

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

CSGC, IITM, Chennai, January 2022

ENS, Paris, May-June 2023

AEI, Potsdam, December 2023

IHP, Paris, March 2024

ENS, Paris, April-June 2025

Éric Gourgoulhon

Laboratoire d'étude de l'Univers et des phénomènes eXtrêmes

CNRS, Observatoire de Paris - PSL, Sorbonne Université

eric.gourgoulhon@obspm.fr

<https://relativite.obspm.fr/blackholes>

— DRAFT —

version of 28 April 2025

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*:
https://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-lemaître-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:
<https://www.lkb.upmc.fr/gravitationalwaves2018/>
- remotely at the *School on Black Holes and Gravitational Waves* organized at the *Centre for Strings, Gravitation and Cosmology* of the *Indian Institute of Technology Madras*, Chennai (India) in January 2022:
<https://physics.iitm.ac.in/~csgc/events/sbhgw>
- at the *École Normale Supérieure*, Paris (France) in May-June 2023, as part of the PSL graduate programs in Physics and in Astrophysics:
<https://relativite.obspm.fr/blackholes/paris23>
- at the *Albert Einstein Institute*, Potsdam (Germany) in December 2023:
<https://relativite.obspm.fr/blackholes/aei23>
- at the *Institut Henri Poincaré*, Paris (France) in March 2024, within the program *Quantum and classical fields interacting with geometry*:
<https://relativite.obspm.fr/blackholes/ihp24>
- at the *École Normale Supérieure*, Paris (France) in April-June 2025, as part of the PSL graduate program in Physics:
<https://relativite.obspm.fr/blackholes/paris25>

In complement to these notes, one may recommend various monographs devoted to black holes: O'Neill (1995) [390], Heusler (1996) [274], Frolov & Novikov (1998) [201], Poisson (2004) [415], Frolov & Zelnikov (2011) [203], Bambi (2017) [29], Chruściel (2020) [118], Grumiller & Sheikh-Jabbari (2022) [245] and King (2023) [315], as well as review articles by Carter (1987) [100], Wald (2001) [504], Chruściel (2002, 2005) [116, 117] and Chruściel, Lopes Costa & Heusler (2012) [122] and recent monographs with various chapters about black holes: Andersson (2020) [14] and Shibata (2016) [455]. In addition, let us point out other lecture notes on black holes: Hawking (1994) [264, 267], Townsend (1997) [482], Compère (2006, 2019) [135, 137], Dafermos and Rodnianski (2008) [148], Deruelle (2009) [160], Andersson, Bäckdahl & Blue (2016) [13] and Reall (2020) [425].

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 [160], to Kip Thorne's book [478], to Carter's article [104] and to Jean Eisenstaedt's articles [179, 181].

The web pages associated to these notes are

<https://relativite.obspm.fr/blackholes>

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

I warmly thank Cyril Pitrou for having organized the Paris 2016 lectures, Fabio Maltoni and Christophe Ringeval for the Louvain-la-Neuve ones, Anastasia Golubtsova and Irina Pirozhenko for the Dubna ones, Bruce Allen, Marie-Anne Bizouard, Nelson Christensen and Pierre-François Cohadon for the Les Houches ones, Chandra Kant Mishra for the Chennai ones, Jean-François Allemand for the Paris 2023 and 2025 ones, Masaru Shibata and Karim Van Aeslt for the Potsdam ones and Dietrich Häfner, Frédéric Hélein and Michał Wrochna for the Paris 2024 ones.

Besides, I am deeply indebted to Imène Belahcene, Jack Borthwick, Brandon Carter, Marc Casals, Udit Narayan Chowdhury, Stéphane Collion, Xiangyang Chen, Sumit Dey, Jean Eisenstaedt, Romain Gervalle, David Hirondel, Ted Jacobson, Michel Le Bellac, Alexandre Le Tiec, Jean-Philippe Nicolas, Jordan Nicoules, Micaela Oertel, Paul Ramond, Nicolas Seroux and Frédéric Vincent for spotting mistakes, correcting typos and making nice suggestions in preliminary versions of the text.

These notes are released under the

Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License



Contents

I Foundations	11
1 General framework	13
1.1 Introduction	13
1.2 Spacetime	13
1.3 Worldlines	17
1.4 Quantities measured by an observer	20
1.5 Einstein equation	22
2 The concept of black hole 1: Horizons as null hypersurfaces	25
2.1 Introduction	25
2.2 Black holes and null hypersurfaces	25
2.3 Geometry of null hypersurfaces	28
2.4 Evolution of the expansion	53
3 The concept of black hole 2: Non-expanding horizons and Killing horizons	57
3.1 Introduction	57
3.2 Non-expanding horizons	57
3.3 Killing horizons	63
3.4 Bifurcate Killing horizons	83
3.5 Summary	89
4 The concept of black hole 3: The global view	91
4.1 Introduction	91
4.2 Conformal completion of Minkowski spacetime	92
4.3 Conformal completions and asymptotic flatness	100
4.4 Black holes	105
5 Stationary black holes	119
5.1 Introduction	119
5.2 Stationary spacetimes	120
5.3 Mass and angular momentum	126
5.4 The event horizon as a Killing horizon	150
5.5 The generalized Smarr formula	160
5.6 The no-hair theorem	169

II Schwarzschild black hole	179
6 Schwarzschild black hole	181
6.1 Introduction	181
6.2 The Schwarzschild-(anti-)de Sitter solution	181
6.3 Radial null geodesics and Eddington-Finkelstein coordinates	187
6.4 Black hole character	199
7 Geodesics in Schwarzschild spacetime: generic and timelike cases	205
7.1 Introduction	205
7.2 Geodesic motion	205
7.3 Timelike geodesics	212
8 Null geodesics and images in Schwarzschild spacetime	231
8.1 Introduction	231
8.2 Main properties of null geodesics	232
8.3 Trajectories of null geodesics in the equatorial plane	239
8.4 Asymptotic direction from some emission point	258
8.5 Images	265
9 Maximal extension of Schwarzschild spacetime	279
9.1 Introduction	279
9.2 Kruskal-Szekeres coordinates	280
9.3 Maximal extension	288
9.4 Carter-Penrose diagram	295
9.5 Einstein-Rosen bridge	305
9.6 Physical relevance of the maximal extension	316
III Kerr black hole	319
10 Kerr black hole	321
10.1 Introduction	321
10.2 The Kerr solution	322
10.3 Extension of the spacetime manifold through $\Delta = 0$	332
10.4 Principal null geodesics	338
10.5 Event horizon	345
10.6 Global quantities	349
10.7 Families of observers in Kerr spacetime	352
10.8 Maximal analytic extension	362
10.9 Further reading	369

11 Geodesics in Kerr spacetime: generic and timelike cases	371
11.1 Introduction	371
11.2 Equations of geodesic motion	372
11.3 Main properties of geodesics	386
11.4 Timelike geodesics	408
11.5 Circular timelike orbits in the equatorial plane	416
11.6 Going further	439
12 Null geodesics and images in Kerr spacetime	441
12.1 Introduction	441
12.2 Main properties of null geodesics	441
12.3 Spherical photon orbits	458
12.4 Black hole shadow and critical curve	477
12.5 Images	493
13 Extremal Kerr black hole	503
13.1 Introduction	503
13.2 Definition and basic properties	504
13.3 Maximal analytic extension	511
13.4 Near-horizon extremal Kerr (NHEK) geometry	521
13.5 Going further	535
IV Dynamical black holes	537
14 Black hole formation 1: dust collapse	539
14.1 Introduction	539
14.2 Lemaître-Tolman equations and their solutions	540
14.3 Oppenheimer-Snyder collapse	548
14.4 Observing the black hole formation	566
14.5 Going further	580
15 Black hole formation 2: Vaidya collapse	581
15.1 Introduction	581
15.2 The ingoing Vaidya metric	581
15.3 Imploding shell of radiation	586
15.4 Configurations with a naked singularity	600
15.5 Going further	612
16 Evolution and thermodynamics of black holes	615
16.1 Introduction	615
16.2 Towards the first law of black hole dynamics	615
16.3 Evolution of the black hole area	633
16.4 The laws of black hole dynamics	640
16.5 Hawking radiation	644

16.6	Black hole thermodynamics	648
16.7	Concluding remarks	651
17	Black hole thermodynamics beyond general relativity	653
17.1	Introduction	653
17.2	Diffeomorphism-invariant theories of gravity	653
17.3	Wald entropy	664
17.4	Covariant phase space formalism	668
17.5	First law for diffeomorphism-invariant theories	682
17.6	What about the second law?	691
18	The quasi-local approach: trapping horizons	693
18.1	Introduction	693
18.2	Trapped surfaces and singularity theorems	693
18.3	Trapping horizons	705
V	Appendices	707
A	Basic differential geometry	709
A.1	Introduction	709
A.2	Differentiable manifolds	709
A.3	Pseudo-Riemannian manifolds	719
A.4	The three basic derivatives	724
A.5	Curvature	732
B	Geodesics	737
B.1	Introduction	737
B.2	Definition and first properties	737
B.3	Existence and uniqueness of geodesics	742
B.4	Geodesics and extremal lengths	748
B.5	Geodesics and symmetries	755
B.6	Geodesics and curvature	757
C	Kerr-Schild metrics	761
C.1	Generic Kerr-Schild spacetimes	761
C.2	Case of Kerr spacetime	763
D	SageMath computations	771
D.1	Introduction	771
D.2	Minkowski spacetime	772
D.3	Anti-de Sitter spacetime	772
D.4	Schwarzschild spacetime	772
D.5	Kerr spacetime	776
D.6	Evolution and thermodynamics	779

E Gyoto computations	781
E.1 Introduction	781
E.2 Image computations	781
F On the Web	783
Bibliography	785

Part I

Foundations

Chapter 1

General framework

Contents

1.1	Introduction	13
1.2	Spacetime	13
1.3	Worldlines	17
1.4	Quantities measured by an observer	20
1.5	Einstein equation	22

1.1 Introduction

This chapter succinctly presents 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 an important part of these lectures (but not always!), we shall assume that the theory of gravitation is general relativity; this means that the spacetime metric obeys the Einstein equation, which is recalled in Sec. 1.5.

This chapter is by no means an introduction to relativistic gravity. In this respect, we recommend the textbooks [86, 108, 162, 256, 370, 463, 498], as well as [161, 228, 336] for the French-speaking reader. The reader might also find useful to start by taking a look at Appendix A, which recaps the concepts from differential geometry employed here.

1.2 Spacetime

1.2.1 General settings

In these lectures we consider a n -dimensional *spacetime*, i.e. a pair (\mathcal{M}, g) , where \mathcal{M} is a smooth real manifold of dimension $n \geq 2$ and g is a Lorentzian metric on \mathcal{M} . In some parts, n will be set to 4 — the standard spacetime dimension — but we shall consider spacetimes with $n \neq 4$ as well.

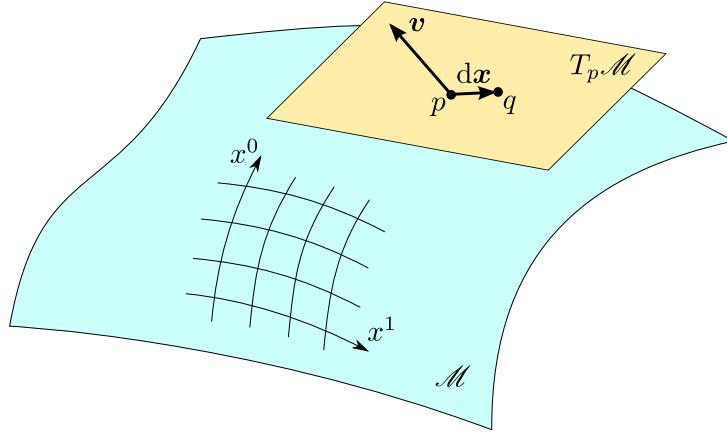


Figure 1.1: A smooth manifold \mathcal{M} : the infinitesimal vector dx connects the nearby points p and q and thus can thought as a displacement within the manifold, while the finite vector v does not correspond to any displacement in the manifold and “lives” in the tangent space $T_p \mathcal{M}$.

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** \mathcal{M} of dimension n is a “space” (precisely a *topological space*) that *locally* resembles \mathbb{R}^n , i.e. can be described by a n -tuple of coordinates (x^0, \dots, x^{n-1}) . However, globally, \mathcal{M} can be very different from \mathbb{R}^n , in particular regarding its topology. It could also happen that many coordinate systems are required to cover \mathcal{M} , while a single one is obviously sufficient for \mathbb{R}^n .

The smooth structure endows the manifold with the concept of **infinitesimal displacement vectors** dx , which connect infinitely close points of \mathcal{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 \mathcal{M} do not live in the manifold but in the **tangent spaces** $T_p \mathcal{M}$, which are defined at each point $p \in \mathcal{M}$. Each $T_p \mathcal{M}$ 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 \mathcal{M} is an orientable manifold (cf. Sec. A.3.4).

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

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

Given a coordinate system (x^α) , the components of g with respect to it are the n^2 scalar fields $(g_{\alpha\beta})$ such that [cf. Eq. (A.40)]

$$\boxed{g = g_{\alpha\beta} dx^\alpha dx^\beta}, \quad (1.2)$$

where dx^α is the differential 1-form of x^α (cf. Sec. A.2.4) and the notation $dx^\alpha dx^\beta$ stands for the symmetric tensor product defined by Eq. (A.39). The scalar square ds^2 of an infinitesimal displacement vector $dx = dx^\alpha \partial_\alpha$ is expressed in terms of the components $(g_{\alpha\beta})$ by the line element formula

$$ds^2 := g(dx, dx) = g_{\alpha\beta} dx^\alpha dx^\beta. \quad (1.3)$$

This formula follows from Eq. (1.2) via the identity $\langle dx^\alpha, dx \rangle = dx^\alpha$ [Eq. (A.18)].

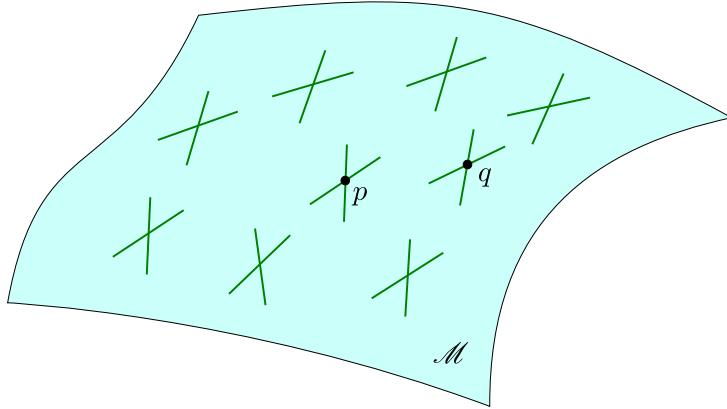


Figure 1.2: A Lorentzian manifold (\mathcal{M}, g) : at each point, the metric tensor g defines privileged directions: those lying in the null cone at p .

Remark 1: In this book, we shall use the bilinear form identity (1.2) rather than the line element (1.3) to present a metric (cf. Box. 3.2 of MTW [370] for a discussion). It turns out to be more convenient when various metric tensors are involved, the notation ds^2 being then ambiguous.

The fact that the signature of g is **Lorentzian**, i.e.

$$\text{sign } g = (-, \underbrace{+,\dots,+}_{n-1 \text{ times}}), \quad (1.4)$$

implies that at each point $p \in \mathcal{M}$, there are privileged directions, along which the line element (1.3) vanishes; they 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 $p \in \mathcal{M}$ that is either timelike or null (cf. Sec. A.3.2) is said to be **causal**. It lies necessarily inside the null cone at p (timelike vector) or along it (null vector).

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 (\mathcal{M}, g) is **time-orientable**, i.e. that it is possible to divide *continuously* all causal (i.e. timelike or null) vectors into two classes, the *future-directed* ones and the *past-directed* ones. More precisely, at each tangent space $T_p \mathcal{M}$, 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 nappe of the null cone at p . This defines an equivalence relation on causal vectors at p , with two equivalence classes. The spacetime (\mathcal{M}, g) is then called **time-orientable** iff some choice of an equivalence class can be performed continuously over all \mathcal{M} . The vectors belonging to the chosen equivalence class are called **future-directed** and the other ones **past-directed**.

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

Lemma 1.1: scalar product of a timelike vector with a causal vector

Let (\mathcal{M}, g) be a time-orientable spacetime and \mathbf{u} a future-directed timelike vector. For any null or timelike (nonzero) vector \mathbf{v} , we have necessarily $g(\mathbf{u}, \mathbf{v}) \neq 0$ and

$$g(\mathbf{u}, \mathbf{v}) < 0 \iff \mathbf{v} \text{ is future-directed} \quad (1.5a)$$

$$g(\mathbf{u}, \mathbf{v}) > 0 \iff \mathbf{v} \text{ is past-directed.} \quad (1.5b)$$

Proof. Without any loss of generality, we may assume that \mathbf{u} is a unit vector: $g(\mathbf{u}, \mathbf{u}) = -1$. Let then $(\mathbf{e}_i)_{1 \leq i \leq n-1}$ be a family of $n-1$ unit spacelike vectors such that $(\mathbf{u}, \mathbf{e}_1, \dots, \mathbf{e}_{n-1})$ is an orthonormal basis of $T_p \mathcal{M}$. We may expand \mathbf{v} on this basis: $\mathbf{v} = v^0 \mathbf{u} + v^i \mathbf{e}_i$. We have necessarily $v^0 \neq 0$, otherwise $\mathbf{v} = v^i \mathbf{e}_i$ would be a spacelike vector, which is excluded by hypothesis. Moreover, the time-orientation of \mathbf{v} is the same as that of \mathbf{u} iff $v^0 > 0$. Since $g(\mathbf{u}, \mathbf{v}) = -v^0$, this establishes (1.5a) and (1.5b). \square

Lemma 1.2: scalar product of a null vector with a causal vector

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

$$g(\mathbf{u}, \mathbf{v}) < 0 \iff \mathbf{v} \text{ is not collinear with } \mathbf{u} \text{ and is future-directed} \quad (1.6a)$$

$$g(\mathbf{u}, \mathbf{v}) = 0 \iff \mathbf{v} \text{ is collinear with } \mathbf{u} \text{ (and thus null)} \quad (1.6b)$$

$$g(\mathbf{u}, \mathbf{v}) > 0 \iff \mathbf{v} \text{ is not collinear with } \mathbf{u} \text{ and is past-directed.} \quad (1.6c)$$

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

$$(v^0)^2 \geq \sum_{i=1}^{n-1} (v^i)^2. \quad (1.7)$$

This implies $|v^0| \geq |v^1|$. If $|v^0| > |v^1|$, then \mathbf{v} cannot be collinear with \mathbf{u} (since this would imply $v^0 = v^1$) and $g(\mathbf{u}, \mathbf{v}) = -v^0 + v^1 \neq 0$, with a sign identical to that of $-v^0$. If $|v^0| = |v^1|$, Eq. (1.7) implies $v^2 = v^3 = \dots = v^{n-1} = 0$. We have then either $\mathbf{v} = v^0(\mathbf{e}_0 + \mathbf{e}_1) = v^0 \mathbf{u}$ or $\mathbf{v} = v^0(\mathbf{e}_0 - \mathbf{e}_1)$. In the first case, \mathbf{v} is collinear with \mathbf{u} and $g(\mathbf{u}, \mathbf{v}) = 0$. In the second case, $g(\mathbf{u}, \mathbf{v}) = -2v^0 \neq 0$. To summarize, the only case where $g(\mathbf{u}, \mathbf{v}) = 0$ is \mathbf{v} being collinear with \mathbf{u} . This establishes (1.6b). In all the other cases, \mathbf{v} is not collinear with \mathbf{u} and the sign of $g(\mathbf{u}, \mathbf{v})$ is that of $-v^0$. Since \mathbf{v} is future-directed if $v^0 > 0$ and past-directed if $v^0 < 0$, this establishes (1.6a) and (1.6c). \square

Two useful properties are immediate consequences of the above lemmas. From Lemma 1.1, we get

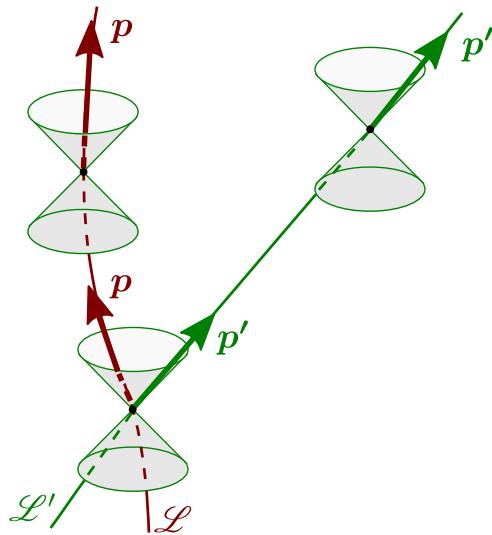


Figure 1.3: Worldlines of a massive particle (\mathcal{L}) and of a massless one (\mathcal{L}').

Property 1.3: no causal orthogonality to a timelike vector

A timelike vector cannot be orthogonal to a timelike vector or to a null vector.

From the part (1.6b) of Lemma 1.2, we get

Property 1.4: orthogonality of two null vectors

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, \mathcal{L} say, and not a point. This curve is called the particle's ***worldline*** and might be thought 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 \mathcal{L} 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 p defined along \mathcal{L} , tangent to \mathcal{L} at each point and future-directed (cf. Fig. 1.3).

One distinguishes two types of particles:

¹When $n \neq 4$, *energy-momentum vector* is definitely a better name than *4-momentum*!

- the **massive particles**, for which \mathcal{L} is a timelike curve, or equivalently, for which \mathbf{p} is a timelike vector:

$$\mathbf{g}(\mathbf{p}, \mathbf{p}) = \mathbf{p} \cdot \mathbf{p} < 0; \quad (1.8)$$

- the **massless particles**, such as the photon, for which \mathcal{L} is a null curve, or equivalently, for which \mathbf{p} is a null vector:

$$\mathbf{g}(\mathbf{p}, \mathbf{p}) = \mathbf{p} \cdot \mathbf{p} = 0. \quad (1.9)$$

In both cases, the **mass** of the particle is defined by²

$$m = \sqrt{-\mathbf{p} \cdot \mathbf{p}}. \quad (1.10)$$

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

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

$$\boxed{\nabla_{\mathbf{p}} \mathbf{p} = 0}, \quad (1.11)$$

or, in index notation,

$$p^\mu \nabla_\mu p^\alpha = 0. \quad (1.12)$$

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

Remark 1: The reverse is not true, i.e. having \mathcal{L} geodesic and \mathbf{p} tangent to \mathcal{L} does not imply (1.11), but the weaker condition $\nabla_{\mathbf{p}} \mathbf{p} = \alpha \mathbf{p}$, with α a scalar field along \mathcal{L} . In this case, one says that \mathbf{p} is a **pregeodesic vector** (cf. Sec. B.2.2 in Appendix B).

For massive particles, Eq. (1.11) can be derived from a variational principle, the action being simply the worldline's length τ as given by the metric tensor:

$$S = \int_A^B d\tau = \int_{\lambda_A}^{\lambda_B} \sqrt{-\mathbf{g} \left(\frac{d\mathbf{x}}{d\lambda}, \frac{d\mathbf{x}}{d\lambda} \right)} d\lambda \quad (1.13)$$

(cf. Sec. B.4.2 for details). For photons, Eq. (1.11) can be derived from Maxwell equations within the geometrical optics approximation (see e.g. Box 5.6 of Ref. [417]), with the assumption that the photon energy-momentum vector is related to the wave 4-vector \mathbf{k} by

$$\mathbf{p} = \hbar \mathbf{k}. \quad (1.14)$$

²Unless specified, we use geometrized units, for which $G = 1$ and $c = 1$.

1.3.3 Massive particles

For a massive particle, the constraint of having the worldline \mathcal{L} timelike has a simple geometrical meaning: \mathcal{L} must always lie inside the light cones of events along \mathcal{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:

$$d\tau = \sqrt{-g(dx, dx)} = \sqrt{-g_{\mu\nu}dx^\mu dx^\nu}, \quad (1.15)$$

where dx is an infinitesimal future-directed³ displacement along \mathcal{L} .

