrm{d}\tau\right)^2 + r^2 \left(\mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\varphi^2\right)}. \tag{14.25}$$

We read on this expression that the *lapse function* (see e.g. Ref. [122]) is $N = 1$ and that the *shift vector* is $\beta^i = (\sqrt{r/2m}, 0, 0)$. Finding a lapse function equal to one reflects simply that the coordinate time τ is some observer proper time: that of the marginally bound radial geodesics discussed above.

Another interesting property of Painlevé-Gullstrand coordinates, which they share with the Lemaître ones, is to be regular at $r = 2m$: despite the vanishing of $g'_{\tau\tau}$ there, as read on (14.19), the determinant of the metric components (14.19) is not vanishing, thanks to the off-diagonal term $g'_{\tau r}$: it is everywhere equal to $-r^4 \sin^2\theta$. See Ref. [184] for a detailed discussion of Painlevé-Gullstrand coordinates.

Historical note: Painlevé-Gullstrand coordinates have been introduced in 1921 by the French mathematician Paul Painlevé [203], as well as by the Swedish physicist and ophtalmologist Allvar Gullstrand (1911 laureate of the Nobel Price in Medicine) in 1922 [136].

14.3 Oppenheimer-Snyder solution

Chapter 15

Black hole formation 2: Vaidya collapse

Chapter 16

Evolution and thermodynamics of black holes

Chapter 17

Black holes and gravitational waves

Chapter 18

The quasi-local approach: trapping horizons

Chapter 19

Higher-dimensional solutions and black holes in alternative theories

Appendix A

Basic differential geometry

Contents

A.1 Introduction

The mathematical language of general relativity is mostly differential geometry. We recall in this appendix basic definitions and results in this field, which we will use throughout these lectures. The reader who has some knowledge of general relativity should be familiar with most of them. We recall them here to make the text fairly self-contained and also to provide definitions with sufficient generality, not limited to the dimension 4 — the standard spacetime dimension. Indeed, even when restricted ourselves to a 4-dimensional spacetime, we have to deal with manifolds whose dimension differs from 4, such as hypersurfaces (e.g. the black hole event horizon) or 2-dimensional surfaces (e.g. cross-sections of a horizon). In the same spirit, we do not stick to Lorentzian metrics (such as the spacetime one) but discuss pseudo-Riemannian metrics, which encompass both Lorentzian metrics and Riemannian ones. Accordingly, in this appendix, $\mathcal{M}$ denotes a generic manifold of any dimension and $\boldsymbol{g}$ a pseudo-Riemannian metric on $\mathcal{M}$.

This appendix is not intended to be some lecture on differential geometry, but a collection of basic definitions and useful results. In particular, contrary to the other parts of these notes, we state many results without proofs, referring the reader to classical textbooks on the topic [171, 176, 177, 199, 23, 53, 102], as well as to the differential geometry sections of the general relativity textbooks [51, 237, 256].

A.2 Differentiable manifolds

A.2.1 Notion of manifold

Given an integer $n \geq 1$, a ***manifold of dimension*** n is a topological space $\mathscr{M}$ obeying the following properties:

1. $\mathscr{M}$ is a ***separated space*** (also called ***Hausdorff space***): any two distinct points of $\mathscr{M}$ admit disjoint open neighbourhoods.

2. $\mathscr{M}$ has a ***countable base***[1]: there exists a countable family $(\mathcal{U}_k)_{k\in\mathbb{N}}$ of open sets of $\mathscr{M}$ such that any open set of $\mathscr{M}$ can be written as the union (possibly infinite) of some members of the above family.

3. Around each point of $\mathscr{M}$, there exists a neighbourhood which is homeomorphic to an open subset of $\mathbb{R}^n$.

Property 1 excludes manifolds with "forks" and is very reasonable from a physical point of view: it allows to distinguish between two points even after a small perturbation. Property 2 excludes "too large" manifolds; in particular it permits setting up the theory of integration on manifolds. It also allows for a smooth manifold of dimension n to be embedded smoothly into the Euclidean space $\mathbb{R}^{2n}$ (Whitney theorem). Property 3 expresses the essence of a manifold: it means that, locally, one can label the points of $\mathscr{M}$ in a continuous way by n real numbers $(x^\alpha)_{\alpha\in\{0,\ldots,n-1\}}$, which are called ***coordinates*** (cf. Fig. A.1). More precisely, given an open subset $\mathcal{U} \subset \mathscr{M}$, a ***coordinate system*** or ***chart*** on $\mathcal{U}$ is a homeomorphism[2]

$$\begin{array}{rccc} \Phi: & \mathcal{U} \subset \mathscr{M} & \longrightarrow & \Phi(\mathcal{U}) \subset \mathbb{R}^n \\ & p & \longmapsto & (x^0, \ldots, x^{n-1}). \end{array} \tag{A.1}$$

Remark 1: In relativity, it is customary to label the n coordinates by an index ranging from 0 to $n-1$. Actually, this convention is mostly used when $\mathscr{M}$ is the spacetime manifold ($n = 4$ in standard general relativity). The computer-oriented reader will have noticed the similarity with the index ranging of arrays in the C/C++ or Python programming languages.

Remark 2: Strictly speaking the definition given above is that of a ***topological manifold***. We are saying *manifold* for short.

Usually, one needs more than one coordinate system to cover $\mathscr{M}$. An ***atlas*** on $\mathscr{M}$ is a set of pairs $(\mathcal{U}_i, \Phi_i)_{i\in I}$, where I is a set (non necessarily finite), $\mathcal{U}_i$ an open set of $\mathscr{M}$ and Φ_i a chart on $\mathcal{U}_i$, such that the union of all $\mathcal{U}_i$ covers $\mathscr{M}$:

$$\bigcup_{i\in I} \mathcal{U}_i = \mathscr{M}. \tag{A.2}$$

[1] In the language of topology, one says that $\mathscr{M}$ is a *second-countable space*.

[2] Let us recall that a ***homeomorphism*** between two topological spaces (here $\mathcal{U}$ and $\Phi(\mathcal{U})$) is a bijective map Φ such that both Φ and Φ^{-1} are continuous.

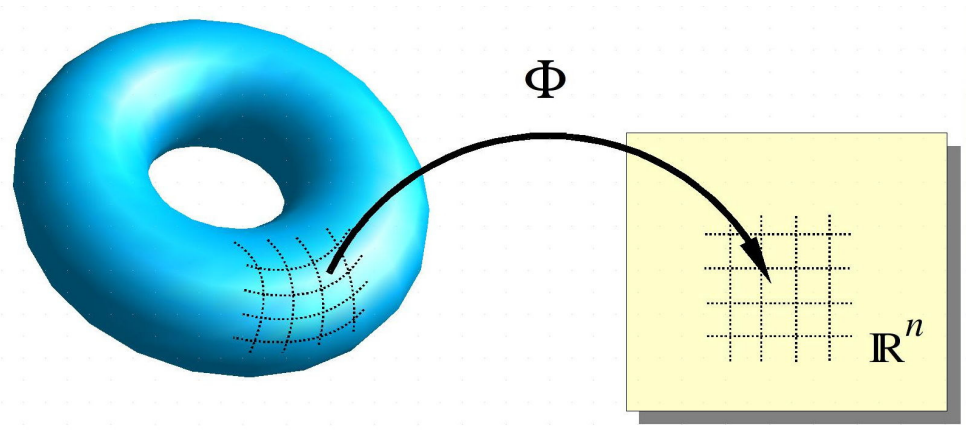


Figure A.1: Locally a manifold resembles $\mathbb{R}^n$ ($n = 2$ on the figure), but this is not necessarily the case at the global level.

The above definition of a manifold lies at the *topological* level (cf. Remark 2), meaning that one has the notion of continuity, but not of differentiability. To get the latter, one should rely on the smooth structure of $\mathbb{R}^n$, via the atlases: a ***smooth manifold***, is a manifold $\mathscr{M}$ equipped with an atlas $(\mathcal{U}_i, \Phi_i)_{i\in I}$ such that for any non-empty intersection $\mathcal{U}_i \cap \mathcal{U}_j$, the mapping

$$\Phi_i \circ \Phi_j^{-1} : \Phi_j(\mathcal{U}_i \cap \mathcal{U}_j) \subset \mathbb{R}^n \longrightarrow \Phi_i(\mathcal{U}_i \cap \mathcal{U}_j) \subset \mathbb{R}^n \tag{A.3}$$

is smooth (i.e. C^∞). Note that the above mapping is from an open set of $\mathbb{R}^n$ to an open set of $\mathbb{R}^n$, so that the invoked differentiability is nothing but that of $\mathbb{R}^n$. Such a mapping is called a ***change of coordinates*** or, in the mathematically-oriented literature, a ***transition map***. The atlas $(\mathcal{U}_i, \Phi_i)_{i\in I}$ is called a ***smooth atlas***. In the following, we consider only smooth manifolds.

Remark 3: When discussing the staticity and no-hair theorems in Chap. 5, we refer to the concept of *analytic manifold*, which is a special case of that of smooth manifold. Indeed, an ***analytic manifold*** is defined as a manifold equipped with an altas for which all the changes of coordinates $\Phi_i \circ \Phi_j^{-1}$ are analytic functions. Let us recall that an ***analytic function*** is a C^∞ function f for which the Taylor series about any point x in its domain converges to f in some neighborhood of x.

Given two smooth manifolds, $\mathscr{M}$ and $\mathscr{M}'$, of respective dimensions n and n', we say that a map $\phi : \mathscr{M} \to \mathscr{M}'$ is ***smooth map*** iff in some (and hence all, thanks to the smoothness of (A.3)) coordinate systems of $\mathscr{M}$ and $\mathscr{M}'$ belonging to the smooth atlases of $\mathscr{M}$ and $\mathscr{M}'$, the coordinates of the image $\phi(p)$ are smooth functions $\mathbb{R}^n \to \mathbb{R}^{n'}$ of the coordinates of p. The map ϕ is said to be a ***diffeomorphism*** iff it is bijective and both ϕ and ϕ^{-1} are smooth. This implies $n = n'$.

Remark 4: Strictly speaking a smooth manifold is a pair $(\mathscr{M}, \mathcal{A})$ where $\mathcal{A}$ is a (maximal) smooth atlas on $\mathscr{M}$. Indeed, a given (topological) manifold $\mathscr{M}$ can have non-equivalent differentiable structures, as shown by Milnor (1956) [187] in the specific case of the unit sphere of dimension 7, $\mathbb{S}^7$: there exist smooth manifolds, the so-called *exotic spheres*, that are homeomorphic to $\mathbb{S}^7$ but

not diffeomorphic to $\mathbb{S}^7$. On the other side, for $n \leq 6$, there is a unique smooth structure for the sphere $\mathbb{S}^n$. Moreover, any manifold of dimension $n \leq 3$ admits a unique smooth structure. Amazingly, in the case of $\mathbb{R}^n$, there exists a unique smooth structure (the standard one) for any $n \neq 4$, but for $n = 4$ (the spacetime case!) there exist uncountably many non-equivalent smooth structures, the so-called *exotic* $\mathbb{R}^4$ [243].

A.2.2 Manifolds with boundary

A (topological) ***manifold with boundary*** $\mathscr{M}$ is defined in the same way as a topological manifold, except that condition 3 in the definition given at the beginning of Sec. A.2.1 is replaced by

3'. Around each point of $\mathscr{M}$, there exists a neighbourhood which is homeomorphic either to an open subset of $\mathbb{R}^n$ or to an open subset[3] the closed half-space

$$\mathbb{H}^n := \left\{(x^1, \ldots, x^n) \in \mathbb{R}^n, \quad x^n \geq 0\right\}. \tag{A.4}$$

A point $p \in \mathscr{M}$ is said to be a ***boundary point*** of $\mathscr{M}$ iff there exists a homeomorphism $\Phi : \mathcal{U} \to \Phi(\mathcal{U}) \subset \mathbb{H}^n$ such that $p \in \mathcal{U}$ and $\Phi(p) \in \partial\mathbb{H}^n$, where

$$\partial\mathbb{H}^n := \left\{(x^1, \ldots, x^n) \in \mathbb{R}^n, \quad x^n = 0\right\}. \tag{A.5}$$

The set of all boundary points of $\mathscr{M}$ is naturally called the ***boundary of*** $\mathscr{M}$ and is denoted by $\partial\mathscr{M}$.

A ***smooth manifold with boundary*** is a manifold with boundary endowed with a smooth atlas, with the understanding that a transition map

$$\Phi_i \circ \Phi_j^{-1} : \Phi_j(\mathcal{U}_i \cap \mathcal{U}_j) \subset \mathbb{H}^n \longrightarrow \Phi_i(\mathcal{U}_i \cap \mathcal{U}_j) \subset \mathbb{H}^n$$

is said to be *smooth* iff it can be extended around each point of its domain (including the points of $\partial\mathbb{H}^n$) into a smooth map from an open subset of $\mathbb{R}^n$ to $\mathbb{R}^n$.

A.2.3 Curves and vectors on a manifold

On a manifold, vectors are defined as tangent vectors to a curve. Given an interval $I \subset \mathbb{R}$, a (smooth) ***curve*** is a subset $\mathscr{L} \subset \mathscr{M}$ that is the image of a smooth map $I \to \mathscr{M}$:

$$\begin{array}{rccc} P: & I & \longrightarrow & \mathscr{M} \\ & \lambda & \longmapsto & p = P(\lambda) \in \mathscr{L}. \end{array} \tag{A.6}$$

Hence $\mathscr{L} = P(I) := \{P(\lambda) |\ \lambda \in I\}$. The function P is called a ***parametrization*** of $\mathscr{L}$ and the real variable λ is called a ***parameter along*** $\mathscr{L}$. Given a coordinate system (x^α)

[3] By *open subset of* $\mathbb{H}^n$, it is meant a set $A \subset \mathbb{H}^n$ that is open with respect to the topology of $\mathbb{H}^n$; A is then not necessarily open when considered as a subset of $\mathbb{R}^n$ (for instance $A = \mathbb{H}^n$).

in a neighbourhood of a point $p \in \mathscr{L}$, the parametrization P is defined by n functions $X^\alpha : I \to \mathbb{R}$ such that

$$x^\alpha(P(\lambda)) = X^\alpha(\lambda). \tag{A.7}$$

Remark 5: In the literature, especially in the mathematical one, a curve is often defined as a map $P : I \to \mathscr{M}$ and not as the image of P. According to this definition, different parametrizations give birth to different curves.

A ***scalar field*** on $\mathscr{M}$ is a function $f : \mathscr{M} \to \mathbb{R}$. In practice, we will always consider smooth scalar fields. At a point $p = P(\lambda) \in \mathscr{L}$, the ***vector tangent to*** $\mathscr{L}$ associated with the parametrization P is the operator $\boldsymbol{v}$ which maps every scalar field f to the real number

$$\boldsymbol{v}(f) = \left.\frac{\mathrm{d}f}{\mathrm{d}\lambda}\right|_{\mathscr{L}} := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[f(P(\lambda+\varepsilon)) - f(P(\lambda)) \right]. \tag{A.8}$$

Given a coordinate system (x^α) around some point $p \in \mathscr{M}$, there are n curves $\mathscr{L}_\alpha$ through p associated with (x^α) and called the ***coordinate lines***: for each $\alpha \in \{0, \ldots, n-1\}$, $\mathscr{L}_\alpha$ is defined as the curve through p parameterized by $\lambda = x^\alpha$ and having constant coordinates x^β for all $\beta \neq \alpha$. The vector tangent to $\mathscr{L}_\alpha$ parameterized by x^α is denoted $\boldsymbol{\partial}_\alpha$. Its action on a scalar field f is by definition

$$\boldsymbol{\partial}_\alpha(f) = \left.\frac{\mathrm{d}f}{\mathrm{d}x^\alpha}\right|_{\mathscr{L}_\alpha} = \left.\frac{\mathrm{d}f}{\mathrm{d}x^\alpha}\right|_{\substack{x^\beta = \mathrm{const} \\ \beta \neq \alpha}}.$$

Considering f as a function of the coordinates $(x^0, \ldots, x^{n-1})$ (whereas strictly speaking it is a function of the points on $\mathscr{M}$) we recognize in the last term the partial derivative of f with respect to x^α. Hence

$$\boxed{\boldsymbol{\partial}_\alpha(f) = \frac{\partial f}{\partial x^\alpha}}. \tag{A.9}$$

Similarly, we may rewrite (A.8) as

$$\begin{aligned} \boldsymbol{v}(f) &= \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[f(X^0(\lambda+\varepsilon), \ldots, X^{n-1}(\lambda+\varepsilon)) - f(X^0(\lambda), \ldots, X^{n-1}(\lambda)) \right] \\ &= \frac{\partial f}{\partial x^\alpha} \frac{\mathrm{d}X^\alpha}{\mathrm{d}\lambda} = \boldsymbol{\partial}_\alpha(f) \frac{\mathrm{d}X^\alpha}{\mathrm{d}\lambda}. \end{aligned}$$

In the above equation, we are using ***Einstein summation convention***: a repeated index implies a summation over all the possible values of this index (here from $\alpha = 0$ to $\alpha = n-1$). The above identity being valid for any scalar field f, we conclude that

$$\boxed{\boldsymbol{v} = v^\alpha \, \boldsymbol{\partial}_\alpha}, \tag{A.10}$$

with the n real numbers

$$v^\alpha := \frac{\mathrm{d}X^\alpha}{\mathrm{d}\lambda}, \qquad 0 \leq \alpha \leq n-1. \tag{A.11}$$

Since every vector tangent to a curve at p is expressible as (A.10), we conclude that the set of all vectors tangent to a curve at p is a vector space of dimension n and that

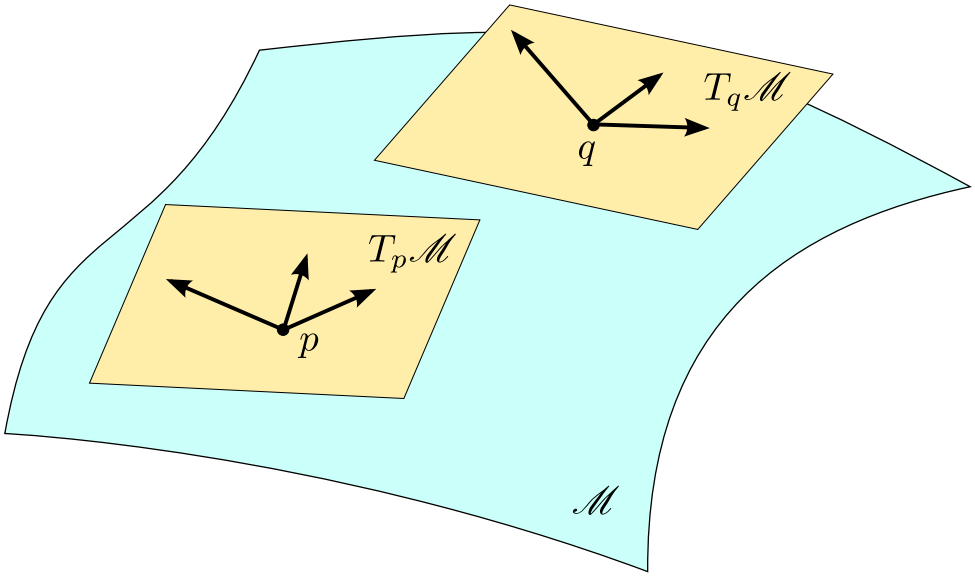


Figure A.2: The vectors at two points p and q on the manifold $\mathscr{M}$ belong to two different vector spaces: the tangent spaces $T_p\mathscr{M}$ and $T_q\mathscr{M}$.

$(\boldsymbol{\partial}_\alpha)$ constitutes a basis of it. This vector space is called the ***tangent vector space to*** $\mathscr{M}$ ***at*** p and is denoted $T_p\mathscr{M}$. The elements of $T_p\mathscr{M}$ are simply called ***vectors*** at p. The basis $(\boldsymbol{\partial}_\alpha)$ is called the ***natural basis*** associated with the coordinates (x^α) and the coefficients v^α in (A.10) are called the ***components of the vector*** $\boldsymbol{v}$ ***with respect to the coordinates*** (x^α). The tangent vector space is represented at two different points in Fig. A.2.

Contrary to what happens for an affine space, one cannot, in general, define a vector connecting two points p and q on a manifold, except if p and q are infinitesimally close to each other. Indeed, in the latter case, we may define the ***infinitesimal displacement vector from*** p ***to*** q as the vector $\mathrm{d}\boldsymbol{x} \in T_p\mathscr{M}$ whose action on a scalar field f is

$$\mathrm{d}\boldsymbol{x}(f) = \mathrm{d}f|_{p\to q} = f(q) - f(p). \tag{A.12}$$

Since p and q are infinitesimally close, there is a unique (piece of) curve $\mathscr{L}$ going from p to q and one has

$$\boxed{\mathrm{d}\boldsymbol{x} = \boldsymbol{v}\,\mathrm{d}\lambda}, \tag{A.13}$$

where λ is a parameter along $\mathscr{L}$, $\boldsymbol{v}$ the associated tangent vector at p and $\mathrm{d}\lambda$ the parameter increment from p to q: $p = P(\lambda)$ and $q = P(\lambda + \mathrm{d}\lambda)$. The relation (A.13) follows immediately from the definition (A.8) of $\boldsymbol{v}$. Given a coordinate system, let (x^α) be the coordinates of p and $(x^\alpha + \mathrm{d}x^\alpha)$ those of q. Then from Eq. (A.12),

$$\mathrm{d}\boldsymbol{x}(f) = \mathrm{d}f = \frac{\partial f}{\partial x^\alpha}\,\mathrm{d}x^\alpha = \mathrm{d}x^\alpha\,\boldsymbol{\partial}_\alpha(f).$$

The scalar field f being arbitrary, we conclude that

$$\boxed{\mathrm{d}\boldsymbol{x} = \mathrm{d}x^\alpha\,\boldsymbol{\partial}_\alpha}. \tag{A.14}$$

In other words, the components of the infinitesimal displacement vector with respect to the coordinates (x^α) are nothing but the infinitesimal coordinate increments $\mathrm{d}x^\alpha$.

A.2.4 Linear forms

A fundamental operation on vectors consists in mapping them to real numbers, and this in a linear way. More precisely, at each point $p \in \mathscr{M}$, one defines a ***linear form*** as a mapping[4]

$$\begin{array}{rccc} \boldsymbol{\omega}: & T_p\mathscr{M} & \longrightarrow & \mathbb{R} \\ & \boldsymbol{v} & \longmapsto & \langle \boldsymbol{\omega}, \boldsymbol{v} \rangle \end{array} \tag{A.15}$$

that is linear: $\langle \boldsymbol{\omega}, \lambda \boldsymbol{v} + \boldsymbol{u} \rangle = \lambda \langle \boldsymbol{\omega}, \boldsymbol{v} \rangle + \langle \boldsymbol{\omega}, \boldsymbol{u} \rangle$ for all $\boldsymbol{u}, \boldsymbol{v} \in T_p\mathscr{M}$ and $\lambda \in \mathbb{R}$. The set of all linear forms at p constitutes a n-dimensional vector space, which is called the ***dual space of*** $T_p\mathscr{M}$ and denoted by $T_p^*\mathscr{M}$. Given the natural basis $(\boldsymbol{\partial}_\alpha)$ of $T_p\mathscr{M}$ associated with some coordinates (x^α), there is a unique basis of $T_p^*\mathscr{M}$, denoted by $(\mathbf{d}x^\alpha)$, such that

$$\boxed{\langle \mathbf{d}x^\alpha, \boldsymbol{\partial}_\beta \rangle = \delta^\alpha{}_\beta}, \tag{A.16}$$

where $\delta^\alpha{}_\beta$ is the ***Kronecker symbol*** : $\delta^\alpha{}_\beta = 1$ if $\alpha = \beta$ and 0 otherwise. The basis $(\mathbf{d}x^\alpha)$ is called the ***dual basis*** of the basis $(\boldsymbol{\partial}_\alpha)$. The notation $(\mathbf{d}x^\alpha)$ stems from the fact that if we apply the linear form $\mathbf{d}x^\alpha$ to the infinitesimal displacement vector (A.14), we get nothing but the number $\mathrm{d}x^\alpha$:

$$\langle \mathbf{d}x^\alpha, \mathrm{d}\boldsymbol{x} \rangle = \langle \mathbf{d}x^\alpha, \mathrm{d}x^\beta\, \boldsymbol{\partial}_\beta \rangle = \mathrm{d}x^\beta \underbrace{\langle \mathbf{d}x^\alpha, \boldsymbol{\partial}_\beta \rangle}_{\delta^\alpha{}_\beta} = \mathrm{d}x^\alpha. \tag{A.17}$$

Remark 6: Do not confuse the linear form $\mathbf{d}x^\alpha$ with the infinitesimal increment $\mathrm{d}x^\alpha$ of the coordinate x^α.

The dual basis can be used to expand any linear form $\boldsymbol{\omega}$, thereby defining its ***components*** ω_α ***with respect to the coordinates*** (x^α):

$$\boldsymbol{\omega} = \omega_\alpha\, \mathbf{d}x^\alpha. \tag{A.18}$$

In terms of components, the action of a linear form on a vector takes then a very simple form:

$$\boxed{\langle \boldsymbol{\omega}, \boldsymbol{v} \rangle = \omega_\alpha v^\alpha}. \tag{A.19}$$

This follows immediately from (A.18), (A.10) and (A.16).

A field of linear forms, i.e. a (smooth) map which associates to each point $p \in \mathscr{M}$ a member of $T_p\mathscr{M}$ is called a ***1-form***. Given a smooth scalar field f on $\mathscr{M}$, there exists a 1-form canonically associated with it, called the ***differential of*** f and denoted $\mathbf{d}f$ or $\boldsymbol{\nabla} f$. At each point $p \in \mathscr{M}$, $\mathbf{d}f$ is the unique linear form which, once applied to the

[4] We are using the same bra-ket notation as in quantum mechanics to denote the action of a linear form on a vector.

infinitesimal displacement vector $\mathrm{d}\boldsymbol{x}$ from p to a nearby point q, gives the change in f between points p and q:

$$\mathrm{d}f := f(q) - f(p) = \langle \mathbf{d}f, \mathrm{d}\boldsymbol{x} \rangle. \tag{A.20}$$

Since $\mathrm{d}f = \partial f / \partial x^\alpha \, \mathrm{d}x^\alpha$, Eq. (A.17) implies that the components of the differential with respect to the dual basis are nothing but the partial derivatives of f with respect to the coordinates (x^α) :

$$\boxed{\mathbf{d}f = \frac{\partial f}{\partial x^\alpha} \, \mathbf{d}x^\alpha}. \tag{A.21}$$

Remark 7: In non-relativistic physics, the concept of ***gradient*** of a scalar field is commonly used instead of the differential, the former being a vector field and the latter a 1-form. This is so because one associates implicitly a vector $\overrightarrow{\boldsymbol{\omega}}$ to any 1-form $\boldsymbol{\omega}$ via the Euclidean scalar product of $\mathbb{R}^3$: $\forall \overrightarrow{v} \in \mathbb{R}^3$, $\langle \boldsymbol{\omega}, \overrightarrow{v} \rangle = \overrightarrow{\boldsymbol{\omega}} \cdot \overrightarrow{v}$. Accordingly, formula (A.20) is rewritten as $\mathrm{d}f = \overrightarrow{\boldsymbol{\nabla}} f \cdot \mathrm{d}\boldsymbol{x}$. But we should keep in mind that, at the fundamental level, the key quantity is the differential 1-form $\boldsymbol{\nabla} f = \mathbf{d}f$, for Eq. (A.20) does not require any metric on the manifold $\mathscr{M}$ to be meaningful. The gradient $\overrightarrow{\boldsymbol{\nabla}} f$ is a derived quantity, obtained from the differential $\boldsymbol{\nabla} f$ by metric duality.

Remark 8: For a fixed value of α, the coordinate x^α can be considered as a scalar field on $\mathscr{M}$. If we apply (A.21) to $f = x^\alpha$, we then get $\mathbf{d}x^\alpha = \mathbf{d}x^\alpha$. Hence the dual basis to the natural basis $(\boldsymbol{\partial}_\alpha)$ is formed by the differentials of the coordinates. This justifies the notation $\mathbf{d}x^\alpha$ used for its elements.

Of course the natural bases are not the only possible bases in the vector space $T_p\mathscr{M}$. One may use a basis $(\boldsymbol{e}_\alpha)$ that is not related to a coordinate system on $\mathscr{M}$, for instance an orthonormal basis with respect to some metric. There exists then a unique basis $(\boldsymbol{e}^\alpha)$ of the dual space $T_p^*\mathscr{M}$ such that[5]

$$\boxed{\langle \boldsymbol{e}^\alpha, \boldsymbol{e}_\beta \rangle = \delta^\alpha{}_\beta}. \tag{A.22}$$

$(\boldsymbol{e}^\alpha)$ is called the ***dual basis*** to $(\boldsymbol{e}_\alpha)$. The relation (A.16) is a special case of (A.22), for which $\boldsymbol{e}_\alpha = \boldsymbol{\partial}_\alpha$ and $\boldsymbol{e}^\alpha = \mathbf{d}x^\alpha$.

A.2.5 Tensors

Tensors are generalizations of both vectors and linear forms. At a point $p \in \mathscr{M}$, a ***tensor of type*** (k, ℓ) with $(k, \ell) \in \mathbb{N}^2$, also called ***tensor*** k ***times contravariant and*** ℓ ***times covariant***, is a mapping

$$\begin{array}{rccc} \boldsymbol{T} : & \underbrace{T_p^*\mathscr{M} \times \cdots \times T_p^*\mathscr{M}}_{k \text{ times}} \times \underbrace{T_p\mathscr{M} \times \cdots \times T_p\mathscr{M}}_{\ell \text{ times}} & \longrightarrow & \mathbb{R} \\ & (\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_k, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_\ell) & \longmapsto & \boldsymbol{T}(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_k, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_\ell) \end{array} \tag{A.23}$$

[5]Notice that, according to the standard usage, the symbol for the vector $\boldsymbol{e}_\alpha$ and that for the linear form $\boldsymbol{e}^\alpha$ differ only by the position of the index α.

that is linear with respect to each of its arguments. The integer $k+\ell$ is called the tensor ***valence***, or sometimes the tensor ***rank*** or ***order***. Let us recall the canonical duality $T_p^{**}\mathscr{M} = T_p\mathscr{M}$, which means that every vector $\boldsymbol{v}$ can be considered as a linear form on the space $T_p^*\mathscr{M}$, via $T_p^*\mathscr{M} \to \mathbb{R}$, $\boldsymbol{\omega} \mapsto \langle \boldsymbol{\omega}, \boldsymbol{v} \rangle$. Accordingly a vector is a tensor of type $(1,0)$. A linear form is a tensor of type $(0,1)$. A tensor of type $(0,2)$ is called a ***bilinear form***. It maps pairs of vectors to real numbers, in a linear way for each vector.

Given a basis $(\boldsymbol{e}_\alpha)$ of $T_p\mathscr{M}$ and the corresponding dual basis $(\boldsymbol{e}^\alpha)$ in $T_p^*\mathscr{M}$, we can expand any tensor $\boldsymbol{T}$ of type (k,ℓ) as

$$\boxed{\boldsymbol{T} = T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell}\, \boldsymbol{e}_{\alpha_1} \otimes \ldots \otimes \boldsymbol{e}_{\alpha_k} \otimes \boldsymbol{e}^{\beta_1} \otimes \ldots \otimes \boldsymbol{e}^{\beta_\ell}}, \tag{A.24}$$

where the ***tensor product*** $\boldsymbol{e}_{\alpha_1} \otimes \ldots \otimes \boldsymbol{e}_{\alpha_k} \otimes \boldsymbol{e}^{\beta_1} \otimes \ldots \otimes \boldsymbol{e}^{\beta_\ell}$ is the tensor of type (k,ℓ) for which the image of $(\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_k, \boldsymbol{v}_1, \ldots, \boldsymbol{v}_\ell)$ as in (A.23) is the real number

$$\prod_{i=1}^{k} \langle \boldsymbol{\omega}_i, \boldsymbol{e}_{\alpha_i} \rangle \;\times\; \prod_{j=1}^{\ell} \langle \boldsymbol{e}^{\beta_j}, \boldsymbol{v}_j \rangle .$$

Notice that all the products in the above formula are simply products in $\mathbb{R}$. The $n^{k+\ell}$ scalar coefficients $T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell}$ in (A.24) are called the ***components of the tensor $\boldsymbol{T}$ with respect to the basis*** $(\boldsymbol{e}_\alpha)$. These components are unique and fully characterize the tensor $\boldsymbol{T}$.

Remark 9: The notations v^α and ω_α already introduced for the components of a vector $\boldsymbol{v}$ [Eq. (A.10)] or a linear form $\boldsymbol{\omega}$ [Eq. (A.18)] are of course the particular cases $(k,\ell) = (1,0)$ or $(k,\ell) = (0,1)$ of (A.24), with, in addition, $\boldsymbol{e}_\alpha = \boldsymbol{\partial}_\alpha$ and $\boldsymbol{e}^\alpha = \mathbf{d}x^\alpha$.

A.2.6 Fields on a manifold

A ***tensor field*** of type (k,ℓ) is a map which associates to each point $p \in \mathscr{M}$ a tensor of type (k,ℓ) on $T_p\mathscr{M}$. By convention, a scalar field is considered as a tensor field of type $(0,0)$. We shall consider only smooth fields.

Given a non-negative integer p, a ***differential form of order*** p, or ***p-form***, is a tensor field of type $(0,p)$, i.e. a field of p-linear forms, that is fully antisymmetric whenever $p \geq 2$. This definition generalizes that of a 1-form given in Sec. A.2.4.

A ***frame field*** or ***moving frame*** is a n-tuple of vector fields $(\boldsymbol{e}_\alpha)$ such that at each point $p \in \mathscr{M}$, $(\boldsymbol{e}_\alpha(p))$ is a basis of the tangent space $T_p\mathscr{M}$. If $n = 4$, a frame field is also called a ***tetrad*** and if $n = 3$, it is called a ***triad***.

Given a vector field $\boldsymbol{v}$ and a scalar field f, the function $\mathscr{M} \to \mathbb{R}$, $p \mapsto \boldsymbol{v}|_p(f)$ clearly defines a scalar field on $\mathscr{M}$, which we denote naturally by $\boldsymbol{v}(f)$. We may then define the ***commutator of two vector fields*** $\boldsymbol{u}$ and $\boldsymbol{v}$ as the vector field $[\boldsymbol{u},\boldsymbol{v}]$ whose action on a scalar field f is

$$[\boldsymbol{u},\boldsymbol{v}](f) := \boldsymbol{u}(\boldsymbol{v}(f)) - \boldsymbol{v}(\boldsymbol{u}(f)). \tag{A.25}$$

With respect to a coordinate system (x^α), it is not difficult, via (A.10), to see that the components of the commutator are

$$\boxed{[\boldsymbol{u},\boldsymbol{v}]^\alpha = u^\mu \frac{\partial v^\alpha}{\partial x^\mu} - v^\mu \frac{\partial u^\alpha}{\partial x^\mu}}. \tag{A.26}$$

A.2.7 Immersions, embeddings and submanifolds

Let $\mathscr{M}$ and $\mathscr{N}$ be two smooth manifolds and

$$\Phi : \ \mathscr{M} \longrightarrow \mathscr{N} \tag{A.27}$$

be a smooth map (cf. Sec. A.2.1). At a given point $p \in \mathscr{M}$, the ***differential*** of Φ is the linear map

$$\mathrm{d}\Phi|_p : \ T_p\mathscr{M} \longrightarrow T_{\Phi(p)}\mathscr{N} \tag{A.28}$$

that "approximates" Φ in the following sense: if $\mathrm{d}\boldsymbol{x} \in T_p\mathscr{M}$ is the infinitesimal displacement vector from p to some (infinitesimally close) point q, then

$$\mathrm{d}\Phi|_p\,(\mathrm{d}\boldsymbol{x}) = \mathrm{d}\boldsymbol{L}, \tag{A.29}$$

where $\mathrm{d}\boldsymbol{L}$ is the infinitesimal displacement vector of $T_{\Phi(p)}\mathscr{N}$ connecting $\Phi(p)$ to $\Phi(q)$ (cf. Fig. ??). In terms of the characterization of vectors by their action on scalar fields [Eq. (A.8)], it is easy to see, thanks to (A.12), that the definition (A.29) is equivalent to

$$\forall \boldsymbol{v} \in T_p\mathscr{M}, \ \forall f \in C^\infty(\mathscr{N},\mathbb{R}), \quad \mathrm{d}\Phi|_p\,(\boldsymbol{v})(f) = \boldsymbol{v}\,(f\circ\Phi)\,. \tag{A.30}$$

The smooth map Φ is called an ***immersion*** iff the differential $\mathrm{d}\Phi|_p$ is injective at any point $p \in \mathscr{M}$. Moreover, Φ is called an ***embedding*** iff (i) Φ is an immersion and (ii) Φ is a homeomorphism $\mathscr{M} \to \Phi(\mathscr{M})$. Note that an embedding is necessarily injective, contrary to an immersion.

A ***submanifold*** of $\mathscr{M}$ is a subset $\mathscr{S} \subset \mathscr{M}$ such that (i) $\mathscr{S}$ is a manifold in the subspace topology and (ii) $\mathscr{S}$ has a smooth structure with respect to which the inclusion map $\iota : \mathscr{S} \to \mathscr{M}$ is an embedding. One can show that $\mathscr{S}$ is a submanifold of $\mathscr{M}$ iff there exists a manifold $\mathscr{S}_0$ (a priori not included in $\mathscr{M}$) and an embedding $\Phi : \mathscr{S}_0 \to \mathscr{M}$, such that $\mathscr{S} = \Phi(\mathscr{S}_0)$.

Remark 10: Scritly speaking, the above definition regards an ***embedded submanifold***; there is also the wider concept of ***immersed submanifold*** (see e.g. Chap 5 of [176]).

One has necessarily $\dim\mathscr{S} \leq \dim\mathscr{M}$. The non-negative integer $m = \dim\mathscr{M} - \dim\mathscr{S}$ is called the ***codimension*** of the submanifold $\mathscr{S}$. A submanifold of codimension 1 is called a ***hypersurface***. A submanifold of dimension 1 is (the image of) a curve in $\mathscr{M}$, but note that not all curves are submanifolds: a curve with self-crossing points is not a submanifold.

A.3 Pseudo-Riemannian manifolds

A.3.1 Metric tensor

A ***pseudo-Riemannian manifold*** is a pair $(\mathscr{M}, \boldsymbol{g})$ where $\mathscr{M}$ is a smooth manifold and $\boldsymbol{g}$ is a ***metric tensor*** on $\mathscr{M}$, i.e. a tensor field obeying the following properties:

1. $\boldsymbol{g}$ is a tensor field of type $(0,2)$: at each point $p \in \mathscr{M}$, $\boldsymbol{g}(p)$ is a bilinear form acting on vectors in the tangent space $T_p\mathscr{M}$:
$$\begin{array}{rccc} \boldsymbol{g}(p): & T_p\mathscr{M} \times T_p\mathscr{M} & \longrightarrow & \mathbb{R} \\ & (\boldsymbol{u}, \boldsymbol{v}) & \longmapsto & \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}). \end{array} \tag{A.31}$$
2. $\boldsymbol{g}$ is ***symmetric***: $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{g}(\boldsymbol{v}, \boldsymbol{u})$.
3. $\boldsymbol{g}$ is ***non-degenerate***: at any point $p \in \mathscr{M}$, a vector $\boldsymbol{u}$ such that $\forall \boldsymbol{v} \in T_p\mathscr{M}$, $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = 0$ is necessarily the null vector.

The properties of being symmetric and non-degenerate are typical of a ***scalar product***. Accordingly, one says that two vectors $\boldsymbol{u}$ and $\boldsymbol{v}$ are ***$\boldsymbol{g}$-orthogonal*** (or simply ***orthogonal*** if there is no ambiguity) iff $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = 0$. Moreover, when there is no ambiguity on the metric (usually the spacetime metric), we are using a dot to denote the scalar product of two vectors taken with $\boldsymbol{g}$:

$$\forall (\boldsymbol{u}, \boldsymbol{v}) \in T_p\mathscr{M} \times T_p\mathscr{M}, \quad \boxed{\boldsymbol{u} \cdot \boldsymbol{v} := \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v})}. \tag{A.32}$$

In a given basis $(\boldsymbol{e}_\alpha)$ of $T_p\mathscr{M}$, the components of $\boldsymbol{g}$ is the matrix $(g_{\alpha\beta})$ defined by formula (A.24) with $(k, \ell) = (0, 2)$:

$$\boldsymbol{g} = g_{\alpha\beta}\, \boldsymbol{e}^\alpha \otimes \boldsymbol{e}^\beta. \tag{A.33}$$

For any pair $(\boldsymbol{u}, \boldsymbol{v})$ of vectors we have then

$$\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}) = g_{\alpha\beta} u^\alpha v^\beta. \tag{A.34}$$

In particular, considering the natural basis associated with some coordinate system (x^α), the scalar square of an infinitesimal displacement vector $\mathrm{d}\boldsymbol{x}$ [cf. Eq. (A.12)] is

$$\mathrm{d}s^2 := \boldsymbol{g}(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{x}) = g_{\alpha\beta} \mathrm{d}x^\alpha\, \mathrm{d}x^\beta. \tag{A.35}$$

This formula, which follows from the value (A.14) of the components of $\mathrm{d}\boldsymbol{x}$, is called the expression of the ***line element*** on the pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$. It is often used to define the metric tensor in general relativity texts. Note that contrary to what the notation may suggest, $\mathrm{d}s^2$ is not necessarily a positive quantity.

A.3.2 Signature and orthonormal bases

An important feature of the metric tensor is its ***signature***: in all bases of $T_p\mathscr{M}$ where the components $(g_{\alpha\beta})$ form a diagonal matrix, there is necessarily the same number, s say, of negative components and the same number, $n-s$, of positive components. The independence of s from the choice of the basis where $(g_{\alpha\beta})$ is diagonal is a basic result of linear algebra named ***Sylvester's law of inertia***. One writes:

$$\text{sign}\, \boldsymbol{g} = (\underbrace{-,\ldots,-}_{s \text{ times}}, \underbrace{+,\ldots,+}_{n-s \text{ times}}). \tag{A.36}$$

If $s = 0$, $\boldsymbol{g}$ is called a ***Riemannian metric*** and $(\mathscr{M}, \boldsymbol{g})$ a ***Riemannian manifold***. In this case, $\boldsymbol{g}$ is ***positive-definite***, which means that

$$\forall \boldsymbol{v} \in T_p\mathscr{M}, \quad \boldsymbol{g}(\boldsymbol{v},\boldsymbol{v}) \geq 0 \tag{A.37}$$

and $\boldsymbol{g}(\boldsymbol{v},\boldsymbol{v}) = 0$ iff $\boldsymbol{v} = 0$. A standard example of Riemannian metric is of course the scalar product of the Euclidean space $\mathbb{R}^n$.

If $s = 1$, $\boldsymbol{g}$ is called a ***Lorentzian metric*** and $(\mathscr{M}, \boldsymbol{g})$ a ***Lorentzian manifold***. One may then have $\boldsymbol{g}(\boldsymbol{v},\boldsymbol{v}) < 0$; vectors for which this occurs are called ***timelike***, whereas vectors for which $\boldsymbol{g}(\boldsymbol{v},\boldsymbol{v}) > 0$ are called ***spacelike***, and those for which $\boldsymbol{g}(\boldsymbol{v},\boldsymbol{v}) = 0$ are called ***null***. The subset of $T_p\mathscr{M}$ formed by all null vectors is termed the ***null cone*** of $\boldsymbol{g}$ at p.

A basis $(\boldsymbol{e}_\alpha)$ of $T_p\mathscr{M}$ is called a ***g-orthonormal basis*** (or simply ***orthonormal basis*** if there is no ambiguity on the metric) iff[6]

$$\begin{aligned} &\boldsymbol{g}(\boldsymbol{e}_\alpha,\boldsymbol{e}_\alpha) = -1 \qquad \text{for} \qquad 0 \leq \alpha \leq s-1 \\ &\boldsymbol{g}(\boldsymbol{e}_\alpha,\boldsymbol{e}_\alpha) = 1 \qquad \text{for} \qquad s \leq \alpha \leq n-1 \\ &\boldsymbol{g}(\boldsymbol{e}_\alpha,\boldsymbol{e}_\beta) = 0 \qquad \text{for} \qquad \alpha \neq \beta. \end{aligned} \tag{A.38}$$

A.3.3 Metric duality

Since the bilinear form $\boldsymbol{g}$ is non-degenerate, its matrix $(g_{\alpha\beta})$ in any basis $(\boldsymbol{e}_\alpha)$ is invertible and the inverse is denoted by $(g^{\alpha\beta})$:

$$\boxed{g^{\alpha\mu} g_{\mu\beta} = \delta^\alpha{}_\beta}. \tag{A.39}$$

The metric $\boldsymbol{g}$ induces an isomorphism between $T_p\mathscr{M}$ (vectors) and $T_p^*\mathscr{M}$ (linear forms) which, in index notation, corresponds to the lowering or raising of the index by contraction with $g_{\alpha\beta}$ or $g^{\alpha\beta}$. In the present book, an index-free symbol will always denote a tensor with a fixed covariance type (such as a vector, a 1-form, a bilinear form, etc.). We will therefore use a different symbol to denote its image under the metric isomorphism. In particular, we denote by an underbar the isomorphism $T_p\mathscr{M} \to T_p^*\mathscr{M}$ and by an arrow the reverse isomorphism $T_p^*\mathscr{M} \to T_p\mathscr{M}$:

[6] No summation on α is implied.

1. For any vector $\boldsymbol{u}$ in $T_p\mathscr{M}$, $\underline{\boldsymbol{u}}$ stands for the unique linear form such that

$$\forall \boldsymbol{v} \in T_p\mathscr{M}, \quad \langle \underline{\boldsymbol{u}}, \boldsymbol{v} \rangle = \boldsymbol{g}(\boldsymbol{u}, \boldsymbol{v}). \tag{A.40}$$

   However, we will omit the underbar on the components of $\underline{\boldsymbol{u}}$, since the position of the index allows us to distinguish between vectors and linear forms, following the standard usage: if the components of $\boldsymbol{u}$ in a given basis $(\boldsymbol{e}_\alpha)$ are denoted by u^α, the components of $\underline{\boldsymbol{u}}$ in the dual basis $(\boldsymbol{e}^\alpha)$ are then denoted by u_α and are given by

$$u_\alpha = g_{\alpha\mu} u^\mu. \tag{A.41}$$

2. For any linear form $\boldsymbol{\omega}$ in $T_p^*\mathscr{M}$, $\overrightarrow{\boldsymbol{\omega}}$ stands for the unique vector of $T_p\mathscr{M}$ such that

$$\forall \boldsymbol{v} \in T_p\mathscr{M}, \quad \boldsymbol{g}(\overrightarrow{\boldsymbol{\omega}}, \boldsymbol{v}) = \langle \boldsymbol{\omega}, \boldsymbol{v} \rangle. \tag{A.42}$$

   As for the underbar, we will omit the arrow on the components of $\overrightarrow{\boldsymbol{\omega}}$ by denoting them ω^α; they are given by

$$\omega^\alpha = g^{\alpha\mu} \omega_\mu. \tag{A.43}$$

3. We extend the arrow notation to *bilinear* forms on $T_p\mathscr{M}$ (type-$(0,2)$ tensor): for any bilinear form $\boldsymbol{T}$, we denote by $\overrightarrow{\boldsymbol{T}}$ the tensor of type $(1,1)$ such that

$$\forall (\boldsymbol{u}, \boldsymbol{v}) \in T_p\mathscr{M} \times T_p\mathscr{M}, \quad \boldsymbol{T}(\boldsymbol{u}, \boldsymbol{v}) = \overrightarrow{\boldsymbol{T}}(\underline{\boldsymbol{u}}, \boldsymbol{v}) = \boldsymbol{u} \cdot \overrightarrow{\boldsymbol{T}}(\boldsymbol{v}), \tag{A.44}$$

   and by $\overrightarrow{\overrightarrow{\boldsymbol{T}}}$ the tensor of type $(2,0)$ defined by

$$\forall (\boldsymbol{u}, \boldsymbol{v}) \in T_p\mathscr{M} \times T_p\mathscr{M}, \quad \boldsymbol{T}(\boldsymbol{u}, \boldsymbol{v}) = \overrightarrow{\overrightarrow{\boldsymbol{T}}}(\underline{\boldsymbol{u}}, \underline{\boldsymbol{v}}). \tag{A.45}$$

   Note that in the second equality of (A.44), we have considered $\overrightarrow{\boldsymbol{T}}$ as an endomorphism $T_p\mathscr{M} \to T_p\mathscr{M}$, which is always possible for a tensor of type $(1,1)$. If $T_{\alpha\beta}$ are the components of $\boldsymbol{T}$ in some basis $(\boldsymbol{e}_\alpha)$, the components of $\overrightarrow{\boldsymbol{T}}$ and $\overrightarrow{\overrightarrow{\boldsymbol{T}}}$ are respectively

$$(\overrightarrow{\boldsymbol{T}})^\alpha{}_\beta = T^\alpha{}_\beta = g^{\alpha\mu} T_{\mu\beta} \tag{A.46}$$

$$(\overrightarrow{\overrightarrow{\boldsymbol{T}}})^{\alpha\beta} = T^{\alpha\beta} = g^{\alpha\mu} g^{\beta\nu} T_{\mu\nu}. \tag{A.47}$$

Remark 1: In mathematical literature, the isomorphism induced by $\boldsymbol{g}$ between $T_p\mathscr{M}$ and $T_p^*\mathscr{M}$ is called the ***musical isomorphism***, because a flat or a sharp symbol is used instead of, respectively, the underbar and the arrow introduced above:

$$\boldsymbol{u}^\flat = \underline{\boldsymbol{u}} \qquad \text{and} \qquad \boldsymbol{\omega}^\sharp = \overrightarrow{\boldsymbol{\omega}}.$$

A.3.4 Levi-Civita tensor

Let us assume that $\mathscr{M}$ is an ***orientable manifold***, i.e. that there exists a n-form[7] on $\mathscr{M}$ (n being $\mathscr{M}$'s dimension) that is continuous on $\mathscr{M}$ and nowhere vanishing. Then, given a metric $\boldsymbol{g}$ on $\mathscr{M}$, one can show that there exist only two n-forms $\boldsymbol{\epsilon}$ such that for any $\boldsymbol{g}$-orthonormal basis $(\boldsymbol{e}_\alpha)$,

$$\boldsymbol{\epsilon}(\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{n-1}) = \pm 1. \tag{A.48}$$

Picking one of these two n-forms amounts to choosing an ***orientation*** for $\mathscr{M}$. The chosen $\boldsymbol{\epsilon}$ is then called the ***Levi-Civita tensor*** associated with the metric $\boldsymbol{g}$. Bases for which the right-hand side of (A.48) is $+1$ are called ***right-handed***, and those for which it is -1 are called ***left-handed***. More generally, given a (not necessarily orthonormal) basis $(\boldsymbol{e}_\alpha)$ of $T_p\mathscr{M}$, one has necessarily $\boldsymbol{\epsilon}(\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{n-1}) \neq 0$ and one says that the basis is ***right-handed*** or ***left-handed*** iff $\boldsymbol{\epsilon}(\boldsymbol{e}_0, \ldots, \boldsymbol{e}_{n-1}) > 0$ or < 0, respectively. The components of $\boldsymbol{\epsilon}$ with respect to $(\boldsymbol{e}_\alpha)$ are given by

$$\boxed{\epsilon_{\alpha_1 \ldots \alpha_n} = \pm\sqrt{|g|}\, [\alpha_1, \ldots, \alpha_n]}, \tag{A.49}$$

where $\pm$ must be $+$ (resp. $-$) for a right-handed basis (resp. left-handed basis), g stands for the determinant of the matrix of $\boldsymbol{g}$'s components with respect to the basis $(\boldsymbol{e}_\alpha)$:

$$\boxed{g := \det(g_{\alpha\beta})} \tag{A.50}$$

and the symbol $[\alpha_1, \ldots, \alpha_n]$ takes the value 0 if any two indices $(\alpha_1, \ldots, \alpha_n)$ are equal and takes the value 1 or -1 if $(\alpha_1, \ldots, \alpha_n)$ is an even or odd permutation, respectively, of $(0, \ldots, n-1)$.

A.3.5 Vector normal to a hypersurface

In a pseudo-Riemannian manifold, one can associate to any hypersurface $\mathscr{S}$ (cf. Sec. A.2.7) a unique normal direction, which can be represented by a vector field $\boldsymbol{n}$ defined on $\mathscr{S}$ as follows. Locally the hypersurface $\mathscr{S}$ can be considered as a level set, i.e. there exists a smooth scalar field $f : \mathscr{M} \to \mathbb{R}$, such that for any point p in the local region of $\mathscr{M}$ considered, the following equivalence holds

$$p \in \mathscr{S} \iff f(p) = 0. \tag{A.51}$$

Then, a vector field $\boldsymbol{v}$ on $\mathscr{M}$ is tangent to $\mathscr{S}$ iff the value of f stays to 0 for a small displacement $\mathrm{d}\lambda$ along $\boldsymbol{v}$; thanks to Eqs. (A.12), (A.13) and (A.10), this is equivalent to

$$\boldsymbol{v}(f) = v^\mu \frac{\partial f}{\partial x^\mu} = 0, \tag{A.52}$$

or to

$$\boldsymbol{g}(\boldsymbol{n}, \boldsymbol{v}) = 0, \tag{A.53}$$

[7] Cf. Sec. A.2.6 for the definition of a n-form.

where we have let appear the gradient vector $\boldsymbol{n} := \overrightarrow{\boldsymbol{\nabla}} f$; in terms of components:

$$n^\alpha = \nabla^\alpha f = g^{\alpha\mu} \frac{\partial f}{\partial x^\mu}. \tag{A.54}$$

The vector field $\boldsymbol{n}$ is called a ***normal*** to $\mathscr{S}$. All normal vectors are collinear to each other.

A.4 The three basic derivatives

Three kinds of derivative operators on tensor fields can be defined on a pseudo-Riemannian manifold. One of them depends on the metric: the *covariant derivative* $\boldsymbol{\nabla}$ (Sec. A.4.1). Another one depends on the choice of a reference vector field: the *Lie derivative* $\boldsymbol{\mathcal{L}}$ (Sec. A.4.2). The third one depends only on the smooth-manifold structure, i.e. is independent of any (metric or vector) field, but, on the other side, it is applicable only to a specific kind of tensor fields: the p-forms; it is the *exterior derivative* $\mathbf{d}$ (Sec. A.4.3).

A.4.1 Covariant derivative

Affine connection on a manifold

Let us denote by $\mathfrak{X}(\mathscr{M})$ the space of smooth vector fields on $\mathscr{M}$. Given a vector field $\boldsymbol{v} \in \mathfrak{X}(\mathscr{M})$, it is not possible from the manifold structure alone to define its variation between two neighbouring points p and q. Indeed a formula like $\mathrm{d}\boldsymbol{v} := \boldsymbol{v}(q) - \boldsymbol{v}(p)$ is meaningless because the vectors $\boldsymbol{v}(q)$ and $\boldsymbol{v}(p)$ belong to two different vector spaces, $T_q\mathscr{M}$ and $T_p\mathscr{M}$ respectively (cf. Fig. A.2). Note that for a scalar field, this problem does not arise [cf. Eq. (A.20)]. The solution is to introduce an extra-structure on the manifold, called an *affine connection* because, by defining the variation of a vector field, it allows one to connect the various tangent spaces on the manifold.

An ***affine connection*** on $\mathscr{M}$ is a mapping

$$\begin{array}{rccc} \boldsymbol{\nabla} & : & \mathfrak{X}(\mathscr{M}) \times \mathfrak{X}(\mathscr{M}) & \longrightarrow & \mathfrak{X}(\mathscr{M}) \\ & & (\boldsymbol{u}, \boldsymbol{v}) & \longmapsto & \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{v} \end{array} \tag{A.55}$$

which satisfies the following properties:

1. $\boldsymbol{\nabla}$ is bilinear (considering $\mathfrak{X}(\mathscr{M})$ as a vector space over $\mathbb{R}$).
2. For any scalar field f and any pair $(\boldsymbol{u}, \boldsymbol{v})$ of vector fields:
$$\boldsymbol{\nabla}_{f\boldsymbol{u}} \boldsymbol{v} = f \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{v}. \tag{A.56}$$
3. For any scalar field f and any pair $(\boldsymbol{u}, \boldsymbol{v})$ of vector fields, the following Leibniz rule holds:
$$\boldsymbol{\nabla}_{\boldsymbol{u}} (f\boldsymbol{v}) = \langle \boldsymbol{\nabla} f, \boldsymbol{u} \rangle \, \boldsymbol{v} + f \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{v}, \tag{A.57}$$
where $\boldsymbol{\nabla} f$ stands for the differential of f as defined in Sec. A.2.4.

The vector $\boldsymbol{\nabla}_{\boldsymbol{u}}\,\boldsymbol{v}$ is called the ***covariant derivative of $\boldsymbol{v}$ along $\boldsymbol{u}$***.

Remark 1: Property 2 is not implied by property 1, for f is a scalar field, not a real number. Actually, property 2 ensures that at a given point $p \in \mathscr{M}$, the value of $\boldsymbol{\nabla}_{\boldsymbol{u}}\,\boldsymbol{v}$ depends only on the vector $\boldsymbol{u}(p) \in T_p\mathscr{M}$ and not on the behaviour of $\boldsymbol{u}$ around p; therefore the role of $\boldsymbol{u}$ is only to give the direction of the derivative of $\boldsymbol{v}$.

Given an affine connection, the variation of a vector field $\boldsymbol{v}$ between two neighbouring points p and q, is defined as

$$\mathrm{d}\boldsymbol{v} := \boldsymbol{\nabla}_{\mathrm{d}\boldsymbol{x}}\,\boldsymbol{v}, \tag{A.58}$$

$\mathrm{d}\boldsymbol{x}$ being the infinitesimal displacement vector connecting p and q [cf Eq. (A.12)]. One says that $\boldsymbol{v}$ is ***parallelly transported from p to q with respect to the connection*** $\boldsymbol{\nabla}$ iff $\mathrm{d}\boldsymbol{v} = 0$.

Given a frame field $(\boldsymbol{e}_\alpha)$ on $\mathscr{M}$, the ***connection coefficients*** of an affine connection $\boldsymbol{\nabla}$ with respect to $(\boldsymbol{e}_\alpha)$ are the scalar fields $\Gamma^\gamma{}_{\alpha\beta}$ defined by the expansion, at each point $p \in \mathscr{M}$, of the vector $\boldsymbol{\nabla}_{\boldsymbol{e}_\beta}\,\boldsymbol{e}_\alpha(p)$ onto the basis $(\boldsymbol{e}_\alpha(p))$:

$$\boxed{\boldsymbol{\nabla}_{\boldsymbol{e}_\beta}\,\boldsymbol{e}_\alpha =: \Gamma^\mu{}_{\alpha\beta}\,\boldsymbol{e}_\mu}. \tag{A.59}$$

An affine connection is entirely defined by the connection coefficients. In other words, there are as many affine connections on a manifold of dimension n as there are possibilities of choosing n^3 scalar fields $\Gamma^\gamma{}_{\alpha\beta}$.

Given $\boldsymbol{v} \in \mathfrak{X}(\mathscr{M})$, one defines a tensor field of type $(1,1)$, $\boldsymbol{\nabla}\boldsymbol{v}$, called the ***covariant derivative of $\boldsymbol{v}$ with respect to the affine connection*** $\boldsymbol{\nabla}$, by the following action at each point $p \in \mathscr{M}$:

$$\begin{array}{rccc} \boldsymbol{\nabla}\boldsymbol{v}(p) \;:\; & T_p^*\mathscr{M} \times T_p\mathscr{M} & \longrightarrow & \mathbb{R} \\ & (\boldsymbol{\omega}, \boldsymbol{u}) & \longmapsto & \langle \boldsymbol{\omega},\, \boldsymbol{\nabla}_{\tilde{\boldsymbol{u}}}\,\boldsymbol{v}(p)\rangle \end{array}, \tag{A.60}$$

where $\tilde{\boldsymbol{u}}$ is any vector field which performs some extension of $\boldsymbol{u}$ around p: $\tilde{\boldsymbol{u}}(p) = \boldsymbol{u}$. As already noted (cf. Remark 1), $\boldsymbol{\nabla}_{\tilde{\boldsymbol{u}}}\,\boldsymbol{v}(p)$ is independent of the choice of $\tilde{\boldsymbol{u}}$, so that the mapping (A.60) is well-defined. By comparing with (A.23), we verify that $\boldsymbol{\nabla}\boldsymbol{v}(p)$ is a tensor of type $(1,1)$.

One can extend the covariant derivative to all tensor fields by (i) demanding that for a scalar field the action of the affine connection is nothing but taking the differential (hence the same notation $\boldsymbol{\nabla} f$) and (ii) using the Leibniz rule. As a result, the covariant derivative of a tensor field $\boldsymbol{T}$ of type (k,ℓ) is a tensor field $\boldsymbol{\nabla}\boldsymbol{T}$ of type $(k,\ell+1)$. Its components with respect a given field frame $(\boldsymbol{e}_\alpha)$ are denoted

$$\nabla_\gamma T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell} := (\boldsymbol{\nabla}\boldsymbol{T})^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell\gamma} \tag{A.61}$$

and are given by

$$\begin{aligned} \nabla_\gamma T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell} \;=\; & \boldsymbol{e}_\gamma(T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell}) + \sum_{i=1}^{k} \Gamma^{\alpha_i}{}_{\sigma\gamma}\, T^{\alpha_1\ldots\overset{i}{\downarrow}\sigma\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell} \\ & - \sum_{i=1}^{\ell} \Gamma^{\sigma}{}_{\beta_i\gamma}\, T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\underset{i}{\uparrow}\sigma\ldots\beta_\ell}, \end{aligned} \tag{A.62}$$

where $\boldsymbol{e}_\gamma(T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell})$ stands for the action of the vector $\boldsymbol{e}_\gamma$ on the scalar field $T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell}$, resulting from the very definition of a vector (cf. Sec. A.2.3). In particular, if $(\boldsymbol{e}_\alpha)$ is a natural frame associated with some coordinate system (x^α), then $\boldsymbol{e}_\alpha = \boldsymbol{\partial}_\alpha$ and the above formula becomes [cf. (A.9)]

$$\begin{aligned}\nabla_\gamma T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell} &= \frac{\partial}{\partial x^\gamma} T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell} + \sum_{i=1}^{k} \Gamma^{\alpha_i}{}_{\sigma\gamma}\, T^{\alpha_1\ldots\overset{i}{\downarrow}\sigma\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell} \\ &\quad - \sum_{i=1}^{\ell} \Gamma^{\sigma}{}_{\beta_i\gamma}\, T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\underset{\uparrow}{\sigma}\ldots\beta_\ell}. \qquad \text{(A.63)}\end{aligned}$$

Remark 2: Notice the position of the index γ in Eq. (A.61): it is the last one on the right-hand side. According to (A.24), $\boldsymbol{\nabla T}$ is then expressed as

$$\boldsymbol{\nabla T} = \nabla_\gamma T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell}\, \boldsymbol{e}_{\alpha_1} \otimes \ldots \otimes \boldsymbol{e}_{\alpha_k} \otimes \boldsymbol{e}^{\beta_1} \otimes \ldots \otimes \boldsymbol{e}^{\beta_\ell} \otimes \boldsymbol{e}^\gamma. \qquad \text{(A.64)}$$

Because $\boldsymbol{e}^\gamma$ is the *last* 1-form of the tensorial product on the right-hand side, the notation $T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell;\gamma}$ instead of $\nabla_\gamma T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell}$ would have been more appropriate. The index convention (A.64) agrees with that of MTW [190] [cf. their Eq. (10.17)].

The ***covariant derivative of the tensor field $\boldsymbol{T}$ along a vector $\boldsymbol{v}$*** is defined by

$$\boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{T} := \boldsymbol{\nabla T}(\underbrace{.,\ldots,.}_{k+\ell \text{ slots}}, \boldsymbol{u}). \qquad \text{(A.65)}$$

The components of $\boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{T}$ are then $v^\mu \nabla_\mu T^{\alpha_1\ldots\alpha_k}{}_{\beta_1\ldots\beta_\ell}$. Note that $\boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{T}$ is a tensor field of the same type as $\boldsymbol{T}$. In the particular case of a scalar field f, we will use the notation $\boldsymbol{v} \cdot \boldsymbol{\nabla} f$ for $\boldsymbol{\nabla}_{\boldsymbol{v}} f$:

$$\boldsymbol{v} \cdot \boldsymbol{\nabla} f := \boldsymbol{\nabla}_{\boldsymbol{v}} f = \langle \boldsymbol{\nabla} f, \boldsymbol{v} \rangle = \boldsymbol{v}(f). \qquad \text{(A.66)}$$

The ***divergence*** with respect to the affine connection $\boldsymbol{\nabla}$ of a tensor field $\boldsymbol{T}$ of type (k, ℓ) with $k \geq 1$ is the tensor field denoted $\boldsymbol{\nabla} \cdot \boldsymbol{T}$ of type $(k-1, \ell)$ and whose components with respect to any frame field are given by

$$(\boldsymbol{\nabla} \cdot \boldsymbol{T})^{\alpha_1\ldots\alpha_{k-1}}{}_{\beta_1\ldots\beta_\ell} = \boldsymbol{\nabla}_\mu T^{\alpha_1\ldots\alpha_{k-1}\mu}{}_{\beta_1\ldots\beta_\ell}. \qquad \text{(A.67)}$$

Remark 3: For the divergence, the contraction is performed on the *last* upper index of $\boldsymbol{T}$.

Levi-Civita connection

On a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ there is a unique affine connection $\boldsymbol{\nabla}$ such that

1. $\boldsymbol{\nabla}$ is ***torsion-free***, i.e. for any scalar field f, $\boldsymbol{\nabla\nabla} f$ is a field of *symmetric* bilinear forms; in components:
$$\nabla_\alpha \nabla_\beta f = \nabla_\beta \nabla_\alpha f. \qquad \text{(A.68)}$$

2. The covariant derivative of the metric tensor vanishes identically:

$$\boxed{\boldsymbol{\nabla} \boldsymbol{g} = 0}. \tag{A.69}$$

$\boldsymbol{\nabla}$ is called the ***Levi-Civita connection associated with*** $\boldsymbol{g}$. In this book, we shall consider only such connections.

With respect to the Levi-Civita connection, the Levi-Civita tensor $\boldsymbol{\epsilon}$ introduced in Sec. A.3.4 shares the same property as $\boldsymbol{g}$:

$$\boxed{\boldsymbol{\nabla} \boldsymbol{\epsilon} = 0}. \tag{A.70}$$

Given a coordinate system (x^α) on $\mathscr{M}$, the connection coefficients of the Levi-Civita connection with respect to the natural basis $(\boldsymbol{\partial}_\alpha)$ are called the ***Christoffel symbols***; they can be evaluated from the partial derivatives of the metric components with respect to the coordinates:

$$\Gamma^\gamma{}_{\alpha\beta} = \frac{1}{2} g^{\gamma\mu} \left(\frac{\partial g_{\mu\beta}}{\partial x^\alpha} + \frac{\partial g_{\alpha\mu}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\mu} \right). \tag{A.71}$$

Note that the Christoffel symbols are symmetric with respect to the lower two indices.