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

$$\boxed{\mathbf{u} := \frac{dx}{d\tau}}. \quad (1.16)$$

By construction, \mathbf{u} is tangent to \mathcal{L} and is a unit timelike vector:

$$\mathbf{u} \cdot \mathbf{u} = -1. \quad (1.17)$$

For a simple particle (no internal structure), the 4-momentum \mathbf{p} is tangent to \mathcal{L} ; it is then necessarily collinear to \mathbf{u} . Since both vectors are future-directed, Eqs. (1.10) and (1.17) lead to

$$\boxed{\mathbf{p} = m \mathbf{u}}. \quad (1.18)$$

1.3.4 Massless particles (photons)

For a massless particle, Eq. (1.15) would lead to $d\tau = 0$ since the displacement dx 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 \mathcal{L} is a parameter λ such that the associated tangent vector,

$$\boxed{\mathbf{v} := \frac{dx}{d\lambda}}, \quad (1.19)$$

is a geodesic vector field: $\nabla_v v = 0$. In general, the tangent vector associated to a given parameter fulfills only $\nabla_v v = \alpha v$, with α a scalar field along \mathcal{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:

$$\lambda' = a\lambda + b, \quad (1.20)$$

with a and b two constants. Given that the photon energy-momentum vector \mathbf{p} is a geodesic vector [Eq. (1.11)], a natural choice of the affine parameter λ is that associated with \mathbf{p} :

$$\boxed{\mathbf{p} = \frac{dx}{d\lambda}}. \quad (1.21)$$

This sets $a = 1$ in the transformation (1.20).

³Cf. Sec. 1.2.2.

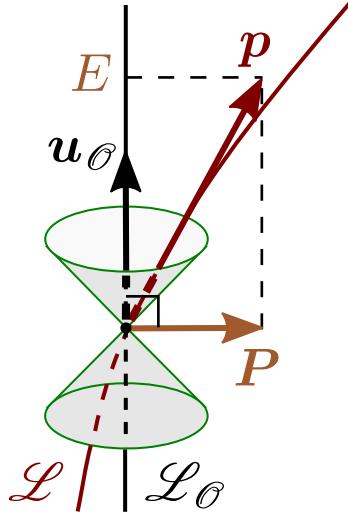


Figure 1.4: Orthogonal decomposition of the energy-momentum vector \mathbf{p} of a particle with respect to the 4-velocity $\mathbf{u}_\mathcal{O}$ of an observer \mathcal{O} , giving birth to the energy E and linear momentum \mathbf{P} as measured by \mathcal{O} .

1.4 Quantities measured by an observer

In the simplest modelization, an *observer* \mathcal{O} in the spacetime (\mathcal{M}, g) is described by a timelike worldline $\mathcal{L}_\mathcal{O}$ that is equipped with an orthonormal basis (e_α) at each point, such that e_0 is future-directed and tangent to $\mathcal{L}_\mathcal{O}$ and (e_α) varies smoothly along $\mathcal{L}_\mathcal{O}$ (see e.g. Sec. 13.6 of Ref. [370] or Chap. 3 of Ref. [227] for an extended discussion). The vector e_0 is then the 4-velocity of \mathcal{O} and the vectors (e_1, e_2, e_3) form an orthonormal basis of the 3-dimensional local rest space of \mathcal{O} . (e_α) is called the *observer's frame*.

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

$$\boxed{\mathbf{p} = E \mathbf{u}_\mathcal{O} + \mathbf{P}}, \quad \text{with} \quad \mathbf{u}_\mathcal{O} \cdot \mathbf{P} = 0, \quad (1.22)$$

where $\mathbf{u}_\mathcal{O} = e_0$ is the 4-velocity of observer \mathcal{O} . By taking the scalar product of Eq. (1.22) with $\mathbf{u}_\mathcal{O}$, we obtain the following expressions for E and \mathbf{P} :

$$\boxed{E = -\mathbf{u}_\mathcal{O} \cdot \mathbf{p}} \quad (1.23)$$

$$\boxed{\mathbf{P} = \mathbf{p} + (\mathbf{u}_\mathcal{O} \cdot \mathbf{p}) \mathbf{u}_\mathcal{O}}. \quad (1.24)$$

The scalar square of Eq. (1.22) leads to

$$\underbrace{\mathbf{p} \cdot \mathbf{p}}_{-m^2} = E^2 \underbrace{\mathbf{u}_\mathcal{O} \cdot \mathbf{u}_\mathcal{O}}_{-1} + 2E \underbrace{\mathbf{u}_\mathcal{O} \cdot \mathbf{P}}_0 + \mathbf{P} \cdot \mathbf{P}, \quad (1.25)$$

where we have used Eq. (1.10) to let appear the particle's mass m . Hence we recover Einstein's relation:

$$\boxed{E^2 = m^2 + \mathbf{P} \cdot \mathbf{P}}. \quad (1.26)$$

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

$$d\mathbf{x} = \mathbf{p} d\lambda, \quad (1.27)$$

where λ is the affine parameter along the particle's worldline whose tangent vector is \mathbf{p} [cf. Eq. (1.21) for a massless particle and Eqs. (1.16) and (1.18) with $\lambda := \tau/m$ for a massive particle]. Substituting (1.22) for \mathbf{p} in (1.27), we get the orthogonal decomposition of $d\mathbf{x}$ with respect to $\mathcal{L}_\mathcal{O}$:

$$d\mathbf{x} = Ed\lambda \mathbf{u}_\mathcal{O} + d\lambda \mathbf{P}. \quad (1.28)$$

\mathcal{O} 's proper time elapsed during the particle's displacement is the coefficient in front of $\mathbf{u}_\mathcal{O}$: $d\tau_\mathcal{O} = Ed\lambda$ and the particle's displacement in \mathcal{O} 's rest frame is the part orthogonal to $\mathbf{u}_\mathcal{O}$: $d\mathbf{X} = d\lambda \mathbf{P}$. By definition, the particle's velocity with respect to \mathcal{O} is

$$\mathbf{V} := \frac{d\mathbf{X}}{d\tau_\mathcal{O}} = \frac{d\lambda \mathbf{P}}{Ed\lambda}. \quad (1.29)$$

Hence the relation

$$\boxed{\mathbf{P} = E \mathbf{V}}. \quad (1.30)$$

By combining with (1.22), we get the following orthogonal decomposition of the particle's 4-momentum:

$$\boxed{\mathbf{p} = E (\mathbf{u}_\mathcal{O} + \mathbf{V})}. \quad (1.31)$$

Relations (1.26), (1.31) and (1.30) are valid for any kind of particle, massive or not. For a massive particle, the energy-momentum vector \mathbf{p} is related to the particle's 4-velocity \mathbf{u} via (1.18). Inserting this relation into (1.23), we obtain

$$\boxed{E = \Gamma m}, \quad (1.32)$$

where

$$\Gamma := -\mathbf{u}_\mathcal{O} \cdot \mathbf{u} \quad (1.33)$$

is the **Lorentz factor** of the particle with respect to the observer. If we depart from units with $c = 1$, Eq. (1.32) becomes the famous relation $E = \Gamma mc^2$. Furthermore, combining (1.30) and (1.32) yields the familiar relation between the linear momentum and the velocity:

$$\boxed{\mathbf{P} = \Gamma m \mathbf{V}}. \quad (1.34)$$

Finally, inserting (1.32) and (1.34) into (1.26) leads to the well-known expression of the Lorentz factor in terms of the velocity:

$$\Gamma = (1 - \mathbf{V} \cdot \mathbf{V})^{-1/2}. \quad (1.35)$$

If we divide Eq. (1.22) by m and use Eqs. (1.18), (1.32) and (1.34) to express respectively \mathbf{p}/m , E/m and \mathbf{P}/m , we get the following orthogonal split of the particle's 4-velocity \mathbf{u} with respect to observer \mathcal{O} :

$$\boxed{\mathbf{u} = \Gamma (\mathbf{u}_\mathcal{O} + \mathbf{V})}. \quad (1.36)$$

For a massless particle (photon), inserting (1.30) into the Einstein relation (1.26) with $m = 0$ yields

$$\boxed{\mathbf{V} \cdot \mathbf{V} = 1}. \quad (1.37)$$

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

$$\boxed{\mathbf{k} = \omega (\mathbf{u}_{\mathcal{O}} + \mathbf{V})}, \quad (1.38)$$

where $\omega = 2\pi\nu$, with ν being the radiation frequency as measured by observer \mathcal{O} . In view of Eq. (1.31) with $\mathbf{p} = \hbar\mathbf{k}$ [Eq. (1.14)], we get the Planck-Einstein relation:

$$\boxed{E = h\nu}. \quad (1.39)$$

1.5 Einstein equation

1.5.1 General form

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

$$\boxed{\mathbf{R} - \frac{1}{2} R \mathbf{g} + \Lambda \mathbf{g} = 8\pi \mathbf{T}}, \quad (1.40)$$

where \mathbf{R} is the Ricci tensor of \mathbf{g} , R is the Ricci scalar of \mathbf{g} (cf. Sec. A.5.3 in Appendix A), Λ is some constant, called the **cosmological constant**, and \mathbf{T} is the energy-momentum tensor of matter and non-gravitational fields. If we let appear the Einstein tensor $\mathbf{G} := \mathbf{R} - (R/2)\mathbf{g}$ [Eq. (A.111)], the Einstein equation is recast as

$$\boxed{\mathbf{G} + \Lambda \mathbf{g} = 8\pi \mathbf{T}}. \quad (1.41)$$

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 \mathbf{G} is identically zero for any metric \mathbf{g} if $n = 2$ (the trace of Eq. (A.112) in Appendix A yields $\mathbf{R} = (R/2)\mathbf{g}$). The exclusion of $n = 2$ also follows by noticing that the Einstein-Hilbert action, which gives birth to Eq. (1.40) for $n \geq 3$, is proportional to the Euler characteristic of \mathcal{M} for $n = 2$ (by virtue of the Gauss-Bonnet theorem), the Euler characteristic being a topological invariant independent of \mathbf{g} .

By taking the trace of (1.40) with respect to \mathbf{g} , it is easy to show that the Einstein equation (1.40) is equivalent to

$$\boxed{\mathbf{R} = \frac{2}{n-2} \Lambda \mathbf{g} + 8\pi \left(\mathbf{T} - \frac{1}{n-2} T \mathbf{g} \right)}, \quad (1.42)$$

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

Remark 2: The spacetime dimension n explicitly appears in the variant (1.42) of the Einstein equation, but not in the original version (1.40). Notice as well that Eq. (1.42) would be ill-posed for $n = 2$ (cf. Remark 1).

The **vacuum Einstein equation with cosmological constant** is Eq. (1.42) with $\mathbf{T} = 0$:

$$\mathbf{R} = \frac{2}{n-2} \Lambda \mathbf{g}. \quad (1.43)$$

In the mathematical literature, a solution to Eq. (1.43) is called an **Einstein metric**. The special case $\Lambda = 0$ is called the **vacuum Einstein equation**:

$$\boxed{\mathbf{R} = 0}. \quad (1.44)$$

It thus corresponds to the vanishing of the Ricci tensor. Solutions of Eq. (1.44) are sometimes called **Ricci-flat metrics**.

Taking the covariant divergence of the Einstein equation (1.41) and invoking the contracted Bianchi identity (A.110) leads to

$$\boxed{\nabla \cdot \vec{\mathbf{T}} = 0}, \quad (1.45)$$

where $\vec{\mathbf{T}}$ is the type-(1, 1) tensor associated by metric duality to \mathbf{T} [cf. Eq. (A.48)]. In index notation, the above equation reads [cf. Eq. (A.67)]

$$\nabla_\mu T^\mu_\alpha = 0.$$

Equation (1.45) is often referred to as the **equation of energy-momentum conservation**.

1.5.2 Electrovacuum Einstein equation

An **electromagnetic field** in the spacetime $(\mathcal{M}, \mathbf{g})$ is a 2-form \mathbf{F} (cf. Sec. A.2.6) such that any particle of mass $m > 0$, electric charge q and 4-velocity \mathbf{u} is subject to the 4-acceleration

$$\nabla_{\mathbf{u}} \mathbf{u} = \frac{q}{m} \vec{\mathbf{F}}(., \mathbf{u}) \iff u^\mu \nabla_\mu u^\alpha = \frac{q}{m} F^\alpha_\mu u^\mu. \quad (1.46)$$

Moreover, in standard electromagnetism, \mathbf{F} is governed by the **Maxwell equations**:

$$\mathbf{d}\mathbf{F} = 0 \quad \text{and} \quad \nabla \cdot \vec{\mathbf{F}} = \mu_0 \mathbf{j}, \quad (1.47)$$

where $\mathbf{d}\mathbf{F}$ is the exterior derivative of \mathbf{F} (cf. Sec. A.4.3), $\vec{\mathbf{F}}$ in the type-(2, 0) antisymmetric tensor (bivector) associated by metric duality to \mathbf{F} [cf. Eqs. (A.49) and (A.50b)], μ_0 is a constant called the **vacuum permeability** and \mathbf{j} is the **electric current density** vector field, which describes the distribution of electric charges in spacetime. In view of formula (A.90c) for the exterior derivative and of definition (A.67) for the divergence operator, the Maxwell equations (1.47) can be expressed in terms of components with respect to some coordinates (x^α) as

$$\frac{\partial F_{\beta\gamma}}{\partial x^\alpha} + \frac{\partial F_{\gamma\alpha}}{\partial x^\beta} + \frac{\partial F_{\alpha\beta}}{\partial x^\gamma} = 0 \quad \text{and} \quad \nabla_\mu F^{\alpha\mu} = \mu_0 j^\alpha. \quad (1.48)$$

Remark 3: Thanks to an identity valid for the divergence of any antisymmetric type-(2, 0) tensor field, the second Maxwell equation can be written in terms of partial derivatives only:

$$\frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^\mu} (\sqrt{-g} F^{\alpha\mu}) = \mu_0 j^\alpha, \quad (1.49)$$

where g stands for the determinant of the components $(g_{\alpha\beta})$ of the metric tensor \mathbf{g} with respect to the coordinates (x^α) .

Remark 4: The second Maxwell equation can be expressed in terms of differential forms, as the first one, by introducing the $(n - 2)$ -form $\star \mathbf{F}$ (n being the spacetime dimension) and the $(n - 1)$ -form $\star \underline{\mathbf{j}}$, which are respectively the Hodge dual of the 2-form \mathbf{F} and the Hodge dual of the 1-form $\underline{\mathbf{j}}$ (the metric dual of the vector field \mathbf{j} , cf. Sec. A.3.3). We shall define the Hodge dual of a p -form in Chap. 5 [Eq. (5.38)]; here, let us simply state the Maxwell equations in terms of it and of the exterior derivative \mathbf{d} :

$$\mathbf{d}\mathbf{F} = 0 \quad \text{and} \quad \mathbf{d}\star\mathbf{F} = \mu_0 \star \underline{\mathbf{j}}. \quad (1.50)$$

A **source-free electromagnetic field** is a 2-form \mathbf{F} that obeys the Maxwell equations (1.47) with $\underline{\mathbf{j}} = 0$:

$$\mathbf{d}\mathbf{F} = 0 \quad \text{and} \quad \nabla \cdot \vec{\mathbf{F}} = 0. \quad (1.51)$$

The energy-momentum tensor \mathbf{T} of a source-free electromagnetic field \mathbf{F} is

$$T_{\alpha\beta} = \frac{1}{\mu_0} \left(F_{\mu\alpha} F^\mu_\beta - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} g_{\alpha\beta} \right), \quad (1.52)$$

the trace of which with respect to \mathbf{g} is

$$T := g^{\mu\nu} T_{\mu\nu} = \frac{4-n}{4\mu_0} F_{\mu\nu} F^{\mu\nu}. \quad (1.53)$$

Remark 5: $T = 0$ only for $n = 4$ (the standard spacetime dimension).

The **electrovacuum Einstein equation** is the Einstein equation (1.42) with $\Lambda = 0$ and \mathbf{T} given by Eq. (1.52); in view of Eq. (1.53), it writes

$$R_{\alpha\beta} = \frac{8\pi}{\mu_0} \left(F_{\mu\alpha} F^\mu_\beta - \frac{1}{2(n-2)} F_{\mu\nu} F^{\mu\nu} g_{\alpha\beta} \right). \quad (1.54)$$

A solution to the **Einstein-Maxwell system** is a triplet $(\mathcal{M}, \mathbf{g}, \mathbf{F})$ such that $(\mathcal{M}, \mathbf{g})$ is a n -dimensional spacetime, \mathbf{F} is a source-free electromagnetic field on \mathcal{M} and (\mathbf{g}, \mathbf{F}) obeys the the electrovacuum Einstein equation (1.54).

Chapter 2

The concept of black hole 1: Horizons as null hypersurfaces

Contents

2.1	Introduction	25
2.2	Black holes and null hypersurfaces	25
2.3	Geometry of null hypersurfaces	28
2.4	Evolution of the expansion	53

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}, 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 emitted

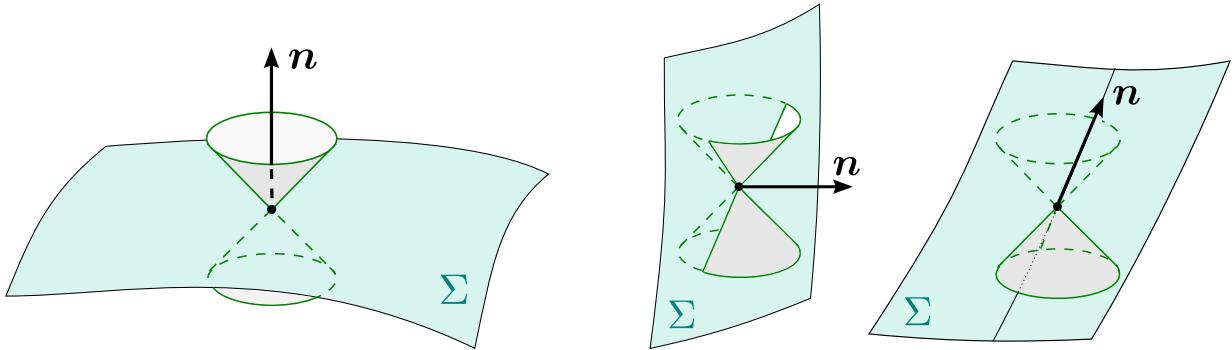


Figure 2.1: The three causal types of hypersurfaces: spacelike (left), timelike (middle) and null (right).

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 \mathcal{M} , for it has to divide \mathcal{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 \mathcal{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 (\mathcal{M}, g) , a (smooth) hypersurface Σ can locally be classified in three categories. The classification, called the ***causal type***, depends on the signature of metric induced by g on Σ – the ***induced metric*** being nothing but the restriction $g|_{\Sigma}$ of g to vector fields tangent to Σ . The hypersurface Σ is said to be

- **spacelike** iff $g|_{\Sigma}$ is positive definite, i.e. iff sign $g|_{\Sigma} = (+, \dots, +)$ ($n - 1$ plus signs), i.e. iff $(\Sigma, g|_{\Sigma})$ is a Riemannian manifold;
- **timelike** iff $g|_{\Sigma}$ is a Lorentzian metric, i.e. iff sign $g|_{\Sigma} = (-, +, \dots, +)$ (1 minus sign and $n - 2$ plus signs), i.e. iff $(\Sigma, g|_{\Sigma})$ is a Lorentzian manifold;
- **null** iff $g|_{\Sigma}$ is degenerate¹ i.e. iff sign $g|_{\Sigma} = (0, +, \dots, +)$ (1 zero and $n - 2$ plus signs).

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

The causal type of the hypersurface Σ can also be deduced from any normal vector² n to it (cf. Fig. 2.1):

- Σ spacelike $\iff n$ timelike;

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

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

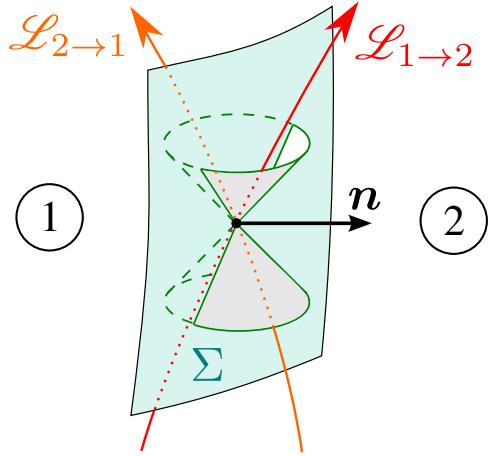


Figure 2.2: A timelike hypersurface is a two-way membrane: $\mathcal{L}_{1 \rightarrow 2}$ is a timelike worldline from Region 1 to Region 2, while $\mathcal{L}_{2 \rightarrow 1}$ is a timelike worldline from Region 2 to Region 1.

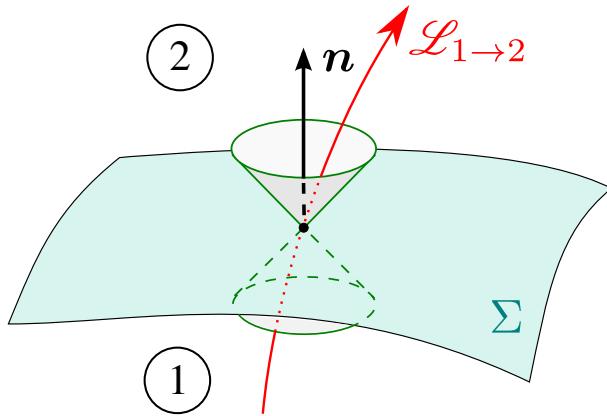


Figure 2.3: A spacelike hypersurface is a one-way membrane: $\mathcal{L}_{1 \rightarrow 2}$ is a timelike worldline from Region 1 to Region 2, while there is no timelike or null worldline from Region 2 to Region 1.

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

These equivalences are easily proved by considering a 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 \mathbf{n} is null iff $\mathbf{n} \cdot \mathbf{n} = 0$, which is nothing but the condition for \mathbf{n} to be tangent to Σ .

A timelike hypersurface is a two-way membrane: it divides (locally) the 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 1 to Region 2, 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

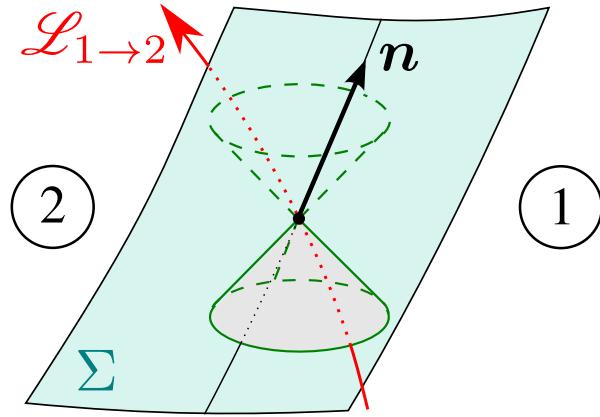


Figure 2.4: A null hypersurface is a one-way membrane: $\mathcal{L}_{1 \rightarrow 2}$ is a timelike worldline from Region 1 to Region 2, while there is no timelike or null worldline from Region 2 to Region 1.

not going from Region 1 to Region 2 must stay on the hypersurface; an example of such null worldline is the one depicted in Fig. 2.4 as the thin black line tangent to the normal 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 generic definition of a black hole implies that its boundary (the event horizon) is a null hypersurface as soon as it is smooth (Property 4.11). 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

Since smooth black hole event horizons are null hypersurfaces, let us examine the geometrical properties of such objects. We shall denote the hypersurface under study by \mathcal{H} , for *horizon*, but the results of this section and of Sec. 2.4 are valid for any null hypersurface.

2.3.1 Hypersurfaces as level sets

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

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

and

$$du \neq 0 \quad \text{on } \mathcal{H}, \quad (2.2)$$

where du is the *differential* of the scalar field u (cf. Sec. A.2.4): $du = (\partial u / \partial x^\alpha) dx^\alpha$ in any coordinate chart (x^α) . Condition (2.2) ensures that \mathcal{H} is a regular hypersurface (an *embedded* submanifold, cf. Sec. A.2.7); without it, \mathcal{H} could be self-intersecting.