For the Levi-Civita connection, the expression for the divergence of a vector takes a rather simple form in a natural basis associated with some coordinates (x^α). Indeed, combining Eqs. (A.67) and (A.63), we get for $\boldsymbol{v} \in \mathfrak{X}(\mathscr{M})$,

$$\boldsymbol{\nabla} \cdot \boldsymbol{v} = \nabla_\mu v^\mu = \frac{\partial v^\mu}{\partial x^\mu} + \Gamma^\mu{}_{\sigma\mu} v^\sigma .$$

Now, from (A.71), we have

$$\Gamma^\mu{}_{\alpha\mu} = \frac{1}{2} g^{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x^\alpha} = \frac{1}{2} \frac{\partial}{\partial x^\alpha} \ln |g| = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\alpha} \sqrt{|g|}, \tag{A.72}$$

where $g := \det(g_{\alpha\beta})$ [Eq. (A.50)]. The last but one equality follows from the general law of variation of the determinant of any invertible matrix A:

$$\boxed{\delta(\ln |\det A|) = \mathrm{tr}(A^{-1} \times \delta A)}, \tag{A.73}$$

where δ denotes any variation (derivative) that fulfills the Leibniz rule, tr stands for the trace and $\times$ for the matrix product. We conclude that

$$\boxed{\boldsymbol{\nabla} \cdot \boldsymbol{v} = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\mu} \left(\sqrt{|g|}\, v^\mu \right)}. \tag{A.74}$$

Similarly, for an antisymmetric tensor field of type $(2,0)$,

$$\nabla_\mu A^{\alpha\mu} = \frac{\partial A^{\alpha\mu}}{\partial x^\mu} + \underbrace{\Gamma^\alpha{}_{\sigma\mu} A^{\sigma\mu}}_{0} + \Gamma^\mu{}_{\sigma\mu} A^{\alpha\sigma} = \frac{\partial A^{\alpha\mu}}{\partial x^\mu} + \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\sigma} \sqrt{|g|}\, A^{\alpha\sigma},$$

where we have used the fact that $\Gamma^\alpha{}_{\sigma\mu}$ is symmetric in (σ, μ), whereas $A^{\sigma\mu}$ is antisymmetric. Hence the simple formula for the divergence of an *antisymmetric* tensor field of $(2,0)$:

$$\boxed{\nabla_\mu A^{\alpha\mu} = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\mu} \left(\sqrt{|g|}\, A^{\alpha\mu} \right)}. \tag{A.75}$$

A.4.2 Lie derivative

As discussed in Sec. A.4.1, the notion of a derivative of a vector field on a manifold $\mathscr{M}$ requires the introduction of some extra-structure on $\mathscr{M}$. In Sec. A.4.1, this extra-structure was an affine connection and in Sec. A.4.1 a metric $\boldsymbol{g}$ (which provides naturally an affine connection: the Levi-Civita one). Another possible extra-structure is a "reference" vector field, with respect to which the derivative is to be defined. This leads to the concept of the *Lie derivative*, which we discuss here.

Lie derivative of a vector field

Consider a vector field $\boldsymbol{u}$ on $\mathscr{M}$, called hereafter the ***flow***. Let $\boldsymbol{v}$ be another vector field on $\mathscr{M}$, the variation of which is to be studied. We can use the flow $\boldsymbol{u}$ to transport the vector $\boldsymbol{v}$ from one point p to a neighbouring one q and then define rigorously the variation of $\boldsymbol{v}$ as the difference between the actual value of $\boldsymbol{v}$ at q and the transported value via $\boldsymbol{u}$. More precisely the definition of the Lie derivative of $\boldsymbol{v}$ with respect to $\boldsymbol{u}$ is as follows (see Fig. A.3). We first define the image $\Phi_\varepsilon(p)$ of the point p by the transport by an infinitesimal "distance" ε along the field lines of $\boldsymbol{u}$ as $\Phi_\varepsilon(p) = q$, where q is the point close to p such that the infinitesimal displacement vector from p to q is $\overrightarrow{pq} = \varepsilon \boldsymbol{u}(p)$ (cf. Sec. A.2.3). We shall call the map $\Phi_\varepsilon : \mathscr{M} \to \mathscr{M}$ hence defined the ***flow map*** along $\boldsymbol{u}$. Besides, if we multiply the vector $\boldsymbol{v}(p)$ by some infinitesimal parameter λ, it becomes an infinitesimal vector at p. Then there exists a unique point p' close to p such that $\lambda \boldsymbol{v}(p) = \overrightarrow{pp'}$. We may transport the point p' to a point q' along the field lines of $\boldsymbol{u}$ by the same "distance" ε as that used to transport p to q: $q' = \Phi_\varepsilon(p')$ (see Fig. A.3). $\overrightarrow{qq'}$ is then an infinitesimal vector at q and we define the transport by the distance ε of the vector $\boldsymbol{v}(p)$ along the field lines of $\boldsymbol{u}$ according to

$$\Phi_\varepsilon^*(\boldsymbol{v}(p)) := \frac{1}{\lambda} \overrightarrow{qq'}. \tag{A.76}$$

$\Phi_\varepsilon^*(\boldsymbol{v}(p))$ is a vector in $T_q\mathscr{M}$. The map $\Phi_\varepsilon^* : T_p\mathscr{M} \to T_q\mathscr{M}$ hence defined is called the ***pushforward*** of the flow map Φ_ε. Actually it is nothing but the *differential* of the flow map $\Phi_\varepsilon : \mathscr{M} \to \mathscr{M}$, as defined in Sec. A.2.7:

$$\Phi_\varepsilon^*(\boldsymbol{v}(p)) = \mathrm{d}\Phi_\varepsilon|_p \left((\boldsymbol{v}(p)\right). \tag{A.77}$$

We may then subtract the vector $\Phi_\varepsilon^*(\boldsymbol{v}(p))$ from the actual value of the field $\boldsymbol{v}$ at $q = \Phi_\varepsilon(p)$ and define the ***Lie derivative*** of $\boldsymbol{v}$ along $\boldsymbol{u}$ at the point p by

$$\boxed{\boldsymbol{\mathcal{L}}_{\boldsymbol{u}} \boldsymbol{v} := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left[\boldsymbol{v} \left(\Phi_\varepsilon(p) \right) - \Phi_\varepsilon^* \left(\boldsymbol{v}(p) \right) \right]}. \tag{A.78}$$

Let us consider a coordinate system (x^α) adapted to the field $\boldsymbol{u}$ in the sense that $\boldsymbol{u} = \boldsymbol{\partial}_0$, where $\boldsymbol{\partial}_0$ is the first vector of the natural basis associated with the coordinates (x^α). We have, from the definitions of points q, p' and q',

$$\begin{aligned} x^\alpha(q) &= x^\alpha(p) + \varepsilon \delta^\alpha{}_0 \\ x^\alpha(p') &= x^\alpha(p) + \lambda v^\alpha(p) \\ x^\alpha(q') &= x^\alpha(p') + \varepsilon \delta^\alpha{}_0, \end{aligned}$$

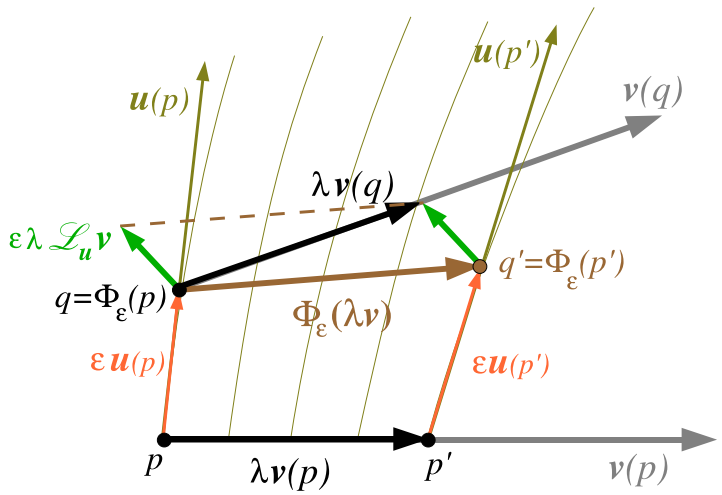


Figure A.3: Geometrical construction of the Lie derivative of a vector field $\boldsymbol{v}$ along a vector field $\boldsymbol{u}$: given a small parameter λ, each extremity of the arrow $\lambda\boldsymbol{v}$ is dragged by some small parameter ε along $\boldsymbol{u}$, to form the vector denoted by $\Phi_\varepsilon^*(\lambda\boldsymbol{v})$. The latter is then compared with the actual value of $\lambda\boldsymbol{v}$ at the point q, the difference (divided by $\lambda\varepsilon$) defining the Lie derivative $\boldsymbol{\mathcal{L}_u}\,\boldsymbol{v}$.

so that

$$(qq')^\alpha = x^\alpha(p') - x^\alpha(p) = \lambda v^\alpha(p).$$

Accordingly, (A.76) and (A.78) result in

$$\begin{aligned}(\boldsymbol{\mathcal{L}_u}\,\boldsymbol{v})^\alpha &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[v^\alpha(q) - v^\alpha(p)\right] \\ &= \lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\left[v^\alpha(x^0+\varepsilon,\ x^1,\ \ldots,\ x^{n-1}) - v^\alpha(x^0,\ x^1,\ \ldots,\ x^{n-1})\right].\end{aligned}$$

Hence, in adapted coordinates, the Lie derivative is simply obtained by taking the partial derivative of the vector components with respect to x^0:

$$\mathcal{L}_{\boldsymbol{u}}\,v^\alpha = \frac{\partial v^\alpha}{\partial x^0}, \tag{A.79}$$

where we have used the standard notation for the components of a Lie derivative: $\mathcal{L}_{\boldsymbol{u}}\,v^\alpha := (\boldsymbol{\mathcal{L}_u}\,\boldsymbol{v})^\alpha$. Besides, using the fact that the components of $\boldsymbol{u}$ are $u^\alpha = (1, 0, \ldots, 0)$ in the adapted coordinate system, we notice that the components of the commutator of $\boldsymbol{u}$ and $\boldsymbol{v}$, as given by (A.26), are

$$[\boldsymbol{u},\boldsymbol{v}]^\alpha = \frac{\partial v^\alpha}{\partial x^0}.$$

This is exactly (A.79): $[\boldsymbol{u},\boldsymbol{v}]^\alpha = \mathcal{L}_{\boldsymbol{u}}\,v^\alpha$. We conclude that the Lie derivative of a vector with respect to another one is actually nothing but the commutator of these two vectors:

$$\boxed{\boldsymbol{\mathcal{L}_u}\,\boldsymbol{v} = [\boldsymbol{u},\boldsymbol{v}]}. \tag{A.80}$$

Thanks to formula (A.26), we may then express the components of the Lie derivative in an arbitrary coordinate system:

$$\boxed{\mathcal{L}_{\boldsymbol{u}}\,v^\alpha = u^\mu\frac{\partial v^\alpha}{\partial x^\mu} - v^\mu\frac{\partial u^\alpha}{\partial x^\mu}}. \tag{A.81}$$

Thanks to the symmetry property of the Christoffel symbols, the partial derivatives in Eq. (A.81) can be replaced by the Levi-Civita connection $\boldsymbol{\nabla}$ associated with some metric $\boldsymbol{g}$, yielding

$$\mathcal{L}_{\boldsymbol{u}}\, v^\alpha = u^\mu \nabla_\mu v^\alpha - v^\mu \nabla_\mu u^\alpha. \tag{A.82}$$

Generalization to any tensor field

If $\boldsymbol{T}$ is tensor field of type $(0,\ell)$ on $\mathscr{M}$ (with $\ell \geq 1$) its ***pullback*** by the flow map Φ_ε is the tensor field $\Phi_\varepsilon^* \boldsymbol{T}$ of type $(0,\ell)$ defined by applying $\boldsymbol{T}$ to pushforwarded vectors:

$$\forall (\boldsymbol{v}_1, \ldots, \boldsymbol{v}_\ell) \in (T_p\mathscr{M})^\ell, \quad \Phi_\varepsilon^* \boldsymbol{T}|_p\,(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_\ell) := \boldsymbol{T}|_{\Phi_\varepsilon(p)}\,(\Phi_\varepsilon^*(\boldsymbol{v}_1), \ldots, \Phi_\varepsilon^*(\boldsymbol{v}_\ell))\,. \tag{A.83}$$

The ***Lie derivative*** of $\boldsymbol{T}$ along $\boldsymbol{u}$ is then defined by comparing the pullback image at some point p to the actual value of $\boldsymbol{\omega}$ at the same point:

$$\boxed{\mathcal{L}_{\boldsymbol{u}}\, \boldsymbol{T} := \lim_{\varepsilon \to 0} \frac{1}{\varepsilon} \left(\Phi_\varepsilon^* \boldsymbol{T} - \boldsymbol{T} \right)}. \tag{A.84}$$

Finally, the Lie derivative is extended to any tensor field by (i) demanding that for a scalar field f, $\mathcal{L}_{\boldsymbol{u}}\, f = \langle \boldsymbol{\nabla} f, \boldsymbol{u} \rangle$ and (ii) using the Leibniz rule. As a result, the Lie derivative $\mathcal{L}_{\boldsymbol{u}}\, \boldsymbol{T}$ of a tensor field $\boldsymbol{T}$ of type (k,ℓ) is a tensor field of the same type, the components of which with respect to a given coordinate system (x^α) are

$$\begin{aligned} \mathcal{L}_{\boldsymbol{u}}\, T^{\alpha_1 \ldots \alpha_k}{}_{\beta_1 \ldots \beta_\ell} &= u^\mu \frac{\partial}{\partial x^\mu} T^{\alpha_1 \ldots \alpha_k}{}_{\beta_1 \ldots \beta_\ell} - \sum_{i=1}^{k} T^{\alpha_1 \ldots \overset{i}{\downarrow} \sigma \ldots \alpha_k}{}_{\beta_1 \ldots \beta_\ell} \frac{\partial u^{\alpha_i}}{\partial x^\sigma} \\ &\quad + \sum_{i=1}^{\ell} T^{\alpha_1 \ldots \alpha_k}{}_{\beta_1 \ldots \underset{i}{\uparrow} \sigma \ldots \beta_\ell} \frac{\partial u^\sigma}{\partial x^{\beta_i}}. \end{aligned} \tag{A.85}$$

In particular, for a 1-form,

$$\mathcal{L}_{\boldsymbol{u}}\, \omega_\alpha = u^\mu \frac{\partial \omega_\alpha}{\partial x^\mu} + \omega_\mu \frac{\partial u^\mu}{\partial x^\alpha}. \tag{A.86}$$

As for the vector case [Eq. (A.81)], the partial derivatives in Eq. (A.85) can be replaced by the covariant derivative $\boldsymbol{\nabla}$ (or any other connection without torsion), yielding

$$\begin{aligned} \mathcal{L}_{\boldsymbol{u}}\, T^{\alpha_1 \ldots \alpha_k}{}_{\beta_1 \ldots \beta_\ell} &= u^\mu \nabla_\mu T^{\alpha_1 \ldots \alpha_k}{}_{\beta_1 \ldots \beta_\ell} - \sum_{i=1}^{k} T^{\alpha_1 \ldots \overset{i}{\downarrow} \sigma \ldots \alpha_k}{}_{\beta_1 \ldots \beta_\ell} \nabla_\sigma u^{\alpha_i} \\ &\quad + \sum_{i=1}^{\ell} T^{\alpha_1 \ldots \alpha_k}{}_{\beta_1 \ldots \underset{i}{\uparrow} \sigma \ldots \beta_\ell} \nabla_{\beta_i} u^\sigma. \end{aligned} \tag{A.87}$$

In adapted coordinates, we have, similarly to Eq. (A.79),

$$\mathcal{L}_{\boldsymbol{u}}\, T^{\alpha_1 \ldots \alpha_k}{}_{\beta_1 \ldots \beta_\ell} = \frac{\partial}{\partial x^0} T^{\alpha_1 \ldots \alpha_k}{}_{\beta_1 \ldots \beta_\ell} \qquad \text{(coordinates adapted to } \boldsymbol{u}\text{)}. \tag{A.88}$$

Note that this formula is a direct consequence of (A.85) since in adapted coordinates, $u^\alpha = (1, 0, \ldots, 0)$, so that $u^\mu \partial/\partial x^\mu = \partial/\partial x^0$ and $\partial u^\alpha/\partial x^\beta = 0$.

Remark 4: We use the same notation Φ_ε^* to denote the pushforward and pullback operators associated with the differentiable map Φ_ε. In the mathematical literature (e.g. [176]), the pushforward operator is usually denoted with the star as a subscript rather than a superscript, i.e. $(\Phi_\varepsilon)_*$, the superscript notation being reserved for the pullback operator.

A.4.3 Exterior derivative

In Sec. A.2.6, we have introduced the *differential forms* as tensor fields of type $(0, q)$, with $q \geq 0$, that are antisymmetric in all their arguments as soon as $q \geq 2$. Otherwise stating, at each point $p \in \mathscr{M}$, a differential form results in a fully antisymmetric multilinear form on the vector space $T_p\mathscr{M}$. A differential form of order q is also called a q-form.

Differential forms play a special role in the theory of integration on a manifold. Indeed, the primary definition of an integral over a manifold of dimension n is the integral of a n-form. The Levi-Civita tensor $\boldsymbol{\epsilon}$ introduced in Sec. A.3.4 is a n-form, whose integral gives the volume with respect to the metric $\boldsymbol{g}$. Regarding physics, it is well known that the electromagnetic field is fundamentally a 2-form (the Faraday tensor $\boldsymbol{F}$); in relativistic hydrodynamics, the vorticity of a fluid is also described by a 2-form (see e.g. [123]).

Being tensor fields, differential forms are subject to the covariant and Lie derivations discussed above. But, in addition, they are subject to a third type of derivation, called *exterior derivation*. The ***exterior derivative*** of a q-form $\boldsymbol{\omega}$ is a $(q+1)$-form which is denoted $\mathbf{d}\boldsymbol{\omega}$ and whose components with respect to a given coordinate system (x^α) are defined by

$$\text{0-form (scalar field)} \quad : \quad (\mathbf{d}\boldsymbol{\omega})_\alpha := \frac{\partial \omega}{\partial x^\alpha} \tag{A.89}$$

$$\text{1-form} \quad : \quad (\mathbf{d}\boldsymbol{\omega})_{\alpha\beta} := \frac{\partial \omega_\beta}{\partial x^\alpha} - \frac{\partial \omega_\alpha}{\partial x^\beta} \tag{A.90}$$

$$\text{2-form} \quad : \quad (\mathbf{d}\boldsymbol{\omega})_{\alpha\beta\gamma} := \frac{\partial \omega_{\beta\gamma}}{\partial x^\alpha} + \frac{\partial \omega_{\gamma\alpha}}{\partial x^\beta} + \frac{\partial \omega_{\alpha\beta}}{\partial x^\gamma} \tag{A.91}$$

$$\text{etc...}$$

It can be easily checked that these formulæ, although expressed in terms of partial derivatives of components in a coordinate system, do define tensor fields. Moreover, the result is clearly antisymmetric (assuming that $\boldsymbol{\omega}$ is), so that we end up with $(q+1)$-forms. Notice that for a scalar field (0-form), the exterior derivative is nothing but the differential 1-form $\mathbf{d}f$ already defined in Sec. A.2.4. Notice also that the definition of the exterior derivative appeals only to the manifold structure. It does not depend upon the metric tensor $\boldsymbol{g}$, nor upon any other extra structure on $\mathscr{M}$.

Remark 5: Although the exterior derivative does not depend on the metric $\boldsymbol{g}$ and hence on the Levi-Civita connection $\boldsymbol{\nabla}$, one may replace all partial derivatives in the formulæ (A.89)-(A.91)

by covariant derivatives:

$$\text{0-form} \quad : \quad (\mathbf{d}\boldsymbol{\omega})_\alpha = \nabla_\alpha \omega \tag{A.92}$$

$$\text{1-form} \quad : \quad (\mathbf{d}\boldsymbol{\omega})_{\alpha\beta} = \nabla_\alpha \omega_\beta - \nabla_\beta \omega_\alpha \tag{A.93}$$

$$\text{2-form} \quad : \quad (\mathbf{d}\boldsymbol{\omega})_{\alpha\beta\gamma} = \nabla_\alpha \omega_{\beta\gamma} + \nabla_\beta \omega_{\gamma\alpha} + \nabla_\gamma \omega_{\alpha\beta} \tag{A.94}$$

The above identities hold thanks to the symmetry of the Christoffel symbols on their last two indices (and thanks to (A.21) for a 0-form).

A fundamental property of the exterior derivation is to be nilpotent:

$$\boxed{\mathbf{d}\mathbf{d}\boldsymbol{\omega} = 0}. \tag{A.95}$$

A q-form $\boldsymbol{\omega}$ is said to be *closed* iff $\mathbf{d}\boldsymbol{\omega} = 0$, and *exact* iff there exists a $(q-1)$-form $\boldsymbol{\sigma}$ such that $\boldsymbol{\omega} = \mathbf{d}\boldsymbol{\sigma}$. From property (A.95), an exact q-form is closed. The Poincaré lemma states that the converse is true, at least locally.

The exterior derivative enters in the well known *Stokes' theorem*: if $\mathcal{D}$ is a submanifold of $\mathscr{M}$ of dimension d that has a boundary (denoted $\partial\mathcal{D}$), then for any $(d-1)$-form $\boldsymbol{\omega}$,

$$\oint_{\partial\mathcal{D}} \boldsymbol{\omega} = \int_{\mathcal{D}} \mathbf{d}\boldsymbol{\omega}. \tag{A.96}$$

Note that $\partial\mathcal{D}$ is a manifold of dimension $d-1$ and $\mathbf{d}\boldsymbol{\omega}$ is a d-form, so that each side of (A.96) is (of course !) a well-defined quantity, as the integral of a q-form over a q-dimensional manifold.

Another very important formula where the exterior derivative enters is the ***Cartan identity***, which states that the Lie derivative of a differential form $\boldsymbol{\omega}$ along a vector field $\boldsymbol{u}$ is expressible as

$$\boxed{\boldsymbol{\mathcal{L}}_{\boldsymbol{u}}\, \boldsymbol{\omega} = \boldsymbol{u} \cdot \mathbf{d}\boldsymbol{\omega} + \mathbf{d}(\boldsymbol{u} \cdot \boldsymbol{\omega})}. \tag{A.97}$$

In the above formula, a dot denotes the contraction on adjacent indices, i.e. $\boldsymbol{u} \cdot \boldsymbol{\omega}$ is the $(q-1)$-form $\boldsymbol{\omega}(\boldsymbol{u}, ., \ldots, .)$, with the $q-1$ last slots remaining free. Notice that in the case of a 1-form, Eq. (A.97) is readily obtained by combining Eqs. (A.86), (A.89) and (A.90).

A.5 Curvature

A.5.1 General definition

The ***Riemann curvature tensor*** of an affine connection $\boldsymbol{\nabla}$ is defined by

$$\begin{array}{rccc} \mathbf{Riem} \ : & \mathfrak{X}^*(\mathscr{M}) \times \mathfrak{X}(\mathscr{M})^3 & \longrightarrow & C^\infty(\mathscr{M}, \mathbb{R}) \\ & (\boldsymbol{\omega}, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{v}) & \longmapsto & \Big\langle \boldsymbol{\omega},\ \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{w} - \boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{\nabla}_{\boldsymbol{u}} \boldsymbol{w} - \boldsymbol{\nabla}_{[\boldsymbol{u},\boldsymbol{v}]} \boldsymbol{w} \Big\rangle, \end{array} \tag{A.98}$$

where $\mathfrak{X}^*(\mathscr{M})$ stands for the space of 1-forms on $\mathscr{M}$, $\mathfrak{X}(\mathscr{M})$ for the space of vector fields on $\mathscr{M}$ and $C^\infty(\mathscr{M}, \mathbb{R})$ for the space of smooth scalar fields on $\mathscr{M}$. The above formula

does define a tensor field on $\mathscr{M}$, i.e. the value of $\mathbf{Riem}(\boldsymbol{\omega}, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{v})$ at a given point $p \in \mathscr{M}$ depends only upon the values of the fields $\boldsymbol{\omega}$, $\boldsymbol{w}$, $\boldsymbol{u}$ and $\boldsymbol{v}$ at p and not upon their behaviours away from p, as the covariant derivatives in Eq. (A.98) might suggest. We denote the components of this tensor in a given basis $(\boldsymbol{e}_\alpha)$, not by $\mathrm{Riem}^\gamma{}_{\delta\alpha\beta}$, but by $R^\gamma{}_{\delta\alpha\beta}$. The definition (A.98) leads then to the following expression, named the ***Ricci identity***:

$$\forall \boldsymbol{w} \in \mathfrak{X}(\mathscr{M}), \quad (\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha)\, w^\gamma = R^\gamma{}_{\mu\alpha\beta}\, w^\mu. \tag{A.99}$$

Remark 1: In view of this identity, one may say that the Riemann tensor measures the lack of commutativity of two successive covariant derivatives of a vector field. On the opposite, for a scalar field and a torsion-free connection, two successive covariant derivatives always commute [cf. Eq. (A.68)].

In a coordinate basis, the components of the Riemann tensor are given in terms of the connection coefficients by

$$\boxed{R^\alpha{}_{\beta\mu\nu} = \frac{\partial \Gamma^\alpha{}_{\beta\nu}}{\partial x^\mu} - \frac{\partial \Gamma^\alpha{}_{\beta\mu}}{\partial x^\nu} + \Gamma^\alpha{}_{\sigma\mu}\Gamma^\sigma{}_{\beta\nu} - \Gamma^\alpha{}_{\sigma\nu}\Gamma^\sigma{}_{\beta\mu}}. \tag{A.100}$$

From the definition (A.98), the Riemann tensor is clearly antisymmetric with respect to its last two arguments $(\boldsymbol{u}, \boldsymbol{v})$:

$$\mathbf{Riem}(.,.,\boldsymbol{u},\boldsymbol{v}) = -\mathbf{Riem}(.,.,\boldsymbol{v},\boldsymbol{u}). \tag{A.101}$$

In addition, it satisfies the cyclic property

$$\mathbf{Riem}(.,\boldsymbol{u},\boldsymbol{v},\boldsymbol{w}) + \mathbf{Riem}(.,\boldsymbol{w},\boldsymbol{u},\boldsymbol{v}) + \mathbf{Riem}(.,\boldsymbol{v},\boldsymbol{w},\boldsymbol{u}) = 0. \tag{A.102}$$

The covariant derivatives of the Riemann tensor obeys the ***Bianchi identity***

$$\boxed{\nabla_\rho R^\alpha{}_{\beta\mu\nu} + \nabla_\mu R^\alpha{}_{\beta\nu\rho} + \nabla_\nu R^\alpha{}_{\beta\rho\mu} = 0}. \tag{A.103}$$

A.5.2 Case of a pseudo-Riemannian manifold

The Riemann tensor of the Levi-Civita connection obeys the additional antisymmetry:

$$\mathbf{Riem}(\boldsymbol{\omega}, \boldsymbol{w}, .,.) = -\mathbf{Riem}(\underline{\boldsymbol{w}}, \overrightarrow{\boldsymbol{\omega}}, .,.). \tag{A.104}$$

Combined with (A.101) and (A.102), this implies the symmetry property

$$\mathbf{Riem}(\boldsymbol{\omega}, \boldsymbol{w}, \boldsymbol{u}, \boldsymbol{v}) = \mathbf{Riem}(\underline{\boldsymbol{u}}, \boldsymbol{v}, \overrightarrow{\boldsymbol{\omega}}, \boldsymbol{w}). \tag{A.105}$$

A pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ with a vanishing Riemann tensor is called a ***flat manifold***; in this case, $\boldsymbol{g}$ is said to be a ***flat metric***. If in addition, it has a Riemannian signature, $\boldsymbol{g}$ is called an ***Euclidean metric***.

A.5.3 Ricci tensor

The ***Ricci tensor*** of the affine connection $\boldsymbol{\nabla}$ is the field of bilinear forms $\boldsymbol{R}$ defined by

$$\begin{array}{rccc} \boldsymbol{R} \; : & \mathfrak{X}(\mathscr{M}) \times \mathfrak{X}(\mathscr{M}) & \longrightarrow & C^\infty(\mathscr{M}, \mathbb{R}) \\ & (\boldsymbol{u}, \boldsymbol{v}) & \longmapsto & \mathbf{Riem}(\boldsymbol{e}^\mu, \boldsymbol{u}, \boldsymbol{e}_\mu, \boldsymbol{v}), \end{array} \tag{A.106}$$

where $(\boldsymbol{e}_\alpha)$ is a vector frame on $\mathscr{M}$ and $(\boldsymbol{e}^\alpha)$ its dual counterpart. This definition is independent of the choice of $(\boldsymbol{e}_\alpha)$. In terms of components:

$$\boxed{R_{\alpha\beta} := R^\mu{}_{\alpha\mu\beta}}. \tag{A.107}$$

Remark 2: Following the standard usage, we denote the components of the Riemann and Ricci tensors by the same letter R, the number of indices allowing us to distinguish between the two tensors. On the other hand, we are using different symbols, **Riem** and $\boldsymbol{R}$, when employing the 'intrinsic' notation.

For the Levi-Civita connection associated with the metric $\boldsymbol{g}$, the property (A.105) implies that the Ricci tensor is symmetric:

$$\boldsymbol{R}(\boldsymbol{u}, \boldsymbol{v}) = \boldsymbol{R}(\boldsymbol{v}, \boldsymbol{u}). \tag{A.108}$$

In addition, one defines the ***Ricci scalar*** (also called ***scalar curvature***) as the trace of the Ricci tensor with respect to the metric $\boldsymbol{g}$:

$$R := g^{\mu\nu} R_{\mu\nu}. \tag{A.109}$$

The Bianchi identity (A.103) implies the divergence-free property

$$\boxed{\boldsymbol{\nabla} \cdot \overrightarrow{\boldsymbol{G}} = 0}, \tag{A.110}$$

where $\overrightarrow{\boldsymbol{G}}$ in the type-$(1,1)$ tensor associated by metric duality [cf. (A.44)] to the ***Einstein tensor***:

$$\boxed{\boldsymbol{G} := \boldsymbol{R} - \frac{1}{2} R \, \boldsymbol{g}}. \tag{A.111}$$

Equation (A.110) is called the ***contracted Bianchi identity***.

A.5.4 Weyl tensor

Let $(\mathscr{M}, \boldsymbol{g})$ be a pseudo-Riemannian manifold of dimension n.

For $n = 1$, the Riemann tensor vanishes identically, i.e. $(\mathscr{M}, \boldsymbol{g})$ is necessarily flat. The reader who has in mind a curved line in the Euclidean plane $\mathbb{R}^2$ might be surprised by the above statement. This is because the Riemann tensor represents the *intrinsic* curvature of a manifold. For a curve, the curvature that is not vanishing is the *extrinsic* curvature, i.e. the curvature resulting from the embedding of the curve in $\mathbb{R}^2$.

For $n = 2$, the Riemann tensor is entirely determined by the knowledge of the Ricci scalar R, according to the formula:

$$R^{\gamma}{}_{\delta\alpha\beta} = \frac{R}{2}\left(\delta^{\gamma}{}_{\alpha}\, g_{\delta\beta} - \delta^{\gamma}{}_{\beta}\, g_{\delta\alpha}\right) \qquad (n = 2). \tag{A.112}$$

For $n = 3$, the Riemann tensor is entirely determined by the knowledge of the Ricci tensor, according to

$$\begin{aligned} R^{\gamma}{}_{\delta\alpha\beta} &= R^{\gamma}{}_{\alpha}\, g_{\delta\beta} - R^{\gamma}{}_{\beta}\, g_{\delta\alpha} + \delta^{\gamma}{}_{\alpha}\, R_{\delta\beta} - \delta^{\gamma}{}_{\beta}\, R_{\delta\alpha} \\ &\quad + \frac{R}{2}\left(\delta^{\gamma}{}_{\beta}\, g_{\delta\alpha} - \delta^{\gamma}{}_{\alpha}\, g_{\delta\beta}\right) \qquad (n = 3). \end{aligned} \tag{A.113}$$

For $n \geq 4$, the Riemann tensor can be split into (i) a "trace-trace" part, represented by the Ricci scalar R [Eq. (A.109)], (ii) a "trace" part, represented by the Ricci tensor $\boldsymbol{R}$ [Eq. (A.107)], and (iii) a "traceless" part, which is constituted by the ***Weyl conformal curvature tensor***, $\boldsymbol{C}$:

$$\begin{aligned} R^{\gamma}{}_{\delta\alpha\beta} &= C^{\gamma}{}_{\delta\alpha\beta} + \frac{1}{n-2}\left(R^{\gamma}{}_{\alpha}\, g_{\delta\beta} - R^{\gamma}{}_{\beta}\, g_{\delta\alpha} + \delta^{\gamma}{}_{\alpha}\, R_{\delta\beta} - \delta^{\gamma}{}_{\beta}\, R_{\delta\alpha}\right) \\ &\quad + \frac{1}{(n-1)(n-2)} R\left(\delta^{\gamma}{}_{\beta}\, g_{\delta\alpha} - \delta^{\gamma}{}_{\alpha}\, g_{\delta\beta}\right). \end{aligned} \tag{A.114}$$

The above relation may be taken as the definition of $\boldsymbol{C}$. It implies that $\boldsymbol{C}$ is traceless: $C^{\mu}{}_{\alpha\mu\beta} = 0$. The other possible traces are zero thanks to the symmetry properties of the Riemann tensor.

Remark 3: The decomposition (A.114) is also meaningful for $n = 3$, but it then implies that the Weyl tensor vanishes identically [compare with (A.113)].

Appendix B

Geodesics

Contents

B.1 Introduction

Geodesics play a key role in general relativity, since they represent the worldlines of test particles and photons (cf. Sec. 1.3). Moreover, in black hole theory, null geodesics play a prominent role, as the generators of event horizons (cf. Sec. 4.4.3). We review here the definition and main properties of geodesics on a generic pseudo-Riemannian manifold, i.e. a manifold equipped with a metric of generic signature, as introduced in Sec. A.3. In particular, the results apply to pure Riemannian manifolds (positive definite metric), as well as to Lorentzian manifolds, i.e. spacetimes. Contrary to Appendix A, proofs of most statements will be provided, since they are quite illustrative.

B.2 Definition and first properties

B.2.1 Geodesics and affine parametrizations

On a Riemannian manifold, i.e. when the metric tensor is positive definite (cf. Sec. A.3.2), a geodesic is the curve of minimal length between two points (at least for close enough points). It is also a curve which is "as straight as possible", in the sense that its tangent vectors are transported parallelly to themselves along it. A typical example is a geodesic in the Euclidean space: this is nothing but a straight line, for which tangent vectors keep

obviously a fixed direction. In a *pseudo*-Riemannian manifold, such as the spacetimes of general relativity, one uses this last property to define geodesics.