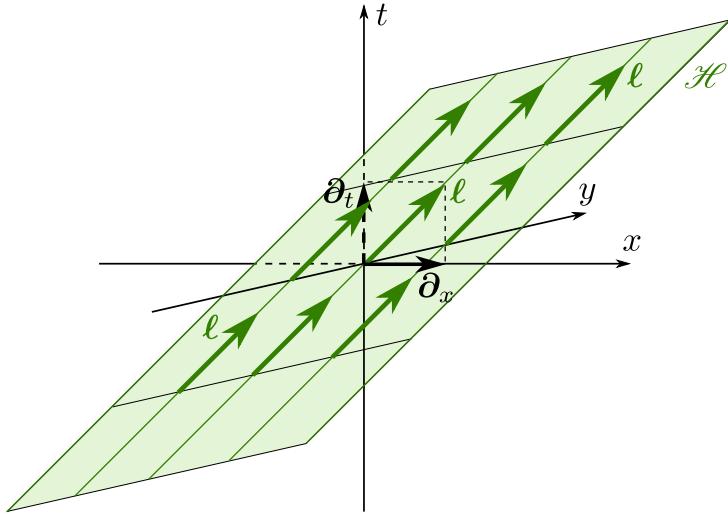


Figure 2.5: Null hyperplane \mathcal{H} of equation $t - x = 0$ in Minkowski spacetime. The dimension along z has been suppressed, so that \mathcal{H} is pictured as a 2-plane.

Example 1 (null hyperplane): A very simple example of null hypersurface is a null hyperplane of the 4-dimensional Minkowski spacetime. The **Minkowski spacetime** is defined by $\mathcal{M} = \mathbb{R}^4$ with \mathbf{g} being a *flat* Lorentzian metric. Natural coordinates are **Minkowskian coordinates** (t, x, y, z) , i.e. coordinates with respect to which the metric components are $g_{\alpha\beta} = \text{diag}(-1, 1, 1, 1)$. The scalar field

$$u(t, x, y, z) = t - x \quad (2.3)$$

defines then a null hyperplane \mathcal{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 \mathcal{H} of a light cone, also called **future light cone**. Note that we have to take out the cone apex from \mathcal{H} , in order to have a regular hypersurface. In the Minkowskian coordinates (t, x, y, z) , the choice of the “retarded time”

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

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

Example 3 (Schwarzschild horizon): Let us consider the 4-dimensional spacetime $(\mathcal{M}, \mathbf{g})$ with \mathcal{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 the metric tensor takes the form³

$$\mathbf{g} = - \left(1 - \frac{2m}{r}\right) dt^2 + \frac{4m}{r} dt dr + \left(1 + \frac{2m}{r}\right) dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2, \quad (2.5)$$

where m is a positive constant. We shall see in Chap. 6 that $(\mathcal{M}, \mathbf{g})$ is actually a part of Schwarzschild spacetime, described in coordinates different from the standard Schwarzschild-Droste ones, $(\bar{t}, \bar{r}, \theta, \varphi)$

³Formula (2.5) follows the notation introduced in Eq. (1.2); see Sec. A.3.1 in Appendix A for an extended discussion, in particular Eq. (A.39). When reading the metric components $(g_{\alpha\beta})$ from (2.5), keep in mind that cross terms involve a factor 2: the $4m/r$ coefficient of $dt dr$ implies that $g_{tr} = 2m/r$.

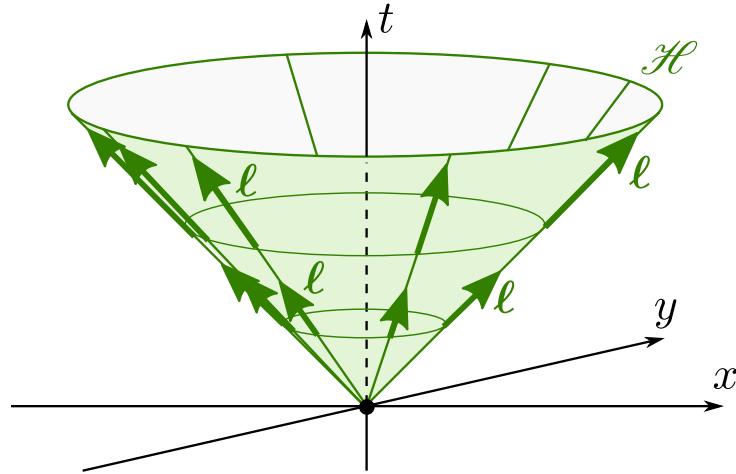


Figure 2.6: Future sheet \mathcal{H} of the light cone of equation $t - \sqrt{x^2 + y^2 + z^2} = 0$ in Minkowski spacetime. The dimension along z has been suppressed, so that \mathcal{H} looks 2-dimensional, whereas it is actually 3-dimensional.

say, the link between the two being $t = \bar{t} + 2m \ln |r/(2m) - 1|$. The present coordinates are called the **ingoing Eddington-Finkelstein coordinates** and have the 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 $r \rightarrow 2m$, as do those of the inverse metric, which are

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

Let us consider the scalar field defined on \mathcal{H} by

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

It is then clear that the hypersurface $u = 0$ is the 3-dimensional “cylinder” \mathcal{H} of equation $r = 2m$ (cf. Fig. 2.7). We shall see below⁴ that \mathcal{H} is indeed a null hypersurface.

2.3.2 Null normals

Let ℓ be a vector field normal to \mathcal{H} . Since \mathcal{H} is a null hypersurface, ℓ is a null vector:

$$\ell \cdot \ell = 0. \quad (2.8)$$

Moreover, we choose ℓ to be future-directed (cf. Sec. 1.2.2).

Remark 1: As a consequence of (2.8), there is no natural normalization of ℓ , contrary to the case of timelike or spacelike hypersurfaces, where one can always choose the normal to be a unit vector (scalar

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

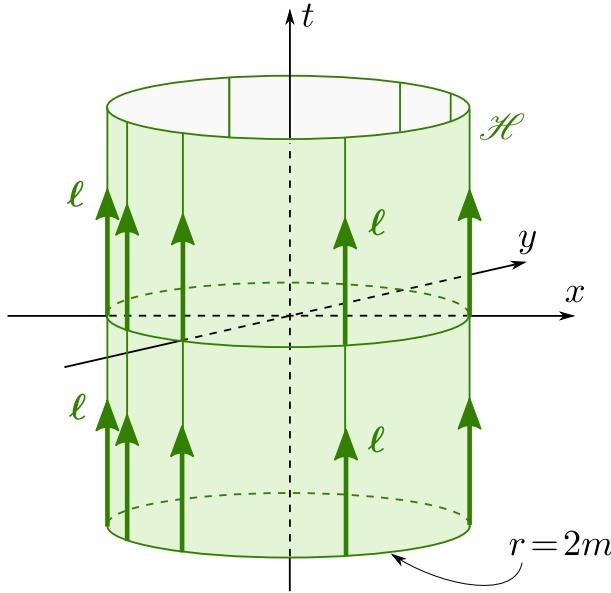


Figure 2.7: Schwarzschild horizon \mathcal{H} introduced in Example 3; The figure is drawn for $\theta = \pi/2$ and is based on coordinates (t, x, y) related to the ingoing Eddington-Finkelstein coordinates (t, r, θ, φ) by $x = r \cos \varphi$ and $y = r \sin \varphi$.

square equal to 1 or -1). It follows that there is no unique choice of ℓ . At this stage, any rescaling $\ell \mapsto \ell' = \alpha \ell$, with α some strictly positive (to preserve the future orientation of ℓ) scalar field on \mathcal{H} , yields a normal vector field ℓ' as valid as ℓ .

The null normal vector field ℓ is a priori defined on \mathcal{H} only and not at points $p \notin \mathcal{H}$. However, it is worth to consider ℓ as a vector field not confined to \mathcal{H} but defined in some open subset of \mathcal{M} around \mathcal{H} . In particular this would permit to define the spacetime covariant derivative $\nabla \ell$, which is not possible if the support of ℓ is restricted to \mathcal{H} . Following Carter [101], a simple way to achieve this is to consider not only a single null hypersurface \mathcal{H} , but a foliation of \mathcal{M} (in the vicinity of \mathcal{H}) by a family of null hypersurfaces, such that \mathcal{H} is an element of this family. Without any loss of generality, we may select the value of the scalar field u defining \mathcal{H} to label these hypersurfaces and denote the family by (\mathcal{H}_u) . The null hypersurface \mathcal{H} is then nothing but the element $\mathcal{H} = \mathcal{H}_{u=0}$ of this family [Eq. (2.1)]. The vector field ℓ can then be viewed as being defined in the part of \mathcal{M} foliated by (\mathcal{H}_u) , such that at each point in this region, ℓ is null and normal to \mathcal{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 (\mathcal{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 (\mathcal{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 (\mathcal{H}_u) is nonunique but all geometrical quantities that we shall introduce hereafter do not depend upon the choice of the foliation \mathcal{H}_u once they are evaluated at \mathcal{H} .

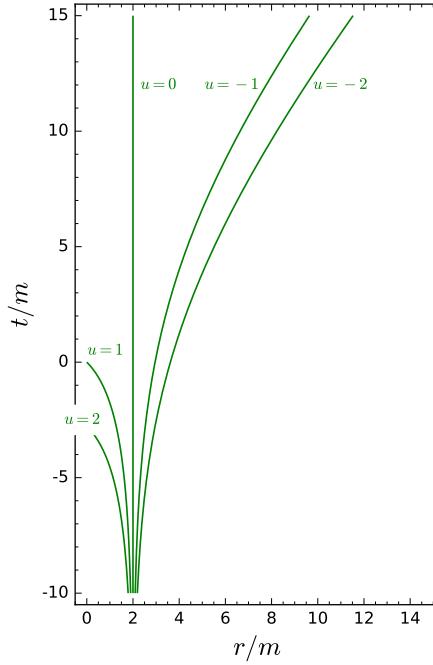


Figure 2.8: Hypersurfaces \mathcal{H}_u defined by $u = \text{const}$ for the Schwarzschild horizon (Example 3).

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

$$\forall \mathbf{v} \in T_p \mathcal{M}, \quad \mathbf{v} \text{ tangent to } \mathcal{H} \iff \nabla_{\mathbf{v}} u = 0 \iff \langle \mathbf{d}u, \mathbf{v} \rangle = 0 \iff \vec{\nabla} u \cdot \mathbf{v} = 0, \quad (2.9)$$

where $\vec{\nabla} u$ is the **gradient** of the scalar field u , i.e. the vector field that is the metric-dual of the 1-form $\mathbf{d}u$ (cf. Sec. A.3.3); in a coordinate chart (x^α) , its components are given by

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

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

$$\boxed{\ell = -e^\rho \vec{\nabla} u}. \quad (2.11)$$

We have chosen the coefficient linking ℓ and $\vec{\nabla} u$ to be strictly negative (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, ℓ is future-directed, as the following example shows:

Example 5 (null hyperplane): We deduce from expression (2.3) for u in Example 1 that $\mathbf{d}u = \mathbf{d}t - \mathbf{d}x$. The gradient vector field obtained by metric duality is $\vec{\nabla} u = -\partial_t - \partial_x$. Choosing for simplicity $\rho = 0$, formula (2.11) yields

$$\ell = \partial_t + \partial_x. \quad (2.12)$$

This vector field is depicted in Fig. 2.5.

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

$$du = dt - \frac{x}{r}dx - \frac{y}{r}dy - \frac{z}{r}dz, \quad \text{with } r := \sqrt{x^2 + y^2 + z^2}.$$

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

$$\ell = \partial_t + \frac{x}{r}\partial_x + \frac{y}{r}\partial_y + \frac{z}{r}\partial_z. \quad (2.13)$$

This vector field is depicted in Fig. 2.6.

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

$$du = \frac{1}{4m}e^{(r-t)/(4m)} \left[-\left(1 - \frac{r}{2m}\right) dt - \left(1 + \frac{r}{2m}\right) dr \right].$$

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

$$\vec{\nabla}u = \frac{1}{4m}e^{(r-t)/(4m)} \left[-\left(1 + \frac{r}{2m}\right)\partial_t + \left(1 - \frac{r}{2m}\right)\partial_r \right].$$

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

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

Equation (2.11) leads then to

$$\ell = \partial_t + \frac{r-2m}{r+2m}\partial_r. \quad (2.15)$$

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

2.3.3 Null geodesic generators

Frobenius identity

Let us take the metric dual of Eq. (2.11); it writes $\underline{\ell} = -e^\rho du = -e^\rho \nabla u$, or, in index notation, $\ell_\alpha = -e^\rho \nabla_\alpha u$. Taking the covariant derivative, we get

$$\nabla_\alpha \ell_\beta = -e^\rho \nabla_\alpha \rho \nabla_\beta u - e^\rho \nabla_\alpha \nabla_\beta u = \nabla_\alpha \rho \ell_\beta - e^\rho \nabla_\alpha \nabla_\beta u$$

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

$$\nabla_\alpha \ell_\beta - \nabla_\beta \ell_\alpha = \nabla_\alpha \rho \ell_\beta - \nabla_\beta \rho \ell_\alpha. \quad (2.16)$$

In the left-hand side there appears the exterior derivative $d\underline{\ell}$ of the 1-form $\underline{\ell}$ [cf. Eq. (A.91b)], while one recognizes in the right-hand side the exterior product of the two 1-forms $d\rho$ and $\underline{\ell}$. Hence we may rewrite (2.16) as

$$\boxed{d\underline{\ell} = d\rho \wedge \underline{\ell}}. \quad (2.17)$$

This reflects the **Frobenius theorem** in its dual formulation (see e.g. Theorem B.3.2 in Wald's textbook [498] or Theorem C.2 in Straumann's textbook [463]): the exterior derivative of the 1-form $\underline{\ell}$ is the exterior product of some 1-form (here $d\rho$) with $\underline{\ell}$ itself if, and only if, $\underline{\ell}$ defines hyperplanes that are integrable in some hypersurface (\mathcal{H} in the present case).

Geodesic generators

Let us contract the Frobenius identity (2.16) with ℓ :

$$\ell^\mu \nabla_\mu \ell_\alpha - \ell^\mu \nabla_\alpha \ell_\mu = \ell^\mu \nabla_\mu \rho \ell_\alpha - \underbrace{\ell^\mu \ell_\mu}_{0} \nabla_\alpha \rho. \quad (2.18)$$

Now, since ℓ is a null vector, $\ell^\mu \nabla_\alpha \ell_\mu = \nabla_\alpha (\ell^\mu \ell_\mu) - \ell_\mu \nabla_\alpha \ell^\mu = -\ell_\mu \nabla_\alpha \ell^\mu$, from which we get

$$\ell^\mu \nabla_\alpha \ell_\mu = 0. \quad (2.19)$$

Hence (2.18) reduces to

$$\ell^\mu \nabla_\mu \ell_\alpha = \kappa \ell_\alpha, \quad (2.20)$$

with

$$\kappa := \ell^\mu \nabla_\mu \rho = \nabla_\ell \rho. \quad (2.21)$$

The metric dual of (2.20) is

$$\boxed{\nabla_\ell \ell = \kappa \ell}. \quad (2.22)$$

This equation implies that the field lines of ℓ are geodesics (cf. Property B.5 in Appendix B). To demonstrate this, we note that a rescaling

$$\ell \mapsto \ell' = \alpha \ell \quad (2.23)$$

with α a positive scalar field can be performed to yield a **geodesic vector field** ℓ' , i.e. a vector field that obeys⁵ Eq. (B.1):

$$\nabla_{\ell'} \ell' = 0. \quad (2.24)$$

Proof. Equations (2.23) and (2.22) imply

$$\nabla_{\ell'} \ell' = \alpha (\nabla_\ell \alpha + \kappa \alpha) \ell. \quad (2.25)$$

Hence, since $\alpha > 0$,

$$\nabla_{\ell'} \ell' = 0 \iff \nabla_\ell \ln \alpha = -\kappa \iff \frac{d}{d\lambda} \ln \alpha = -\kappa,$$

where λ is the parameter associated to ℓ along its field lines \mathcal{L} : $\ell = dx/d\lambda|_{\mathcal{L}}$. Therefore, setting $\alpha(\lambda) = \exp(-\int_0^\lambda \kappa(\bar{\lambda}) d\bar{\lambda})$ along each field line \mathcal{L} ensures that $\ell' = \alpha \ell$ fulfills Eq. (2.24). \square

Because of (2.24), the field lines of ℓ' are null geodesics and ℓ' is the tangent vector to them associated with some *affine* parameter λ' . On the other side, if $\kappa \neq 0$, ℓ 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 ℓ (cf. Sec. B.2.2 in Appendix B).

Since ℓ is collinear to ℓ' , it obviously shares the same field lines, which have just been shown to be null geodesics. Hence we may state:

⁵A vector field that obeys the weaker condition (2.22), with κ possibly nonzero, is called a **pregeodesic vector field**, cf. Sec. B.2.2.

Property 2.1: geodesic generators of a null hypersurface

Any null hypersurface \mathcal{H} is ruled by a family of null geodesics, called the **null geodesic generators** of \mathcal{H} , and each vector field ℓ normal to \mathcal{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 ℓ are null curves that are tangent to \mathcal{H} , one 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 geodesic, given that all geodesics of Minkowski spacetime are straight lines.

As a by-product of Eq. (2.25), we get the behavior of the non-affinity coefficient under a rescaling of the null normal:

$$\ell' = \alpha \ell \implies \kappa' = \alpha \kappa + \nabla_\ell \alpha. \quad (2.26)$$

Example 8 (null hyperplane): It is clear on expression (2.12) for ℓ that the covariant derivative $\nabla \ell$ vanishes identically. In particular $\nabla_\ell \ell = 0$. Equation (2.22) then implies $\kappa = 0$, which is in agreement with Eq. (2.21) and the choice $\rho = 0$ performed in Example 5. The null geodesic generators of \mathcal{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 ℓ and the fact that $\nabla_\beta \ell^\alpha = \partial_\beta \ell^\alpha$ in the Minkowskian coordinates (t, x, y, z) , we get

$$\nabla_\beta \ell^\alpha = \frac{1}{r^3} \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & y^2 + z^2 & -xy & -xz \\ 0 & -xy & x^2 + z^2 & -yz \\ 0 & -xz & -yz & x^2 + y^2 \end{pmatrix} \quad (\alpha = \text{row index}; \beta = \text{column index}). \quad (2.27)$$

We obtain then $\ell^\mu \nabla_\mu \ell^\alpha = 0$. From Eq. (2.22), we conclude that $\kappa = 0$, which is in agreement with Eq. (2.21) and the choice $\rho = 0$ performed in Example 6. The null geodesic generators of \mathcal{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) $\nabla_\ell t = 1$ 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 ℓ as given by (2.15) is (cf. Sec. D.4.1 for the computation)

$$\nabla_\beta \ell^\alpha = \begin{pmatrix} \frac{m}{r^2} & \frac{m}{r^2} \frac{3r+2m}{r+2m} & 0 & 0 \\ \frac{m}{r^2} \frac{r-2m}{r+2m} & \frac{m}{r^2} \frac{3r^2-4m(r+m)}{(r+2m)^2} & 0 & 0 \\ 0 & 0 & \frac{r-2m}{r(r+2m)} & 0 \\ 0 & 0 & 0 & \frac{r-2m}{r(r+2m)} \end{pmatrix} \quad (\alpha = \text{row index}; \beta = \text{column index}). \quad (2.28)$$

Contracting with ℓ^β , we obtain

$$\nabla_\ell \ell = \frac{4m}{(r+2m)^2} \partial_t + \frac{4m(r-2m)}{(r+2m)^3} \partial_r = \frac{4m}{(r+2m)^2} \ell.$$

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

$$\kappa = \frac{1}{4m}. \quad (2.29)$$

This value agrees with $\kappa = \nabla_\ell \rho$ [Eq. (2.21)] 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 ℓ (since $\nabla_\ell t = 1$ 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 kinds of null hypersurfaces. For instance the interior of a future null cone in Minkowski spacetime is a region from which no particle may escape, but since the null cone is expanding indefinitely, particles can travel arbitrarily 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.

Let \mathcal{H} be a null hypersurface of a n -dimensional spacetime (\mathcal{M}, g) with $n \geq 3$. We define a ***cross-section*** of \mathcal{H} as a submanifold \mathcal{S} of \mathcal{H} of dimension $n-2$ such that (i) the null normal ℓ is nowhere tangent to \mathcal{S} and (ii) each null geodesic generator of \mathcal{H} intersects \mathcal{S} at most once. If each null geodesic generator intersects \mathcal{S} exactly once, we shall say that \mathcal{S} is a ***complete cross-section*** (cf. Fig. 2.9).

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 \mathcal{H} delimitates some region of spacetime, we shall assume that its cross-sections are ***closed manifolds***, i.e. are compact without boundary⁶. The simplest example is the sphere, more precisely the $(n-2)$ -dimensional sphere \mathbb{S}^{n-2} , as illustrated in Fig. 2.9. This example is the one relevant for standard 4-dimensional stationary black holes, thanks a topology theorem that we shall discuss in Sec. 5.2.3. But for $n > 4$, other compact topologies, like that of a torus, are allowed for cross-sections of a stationary black hole.

A first important property of cross-sections, independently of whether they are closed or not, is:

⁶Strictly speaking, a *closed manifold* is simply a *compact manifold*, since a topological manifold, as defined in Sec. A.2.1, has no boundary. Indeed, the concept of *manifold with boundary* requires a separate definition (cf. Sec. A.2.2), so that a manifold with boundary is not a manifold. Hence, we may consider the terminology *closed manifold* as stressing the no-boundary aspect of a compact manifold.

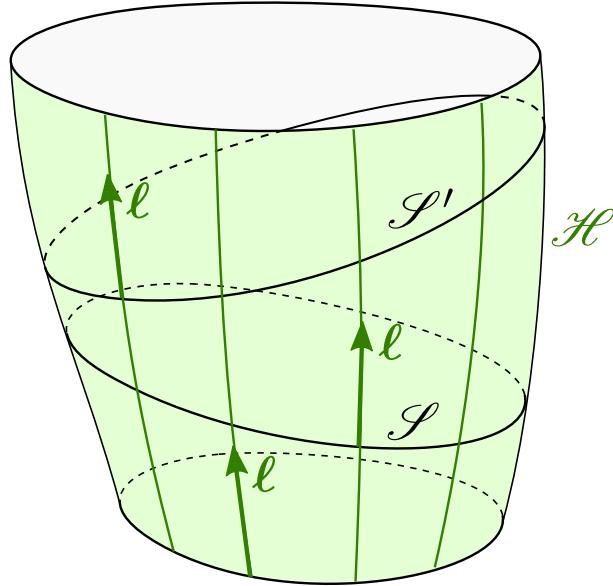


Figure 2.9: The null hypersurface \mathcal{H} and two (complete) cross-sections \mathcal{S} and \mathcal{S}' . The green curves represent some null geodesic generators, with the null normal ℓ tangent to them.

Property 2.2: spacelike character of cross-sections

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

The spacelike character of \mathcal{S} follows from

Lemma 2.3: tangent vectors to a null hypersurface

Every nonzero tangent vector v to a null hypersurface is either spacelike or null. Moreover, if it is null, v is tangent to a null geodesic generator, i.e. v is normal to the hypersurface.

Proof. Tangent vectors to a null hypersurface \mathcal{H} are by definition vectors v such that $g(\ell, v) = 0$, where ℓ is the normal to \mathcal{H} . Since ℓ is null, it follows then from Property 1.3 (Sec. 1.2.2) that v cannot be timelike. Besides, if v is null, Property 1.4 (Sec. 1.2.2) implies that it must be collinear to ℓ . \square

Proof of Property 2.2. Let $p \in \mathcal{S}$ and $v \in T_p \mathcal{M}$ be a nonzero tangent vector to \mathcal{S} . The above lemma implies that 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 v is necessarily spacelike, which proves that \mathcal{S} is a spacelike submanifold. \square

Example 11 (light cone): The future sheet \mathcal{H} of the Minkowski-spacetime light cone considered in Examples 2, 6 and 9 has the topology $\mathbb{R} \times \mathbb{S}^2$ since we have excluded the cone apex from \mathcal{H} . The

complete cross-sections are then compact and it is natural to choose them as the spheres 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 $\mathbb{R} \times \mathbb{S}^2$ (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 ℓ . 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

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

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

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

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

$$\forall (\mathbf{u}, \mathbf{v}) \in T_p \mathcal{S} \times T_p \mathcal{S}, \quad q(\mathbf{u}, \mathbf{v}) := g(\mathbf{u}, \mathbf{v}). \quad (2.30)$$

Saying that \mathcal{S} is spacelike (Property 2.2) is equivalent to saying that q is positive definite, i.e. $q(\mathbf{v}, \mathbf{v}) \geq 0$ for all $\mathbf{v} \in T_p \mathcal{S}$, with $q(\mathbf{v}, \mathbf{v}) = 0 \iff \mathbf{v} = 0$. In other words, (\mathcal{S}, q) is a **Riemannian manifold** (cf Sec. A.3.2 in Appendix A).