Let us first recall the basic definitions given in Sec. A.2.3: a ***curve***[1] is the image $\mathscr{L} = P(I)$ of a map (called a ***parametrization*** of the curve) $P : I \to \mathscr{M}$, $\lambda \mapsto P(\lambda)$, where I is an interval of $\mathbb{R}$ and the variable λ is called a ***parameter*** of the curve. Moreover, we exclude the case where $\mathscr{L}$ is reduced to a single point of $\mathscr{M}$, i.e. where P is a constant function. We are now in position to define a geodesic as a "straight" curve:

A smooth curve $\mathscr{L}$ of a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ is called a ***geodesic*** iff it admits a parametrization P whose associated tangent vector field $\boldsymbol{v}$ is transported parallelly to itself along $\mathscr{L}$:

$$\boxed{\boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{v} = 0}, \tag{B.1}$$

where $\boldsymbol{\nabla}$ is the Levi-Civita connection of the metric $\boldsymbol{g}$. The parametrization P is then called an ***affine parametrization*** and the corresponding parameter λ is called an ***affine parameter*** of the geodesic $\mathscr{L}$. Note that the relation between $\boldsymbol{v}$ and λ is

$$\boldsymbol{v} = \frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}, \tag{B.2}$$

where $\mathrm{d}\boldsymbol{x}$ is the infinitesimal displacement along $\mathscr{L}$ corresponding to the change $\mathrm{d}\lambda$ in the parameter λ (cf. Eq. (A.13)).

The qualifier *affine* in the above definition stems from the following property:

Any two affine parametrizations of a geodesic $\mathscr{L}$ are necessarily related by an affine transformation:

$$\lambda' = a\lambda + b, \tag{B.3}$$

where a and b are two real constants such that $a \neq 0$.

Proof. Let $P : I \to \mathscr{L} \subset \mathscr{M}$, $\lambda \mapsto P(\lambda)$ and $P' : I' \to \mathscr{L}$, $\lambda' \mapsto P'(\lambda')$ be two parametrizations of $\mathscr{L}$. They are necessarily related by a diffeomorphism $I \to I'$, $\lambda \mapsto \lambda'(\lambda)$. It follows from Eq. (B.2) that the tangent vector fields $\boldsymbol{v}$ and $\boldsymbol{v'}$ associated with these two parametrizations are related by

$$\boldsymbol{v} = \frac{\mathrm{d}\lambda'}{\mathrm{d}\lambda}\, \boldsymbol{v'}. \tag{B.4}$$

Using the rules 2 and 3 governing the connection $\boldsymbol{\nabla}$ (cf. Sec. A.4.1), we get then

$$\boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{v} = \frac{\mathrm{d}^2\lambda'}{\mathrm{d}\lambda^2}\, \boldsymbol{v'} + \left(\frac{\mathrm{d}\lambda'}{\mathrm{d}\lambda}\right)^2 \boldsymbol{\nabla}_{\boldsymbol{v'}} \boldsymbol{v'}. \tag{B.5}$$

If both parametrizations are affine, then $\boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{v} = 0$ and $\boldsymbol{\nabla}_{\boldsymbol{v'}} \boldsymbol{v'} = 0$, so that the above identity reduces to $\mathrm{d}^2\lambda'/\mathrm{d}\lambda^2 = 0$, which implies the affine law (B.3). □

[1] As already noticed (cf. Remark 5 p. 481), in the mathematical literature, it is common to define a curve as the parametrization itself, and not as its image.

Remark 1: Because of (B.1), a geodesic is also called an ***autoparallel curve***. It is also sometimes called a ***zero-acceleration curve***, the vector $\boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{v}$ being considered as the "acceleration" of (the parametrization P of) the curve $\mathscr{L}$; this is of course by extension of the concept of 4-acceleration $\boldsymbol{a} := \boldsymbol{\nabla}_{\boldsymbol{u}}\boldsymbol{u}$ of a timelike worldline with 4-velocity $\boldsymbol{u}$, the latter being nothing but the tangent vector associated with the parametrization of the worldline by its proper time (cf. Sec. 1.3.3).

An important property of geodesics is

Let $\mathscr{L}$ be a geodesic of $(\mathscr{M}, \boldsymbol{g})$ and $\boldsymbol{v}$ a tangent vector field associated with an affine parametrization of $\mathscr{L}$. Then the scalar square of $\boldsymbol{v}$ with respect to the metric $\boldsymbol{g}$ is constant along $\mathscr{L}$:

$$\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v}) = \text{const.} \tag{B.6}$$

Proof. The variation of $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})$ along $\mathscr{L}$ is given by

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})\right) &= \boldsymbol{v}\left(\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})\right) = \boldsymbol{\nabla}_{\boldsymbol{v}}\left(\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})\right) = v^\mu \nabla_\mu (g_{\rho\sigma} v^\rho v^\sigma) \\ &= v^\mu \underbrace{\nabla_\mu g_{\rho\sigma}}_{0} v^\rho v^\sigma + g_{\rho\sigma} \underbrace{v^\mu \nabla_\mu v^\rho}_{0} v^\sigma + g_{\rho\sigma} v^\rho \underbrace{v^\mu \nabla_\mu v^\sigma}_{0} = 0, \end{aligned}$$

where we have used the fact that $\boldsymbol{\nabla}$ is the Levi-Civita connection of $\boldsymbol{g}$ [Eq. (A.69)] and $\boldsymbol{v}$ obeys the geodesic equation (B.1). □

The constancy of $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})$ has an interesting corollary: the tangent vector $\boldsymbol{v}$ cannot change its type along $\mathscr{L}$. Hence:

In a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$, a geodesic $\mathscr{L}$ belongs necessarily to one of the following three categories:

- ***timelike geodesic***: tangent vectors are timelike at all points of $\mathscr{L}$;
- ***null geodesic***: tangent vectors are null at all points of $\mathscr{L}$;
- ***spacelike geodesic***: tangent vectors are spacelike at all points of $\mathscr{L}$.

This is in sharp contrast with generic curves, which, for instance, can be timelike on some portions and null or spacelike on other parts.

In the timelike case, or the spacelike one, the tangent vector field $\boldsymbol{v}$ can be rescaled by the constant $\sqrt{|\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})|}$ to get a unit tangent vector field, i.e. a tangent vector field $\boldsymbol{u}$ which obeys $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = -1$ (timelike geodesic) or $\boldsymbol{g}(\boldsymbol{u}, \boldsymbol{u}) = 1$ (spacelike geodesic). Moreover, in doing so, the affine character of the parametrization is preserved. Indeed, the rescaling amounts to choosing the constant a in the affine law (B.3) such that $a = \sqrt{|\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})|}$. Thus, we have

A timelike or spacelike geodesic of a Lorentzian manifold has an affine parametrization, the tangent vector of which is a unit vector (i.e. of scalar square ± 1 with respect to $\boldsymbol{g}$). Moreover, this parametrization is unique up to some choice of origin (choice of b in (B.3)) and of orientation ($a = \pm 1$ in (B.3)).

We shall see in Sec. B.2.2 that for a timelike geodesic, the affine parameter with unit tangent vector is nothing but the *proper time*, while for a spacelike geodesic, it is the *arc length*.

B.2.2 Generic parametrizations of geodesics

Geodesics can be characterized by any of their tangent vectors, i.e. tangent vectors not necessarily associated with an affine parametrization, as follows:

A curve $\mathscr{L}$ is a geodesic iff the tangent vector field $\boldsymbol{v}$ associated with any parametrization of $\mathscr{L}$ obeys

$$\boldsymbol{\nabla}_{\boldsymbol{v}} \boldsymbol{v} = \kappa \, \boldsymbol{v}, \tag{B.7}$$

where κ is a scalar field along $\mathscr{L}$.

Proof. Let $P : I \to \mathscr{L}$, $\lambda \mapsto P(\lambda)$ be the parametrization of $\mathscr{L}$ corresponding to the tangent vector field $\boldsymbol{v}$: $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$. If $\mathscr{L}$ is a geodesic, then there exists a parametrization $\lambda' \mapsto P'(\lambda')$ whose tangent vector, $\boldsymbol{v'}$ say, obeys $\boldsymbol{\nabla}_{\boldsymbol{v'}} \boldsymbol{v'} = 0$. Since the accelerations of any two parametrizations of $\mathscr{L}$ are related by Eq. (B.5), we deduce that $\boldsymbol{v}$ obeys (B.7) with

$$\kappa = \left(\frac{\mathrm{d}\lambda'}{\mathrm{d}\lambda} \right)^{-1} \frac{\mathrm{d}^2 \lambda'}{\mathrm{d}\lambda^2}.$$

Conversely, if $\boldsymbol{v}$ obeys (B.7) with $\kappa = \kappa(\lambda)$, then Eq. (B.5) implies that $\boldsymbol{\nabla}_{\boldsymbol{v'}} \boldsymbol{v'} = 0$, i.e. that $\mathscr{L}$ is a geodesic, provided that the change of parametrization $\lambda' = \lambda'(\lambda)$ fulfils

$$\kappa(\lambda) \frac{\mathrm{d}\lambda'}{\mathrm{d}\lambda} - \frac{\mathrm{d}^2 \lambda'}{\mathrm{d}\lambda^2} = 0.$$

This differential equation has the following solution:

$$\lambda' = a \int_{\lambda_1}^{\lambda} \left[\exp \left(\int_{\lambda_0}^{\tilde{\lambda}} \kappa(\tilde{\tilde{\lambda}}) \mathrm{d}\tilde{\tilde{\lambda}} \right) \mathrm{d}\tilde{\lambda} \right] + b,$$

where a, b, λ_0 and λ_1 are constants, with $a \neq 0$ and $\lambda_0, \lambda_1 \in I$. □

The above property motivates the following definitions:

A vector field $\boldsymbol{v}$ obeying (B.7) is called a ***pregeodesic vector field***. The scalar field κ is then called the ***non-affinity coefficient*** of $\boldsymbol{v}$. If $\kappa = 0$, $\boldsymbol{v}$ is naturally called a ***geodesic vector field***.

Note that the property established above is equivalent to stating that the field lines of a pregeodesic vector field are geodesics.

An easy consequence of Eq. (B.7) is the following evolution law for the scalar square of the tangent vector:

Along a geodesic $\mathscr{L}$, the scalar square $\boldsymbol{g}(\boldsymbol{v},\boldsymbol{v})$ of the tangent vector $\boldsymbol{v}$ associated with any parametrization of $\mathscr{L}$ evolves according to

$$\boldsymbol{\nabla}_{\boldsymbol{v}} \left[\boldsymbol{g}(\boldsymbol{v},\boldsymbol{v})\right] = 2\kappa\, \boldsymbol{g}(\boldsymbol{v},\boldsymbol{v}), \tag{B.8}$$

where κ is the non-affinity coefficient of $\boldsymbol{v}$.

Proof. One has, using $\boldsymbol{\nabla g} = 0$ [Eq. (A.69)] and Eq. (B.7),

$$v^\mu \nabla_\mu (g_{\rho\sigma} v^\rho v^\sigma) = v^\mu \underbrace{\nabla_\mu g_{\rho\sigma}}_{0} v^\rho v^\sigma + g_{\rho\sigma} \underbrace{v^\mu \nabla_\mu v^\rho}_{\kappa v^\rho} v^\sigma + g_{\rho\sigma} v^\rho \underbrace{v^\mu \nabla_\mu v^\sigma}_{\kappa v^\sigma} = 2\kappa g_{\rho\sigma} v^\rho v^\sigma,$$

hence the law (B.8). □

We recover of course (B.6) in the special case $\kappa = 0$ ($\boldsymbol{v}$ geodesic vector).

Remark 2: If λ is the parameter associated with $\boldsymbol{v}$, i.e. $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$, we may introduce the scalar function $V(\lambda) := \boldsymbol{g}(\boldsymbol{v},\boldsymbol{v})$ and rewrite (B.8) as a first-order ordinary differential equation for it [cf. Eq. (A.8)]:

$$\frac{\mathrm{d}V}{\mathrm{d}\lambda} = 2\kappa(\lambda) V(\lambda). \tag{B.9}$$

A consequence of (B.8) is

On a Lorentzian manifold, the parametrization of a timelike geodesic by the proper time ($\lambda = \tau$) is an affine parametrization. Similarly, on a Lorentzian or Riemannian manifold, the parametrization of a spacelike geodesic by the arc length ($\lambda = s$) is an affine parametrization.

Proof. The tangent vector associated with the proper time τ along a timelike geodesic is nothing but the 4-velocity $\boldsymbol{u}$ (cf. Sec. 1.3.3), which is of constant scalar square: $\boldsymbol{g}(\boldsymbol{u},\boldsymbol{u}) = -1$, so that Eq. (B.8) reduces to $0 = -2\kappa$, hence $\kappa = 0$, which implies that we are dealing with an affine parametrization. Similarly, the tangent vector associated with the arc length s along a spacelike geodesic has a scalar square everywhere equal to 1, leading to the same conclusion. □

B.3 Existence and uniqueness of geodesics

B.3.1 The geodesic equation

Let $\mathscr{L}$ be a curve in a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ of dimension n, such that $\mathscr{L}$ is contained in the domain of a coordinate chart $(x^\alpha)_{0\leq\alpha\leq n-1}$. Then any parametrization of $\mathscr{L}$, $P : I \to \mathscr{L}$, $\lambda \mapsto P(\lambda)$, can be described by n functions $X^\alpha : I \to \mathbb{R}$ according to Eq. (A.7): $x^\alpha(P(\lambda)) = X^\alpha(\lambda)$. The curve $\mathscr{L}$ is a geodesic iff there exists a parametrization of $\mathscr{L}$ for which the functions X^α fulfils the following system of n second-order differential equations, called the ***geodesic equation***:

$$\boxed{\frac{\mathrm{d}^2X^\alpha}{\mathrm{d}\lambda^2} + \Gamma^\alpha{}_{\mu\nu}\frac{\mathrm{d}X^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}X^\nu}{\mathrm{d}\lambda} = 0}, \qquad 0 \leq \alpha \leq n-1, \tag{B.10}$$

where the $\Gamma^\alpha{}_{\mu\nu}$'s are the Christoffel symbols of the metric $\boldsymbol{g}$ with respect to the coordinates (x^α), as given by Eq. (A.71).

Proof. Notice first that the components with respect to the chart (x^α) of the tangent vector field $\boldsymbol{v}$ associated with the parameter λ are [cf. Eq. (A.11)]

$$v^\alpha = \frac{\mathrm{d}X^\alpha}{\mathrm{d}\lambda}. \tag{B.11}$$

On the other side, the components of $\boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{v}$ are

$$v^\mu\nabla_\mu v^\alpha = v^\mu\frac{\partial v^\alpha}{\partial x^\mu} + \Gamma^\alpha{}_{\mu\nu}v^\mu v^\nu = \boldsymbol{v}(v^\alpha) + \Gamma^\alpha{}_{\mu\nu}v^\mu v^\nu = \frac{\mathrm{d}v^\alpha}{\mathrm{d}\lambda} + \Gamma^\alpha{}_{\mu\nu}v^\mu v^\nu,$$

where we have used successively Eqs. (A.63), (A.10) and (A.8). The above relation, along with (B.11), shows that the left-hand side of Eq. (B.10) is nothing but the component α of $\boldsymbol{\nabla}_{\boldsymbol{v}}\boldsymbol{v}$. The conclusion follows from the very definition of a geodesic given in Sec. B.2.1. □

Note that if a solution of the geodesic equation (B.10) is found, the parameter λ is necessarily an affine parameter. For a generic parameter, the differential equation becomes (B.10) with the right-hand side replaced by $\kappa\mathrm{d}X^\alpha/\mathrm{d}\lambda$, which is the coordinate expression of the right-hand side $\kappa\boldsymbol{v}$ in Eq. (B.7). Hence, we have

A curve $\mathscr{L}$ in the domain of a chart (x^α) is a geodesic iff some (actually all) coordinate expression $x^\alpha = X^\alpha(\lambda)$ of $\mathscr{L}$ fulfils the following system of n second-order differential equations, usually called the ***pregeodesic equation***,

$$\boxed{\frac{\mathrm{d}^2X^\alpha}{\mathrm{d}\lambda^2} + \Gamma^\alpha{}_{\mu\nu}\frac{\mathrm{d}X^\mu}{\mathrm{d}\lambda}\frac{\mathrm{d}X^\nu}{\mathrm{d}\lambda} = \kappa(\lambda)\frac{\mathrm{d}X^\alpha}{\mathrm{d}\lambda}}, \qquad 0 \leq \alpha \leq n-1. \tag{B.12}$$

for some real-valued function $\kappa(\lambda)$.

B.3.2 Existence and uniqueness

We may use the geodesic equation to prove the following existence and uniqueness theorem:

Given a point p in a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ and a vector $\boldsymbol{V}$ in the tangent space to $\mathscr{M}$ at p, i.e. $\boldsymbol{V} \in T_p\mathscr{M}$, there exists a geodesic $\mathscr{L}$ through p such that $\boldsymbol{V}$ is the value at p of the tangent vector of some affine parametrization of $\mathscr{L}$:

$$\boldsymbol{V} = \left.\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\lambda}\right|_p . \tag{B.13}$$

Moreover, this geodesic is unique, in the sense that any geodesic $\mathscr{L}'$ sharing the same property coincides with $\mathscr{L}$ in some open neighbourhood of p.

Proof. Let (x^α) be a coordinate chart of $\mathscr{M}$ around p. Let (V^α) be the components of $\boldsymbol{V}$ in the basis of $T_p\mathscr{M}$ induced by the coordinate frame $(\boldsymbol{\partial}_\alpha)$ associated with (x^α):

$$\boldsymbol{V} = V^\alpha \left.\boldsymbol{\partial}_\alpha\right|_p .$$

A geodesic through p having $\boldsymbol{V}$ as tangent vector at p is then obtained as a solution $(X^\alpha(\lambda))$ of the system (B.10) with the initial conditions [cf. Eq. (B.11)]

$$X^\alpha(0) = x^\alpha(p) \quad \text{and} \quad \frac{\mathrm{d}X^\alpha}{\mathrm{d}\lambda}(0) = V^\alpha. \tag{B.14}$$

The system (B.10) + (B.14) constitutes a well-posed Cauchy problem and standard results about ordinary differential equations, e.g. the Picard-Lindelöf (or Cauchy–Lipschitz) theorem, guarantee the existence and uniqueness of the solution. □

A few definitions follow naturally:

A geodesic $\mathscr{L}$ is said to be ***inextendible*** or ***maximal*** iff there does not exist any geodesic $\mathscr{L}'$ such that $\mathscr{L} \subset \mathscr{L}'$ and $\mathscr{L}' \neq \mathscr{L}$.

A geodesic $\mathscr{L}$ is ***complete*** iff the interval spanned by any of its affine parameters is the whole real line: $I = \mathbb{R}$. A geodesic that is not complete is called ***incomplete***.

It is easy to show that

Any complete geodesic is inextendible.

Proof. Let $\mathscr{L}$ be a complete geodesic. Let us consider any geodesic $\mathscr{L}'$ such that $\mathscr{L} \subset \mathscr{L}'$. Let λ and λ' be affine parameters of respectively $\mathscr{L}$ and $\mathscr{L}'$. Since $\mathscr{L} \subset \mathscr{L}'$, λ' is also an affine parameter of $\mathscr{L}$ and we must have, along $\mathscr{L}$, $\lambda' = a\lambda + b$ with $a \neq 0$ [Eq. (B.3)]. Since the range of λ is $(-\infty, +\infty)$, for $\mathscr{L}$ is complete, this implies that the range of λ' on $\mathscr{L}$ is $(-\infty, +\infty)$ as well, which make impossible to have points in $\mathscr{L}' \setminus \mathscr{L}$. Hence $\mathscr{L}' = \mathscr{L}$, i.e. $\mathscr{L}$ is inextendible. □

On physical grounds, one may consider that any timelike geodesic in a given spacetime must be complete. Otherwise, this would mean that there exists a worldline $\mathscr{L}$ of a freely falling observer that "ends" at some finite proper time. This would be the signature of either (i) the possibility to extend the spacetime into a larger one or (ii) the ending of worldline $\mathscr{L}$ at some (curvature) singularity. A spacetime in which this does not occur is called timelike geodesically complete:

The pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ is said to be ***geodesically complete*** iff every inextendible geodesic is complete.

Remark 1: A well-known theorem of differential geometry, namely the ***Hopf-Rinow theorem***, states that a connected *Riemannian* manifold is geodesically complete iff it is *complete* as a *metric space* for the distance function $d(p, q)$ defined as the infimum of the length[2] of all curves from p to q (see e.g. Ref. [177]). However, there is no such theorem for a *Lorentzian* manifold, for the metric does not induce any distance function turning the manifold into a metric space.

The following proposition strengthens the existence and uniqueness result obtained above:

Given a point p in a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ and a nonzero vector $\boldsymbol{V}$ in the tangent space to $\mathscr{M}$ at p, i.e. $\boldsymbol{V} \in T_p\mathscr{M}$, there exists a unique inextendible geodesic through p, which we shall denote by $\mathscr{L}_{\boldsymbol{V}}$, such that $\boldsymbol{V}$ is the value at p of the tangent vector of some affine parametrization of $\mathscr{L}_{\boldsymbol{V}}$. We shall then denote by $P_{\boldsymbol{V}}$ the unique affine parametrization of $\mathscr{L}_{\boldsymbol{V}}$ such that

$$P_{\boldsymbol{V}}(0) = p \quad \text{and} \quad \boldsymbol{v}|_p = \boldsymbol{V}, \tag{B.15}$$

where $\boldsymbol{v}$ is the tangent vector field of $P_{\boldsymbol{V}}$.

We refer to O'Neill's textbook [199], p. 68 for the proof.

B.3.3 Exponential map

One can make use of geodesics to map a tangent space to the base manifold:

Given a point p in a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$, let $\mathcal{E}_p$ be the subset of the tangent space $T_p\mathscr{M}$ defined by $\boldsymbol{V} \in \mathcal{E}_p$ iff either $\boldsymbol{V} = 0$ or the affine parametrization $P_{\boldsymbol{V}}$ of the geodesic $\mathscr{L}_{\boldsymbol{V}}$ has a domain large enough to include the interval $[0, 1]$. The

[2] The length of a curve is defined by Eq. (B.24) below.

exponential map at p is then defined as

$$\begin{aligned} \exp_p : \ \mathcal{E}_p \subset T_p\mathscr{M} &\longrightarrow \mathscr{M} \\ \boldsymbol{V} &\longmapsto \begin{cases} p & \text{if } \boldsymbol{V} = 0 \\ P_{\boldsymbol{V}}(1) \in \mathscr{L}_{\boldsymbol{V}} & \text{if } \boldsymbol{V} \neq 0 \end{cases} \end{aligned} \tag{B.16}$$

In other words, $\exp_p$ maps a nonzero vector $\boldsymbol{V}$ in the tangent space to $\mathscr{M}$ at p to the point of $\mathscr{M}$ of affine parameter $\lambda = 1$ along the unique geodesic through p, the parameter λ being such that (i) $\lambda = 0$ corresponds to p and (ii) the associated tangent vector $\mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ at p is $\boldsymbol{V}$.

Note that if $(\mathscr{M}, \boldsymbol{g})$ is geodesically complete, $\mathcal{E}_p = T_p\mathscr{M}$ for every point $p \in \mathscr{M}$.

An immediate property of the exponential map is

If $\boldsymbol{V} \in \mathcal{E}_p \setminus \{0\}$, for any $t \in [0,1]$, $\exp_p(t\boldsymbol{V})$ lies on the same geodesic $\mathscr{L}_{\boldsymbol{V}}$ as $\exp_p(\boldsymbol{V})$, at the parameter $\lambda = t$ of the parametrization $P_{\boldsymbol{V}}$:

$$\forall t \in [0,1], \quad \exp_p(t\boldsymbol{V}) = P_{\boldsymbol{V}}(t). \tag{B.17}$$

Proof. For $t = 0$, the property follows from the definition of $\exp_p$, since $P_{\boldsymbol{V}}(0) = p$. If $t \neq 0$, the nonzero vector $t\boldsymbol{V}$ is collinear to $\boldsymbol{V}$ and the uniqueness property of geodesics (cf. Sec. B.3.2) implies that $\mathscr{L}_{t\boldsymbol{V}} = \mathscr{L}_{\boldsymbol{V}}$. By virtue of the transformation law (B.4), $t\boldsymbol{V}$ is the tangent vector to $\mathscr{L}_{\boldsymbol{V}}$ corresponding to the affine parameter $\lambda' = t^{-1}\lambda$, where λ is the affine parameter whose tangent vector field obeys $\boldsymbol{v}|_p = \boldsymbol{V}$. From the definition of $\exp_p$, we have then $\exp_p(t\boldsymbol{V}) = P_{t\boldsymbol{V}}(\lambda' = 1) = P_{\boldsymbol{V}}(\lambda = t \times 1)$, hence (B.17). □

The exponential map realizes a local identification of the manifold with its tangent space at a given point:

For any $p \in \mathscr{M}$, there exists a neighbourhood U of 0 in the tangent space $T_p\mathscr{M}$ and a neighbourhood $\mathscr{U}$ of p in the manifold $\mathscr{M}$ such that the exponential map $\exp_p$ is a diffeomorphism from U to $\mathscr{U}$.

Proof. It is clear from its definition that $\exp_p$ is a smooth map, at least on some neighbourhood U' of 0 in $T_p\mathscr{M}$. We may then consider the differential of $\exp_p$ at 0, $\mathrm{d}\exp_p\big|_0$. By virtue of the inverse function theorem for manifolds (see e.g. Theorem 4.5 in Ref. [176]), it suffices to show that $\mathrm{d}\exp_p\big|_0$ is invertible to complete the proof. By definition of the differential of a map (cf. Sec. A.2.7) and since $\exp_p : \mathcal{E}_p \subset T_p\mathscr{M} \to \mathscr{M}$ and $\exp_p(0) = p$, $\mathrm{d}\exp_p\big|_0$ carries an infinitesimal displacement vector of[3] $T_0(T_p\mathscr{M})$, $\boldsymbol{\varepsilon}$ say, connecting 0 to a nearby element of $T_p\mathscr{M}$, $\boldsymbol{\varepsilon}'$ say, to the infinitesimal vector $\boldsymbol{E} \in T_p\mathscr{M}$ connecting $p = \exp_p(0)$ to $q = \exp_p(\boldsymbol{\varepsilon}')$:

$$\begin{aligned} \mathrm{d}\exp_p\big|_0 : \ T_0(T_p\mathscr{M}) &\longrightarrow T_p\mathscr{M} \\ \boldsymbol{\varepsilon} &\longmapsto \boldsymbol{E} = \overrightarrow{pq}. \end{aligned}$$

[3] Here the vector space $T_p\mathscr{M}$ is considered as a n-dimensional smooth manifold, and $T_0(T_p\mathscr{M})$ stands for its tangent space at 0 (the zero vector of $T_p\mathscr{M}$).

Now, since $T_p\mathscr{M}$ is a vector space, we have the canonical identification $T_0(T_p\mathscr{M}) \simeq T_p\mathscr{M}$, from which $\boldsymbol{\varepsilon'} = \boldsymbol{\varepsilon}$. Without any loss of generality, we may write $\boldsymbol{\varepsilon} = \varepsilon \boldsymbol{V}$, where ε is infinitesimal small and $\boldsymbol{V} \in T_p\mathscr{M}$. Then $q = \exp_p(\varepsilon \boldsymbol{V}) = P_{\boldsymbol{V}}(\varepsilon)$, where the second identity results from (B.17). We have thus

$$\mathrm{d}\exp_p\big|_0 (\varepsilon \boldsymbol{V}) = \boldsymbol{E} = \overrightarrow{pq} = \overrightarrow{P_{\boldsymbol{V}}(0) P_{\boldsymbol{V}}(\varepsilon)}.$$

According to the definition of $P_{\boldsymbol{V}}$, the infinitesimal vector $\overrightarrow{P_{\boldsymbol{V}}(0) P_{\boldsymbol{V}}(\varepsilon)}$ along the geodesic $\mathscr{L}_{\boldsymbol{V}}$ is $\varepsilon \boldsymbol{V}$, hence

$$\mathrm{d}\exp_p\big|_0 (\varepsilon \boldsymbol{V}) = \varepsilon \boldsymbol{V}.$$

Since the differential $\mathrm{d}\exp_p\big|_0$ is a linear map, we get $\mathrm{d}\exp_p\big|_0 (\boldsymbol{V}) = \boldsymbol{V}$. The vector $\boldsymbol{V}$ being arbitrary, we conclude that $\mathrm{d}\exp_p\big|_0$ is nothing but the identity map of the vector space $T_p\mathscr{M}$:

$$\mathrm{d}\exp_p\big|_0 = \mathrm{id}_{T_p\mathscr{M}}\,.$$

In particular, $\mathrm{d}\exp_p\big|_0$ is invertible. □

B.3.4 Normal coordinates

Given $p \in \mathscr{M}$, a ***normal neighbourhood*** of p is a neighbourhood $\mathscr{U}$ of p that is the image of a starshaped neighbourhood of $0 \in T_p\mathscr{M}$ under the local diffeomorphism $\exp_p$ given by the above proposition. By ***starshaped neighbourhood*** of 0, it is meant a neighbourhood U of 0 such that $\boldsymbol{V} \in U$ implies $t\boldsymbol{V} \in U$ for any $t \in [0,1]$.

On a normal neighbourhood, one may define coordinates linked to geodesics as follows.

Let $\mathscr{U}$ be a normal neighbourhood of $p \in \mathscr{M}$ and $(\boldsymbol{E}_\alpha)_{0 \leq \alpha \leq n-1}$ be a basis of $T_p\mathscr{M}$. If $(\boldsymbol{E}^\alpha)$ stands for the basis of $T_p^*\mathscr{M}$ dual to $(\boldsymbol{E}_\alpha)$, the map

$$\begin{aligned} \Phi : \ \mathscr{U} &\longrightarrow \mathbb{R}^n \\ q &\longmapsto \big(\langle \boldsymbol{E}^0, \exp_p^{-1}(q)\rangle, \ldots, \langle \boldsymbol{E}^{n-1}, \exp_p^{-1}(q)\rangle\big) \end{aligned} \tag{B.18}$$

is a coordinate chart on $\mathscr{U}$, which is called ***geodesic normal coordinates***, often shorten as ***normal coordinates***.

In other words, normal coordinates (x^α) on $\mathscr{U}$ are such that the tangent vector $x^\mu(q)\boldsymbol{E}_\mu \in T_p\mathscr{M}$ has precisely q as image by the exponential map:

$$\forall q \in \mathscr{U}, \quad \exp_p(x^\mu(q)\boldsymbol{E}_\mu) = q. \tag{B.19}$$

Remark 2: Some authors, e.g. [199], add the condition that the basis $(\boldsymbol{E}_\alpha)$ is orthonormal (with respect to the metric $\boldsymbol{g}$) in the definition of normal coordinates. We follow here the more general definition of [167, 26, 146, 237]. The name ***Riemann normal coordinates*** is also commonly

encountered in the literature, either for normal coordinates as defined here (e.g. [190, 256]) or for those with the basis orthonormality requirement (e.g. [216]).

A characteristic feature of normal coordinates is that, in terms of them, geodesics through p look like straight lines through 0 in $\mathbb{R}^n$:

In a normal coordinate system (x^α), the equation of the unique geodesic $\mathscr{L}_{\boldsymbol{V}}$ through p admitting $\boldsymbol{V} \in T_p\mathscr{M}$ as tangent vector at p is (as long as $\mathscr{L}_{\boldsymbol{V}}$ remains in the normal neighbourhood $\mathscr{U}$)

$$x^\alpha = X^\alpha(\lambda) = \lambda V^\alpha, \tag{B.20}$$

where the V^α's are the components of $\boldsymbol{V}$ with respect to the basis $(\boldsymbol{E}_\alpha)$ defining the normal coordinates: $\boldsymbol{V} = V^\mu \boldsymbol{E}_\mu$.

Proof. Let λ be the affine parameter of $\mathscr{L}_{\boldsymbol{V}}$ corresponding to the parametrization $P_{\boldsymbol{V}}$. The coordinate equation of $\mathscr{L}_{\boldsymbol{V}}$ is then $x^\alpha = X^\alpha(\lambda)$ with [cf. Eq. (B.18)]

$$X^\alpha(\lambda) = x^\alpha\left(P_{\boldsymbol{V}}(\lambda)\right) = \langle \boldsymbol{E}^\alpha, \exp_p^{-1}\left(P_{\boldsymbol{V}}(\lambda)\right)\rangle.$$

Now, according to Eq. (B.17), $P_{\boldsymbol{V}}(\lambda) = \exp_p(\lambda \boldsymbol{V})$. Hence

$$X^\alpha(\lambda) = \langle \boldsymbol{E}^\alpha, \exp_p^{-1} \circ \exp_p(\lambda \boldsymbol{V})\rangle = \langle \boldsymbol{E}^\alpha, \lambda \boldsymbol{V}\rangle = \lambda \langle \boldsymbol{E}^\alpha, \boldsymbol{V}\rangle = \lambda V^\alpha.$$

□

Let (x^α) be a normal coordinate system around $p \in \mathscr{M}$ associated with a basis $(\boldsymbol{E}_\alpha)$ of $T_p\mathscr{M}$. Then

- the coordinate frame $(\boldsymbol{\partial}_\alpha)$ associated with (x^α) coincides with $(\boldsymbol{E}_\alpha)$ at p:

$$\boldsymbol{\partial}_\alpha|_p = \boldsymbol{E}_\alpha; \tag{B.21}$$

- the values at p of the components $(g_{\alpha\beta})$ of the metric tensor $\boldsymbol{g}$ with respect to (x^α) are

$$g_{\alpha\beta}(p) = \boldsymbol{g}|_p\,(\boldsymbol{E}_\alpha, \boldsymbol{E}_\beta); \tag{B.22}$$

- the Christoffel symbols of $\boldsymbol{g}$ with respect to the coordinates (x^α) vanish at p:

$$\Gamma^\alpha{}_{\beta\gamma}(p) = 0. \tag{B.23}$$

Proof. Let $\mathscr{U}$ be the normal neighbourhood covered by (x^α) and $\boldsymbol{V} \in \exp_p^{-1}(\mathscr{U}) \subset T_p\mathscr{M}$. The tangent vector field to the geodesic $\mathscr{L}_{\boldsymbol{V}}$ corresponding to the parametrization $P_{\boldsymbol{V}}$ is $\boldsymbol{v} = \dot{X}^\mu \boldsymbol{\partial}_\mu$ with $\dot{X}^\mu$ obtained by deriving (B.20) with respect to λ: $\dot{X}^\mu = V^\mu$. Hence $\boldsymbol{v} = V^\mu \boldsymbol{\partial}_\mu$. Now, from the very definition of $P_{\boldsymbol{V}}$, $\boldsymbol{v}|_p = \boldsymbol{V} = V^\mu \boldsymbol{E}_\mu$. We have therefore

$$V^\mu\, \boldsymbol{\partial}_\mu|_p = V^\mu \boldsymbol{E}_\mu.$$

This identity being fulfilled for any V^μ, Eq. (B.21) follows. Equation. (B.22) is an immediate consequence of Eq. (B.21), since $g_{\alpha\beta} = \boldsymbol{g}(\boldsymbol{\partial}_\alpha, \boldsymbol{\partial}_\beta)$. Finally, with the functions $X^\alpha(\lambda)$ given by (B.20), the geodesic equation (B.10) reduces to ${\Gamma^\alpha}_{\mu\nu} V^\mu V^\nu = 0$. In particular, at p, we get

$$ {\Gamma^\alpha}_{\mu\nu}(p) V^\mu V^\nu = 0. $$

This identity must hold for any V^α. It expresses therefore that, for each value of α, the quadratic form $\boldsymbol{V} \mapsto {\Gamma^\alpha}_{\mu\nu}(p) V^\mu V^\nu$ is identically zero on $T_p\mathscr{M}$. Since the Christoffel symbols ${\Gamma^\alpha}_{\mu\nu}$ are symmetric in $\mu\nu$, it is equivalent to say that, for each value of α, the symmetric bilinear form $(\boldsymbol{U}, \boldsymbol{V}) \mapsto {\Gamma^\alpha}_{\mu\nu}(p) U^\mu V^\nu$ is identically zero, which amounts to ${\Gamma^\alpha}_{\mu\nu}(p) = 0$, i.e. Eq. (B.23). □

B.4 Geodesics and variation of length

B.4.1 Length of a curve

Geodesics in a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ have been defined in Sec. B.2.1 as the "straightest lines", i.e. as autoparallel curves with respect to the Levi-Civita connection of $\boldsymbol{g}$. Here, we make some attempt to connect them with the first feature mentioned in Sec. B.2.1, namely, in a pure Riemannian manifold, geodesics are locally the curves of *minimal length*. We have first to define the length of a curve. Of course, when the metric is not definite positive, one cannot use the integral of the norm of infinitesimal displacements along the curve, i.e. $\mathrm{d}s := \sqrt{\boldsymbol{g}(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{x})}$, since $\boldsymbol{g}(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{x})$ can be negative. Rather, it is quite natural to employ instead $\mathrm{d}s := \sqrt{|\boldsymbol{g}(\mathrm{d}\boldsymbol{x}, \mathrm{d}\boldsymbol{x})|}$. Using $\mathrm{d}\boldsymbol{x} = \boldsymbol{v}\,\mathrm{d}\lambda$ [Eq. (A.13)], we end up with the following definition:

The ***length*** of a curve $\mathscr{L}$ connecting two points p and q in a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$ is

$$ L_{(p,q)}(\mathscr{L}) := \int_{\lambda_p}^{\lambda_q} \sqrt{|\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v})|}\, \mathrm{d}\lambda, \tag{B.24} $$

where λ is some parameter along $\mathscr{L}$, λ_p (resp. λ_q) being its value at p (resp. q), $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ is the associated tangent vector field, and we assume $\lambda_q \geq \lambda_p$.

Thanks to the transformation law (B.4), it is easy to check that the value of $L_{(p,q)}(\mathscr{L})$ is independent from the choice of the parametrization of $\mathscr{L}$, i.e. for a fixed pair of points (p, q), it is a function of $\mathscr{L}$ only.

When $\mathscr{L}$ is included in the domain of a coordinate chart (x^α), so that its equation is $x^\alpha = X^\alpha(\lambda)$, we may rewrite (B.24) as [cf. Eq. (B.11)]

$$ L_{(p,q)}(\mathscr{L}) := \int_{\lambda_p}^{\lambda_q} \sqrt{\left| g_{\mu\nu}(X^\rho(\lambda)) \dot{X}^\mu \dot{X}^\nu \right|}\, \mathrm{d}\lambda, \tag{B.25} $$

where $\dot{X}^\alpha := \mathrm{d}X^\alpha/\mathrm{d}\lambda$ and $g_{\mu\nu}(X^\rho(\lambda))$ stands for the components of the metric tensor $\boldsymbol{g}$ with respect to the coordinates (x^α) at the point of coordinates $X^\rho(\lambda)$.

From the very definition of $L_{(p,q)}(\mathscr{L})$, it is obvious that

$$L_{(p,q)}(\mathscr{L}) \geq 0. \tag{B.26}$$

Moreover, if it exists, any null curve from p to q achieves the absolute minimum of the length, without having to be a geodesic:

$$\mathscr{L} \text{ null } \Longrightarrow L_{(p,q)}(\mathscr{L}) = 0. \tag{B.27}$$

B.4.2 Timelike and spacelike geodesics as stationary points of the length functional

The property (B.27) implies that, in a pseudo-Riemannian manifold, the curve that minimizes the length between two points is not necessarily a geodesic. A typical example is the null helix in Minkowski spacetime, discussed in Remark 2 on p. 35. Moreover, when $\boldsymbol{g}$ is not positive definite, it could be relevant to consider curves of *maximal* length between two points, i.e. to search for an extremum, be it a minimum or a maximum.

To find the curves of extremal length, it is quite natural to study the behaviour of the length as a variational problem, i.e. to consider $L_{(p,q)}(\mathscr{L})$ as an "action" and to write the Euler-Lagrange equation for the "Lagrangian" defined as the integrand of (B.25):

$$\mathcal{L}(X^\alpha, \dot{X}^\alpha) = \sqrt{\left| g_{\mu\nu}(X^\rho)\dot{X}^\mu \dot{X}^\nu \right|}. \tag{B.28}$$

Before proceeding, a few caveats must be made. First of all, the Euler-Lagrange equation locate only *stationary* points of the action (here the length $L_{(p,q)}(\mathscr{L})$), i.e. points where the action does not vary to first order in small changes of the curve. Such points are not necessarily extrema: they can be *saddle* points, as we shall see. Secondly, because of the square root in (B.28), the Lagrangian is not differentiable at points where $g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu = 0$. This corresponds to points where the curve $\mathscr{L}$ is null. We shall therefore exclude such curves in our analysis (we shall return to null curves in Sec. B.4.3). But then $g_{\mu\nu}\dot{X}^\mu \dot{X}^\nu$ has to be either always positive along $\mathscr{L}$ (i.e. $\mathscr{L}$ is spacelike) or always negative (i.e. $\mathscr{L}$ is timelike); indeed, by continuity it cannot change sign without going through zero. We shall then apply the variational principle separately to two subsets of curves connecting p and q: the timelike ones and the spacelike ones. The calculations will be conducted in parallel by introducing the sign parameter $\epsilon = -1$ for timelike curves and $\epsilon = +1$ for spacelike ones. One can then get rid of the absolute value in the Lagrangian, which becomes

$$\mathcal{L}(X^\alpha, \dot{X}^\alpha) = \sqrt{\epsilon g_{\mu\nu}(X^\rho)\dot{X}^\mu \dot{X}^\nu}. \tag{B.29}$$

Asking that the length (B.25) is stationary with respect to small changes in the curve connecting p and q is equivalent to the Euler-Lagrange equation:

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial \mathcal{L}}{\partial \dot{X}^\alpha}\right) - \frac{\partial \mathcal{L}}{\partial X^\alpha} = 0. \tag{B.30}$$

We have

$$\frac{\partial}{\partial X^\alpha}\left(g_{\mu\nu}(X^\rho)\dot{X}^\mu\dot{X}^\nu\right) = \frac{\partial g_{\mu\nu}}{\partial x^\alpha}\dot{X}^\mu\dot{X}^\nu, \tag{B.31}$$

with the understanding that $\partial g_{\mu\nu}/\partial x^\alpha$ shall be taken at the point $X^\rho(\lambda)$. Hence, given the Lagrangian (B.29),

$$\frac{\partial \mathcal{L}}{\partial X^\alpha} = \frac{\epsilon}{2\mathcal{L}}\frac{\partial g_{\mu\nu}}{\partial x^\alpha}\dot{X}^\mu\dot{X}^\nu. \tag{B.32}$$

Besides,

$$\frac{\partial}{\partial \dot{X}^\alpha}\left(g_{\mu\nu}(X^\rho)\dot{X}^\mu\dot{X}^\nu\right) = g_{\alpha\nu}\dot{X}^\nu + g_{\mu\alpha}\dot{X}^\mu = 2g_{\alpha\mu}\dot{X}^\mu. \tag{B.33}$$

Hence

$$\frac{\partial \mathcal{L}}{\partial \dot{X}^\alpha} = \frac{\epsilon}{\mathcal{L}}g_{\alpha\mu}\dot{X}^\mu,$$

from which,

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial \mathcal{L}}{\partial \dot{X}^\alpha}\right) = -\frac{\epsilon}{\mathcal{L}^2}\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\lambda}g_{\alpha\mu}\dot{X}^\mu + \frac{\epsilon}{\mathcal{L}}\frac{\partial g_{\alpha\mu}}{\partial x^\nu}\dot{X}^\nu\dot{X}^\mu + \frac{\epsilon}{\mathcal{L}}g_{\alpha\mu}\ddot{X}^\mu. \tag{B.34}$$

In view of (B.32) and (B.34), the Euler-Lagrange equation (B.30) becomes, after multiplication by $\mathcal{L}/\epsilon$,

$$-\frac{1}{\mathcal{L}}\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\lambda}g_{\alpha\mu}\dot{X}^\mu + \frac{\partial g_{\alpha\mu}}{\partial x^\nu}\dot{X}^\mu\dot{X}^\nu + g_{\alpha\mu}\ddot{X}^\mu - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\alpha}\dot{X}^\mu\dot{X}^\nu = 0.$$

Now, playing with the names of repeated indices and using the symmetry of $g_{\alpha\beta}$, we can rewrite the second term as

$$\frac{\partial g_{\alpha\mu}}{\partial x^\nu}\dot{X}^\mu\dot{X}^\nu = \frac{1}{2}\left(\frac{\partial g_{\alpha\nu}}{\partial x^\mu}\dot{X}^\nu\dot{X}^\mu + \frac{\partial g_{\mu\alpha}}{\partial x^\nu}\dot{X}^\mu\dot{X}^\nu\right) = \frac{1}{2}\left(\frac{\partial g_{\alpha\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\alpha}}{\partial x^\nu}\right)\dot{X}^\mu\dot{X}^\nu. \tag{B.35}$$

Accordingly, we get

$$g_{\alpha\mu}\ddot{X}^\mu + \frac{1}{2}\left(\frac{\partial g_{\alpha\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\alpha}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\alpha}\right)\dot{X}^\mu\dot{X}^\nu = \kappa g_{\alpha\mu}\dot{X}^\mu, \tag{B.36}$$

where we have introduced

$$\kappa := \frac{1}{\mathcal{L}}\frac{\mathrm{d}\mathcal{L}}{\mathrm{d}\lambda}. \tag{B.37}$$

If we multiply Eq. (B.36) by the matrix $g^{\alpha\beta}$ (the components of the inverse metric) and use $g^{\alpha\beta}g_{\alpha\mu} = \delta^\beta{}_\mu$ as well as the expression (A.71) of the Christoffel symbols, we get exactly the pregeodesic equation (B.12). Hence we conclude

> Among all timelike (resp. spacelike) curves connecting two points p and q, a curve has a stationary length iff it is a timelike (resp. spacelike) geodesic.

For a timelike geodesic, and for points p and q not too far (in the same normal neighbourhood), the stationary length corresponds actually to a *maximum*:

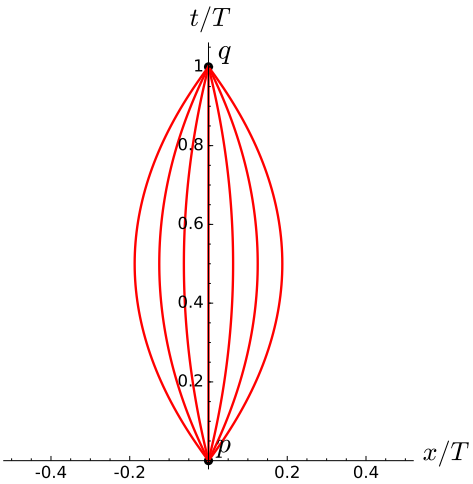


Figure B.1: Timelike curves $\mathscr{L}_h$ connecting the point p of coordinates $(0,0,0,0)$ to the point q of coordinates $(T,0,0,0)$ in Minkowski spacetime. From the left to right, the depicted curves correspond to h spanning $[-3/4, 3/4]$, with the step $\delta h = 1/4$.

Let $(\mathscr{M}, \boldsymbol{g})$ be a Lorentzian manifold, $p \in \mathscr{M}$ and $\mathscr{U}$ some normal neighbourhood[a] of p. For any point $q \in \mathscr{U}$ such that there exists a timelike curve in $\mathscr{U}$ from p to q, the geodesic from p to q is the unique timelike curve of largest length in $\mathscr{U}$ connecting p to q.

[a] See Sec. B.3.4.

We shall not provide a full proof here but refer instead to the proof of Proposition 5.34 in O'Neill's textbook [199]. We shall only illustrate the property on a specific example in flat spacetime (Example 1 below).

If one interprets timelike curves as worldlines and the length as the proper time (cf. Sec. 1.3.3), the above maximum can be viewed as a generalization of the standard "twin paradox" of special relativity: when they meet again, the twin who followed the geodesic (i.e. some inertial trajectory) is older than his brother, who made a round trip.

Example 1 (Timelike geodesic in Minkowski spacetime): Let us suppose that $(\mathscr{M}, \boldsymbol{g})$ is the 4-dimensional Minkowski spacetime. All geodesics are then (segments of) straight lines. If p and q are connected by a timelike geodesic $\mathscr{L}$, we may consider a Minkowskian coordinate system $(x^\alpha) = (t, x, y, z)$ such that $x^\alpha(p) = (0,0,0,0)$ and $x^\alpha(q) = (T,0,0,0)$, for some $T > 0$. t is then the proper time along $\mathscr{L}$ and $L_{(p,q)}(\mathscr{L}) = T$. Let us consider the one-parameter family of curves $(\mathscr{L}_h)_{h\in(-1,1)}$ defined by $x^\alpha = X^\alpha(\lambda)$ with $\lambda \in [0,T]$ and

$$X^0(\lambda) = \lambda, \quad X^1(\lambda) = \frac{h}{T}\lambda(T-\lambda), \quad X^2(\lambda) = 0, \quad X^3(\lambda) = 0.$$

Note that $X^0(\lambda) = \lambda$ means that the curve parameter coincides with the time coordinate: $\lambda = t$. We have $\mathscr{L}_0 = \mathscr{L}$ and for $h \neq 0$, $\mathscr{L}$ is an arc of parabola from p to q in the (t,x) plane (cf. Fig. B.1); the dimensionless parameter h is related to the curve's maximal extension along x by

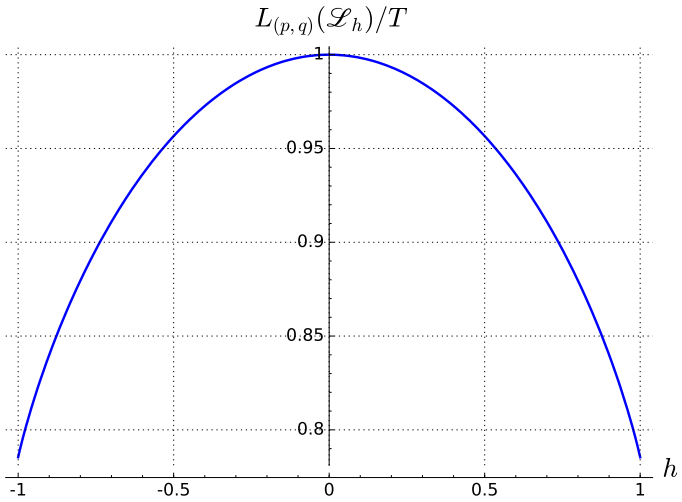


Figure B.2: Length of the timelike curve $\mathscr{L}_h$ connecting the point p of coordinates $(0,0,0,0)$ to the point q of coordinates $(T,0,0,0)$ in Minkowski spacetime, as a function of the parameter h measuring the deviation from the timelike geodesic $\mathscr{L} = \mathscr{L}_0$.

$x_{\max} = hT/4$. We have

$$\dot{X}^0(\lambda) = 1, \quad \dot{X}^1(\lambda) = h\left(1 - 2\frac{\lambda}{T}\right), \quad \dot{X}^2(\lambda) = 0, \quad \dot{X}^3(\lambda) = 0.$$

Given that $(g_{\alpha\beta}) = \mathrm{diag}(-1,1,1,1)$, it follows that $g_{\mu\nu}\dot{X}^\mu\dot{X}^\nu = -1 + h^2(1-2\lambda/T)^2$. Since $\lambda \in [0,T]$, this shows that $\mathscr{L}_h$ is a timelike curve as long as $-1 \leq h \leq 1$. Its length is

$$L_{(p,q)}(\mathscr{L}_h) = \int_0^T \sqrt{1 - h^2\left(1 - 2\frac{\lambda}{T}\right)^2}\,\mathrm{d}\lambda = \frac{T}{2h}\int_{-h}^{h}\sqrt{1-u^2}\,\mathrm{d}u = \frac{T}{2h}\int_{-\arcsin h}^{\arcsin h}\cos^2\theta\,\mathrm{d}\theta.$$

Evaluating the integral leads to

$$L_{(p,q)}(\mathscr{L}_h) = \frac{T}{2}\left(\sqrt{1-h^2} + \frac{\arcsin h}{h}\right).$$

Note that $\arcsin h/h$ is well-defined at $h = 0$, since $\lim_{h\to 0}\arcsin h/h = 1$. The graph of $L_{(p,q)}(\mathscr{L}_h)$ as a function h is plotted in Fig. B.2. We see clearly that $h = 0$, i.e. the geodesic $\mathscr{L}$, corresponds to the maximal length.

For a spacelike geodesic in a Lorentzian manifold, the stationary length corresponds neither to a maximum nor a minimum, but rather to a *saddle point*, as the example below illustrates.

Example 2 (Spacelike geodesic in Minkowski spacetime): As in Example 1, we consider Minkowski spacetime, but this time, $\mathscr{L}$ is assumed to be a spacelike geodesic from p to q. Since $\mathscr{L}$ is necessarily a straight line segment, without any loss of generality, we may introduce a Minkowskian coordinate system $(x^\alpha) = (t,x,y,z)$ such that $x^\alpha(p) = (0,0,0,0)$ and $x^\alpha(q) = (0,L,0,0)$ for some $L > 0$, which is nothing but the length $L_{(p,q)}(\mathscr{L})$ of the geodesic $\mathscr{L}$. Any spacelike curve $\mathscr{L}'$ connecting p and q and lying in the hyperplane Σ defined by $t = 0$ obeys

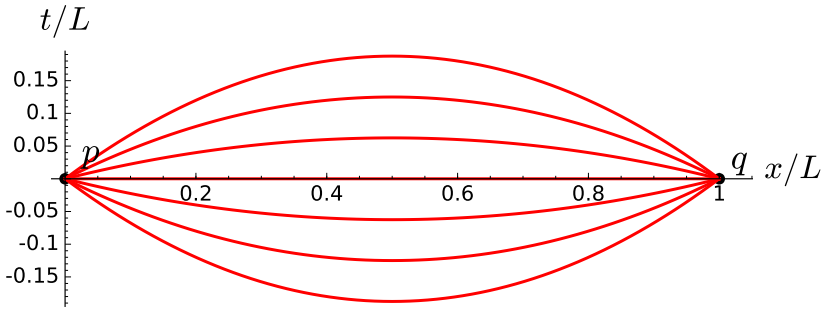


Figure B.3: Spacelike curves $\mathscr{L}_h$ connecting the point p of coordinates $(0,0,0,0)$ to the point q of coordinates $(0,L,0,0)$ in Minkowski spacetime. From the bottom to the top, the depicted curves correspond to h spanning $[-3/4,3/4]$, with the step $\delta h = 1/4$.

$L_{(p,q)}(\mathscr{L}') \geq L_{(p,q)}(\mathscr{L})$ since Σ, equipped with the metric induced by $\boldsymbol{g}$, is a 3-dimensional Euclidean space.

Let us consider some one-parameter family of curves $(\mathscr{L}_h)_{h\in(-1,1)}$ lying in the orthogonal complement of Σ through p and q, namely the curves defined $x^\alpha = X^\alpha(\lambda)$ with $\lambda \in [0,L]$ and

$$X^0(\lambda) = \frac{h}{L}\lambda(L-\lambda), \quad X^1(\lambda) = \lambda, \quad X^2(\lambda) = 0, \quad X^3(\lambda) = 0.$$

As in Example 1, we have $\mathscr{L}_0 = \mathscr{L}$ and for $h \neq 0$, the $\mathscr{L}_h$'s are arcs of parabola from p to q, which remain spacelike as long as $-1 < h < 1$ (cf. Fig. B.3). The computations are similar to those of Example 1, leading to

$$L_{(p,q)}(\mathscr{L}_h) = \frac{L}{2}\left(\sqrt{1-h^2} + \frac{\arcsin h}{h}\right).$$

$L_{(p,q)}(\mathscr{L}_h)/L$ is exactly the same of function of h as $L_{(p,q)}(\mathscr{L}_h)/T$ in Example 1. In view of Fig. B.2, we therefore assert that $L_{(p,q)}(\mathscr{L}_h) \leq L_{(p,q)}(\mathscr{L})$.

We conclude that the spacelike geodesic $\mathscr{L}$ corresponds to a saddle point of the length functional: it is a minimum among the curves lying in the (x,y,z) hyperplane but a maximum among those lying in the (t,x) plane.

B.4.3 All geodesics as stationary points of some action

We have excluded null geodesics from the above variational analysis by invoking the necessary smoothness of the Lagrangian (B.28). We may further convince ourselves that null geodesics would not have fit in the analysis by noticing the division by $\mathcal{L}$ in Eq. (B.37), which excludes $\mathcal{L} = 0$. However, it is possible to get all geodesics, including the null ones, from a variational principle; one has to start from a different action, namely

$$S_{(p,q)}(P) := \frac{1}{2}\int_{\lambda_p}^{\lambda_q} g_{\mu\nu}(X^\rho)\dot{X}^\mu\dot{X}^\nu \,\mathrm{d}\lambda, \tag{B.38}$$

where P is a parametrization of the curve $\mathscr{L}$, λ the corresponding parameter and $x^\alpha = X^\alpha(\lambda)$ the coordinate expression of P.

The Lagrangian in (B.38) is

$$\mathcal{L}_2(X^\alpha, \dot{X}^\alpha) = \frac{1}{2} g_{\mu\nu}(X^\rho) \dot{X}^\mu \dot{X}^\nu. \tag{B.39}$$

We notice that it is always differentiable, even when $g_{\mu\nu}\dot{X}^\mu\dot{X}^\nu = 0$, i.e. it allows for null curves. However, the price to pay is that, contrary to the length (B.25), the action depends on the parametrization of the curve, hence the notation $S_{(p,q)}(P)$ rather than $S_{(p,q)}(\mathscr{L})$. For this reason, $S_{(p,q)}(P)$ is not expected to have any significant physical meaning, contrary to $L_{(p,q)}(\mathscr{L})$, which is the proper time along timelike curves.

Searching for stationary points of the action (B.38) is straightforward. Indeed, given Eqs. (B.31) and (B.33), we have

$$\frac{\partial \mathcal{L}_2}{\partial X^\alpha} = \frac{1}{2} \frac{\partial g_{\mu\nu}}{\partial x^\alpha} \dot{X}^\mu \dot{X}^\nu \qquad \text{and} \qquad \frac{\partial \mathcal{L}_2}{\partial \dot{X}^\alpha} = g_{\alpha\mu} \dot{X}^\mu,$$

so that

$$\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\frac{\partial \mathcal{L}_2}{\partial \dot{X}^\alpha} \right) = \frac{\partial g_{\alpha\mu}}{\partial x^\nu} \dot{X}^\nu \dot{X}^\mu + g_{\alpha\mu} \ddot{X}^\mu.$$

Using the identity (B.35), the Euler-Lagrange equation (B.30) (with $\mathcal{L}$ substituted by $\mathcal{L}_2$) turns out to be equivalent to the geodesic equation (B.10). We conclude that

> In a pseudo-Riemannian manifold $(\mathscr{M}, \boldsymbol{g})$, a curve $\mathscr{L}$ equipped with a parametrization P is a stationary point of the action (B.38) iff $\mathscr{L}$ is a geodesic and P an affine parametrization of it.

Remark 1: The variational principle applied to the action (B.38) leads directly to the geodesic equation (B.10), which implies that the involved parametrization is affine. On the contrary, the variation of the length functional (B.25), leads only to the pregeodesic equation (B.12) (cf. the computation in Sec. B.4.2), which permits a generic parametrization of the geodesic, in agreement with the fact that the length is parametrization-independent, contrary to the action (B.38).

Remark 2: The factor 1/2 in Eq. (B.38) does not play any role in the variational principle, so we could have dropped it. However, thanks to it, the momentum conjuguate to X^α takes a simple form:

$$\Pi_\alpha := \frac{\partial \mathcal{L}}{\partial \dot{X}^\alpha} = g_{\alpha\mu} \dot{X}^\mu. \tag{B.40}$$

The Lagrangian (B.39) can be then written $\mathcal{L}_2 = 1/2\, \Pi_\mu \dot{X}^\mu$ and the Hamiltonian deduced from it by the standard Legendre transformation is $\mathcal{H} = \Pi_\mu \dot{X}^\mu - \mathcal{L}_2 = 1/2\, \Pi_\mu \dot{X}^\mu$, i.e.

$$\mathcal{H}(X^\alpha, \Pi_\alpha) = \frac{1}{2} g^{\mu\nu}(X^\rho) \Pi_\mu \Pi_\nu. \tag{B.41}$$

Such a Hamiltonian has been used by Carter [41] to study the geodesics in Kerr spacetime, discovering the famous *Carter constant*.

B.5 Geodesics and symmetries

B.5.1 Geodesics in presence of a Killing vector

As a reminiscence of Noether's theorem, symmetries in a pseudo-Riemannian manifold lead to conserved quantities along geodesics. Let us first recall that 1-dimensional groups of symmetry and the related concept of Killing vector field have been introduced in Sec. 3.3.1. In terms of them, we may state the following conservation law:

If the pseudo-Riemannian manifold $(\mathcal{M}, \boldsymbol{g})$ admits a 1-dimensional group of symmetry of generator $\boldsymbol{\xi}$, i.e. $\boldsymbol{\xi}$ is a Killing vector field of $(\mathcal{M}, \boldsymbol{g})$, then along any geodesic $\mathscr{L}$, the $\boldsymbol{g}$-scalar product of $\boldsymbol{\xi}$ by any tangent vector field $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ associated with an affine parameter λ of $\mathscr{L}$ is constant:

$$\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{v}) = \text{const.} \tag{B.42}$$

Proof. The variation of $\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{v})$ along $\mathscr{L}$ is, according to Eq. (A.8),

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\lambda} \boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{v}) &= \boldsymbol{v}\left(\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{v})\right) = \boldsymbol{\nabla}_{\boldsymbol{v}}\left(\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{v})\right) \\ &= v^\sigma \nabla_\sigma (g_{\mu\nu} \xi^\mu v^\nu) = v^\sigma \nabla_\sigma (\xi_\nu v^\nu) = \nabla_\sigma \xi_\nu \, v^\sigma v^\nu + \xi_\nu v^\sigma \nabla_\sigma v^\nu \\ &= \frac{1}{2} (\underbrace{\nabla_\sigma \xi_\nu + \nabla_\nu \xi_\sigma}_{0}) v^\sigma v^\nu + \xi_\nu \underbrace{v^\sigma \nabla_\sigma v^\nu}_{0} = 0, \end{aligned} \tag{B.43}$$

where the first zero holds because $\boldsymbol{\xi}$ obeys the Killing equation (3.26) and the second one holds thanks to Eq. (B.1), which expresses that $\mathscr{L}$ is a geodesic and $\boldsymbol{v}$ the tangent vector associated with some affine parameter. □

Remark 1: If the tangent vector $\boldsymbol{v}$ is associated with a generic (not necessarily affine) parameter of $\mathscr{L}$, the second zero in Eq. (B.43) must be replaced by κv^ν, where κ is the non-affinity coefficient of $\boldsymbol{v}$ [cf. Eq. (B.7)]. Accordingly $\boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{v})$ is no longer constant along $\mathscr{L}$ but rather evolves according to

$$\frac{\mathrm{d}}{\mathrm{d}\lambda} \boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{v}) = \kappa \, \boldsymbol{g}(\boldsymbol{\xi}, \boldsymbol{v}). \tag{B.44}$$

Note that κ a priori varies along $\mathscr{L}$, so that the integration of this first-order differential equation depends of the precise form of the function $\kappa(\lambda)$.

B.5.2 Geodesics in presence of a Killing tensor

While the concept of Killing vector is by definition tight to a spacetime symmetry (isometry), there is a generalization of the Killing equation (3.26) to tensors of higher ranks, which is not directly related to any symmetry of the metric tensor. It is however interesting since it leads to conserved quantities along geodesics.

A ***Killing tensor*** (also called a ***Stäckel-Killing tensor*** [45]) of rank $p \geq 1$ in the pseudo-Riemannian manifold $(\mathcal{M}, \boldsymbol{g})$ is a tensor field $\boldsymbol{K}$ of type $(0,p)$ that is fully symmetric and whose covariant derivative obeys

$$\nabla_{(\alpha_1} K_{\alpha_2 \ldots \alpha_{p+1})} = 0. \tag{B.45}$$

Example 3: A trivial example of Killing tensor is the metric tensor $\boldsymbol{g}$ itself. If $(\mathcal{M}, \boldsymbol{g})$ admits a Killing vector $\boldsymbol{\xi}$, another example is $\boldsymbol{K} = \underline{\boldsymbol{\xi}}$ (the 1-form associated to $\boldsymbol{\xi}$ by metric duality), since for $p = 1$, Eq. (B.45) reduces to the Killing equation (3.26). An example for $p = 2$ is then $\boldsymbol{K} = \underline{\boldsymbol{\xi}} \otimes \underline{\boldsymbol{\xi}}$ (by the Leibniz rule + the Killing equation). Similarly, $\boldsymbol{K} = \underline{\boldsymbol{\xi}} \otimes \underline{\boldsymbol{\xi}} \otimes \underline{\boldsymbol{\xi}}$ is a Killing tensor with $p = 3$, etc. A less trivial example is the Walker-Penrose Killing tensor of Kerr spacetime discussed in Sec. 11.2.4.

If a spacetime is endowed with a Killing tensor that is not trivial, i.e. neither formed from $\boldsymbol{g}$ nor any Killing vector as in the examples above, one often says that this spacetime has some ***hidden symmetry*** (see e.g. the review article [109] for an extended discussion). This is because, as Killing vectors, Killing tensors give birth to conserved quantities along geodesics:

If $\boldsymbol{K}$ is a Killing tensor of rank p on the the pseudo-Riemannian manifold $(\mathcal{M}, \boldsymbol{g})$, then along any geodesic $\mathscr{L}$, the scalar $\boldsymbol{K}(\boldsymbol{v}, \ldots, \boldsymbol{v})$, where $\boldsymbol{v} = \mathrm{d}\boldsymbol{x}/\mathrm{d}\lambda$ is any tangent vector field associated with an affine parameter λ of $\mathscr{L}$, is constant:

$$\boldsymbol{K}(\boldsymbol{v}, \ldots, \boldsymbol{v}) = \text{const}. \tag{B.46}$$

Proof. The variation of $\boldsymbol{K}(\boldsymbol{v}, \ldots, \boldsymbol{v})$ along $\mathscr{L}$ is given by

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\lambda}\left(\boldsymbol{K}(\boldsymbol{v}, \ldots, \boldsymbol{v})\right) &= \boldsymbol{\nabla}_{\boldsymbol{v}}\left(\boldsymbol{K}(\boldsymbol{v}, \ldots, \boldsymbol{v})\right) = v^\mu \nabla_\mu (K_{\nu_1 \ldots \nu_p} v^{\nu_1} \cdots v^{\nu_p}) \\
&= \underbrace{v^\mu \nabla_\mu K_{\nu_1 \ldots \nu_p} v^{\nu_1} \cdots v^{\nu_p}}_{0} \\
&\quad + K_{\nu_1 \ldots \nu_p} \underbrace{v^\mu \nabla_\mu v^{\nu_1}}_{0} \cdots v^{\nu_p} + \cdots + K_{\nu_1 \ldots \nu_p} v^{\nu_1} \cdots \underbrace{v^\mu \nabla_\mu v^{\nu_p}}_{0} \\
&= 0,
\end{aligned}$$

where the first zero results from (B.45), while the zeros in the line below arise from the geodesic equation (B.1) obeyed by $\boldsymbol{v}$. □

Example 4: Since we have already noticed that the metric tensor $\boldsymbol{g}$ is a Killing tensor (Example 3), the property (B.6) appears as a special case of (B.46). For the Walker-Penrose Killing tensor $\boldsymbol{K}$ of Kerr spacetime, the conserved quantity $\boldsymbol{K}(\boldsymbol{v}, \boldsymbol{v})$ leads to the Carter constant (cf. Chap. 11).