Example 14 (Schwarzschild horizon): The metric induced by g on the cross-section \mathcal{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), given that $(x^a) = (\theta, \varphi)$ is a coordinate system on \mathcal{S} :

$$q = 4m^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2). \quad (2.31)$$

An important consequence of \mathcal{S} being spacelike is (cf. Fig. 2.10):

Property 2.4: orthogonal complement of a cross-section tangent space

At each point p of a cross-section \mathcal{S} of a null hypersurface \mathcal{H} , the tangent space $T_p \mathcal{S}$ has an **orthogonal complement** $T_p^\perp \mathcal{S}$, i.e. a timelike 2-dimensional vector plane such that (i) the tangent space to \mathcal{M} at p , $T_p \mathcal{M}$, is the direct sum of $T_p \mathcal{S}$ and $T_p^\perp \mathcal{S}$:

$$T_p \mathcal{M} = T_p \mathcal{S} \oplus T_p^\perp \mathcal{S} \quad (2.32)$$

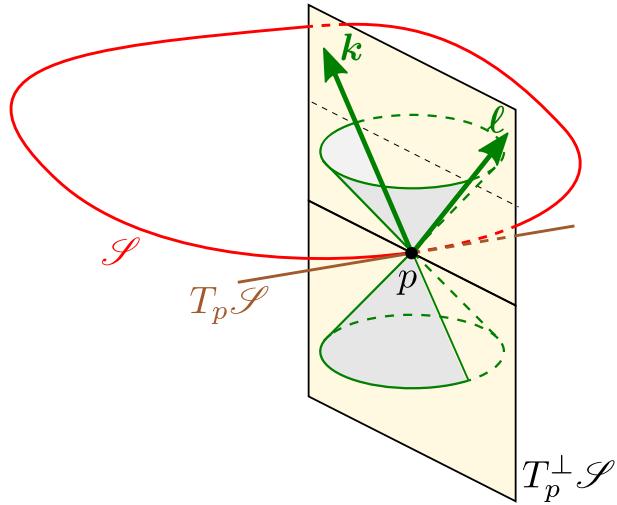


Figure 2.10: The tangent space $T_p \mathcal{S}$ to the cross-section \mathcal{S} and its 2-dimensional orthogonal complement $T_p^\perp \mathcal{S}$. Only the dimensionality of the latter is respected in the figure: \mathcal{S} and $T_p \mathcal{S}$ are depicted as 1-dimensional objects, while they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

and (ii) every vector in $T_p^\perp \mathcal{S}$ is orthogonal to \mathcal{S} .

Proof. Using the Gram-Schmidt process to construct a g -orthogonal basis of $T_p \mathcal{M}$, starting from a q -orthogonal basis of $T_p \mathcal{S}$, one naturally ends up with a complement of $T_p \mathcal{S}$ that is 2-dimensional and such that $\text{sign } g|_{T_p^\perp \mathcal{S}} = (-, +)$ (timelike signature), in order to ensure that the signature of g is $(-, +, \dots, +)$. \square

Since $\mathcal{S} \subset \mathcal{H}$, the null normal ℓ to \mathcal{H} is orthogonal to any vector tangent to \mathcal{S} , which implies $\ell \in T_p^\perp \mathcal{S}$. One can supplement ℓ by another null vector to form a basis of $T_p^\perp \mathcal{S}$:

Property 2.5: complement null normal to a cross-section

At each point p of a cross-section \mathcal{S} of a null hypersurface \mathcal{H} with future-directed null normal ℓ , there exists a unique future-directed null vector k transverse to \mathcal{H} such that

$$k \cdot \ell = -1 \quad \text{and} \quad T_p^\perp \mathcal{S} = \text{Span}(\ell, k). \quad (2.33)$$

Proof. As a timelike plane, $T_p^\perp \mathcal{S}$ has two independent null directions, which can be seen as the two intersections of the null cone at p with the 2-plane $T_p^\perp \mathcal{S}$ (cf. Fig. 2.10). Let us denote by k a future-directed null vector in the null direction of $T_p^\perp \mathcal{S}$ that is not along ℓ . By a proper rescaling $k \mapsto \alpha k$, we may choose k so that $k \cdot \ell = -1$ (cf. Lemma 1.2). Since ℓ and k are non-collinear vectors of $T_p^\perp \mathcal{S}$ and $\dim T_p^\perp \mathcal{S} = 2$, they constitute a basis of $T_p^\perp \mathcal{S}$. \square

Remark 3: While ℓ is independent of the choice of the cross-section \mathcal{S} of \mathcal{H} through p , k does depend on \mathcal{S} .

Property 2.6: extension of q

The bilinear form q , which has been defined only on $T_p\mathcal{S}$ via (2.30), can be extended to all vectors of $T_p\mathcal{M}$ by setting

$$q := g + \underline{\ell} \otimes \underline{k} + \underline{k} \otimes \underline{\ell}, \quad (2.34)$$

or, in index notation,

$$q_{\alpha\beta} := g_{\alpha\beta} + \ell_\alpha k_\beta + k_\alpha \ell_\beta. \quad (2.35)$$

It obeys

$$\forall (\mathbf{u}, \mathbf{v}) \in T_p\mathcal{M} \times T_p\mathcal{M}, \quad q(\mathbf{u}, \mathbf{v}) = q(\mathbf{u}^\parallel, \mathbf{v}^\parallel), \quad (2.36)$$

where \mathbf{u}^\parallel and \mathbf{v}^\parallel are the tangent parts of \mathbf{u} and \mathbf{v} to \mathcal{S} , which are uniquely defined by the direct sum (2.32):

$$\mathbf{u} = \mathbf{u}^\parallel + \mathbf{u}^\perp \quad \text{and} \quad \mathbf{v} = \mathbf{v}^\parallel + \mathbf{v}^\perp, \quad \text{with} \quad \mathbf{u}^\parallel, \mathbf{v}^\parallel \in T_p\mathcal{S}, \quad \mathbf{u}^\perp, \mathbf{v}^\perp \in T_p^\perp\mathcal{S}. \quad (2.37)$$

Proof. Thanks to (2.33), we may write the orthogonal decompositions (2.37) as

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

Using $\ell \cdot \ell = 0$, $\mathbf{k} \cdot \mathbf{k} = 0$ and $\ell \cdot \mathbf{k} = -1$, we have then $\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\parallel \cdot \mathbf{v}^\parallel - u^0 v^1 - u^1 v^0$, so that q defined by (2.34) obeys

$$\begin{aligned} q(\mathbf{u}, \mathbf{v}) &= \mathbf{u} \cdot \mathbf{v} + (\ell \cdot \mathbf{u})(\mathbf{k} \cdot \mathbf{v}) + (\mathbf{k} \cdot \mathbf{u})(\ell \cdot \mathbf{v}) \\ &= \mathbf{u}^\parallel \cdot \mathbf{v}^\parallel - u^0 v^1 - u^1 v^0 + u^1 v^0 + u^0 v^1 = \mathbf{u}^\parallel \cdot \mathbf{v}^\parallel. \end{aligned}$$

□

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

$$\mathbf{k} = \frac{1}{2} \partial_t - \frac{x}{2r} \partial_x - \frac{y}{2r} \partial_y - \frac{z}{2r} \partial_z.$$

Evaluating q via (2.34), given expression (2.13) for ℓ , we get the following components of 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: $x = r \sin \theta \cos \varphi$, $y = r \sin \theta \sin \varphi$ and $z = r \cos \theta$, the components of q become

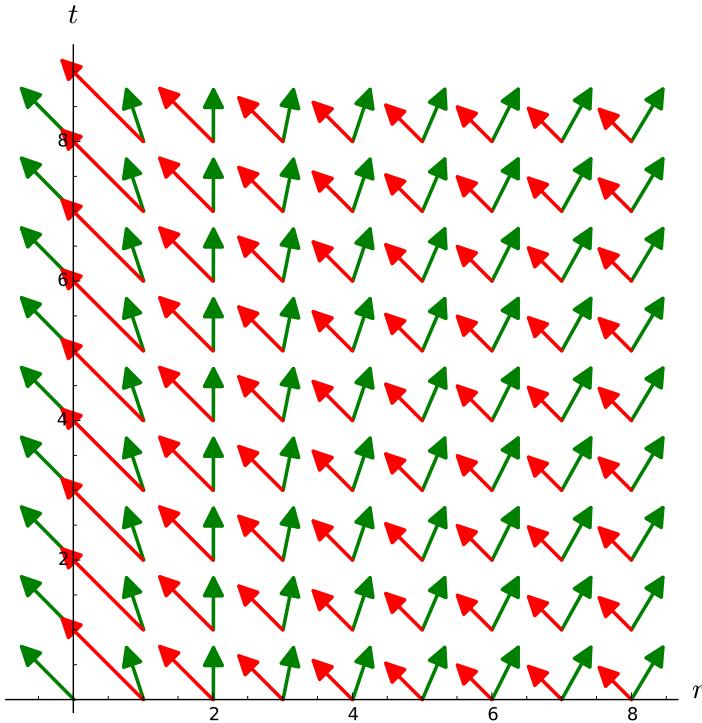


Figure 2.11: Null vector fields ℓ (green) and k (red) corresponding to Example 16 (Schwarzschild horizon). The plot is a 2-dimensional slice $\theta = \text{const}$ and $\varphi = \text{const}$ of the spacetime \mathcal{M} , with t and r labelled in units of m . Note that since k diverges at $r = 0$ [cf. Eq. (2.39)], it is not represented there.

instead

$$q_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}. \quad (2.38)$$

and we recognize in $q_{ab} = \text{diag}(r^2, r^2 \sin^2 \theta)$ 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 ℓ that the null vector k obeying $k \cdot \ell = -1$ and orthogonal to the sphere \mathcal{S} introduced in Example 12 is

$$\mathbf{k} = \left(\frac{1}{2} + \frac{m}{r} \right) \partial_t - \left(\frac{1}{2} + \frac{m}{r} \right) \partial_r. \quad (2.39)$$

The vector field \mathbf{k} is depicted in Fig. 2.11. We have (cf. the notebook D.4.1)

$$\underline{\ell} = \frac{2m-r}{2m+r} dt + dr \quad \text{and} \quad \underline{\mathbf{k}} = - \left(\frac{1}{2} + \frac{m}{r} \right) dt - \left(\frac{1}{2} + \frac{m}{r} \right) dr, \quad (2.40)$$

so that Eq. (2.34) leads to the following components of q in terms of the ingoing Eddington-Finkelstein

coordinates $x^\alpha = (t, r, \theta, \varphi)$:

$$q_{\alpha\beta} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \theta \end{pmatrix}. \quad (2.41)$$

The pair (ℓ, k) forms a null basis of $T_p^\perp \mathcal{S}$ [cf. Eq. (2.33)]. One can construct from it an *orthonormal* basis (n, s) as follows:

$$\begin{cases} n = \frac{1}{2}\ell + k \\ s = \frac{1}{2}\ell - k. \end{cases} \iff \begin{cases} \ell = n + s \\ k = \frac{1}{2}(n - s). \end{cases} \quad (2.42)$$

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

$$n \cdot n = -1, \quad s \cdot s = 1 \quad \text{and} \quad n \cdot s = 0. \quad (2.43)$$

In other words, (n, s) is an orthonormal basis of the Lorentzian plane $(T_p^\perp \mathcal{S}, g)$.

If we substitute (2.42) for ℓ and k in Eq. (2.34), we get

$$q = g + \underline{n} \otimes \underline{n} - \underline{s} \otimes \underline{s}. \quad (2.44)$$

Property 2.7: orthogonal projector onto a cross-section

The metric dual^a of the bilinear form q defined by Eq. (2.34), i.e. the tensor of type $(1, 1)$ defined by

$$\overrightarrow{q} := \text{Id} + \underline{\ell} \otimes \underline{k} + \underline{k} \otimes \underline{\ell}, \quad (2.45)$$

or, in index notation,

$$q^\alpha_\beta := \delta^\alpha_\beta + \ell^\alpha k_\beta + k^\alpha \ell_\beta, \quad (2.46)$$

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

$$\forall v \in T_p \mathcal{M}, \quad \overrightarrow{q}(v) = v^\parallel. \quad (2.47)$$

^aSee Eq. (A.48) of Appendix A for the arrow notation.

Proof. This follows readily from the decomposition $v = v^\parallel + v^0 \ell + v^1 k$ used above. \square

In particular, we have

$$\overrightarrow{q}(\ell) = 0 \quad \text{and} \quad \overrightarrow{q}(k) = 0. \quad (2.48)$$

2.3.5 Expansion along the null normal

One defines the expansion of the cross-section \mathcal{S} along the vector field ℓ as follows.

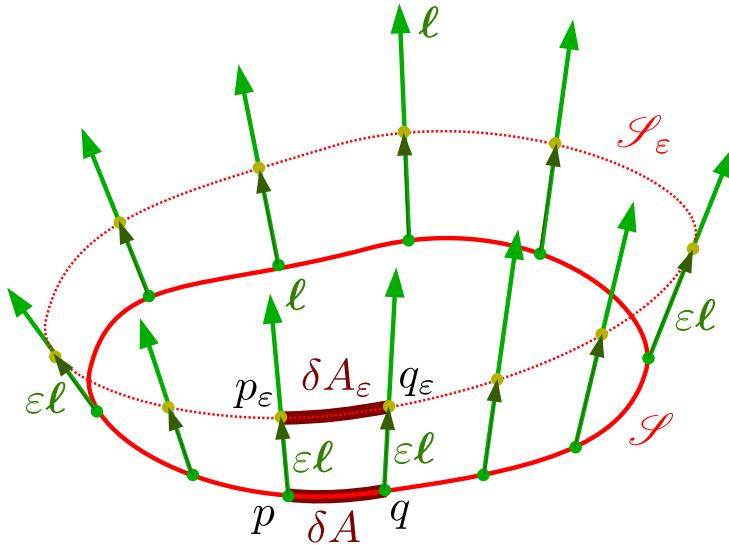


Figure 2.12: Lie dragging of the surface \mathcal{S} along ℓ by the small parameter ε . \mathcal{S} is drawn as a 1-dimensional submanifold, while it is actually a $(n - 2)$ -dimensional one, n being the spacetime dimension.

Given an infinitesimal parameter $\varepsilon \geq 0$, take a point $p \in \mathcal{S}$ and displace it by the (infinitesimal) vector $\varepsilon\ell$, thereby getting a nearby point p_ε (cf. Fig. 2.12). Since ℓ is tangent to \mathcal{H} and $p \in \mathcal{H}$, we have $p_\varepsilon \in \mathcal{H}$. By repeating this for each point in \mathcal{S} , keeping the value of ε fixed, we define a new $(n - 2)$ -dimensional surface, \mathcal{S}_ε say (cf. Fig. 2.12). One says that \mathcal{S}_ε is obtained from \mathcal{S} by **Lie dragging along ℓ by the parameter ε** . Note that $\mathcal{S}_{\varepsilon=0} = \mathcal{S}$. Since $p_\varepsilon \in \mathcal{H}$ for every $p \in \mathcal{S}$, we have $\mathcal{S}_\varepsilon \subset \mathcal{H}$. Because the null direction ℓ is transverse to \mathcal{S}_ε by construction, it follows that \mathcal{S}_ε is spacelike (cf. Lemma 2.3 in Sec. 2.3.4). At each point $p \in \mathcal{S}$, the **expansion of \mathcal{S} along ℓ** is defined from the rate of change $\theta_{(\ell)}$ of the area^a δA of a surface element δS of \mathcal{S} around p :

$$\boxed{\theta_{(\ell)} := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \frac{\delta A_\varepsilon - \delta A}{\delta A}}. \quad (2.49)$$

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 ℓ by the parameter ε (cf. Fig. 2.12).

^aWe 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 $n \geq 5$.

Remark 4: The reader may wonder why the expansion is not denoted by something like $\theta_{(\ell)}(\mathcal{S})$, since its definition depends explicitly on \mathcal{S} . We shall show below that, because \mathcal{H} is a null hypersurface, $\theta_{(\ell)}$ is actually independent of the choice of the cross-section \mathcal{S} .

In formula (2.49), the area δA is measured with respect to the metric q of \mathcal{S} and similarly the area δA_ε is measured by the Riemannian metric induced by g on \mathcal{S}_ε . For instance, if the surface element $\delta S \subset \mathcal{S}$ is a $(n - 2)$ -dimensional parallelogram delimited by some infinitesimal

displacement vectors $d\mathbf{x}_{(2)}, \dots, d\mathbf{x}_{(n-1)}$, the area of δS is

$$\delta A = \mathcal{S}\epsilon(d\mathbf{x}_{(2)}, \dots, d\mathbf{x}_{(n-1)}), \quad (2.50)$$

where $\mathcal{S}\epsilon$ is the Levi-Civita tensor associated with the metric q of \mathcal{S} (cf. Sec. A.3.4 in Appendix A). The latter is connected to the Levi-Civita tensor of g by the following property.

Property 2.8: area $(n - 2)$ -form of a cross-section

The Levi-Civita tensor $\mathcal{S}\epsilon$ associated with the metric q of \mathcal{S} is expressible in terms of the spacetime Levi-Civita tensor ϵ as

$$\boxed{\mathcal{S}\epsilon = \epsilon(\mathbf{n}, \mathbf{s}, \dots) = \epsilon(\mathbf{k}, \ell, \dots)}, \quad (2.51)$$

or, in index notation,

$$\mathcal{S}\epsilon_{\alpha_1 \dots \alpha_{n-2}} = n^\mu s^\nu \epsilon_{\mu\nu\alpha_1 \dots \alpha_{n-2}} = k^\mu \ell^\nu \epsilon_{\mu\nu\alpha_1 \dots \alpha_{n-2}},$$

where (\mathbf{n}, \mathbf{s}) is the orthonormal basis of $T_p^\perp \mathcal{S}$ associated to the null basis (ℓ, \mathbf{k}) by Eq. (2.42).

Proof. We note that $\epsilon(\mathbf{n}, \mathbf{s}, \dots)$ defines a fully antisymmetric $(n - 2)$ -linear form on $T_p \mathcal{S}$. Since the space of such forms is 1-dimensional (for $\dim T_p \mathcal{S} = n - 2$), there exists necessarily some proportionality factor a such that $\epsilon(\mathbf{n}, \mathbf{s}, \dots) = a \mathcal{S}\epsilon$. Now $\epsilon(\mathbf{n}, \mathbf{s}, d\mathbf{x}_{(2)}, \dots, d\mathbf{x}_{(n-1)})$ is the volume of the n -parallelepiped constructed on the vectors $\mathbf{n}, \mathbf{s}, d\mathbf{x}_{(2)}, \dots, d\mathbf{x}_{(n-1)}$. Given that \mathbf{n} and \mathbf{s} are unit-length vectors for the metric g , we have actually

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

This implies that $a = 1$, thereby establishing the first equality in Eq. (2.51). The second equality follows by substituting (2.42) for \mathbf{n} and \mathbf{s} in (2.51) and using the multilinearity and antisymmetry of ϵ . \square

In the vicinity of \mathcal{S} , let us consider a spacetime coordinate system $(x^\alpha) = (u, v, x^2, \dots, x^{n-1})$ that is adapted to \mathcal{S} and ℓ in the sense that \mathcal{H} is (locally) the set $u = 0$, \mathcal{S} is the set $(u, v) = (0, 0)$ and

$$\ell = \partial_v. \quad (2.52)$$

Then, from the very definition of the Lie dragging of \mathcal{S} along ℓ , \mathcal{S}_ε is the set $(u, v) = (0, \varepsilon)$ and $(x^a) = (x^2, \dots, x^{n-1})$ can be viewed as a coordinate system⁷ on both \mathcal{S} and \mathcal{S}_ε . Let us choose the $n - 2$ infinitesimal displacement vectors in (2.50) along the coordinate lines of this system: $d\mathbf{x}_{(a)} = dx^a \partial_a$ (no summation on a). Then expression (2.50) for the area of δS becomes

$$\begin{aligned} \delta A &= dx^2 \cdots dx^{n-1} \mathcal{S}\epsilon(\partial_2, \dots, \partial_{n-1}) = dx^2 \cdots dx^{n-1} \mathcal{S}\epsilon_{2\dots(n-1)} \\ \delta A &= \sqrt{q} dx^2 \cdots dx^{n-1}, \end{aligned} \quad (2.53)$$

⁷Let 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 $n - 1$.

where we have used the multilinearity of $\mathcal{S}\epsilon$ and Eq. (A.52) for the components of the Levi-Civita tensor $\mathcal{S}\epsilon$, q standing for the determinant of the metric q with respect to the coordinates (x^a) . By the very definition of the Lie dragging, the surface element δS_ε on \mathcal{S}_ε is defined by the same values of the coordinates (x^a) as δS . In particular, the coordinate increments dx^2, \dots, dx^{n-1} on \mathcal{S}_ε take the same values as on \mathcal{S} . Therefore, the area of δS_ε is

$$\delta A_\varepsilon = \sqrt{q(\varepsilon)} dx^2 \cdots dx^{n-1}, \quad (2.54)$$

where $q(\varepsilon)$ stands for the determinant of the components with respect to the coordinates (x^a) of the metric q_ε induced by g on \mathcal{S}_ε . Since \mathcal{S}_ε is spacelike (cf. above), q_ε is positive definite, so that $q(\varepsilon) > 0$. In view of (2.53)-(2.54), the definition (2.49) of the expansion of \mathcal{S} along ℓ can be rewritten as

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

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

$$\theta_{(\ell)} = \frac{1}{2} \frac{d}{d\varepsilon} \ln q. \quad (2.55)$$

Given that \mathcal{S}_ε is deduced from \mathcal{S} by Lie dragging along ℓ and ε is the value of the coordinate v associated with ℓ [cf. Eq. (2.52)], we may rewrite this formula as the Lie derivative of $\ln q$ along ℓ :

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

(2.56)

Example 17 (light cone): For the light cone in Minkowski spacetime, it is easy to evaluate $\theta_{(\ell)}$ 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 $q = r^2 d\theta^2 + r^2 \sin^2 \theta d\varphi^2$ [cf. Eq. (2.38)]. We have then $q = \det(q_{ab}) = r^4 \sin^2 \theta$. Moreover, the coordinate v can be chosen as $v = t - t_0$ since t is an (affine) parameter associated with ℓ (cf. Example 9). Given that $t = r$ on \mathcal{H} , we have $\varepsilon = v = r - t_0$, so that Eq. (2.55) yields

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

i.e.

$$\theta_{(\ell)} = \frac{2}{r}. \quad (2.57)$$

Example 18 (Schwarzschild horizon): As above, we have $q = r^4 \sin^2 \theta$ [cf. Eq. (2.5)], so that Eq. (2.56) yields

$$\theta_{(\ell)} = \frac{1}{2} \mathcal{L}_\ell \ln q = \frac{1}{2} \ell^\mu \frac{\partial}{\partial x^\mu} \ln q = \underbrace{\frac{\partial}{\partial t} \ln(r^2 \sin \theta)}_0 + \frac{r-2m}{r+2m} \frac{\partial}{\partial r} \ln(r^2 \sin \theta) = \frac{2}{r} \frac{r-2m}{r+2m}, \quad (2.58)$$

where we have used the components ℓ^μ read on (2.15). The above expression is valid for any hypersurface of the family \mathcal{H}_u . For the case of the Schwarzschild horizon, $r = 2m$ and Eq. (2.58) yields a vanishing expansion:

$$\theta_{(\ell)} = 0. \quad (2.59)$$

Note that for large r , Eq. (2.58) yields $\theta_{(\ell)} \sim 2/r$, i.e. we recover the flat spacetime result (2.57), which is consistent with the fact that for large r , \mathcal{H}_u is close to a Minkowskian light cone (cf. Fig. 2.8). Note also that Eq. (2.58) yields $\theta_{(\ell)} < 0$ for $r < 2m$ and $\theta_{(\ell)} > 0$ 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 \mathcal{H} : $A = 4\pi r^2$ from Eq. (2.41) and ℓ 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. (2.56) can be rewritten as

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

where Q is the matrix representing the components of q with respect to the coordinates $(x^a) = (x^2, \dots, x^{n-1})$: $Q = (q_{ab})$. Since $Q^{-1} = (q^{ab})$, there comes

$$\boxed{\theta_{(\ell)} = \frac{1}{2} q^{ab} \mathcal{L}_\ell q_{ab}}. \quad (2.60)$$

The Lie derivative along ℓ of the metric q of the cross-section \mathcal{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 $\mathcal{S} \rightarrow \mathcal{H}$ that corresponds to the displacement of points of \mathcal{S} by some infinitesimal quantity ε along ℓ . Using the notations of Fig. 2.12, we have then $p_\varepsilon = \Phi_\varepsilon(p)$, $q_\varepsilon = \Phi_\varepsilon(q)$ and $\mathcal{S}_\varepsilon = \Phi_\varepsilon(\mathcal{S})$. The **Lie derivative along ℓ of q** is then the field $\mathcal{L}_\ell q$ of bilinear forms on \mathcal{S} defined by the following action on any pair of vectors (u, v) tangent to \mathcal{S} at the same point p :

$$\boxed{\mathcal{L}_\ell q(u, v) := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} [q_\varepsilon(\Phi_{\varepsilon*} u, \Phi_{\varepsilon*} v) - q(u, v)]}, \quad (2.61)$$

where, as above, q_ε is the metric induced by g on the $(n-2)$ -surface \mathcal{S}_ε deduced from \mathcal{S} by Lie dragging along ℓ by the quantity ε and $\Phi_{\varepsilon*} u$ (resp. $\Phi_{\varepsilon*} v$) is the tangent vector to \mathcal{S}_ε at $\Phi_\varepsilon(p)$ that is the pushforward of u (resp. v) by the map Φ_ε (cf. Sec. A.2.8).

Remark 5: Since q is nothing but the metric induced by the spacetime metric g on cross-sections of \mathcal{H} , we may rewrite the above formula as

$$\mathcal{L}_\ell q(u, v) := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} [g|_{\Phi_\varepsilon(p)}(\Phi_{\varepsilon*} u, \Phi_{\varepsilon*} v) - g|_p(u, v)]. \quad (2.62)$$

One may wonder about the link between the Lie derivative $\mathcal{L}_\ell q$ defined by Eq. (2.61), which is a tensor field on \mathcal{S} , and the Lie derivative along ℓ of the spacetime extension q introduced by Eq. (2.34). For the sake of clarity, let us denote here the latter by \bar{q} . More precisely, we may consider that \bar{q} is a field defined in some neighborhood of the portion of \mathcal{H} sliced by $\bigcup_\varepsilon \mathcal{S}_\varepsilon$ via Eq. (2.34), with k defined at each point $p \in \mathcal{S}_\varepsilon$ as the unique null vector of $T_p^\perp \mathcal{S}_\varepsilon$ obeying $\ell \cdot k = -1$. Let u and v be vector fields on \mathcal{H} that are tangent to the cross-sections \mathcal{S}_ε . Applying the bilinear form $\mathcal{L}_\ell \bar{q}$ to them and using the Leibniz rule to expand $\mathcal{L}_\ell [\bar{q}(u, v)]$ yields

$$\mathcal{L}_\ell \bar{q}(u, v) = \mathcal{L}_\ell [\bar{q}(u, v)] - \bar{q}(\mathcal{L}_\ell u, v) - \bar{q}(u, \mathcal{L}_\ell v). \quad (2.63)$$

Now, since \mathbf{u} and \mathbf{v} are tangent to \mathcal{S}_ε , we may write $\bar{\mathbf{q}}(\mathbf{u}, \mathbf{v}) = \mathbf{q}(\mathbf{u}, \mathbf{v})$. 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 \mathcal{S}_ε are Lie-dragged along ℓ , the vectors $\mathcal{L}_\ell \mathbf{u}$ and $\mathcal{L}_\ell \mathbf{v}$ are also tangent to \mathcal{S}_ε . Therefore, we have $\bar{\mathbf{q}}(\mathcal{L}_\ell \mathbf{u}, \mathbf{v}) = \mathbf{q}(\mathcal{L}_\ell \mathbf{u}, \mathbf{v})$ and $\bar{\mathbf{q}}(\mathbf{u}, \mathcal{L}_\ell \mathbf{v}) = \mathbf{q}(\mathbf{u}, \mathcal{L}_\ell \mathbf{v})$ as well. Thus, we may rewrite (2.63) as

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

The right-hand side is identical to what would be obtained by expressing $\mathcal{L}_\ell \mathbf{q}(\mathbf{u}, \mathbf{v})$ via the Leibniz rule. Hence we conclude that $\mathcal{L}_\ell \bar{\mathbf{q}}(\mathbf{u}, \mathbf{v}) = \mathcal{L}_\ell \mathbf{q}(\mathbf{u}, \mathbf{v})$. Since this identity holds for a pair (\mathbf{u}, \mathbf{v}) of vectors tangent to \mathcal{S}_ε , we may express it for any pair of vectors, i.e. not necessarily tangent to \mathcal{S}_ε by introducing the orthogonal projector $\bar{\mathbf{q}}$ onto \mathcal{S}_ε [cf. Eq. (2.45)]:

$$\mathcal{L}_\ell \bar{\mathbf{q}}(\bar{\mathbf{q}}(\mathbf{u}), \bar{\mathbf{q}}(\mathbf{v})) = \mathcal{L}_\ell \mathbf{q}(\bar{\mathbf{q}}(\mathbf{u}), \bar{\mathbf{q}}(\mathbf{v})), \quad (2.64)$$

or, in index notation,

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

Taking the trace with respect to \mathbf{g} , we get $\mathcal{L}_\ell \bar{q}_{\mu\nu} \bar{q}^\mu{}_\sigma \bar{q}^\nu{}^\sigma = \mathcal{L}_\ell q_{ab} \bar{q}^a{}_\sigma \bar{q}^{b\sigma}$. Now, since $\bar{\mathbf{q}}$ is symmetric and $\bar{\mathbf{q}}$ 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

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

where the second equality follows from $\bar{q}^{ab} = q^{ab}$. Hence we may rewrite (2.60) as

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

(2.65)

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

Substituting (2.34) for $q_{\mu\nu}$, and using the Leibniz rule, we get

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

where the last equality follows from $q^{\mu\nu} \ell_\mu = 0$ and $q^{\mu\nu} k_\mu = 0$ [Eq. (2.48)]. Using the Killing expression (A.87) of the Lie derivative of \mathbf{g} : $\mathcal{L}_\ell g_{\mu\nu} = \nabla_\mu \ell_\nu + \nabla_\nu \ell_\mu$, and the symmetry of $q^{\mu\nu}$, we arrive at

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

(2.66)

We can transform this relation further by expressing $q^{\mu\nu}$ via (2.34):

$$\begin{aligned} \theta_{(\ell)} &= (g^{\mu\nu} + \ell^\mu k^\nu + k^\mu \ell^\nu) \nabla_\mu \ell_\nu = \nabla_\mu \ell^\mu + k^\nu \underbrace{\ell^\mu \nabla_\mu \ell_\nu}_{\kappa \ell_\nu} + k^\mu \ell^\nu \nabla_\mu \ell_\nu \\ &= \nabla_\mu \ell^\mu + \kappa \underbrace{k^\nu \ell_\nu}_{-1} + \frac{1}{2} \underbrace{k^\mu \nabla_\mu (\ell_\nu \ell^\nu)}_0 = \nabla_\mu \ell^\mu - \kappa, \end{aligned}$$

where we have used respectively the properties (2.22), (2.19) and (2.33). Denoting the divergence of ℓ by $\nabla \cdot \ell = \nabla_\mu \ell^\mu$, we may write

$$\theta_{(\ell)} = \nabla \cdot \ell - \kappa.$$

(2.67)

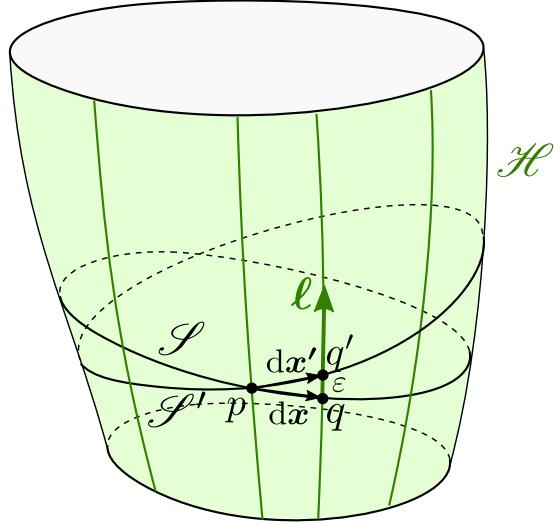


Figure 2.13: Two cross-sections \mathcal{S} and \mathcal{S}' through the same point p of \mathcal{H} .

Remark 6: Contrary to $\theta_{(\ell)}$ or κ , the quantity $\nabla \cdot \ell$ depends a priori on the extension of ℓ outside \mathcal{H} (cf. the discussion in Sec. 2.3.2). For Eq. (2.67) to hold, we have supposed that ℓ remains null outside \mathcal{H} , so that $k^\mu \nabla_\mu (\ell_\nu \ell^\nu)$, which is a derivative in a direction transverse to \mathcal{H} , could be set to zero in the computation leading to (2.67).

Example 19 (light cone): $\nabla \cdot \ell$ is easily computed by taking the trace of (2.27) and we have $\kappa = 0$ (cf. Example 9), so that (2.67) yields

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

Hence we recover the result obtained in Example 17.

Example 20 (Schwarzschild horizon): Here also, $\nabla \cdot \ell$ is easily computed by taking the trace of (2.28):

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

Given the value $\kappa = 4m/(r + 2m)^2$ found in Example 10, formula (2.67) leads to

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

Hence we recover the result (2.58).

We notice that the right-hand side of (2.67) is independent of the explicit choice of the cross-section \mathcal{S} : clearly both $\nabla \cdot \ell$ and κ depends only on the null normal ℓ of \mathcal{H} . This justifies the notation $\theta_{(\ell)}$, which does not refer to \mathcal{S} (cf. Remark 4 in page 43). This can be understood geometrically as follows. Let $p \in \mathcal{H}$ be a point where one would like to evaluate $\theta_{(\ell)}$. Let \mathcal{S} and \mathcal{S}' be two distinct cross-sections of \mathcal{H} going through p (cf. Fig. 2.13). Let q be

a point of \mathcal{S} infinitely close to p and let q' be the point of \mathcal{S}' located on the same null geodesic generator as q , i.e. $\overrightarrow{qq'} = \varepsilon\ell$, with ε infinitely small. Let $d\mathbf{x}$ (resp. $d\mathbf{x}'$) be the infinitesimal vector connecting p to q (resp. p to q'). We have then $d\mathbf{x}' = d\mathbf{x} + \varepsilon\ell$, the scalar square of which is

$$d\mathbf{x}' \cdot d\mathbf{x}' = d\mathbf{x} \cdot d\mathbf{x} + 2\varepsilon \underbrace{d\mathbf{x} \cdot \ell}_0 + \varepsilon^2 \underbrace{\ell \cdot \ell}_0,$$

where we have used the fact that ℓ is normal to any tangent vector to \mathcal{H} , such as $d\mathbf{x}$ and ℓ itself. Hence $d\mathbf{x}' \cdot d\mathbf{x}' = d\mathbf{x} \cdot d\mathbf{x}$. 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 \mathcal{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 $\theta_{(\ell)}$ at p does not depend on the choice of δS . We conclude:

Property 2.9: independence of the expansion from the cross-section

The expansion $\theta_{(\ell)}$ at a point p of a null hypersurface \mathcal{H} does not depend on the choice of the cross-section through p . It depends only on the null normal ℓ of \mathcal{H} . Accordingly, from now on, we shall call $\theta_{(\ell)}$ the **expansion of the null hypersurface \mathcal{H} along ℓ** .

The dependency of the expansion on ℓ is given by:

Property 2.10: rescaling of the expansion upon a change of null normal

Two null normals ℓ and ℓ' to a null hypersurface \mathcal{H} are necessarily collinear: $\ell' = \alpha\ell$, where α is a non-vanishing scalar field on \mathcal{H} . We have then the following scaling law:

$$\ell' = \alpha\ell \implies \theta_{(\ell')} = \alpha\theta_{(\ell)}. \quad (2.68)$$

Proof. This follows immediately from expression (2.56) for $\theta_{(\ell)}$, given that the metric q is independent of ℓ and $\mathcal{L}_{\alpha\ell} \ln q = \alpha\mathcal{L}_\ell \ln q$. One can also get the result from Eq. (2.66): $\theta_{(\ell')} = q^{\mu\nu}\nabla_\mu(\alpha\ell_\nu) = \alpha q^{\mu\nu}\nabla_\mu\ell_\nu = \alpha\theta_{(\ell)}$, since $q^{\mu\nu}\ell_\nu = 0$ [Eq. (2.48)]. \square

Remark 7: The reader may check that the rescaling laws (2.26) and (2.68) for respectively κ and $\theta_{(\ell)}$ are compatible with expression (2.67) for $\theta_{(\ell)}$, given that $\nabla \cdot \ell' = \alpha\nabla \cdot \ell + \nabla_\ell \alpha$.

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

$$\boxed{\theta_{(\ell)} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \frac{\delta A_\varepsilon - \delta A}{\delta A} = \frac{1}{2} \mathcal{L}_\ell \ln q = \frac{1}{2} q^{\mu\nu} \mathcal{L}_\ell q_{\mu\nu} = q^{\mu\nu} \nabla_\mu \ell_\nu = \nabla \cdot \ell - \kappa}, \quad (2.69)$$

with the reminder that the last equality is valid insofar as the vector field ℓ is null in some entire open neighborhood of \mathcal{H} (and not only on \mathcal{H}), as stressed in Remark 6.

2.3.6 Deformation rate and shear tensor

Let us consider a cross-section \mathcal{S} of the null hypersurface \mathcal{H} . The **deformation rate** Θ of \mathcal{S} **along** ℓ is defined from the Lie derivative along ℓ of the induced metric q on \mathcal{S} [cf. Eq. (2.61)] as

$$\Theta := \frac{1}{2} \vec{q}^* \mathcal{L}_\ell q, \quad (2.70)$$

where \vec{q}^* stands for the action of the orthogonal projector \vec{q} onto \mathcal{S} on the bilinear form $\mathcal{L}_\ell q$. This action extends $\mathcal{L}_\ell q$, which is defined a priori only for vectors of $T_p \mathcal{S}$ by Eq. (2.61), to all vectors of $T_p \mathcal{M}$, for any $p \in \mathcal{S}$, via

$$\forall (\mathbf{u}, \mathbf{v}) \in T_p \mathcal{M} \times T_p \mathcal{M}, \quad \vec{q}^* \mathcal{L}_\ell q (\mathbf{u}, \mathbf{v}) := \mathcal{L}_\ell q (\vec{q}(\mathbf{u}), \vec{q}(\mathbf{v})). \quad (2.71)$$

Since q is symmetric, it is clear from the above definition that Θ is a field of symmetric bilinear forms. Thanks to the identity (2.64), we can use for q the spacetime extension (2.34) and write the index-notation version of the definition (2.70) as

$$\Theta_{\alpha\beta} = \frac{1}{2} q^\mu{}_\alpha q^\nu{}_\beta \mathcal{L}_\ell q_{\mu\nu}. \quad (2.72)$$

Substituting $g_{\mu\nu} + \ell_\mu k_\nu + k_\mu \ell_\nu$ for $q_{\mu\nu}$ [Eq. (2.34)], we get

$$\Theta_{\alpha\beta} = \frac{1}{2} q^\mu{}_\alpha q^\nu{}_\beta (\mathcal{L}_\ell g_{\mu\nu} + k_\nu \mathcal{L}_\ell \ell_\mu + \ell_\mu \mathcal{L}_\ell k_\nu + \ell_\nu \mathcal{L}_\ell k_\mu + k_\mu \mathcal{L}_\ell \ell_\nu) = \frac{1}{2} q^\mu{}_\alpha q^\nu{}_\beta \mathcal{L}_\ell g_{\mu\nu},$$

where the last equality follows from $q^\mu{}_\alpha \ell_\mu = 0$ and $q^\mu{}_\alpha k_\mu = 0$ [Eq. (2.48)]. Now, using the Killing expression (A.87) and the Frobenius identity (2.16), we may write

$$\mathcal{L}_\ell g_{\mu\nu} = \nabla_\mu \ell_\nu + \nabla_\nu \ell_\mu = 2\nabla_\mu \ell_\nu + \nabla_\nu \rho \ell_\mu - \nabla_\mu \rho \ell_\nu.$$

Given that $q^\mu{}_\alpha \ell_\mu = 0$, we obtain

$$\Theta_{\alpha\beta} = q^\mu{}_\alpha q^\nu{}_\beta \nabla_\mu \ell_\nu. \quad (2.73)$$

Let us substitute (2.45) for the projector \vec{q} :

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

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

$$\nabla_\alpha \ell_\beta = \Theta_{\alpha\beta} + \omega_\alpha \ell_\beta - \ell_\alpha k^\mu \nabla_\mu \ell_\beta, \quad (2.74)$$

where we have let appear the 1-form ω defined by

$$\omega_\alpha := -k^\mu \nabla_\nu \ell_\mu \Pi^\nu{}_\alpha = -k^\mu \nabla_\alpha \ell_\mu - k^\mu k^\nu \nabla_\mu \ell_\nu \ell_\alpha. \quad (2.75)$$

Here Π is the projector onto \mathcal{H} along k :

$$\Pi^\alpha{}_\beta = \delta^\alpha{}_\beta + k^\alpha \ell_\beta. \quad (2.76)$$

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

$$\forall \mathbf{v} \in T_p \mathcal{H}, \quad \langle \omega, \mathbf{v} \rangle = -\mathbf{k} \cdot \nabla_{\mathbf{v}} \ell. \quad (2.77)$$

In particular, for $\mathbf{v} = \ell$, we get $\langle \omega, \ell \rangle = -\mathbf{k} \cdot \nabla_\ell \ell$. Given that $\nabla_\ell \ell = \kappa \ell$ [Eq. (2.22)] and $\mathbf{k} \cdot \ell = -1$ [Eq. (2.33)], we arrive at the simple formula:

$$\boxed{\langle \omega, \ell \rangle = \kappa}. \quad (2.78)$$

Remark 8: Thanks to the projector Π involved in its definition, the 1-form ω does not depend on the extension of the vector field ℓ away from \mathcal{H} . The same property holds for Θ . On the contrary, the tensor field ∇_ℓ , which appears in the left-hand side of formula (2.74) and in the last term of its right-hand side, depends on the extension of ℓ away from \mathcal{H} .

By comparing (2.65) and (2.72), we notice that the trace of Θ is nothing but the expansion $\theta_{(\ell)}$:

$$\boxed{\theta_{(\ell)} = g^{\mu\nu} \Theta_{\mu\nu} = q^{\mu\nu} \Theta_{\mu\nu} = \Theta^\mu_\mu}. \quad (2.79)$$

The trace-free part of Θ is called the *shear tensor of \mathcal{S} along ℓ* :

$$\boxed{\sigma := \Theta - \frac{1}{n-2} \theta_{(\ell)} \mathbf{q}}, \quad (2.80)$$

or, in index notation:

$$\sigma_{\alpha\beta} = \Theta_{\alpha\beta} - \frac{1}{n-2} \theta_{(\ell)} q_{\alpha\beta}. \quad (2.81)$$

Remark 9: The $1/(n-2)$ factor arises from the trace of \mathbf{q} , which is $n-2$. This follows immediately from \mathbf{q} being a metric tensor on the $(n-2)$ -dimensional manifold \mathcal{S} ; this can also be recovered from the spacetime extension (2.34) of \mathbf{q} : $q^\mu_\mu = \delta^\mu_\mu + 2\ell_\mu k^\mu$, with $\delta^\mu_\mu = n$ and $\ell_\mu k^\mu = -1$.

By construction, we have $\sigma^\mu_\mu = g^{\mu\nu} \sigma_{\mu\nu} = q^{\mu\nu} \sigma_{\mu\nu} = 0$. Note that, as \mathbf{q} , the tensor fields Θ and σ are tangent to \mathcal{S} , in the sense that

$$\boxed{\forall \mathbf{v} \in T_p^\perp \mathcal{S}, \quad \Theta(\mathbf{v}, .) = \sigma(\mathbf{v}, .) = 0}, \quad (2.82)$$

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

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 ℓ (i.e. the θ and φ are constant along the field lines of ℓ on \mathcal{H} and $\ell = \partial/\partial t$ in these coordinates, in other words, $\ell^\alpha = (1, 0, 0)$), we have [cf. formula (A.85) in Appendix A]

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

where the second equality follows from $t = r$ on \mathcal{H} . Given that $q_{ab} = \text{diag}(r^2, r^2 \sin^2 \theta)$ [cf. Eq. (2.38)], we obtain

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

Hence (2.70) yields

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

Taking the trace, we get immediately $\theta_{(\ell)} = 2/r$, i.e. we recover the result of Examples 17 and 19. From (2.80), we get a vanishing shear:

$$\sigma = 0.$$

Example 22 (Schwarzschild horizon): The Lie derivative of q , as given by Eq. (2.41), along ℓ is (cf. Appendix D for the computation):

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

Since $\vec{q}^* q = q$, Eq. (2.70) yields

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

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

$$\Theta = 0. \quad (2.83)$$

Property 2.11: rescaling of the deformation rate and shear upon a change of null normal

Two null normals ℓ and ℓ' to a null hypersurface \mathcal{H} are necessarily collinear: $\ell' = \alpha \ell$, where α is a non-vanishing scalar field on \mathcal{H} . The deformation rates Θ and Θ' and shear tensors σ and σ' of a given cross-section \mathcal{S} of \mathcal{H} along respectively ℓ and ℓ' obey the following scaling law:

$$\ell' = \alpha \ell \implies \Theta' = \alpha \Theta \text{ and } \sigma' = \alpha \sigma. \quad (2.84)$$

Proof. Equation (2.73) yields $\Theta'_{\alpha\beta} = q^\mu_\alpha q^\nu_\beta \nabla_\mu (\alpha \ell_\nu) = \alpha q^\mu_\alpha q^\nu_\beta \nabla_\mu \ell_\nu + q^\mu_\alpha q^\nu_\beta \ell_\nu \nabla_\mu \alpha = \alpha \Theta_{\alpha\beta} + 0$ since $q^\nu_\beta \ell_\nu = 0$ [Eq. (2.48)]. The result for the shear tensor follows then from the definition (2.80) along with the scaling law (2.68) for the expansion. \square

Remark 10: The scaling law (2.84) shows explicitly that the deformation rate Θ and shear tensor σ depend on the choice of the null normal ℓ to \mathcal{H} . Both tensors also depend on the choice of the cross-section \mathcal{S} of \mathcal{H} , contrary to the expansion $\theta_{(\ell)}$ (Property 2.9).