Appendix C

Kerr-Schild metrics

Contents

C.1 Generic Kerr-Schild spacetimes

C.1.1 Definition

A spacetime $(\mathscr{M}, \boldsymbol{g})$ is said to have a ***Kerr-Schild metric*** iff the metric tensor $\boldsymbol{g}$ can be written

$$\boxed{\boldsymbol{g} = \boldsymbol{f} + 2H\underline{\boldsymbol{k}} \otimes \underline{\boldsymbol{k}}}, \tag{C.1}$$

or equivalently (in index notation):

$$\boxed{g_{\alpha\beta} = f_{\alpha\beta} + 2Hk_\alpha k_\beta}, \tag{C.2}$$

where $\boldsymbol{f}$ is a flat Lorentzian metric on $\mathscr{M}$ (Minkowski metric), H is a scalar field on $\mathscr{M}$ and $\underline{\boldsymbol{k}}$ is a 1-form on $\mathscr{M}$ such that the vector associated to it via $\boldsymbol{f}$ is a null vector of the metric $\boldsymbol{f}$:

$$f^{\mu\nu}k_\mu k_\nu = 0, \tag{C.3}$$

where $f^{\mu\nu}$ stands for the components of the inverse of the metric $\boldsymbol{f}$ (i.e. $f^{\alpha\mu}f_{\mu\beta} = \delta^\alpha{}_\beta$).

A motivation for studying Kerr-Schild metrics is that the inverse metric has a simple expression:

$$g^{\alpha\beta} = f^{\alpha\beta} - 2Hk^\alpha k^\beta, \tag{C.4}$$

where

$$k^\alpha := f^{\alpha\mu}k_\mu. \tag{C.5}$$

Proof: we have successively:

$$\begin{aligned}(f^{\alpha\mu} - 2Hk^\alpha k^\mu)g_{\mu\beta} &= (f^{\alpha\mu} - 2Hk^\alpha k^\mu)(f_{\mu\beta} + 2Hk_\mu k_\beta) \\ &= \underbrace{f^{\alpha\mu} f_{\mu\beta}}_{\delta^\alpha{}_\beta} + 2H \underbrace{f^{\alpha\mu} k_\mu}_{k^\alpha} k_\beta - 2Hk^\alpha \underbrace{k^\mu f_{\mu\beta}}_{k_\beta} - 4H^2 k^\alpha \underbrace{k^\mu k_\mu}_{0} k_\beta \\ &= \delta^\alpha{}_\beta,\end{aligned}$$

which establishes Eq. (C.4). □

Given (C.4), it is easy to see that the vector field $\boldsymbol{k}$ associated to the 1-form $\underline{\boldsymbol{k}}$ by $\boldsymbol{g}$-duality (cf. Sec. A.3.3) is the same as the vector field obtained by $\boldsymbol{f}$-duality:

$$g^{\alpha\mu}k_\mu = (f^{\alpha\mu} - 2Hk^\alpha k^\mu)k_\mu = \underbrace{f^{\alpha\mu}k_\mu}_{k^\alpha} - 2Hk^\alpha \underbrace{k^\mu k_\mu}_{0} = k^\alpha.$$

Accordingly, we may write the components of $\boldsymbol{k}$ simply as k^α without having to specify whether the index has been raised with the metric $\boldsymbol{g}$ or with the metric $\boldsymbol{f}$:

$$k^\alpha = f^{\alpha\mu}k_\mu = g^{\alpha\mu}k_\mu. \tag{C.6}$$

It follows immediately that $\boldsymbol{k}$ is a null vector field for both metrics:

$$\boxed{\boldsymbol{g}(\boldsymbol{k},\boldsymbol{k}) = \boldsymbol{f}(\boldsymbol{k},\boldsymbol{k}) = 0}. \tag{C.7}$$

If $(\mathscr{M}, \boldsymbol{g})$ is a spacetime of Kerr-Schild type, then ***Kerr-Schild coordinates*** are coordinates $(x^\alpha) = (t, x, y, z)$ that are Minkowskian for $\boldsymbol{f}$, i.e. coordinates in which the components of the flat metric $\boldsymbol{f}$ take the form

$$f_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = -\mathrm{d}t^2 + \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2. \tag{C.8}$$

C.1.2 Basic property

Let $\boldsymbol{g}$ be a Kerr-Schild metric. If $\boldsymbol{g}$ obeys the vacuum Einstein equation, i.e. if the Ricci tensor of $\boldsymbol{g}$ vanishes identically:

$$R_{\alpha\beta} = 0, \tag{C.9}$$

then the scalar field H appearing in Eq. (C.1) can be chosen so that $\boldsymbol{k}$ is a geodesic vector field[a]:

$$\boxed{k^\mu \nabla_\mu k^\alpha = 0}, \tag{C.10}$$

where ∇ stands for the covariant derivative associated with $\boldsymbol{g}$.

[a] See Sec. B.2.2 for the definition of a geodesic vector field.

The proof of the above proposition can be found in Ref. [166].

C.2 Case of Kerr spacetime

C.2.1 Kerr-Schild form

Let consider the Kerr spacetime $(\mathscr{M}, \boldsymbol{g})$, where $\mathscr{M}$ is the manifold (10.28): $\mathscr{M} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathscr{R}$ and $\boldsymbol{g}$ is the metric tensor given by Eq. (10.36) in terms of the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ introduced in Sec. 10.3.3. Let us show that $\boldsymbol{g}$ is a Kerr-Schild metric, with the associated null vector field $\boldsymbol{k}$ being nothing but the vector field generating the ingoing principal null geodesics $\mathscr{L}^{\rm in}_{(v,\theta,\tilde{\varphi})}$ discussed in Sec. 10.4. Its expression in terms of the Kerr coordinates is given by Eq. (10.43):

$$\boxed{\boldsymbol{k} = \boldsymbol{\partial}_{\tilde{t}} - \boldsymbol{\partial}_{\tilde{r}}}. \tag{C.11}$$

In other words, the components of $\boldsymbol{k}$ with respect to the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ are $k^\alpha = (1, -1, 0, 0)$. The 1-form $\underline{\boldsymbol{k}}$ associated to $\boldsymbol{k}$ by $\boldsymbol{g}$-duality is given by Eq. (10.44):

$$\underline{\boldsymbol{k}} = -\mathrm{d}\tilde{t} - \mathrm{d}r + a \sin^2\theta \, \mathrm{d}\tilde{\varphi}. \tag{C.12}$$

Equivalently, $k_\alpha = (-1, -1, 0, a \sin^2\theta)$. Let us then introduce the symmetric bilinear form

$$\boldsymbol{f} := \boldsymbol{g} - 2H \underline{\boldsymbol{k}} \otimes \underline{\boldsymbol{k}}, \tag{C.13}$$

where H is the following scalar field on $\mathscr{M}$:

$$\boxed{H := \frac{mr}{\rho^2}}, \tag{C.14}$$

with $\rho^2 := r^2 + a^2 \cos^2\theta$ [Eq. (10.9)]. The expression of $\boldsymbol{f}$ in terms of the Kerr coordinates is deduced from that of $\boldsymbol{g}$ [Eq. (10.36)] and that of $\underline{\boldsymbol{k}}$ [Eq. (C.12)]:

$$\boxed{f_{\mu\nu}\, \mathrm{d}x^\mu \mathrm{d}x^\nu = -\mathrm{d}\tilde{t}^2 + \mathrm{d}r^2 - 2a \sin^2\theta \, \mathrm{d}r \, \mathrm{d}\tilde{\varphi} + \rho^2 \, \mathrm{d}\theta^2 + (r^2 + a^2) \sin^2\theta \, \mathrm{d}\tilde{\varphi}^2}. \tag{C.15}$$

It is easy to check that $f^{\alpha\beta} := g^{\alpha\beta} + 2Hk^\alpha k^\beta$ defines an inverse of $\boldsymbol{f}$: $f^{\alpha\mu} f_{\mu\beta} = \delta^\alpha{}_\beta$ (computation similar to that in Sec. C.1.1). Hence the symmetric bilinear form $\boldsymbol{f}$ is non-degenerate; this implies that $\boldsymbol{f}$ is a *metric tensor* on $\mathscr{M}$ (cf. Sec. A.3.1). Given the components (C.12), it is immediate to check that $\boldsymbol{k}$ is a null vector for $\boldsymbol{f}$ as well: $\boldsymbol{f}(\boldsymbol{k}, \boldsymbol{k}) = 0$. Moreover, $\boldsymbol{f}$ is a *flat* metric, since a direct computation of its Riemann tensor (cf. the notebook D.4.4) reveals that

$$\mathbf{Riem}(\boldsymbol{f}) = 0. \tag{C.16}$$

In view of the definition given in Sec. C.1.1, we conclude that

> The Kerr metric $\boldsymbol{g}$ is a Kerr-Schild metric, i.e. it can be written in the form (C.1) with the flat metric $\boldsymbol{f}$ given by Eq. (C.15), the scalar field H given by Eq. (C.14) and the null vector $\boldsymbol{k}$ given by Eq. (C.11), $\boldsymbol{k}$ being the tangent vector field to the ingoing principal null geodesics.

In Sec. 10.4, we have already noticed that $\boldsymbol{k}$ is a geodesic vector: $\boldsymbol{\nabla}_{\boldsymbol{k}}\,\boldsymbol{k} = 0$ [Eq. (10.45)], in agreement with (C.10).

Remark 1: The Kerr metric can also be brought to the Kerr-Schild form by using the tangent vector field to the *outgoing* principal null geodesics. Hence the Kerr-Schild decomposition (C.1) is not unique for the Kerr metric.

C.2.2 Kerr-Schild coordinates on Kerr spacetime

It is not immediately obvious that the metric $\boldsymbol{f}$ given by Eq. (C.15) is a flat Lorentzian metric. Let us introduce coordinates in which $\boldsymbol{f}$ takes a manifestly Minkowskian form, i.e. Kerr-Schild coordinates, according to the nomenclature introduced in Sec. C.1.1.

Actually, if $a \neq 0$, one cannot introduce a Kerr-Schild coordinate system on the whole spacetime manifold $\mathscr{M} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathscr{R}$ as defined by Eq. (10.28). One has to split it in two parts:

$$\mathscr{M} := \mathscr{M}_+ \cup \mathscr{M}_-, \tag{C.17a}$$

$$\mathscr{M}_+ := \mathbb{R} \times [0, +\infty) \times \mathbb{S}^2 \setminus \mathscr{R} \tag{C.17b}$$

$$\mathscr{M}_- := \mathbb{R} \times (-\infty, 0] \times \mathbb{S}^2 \setminus \mathscr{R}. \tag{C.17c}$$

In other words, $\mathscr{M}_+$ is the part $r \geq 0$ of $\mathscr{M}$, while $\mathscr{M}_-$ is the part $r \leq 0$. Note that $\mathscr{M}_+$ and $\mathscr{M}_-$ are submanifolds with boundaries (cf. Sec. A.2.2), which overlap at $r = 0$. In terms of the domains introduced in Sec. 10.2.1, $\mathscr{M}_+$ contains $\mathscr{M}_{\mathrm{I}}$, $\mathscr{M}_{\mathrm{II}}$ and a part of $\mathscr{M}_{\mathrm{III}}$, while $\mathscr{M}_-$ is entirely included in $\mathscr{M}_{\mathrm{III}}$. The ***Kerr-Schild coordinates*** $(\tilde{t}, x, y, z)$ on $\mathscr{M}_+$ are defined from the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ by the following formulas:

$$\tilde{t} = \tilde{t} \tag{C.18a}$$

$$x = (r\cos\tilde{\varphi} - a\sin\tilde{\varphi})\sin\theta \tag{C.18b}$$

$$y = (r\sin\tilde{\varphi} + a\cos\tilde{\varphi})\sin\theta \tag{C.18c}$$

$$z = r\cos\theta. \tag{C.18d}$$

Remark 2: As we shall see in Sec. C.2.3, the Kerr-Schild coordinates are singular at $r = 0$, so strictly speaking, we should have omitted $r = 0$ from the definition of $\mathscr{M}_+$ and $\mathscr{M}_-$.

Remark 3: For $a = 0$, Eqs. (C.18b)-(C.18d) reduce to the standard relations between Cartesian and spherical coordinates in Euclidean space.

Remark 4: Equations (C.18b)-(C.18c) can be combined into a single relation:

$$x + iy = (r + ia)\mathrm{e}^{i\tilde{\varphi}}\sin\theta. \tag{C.19}$$

From Eqs. (C.18b)-(C.18c), we get

$$x^2 + y^2 = (r^2 + a^2)\sin^2\theta. \tag{C.20}$$

Combining with Eq. (C.18d) yields:

$$\boxed{\frac{x^2+y^2}{r^2+a^2}+\frac{z^2}{r^2}=1}. \tag{C.21}$$

This is a quadratic equation in r^2. Solving it results in

$$r=\sqrt{\frac{1}{2}\left(x^2+y^2+z^2-a^2+\sqrt{(x^2+y^2+z^2-a^2)^2+4a^2z^2}\right)}. \tag{C.22}$$

The components of $\boldsymbol{f}$ in terms on the coordinates $(x^\alpha)=(\tilde{t},x,y,z)$ are obtained via the tensor change-of-components formula with the transformation (C.18) (cf. the notebook D.4.4):

$$f_{\mu\nu}\,\mathrm{d}x^\mu\mathrm{d}x^\nu=-\mathrm{d}\tilde{t}^2+\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2. \tag{C.23}$$

This proves that $(\tilde{t},x,y,z)$ are Kerr-Schild coordinates, as announced.

The expression of the vector $\boldsymbol{k}$ in terms of the Kerr-Schild coordinates is obtained similarly:

$$\boldsymbol{k}=\boldsymbol{\partial}_{\tilde{t}}-\frac{rx+ay}{r^2+a^2}\,\boldsymbol{\partial}_x-\frac{ry-ax}{r^2+a^2}\,\boldsymbol{\partial}_y-\frac{z}{r}\,\boldsymbol{\partial}_z. \tag{C.24}$$

In this formula, r is to be considered as the function of (x,y,z) given by Eq. (C.22). For the associated 1-form, we get

$$\underline{\boldsymbol{k}}=-\mathbf{d}\tilde{t}-\frac{rx+ay}{r^2+a^2}\,\mathbf{d}x-\frac{ry-ax}{r^2+a^2}\,\mathbf{d}y-\frac{z}{r}\,\mathbf{d}z. \tag{C.25}$$

The scalar factor H can be re-expressed from Eq. (C.14) in terms of z and r:

$$H=\frac{mr^3}{r^4+a^2z^2}. \tag{C.26}$$

Remark 5: If $a=0$ (Schwarzschild limit), we get

$$r=\sqrt{x^2+y^2+z^2},\quad H=\frac{m}{r}\quad\text{and}\quad \underline{\boldsymbol{k}}=-\mathbf{d}\tilde{t}-\frac{x}{r}\,\mathbf{d}x-\frac{y}{r}\,\mathbf{d}y-\frac{z}{r}\,\mathbf{d}z. \tag{C.27}$$

Remark 6: For $a\neq 0$, the relations (C.27) hold at first order in the limit $r\gg a$, or equivalently in the limit $\sqrt{x^2+y^2+z^2}\gg a$.

The explicit form of the components $g_{\mu\nu}$ of the Kerr metric in Kerr-Schild coordinates can be read off by expanding the line element

$$\boxed{\begin{aligned} g_{\mu\nu}\,\mathrm{d}x^\mu\,\mathrm{d}x^\nu = \;& -\mathrm{d}\tilde{t}^2+\mathrm{d}x^2+\mathrm{d}y^2+\mathrm{d}z^2 \\ & +\frac{2mr^3}{r^4+a^2z^2}\left(\mathrm{d}\tilde{t}+\frac{rx+ay}{r^2+a^2}\,\mathrm{d}x+\frac{ry-ax}{r^2+a^2}\,\mathrm{d}y+\frac{z}{r}\,\mathrm{d}z\right)^2 \end{aligned}}, \tag{C.28}$$

which is obtained by combining Eqs (C.2), (C.23), (C.26) and (C.25).

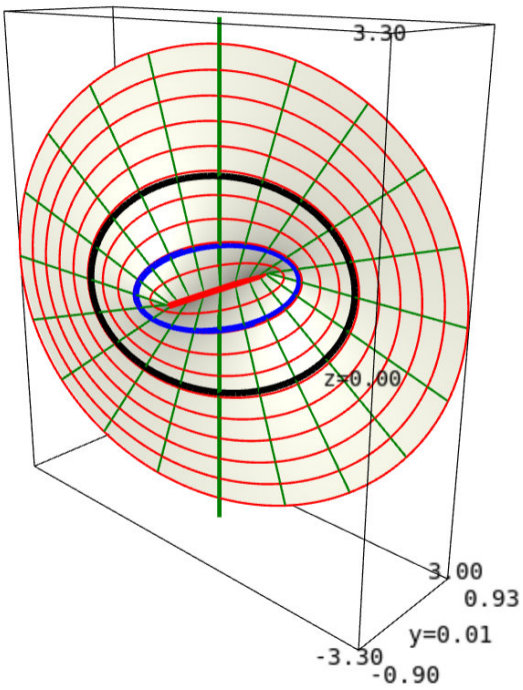


Figure C.1: Surface $\tilde{t} = \text{const}$, $\tilde{\varphi} \in \{0, \pi\}$ and $r \geq 0$ of the $a = 0.9\,m$ Kerr spacetime depicted in terms of the Kerr-Schild coordinates (x, y, z). The drawing is limited to $r \leq 3m$. The vertical thick green line is the axis of rotation. On the right of it $\tilde{\varphi} = 0$, while on the left of it $\tilde{\varphi} = \pi$. The red lines are curves $r = \text{const}$, while the green ones are curves $\theta = \text{const}$, which can be thought of as the traces of the ingoing principal null geodesics. The thick black curve marks the black hole event horizon and the thick blue curve the Cauchy horizon. The thick red segment along the y-axis marks the intersection of the surface with the disk $r = 0$. *[Figure produced with the notebook D.4.4]*

Remark 7: It is clear on (C.28) that all metric components in Kerr-Schild coordinates are regular both at the the black hole event horizon ($r = m + \sqrt{m^2 - a^2}$, cf. Sec. 10.5.2) and the Cauchy horizon ($r = m - \sqrt{m^2 - a^2}$, cf. Sec. 10.8.3). This property, which is shared by the Kerr coordinates, is in sharp contrast with the metric components in Boyer-Lindquist coordinates, which are singular at both horizons (cf. Sec. 10.2.6).

Finally, the axisymmetry Killing vector $\boldsymbol{\eta} = \boldsymbol{\partial}_{\tilde{\varphi}}$ of Kerr spacetime [Eq. (10.39)] has the following expression in terms of Kerr-Schild coordinates:

$$\boldsymbol{\eta} = -y\,\boldsymbol{\partial}_x + x\,\boldsymbol{\partial}_y. \tag{C.29}$$

This is easily established from the chain rule and the partial derivative with respect to $\tilde{\varphi}$ of expressions (C.18). We notice on Eq. (C.29) that $\boldsymbol{\eta}$ is also a Killing vector for the flat metric $\boldsymbol{f}$, namely the Killing vector generating spatial rotations about the z-axis.

The identity (C.21) shows that, in the Euclidean space spanned by the (x, y, z) coordinates, the surfaces of constant $r \neq 0$ are confocal[1] ellipsoids of revolution. This is depicted in Fig. C.1, which represents a slice $\tilde{t} = \text{const}$ and $\tilde{\varphi} = 0$ or π in terms of the

[1] In any plane containing the axis of symmetry $x^2 + y^2 = 0$, the trace of the ellipsoids are ellipses that share the same foci, located at the abscissas $\pm a$ along the $z = 0$ axis.

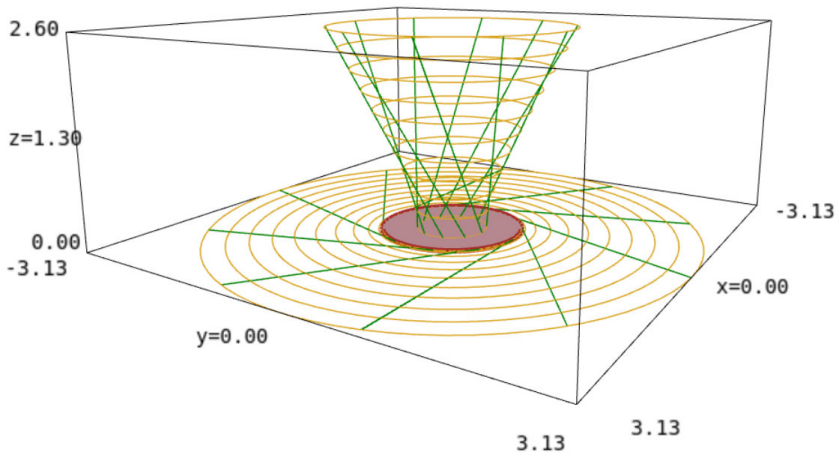


Figure C.2: Surfaces $(\tilde{t},\theta) = \text{const}$ of the $a = 0.9\,m$ Kerr spacetime depicted in terms of the Kerr-Schild coordinates (x, y, z). The disk-like surface in the plane $z = 0$ is for $\theta = \pi/2$, while the cone-like surface is for $\theta = \pi/6$. The pale brown lines are curves $(r,\theta) = \text{const}$, while the green ones are curves $(\theta,\tilde{\varphi}) = \text{const}$. The latter can be thought of as the traces of the ingoing principal null geodesics. The central pink disk is the (double) disk $r = 0$, the boundary of which is the curvature singularity. *[Figure produced with the notebook D.4.4]*

(x, y, z) coordinates. Note that this slice is not a plane but a warped surface, with a kink along the red segment $-a < y < a$ at $(x, z) = (0, 0)$, which is the intersection of the slice with the double disk $r = 0$ (to be discussed below). Peculiar $r = \text{const}$ surfaces are the black hole event horizon ($r = r_+ = m + \sqrt{m^2 - a^2}$, cf. Sec. 10.5.2) and the Cauchy horizon ($r = r_- = m - \sqrt{m^2 - a^2}$, cf. Sec. 10.8.3). They are depicted in respectively black and blue in Fig. C.1.

Since the ingoing principal null geodesics $\mathscr{L}^{\text{in}}_{(v,\theta,\tilde{\varphi})}$ (cf. Sec. 10.4) are curves $(\theta,\tilde{\varphi}) = \text{const}$, their traces in the 3-space of Fig. C.1 are the green lines that terminates at the disk $r = 0$ (the red segment along the y-axis). Another view of the ingoing principal null geodesics is provided by Fig. C.2, which shows two surfaces $(\tilde{t},\theta) = \text{const}$ in terms of the Kerr-Schild coordinates (x, y, z), namely the surfaces $\theta = \pi/6$ and $\theta = \pi/2$. We notice that, although they are straight lines in terms of the Kerr-Schild coordinates, the ingoing principal null geodesics are winding around the rotation axis in the direction of the black hole rotation, which is indicated by $\boldsymbol{\eta}$ [cf. Eq. (C.29)].

C.2.3 The double-disk $r = 0$

For $r = 0$, the system (C.18) reduces to

$$\tilde{t} = \tilde{t} \tag{C.30a}$$
$$x = -a\sin\theta\sin\tilde{\varphi} \tag{C.30b}$$
$$y = a\sin\theta\cos\tilde{\varphi} \tag{C.30c}$$
$$z = 0 \tag{C.30d}$$

For $a \neq 0$ and a fixed value of $\tilde{t}$, the subset $r = 0$ of the Kerr spacetime $\mathscr{M}$ is the set $\mathcal{S}_{0,t}$ discussed in Sec. 10.2.2 [cf. Eq. (10.14)], where t is related to $\tilde{t}$ via Eq. (10.35a), which

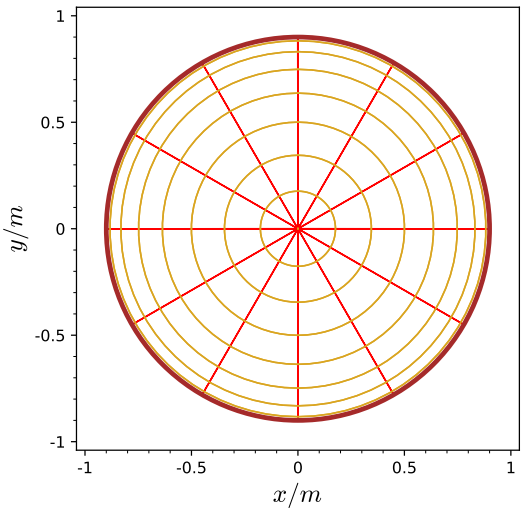


Figure C.3: Disk (actually double disk) $r = 0$ of the $a = 0.9\,m$ Kerr spacetime depicted in terms of the Kerr-Schild coordinates (x, y). The pale brown circles are the curves $\theta = \text{const}$, while the red segments are the curves $\tilde{\varphi} = \text{const}$. The disk boundary at $\sqrt{x^2 + y^2} = a$ is the curvature singularity of Kerr spacetime. *[Figure produced with the notebook D.4.4]*

for $r = 0$, is basically $\tilde{t} = t + \text{const}$. $\mathcal{S}_{0,t}$ is topologically a 2-sphere minus its equator. It is therefore disconnected, with two connected components: the two open hemispheres, $\mathcal{S}^+_{0,t}$ and $\mathcal{S}^-_{0,t}$ say. As shown in Sec. 10.2.2, $\mathcal{S}^+_{0,t}$ and $\mathcal{S}^-_{0,t}$ are actually two disks that, with respect to the metric induced by $\boldsymbol{g}$, (i) have radius a and (ii) are flat. The disk $\mathcal{S}^+_{0,t}$ is spanned by the coordinates $(\theta, \tilde{\varphi})$ with $\theta \in [0, \pi/2)$, while $\mathcal{S}^-_{0,t}$ is spanned by the coordinates $(\theta, \tilde{\varphi})$ with $\theta \in (\pi/2, \pi]$. Let $p \in \mathcal{S}^+_{0,t}$ be a point of Kerr coordinates $(\tilde{t}, 0, \theta, \tilde{\varphi})$. The point q of coordinates $(\tilde{t}, 0, \pi - \theta, \tilde{\varphi})$ belongs to $\mathcal{S}^-_{0,t}$; it is therefore distinct from p, since $\mathcal{S}^+_{0,t}$ and $\mathcal{S}^-_{0,t}$ are two disjoint sets. Now, the transformation (C.30) maps both p and q to the same value $(\tilde{t}, -a \sin\theta \sin\tilde{\varphi}, a \sin\theta \cos\varphi, 0)$ of the Kerr-Schild coordinates $(\tilde{t}, x, y, z)$. We conclude that, for $r = 0$, the Kerr-Schild coordinate system fails to establish a one-to-one correspondence between the manifold points and some open subset of $\mathbb{R}^4$. Actually, as it is clear from (C.30), the two disks $\mathcal{S}^+_{0,t}$ and $\mathcal{S}^-_{0,t}$ are mapped by Kerr-Schild coordinates to a single disk, namely the disk of radius a centered at $(x, y, z) = (0, 0, 0)$ in the plane $z = 0$. This disk is shown in Fig. C.2 (pink central disk) and in Fig. C.3; its section $\tilde{\varphi} = 0$ or π is depicted as the red segment in Fig. C.1. The disk boundary is the circle $z = 0$, $\sqrt{x^2 + y^2} = a$; it is not part of $\mathscr{M}$, being the curvature singularity of Kerr spacetime (cf. Sec. 10.2.6).

C.2.4 Kerr-Schild coordinates on the $r \leq 0$ part

On the domain $\mathscr{M}_-$, i.e. for $r \leq 0$, one can introduce another patch of Kerr-Schild coordinates, $(\tilde{t}, x', y', z')$ say, by formulas similar to (C.18). A difference is when expressing

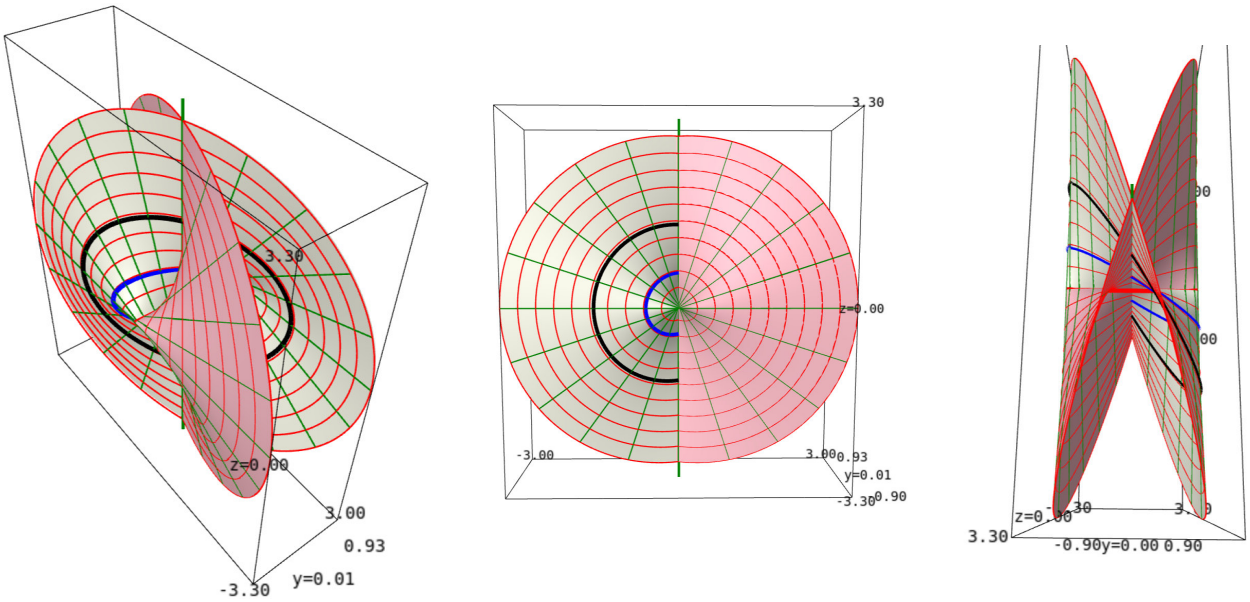


Figure C.4: Three views of the immersion of the full $\tilde{t} = \text{const}$ and $\tilde{\varphi} \in \{0, \pi\}$ surface of the $a = 0.9\,m$ Kerr spacetime in the Euclidean space $\mathbb{R}^3$, using the Kerr-Schild coordinates (x, y, z) for the $r \geq 0$ part (drawn in grey) and the Kerr-Schild coordinates (x', y', z') for the $r \leq 0$ part (drawn in pink). The red lines are curves $(r, \tilde{\varphi}) = \text{const}$, while the green straight lines are ingoing principal null geodesics, which obey $(\theta, \tilde{\varphi}) = \text{const}$. The thick (resp. blue) black curve marks the black hole event horizon (resp. Cauchy horizon). *[Figure produced with the notebook D.4.4; see this notebook for an interactive 3D view]*

the square root for r as in Eq. (C.22), one has to take the minus sign, so that

$$r := -\sqrt{\frac{1}{2}\left(x'^2 + y'^2 + z'^2 - a^2 + \sqrt{(x'^2 + y'^2 + z'^2 - a^2)^2 + 4a^2 z'^2}\right)}. \tag{C.31}$$

The two Kerr-Schild coordinate domains $\mathscr{M}_+$ and $\mathscr{M}_-$ are connected through the $r = 0$ hypersurface. On a $\tilde{t} = \text{const}$ slice, this means being connected through the double disk $\mathcal{S}_{0,t}$. Such a connection is depicted in Fig. C.4, which represents the (non-isometric) immersion in $\mathbb{R}^3$ of the surface $\tilde{t} = \text{const}$ and $\tilde{\varphi} \in \{0, \pi\}$, with r ranging from $-\infty$ to $+\infty$ (actually from $-3m$ to $3m$ on the figure). The immersion is not an embedding[2] because the sheet $r \geq 0$ (in grey) intersects the sheet $r \leq 0$ (in pink) along the rotation axis, while the only intersection of $\mathscr{M}_+$ and $\mathscr{M}_-$ is along the double disk $\mathcal{S}_{0,t}$ (reduced to the segment $-a < y < a$ at $(x, z) = (0, 0)$ in Fig. C.4). In other words, the intersection of the grey and pink sheets along the z-axis in Fig. C.4 is spurious (does not correspond to an intersection in the physical spacetime), while the intersection along the y-axis is physical. From the central and right views in Fig. C.4, one sees clearly that the ingoing principal null geodesics go smoothly from the $r > 0$ region to the $r < 0$ region through $\mathcal{S}_{0,t}$.

[2] See Sec. A.2.7 for the definitions of *immersion* and *embedding*.

C.2.5 Link with Boyer-Lindquist coordinates

The link between the Kerr-Schild coordinates $(\tilde{t}, x, y, z)$ and the Boyer-Lindquist coordinates (t, r, θ, φ) is obtained by combining Eqs. (C.18) with Eqs. (10.35).

Historical note: Kerr-Schild coordinates on Kerr spacetime have been introduced by Roy Kerr in the famous 1963 paper [164] announcing the discovery of the Kerr metric. They have been discussed further by Robert Boyer and Richard Lindquist in 1967 [30], Brandon Carter in 1968 [41] and Stephen Hawking and George Ellis in 1973 [146]. Generic Kerr-Schild metrics have been introduced and studied by Roy Kerr and Alfred Schild in 1965 [166].

Appendix D

SageMath computations

Contents

D.1 Introduction

SageMath (https://www.sagemath.org/) is a modern free, open-source mathematics software system, which is based on the Python programming language. It makes use of over 90 open-source packages, among which are Maxima, Pynac and SymPy (symbolic calculations), GAP (group theory), PARI/GP (number theory), Singular (polynomial computations), and matplotlib (high quality 2D figures). SageMath provides a uniform Python interface to all these packages. However, SageMath is more than a mere interface: it contains a large and increasing part of original code (more than 750,000 lines of Python and Cython, involving 5344 classes). SageMath was created in 2005 by William Stein [236] and since then its development has been sustained by more than a hundred researchers (mostly mathematicians). Very good introductory textbooks about SageMath are [265, 162, 15].

The SageManifolds project (https://sagemanifolds.obspm.fr/) provides SageMath with capabilities for differential geometry and tensor calculus, where are used here to perform some computations relative to black hole spacetimes.