2.4 Evolution of the expansion

2.4.1 Null Raychaudhuri equation

Let us derive an evolution equation for the expansion $\theta_{(\ell)}$; it is quite natural to consider the evolution along the null generators of \mathcal{H} , i.e. to evaluate the quantity $\nabla_\ell \theta_{(\ell)}$, all the more that ℓ is by hypothesis future-directed. The starting point is the contracted Ricci identity [Eq. (A.107)] applied to ℓ :

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

where $R_{\mu\alpha}$ stands for the Ricci tensor of \mathbf{g} . Substituting Eq. (2.74) for $\nabla_\alpha \ell^\mu$ and $\theta_{(\ell)} + \kappa$ for $\nabla_\mu \ell^\mu = \nabla \cdot \ell$ [cf. Eq. (2.67)] yields

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

Expanding the left-hand side and using again Eqs. (2.67) and (2.74) results in

$$\begin{aligned} \nabla_\mu \Theta_\alpha^\mu + \ell^\mu \nabla_\mu \omega_\alpha - \nabla_\alpha (\theta_{(\ell)} + \kappa) + (\theta_{(\ell)} + \kappa) \omega_\alpha - \Theta_{\alpha\mu} k^\nu \nabla_\nu \ell^\mu \\ - [\omega_\mu k^\nu \nabla_\nu \ell^\mu + \nabla_\mu (k^\nu \nabla_\nu \ell^\mu)] \ell_\alpha = R_{\mu\alpha} \ell^\mu. \end{aligned} \quad (2.85)$$

The above relation is a 1-form identity that has interesting consequences. In Chap. 3, we shall apply it to tangent vectors to a cross-section in order to get the so-called zeroth law of black hole mechanics. Here we shall instead apply it to the normal vector field ℓ (i.e. contract it with ℓ^α); since $\omega_\alpha \ell^\alpha = \kappa$ [Eq. (2.78)] and $\ell_\alpha \ell^\alpha = 0$, we get

$$\ell^\nu \nabla_\mu \Theta_\nu^\mu + \ell^\nu \ell^\mu \nabla_\mu \omega_\nu - \ell^\mu \nabla_\mu (\theta_{(\ell)} + \kappa) + \kappa (\theta_{(\ell)} + \kappa) = R_{\mu\nu} \ell^\mu \ell^\nu. \quad (2.86)$$

Now, using $\Theta_\nu^\mu \ell^\nu = 0$ [Eq. (2.82)] and Eq. (2.74) to express $\nabla_\mu \ell_\nu$, we can write

$$\begin{aligned} \ell^\nu \nabla_\mu \Theta_\nu^\mu &= \nabla_\mu (\underbrace{\Theta_\nu^\mu \ell^\nu}_0) - \Theta_\nu^\mu \nabla_\mu \ell^\nu = -\Theta^{\mu\nu} \nabla_\mu \ell_\nu = -\Theta^{\mu\nu} (\Theta_{\mu\nu} + \omega_\mu \ell_\nu - \ell_\mu k^\sigma \nabla_\sigma \ell_\nu) \\ &= -\Theta^{\mu\nu} \Theta_{\mu\nu}. \end{aligned}$$

On the other side, thanks to Eqs. (2.78) and (2.22), we have

$$\ell^\nu \ell^\mu \nabla_\mu \omega_\nu = \ell^\mu \nabla_\mu (\underbrace{\omega_\nu \ell^\nu}_\kappa) - \omega_\nu \underbrace{\ell^\mu \nabla_\mu \ell^\nu}_{\kappa \ell^\nu} = \ell^\mu \nabla_\mu \kappa - \kappa^2.$$

Accordingly Eq. (2.86) simplifies to

$$-\Theta^{\mu\nu} \Theta_{\mu\nu} - \ell^\mu \nabla_\mu \theta_{(\ell)} + \kappa \theta_{(\ell)} = R_{\mu\nu} \ell^\mu \ell^\nu. \quad (2.87)$$

The first term in the left-hand side can be evaluated by expressing Θ in terms of $\theta_{(\ell)}$ and the shear tensor σ [Eq. (2.80)]:

$$\begin{aligned} \Theta_{\mu\nu} \Theta^{\mu\nu} &= \left(\sigma_{\mu\nu} + \frac{1}{n-2} \theta_{(\ell)} q_{\mu\nu} \right) \left(\sigma^{\mu\nu} + \frac{1}{n-2} \theta_{(\ell)} q^{\mu\nu} \right) \\ &= \underbrace{\sigma_{\mu\nu} \sigma^{\mu\nu}}_{\sigma_{ab} \sigma^{ab}} + \frac{2}{n-2} \theta_{(\ell)} \underbrace{q^{\mu\nu} \sigma_{\mu\nu}}_0 + \frac{1}{(n-2)^2} \theta_{(\ell)}^2 \underbrace{q_{\mu\nu} q^{\mu\nu}}_{n-2} = \sigma_{ab} \sigma^{ab} + \frac{1}{n-2} \theta_{(\ell)}^2. \end{aligned}$$

Hence Eq. (2.87) leads to:

Property 2.12: null Raychaudhuri equation

Any normal vector field ℓ to a null hypersurface \mathcal{H} of a n -dimensional spacetime (\mathcal{M}, g) obeys the **null Raychaudhuri equation**:

$$\nabla_\ell \theta_{(\ell)} = \kappa \theta_{(\ell)} - \frac{1}{n-2} \theta_{(\ell)}^2 - \sigma_{ab} \sigma^{ab} - R(\ell, \ell). \quad (2.88)$$

If ℓ is future-directed, this is an evolution equation for $\theta_{(\ell)}$ along the null generators of \mathcal{H} .

Remark 1: Actually Eq. (2.88) is a particular case of what is generally called the null Raychaudhuri equation, namely the case where the vorticity of the vector field ℓ vanishes. This applies because ℓ is hypersurface-orthogonal, being normal to \mathcal{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 \mathcal{M} . A null vector field ℓ tangent to the geodesics of the congruence has a priori some vorticity w and a term $+w_{ab}w^{ab}$ must be added to the right-hand side of Eq. (2.88) (see e.g. Eq. (4.35) of Ref. [265]).

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

If the spacetime (\mathcal{M}, g) is ruled by general relativity, i.e. if g obeys Einstein equation (1.42), we may express the term involving the Ricci tensor in terms of the total energy-momentum tensor T :

$$R(\ell, \ell) = \frac{2}{n-2} \Lambda \underbrace{g(\ell, \ell)}_0 + 8\pi [T(\ell, \ell) - \frac{1}{n-2} T \underbrace{g(\ell, \ell)}_0] = 8\pi T(\ell, \ell). \quad (2.89)$$

The null Raychaudhuri equation becomes then

$$\nabla_\ell \theta_{(\ell)} = \kappa \theta_{(\ell)} - \frac{1}{n-2} \theta_{(\ell)}^2 - \sigma_{ab} \sigma^{ab} - 8\pi T(\ell, \ell). \quad (2.90)$$

Remark 3: The cosmological constant Λ does not appear in the null Raychaudhuri equation (2.90), despite the latter involves the Einstein equation with Λ [Eq. (1.42)].

Remark 4: We have stressed above that the shear tensor σ does depend on the choice of a cross-section of \mathcal{H} , in addition to ℓ (cf. Remark 10 on p. 52). The null Raychaudhuri equation (2.88) shows that the “shear square” $\sigma_{ab} \sigma^{ab}$ does *not* depend on the choice of a cross-section, for all the other terms in (2.88) are independent of it.

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 $\sigma = 0$, hence $\sigma_{ab} \sigma^{ab} = 0$. Moreover, the Ricci tensor of Minkowski spacetime vanishes identically. The null Raychaudhuri equation (2.88) reduces then to

$$\nabla_\ell \theta_{(\ell)} = -\frac{1}{2} \theta_{(\ell)}^2,$$

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

Example 24 (Schwarzschild horizon): For the Schwarzschild horizon \mathcal{H} , the null Raychaudhuri equation is trivially satisfied, i.e. each of its terms vanishes identically: $\theta_{(\ell)} = 0$ on \mathcal{H} [Eq. (2.59)], which implies $\nabla_\ell \theta_{(\ell)} = 0$ since ℓ is tangent to \mathcal{H} , $\sigma = 0$ since $\Theta = 0$ [Eq. (2.83)] and the Ricci tensor of the metric (2.5) is zero (cf. the notebook D.4.1).

2.4.2 Null convergence condition

If one chooses the null normal ℓ associated to an affine parametrization of the null generators of \mathcal{H} , then $\kappa = 0$ and the null Raychaudhuri equation (2.88) reduces to

$$\nabla_\ell \theta_{(\ell)} = -\frac{1}{n-2} \theta_{(\ell)}^2 - \sigma_{ab} \sigma^{ab} - \mathbf{R}(\ell, \ell). \quad (2.91)$$

The first term in the right-hand side is manifestly non-positive. The same property holds for the second term, since

$$\sigma_{ab} \sigma^{ab} \geq 0. \quad (2.92)$$

To prove (2.92), it suffices to consider a q -orthonormal basis of $T_p \mathcal{S}$. Since q is a Riemannian metric, the components of its inverse in that basis are $q^{ab} = \text{diag}(1, \dots, 1) = \delta^{ab}$. Hence $\sigma^{ab} = q^{ac} q^{bd} \sigma_{cd} = \delta^{ac} \delta^{bd} \sigma_{cd} = \sigma_{ab}$, so that

$$\sigma_{ab} \sigma^{ab} = \sum_{a=2}^{n-1} \sum_{b=2}^{n-1} (\sigma_{ab})^2 \geq 0. \quad (2.93)$$

Regarding the last term in the right-hand side of Eq. (2.91), it is non-positive if the Ricci tensor obeys the so-called **null convergence condition**:

$$\boxed{\mathbf{R}(\ell, \ell) \geq 0 \quad \text{for any null vector } \ell}. \quad (2.94)$$

If general relativity is assumed, i.e. if the Einstein equation (1.40) is fulfilled, we have $\mathbf{R}(\ell, \ell) = 8\pi \mathbf{T}(\ell, \ell)$ [Eq. (2.89)], where \mathbf{T} is the total energy-momentum tensor of the matter and non-gravitational fields. Accordingly, the null convergence condition is equivalent to the so-called **null energy condition**:

$$\boxed{\mathbf{T}(\ell, \ell) \geq 0 \quad \text{for any null vector } \ell}_{\text{GR}}, \quad (2.95)$$

where the index ‘GR’ stands for *general relativity*. The null energy condition (2.95) is a pretty weak physical requirement: it is satisfied by

- vacuum: $\mathbf{T} = 0$;
- any “reasonable” matter model, such as a perfect fluid with a proper energy density ρ and pressure p obeying $\rho + p \geq 0$ ⁸;

⁸Indeed, for the perfect-fluid energy-momentum tensor $\mathbf{T} = (\rho + p)\underline{\mathbf{u}} \otimes \underline{\mathbf{u}} + p\mathbf{g}$, one has $\mathbf{T}(\ell, \ell) = (\rho + p)(\mathbf{u} \cdot \ell)^2$ with $(\mathbf{u} \cdot \ell)^2 \geq 0$.

- any electromagnetic field⁹;
- any massless scalar field [265, 320];
- “dark energy” modeled by $\mathbf{T} = -\frac{\Lambda}{8\pi} \mathbf{g}$.

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

$$\mathbf{T}(\mathbf{u}, \mathbf{u}) \geq 0 \quad \text{for any timelike vector } \mathbf{u}. \quad (2.96)$$

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

If the null convergence condition holds, all the terms in the right-hand side of the reduced Raychaudhuri equation (2.91) are non-positive and we may conclude:

Property 2.13: non-increase of the expansion

If the null convergence condition (2.94) is fulfilled on a null hypersurface \mathcal{H} – for general relativity, this is equivalent to demanding that the null energy condition (2.95) holds on \mathcal{H} –, then the expansion $\theta_{(\ell)}$ along any null normal ℓ associated to an affine parametrization of \mathcal{H} 's null generators cannot increase along ℓ , i.e. it obeys

$$\nabla_\ell \theta_{(\ell)} \leq 0. \quad (2.97)$$

This property justifies the name *convergence condition* given to $R(\ell, \ell) \geq 0$: if $\theta_{(\ell)} < 0$ in some part of \mathcal{H} , so that nearby null generators are converging, Property 2.13 shows that the null generators cannot subsequently diverge.

⁹This is readily seen from the electromagnetic energy-momentum tensor (1.52), which yields $\mathbf{T}(\ell, \ell) = \mu_0^{-1} \mathbf{E} \cdot \mathbf{E}$ where $E^\alpha := F^\alpha_\mu \ell^\mu$. Thanks to the antisymmetry of \mathbf{F} , the vector \mathbf{E} obeys $\mathbf{E} \cdot \ell = 0$. It follows from Property 1.3 that \mathbf{E} cannot be timelike. Hence $\mathbf{E} \cdot \mathbf{E} \geq 0$ and $\mathbf{T}(\ell, \ell) \geq 0$.

Chapter 3

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

Contents

3.1	Introduction	57
3.2	Non-expanding horizons	57
3.3	Killing horizons	63
3.4	Bifurcate Killing horizons	83
3.5	Summary	89

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*, which occur in spacetimes that are globally stationary (Sects. 3.3 and 3.4). Actually, we shall see in Chap. 5 that, under some rather general hypotheses, 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 boundary¹) and have a constant area, i.e. a vanishing expansion. Hence the definition:

A **non-expanding horizon** is a null hypersurface \mathcal{H} having the topology

$$\mathcal{H} \simeq \mathbb{R} \times \mathcal{S}, \quad (3.1)$$

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

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

Remark 1: Given the scaling law (2.68), if $\theta_{(\ell)} = 0$ for some normal ℓ , then $\theta_{(\ell')} = 0$ for any other normal ℓ' . Hence the definition of a non-expanding horizon does not depend on the choice of the null normal.

The above definition captures only the event horizon of black holes in equilibrium, to be discussed in detail in Chap. 5. For a black hole out of equilibrium, one has generically $\theta_{(\ell)} > 0$, as we shall see in Chap. 16.

Example 1 (Schwarzschild horizon): In view of Eq. (2.59), 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 Hájíček in 1973 under the name of *totally geodesic null hypersurface* [250] or *perfect horizon* [251, 252]. The terminology *non-expanding horizon* is due to Abhay Ashtekar, Stephen Fairhurst and Badri Krishnan in 2000 [25] (see also [23]).

3.2.2 Invariance of the area

Given a complete cross-section \mathcal{S} of \mathcal{H} , the area of \mathcal{S} , with respect to the spacetime metric \mathbf{g} , is [cf. Eqs. (2.50) and (2.53)]

$$A = \int_{\mathcal{S}} \mathcal{S}\epsilon(d\mathbf{x}_{(2)}, \dots, d\mathbf{x}_{(n-1)}) = \int_{\mathcal{S}} \sqrt{q} dx^2 \cdots dx^{n-1}, \quad (3.3)$$

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

¹Cf. the discussion in Sec. 2.3.4.

Property 3.1: area of a non-expanding horizon

The area A does not depend on the choice of the complete cross-section \mathcal{S} of \mathcal{H} . For this reason, A is called the ***area of the non-expanding horizon*** \mathcal{H} .

Proof. Let \mathcal{S}' be a second complete cross-section of \mathcal{H} . Let us assume first that $\mathcal{S}' \cap \mathcal{S} = \emptyset$. Then each null geodesic generator \mathcal{L} of \mathcal{H} intersects \mathcal{S} and \mathcal{S}' in two distinct points and it is possible to chose a parameter λ of \mathcal{L} such that $\lambda = 0$ on \mathcal{S} and $\lambda = 1$ on \mathcal{S}' . Let $\ell = dx/d\lambda|_{\mathcal{L}}$ be the associated tangent vector. We may then say that the cross-section \mathcal{S}' is deduced from \mathcal{S} by the Lie dragging of \mathcal{S} along ℓ by the parameter increase $\delta\lambda = 1$. More precisely, we may consider that \mathcal{S}' is deduced from \mathcal{S} by a continuous deformation, represented by a 1-parameter family (\mathcal{S}_λ) of cross-sections such that $\mathcal{S}_0 = \mathcal{S}$ and $\mathcal{S}_1 = \mathcal{S}'$. Associated with this family is a real-valued function $\lambda \mapsto A(\lambda)$ given the area of each element \mathcal{S}_λ . By the very definition of the expansion along ℓ [Eq. (2.49)], we have then

$$\frac{dA}{d\lambda} = \int_{\mathcal{S}_\lambda} \theta_{(\ell)} \delta A.$$

If \mathcal{H} is a non-expanding horizon, then $\theta_{(\ell)} = 0$ and it follows that $A(\lambda)$ is a constant function. Hence the area of \mathcal{S}' is equal to that of \mathcal{S} . If \mathcal{S} and \mathcal{S}' have a non-empty intersecting part, the argument can be repeated on each of the non-intersecting parts, given that the area of the part $\mathcal{S} \cap \mathcal{S}'$ is obviously the same for \mathcal{S} and \mathcal{S}' . \square

Example 3 (Schwarzschild horizon): The area of the Schwarzschild horizon is readily computed from the metric (2.31): $q = 4m^2 (d\theta^2 + \sin^2 \theta d\varphi^2)$, yielding $\sqrt{q} = 4m^2$; we get

$$A = 16\pi m^2. \quad (3.4)$$

3.2.3 Trapped surfaces

If there exists some natural concept of *outer/inner* region with respect to \mathcal{H} , for instance the outer region being the one having some asymptotically flat end, and if the transverse null normals \mathbf{k} to cross-sections point to the inner region, then the property $\theta_{(\ell)} = 0$ means that any cross-section \mathcal{S} of the non-expanding horizon \mathcal{H} is a ***marginally outer trapped surface*** (often abridged as **MOTS**). This definition is due to Hawking [260], an ***outer trapped surface*** would be one for which $\theta_{(\ell)} \leq 0$.

The MOTS definition is related to, but distinct from, the definition of a marginally trapped surface by Penrose [404]: a $(n - 2)$ -dimensional submanifold \mathcal{S} of \mathcal{M} is a ***trapped surface*** iff (i) \mathcal{S} is closed (i.e. compact without boundary), (ii) \mathcal{S} is spacelike and (iii) the two systems of null geodesics emerging orthogonally from \mathcal{S} towards the future converge locally at \mathcal{S} , i.e. they have negative expansions:

$$\theta_{(\ell)} < 0 \quad \text{and} \quad \theta_{(k)} < 0, \quad (3.5)$$

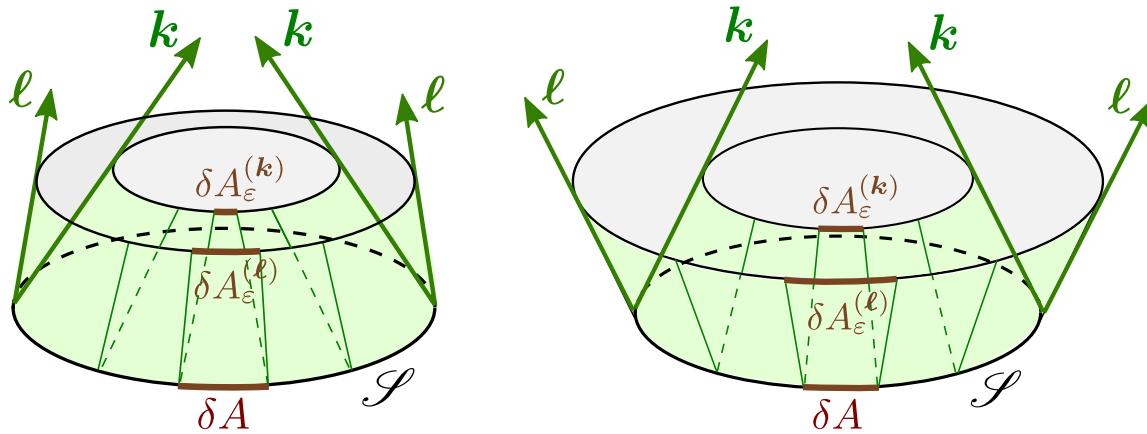


Figure 3.1: Trapped surface (left): $\delta A_\varepsilon^{(k)} < \delta A$ and untrapped surface (right): $\delta A_\varepsilon^{(k)} > \delta A$, both surfaces having $\delta A_\varepsilon^{(k)} < \delta A$.

where the expansion along k is defined in the same way as that along ℓ [cf. Eq. (2.69)]:

$$\theta_{(k)} := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \frac{\delta A_\varepsilon^{(k)} - \delta A}{\delta A} = \frac{1}{2} \mathcal{L}_k \ln q = \frac{1}{2} q^{\mu\nu} \mathcal{L}_k q_{\mu\nu} = q^{\mu\nu} \nabla_\mu k_\nu, \quad (3.6)$$

$\delta A_\varepsilon^{(k)}$ being the area of the surface element that is deduced from the surface element of area δA on \mathcal{S} by the Lie dragging along k by a parameter ε (cf. Fig. 3.1). The limit case $\theta_{(\ell)} = 0$ and $\theta_{(k)} < 0$ correspond to the so-called **marginally trapped surface**. In flat spacetime (Minkowski), given any closed spacelike surface, one has $\theta_{(\ell)} > 0$ and $\theta_{(k)} < 0$ (cf. Fig. 3.1 right), so there is no trapped surface. We shall discuss trapped surfaces in more detail in Sec. 18.2.1.

Cross-sections of a non-expanding horizon are usually marginally trapped surfaces (cf. the example below). However, there exist some pathological situations for which $\theta_{(k)} > 0$ at some points of \mathcal{S} [216].

Example 4: Let us consider a cross-section \mathcal{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 (2.40), we get (cf. the notebook D.4.1 for the computation) $\theta_{(k)} = -(r + 2m)/r^2$. In particular, on \mathcal{S} ($r = 2m$), $\theta_{(k)} = -1/m$. Hence $\theta_{(k)} < 0$. Since we had already $\theta_{(\ell)} = 0$ [cf. Eq. (2.59)], we conclude that \mathcal{S} is a marginally trapped surface. This could also have been inferred from Fig. 2.11, since according to the metric (2.41), the area of the cross-sections of \mathcal{H} is nothing but $4\pi r^2$ and k points to decreasing values of r , while, on \mathcal{H} , ℓ points to a fixed value of r .

3.2.4 Vanishing of the deformation rate tensor

If \mathcal{H} is a non-expanding horizon, we may set $\theta_{(\ell)} = 0$ in the null Raychaudhuri equation (2.88); it reduces then to

$$\sigma_{ab} \sigma^{ab} + \mathbf{R}(\ell, \ell) = 0, \quad (3.7)$$

where \mathbf{R} is the Ricci tensor of the metric \mathbf{g} . Let us assume that the null convergence condition discussed in Sec. 2.4.2 holds: $\mathbf{R}(\ell, \ell) \geq 0$ [Eq. (2.94)]. Then, given the property $\sigma_{ab} \sigma^{ab} \geq 0$

[Eq. (2.92)], Eq. (3.7) implies both $\sigma_{ab}\sigma^{ab} = 0$ and

$$\boxed{\mathbf{R}(\ell, \ell) = 0}. \quad (3.8)$$

The identity $\sigma_{ab}\sigma^{ab} = 0$ is possible only if each of the terms σ_{ab} in the sum (2.93) is zero. Hence we have necessarily $\boldsymbol{\sigma} = 0$. Since we had already $\theta_{(\ell)} = 0$ (\mathcal{H} non-expanding), this implies that the full deformation rate tensor Θ vanishes identically [cf. Eq. (2.80)]. Moreover, given the scaling law (2.84), this holds for any null normal ℓ . We can thus conclude:

Property 3.2: invariance of the cross-section metric along the null generators

Provided that the null convergence condition (2.94) holds — which occurs in general relativity if the null energy condition (2.95) holds —, the deformation rate of any cross-section \mathcal{S} of a non-expanding horizon \mathcal{H} along any null normal ℓ is identically zero:

$$\boxed{\Theta := \frac{1}{2} \vec{q}^* \mathcal{L}_\ell q = 0}. \quad (3.9)$$

In other words, the whole metric q (and not only the area form \mathcal{E} , as a mere $\theta_{(\ell)} = 0$ would suggest) of any cross-section \mathcal{S} is invariant along the geodesic generators of \mathcal{H} .

Example 5 (Schwarzschild horizon): We had already noticed that, for the Schwarzschild horizon, $\Theta = 0$ [Eq. (2.83) in Example 22 of Chap. 2].

3.2.5 Induced affine connection

Since \mathcal{H} is a null hypersurface, the “metric” $g|_{\mathcal{H}}$ induced on it by the spacetime metric g is degenerate. As a consequence, there is a priori no unique connection on \mathcal{H} associated with $g|_{\mathcal{H}}$. However, when \mathcal{H} is a non-expanding horizon and the null convergence condition holds on \mathcal{H} , so that $\Theta = 0$ (Property 3.2), the spacetime connection ∇ induces a unique torsion-free connection $\mathcal{H}\nabla$ on \mathcal{H} as follows. Let u and v be two vector fields on \mathcal{H} . We have, using (2.74) to express $\nabla_\nu \ell_\mu$ in terms of $\Theta_{\nu\mu}$:

$$\ell_\mu u^\nu \nabla_\nu v^\mu = u^\nu \nabla_\nu (\underbrace{\ell_\mu v^\mu}_0) - v^\mu u^\nu \nabla_\nu \ell_\mu = - \underbrace{\Theta_{\nu\mu}}_0 v^\mu u^\nu - \omega_\nu u^\nu \underbrace{\ell_\mu v^\mu}_0 + v^\mu \underbrace{u^\nu \ell_\nu}_0 k^\sigma \nabla_\sigma \ell_\mu = 0.$$

Hence ℓ is orthogonal to the vector field $\nabla_u v$. It follows immediately that $\nabla_u v$ is tangent to \mathcal{H} . We conclude:

Property 3.3: induced affine connection

Under the hypotheses of Property 3.2, the operator

$$\begin{aligned} \mathcal{H}\nabla : \mathfrak{X}(\mathcal{H}) \times \mathfrak{X}(\mathcal{H}) &\longrightarrow \mathfrak{X}(\mathcal{H}) \\ (\mathbf{u}, \mathbf{v}) &\longmapsto \nabla_u v, \end{aligned} \quad (3.10)$$

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

As a consequence of the identity ${}^{\mathcal{H}}\nabla_u v = \nabla_u v$ for tangent vector fields to \mathcal{H} , $(\mathcal{H}, {}^{\mathcal{H}}\nabla)$ is a **totally geodesic submanifold** of (\mathcal{M}, g) , i.e. any geodesic of $(\mathcal{H}, {}^{\mathcal{H}}\nabla)$ is also a geodesic of (\mathcal{M}, g) (cf. the historical note on p. 58).

Property 3.4: horizon-intrinsic derivative of the null normal

Under the hypotheses of Property 3.2, the covariant derivative of the null normal ℓ with respect to the affine connection ${}^{\mathcal{H}}\nabla$ takes the form

$${}^{\mathcal{H}}\nabla\ell = \ell \otimes {}^{\mathcal{H}}\omega, \quad (3.11)$$

where ${}^{\mathcal{H}}\omega$ is the tensor field on \mathcal{H} defined as the restriction of the 1-form ω introduced in Sec. 2.3.6 to tangent vectors to \mathcal{H} [cf. Eq. (2.77)]. In other words, ${}^{\mathcal{H}}\omega$ is the pullback $\iota^*\omega$ of ω by the inclusion map $\iota : \mathcal{H} \rightarrow \mathcal{M}$ (cf. Sec. A.2.8).

Proof. By definition of the covariant derivative with respect to an affine connection [cf. Eq. (A.62) in Appendix A], ${}^{\mathcal{H}}\nabla\ell$ is a tensor field of type $(1, 1)$ on \mathcal{H} , the action of which on any pair (a, u) formed by a 1-form a on \mathcal{H} and a vector field u on \mathcal{H} is

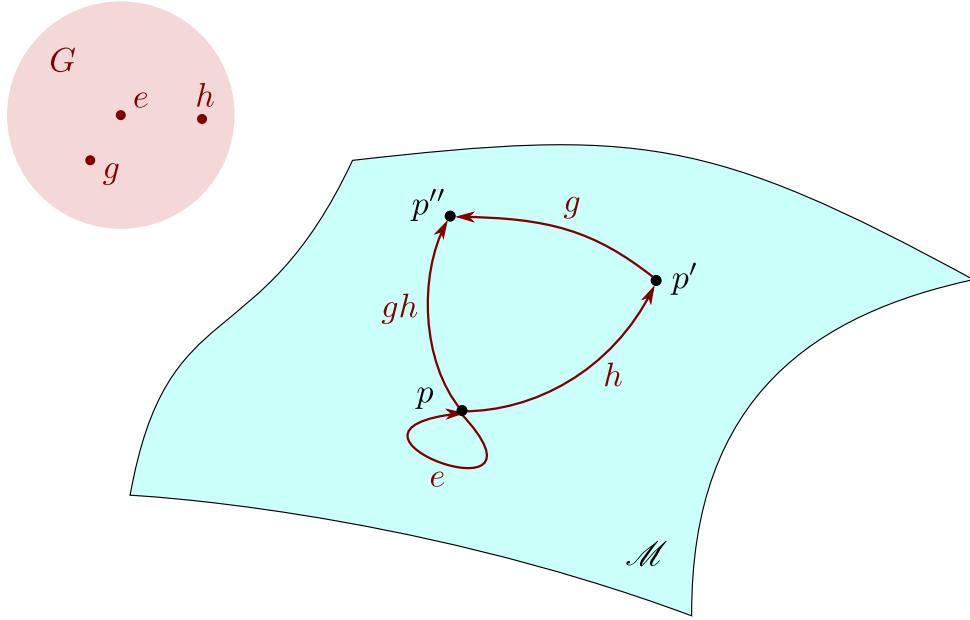
$${}^{\mathcal{H}}\nabla\ell(a, u) = \langle a, {}^{\mathcal{H}}\nabla_u \ell \rangle.$$

Let \bar{a} be the extension of a to a 1-form on \mathcal{M} defined by $\bar{a} = a \circ \Pi$, where Π is the projector onto \mathcal{H} along k [Eq. (2.76)]. Since ${}^{\mathcal{H}}\nabla_u \ell$ is tangent to \mathcal{H} , we have $\langle a, {}^{\mathcal{H}}\nabla_u \ell \rangle = \langle \bar{a}, {}^{\mathcal{H}}\nabla_u \ell \rangle$. In view of (3.10) and using Eq. (2.74) to express $\nabla\ell$, we then get

$$\begin{aligned} {}^{\mathcal{H}}\nabla\ell(a, u) &= \langle \bar{a}, \nabla_u \ell \rangle = \bar{a}_\mu u^\nu \nabla_\nu \ell^\mu = \bar{a}_\mu u^\nu \left(\underbrace{\Theta^\mu_\nu}_0 + \omega_\nu \ell^\mu - \ell_\nu k^\rho \nabla_\rho \ell^\mu \right) \\ &= \bar{a}_\mu \ell^\mu \omega_\nu u^\nu - \underbrace{\ell_\nu u^\nu}_{0} \bar{a}_\mu k^\rho \nabla_\rho \ell^\mu = \langle \bar{a}, \ell \rangle \langle \omega, u \rangle = \langle a, \ell \rangle \langle {}^{\mathcal{H}}\omega, u \rangle. \end{aligned}$$

Given the definition of a tensor product, this proves Eq. (3.11). □

A priori the 1-form ${}^{\mathcal{H}}\omega$ depends upon the choice of the cross-section \mathcal{S} of \mathcal{H} , via the vector k involved in Eq. (2.77): $\langle {}^{\mathcal{H}}\omega, v \rangle = -k \cdot \nabla_v \ell$ for any $v \in T_p \mathcal{H}$. Formula (3.11) shows that for a non-expanding horizon, this is not the case: ${}^{\mathcal{H}}\omega$ is a quantity intrinsic to \mathcal{H} and to the value of ℓ on \mathcal{H} . Moreover, under a change of null normal, $\ell \mapsto \ell' = \alpha \ell$, it remains constant up to the addition of an exact 1-form: ${}^{\mathcal{H}}\omega' = {}^{\mathcal{H}}\omega + d \ln \alpha$. The 1-form ${}^{\mathcal{H}}\omega$ is usually called the **connection 1-form** or **rotation 1-form** of (\mathcal{H}, ℓ) [23, 229]; the term *rotation* stems from the angular momentum of the non-expanding horizon being given by an integral involving ${}^{\mathcal{H}}\omega$, as we shall see in Chap. 18.

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

3.2.6 Going further

See Refs. [26, 229, 296] for more about non-expanding horizons, in particular for a subclass of them called *isolated horizons*.

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, such as stationarity and axisymmetry, are described in a coordinate-independent way by means of a group acting on the spacetime manifold \mathcal{M} . Through this action, each element of the group displaces points within \mathcal{M} and one demands that the metric g is invariant under such displacement. More precisely, given a group G , a **group action** of G on \mathcal{M} is a map²

$$\begin{aligned} \Phi : G \times \mathcal{M} &\longrightarrow \mathcal{M} \\ (g, p) &\longmapsto \Phi(g, p) =: \Phi_g(p) \end{aligned} \tag{3.12}$$

such that (cf. Fig. 3.2)

²Do not confuse the generic element g of group G with the metric tensor g .

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

The **orbit** of a point $p \in \mathcal{M}$ is the set $\{g(p), g \in G\} \subset \mathcal{M}$, i.e. the set of points which are connected to p by some transformation belonging to G . 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 so-called “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

$$\Phi_t := \Phi_{g_t}. \quad (3.13)$$

If G is a 1-dimensional Lie group, the orbit of a given point $p \in \mathcal{M}$ under the group action is then either $\{p\}$ (when p is a fixed point of the group action) or a curve \mathcal{L}_p of \mathcal{M} . In the latter case, t is a natural parameter along \mathcal{L}_p (cf. Fig. 3.3). The tangent vector ξ to \mathcal{L}_p corresponding to that parameter is called the **generator of the group** G associated with the t -parametrization of G . At each point of the orbit \mathcal{L}_p , it is given by (cf. Sec. A.2.3 and Fig. 3.3)

$$\xi = \left. \frac{dx}{dt} \right|_{\mathcal{L}_p}. \quad (3.14)$$

We can extend ξ to a vector field on all \mathcal{M} , by varying p over \mathcal{M} and setting $\xi = 0$ if p is a fixed point of the group action. For any $p \in \mathcal{M}$ and dt infinitesimal, the (infinitesimal) vector connecting p to $\Phi_{dt}(p)$ is then $dx = dt \xi$ (cf. Fig. 3.3), so that we may state:

Property 3.5: infinitesimal transformations as displacements along the generator

A 1-dimensional group action limited to infinitesimal transformations of parameter dt around the identity ($dt = 0$) amounts to displacements by the infinitesimal vector $dt \xi$, where ξ is the generator of the group.

Given a spacetime (\mathcal{M}, g) and a 1-dimensional Lie group G , one says that G is a **symmetry group** or a **isometry group** of (\mathcal{M}, g) , or equivalently that (\mathcal{M}, g) is **invariant under the action of** G , iff there is an action Φ of G on \mathcal{M} such that for any value of the parameter t of G , Φ_t is an **isometry** of (\mathcal{M}, g) , i.e. Φ_t preserves the “scalar products” (and hence the “distances”) in the following sense: for any $p \in \mathcal{M}$ and any pair of points (q, r) infinitely close to p , one has

$$g|_{\Phi_t(p)}(dx', dy') = g|_p(dx, dy), \quad (3.15)$$

for the infinitesimal displacement vectors $dx := \overrightarrow{pq}$, $dy := \overrightarrow{pr}$, $dx' := \overrightarrow{\Phi_t(p)\Phi_t(q)}$ and $dy' := \overrightarrow{\Phi_t(p)\Phi_t(r)}$ (cf. Sec. 1.2). Now, by definition, dx' is nothing but the pushforward of the vector $dx \in T_p \mathcal{M}$ to the tangent space $T_{\Phi_t(p)} \mathcal{M}$ by the map Φ_t (cf. Sec. A.2.8), and similarly dy' is the pushforward of dy by Φ_t :

$$dx' = \Phi_{t*}(dx) \quad \text{and} \quad dy' = \Phi_{t*}(dy).$$

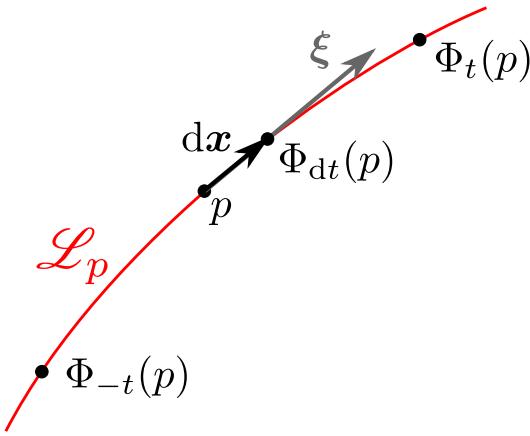


Figure 3.3: Orbit \mathcal{L}_p of a point p under the action Φ of a 1-dimensional Lie group, parameterized by $t \in \mathbb{R}$. The tangent vector $\xi = d\mathbf{x}/dt|_{\mathcal{L}_p}$ is the group generator associated with this parameter.

By rescaling by infinitely small parameters (using the bilinearity of \mathbf{g}), it is clear that (3.15) holds for finite vectors as well, so that we may say that Φ_t is an isometry of $(\mathcal{M}, \mathbf{g})$ iff

$$\forall p \in \mathcal{M}, \forall (\mathbf{u}, \mathbf{v}) \in (T_p \mathcal{M})^2, \quad \mathbf{g}|_{\Phi_t(p)}(\Phi_{t*}\mathbf{u}, \Phi_{t*}\mathbf{v}) = \mathbf{g}|_p(\mathbf{u}, \mathbf{v}), \quad (3.16)$$

where $\Phi_{t*}\mathbf{u}$ (resp. $\Phi_{t*}\mathbf{v}$) is the pushforward of the vector $\mathbf{u} \in T_p \mathcal{M}$ (resp. $\mathbf{v} \in T_p \mathcal{M}$) to the tangent space $T_{\Phi_t(p)} \mathcal{M}$ by Φ_t (cf. Sec. A.2.8). Given the definition (A.33) of the pullback of a multilinear form, we may reexpress the isometry condition (3.16) in terms of the pullback of \mathbf{g} by Φ_t :

$$\Phi_t^* \mathbf{g} = \mathbf{g}. \quad (3.17)$$

This is equivalent to the vanishing of the Lie derivative of \mathbf{g} along the generators of G :

Property 3.6: characterization of continuous spacetime isometries

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

$$\boxed{\mathcal{L}_\xi \mathbf{g} = 0}. \quad (3.18)$$

The vector field ξ is then called a **Killing vector** of $(\mathcal{M}, \mathbf{g})$, Eq. (3.18) being equivalent to the so-called **Killing equation**:

$$\boxed{\nabla_\alpha \xi_\beta + \nabla_\beta \xi_\alpha = 0}. \quad (3.19)$$

Proof. According the definition (A.82) of the Lie derivative, we have

$$\mathcal{L}_\xi \mathbf{g} := \lim_{t \rightarrow 0} \frac{1}{t} (\Phi_t^* \mathbf{g} - \mathbf{g}).$$

The isometry condition (3.17) implies then $\mathcal{L}_\xi \mathbf{g} = 0$. The reverse is true by integration. The equivalence between Eqs. (3.18) and (3.19) immediately results from expression (A.87) for the Lie derivative of \mathbf{g} . \square

Property 3.7: isometry in adapted coordinates

In terms of the components $g_{\alpha\beta}$ of \mathbf{g} with respect to coordinates $(x^\alpha) = (t, x^1, \dots, x^{n-1})$ adapted to the Killing vector ξ , i.e. such that $\xi = \partial_t$, the isometry condition (3.18) is equivalent to

$$\frac{\partial g_{\alpha\beta}}{\partial t} = 0. \quad (3.20)$$

t is then called an *ignorable coordinate*.

Proof. This is a direct consequence of the identity (A.85). \square

3.3.2 Definition and examples of Killing horizons

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

Thus the existence of a Killing horizon requires that the spacetime $(\mathcal{M}, \mathbf{g})$ 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. Since the normal to a null hypersurface is tangent to its null geodesic generators and a Killing vector is tangent to the orbits of the isometry group, a definition equivalent to the above one is:

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

It immediately follows from the definition that the Killing vector ξ is non-vanishing on \mathcal{H} (a normal to a hypersurface cannot vanish) and is null on \mathcal{H} :

$$\xi|_{\mathcal{H}} \neq 0 \quad \text{and} \quad \xi \cdot \xi|_{\mathcal{H}} = 0. \quad (3.21)$$

We shall see in Chap. 5 that in a stationary spacetime, under some rather generic hypotheses, a (connected part of 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 \mathcal{H} discussed in Examples 1, 5 and 8 of Chap. 2. \mathcal{H} is defined by the equation $t = x$. The vector field

$$\xi := \partial_t + \partial_x \quad (3.22)$$

is a Killing vector of Minkowski spacetime: ξ is the generator of translations in the direction $\partial_t + \partial_x$, and these translations constitute a 1-dimensional subgroup of the Poincaré group – the symmetry group of Minkowski spacetime. We note that ξ coincides with the null vector ℓ defined by Eq. (2.12). Since ℓ is normal to \mathcal{H} , we conclude immediately that \mathcal{H} is a Killing horizon with respect to ξ .

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

$$\xi := x\partial_t + t\partial_x. \quad (3.23)$$

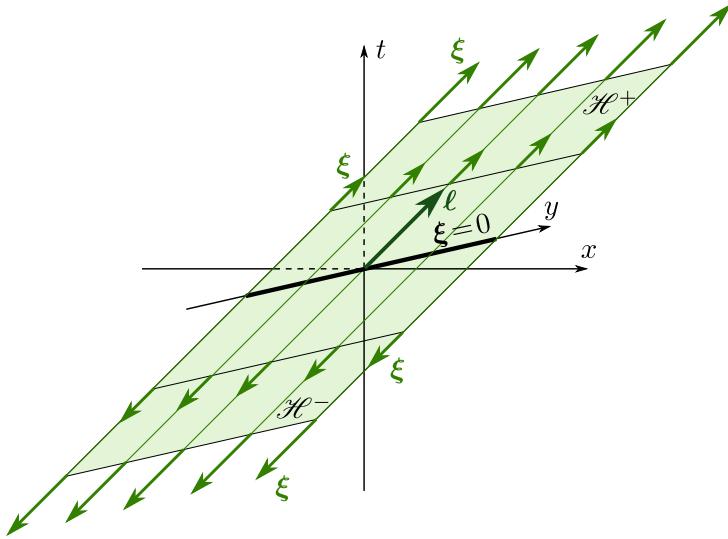


Figure 3.4: Null half-hyperplanes \mathcal{H}^+ and \mathcal{H}^- as Killing horizons for the Killing vector field $\xi = x\partial_t + t\partial_x$ generating Lorentz boosts in Minkowski spacetime. The green lines are the null geodesic generators of \mathcal{H} , while the thick black line (actually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

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

$$\xi \stackrel{\mathcal{H}}{=} t(\partial_t + \partial_x) \stackrel{\mathcal{H}}{=} t\ell,$$

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

$$\mathcal{H}^+ := \{p \in \mathcal{H}, \quad t(p) > 0\} \quad \text{and} \quad \mathcal{H}^- := \{p \in \mathcal{H}, \quad t(p) < 0\}, \quad (3.24)$$

so that each of them is a Killing horizon with respect to ξ (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 \mathcal{H} considered above, but this time with the Killing vector

$$\xi := y(\partial_t + \partial_x) + (t - x)\partial_y. \quad (3.25)$$

This vector is indeed the generator of null rotations leaving the plane $\text{Span}(\ell, \partial_z)$ strictly invariant (cf. e.g. Sec. 6.4.5 of Ref. [227]), ℓ being the null normal of \mathcal{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 (3.25) obeys Killing equation (3.19). On \mathcal{H} , $t - x = 0$, so that (3.25) reduces to

$$\xi \stackrel{\mathcal{H}}{=} y(\partial_t + \partial_x) \stackrel{\mathcal{H}}{=} y\ell.$$

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

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

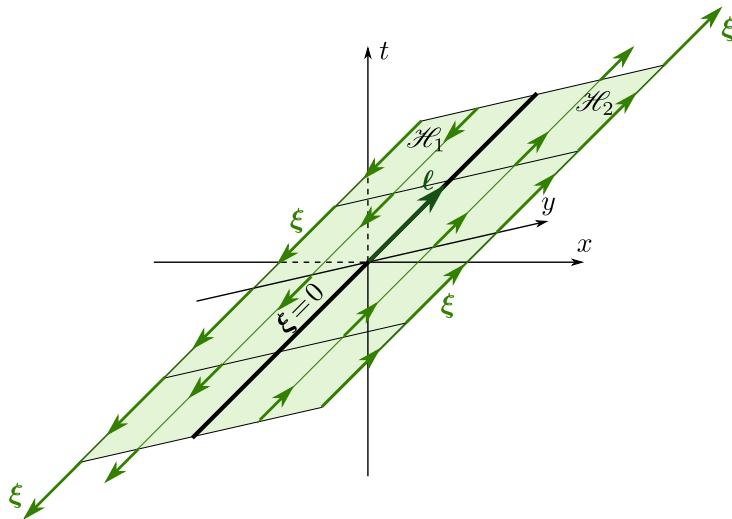


Figure 3.5: Null half-hyperplanes \mathcal{H}_1 and \mathcal{H}_2 as Killing horizons for the Killing vector field $\xi = y(\partial_t + \partial_x) + (t - x)\partial_y$ generating null rotations in Minkowski spacetime. The green lines are the null geodesic generators of \mathcal{H} , while the thick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

each of them being a Killing horizon with respect to ξ (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** [88, 91]: 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 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 ℓ of the family of hypersurfaces \mathcal{H}_u and the fact that the Schwarzschild horizon \mathcal{H} is defined by $r = 2m$, we have

$$\ell \stackrel{\mathcal{H}}{=} \partial_t. \quad (3.26)$$

Now the vector field ∂_t is clearly a Killing vector of metric g as given by (2.5), since none of the metric components $g_{\alpha\beta}$ depends upon t . Hence (3.26) shows that the Schwarzschild horizon is a Killing horizon. By the way, Eq. (3.26) was our motivation for the choice of the null normal ℓ performed in Example 7 of Chap. 2.

Historical note : The concept of Killing horizon has been introduced by Brandon Carter in 1966 [87, 88] and developed in an article published in 1969 [91]. 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 [70], leading to the concept of *bifurcate Killing horizon*, to be discussed in Sec. 3.4 (cf. the historical note on p. 87).

3.3.3 Killing horizons as non-expanding horizons

Let us consider a cross-section \mathcal{S} of a Killing horizon \mathcal{H} with respect to a Killing vector ξ and let us select the null normal ℓ to coincide with ξ on \mathcal{H} : $\ell \stackrel{\mathcal{H}}{=} \xi$. Since ξ is an isometry generator, it is pretty obvious that the metric q induced by g on \mathcal{S} will not evolve when Lie dragged along ξ and hence that the deformation rate Θ of \mathcal{S} along ξ , as defined by Eq. (2.70), vanishes identically. One can establish this rigorously from expression (2.73) for Θ . Indeed, without any loss of generality, we may express the null vector field ℓ in a neighborhood of \mathcal{H} as $\ell = \xi + uw$ where u is a scalar field defining \mathcal{H} as the level set $u = 0$, i.e. a scalar field obeying Eqs. (2.1)-(2.2). Then Eq. (2.73) leads to

$$\Theta_{\alpha\beta} = q^\mu_\alpha q^\nu_\beta \nabla_\mu \xi_\nu + \underbrace{u}_{0} q^\mu_\alpha q^\nu_\beta \nabla_\mu w_\nu + \underbrace{q^\mu_\alpha \nabla_\mu u}_{0} q^\nu_\beta w_\nu = q^\mu_\alpha q^\nu_\beta \nabla_\mu \xi_\nu,$$

where the identifications with zero hold on \mathcal{S} , the second one because u is constant (being equal to zero) on \mathcal{S} . Now thanks to the Killing equation (3.19), $q^\mu_\alpha q^\nu_\beta \nabla_\mu \xi_\nu$ is an antisymmetric tensor field, while $\Theta_{\alpha\beta}$ is symmetric (cf. Sec. 2.3.6). It follows that $\Theta_{\alpha\beta} = 0$. In view of the scaling law (2.84), we can extend this result to the deformation rate of \mathcal{S} along any null normal to \mathcal{H} and therefore state:

Property 3.8: vanishing of the deformation rate of cross-sections of a Killing horizon

On a Killing horizon \mathcal{H} , the deformation rate Θ of any cross-section along any null normal ℓ is identically zero:

$$\boxed{\Theta = 0}. \quad (3.27)$$

This implies that the expansion of \mathcal{H} along any null normal ℓ vanishes:

$$\theta_{(\ell)} = 0. \quad (3.28)$$

In view of the definition of a non-expanding horizon (cf. Sec. 3.2.1), an immediate corollary is

Property 3.9: Killing horizons as non-expanding horizons

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

Remark 1: Θ vanishes for any Killing horizon [Eq. (3.27)], while to get the same result on a generic non-expanding horizon, one has to assume that the null convergence condition holds on \mathcal{H} (Property 3.2).