There are basically two ways to use SageMath:

- Install it on your computer, by downloading the sources or a binary version from https://www.sagemath.org/ (the SageManifolds extensions towards differential geometry are fully integrated in version 7.5 and higher)

- Use it online via CoCalc: https://cocalc.com/

The SageMath notebooks (Jupyter format) accompanying these lecture notes are available at the nbviewer.jupyter.org links provided below. Clicking on the link leads to a read-only view of the notebook. Then, by clicking on the *Execute on Binder* button (the three interlaced circles in the top right menu), one gets access to a freely modifiable and executable version. All the notebooks are also collected on the page
https://luth.obspm.fr/~luthier/gourgoulhon/bh16/sage.html

D.2 Minkowski spacetime

D.2.1 Conformal completion of Minkowski spacetime

This notebook accompanies Chap. 4; in particular, it provides many figures for Sec. 4.2.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/conformal_Minkowski.ipynb

D.3 Schwarzschild spacetime

D.3.1 The Schwarzschild horizon

This notebook accompanies Chap. 2 in treating the future event horizon of Schwarzschild spacetime in Eddington-Finkelstein coordinates as an example of null hypersurface:

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Schwarzschild_horizon.ipynb

D.3.2 Solving Einstein equation: Kottler solution

This notebook accompanies Chap. 6: it computes the Kottler solution by solving the Einstein equation for vacuum spherically symmetric spacetimes with a cosmological constant Λ, yielding Schwarzschild solution in the special case $\Lambda = 0$.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kottler_solution.ipynb

D.3.3 Kretschmann scalar of Schwarzschild spacetime

This notebook accompanies Chap. 6: it computes the Riemann curvature tensor of Schwarzschild metric and evaluates the Kretschmann scalar as defined by Eq. (6.46).

https://nbviewer.jupyter.org/github/sagemanifolds/SageManifolds/blob/master/Notebooks/SM_basic_Schwarzschild.ipynb

D.3.4 Radial null geodesics in Schwarzschild spacetime

This notebook accompanies Chap. 6: it provides figures based on Schwarzschild-Droste coordinates and ingoing Eddington-Finkelstein coordinates.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Schwarz_radial_null_geod.ipynb

D.3.5 Radial timelike geodesics in Schwarzschild spacetime

This notebook accompanies Chap. 7: it provides figures as well as the computation of the integral leading to of Eq. (7.42).

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/ges_radial_free_fall.ipynb

D.3.6 Timelike orbits in Schwarzschild spacetime

This notebook accompanies Chap. 7: it provides figures of timelike orbits in the equatorial plane.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/ges_orbits.ipynb

D.3.7 Effective potential for null geodesics in Schwarzschild spacetime

This notebook accompanies Chap. 8, providing the plot of the effective potential $U(r)$.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/ges_effective_potential_null.ipynb

D.3.8 Null geodesics in Schwarzschild spacetime

This notebook accompanies Chap. 8, computing and plotting various null geodesics.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/ges_null_geod.ipynb

D.3.9 Periastron and apoastron of null geodesics in Schwarzschild spacetime

This notebook accompanies Chap. 8, computing periastrons and apoastrons along null geodesics.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/ges_null_periastron.ipynb

D.3.10 Critical null geodesics in Schwarzschild spacetime

This notebook accompanies Chap. 8, plotting critical null geodesics.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/ges_null_critical_geod.ipynb

D.3.11 Elliptic integrals for null geodesics in Schwarzschild spacetime

This notebook accompanies Chap. 8, computing the trace of null geodesics in the equatorial plane via elliptic integrals.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/gis_elliptic_int.ipynb

D.3.12 Null geodesics in Schwarzschild spacetime with $b < b_{\rm c}$

This notebook accompanies Chap. 8, computing various quantities that are relevant for null geodesics with an impact parameter lower than the critical one.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/gis_paramaters_b_lt_bc.ipynb

D.3.13 Multiple images in Schwarzschild spacetime

This notebook accompanies Chap. 8, computing null geodesics that depart in a fixed direction (the observer one).

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/ges_null_images.ipynb

D.3.14 Emission from a point source in Schwarzschild spacetime

This notebook accompanies Chap. 8, computing quantities related to the emission by a static observer.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/gis_emission.ipynb

D.3.15 Images of an accretion disk around a Schwarzschild black hole

This notebook accompanies Chap. 8, computing null geodesics illustrating the formation of images of an accretion disk.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/gis_disk_image.ipynb

D.3.16 Kruskal-Szekeres coordinates in Schwarzschild spacetime

This notebook accompanies Chap. 9: it provides the figures based on Kruskal-Szekeres coordinates.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Schwarz_Kruskal_Szekeres.ipynb

D.3.17 Standard (singular) Carter-Penrose diagram of Schwarzschild spacetime

This notebook accompanies Chap. 9: it provides the standard Carter-Penrose diagram shown in Fig. 9.10.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Schwarz_conformal_std.ipynb

D.3.18 Regular Carter-Penrose diagram of Schwarzschild spacetime

This notebook accompanies Chap. 9: it provides the regular Carter-Penrose diagram shown in Fig. 9.11.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Schwarz_conformal.ipynb

D.3.19 Einstein-Rosen bridge in Schwarzschild spacetime

This notebook accompanies Chap. 9: it provides the isometric embedding diagrams shown in Figs. 9.14 to 9.16, as well as the associated Kruskal diagram of Fig. 9.12.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Einstein-Rosen_bridge.ipynb

D.4 Kerr spacetime

D.4.1 Kerr metric as a solution of Einstein equation

This notebook accompanies Chap. 10: the Kerr metric, expressed in Boyer-Lindquist coordinates, is shown to be a solution of the vacuum Einstein equation. Moreover, the Kretschmann scalar is computed.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_solution.ipynb

D.4.2 Kerr spacetime in Kerr coordinates

This notebook accompanies Chap. 10: the Kerr metric is expressed in Kerr coordinates, the vacuum Einstein equation is checked, the outgoing and ingoing principal null geodesics are considered and the black hole surface gravity is computed.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_in_Kerr_coord.ipynb

D.4.3 Plot of principal null geodesics in Kerr spacetime

This notebook provides some figures for Chap. 10.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_princ_null_geod.ipynb

D.4.4 Kerr-Schild coordinates on Kerr spacetime

This notebook accompanies Appendix. C: the Kerr metric is expressed in a Kerr-Schild form and Kerr-Schild coordinates are introduced.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_Schild.ipynb

D.4.5 ZAMO frame on Kerr spacetime

This notebook accompanies Chap. 10, providing various formulas relative to the orthonormal frame carried by a Zero-angular-momentum observer (ZAMO).

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_ZAMO_frame.ipynb

D.4.6 Carter frame on Kerr spacetime

This notebook accompanies Chap. 10, providing various formulas relative to the orthonormal frame carried by a Carter observer.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_Carter_frame.ipynb

D.4.7 Walker-Penrose Killing tensor on Kerr spacetime

This notebook accompanies Chap. 11; it shows that the tensor $\boldsymbol{K}$ defined by Eq. (11.16) is actually a Killing tensor.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_Killing_tensor.ipynb

D.4.8 Timelike and null geodesics in Kerr spacetime

This notebook accompanies Chap. 11, computing and plotting null and timelike geodesics.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_geod_plots.ipynb

D.4.9 Circular equatorial orbits in Kerr spacetime

This notebook provides some figures for Chap. 11.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_circular_orbits.ipynb

D.4.10 Zero-energy null geodesics in Kerr spacetime

This notebook provides some figures for Chap. 12.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_null_geod_zero_ener.ipynb

D.4.11 Existence and stability of spherical photon orbits in Kerr spacetime

This notebook provides some figures for Chap. 12.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_spher_photon_existence.ipynb

D.4.12 Plots of spherical photon orbits geodesics in Kerr spacetime

This notebook provides some figures for Chap. 12.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_spher_null_geod.ipynb

D.4.13 Plots of null geodesics in Kerr spacetime

This notebook provides some figures for Chap. 12.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_null_geod_plots.ipynb

D.4.14 Shadow and critical curve of a Kerr black hole

This notebook provides some figures for Chap. 12.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_shadow.ipynb

D.4.15 Images of an accretion disk around a Kerr black hole

This notebook provides some figures for Chap. 12.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_images.ipynb

D.4.16 Critical curve of a Kerr black hole onto the EHT image of M87*

This notebook generates Fig. 12.30.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_image_M87.ipynb

D.4.17 Extremal Kerr spacetime

This notebook accompanies Chap. 13.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_extremal.ipynb

D.4.18 Maximal extension of the extremal Kerr spacetime

This notebook accompanies Chap. 13.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_extremal_extended.ipynb

D.5 Evolution and thermodynamics

D.5.1 Lemaître-Tolman equations

This notebook accompanies Chap. 14: it provides the derivation of the Lemaître-Tolman equations from the Einstein equation expressed in Lemaître synchronous coordinates.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Lemaitre_Tolman.ipynb

D.5.2 Trapping horizon in Vaidya spacetime

This notebook accompanies Chap. 18: the Vaidya metric is expressed in Eddington-Finkelstein coordinates, the Einstein equation is checked, the outgoing and ingoing radial null geodesics are computed and the trapping horizon and the event horizon are drawn in

a spacetime diagram.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Vaidya.ipynb

Appendix E

Gyoto computations

Contents

E.1 Introduction

Gyoto (https://gyoto.obspm.fr) is a free open-source C++ code for computing orbits and ray-traced images in general relativity [253]. It has a Python interface and has the capability to integrate geodesics not only in analytical spacetimes (such as Kerr) but also in numerical ones, i.e. in spacetimes arising from numerical relativity.

E.2 Image computations

Here we provide the Gyoto input files, in XML format, that have been used to produce the images shown in Chaps. 8 and 12. To generate the images, it suffices to run Gyoto as

```
gyoto input.xml output.fits
```

By default, the computation is performed in parallel on 8 threads; you can adapt to your CPU by changing the field `NThreads` in the file `input.xml`. The output image is in FITS format and can be converted to PNG or JPEG by most image processing programs, such as GIMP.

E.2.1 Accretion disk around a Schwarzschild black hole

The input XML files for generating the images shown in Fig. 8.25 are the files `gis_disk*.xml` in the directory
https://github.com/egourgoulhon/BHLectures/tree/master/gyoto

E.2.2 Accretion disk around a Kerr black hole

The input XML files for generating the images shown in Figs. 12.27 and 12.28 are the files `gik_a*.xml` in the directory
https://github.com/egourgoulhon/BHLectures/tree/master/gyoto

Appendix F

On the Web

Here is a selection of scientific web pages related to black holes:

- Movies of binary black holes mergers computed by the SXS team:
  https://www.black-holes.org/explore/movies
- Movies from computations of the Center for Computational Relativity and Gravitation, Rochester Institute of Technology:
  https://ccrg.rit.edu/movies
- Journey around a black hole (Alain Riazuelo)
  http://www2.iap.fr/users/riazuelo/bh/
- Kerr black holes images and videos (David Madore)
  http://www.madore.org/~david/math/kerr.html
- Spherical photon orbits around a Kerr black hole (Edward Teo):
  http://phyweb.physics.nus.edu.sg/~phyteoe/kerr/
- Kerr Spherical Photon Orbits (Leo C. Stein):
  https://duetosymmetry.com/tool/kerr-circular-photon-orbits/
- Scratch pad for visualizing bound, timelike geodesics in Kerr spacetime:
  http://nielswarburton.net/geodesics/interactive/Kerr_geodesic.html
- Gyoto gallery (images and movies of accretion flows around black holes and alternative compact objects; see also Appendix E):
  https://gyoto.obspm.fr/gallery/

Bibliography

[1] M.A. Abramowicz and P.C. Fragile: *Foundations of Black Hole Accretion Disk Theory*, Living Rev. Relativ. **16**, 1 (2013)
DOI: 10.12942/lrr-2013-1

[2] M. Abramowitz and I. Stegun: *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, NBS, 10th printing (1972); updated in the *NIST Digital Library of Mathematical Functions*, available online at
https://dlmf.nist.gov/

[3] R. Abuter et al. (GRAVITY Collaboration): *Detection of the gravitational redshift in the orbit of the star S2 near the Galactic centre massive black hole*, Astron. Astrophys. **615**, L15 (2018)
DOI: 10.1051/0004-6361/201833718 arXiv: 1807.09409

[4] R. Abuter et al. (GRAVITY Collaboration): *Detection of the Schwarzschild precession in the orbit of the star S2 near the Galactic centre massive black hole*, Astron. Astrophys. **636**, L5 (2020)
DOI: 10.1051/0004-6361/202037813 arXiv: 2004.07187

[5] K. Akiyama et al. (the Event Horizon Telescope Collaboration): *First M87 Event Horizon Telescope Results. I. The Shadow of the Supermassive Black Hole*, Astrophys. J. Lett. **875**, L1 (2019)
DOI: 10.3847/2041-8213/ab0ec7 arXiv: 1906.11238

[6] K. Akiyama et al. (the Event Horizon Telescope Collaboration): *First M87 Event Horizon Telescope Results. V. Physical Origin of the Asymmetric Ring*, Astrophys. J. Lett. **875**, L5 (2019)
DOI: 10.3847/2041-8213/ab0f43 arXiv: 1906.11242

[7] K. Akiyama et al. (the Event Horizon Telescope Collaboration): *First M87 Event Horizon Telescope Results. VI. The Shadow and Mass of the Central Black Hole*, Astrophys. J. Lett. **875**, L6 (2019)
DOI: 10.3847/2041-8213/ab1141 arXiv: 1906.11243

[8] M. Alcubierre: *Introduction to 3+1 Numerical Relativity*, Oxford Univ. Press, Oxford (2008)
DOI: 10.1093/acprof:oso/9780199205677.001.0001

[9] S. Alexakis, A. D. Ionescu, and S. Klainerman: *Rigidity of stationary black holes with small angular momentum on the horizon*, Duke Math. J. **163**, 2603 (2014)
DOI: 10.1215/00127094-2819517

[10] L. Andersson, T. Bäckdahl, and P. Blue: *Geometry of black hole spacetimes*, lectures notes, 2014 Summer School on *Asymptotic Analysis in General Relativity*, Institut Fourier, Grenoble
https://arxiv.org/abs/1610.03540

[11] A. Ashtekar, C. Beetle, and J. Lewandowski: *Geometry of generic isolated horizons*, Class. Quantum Grav. **19**, 1195 (2002)
DOI: 10.1088/0264-9381/19/6/311 arXiv: gr-qc/0111067

[12] A. Ashtekar, S. Fairhurst, and B. Krishnan: *Isolated horizons: Hamiltonian evolution and the first law*, Phys. Rev. D **62**, 104025 (2000).

[13] A. Ashtekar and B. Krishnan: *Isolated and dynamical horizons and their applications*, Living Rev. Relativity **7**, 10 (2004)
http://www.livingreviews.org/lrr-2004-10

[14] C. Bambi: *Black Holes: A Laboratory for Testing Strong Gravity*, Springer, Singapore (2017)
DOI: 10.1007/978-981-10-4524-0

[15] G.V. Bard: *Sage for Undergraduates*, Americ. Math. Soc. (2015);
preprint freely downloadable from http://www.gregorybard.com/

[16] J. M. Bardeen: *A Variational Principle for Rotating Stars in General Relativity*, Astrophys. J. **162**, 71 (1970)
DOI: 10.1086/150635

[17] J. M. Bardeen: *Timelike and Null Geodesics in the Kerr Metric*, in *Black Holes – Les astres occlus*, edited by C. DeWitt and B. DeWitt, Gordon and Breach, New York (1973), p. 215.

[18] J.W. Bardeen, B. Carter, and S.W. Hawking: *The four laws of black hole mechanics*, Commun. Math. Phys., **31**, 161 (1973).

[19] J.W. Bardeen, W.H. Press, and S.A. Teukolsky: *Rotating Black Holes: Locally Nonrotating Frames, Energy Extraction, and Scalar Synchrotron Radiation*, Astrophys. J. **178**, 347 (1972)
DOI: 10.1086/151796

[20] T.W. Baumgarte and S.L. Shapiro: *Numerical Relativity: Solving Einstein's Equations on the Computer*, Cambridge Univ. Press, Cambridge (2010)
DOI: 10.1017/CBO9781139193344

[21] K. Beckwith and C. Done: *Extreme gravitational lensing near rotating black holes*, Month. Not. Roy. Astron. Soc. **359**, 1217 (2005)
DOI: 10.1111/j.1365-2966.2005.08980.x arXiv: astro-ph/0411339

[22] R. Bender et al.: *HST STIS Spectroscopy of the Triple Nucleus of M31: Two Nested Disks in Keplerian Rotation around a Supermassive Black Hole*, Astrophys. J. **631**, 280 (2005)
DOI: 10.1086/432434 arXiv: astro-ph/0509839

[23] M. Berger: *A Panoramic View of Riemannian Geometry*, Springer, Berlin (2003).

[24] J. Bičák and Z. Stuchlík: *On the Latitudinal and Radial Motion in the Field of a Rotating Black Hole*, Bul. Astron. Inst. Czechoslovakia **27**, 129 (1976)
https://ui.adsabs.harvard.edu/abs/1976BAICz..27..129B

[25] D. Bini, A. Geralico, and R.T. Jantzen: *Gyroscope precession along general timelike geodesics in a Kerr black hole spacetime*, Phys. Rev. D **95**, 124022 (2017)
DOI: 10.1103/PhysRevD.95.124022 arXiv: 1703.09525

[26] R.L. Bishop and S.I. Goldberg: *Tensor Analysis on Manifolds*, Macmillan, New York (1968); reprinted by Dover, New York (1980).

[27] G.S. Bisnovatyi-Kogan and O.Yu. Tsupko: *Strong gravitational lensing by Schwarzschild black holes*, Astrofizika **51**, 125 (2008); English translation in Astrophysics **51**, 99 (2008)
DOI: 10.1007/s10511-008-0011-8 arXiv: 0803.2468

[28] H. Bondi: *Spherically Symmetrical Models in General Relativity*, Month. Not. Roy. Astron. Soc. **107**, 410 (1947); reprinted in Gen. Relativ. Gravit. **31**, 1783 (1999)
DOI: 10.1023/A:1026726520289

[29] R.H. Boyer: *Geodesic Killing orbits and bifurcate Killing horizons*, Proc. Roy. Soc. A **311**, 245 (1969)
DOI: 10.1098/rspa.1969.0116

[30] R.H. Boyer and R.W. Lindquist: *Maximal Analytic Extension of the Kerr Metric*, J. Math. Phys. **8**, 265 (1967)
DOI: 10.1063/1.1705193

[31] R.H. Boyer and T.G. Price: *An interpretation of the Kerr metric in general relativity*, Math. Proc. Cambridge Phil. Soc. **61**, 531 (1965)
DOI: 10.1017/S0305004100004096

[32] D. Brill: *History of a black hole horizon*, Grav. Cosmo. **20**, 165 (2014)
DOI: 10.1134/S0202289314030050

[33] L.M. Burko, G. Khanna, and A. Zenginoğlu: *Cauchy-horizon singularity inside perturbed Kerr black holes*, Phys. Rev. D **93**, 041501(R) (2016)
DOI: 10.1103/PhysRevD.93.041501

[34] P.F. Byrd and M.D. Friedman: *Handbook of Elliptic Integrals for Engineers and Scientists*, 2nd ed., Springer (Berlin) (1971)
DOI: 10.1007/978-3-642-65138-0

[35] A. Čadež and U. Kostić: *Optics in the Schwarzschild spacetime*, Phys. Rev. D **72**, 104024 (2005)
DOI: 10.1103/PhysRevD.72.104024

[36] V. Cardoso: *The hole picture*, Nature Review Phys. **1**, 701 (2019)
DOI: 10.1038/s42254-019-0119-2 arXiv: 1910.04173

[37] V. Cardoso, F. Duque, and A. Foschi: *The light ring and the appearance of matter accreted by black holes*, preprint (2021)
arXiv: 2102.07784

[38] S.M. Carroll: *Spacetime and Geometry: An Introduction to General Relativity*, Addison Wesley (Pearson Education), San Fransisco (2004);
http://preposterousuniverse.com/spacetimeandgeometry/

[39] B. Carter: *Complete Analytic Extension of the Symmetry Axis of Kerr's Solution of Einstein's Equations*, Phys. Rev. **141**, 1242 (1966)
DOI: 10.1103/PhysRev.141.1242

[40] B. Carter: *Stationary Axisymmetric Systems in General Relativity*, Ph. D. Thesis, Cambridge University (1967); available at
https://luth.obspm.fr/~luthier/carter/Thesis/

[41] B. Carter: *Global Structure of the Kerr Family of Gravitational Fields*, Phys. Rev. **174**, 1559 (1968)
DOI: 10.1103/PhysRev.174.1559

[42] B. Carter: *Hamilton-Jacobi and Schrodinger Separable Solutions of Einstein's Equations*, Commun. Math. Phys. **10**, 280 (1968)
DOI: 10.1007/BF03399503

[43] B. Carter: *Killing horizons and orthogonally transitive groups in space-time*, J. Math. Phys. **10**, 70 (1969).

[44] B. Carter: *Axisymmetric Black Hole Has Only Two Degrees of Freedom*, Phys. Rev. Lett. **26**, 331 (1971)
DOI: 10.1103/PhysRevLett.26.331

[45] B. Carter: *Black hole equilibrium states Part I. Analytic and Geometric Properties of the Kerr Solutions*, in *Black Holes – Les astres occlus*, edited by C. DeWitt and B. DeWitt, Gordon and Breach, New York (1973), p. 57; reprinted (with corrections) in Gen. Relativ. Gravit. **41**, 2873 (2009)
DOI: 10.1007/s10714-009-0888-5

[46] B. Carter: *Black hole equilibrium states Part II. General theory of stationary black hole states*, in *Black Holes – Les astres occlus*, edited by C. DeWitt and B. DeWitt, Gordon and Breach, New York (1973), p. 125; reprinted (with corrections) in Gen. Relativ. Gravit. **42**, 653 (2010)
DOI: 10.1007/s10714-009-0920-9

[47] B. Carter: *Mathematical foundations of the theory of relativistic stellar and black hole configurations*, in *Gravitation in Astrophysics*, Eds. B. Carter and J.B. Hartle, Plenum Press, New York (1987).

[48] B. Carter: *Extended tensorial curvature analysis for embeddings and foliations*, Contemp. Math. **203**, 207 (1997).

[49] B. Carter: *Has the black hole equilibrium problem been solved?*, in *The Eighth Marcel Grossmann Meeting*, edited by T. Piran, World Scientific, Singapore (1999)
arXiv: gr-qc/9712038

[50] S. Chandrasekhar: *The Mathematical Theory of Black Holes*, Oxford Univ. Press, New York (1983).

[51] Y. Choquet-Bruhat: *General Relativity and Einstein's Equations*, Oxford Univ. Press, New York (2009).

[52] Y. Choquet-Bruhat: *Introduction to General Relativity, Black Holes, and Cosmology*, Oxford Univ. Press, New York (2015).

[53] Y. Choquet-Bruhat, C. De Witt-Moretten, and M. Dillard-Bleick: *Analysis, Manifolds and Physics*, North-Holland, Amsterdam (1977).

[54] P.T. Chruściel: *On Rigidity of Analytic Black Holes*, Commun. Math. Phys. **189**, 1 (1997)
DOI: 10.1007/s002200050187

[55] P.T. Chruściel: *Black holes*, in *Proceedings of the Tübingen conference on conformal structure of space-time*, edited by J. Frauendiener, H. Friedrich, Lecture Notes in Physics **604**, 61, Springer, Berlin (2002).

[56] P.T. Chruściel: *Black holes – An Introduction*, in *100 Years of Relativity*, edited by A. Ashtekar, World Scientific, Singapore (2005), p. 93
DOI: 10.1142/9789812700988_0004

[57] P.T. Chruściel: *Geometry of Black Holes*, Oxford Univ. Press, Oxford (2020)
DOI: 10.1093/oso/9780198855415.001.0001
https://global.oup.com/academic/product/geometry-of-black-holes-9780198855415

[58] P. Chruściel, E. Delay, G. Galloway, and R. Howard: *Regularity of horizons and the area theorem*, Annales Henri Poincaré **2**, 109 (2001)
DOI: 10.1007/PL00001029

[59] P. Chruściel and G.J. Galloway: *Uniqueness of static black holes without analyticity*, Class. Quantum Grav. **27**, 152001 (2010)
DOI: 10.1088/0264-9381/27/15/152001

[60] P.T. Chruściel, J. Lopes Costa, and M. Heusler: *Stationary Black Holes: Uniqueness and Beyond*, Living Rev. Relativity **15**, 7 (2012)
http://www.livingreviews.org/lrr-2012-7

[61] P.T. Chruściel, M. Maliborski, and N. Yunes: *Structure of the singular ring in Kerr-like metrics*, Phys. Rev. D **101**, 104048 (2020)
DOI: 10.1103/PhysRevD.101.104048 arXiv: 1912.06020

[62] P.T. Chruściel and L. Nguyen: *A Uniqueness Theorem for Degenerate Kerr–Newman Black Holes*, Ann. Henri Poincaré **11**, 585 (2010)
DOI: 10.1007/s00023-010-0038-3

[63] P.T. Chruściel, C.R. Ölz, and S.J Szybka: *Space-time diagrammatics*, Phys. Rev. D **86**, 124041 (2012)
DOI: 10.1103/PhysRevD.86.124041 arXiv: 1211.1718

[64] M.I. Cohen, J.D. Kaplan, and M.A. Scheel: *Toroidal horizons in binary black hole inspirals*, Phys. Rev. D **85**, 024031 (2012)
DOI: 10.1103/PhysRevD.85.024031

[65] M.I. Cohen, H.P. Pfeiffer, and M.A. Scheel: *Revisiting event horizon finders*, Class. Quantum Grav. **26**, 035005 (2009)
DOI: 10.1088/0264-9381/26/3/035005

[66] P. Collas and D. Klein: *Embeddings and time evolution of the Schwarzschild wormhole*, Amer. J. Phys. **80**, 203 (2012)
DOI: 10.1119/1.3672848

[67] G. Compère: *An introduction to the mechanics of black holes*, Lecture notes, Second Modave Summer School in Mathematical Physics (2006)
arXiv: gr-qc/0611129

[68] G. Compère: *Advanced Lectures on General Relativity*, Springer (Berlin)
DOI: 10.1007/978-3-030-04260-8 arXiv: 1801.07064

[69] G. Contopoulos: *Orbits through the ergosphere of a Kerr black hole*, Gen. Relat. Grav. **16**, 43 (1984)
DOI: 10.1007/BF00764017

[70] R.M. Corless, G.H. Gonnet, D.E.G. Hare, D.J. Jeffrey, and D.E. Knuth: *On the Lambert W function*, Adv. Comput. Math. **5**, 329 (1996)
DOI: 10.1007/BF02124750

[71] P.V.P. Cunha and C.A.R. Herdeiro: *Shadows and strong gravitational lensing: a brief review*, Gen. Relat. Grav. **50**, 42 (2018)
DOI: 10.1007/s10714-018-2361-9 arXiv: 1801.00860

[72] P.V.P. Cunha, C.A.R. Herdeiro and E. Radu: *Fundamental photon orbits: Black hole shadows and spacetime instabilities*, Phys. Rev. D **96**, 024039 (2017)
DOI: 10.1103/PhysRevD.96.024039 arXiv: 1705.05461

[73] C.T. Cunningham and J.M. Bardeen: *The Optical Appearance of a Star Orbiting an Extreme Kerr Black Hole* Astrophys. J. **183**, 237 (1973)
DOI: 10.1086/152223

[74] H.D. Curtis: *Descriptions of 762 Nebulae and Clusters Photographed with the Crossley Reflector*, Publications of Lick Observatory **13**, 9 (1918)
URL: https://ui.adsabs.harvard.edu/abs/1918PLicO..13....9C

[75] M. Dafermos and I. Rodnianski: *Lecture on black holes and linear waves*, in Clay Math. Proc. **17**: *Evolution Equations*, edited by D. Ellwood, I. Rodnianski, G. Staffilani and J. Wunsch, AMS / Clay Mathematical Institute, Cambridge (2013)
http://people.maths.ox.ac.uk/cmi/library/proceedings/cmip017c.pdf

[76] T. Damour: *Quelques propriétés mécaniques, électromagnétiques, thermodynamiques et quantiques des trous noirs*, Thèse de doctorat d'État, Université Paris 6 (1979); available at
http://pagesperso.ihes.fr/~damour/Articles/

[77] T. Damour: *Surface effects in black hole physics*, in *Proceedings of the Second Marcel Grossmann Meeting on General Relativity*, Ed. R. Ruffini, North Holland (1982), p. 587; available at
http://pagesperso.ihes.fr/~damour/Articles/

[78] T. Damour: *The Entropy of Black Holes: A Primer*, in *Poincaré Seminar 2003: Bose-Einstein Condensation – Entropy*, edited by J. Dalibard, B. Duplantier and V. Rivasseau, Birkhaäuser, Basel (2004), p. 227

[79] C. G. Darwin: *The gravity field of a particle*, Proc. Roy. Soc. A **249**, 180 (1959)
DOI: 10.1098/rspa.1959.0015

[80] C. G. Darwin: *The gravity field of a particle. II*, Proc. Roy. Soc. A **263**, 39 (1961)
DOI: 10.1098/rspa.1961.0142

[81] B. Datt: *On a Class of Solutions of the Gravitation Equations of Relativity*, Zeitschrift für Physik **108**, 314 (1938); reprinted in Gen. Relativ. Gravit. **31**, 1619 (1999)
DOI: 10.1023/A:1026742707143

[82] F. de Felice: *Equatorial geodesic motion in the gravitational field of a rotating source*, Nuovo Cimento B **57**, 351 (1968)
DOI: 10.1007/BF02710207

[83] C. De Jans: *Sur le mouvement d'une particule matérielle dans un champ de gravitation à symétrie sphérique*, Mem. Acad. Roy. Belgique Cl. Sci. **7**, 1 (1923).

[84] N. Deruelle: *Black Holes in General Relativity*, lectures at Institut de Physique Théorique, CEA, Saclay (France) (2009); slides and notes available at
http://ipht.cea.fr/Docspht/search/article.php?id=t09/346

[85] N. Deruelle and J.-P. Uzan: *Théories de la Relativité*, Belin, Paris (2014).

[86] N. Deruelle and J.-P. Uzan: *Relativity in Modern Physics*, Oxford Univ. Press, Oxford (2018)
DOI: 10.1093/oso/9780198786399.001.0001

[87] C. DeWitt and B. DeWitt (eds.): *Black Holes – Les astres occlus*, Gordon and Breach, New York (1973).

[88] J. Dexter and E. Agol: *A Fast New Public Code for Computing Photon Orbits in a Kerr Spacetime*, Astrophys. J. **696**, 1616 (2009)
DOI: 10.1088/0004-637X/696/2/1616 arXiv: 0903.0620

[89] V.I. Dokuchaev: *Is there life inside black holes?*, Class. Quantum Grav. **28**, 235015 (2011)
DOI: 10.1088/0264-9381/28/23/235015 arXiv: 1103.6140

[90] V.I. Dokuchaev: *To see the invisible: Image of the event horizon within the black hole shadow*, Int. J. Mod. Phys. D **28**, 1941005 (2019)
DOI: 10.1142/S0218271819410050 arXiv: 1812.06787

[91] V.I. Dokuchaev and N.O. Nazarova: *Event Horizon Image within Black Hole Shadow*, J. Exp. Theor. Phys. **128**, 578 (2019)
DOI: 10.1134/S1063776119030026 arXiv: 1804.08030

[92] V.I. Dokuchaev and N.O. Nazarova: *Silhouettes of invisible black holes*, Physics-Uspekhi **63**, 583 (2020)
DOI: 10.3367/UFNe.2020.01.038717 arXiv: 1911.07695

[93] V.I. Dokuchaev and N.O. Nazarova: *Visible shapes of black holes M87* and SgrA**, preprint
arXiv: 2007.14121

[94] S. Drasco and S.A. Hughes: *Rotating black hole orbit functionals in the frequency domain*, Phys. Rev. D **69**, 044015 (2004)
DOI: 10.1103/PhysRevD.69.044015 arXiv: astro-ph/0308479

[95] S. Drasco and S.A. Hughes: *Gravitational wave snapshots of generic extreme mass ratio inspirals*, Phys. Rev. D **73**, 024027 (2006)
DOI: 10.1103/PhysRevD.73.024027 arXiv: gr-qc/0509101

[96] J. Droste: *The Field of a Single Centre in Einstein's Theory of Gravitation, and the Motion of a Particle in that Field*, Koninklijke Nederlandsche Akademie van Wetenschappen, Proceedings **19**, 197 (1917); reprinted in Gen. Relativ. Gravit. **34**, 1545 (2002)
DOI: 10.1023/A:1020747322668

[97] A.S. Eddington: *A comparison of Whitehead's and Einstein's Formulæ*, Nature **113**, 192 (1924)
DOI: 10.1038/113192a0

[98] A. Einstein and N. Rosen: *The Particle Problem in the General Theory of Relativity*, Phys. Rev. **48**, 73 (1935)
DOI: 10.1103/PhysRev.48.73

[99] J. Eisenstaedt: *Histoire et Singularités de la Solution de Schwarzschild (1915-1923)*, Archive for History of Exact Sciences **27**, 157 (1982)
DOI: 10.1007/BF00348347

[100] J. Eisenstaedt: *Trajectoires et Impasses de la Solution de Schwarzschild*, Archive for History of Exact Sciences **37**, 275 (1987)
DOI: 10.1007/BF00417007

[101] J. Eisenstaedt: *Lemaître and the Schwarzschild Solution*, in *The Attraction of Gravitation: New Studies in the History of General Relativity*, Proc. Third International Conference on the History and Philosophy of General Relativity. Einstein Studies, Vol. 5, edited by J. Earman, M. Janssen, and J.D. Norton, Birkhäuser, Boston (1993), p. 353.

[102] H. Eschrig: *Topology and Geometry for Physics*, Springer, Berlin (2011).

[103] H. Falcke: *Imaging black holes: past, present and future*, J. Phys. Conf. Ser. **942**, 012001 (2017)
DOI: 10.1088/1742-6596/942/1/012001 arXiv: 1801.03298

[104] H. Falcke, F. Melia, and E. Agol: *Viewing the Shadow of the Black Hole at the Galactic Center*, Astrophys. J. **528**, L13 (2000)
DOI: 10.1086/312423 arXiv: astro-ph/9912263

[105] J.R. Farah, D.W. Pesce, M.D. Johnson, and L. Blackburn: *On the Approximation of the Black Hole Shadow with a Simple Polar Curve*, Astrophys. J. **900**, 77 (2020)
DOI: 10.3847/1538-4357/aba59a arXiv: 2007.06732

[106] D. Finkelstein: *Past-Future Asymmetry of the Gravitational Field of a Point Particle*, Phys. Rev. **110**, 965 (1958)
DOI: 10.1103/PhysRev.110.965

[107] L. Flamm: *Beiträge zur Einsteinschen Gravitationstheorie*, Physikalische Zeitschrift **17**, 448 (1916); English translation in Gen. Relativ. Gravit. **47**, 72 (2015)
DOI: 10.1007/s10714-015-1908-2

[108] J. Frauendiener: *Conformal Infinity*, Living Rev. Relativity **7**, 1 (2004)
DOI: 10.12942/lrr-2004-1

[109] V. P. Frolov, P. Krtouš, and D. Kubizňák: *Black holes, hidden symmetries, and complete integrability*, Living Rev. Relativity **20**, 6 (2017)
DOI: 10.1007/s41114-017-0009-9

[110] V.P. Frolov and I.D. Novikov: *Black Hole Physics: Basic Concepts and New Developments*, Kluwer Academic Publisher, Dordrecht (1998); reprinted by Springer, Dordrecht (1998).
DOI: 10.1007/978-94-011-5139-9

[111] V.P. Frolov and A. Zelnikov: *Introduction to Black Hole Physics*, Oxford Univ. Press, New York (2011).

[112] C. Fronsdal: *Completion and Embedding of the Schwarzschild Solution*, Phys. Rev. **116**, 778 (1959)
DOI: 10.1103/PhysRev.116.778

[113] R. Fujita & W. Hikida: *Analytical solutions of bound timelike geodesic orbits in Kerr spacetime*, Class. Quantum Grav. **26**, 135002 (2009)
DOI: 10.1088/0264-9381/26/13/135002 arXiv: 0906.1420

[114] K. Gebhardt, J. Adams, D. Richstone, T.R. Lauer, S.M. Faber, K. Gültekin, J. Murphy, and S. Tremaine: *The Black Hole Mass in M87 from Gemini/NIFS Adaptive Optics Observations*, Astrophys. J. **729**, 119 (2011)
DOI: 10.1088/0004-637X/729/2/119

[115] R. Genzel, F. Eisenhauer and S. Gillessen: *The Galactic Center massive black hole and nuclear star cluster*, Rev. Mod. Phys. **82**, 3121 (2010)
DOI: 10.1103/RevModPhys.82.3121

[116] R. Geroch and J.B. Hartle: *Distorted black holes*, J. Math. Phys. **23**, 680 (1982).

[117] G. W. Gibbons and M. Vyska: *The application of Weierstrass elliptic functions to Schwarzschild null geodesics*, Class. Quantum Grav. **29**, 065016 (2012)
DOI: 10.1088/0264-9381/29/6/065016 arXiv: 1110.6508

[118] G. Gibbons and C.M. Will: *On the Multiple Deaths of Whitehead's Theory of Gravity*, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **39**, 41 (2008)
DOI: 10.1016/j.shpsb.2007.04.004

[119] K. Glampedakis and D. Kennefick: *Zoom and whirl: Eccentric equatorial orbits around spinning black holes and their evolution under gravitational radiation reaction* Phys. Rev. D **66**, 044002 (2002) DOI: 10.1103/PhysRevD.66.044002

[120] K. Glampedakis and G. Pappas: *Can supermassive black hole shadows test the Kerr metric?*, preprint (2021)
arXiv: 2102.13573

[121] L. Gou, J.E. McClintock, R.A. Remillard, J.F. Steiner, M.J. Reid, J.A. Orosz, R. Narayan, M. Hanke, and J. García: *Confirmation via the Continuum-Fitting Method that the Spin of the Black Hole in Cygnus X-1 is Extreme*, Astrophys. J. **790**, 29 (2014)
DOI: 10.1088/0004-637X/790/1/29

[122] E. Gourgoulhon: *3+1 Formalism in General Relativity; Bases of Numerical Relativity*, Springer, Berlin (2012)
https://relativite.obspm.fr/3p1

[123] E. Gourgoulhon: *Special Relativity in General Frames: From Particles to Astrophysics*, Springer, Berlin (2013)
https://relativite.obspm.fr/sperel

[124] E. Gourgoulhon: *Relativité générale*, Master course lecture notes
https://luth.obspm.fr/~luthier/gourgoulhon/fr/master/relat.html

[125] E. Gourgoulhon and J.L. Jaramillo: *A 3+1 perspective on null hypersurfaces and isolated horizons*, Phys. Rep. **423**, 159 (2006)
DOI: 10.1016/j.physrep.2005.10.005

[126] E. Gourgoulhon and J. L. Jaramillo: *New theoretical approaches to black holes*, in Proceedings of the Conference *Jean-Pierre Lasota, X-ray Binaries, Accretion Disks and Compact Stars* held in Trzebieszowice, Poland (7-13 October 2007), edited by M. Abramowicz and O. Straub, New Astronomy Reviews **51**, 791 (2008)
DOI: 10.1016/j.newar.2008.03.026

[127] I.S. Gradshteyn, I.M. Ryzhik, Y.V. Geronimus, M.Y. Tseytlin, and A. Jeffrey: *Table of Integrals, Series, and Products* (8 ed.), Academic Press (2015).

[128] S.E. Gralla: *Can the EHT M87 results be used to test general relativity?*, Phys. Rev. D **103**, 024023 (2021)
DOI: 10.1103/PhysRevD.103.024023 arXiv: 2010.08557

[129] S.E. Gralla, D.E. Holz, and R.M. Wald: *Black hole shadows, photon rings, and lensing rings*, Phys. Rev. D **100**, 024018 (2019)
DOI: 10.1103/PhysRevD.100.024018 arXiv: 1906.00873

[130] S.E. Gralla and A. Lupsasca: *Lensing by Kerr black holes*, Phys. Rev. D **101**, 044031 (2020)
DOI: 10.1103/PhysRevD.101.044031 arXiv: 1910.12873

[131] S.E. Gralla and A. Lupsasca: *Null geodesics of the Kerr exterior*, Phys. Rev. D **101**, 044032 (2020)
DOI: 10.1103/PhysRevD.101.044032 arXiv: 1910.12881

[132] S.E. Gralla and A. Lupsasca: *Observable shape of black hole photon rings*, Phys. Rev. D **102**, 124003 (2020)
DOI: 10.1103/PhysRevD.102.124003 arXiv: 2007.10336

[133] S.E. Gralla, A. Lupsasca and A. Strominger: *Observational signature of high spin at the Event Horizon Telescope*, Month. Not. Roy. Astron. Soc. **475**, 3829 (2018)
DOI: 10.1093/mnras/sty039 arXiv: 1710.11112

[134] A.A. Grib and Y.V. Pavlov: *Black holes and particles with zero or negative energy*, Theor. Math. Phys. **190**, 268 (2017)
DOI: 10.1134/S0040577917020088 arXiv: 1601.02592

[135] R. Grossman, J. Levin, and G. Perez-Giz: *Harmonic structure of generic Kerr orbits*, Phys. Rev. D **85**, 023012 (2012)
DOI: 10.1103/PhysRevD.85.023012

[136] A. Gullstrand: *Allgemeine Lösung des statischen Einkörperproblems in der Einsteinschen Gravitationstheorie*, Arkiv. Mat. Astron. Fys. **16**, 1 (1922).

[137] Y. Hagihara: *Theory of the Relativistic Trajectories in a Gravitational Field of Schwarzschild*, Japanese Journal of Astronomy and Geophysics **8**, 67 (1931); available at
https://ui.adsabs.harvard.edu/abs/1930JaJAG...8...67H

[138] P. Hájiček: *Exact models of charged black holes. I. Geometry of totally geodesic null hypersurface*, Commun. Math. Phys. **34**, 37 (1973).
DOI: 10.1007/BF01646541

[139] P. Hájiček: *Exact models of charged black holes. II. Axisymmetric stationary horizons*, Commun. Math. Phys. **34**, 53 (1973).
DOI: 10.1007/BF01646542

[140] P. Hájiček: *Can outside fields destroy black holes ?*, J. Math. Phys. **15**, 1554 (1974).
DOI: 10.1063/1.1666846

[141] J. Haláček and T. Ledvinka: *The analytic conformal compactification of the Schwarzschild spacetime*, Class. Quantum Grav. **31**, 015007 (2014).
DOI: 10.1088/0264-9381/31/1/015007

[142] J.B. Hartle: *Gravity: An Introduction to Einstein's General Relativity*, Addison Wesley (Pearson Education), San Fransisco (2003);
http://wps.aw.com/aw_hartle_gravity_1/

[143] S.W. Hawking: *Gravitational radiation from colliding black holes*, Phys. Rev. Lett. **26**, 1344 (1971).

[144] S.W. Hawking: *Black holes in general relativity*, Commun. Math. Phys. **25**, 152 (1972)
DOI: 10.1007/BF01877517

[145] S.W. Hawking: *The event horizon*, in *Black Holes – Les astres occlus*, edited by C. DeWitt and B. DeWitt, Gordon and Breach, New York (1973), p. 1.

[146] S.W. Hawking and G.F.R. Ellis: *The large scale structure of space-time*, Cambridge University Press, Cambridge (1973).

[147] S.W. Hawking: *The Nature of Space and Time* (1994)
arXiv: hep-th/9409195

[148] S.W. Hawking and R. Penrose: *The Nature of Space and Time*, Princeton Univ. Press, Princeton (2015).

[149] C. Heinicke and F.W. Hehl: *Schwarzschild and Kerr solutions of Einstein's field equation: An Introduction*, Int. J. Mod. Phys. D **24**, 1530006 (2015)
DOI: 10.1142/S0218271815300062

[150] C.A.R. Herdeiro and J.P.S. Lemos: *O buraco negro cinquenta anos depois: A génese do nome*, Gazeta de Fisica **41**, 2 (2018); English translation in *The black hole fifty years after: Genesis of the name*
arXiv: 1811.06587

[151] W.C. Hernandez and C.W. Misner: *Observer Time as a Coordinate in Relativistic Spherical Hydrodynamics*, Astrophys. J. **143**, 452 (1966)
DOI: 10.1086/148525

[152] D. Hilbert: *Die Grundlagen der Physik (Zweite Mitteilung)*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalishe Klasse, Jahr 1917, 53 (1917); available at
http://resolver.sub.uni-goettingen.de/purl?PPN252457811_1917 and
http://www.digizeitschriften.de/dms/resolveppn/?PID=GDZPPN002504561

[153] D. Hilbert: *Die Grundlagen der Physik II*, lectures given at Göttingen University in winter 1916-1917, lecture notes written by R. Bär; reprinted in [226], p. 162-307.

[154] S. Husa and J. Winicour: *Asymmetric merger of black holes* Phys. Rev. D **60**, 084019 (1999)
DOI: 10.1103/PhysRevD.60.084019

[155] W. Israel: *Event Horizons in Static Vacuum Space-Times*, Phys. Rev. **164**, 1776 (1967)
DOI: 10.1103/PhysRev.164.1776

[156] W. Israel: *Event horizons in static electrovac space-times*, Commun. Math. Phys. **8**, 245 (1968)
DOI: 10.1007/BF01645859

[157] J.L. Jaramillo: *An Introduction to Local Black Hole Horizons in the 3+1 Approach of General Relativity*, in *Black Holes: New Horizon*, edited by S.A. Hayward, World Scientific, Singapore (2013)
arXiv: 1108.2408

[158] M. Jaroszyński and A. Kurpiewski: *Optics near Kerr black holes: spectra of advection dominated accretion flows*, Astron. Astrophys. **326**, 419 (1997)
https://ui.adsabs.harvard.edu/abs/1997A&A...326..419J arXiv: astro-ph/9705044

[159] T. Johannsen: *Sgr A* and general relativity*, Class. Quantum Grav. **33**, 113001 (2016)
DOI: 10.1088/0264-9381/33/11/113001

[160] T. Johannsen and D. Psaltis: *Testing the No-Hair Theorem with Observations in the Electromagnetic Spectrum: II. Black-Hole Images*, Astrophys. J. **718**, 446 (2010)
DOI: 10.1088/0004-637X/718/1/446 arXiv: 1005.1931

[161] M.D. Johnson, A. Lupsasca, A. Strominger, G.N. Wong, S. Hadar, D. Kapec, R. Narayan, A. Chael, C.F. Gammie, P. Galison, D.C. Palumbo, S.S. Doeleman, L. Blackburn, M. Wielgus, D.W. Pesce, J.R. Farah, and J.M. Moran: *Universal interferometric signatures of a black hole's photon ring*, Science Advances **6**, eaaz1310 (2020)
DOI: 10.1126/sciadv.aaz1310 arXiv: 1907.04329

[162] D. Joyner and W. Stein: *Sage Tutorial*, CreateSpace (2014).

[163] J. Katz: *A note on Komar's anomalous factor*, Class. Quantum Grav. **2**, 423 (1985)
DOI: 10.1088/0264-9381/2/3/018

[164] R.P. Kerr: *Gravitational Field of a Spinning Mass as an Example of Algebraically Special Metrics*, Phys. Rev. Lett. **11**, 237 (1963)
DOI: 10.1103/PhysRevLett.11.237

[165] R.P. Kerr: *Discovering the Kerr and Kerr-Schild metrics*, in *The Kerr spacetime*, edited by D.L. Wiltshire, M. Visser, and S.M. Scott, Cambridge Univ. Press, Cambridge (2009), p. 38
arXiv: 0706.1109

[166] R.P. Kerr and A. Schild: *A New Class of Vacuum Solutions of the Einstein Field Equations*, in *Atti del Convegno sulla Relativita Generale: Problemi dell'Energia e Onde Gravitazionali*, G. Barbèra Editore, Firenze (1965), p. 1; reprinted in Gen. Relativ. Gravit. **41**, 2485 (2009)
DOI: 10.1007/s10714-009-0857-z

[167] S. Kobayashi and K. Nomizu: *Foundations of Differential Geometry*, vol. 1, Interscience, New York (1963).

[168] F. Kottler: *Über die physikalischen Grundlagen der Einsteinschen Gravitationstheorie*, Annalen der Physik **56**, 401 (1918)
DOI: 10.1002/andp.19183611402

[169] A. Krasuiński: *Editor's Note: The Expanding Universe*, Gen. Relativ. Gravit. **29**, 637 (1997)

[170] M. D. Kruskal: *Maximal Extension of Schwarzschild Metric*, Phys. Rev. **119**, 1743 (1960)
DOI: 10.1103/PhysRev.119.1743

[171] J. Lafontaine: *An Introduction to Differential Manifolds*, Springer, Cham (2015).

[172] D. Langlois: *Relativité générale*, Vuibert, Paris (2013).

[173] J.-P. Lasota, E. Gourgoulhon, M. Abramowicz, A. Tchekhovskoy, and R. Narayan: *Extracting black-hole rotational energy: The generalized Penrose process*, Phys. Rev. D **89**, 024041 (2014)
DOI: 10.1103/PhysRevD.89.024041

[174] M. von Laue: *Die Relativitätstheorie. Zweiter Band: Die Allgemeine Relativitätstheorie Und Einsteins Lehre Von Der Schwerkraft*, Friedr. Vieweg & Sohn, Braunschweig (1921); available at
https://archive.org/details/dierelativitts02laueuoft/page/n5/mode/2up

[175] M. Le Bellac: *Quantum Physics*, Cambridge Univ. Press, Cambridge (2006)
DOI: 10.1017/CBO9780511616471

[176] J. M. Lee: *Introduction to Smooth Manifolds*, 2nd edition, Springer, New-York (2013)
DOI: 10.1007/978-1-4419-9982-5

[177] J. M. Lee: *Introduction to Riemannian Manifolds*, 2nd edition, Springer, Cham (2018)
DOI: 10.1007/978-3-319-91755-9

[178] G. Lemaître: *L'univers en expansion*, Publication du Laboratoire d'Astronomie et de Géodésie de l'Université de Louvain **9**, 171 (1932); reprinted in Annales de la Société Scientifique de Bruxelles A **53**, 51 (1933); English translation in Gen. Relativ. Gravit. **29**, 641 (1997)
DOI: 10.1023/A:1018855621348

[179] J.-P. Luminet: *Image of a spherical black hole with thin accretion disk*, Astron. Astrophys. **75**, 228 (1979).

[180] J.-P. Luminet: *An Illustrated History of Black Hole Imaging: Personal Recollections (1972-2002)*, preprint.
arXiv: 1902.11196

[181] J. Maldacena and L. Susskind: *Cool horizons for entangled black holes*, Fortschritte der Physik **61**, 781 (2013)
DOI: 10.1002/prop.201300020 arXiv: 1306.0533

[182] J.-A. Marck: *Journey to a black hole*, movie (1991)
https://www.youtube.com/watch?v=5Oqop50ltrM
see also https://images.cnrs.fr/video/46

[183] J.-A. Marck: *Short-cut method of solution of geodesic equations for Schwarzchild black hole*, Class. Quantum Grav. **13**, 393 (1996)
DOI: 10.1088/0264-9381/13/3/007 arXiv: gr-qc/9505010

[184] K. Martel and E. Poisson: *Regular coordinate systems for Schwarzschild and other spherical spacetimes*, Amer. J. Phys. **69**, 476 (2001).

[185] R.A. Matzner, H.E. Seidel, S.L. Shapiro, L. Smarr, W.-M. Suen, S.A. Teukolsky, and J. Winicour: *Geometry of a Black Hole Collision*, Science **270**, 941 (1995)
DOI: 10.1126/science.270.5238.941

[186] F. Mertens, A.P. Lobanov, R.C. Walker, and P.E. Hardee: *Kinematics of the jet in M87 on scales of 100-1000 Schwarzschild radii*, Astron. Astrophys. **595**, A54 (2016)
DOI: 10.1051/0004-6361/201628829 arXiv: 1608.05063

[187] J. W. Milnor: *On manifolds homeomorphic to the 7-sphere*, Ann. Math. **64**, 399 (1956).

[188] Y. Mino: *Perturbative approach to an orbital evolution around a supermassive black hole*, Phys. Rev. D **67**, 084027 (2003)
DOI: 10.1103/PhysRevD.67.084027 arXiv: gr-qc/0302075

[189] C.W. Misner and D.H. Sharp: *Relativistic Equations for Adiabatic, Spherically Symmetric Gravitational Collapse*, Phys. Rev. **136**, B571 (1964)
DOI: 10.1103/PhysRev.136.B571

[190] C.W. Misner, K.S. Thorne, and J.A. Wheeler: *Gravitation*, Freeman, New York (1973).

[191] V. Moncrief and J. Isenberg: *Symmetries of higher dimensional black holes*, Class. Quantum Grav. **25**, 195015 (2008)
DOI: 10.1088/0264-9381/25/19/195015

[192] G. Muñoz: *Orbits of massless particles in the Schwarzschild metric: Exact solutions*, Amer. J. Phys **82**, 564 (2014)
DOI: 10.1119/1.4866274

[193] R. Narayan, M.D. Johnson, Michael D., and C.F. Gammie: *The Shadow of a Spherically Accreting Black Hole*, Astrophys. J. Lett. **885**, L33 (2019)
DOI: 10.3847/2041-8213/ab518c arXiv: 1910.02957

[194] J. Nash: *The Imbedding Problem for Riemannian Manifolds*, Ann. Math. **63**, 20 (1956)
DOI: 10.2307/1969989

[195] J.-P. Nicolas: *The conformal approach to asymptotic analysis*
arXiv: 1508.02592

[196] I.D. Novikov: *On the evolution of a semiclosed world* (in Russian), Astronomicheskii Zhurnal **40**, 772 (1963); English translation in Soviet Astronomy **7**, 587 (1964)
URL: `https://ui.adsabs.harvard.edu/abs/1964SvA.....7..587N%2F/`

[197] I.D. Novikov and V.P. Frolov: *Physics of Black Holes*, Kluwer, Dordrecht (1989)

[198] I.D. Novikov and K.S. Thorne: *Astrophysics of Black Holes*, in *Black Holes – Les astres occlus*, edited by C. DeWitt and B. DeWitt, Gordon and Breach, New York (1973), p. 343.

[199] B. O'Neill: *Semi-Riemannian Geometry, with Applications to Relativity*, Academic Press, New York (1983).

[200] B. O'Neill: *The geometry of Kerr black holes*, A.K. Peters, Wellesley (MA) (1995); reprinted by Dover, Mineola (NY) (2014).

[201] J.A. Orosz, J.E. McClintock, J.P. Aufdenberg, R.A. Remillard, M.J. Reid, R. Narayan, and L. Gou: *The Mass of the Black Hole in Cygnus X-1*, Astrophys. J. **742**, 84 (2011)
DOI: 10.1088/0004-637X/742/2/84

[202] D.N. Page and K.S. Thorne: *Disk-Accretion onto a Black Hole. Time-Averaged Structure of Accretion Disk*, Astrophys. J. **191**, 499 (1974)
DOI: 10.1086/152990

[203] P. Painlevé: *La mécanique classique et la théorie de la relativité*, Comptes Rendus Acad. Sci. (Paris) **173**, 677 (1921).

[204] E. Papantonopoulos (ed.): *Physics of Black Holes*, Springer, Berlin (2009).

[205] R. Penrose: *Asymptotic Properties of Fields and Space-Times*, Phys. Rev. Lett. **10**, 66 (1963)
DOI: 10.1103/PhysRevLett.10.66

[206] R. Penrose: *Conformal treatment of infinity*, in *Relativity, groups and topology*, edited by B. DeWitt and C. DeWitt, Gordon and Breach, New York (1964), p. 565; reprinted in Gen. Relativ. Gravit. **43**, 901 (2011)
DOI: 10.1007/s10714-010-1110-5

[207] R. Penrose: *Gravitational collapse and space-time singularities*, Phys. Rev. Lett. **14**, 57 (1965).

[208] R. Penrose: *Zero Rest-Mass Fields Including Gravitation: Asymptotic Behaviour*, Proc. Roy. Soc. A **284**, 159 (1965)
http://www.jstor.org/stable/2415306

[209] R. Penrose: *Structure of Space-Time*, in *Battelle Rencontres: 1967 Lectures in Mathematics and Physics*, edited by C.M. DeWitt and J.A. Wheeler, Benjamin, New York (1968), p. 121.

[210] R. Penrose: *Gravitational Collapse: The Role of General Relativity*, Rivista del Nuovo Cimento, Numero Speziale **1**, 252 (1969); reprinted in Gen. Relativ. Gravit. **34**, 1141 (2002)
DOI: 10.1023/A:1016578408204

[211] R. Penrose and G.R. Floyd: *Extraction of Rotational Energy from a Black Hole*, Nature Phys. Sci. **229**, 177 (1971)
DOI: 10.1038/physci229177a0

[212] G. Perez-Giz: *From Measure Zero to Measure Hero: Periodic Kerr Orbits and Gravitational Wave Physics*, PhD thesis, Columbia University, New York (2011)
DOI: 10.7916/D8P84JVB

[213] J. Plebański and A. Krasiński: *An Introduction to General Relativity and Cosmology*, Cambridge Univ. Press, Cambridge (2006).

[214] E. Poisson: *A Relativist's Toolkit, The Mathematics of Black-Hole Mechanics*, Cambridge Univ. Press, Cambridge (2004);
http://www.physics.uoguelph.ca/poisson/toolkit/

[215] E. Poisson and W. Israel: *Internal structure of black holes*, Phys. Rev. D **41**, 1796 (1990)
DOI: 10.1103/PhysRevD.41.1796

[216] E. Poisson and C.M. Will: *Gravity: Newtonian, Post-Newtonian, Relativistic*, Cambridge Univ. Press, Cambridge (2014);
https://www.physics.uoguelph.ca/poisson/research/gravity.html

[217] H. Reall: *Black Holes*, lecture notes for the course on Black Holes in Part III of the Cambridge Mathematical Tripos (2020), available at
http://www.damtp.cam.ac.uk/user/hsr1000/teaching.html

[218] M.J. Reid, J.E. McClintock, R. Narayan, L. Gou, R.A. Remillard, and J.A. Orosz: *The trigonometric parallax of Cygnus X-1*, Astrophys. J. **742**, 83 (2011)
DOI: 10.1088/0004-637X/742/2/83 arXiv: 1106.3688

[219] C.S. Reynolds: *Observational Constraints on Black Hole Spin*, Ann. Rev. Astron. Astrophys., in press (2021)
arXiv: 2011.08948

[220] A. Riazuelo: *Trous noirs et trous de vers*, Pour la Science **372**, 36 (2008).

[221] A. Riazuelo: *Seeing relativity – I. Ray tracing in a Schwarzschild metric to explore the maximal analytic extension of the metric and making a proper rendering of the stars*, Int. J. Mod. Phys. D **28**, 1950042 (2019)
DOI: 10.1142/S0218271819500421 arXiv: 1511.06025

[222] A. Riazuelo: *Seeing relativity – III. Journeying within the Kerr metric toward the negative gravity region*, Int. J. Mod. Phys. D **29**, 2050109 (2020)
DOI: 10.1142/S0218271820501096 arXiv: 2008.04384

[223] W. Rindler: *Visual Horizons in World-Models*, Month. Not. Roy. Astron. Soc. **116**, 662 (1956); reprinted in Gen. Relativ. Gravit. **34**, 133 (2002)
DOI: 10.1023/A:1015347106729

[224] D.C. Robinson: *Uniqueness of the Kerr Black Hole*, Phys. Rev. Lett. **34**, 905 (1975)
DOI: 10.1103/PhysRevLett.34.905

[225] I. Robinson and A. Trautman: *Some Spherical Gravitational Waves in General Relativity*, Proc. Roy. Soc. Lond. **A 265**, 463 (1962)
DOI: 10.1098/rspa.1962.0036

[226] T. Sauer and U. Majer (eds.): *David Hilbert's Lectures on the Foundations of Physics, 1915-1927*, Springer, Berlin (2009)
DOI: 10.1007/b12915

[227] M.A. Scheel, M. Boyle, T. Chu, L.E. Kidder, K.D. Matthews, and H.P. Pfeiffer: *High-accuracy waveforms for binary black hole inspiral, merger, and ringdown*, Phys. Rev. D **79**, 024003 (2009)
DOI: 10.1103/PhysRevD.79.024003

[228] W. Schmidt: *Celestial mechanics in Kerr spacetime*, Class. Quantum Grav. **19**, 2743 (2002)
DOI: 10.1088/0264-9381/19/10/314

[229] K. Schwarzschild: *Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie*, Sitzungsberichte der Königlich Preussischen Akademie der Wissenschaften zu Berlin, Phys.-Math. Klasse 1916, 189 (1916); English translation in Gen. Relativ. Gravit. **35**, 951 (2003)
DOI: 10.1023/A:1022971926521 arXiv: physics/9905030

[230] O. Semerák: *Stationary frames in the Kerr field*, Gen. Relat. Grav. **25**, 1041 (1993)
DOI: 10.1007/BF00763554

[231] J.M.M. Senovilla and D. Garfinkle: *The 1965 Penrose singularity theorem*, Class. Quantum Grav. **32**, 124008 (2015)
DOI: 10.1088/0264-9381/32/12/124008 arXiv: 1410.5226

[232] M. Shibata: *Numerical Relativity*, World Scientific, Singapore (2015)
DOI: 10.1142/9692

[233] M. Siino: *Topological Appearance of Event Horizon*, Prog. Theor. Phys. **99**, 1 (1998).

[234] M. Siino: *Topology of event horizons* Phys. Rev. D **58**, 104016 (1998)
DOI: 10.1103/PhysRevD.58.104016

[235] L.C. Stein and N. Warburton: *Location of the last stable orbit in Kerr spacetime*, Phys. Rev. D **101**, 064007 (2020)
DOI: 10.1103/PhysRevD.101.064007 arXiv: 1912.07609

[236] W. Stein and D. Joyner: *SAGE: System for Algebra and Geometry Experimentation*, Commun. Comput. Algebra **39**, 61 (2005).

[237] N. Straumann: *General Relavity*, 2nd edition, Springer, Dordrecht (2013)
DOI: 10.1007/978-94-007-5410-2

[238] Z. Stuchlík: *The Radial Motion of Photons in Kerr Metric*, Bul. Astron. Inst. Czechoslovakia **32**, 40 (1981)
https://ui.adsabs.harvard.edu/abs/1981BAICz..32...40S

[239] D. Sudarsky and R.M. Wald: *Extrema of mass, stationarity, and staticity, and solutions to the Einstein–Yang–Mills equations*, Phys. Rev. D **46**, 1453 (1992)
DOI: 10.1103/PhysRevD.46.1453

[240] J.L. Synge: *The Gravitational Field of a Particle*, Proc. Roy. Irish Acad. A **53**, 83 (1950)
http://www.jstor.org/stable/20488511

[241] J.L. Synge: *The Escape of Photons from Gravitationally Intense Stars*, Month. Not. Roy. Astron. Soc. **131**, 463 (1966)
DOI: 10.1093/mnras/131.3.463

[242] G. Szekeres: *On the Singularities of a Riemannian Manifold*, Publicationes Mathematicae Debrecen **7**, 285 (1960); reprinted in Gen. Relativ. Gravit. **34**, 2001 (2002)
DOI: 10.1023/A:1020744914721

[243] C. H. Taubes: *Gauge theory on asymptotically periodic 4-manifolds*, J. Differential Geom. **25**, 363 (1987).

[244] E. Teo: *Spherical Photon Orbits Around a Kerr Black Hole*, Gen. Relativ. Gravit. **35**, 1909 (2003)
DOI: 10.1023/A:1026286607562

[245] S.A. Teukolsky: *The Kerr metric*, Class. Quantum Grav. **32** 124006 (2015)
DOI: 10.1088/0264-9381/32/12/124006

[246] K.S. Thorne: *Black Holes and Time Warps*, Norton (1994).

[247] K.S. Thorne and D. MacDonald: *Electrodynamics in curved spacetime: 3 + 1 formulation*, Month. Not. Roy. Astron. Soc. **198**, 339 (1982)
DOI: 0.1093/mnras/198.2.339

[248] R.C. Tolman: *Effect of Inhomogeneity on Cosmological Models*, Proc. Nat. Acad. Sci. USA **20**, 169 (1934); reprinted in Gen. Relativ. Gravit. **29**, 935 (1997)
DOI: 10.1023/A:1018891418565

[249] P.K. Townsend: *Black Holes*, Lecture notes for a course taught in Part III of the Cambridge University Mathematical Tripos (1997)
arXiv: gr-qc/9707012

[250] W. G. Unruh: *Universal coordinates for Schwarzschild black holes* (2014)
arXiv: 1401.3393

[251] K. Van Aelst: *Note on equatorial geodesics in circular spacetimes*, Class. Quantum Grav. **37**, 207001 (2020)
DOI: 10.1088/1361-6382/aba80c

[252] S.U. Viergutz: *Image generation in Kerr geometry. I. Analytical investigations on the stationary emitter-observer problem*
Astron. Astrophys. **272**, 355 (1993)
https://ui.adsabs.harvard.edu/abs/1993A&A...272..355V

[253] F.H. Vincent, T. Paumard, E. Gourgoulhon, and G. Perrin: *GYOTO: a new general relativistic ray-tracing code*, Class. Quantum Grav. **28**, 225011 (2011)
DOI: 10.1088/0264-9381/28/22/225011 arXiv: 1109.4769

[254] F. H. Vincent, M. Wielgus, M. A. Abramowicz, E. Gourgoulhon, J.-P. Lasota, T. Paumard, and G. Perrin: *Geometric modeling of M87* as a Kerr black hole or a non-Kerr compact object*, Astron. Astrophys. **646**, A37 (2021)
DOI: 10.1051/0004-6361/202037787 arXiv: 2002.09226

[255] M. Visser: *The Kerr spacetime — a brief introduction*, in *The Kerr spacetime*, edited by D.L. Wiltshire, M. Visser, and S.M. Scott, Cambridge Univ. Press, Cambridge (2009), p. 3
arXiv: 0706.0622

[256] R.M. Wald: *General relativity*, University of Chicago Press, Chicago (1984).

[257] R.M. Wald: *The thermodynamics of black holes*, Living Rev. Relativity **4**, 6 (2001)
http://www.livingreviews.org/lrr-2001-6

[258] M. Walker and R. Penrose: *On Quadratic First Integrals of the Geodesic Equations for Type* {22} *Spacetimes*, Commun. math. Phys. **18**, 265 (1970)
DOI: 10.1007/BF01649445

[259] R.C. Walker, P.E. Hardee, F.B. Davies, C. Ly, and W. Junor: *The Structure and Dynamics of the Subparsec Jet in M87 Based on 50 VLBA Observations over 17 Years at 43 GHz*, Astrophys. J. **855** 128 (2018)
DOI: 10.3847/1538-4357/aaafcc arXiv: 1802.06166

[260] J.L. Walsh, A.J. Barth, L.C. Ho, and M. Sarzi: *The M87 Black Hole Mass from Gas-dynamical Models of Space Telescope Imaging Spectrograph Observations*, Astrophys. J. **770**, 86 (2013)
DOI: 10.1088/0004-637X/770/2/86 arXiv: 1304.7273

[261] N. Warburton, L. Barack, and N. Sago: *Isofrequency pairing of geodesic orbits in Kerr geometry*, Phys. Rev. D **87**, 084012 (2013)
DOI: 10.1103/PhysRevD.87.084012 arXiv: 1301.3918

[262] H. Weyl: *Über die statischen kugelsymmetrischen Lösungen von Einsteins "kosmologischen" Gravitationsgleichungen*, Physikalische Zeitschrift **20**, 31 (1919).

[263] M. Wielgus et al.: *Monitoring the Morphology of M87* in 2009-2017 with the Event Horizon Telescope*, Astrophys. J. **901**, 67 (2020)
DOI: 10.3847/1538-4357/abac0d arXiv: 2009.11842

[264] D.C. Wilkins: *Bound Geodesics in the Kerr Metric*, Phys. Rev. D **5**, 814 (1972)
DOI: 10.1103/PhysRevD.5.814

[265] P. Zimmermann et al.: *Computational Mathematics with SageMath*, SIAM (2018) freely downloadable from http://sagebook.gforge.inria.fr/english.html

[266] R.L. Znajek: *Black hole electrodynamics and the Carter tetrad*, Month. Not. Roy. Astron. Soc. **179**, 457 (1977)
DOI: 10.1093/mnras/179.3.457

Index