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 ℓ coinciding with the Killing vector ξ on a Killing horizon \mathcal{H} . According to the definition (2.22), we have

$$\nabla_\xi \xi \stackrel{\mathcal{H}}{=} \kappa \xi. \quad (3.29)$$

Let us consider the metric dual of this relation, i.e. $\xi^\mu \nabla_\mu \xi_\alpha \stackrel{\mathcal{H}}{=} \kappa \xi_\alpha$, and use the Killing equation (3.19) as $\nabla_\mu \xi_\alpha = -\nabla_\alpha \xi_\mu$; we get

$$\xi^\mu \nabla_\alpha \xi_\mu \stackrel{\mathcal{H}}{=} -\kappa \xi_\alpha.$$

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

$$\nabla_\alpha (\xi_\mu \xi^\mu) \stackrel{\mathcal{H}}{=} -2\kappa \xi_\alpha. \quad (3.30)$$

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

$$\boxed{d(\underline{\xi} \cdot \underline{\xi}) \stackrel{\mathcal{H}}{=} -2\kappa \underline{\xi}}. \quad (3.31)$$

Remark 2: If $u := \underline{\xi} \cdot \underline{\xi}$ is a regular scalar field in the vicinity of \mathcal{H} , in the sense that $du \neq 0$, it is not surprising to have du proportional to $\underline{\xi}$ on \mathcal{H} . Indeed, if $du \neq 0$, the hypersurface \mathcal{H} can be considered as the level set $u = 0$ for $\underline{\xi}$ is null on \mathcal{H} (cf. Sec. 2.3.1). It follows that the gradient $\vec{\nabla} u$ must be collinear to the normal $\underline{\xi}$ to \mathcal{H} (cf. Sec. 2.3.2) or equivalently $du \stackrel{\mathcal{H}}{=} \alpha \underline{\xi}$. Equation (3.31) simply shows that the proportionality factor is $\alpha = -2\kappa$.

Another interesting relation is obtained from the Frobenius theorem applied to $\underline{\xi}$. Indeed, since on \mathcal{H} , $\underline{\xi}$ is normal to a hypersurface (\mathcal{H}), the Frobenius theorem in its dual formulation (see e.g. Theorem B.3.2 in Wald's textbook [498] or Theorem C.2 in Straumann's textbook [463]) states that there exists a 1-form a such that

$$d\underline{\xi} \stackrel{\mathcal{H}}{=} a \wedge \underline{\xi}, \quad (3.32)$$

or equivalently

$$\nabla_\alpha \xi_\beta - \nabla_\beta \xi_\alpha \stackrel{\mathcal{H}}{=} a_\alpha \xi_\beta - a_\beta \xi_\alpha. \quad (3.33)$$

Remark 3: In the case of the vector ℓ , which is normal to \mathcal{H} by definition, the Frobenius identity is Eq. (2.16): $\nabla_\alpha \ell_\beta - \nabla_\beta \ell_\alpha = \nabla_\alpha \rho \ell_\beta - \nabla_\beta \rho \ell_\alpha$. Since $\ell \stackrel{\mathcal{H}}{=} \underline{\xi}$, we may write

$$\nabla_\alpha \ell_\beta - \nabla_\beta \ell_\alpha \stackrel{\mathcal{H}}{=} \nabla_\alpha \rho \xi_\beta - \nabla_\beta \rho \xi_\alpha.$$

But in general, $\nabla_\alpha \ell_\beta \neq \nabla_\alpha \xi_\beta$ on \mathcal{H} , since ℓ and ξ do not coincide outside \mathcal{H} . Accordingly, one cannot identify the left-hand side of the above equation with the left-hand side of Eq. (3.33), so that the 1-form a is not $\nabla \rho$.

Thanks to the Killing equation (3.19), we may reshape (3.33) to

$$2\nabla_\alpha \xi_\beta \stackrel{\mathcal{H}}{=} a_\alpha \xi_\beta - a_\beta \xi_\alpha. \quad (3.34)$$

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

$$2\xi^\mu \nabla_\mu \xi_\alpha \stackrel{\mathcal{H}}{=} a_\mu \xi^\mu \xi_\alpha - \underbrace{\xi_\mu \xi^\mu}_{\stackrel{\mathcal{H}}{=} 0} a_\alpha.$$

In view of Eq. (3.29), the left-hand side of this equation is $2\kappa\xi_\alpha$. Hence we obtain

$$a_\mu\xi^\mu \stackrel{\mathcal{H}}{=} 2\kappa. \quad (3.35)$$

Besides, taking the square of (3.34) leads to

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

where we have used Eq. (3.35). Hence

$$\boxed{\kappa^2 \stackrel{\mathcal{H}}{=} -\frac{1}{2}\nabla_\mu\xi_\nu\nabla^\mu\xi^\nu}. \quad (3.36)$$

This is an explicit expression of κ in terms of the Killing vector field ξ . However, in actual calculations, it is generally preferable to employ formula (3.31) to evaluate κ , because it does not involve the computation of any covariant derivative, contrary to formula (3.36).

3.3.5 The zeroth law of black hole dynamics

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 ℓ the null normal to \mathcal{H} that coincides with the Killing vector field: $\ell \stackrel{\mathcal{H}}{=} \xi$. The vector field ℓ is then a symmetry generator on \mathcal{H} , which implies

$$\mathcal{L}_\ell\kappa = 0. \quad (3.37)$$

This means that κ is constant along any field line of ℓ (i.e. any null geodesic generator of \mathcal{H}). It could however vary from one field line to another. To prove that this is not the case, let us consider a complete cross-section \mathcal{S} of \mathcal{H} and show that κ is constant on \mathcal{S} . The starting point is applying the 1-form identity (2.85) to a generic vector field v tangent to \mathcal{S} , i.e. contract (2.85) with v^α . Since $\ell_\mu v^\mu = 0$, we get

$$v^\nu\nabla_\mu\Theta^\mu_\nu + v^\nu\ell^\mu\nabla_\mu\omega_\nu - v^\mu\nabla_\mu(\theta_{(\ell)} + \kappa) + (\theta_{(\ell)} + \kappa)\omega_\mu v^\mu - \Theta_{\mu\nu}v^\mu k^\sigma\nabla_\sigma\ell^\nu = R_{\mu\nu}\ell^\mu v^\nu. \quad (3.38)$$

Now, since \mathcal{H} is a Killing horizon, we have $\Theta = 0$ [Eq. (3.27)] and in particular $\theta_{(\ell)} = 0$, so that many terms in the above equation vanish. The term $\nabla_\mu\Theta^\mu_\nu$ requires some special care though, because it potentially involves derivatives of $\vec{\Theta}$ in directions transverse³ to \mathcal{H} . Let us then introduce in the vicinity of \mathcal{S} a spacetime coordinate system $(x^\alpha) = (u, v, x^2, \dots, x^{n-1})$

³Recall that, in our setting, tensor fields are extended beyond \mathcal{H} by considering \mathcal{H} as the element $u = 0$ of a family $(\mathcal{H}_u)_{u \in \mathbb{R}}$ of null hypersurfaces (cf. Sec. 2.3.2). We cannot assume that each \mathcal{H}_u is a Killing horizon (because typically the scalar $\xi \cdot \xi$ vanishes on a single hypersurface, not on an open subset of \mathcal{M}), so $\vec{\Theta}$ is a priori not zero outside \mathcal{H} .

adapted to \mathcal{S} , as the one introduced in Sec. 2.3.5 [cf. Eq. (2.52)], namely a coordinate system such that \mathcal{S} is the set $(u, v) = (0, 0)$. Then the components of $\vec{\Theta}$ with respect to (x^α) necessarily verify $\Theta^0{}_\alpha = 0$ and $\Theta^1{}_\alpha = 0$. Indeed, since $\vec{\Theta}$ is a tensor field tangent to \mathcal{S} , on which u and v are constant, we have $\Theta^\mu{}_\alpha \nabla_\mu u = 0$ and $\Theta^\mu{}_\alpha \nabla_\mu v = 0$ with $\nabla_\mu u = \partial_\mu u = \delta^0{}_\mu$ and $\nabla_\mu v = \partial_\mu v = \delta^1{}_\mu$ by definition of (x^α) . Expressing the covariant derivative via Eq. (A.64), we get

$$\nabla_\mu \Theta^\mu{}_\nu = \partial_a \Theta^a{}_\nu + \Gamma^\mu{}_{\mu\sigma} \Theta^\sigma{}_\nu - \Gamma^\sigma{}_{\mu\nu} \Theta^\mu{}_\sigma,$$

where the sum of the partial derivatives has been limited to the indices $a \in \{2, \dots, n-1\}$ since $\Theta^0{}_\nu = \Theta^1{}_\nu = 0$. Given that the term $\partial_a \Theta^a{}_\nu$ involves only the variation of $\vec{\Theta}$ in directions tangent to \mathcal{S} , we conclude that $\nabla_\mu \Theta^\mu{}_\nu = 0$ if $\vec{\Theta} = 0$ on \mathcal{S} . Hence, for a Killing horizon, Eq. (3.38) reduces to

$$(\ell^\nu \nabla_\nu \omega_\mu + \kappa \omega_\mu) v^\mu - v^\mu \nabla_\mu \kappa = R_{\mu\nu} \ell^\mu v^\nu. \quad (3.39)$$

The terms in the parentheses are related to the Lie derivative of ω along ℓ ; indeed formula (A.86) gives:

$$\begin{aligned} \mathcal{L}_\ell \omega_\mu &= \ell^\nu \nabla_\nu \omega_\mu + \omega_\nu \nabla_\mu \ell^\nu = \ell^\nu \nabla_\nu \omega_\mu + \omega_\nu (\Theta_\mu{}^\nu + \omega_\mu \ell^\nu - \ell_\mu k^\sigma \nabla_\sigma \ell^\nu) \\ &= \ell^\nu \nabla_\nu \omega_\mu + \kappa \omega_\mu - \omega_\nu \ell_\mu k^\sigma \nabla_\sigma \ell^\nu, \end{aligned}$$

where we have expressed $\nabla_\mu \ell^\nu$ via Eq. (2.74) and have used $\Theta_\mu{}^\nu = 0$ and $\omega_\nu \ell^\nu = \kappa$ [Eq. (2.78)]. Given that $\ell_\mu v^\mu = 0$, Eq. (3.39) becomes

$$\langle \mathcal{L}_\ell \omega, v \rangle - \nabla_v \kappa = R(\ell, v). \quad (3.40)$$

It is judicious to express $\mathcal{L}_\ell \omega$ in terms of $\mathcal{L}_\ell \mathcal{H}\omega$, where $\mathcal{H}\omega$ is the restriction of ω to tangent vectors to \mathcal{H} (pullback to \mathcal{H}) introduced in Sec. 3.2.5. The reason is that, contrary to ω , $\mathcal{H}\omega$ is a geometric quantity intrinsic to \mathcal{H} and ℓ , independent of k and hence of the cross-section \mathcal{S} (cf. Property 3.4 and the discussion below it). It has therefore to obey the spacetime symmetry generated by $\xi \stackrel{\mathcal{H}}{=} \ell$, i.e. one has

$$\mathcal{L}_\ell \mathcal{H}\omega = 0. \quad (3.41)$$

Now, using the Leibniz rule twice, we get

$$\langle \mathcal{L}_\ell \omega, v \rangle = \mathcal{L}_\ell \underbrace{\langle \omega, v \rangle}_{\langle \mathcal{H}\omega, v \rangle} - \underbrace{\langle \omega, \mathcal{L}_\ell v \rangle}_{\langle \mathcal{H}\omega, \mathcal{L}_\ell v \rangle} = \langle \mathcal{L}_\ell \mathcal{H}\omega, v \rangle,$$

where the identities $\langle \omega, v \rangle = \langle \mathcal{H}\omega, v \rangle$ and $\langle \omega, \mathcal{L}_\ell v \rangle = \langle \mathcal{H}\omega, \mathcal{L}_\ell v \rangle$ hold by the very definition of $\mathcal{H}\omega$, since v and $\mathcal{L}_\ell v$ are tangent⁴ to \mathcal{H} . It follows then from (3.41) that $\langle \mathcal{L}_\ell \omega, v \rangle = 0$; hence the first term in Eq. (3.40) vanishes identically and there remains only:

$$\nabla_v \kappa = -R(\ell, v). \quad (3.42)$$

⁴ $\mathcal{L}_\ell v$ is tangent to \mathcal{H} for both ℓ and v are tangent to \mathcal{H} .

To go further, we shall assume the **null dominance condition** [420], namely that there exists a scalar field f such that

the vector $\mathbf{W} := -\vec{\mathbf{G}}(\ell) - f\ell$ is zero or future-directed (null or timelike)
for any future-directed null vector ℓ

(3.43)

In the above equation, $\vec{\mathbf{G}}$ stands for the type-(1, 1) tensor associated by metric duality to the Einstein tensor \mathbf{G} [Eq. (A.111)], so that the expression of \mathbf{W} in index notation is [cf. Eq. (A.50a)]

$$W^\alpha := -G_\mu^\alpha \ell^\mu - f \ell^\alpha. \quad (3.44)$$

Note that the null dominance condition implies the null convergence condition (2.94) since

$$\mathbf{R}(\ell, \ell) = \mathbf{R}(\ell, \ell) - \frac{R}{2} \underbrace{\mathbf{g}(\ell, \ell)}_0 + f \underbrace{\mathbf{g}(\ell, \ell)}_0 = \mathbf{G}(\ell, \ell) + f \mathbf{g}(\ell, \ell) = -\mathbf{W} \cdot \ell \geq 0, \quad (3.45)$$

the inequality holding because both \mathbf{W} and ℓ are future-directed (cf. Lemma 1.2).

If gravity is described by general relativity, i.e. if the metric \mathbf{g} fulfills the Einstein equation (1.41), then the null dominance condition with $f = \Lambda$ is equivalent to the **null dominant energy condition**:

The vector $\mathbf{W} := -\vec{\mathbf{T}}(\ell)$ is zero or future-directed (null or timelike)
for any future-directed null vector ℓ

, (3.46)

GR

where \mathbf{T} is the energy-momentum tensor of matter and non-gravitational fields. By continuity, the null dominant energy condition is implied by the standard **dominant energy condition**:

The vector $\mathbf{W} := -\vec{\mathbf{T}}(\mathbf{u})$ is zero or future-directed (null or timelike)
for any future-directed timelike vector \mathbf{u}

. (3.47)

Physically, the dominant energy condition states that, with respect to any observer (represented by its 4-velocity \mathbf{u} , which is future-directed timelike), the energy of matter and non-gravitational fields does not move faster than light (see Ref. [103] for an extended discussion).

Remark 4: While the name *null convergence condition* for $\mathbf{R}(\ell, \ell) \geq 0$ [Eq. (2.94)] is standard in the literature (e.g. [265, 451, 453]), the name *null dominance condition* for (3.43) is not standard. We are using it to distinguish from the *null dominant energy condition* (3.46), which a condition on the matter energy-momentum tensor, while (3.43) is a pure geometrical identity, independent of the Einstein equation. In this way, *null dominance condition* is on the same footing as *null convergence condition*.

Coming back to Eq. (3.42), we note that its right-hand side is nothing but the scalar product of the vector \mathbf{W} defined by Eq. (3.43) with \mathbf{v} :

$$\mathbf{W} \cdot \mathbf{v} = -(\vec{\mathbf{G}}(\ell) + f\ell) \cdot \mathbf{v} = -\mathbf{R}(\ell, \mathbf{v}) + \left(\frac{R}{2} - f \right) \underbrace{\ell \cdot \mathbf{v}}_0 = -\mathbf{R}(\ell, \mathbf{v}).$$

If we assume the null dominance condition, the null convergence condition holds, so that $R(\ell, \ell) = 0$ on \mathcal{H} [Eq. (3.8)]. Then, according to Eq. (3.45), $\ell \cdot W = -R(\ell, \ell) = 0$. This implies that the vector W is tangent to \mathcal{H} . The latter being a null hypersurface, W must then be either collinear to ℓ or spacelike (cf. Lemma 2.3 in Sec. 2.3.4). Now, according to the null dominance condition (3.43), W cannot be spacelike. We conclude that W is collinear to ℓ . Consequently, we have $W \cdot v = 0$. Hence the right-hand side of Eq. (3.42) vanishes identically and we are left with $\nabla_v \kappa = 0$. Since v is a generic vector field tangent to \mathcal{S} and \mathcal{S} is connected (for it is a complete cross-section of a Killing horizon, which is connected by definition), this implies that κ is constant over \mathcal{S} . Given that κ is constant along each null geodesic generator of \mathcal{H} , this completes the demonstration that κ is constant over \mathcal{H} . More precisely, we have established the following property:

Property 3.10: zeroth law of black hole dynamics

If the null dominance condition (3.43) is fulfilled on a Killing horizon \mathcal{H} — which is guaranteed in general relativity if the null dominant energy condition (3.46) holds —, then the non-affinity coefficient κ of the null normal coinciding with the Killing vector ξ defining \mathcal{H} is constant over \mathcal{H} :

$$\kappa = \text{const.} \quad (3.48)$$

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 (3.48) is known as the **zeroth law of black hole dynamics**. More precisely, the zeroth law — to be discussed in detail in Chap. 16 — states that the surface gravity of a black hole in equilibrium is constant and we shall see in Chap. 5 that (any connected part of) the event horizon of a black hole in equilibrium is a Killing horizon.

Remark 5: The standard proof of the zeroth law, based on taking the covariant derivative of Eq. (3.29) [39, 100, 498] or of Eq. (3.31) [274] and on an identity expressing the second order derivative of a Killing vector in terms of the Riemann tensor [Eq. (3.83) below], is quite long (see e.g. pp. 333-334 of Wald's textbook [498] and the remark at the end of Sec. 5.5.2 of Poisson's textbook [415]). The proof presented above is shorter but it relies on the concept of induced affine connection on a non-expanding horizon (Sec. 3.2.5) and the associated rotation 1-form $\mathcal{H}\omega$ (cf. Property 3.4). This proof is actually adapted from a proof presented by Damour [150, 151] (cf. historical note below); see also Ref. [22] for a related proof regarding isolated horizons.

Remark 6: The constancy of κ on a Killing horizon can also be proved without the null dominance condition, but at the price of additional hypotheses: either the Killing horizon \mathcal{H} is part of a so-called bifurcate Killing horizon, as we shall see in Sec. 3.4.3 (Property 3.16) or the spacetime is axisymmetric, in addition to be stationary, and the two Killing vectors associated with stationarity and axisymmetry are orthogonal to $(n - 2)$ -dimensional surfaces [95, 421].

Example 11 (null hyperplane as a translation-Killing horizon): For the null hyperplane \mathcal{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, which is obviously constant over \mathcal{H} .

Example 12 (null hyperplane as a boost-Killing horizon): Let us consider each of the null half-hyperplanes \mathcal{H}^+ and \mathcal{H}^- of Example 7, which are Killing horizons with respect to the boost Killing vector $\xi = x\partial_t + t\partial_x$. On \mathcal{H}^+ , the future-directed null normal coinciding with this Killing vector is $\ell^+ = t\ell$, ℓ being the geodesic null normal defined by $\ell := \partial_t + \partial_x$ [cf. Eq. (2.12)]. Using $\kappa_\ell = 0$ and the scaling law (2.26), we get the non-affinity coefficient of ℓ^+ as $\kappa_+ = \nabla_\ell t = \partial_t t + \partial_x t$, i.e.

$$\kappa_+ = 1.$$

On \mathcal{H}^- , ξ is past-directed (cf. Fig. 3.4). Sticking to future-directed null normals, we shall then consider \mathcal{H}^- as a Killing horizon with respect to the Killing vector field $-\xi$. The future-directed null normal coinciding with $-\xi$ on \mathcal{H}^- is then $\ell^- = -t\ell$, from which we deduce the non-affinity coefficient of ℓ^- : $\kappa_- = \nabla_\ell(-t) = \partial_t(-t) + \partial_x(-t)$, i.e.

$$\kappa_- = -1.$$

We check that κ_+ (resp. κ_-) is constant over the Killing horizon \mathcal{H}^+ (resp. \mathcal{H}^-), in agreement with Property 3.10.

Example 13 (null hyperplane as a null-rotation-Killing horizon): In Example 8, we have introduced the Killing horizons \mathcal{H}_1 and \mathcal{H}_2 with respect to the null-rotation Killing vector $\xi = y(\partial_t + \partial_x) + (t-x)\partial_y$ of Minkowski spacetime. On \mathcal{H}_1 , ξ is past-directed (cf. Fig. 3.5), so that we shall actually consider \mathcal{H}_1 as a Killing horizon with respect to the Killing vector field $-\xi$. The future-directed null normal coinciding with $-\xi$ on \mathcal{H}_1 is then $\ell_1 = -y\ell$. Since it is clearly constant along the null geodesic generators of \mathcal{H}_1 , we have $\nabla_{\ell_1}\ell_1 = 0$, hence the associated non-affinity coefficient vanishes: $\kappa_1 = 0$. On \mathcal{H}_2 , ξ is future-directed (cf. Fig. 3.5) and the null normal coinciding with it is $\ell_2 = y\ell$, 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. (2.29)] that on a Schwarzschild horizon $\kappa = 1/(4m)$, 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: $\kappa = 1/(4m)$.

Example 15 (Cubic Galileon black hole as a counter-example): It has been found recently that the surface gravity of rotating stationary black holes in a scalar-tensor theory of gravity known as the *cubic Galileon* is not constant [240]. This evades the zeroth law (Property 3.10) because the null dominance condition is not satisfied by these solutions.

Historical note : The constancy of κ for a Killing horizon has been proven by Stephen Hawking in his lecture at the famous 1972 Les Houches Summer School [260] (p. 43). It has also been proven without requiring the null dominance condition, but assuming axisymmetry and orthogonal transitivity (cf. Remark 6 on p. 74) by Brandon Carter in his lecture at the same summer school [95] (Theorem 8, p. 167). A third proof of the constancy of κ , using the null dominance 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* [39]. Yet another proof has been provided in 1979 by Thibault Damour [150, 151], who developed a fluid-bubble approach to the dynamics of an event horizon. In Damour's framework, the pullback of the 1-form $-\omega/(8\pi)$ to the cross-section \mathcal{S} is considered as the momentum surface density of the fluid bubble and Eq. (3.38) is turned into a 2-dimensional Navier-Stokes equation (see e.g. Sec. 6.3 of Ref. [229] for details), where $\kappa/(8\pi)$ plays the role of the bubble's surface pressure, so that it must be constant in equilibrium. As mentioned in Remark 5, the proof presented above is derived from Damour's one.

3.3.6 Classification of Killing horizons

Since κ is constant on a Killing horizon \mathcal{H} (assuming the null dominance condition), we may use it to classify Killing horizons in two categories, depending whether κ vanishes or not:

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

Example 16 (Killing horizons in Minkowski spacetime): 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.

Example 17 (Schwarzschild and Kerr horizons): 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.

The next example regards the anti-de Sitter spacetime and will play some role in the study of the extremal ($a = m$) Kerr black hole in Chap. 13.

Example 18 (Poincaré horizon in AdS₄): The 4-dimensional **anti-de Sitter spacetime** (AdS₄) is (\mathcal{M}, g) with $\mathcal{M} \simeq \mathbb{R}^4$ and g is the metric whose components in the so-called *global static coordinates* $(\tau, r, \theta, \varphi)$ are given by

$$g = \ell^2 \left[-(1 + r^2) d\tau^2 + \frac{dr^2}{1 + r^2} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right], \quad (3.49)$$

where ℓ is a positive constant. Note that τ spans \mathbb{R} , r spans $(0, +\infty)$, while (θ, φ) are standard spherical coordinates on \mathbb{S}^2 : $\theta \in (0, \pi)$ and $\varphi \in (0, 2\pi)$. The metric (3.49) is a solution to the vacuum Einstein equation (1.43) with the negative cosmological constant $\Lambda = -3/\ell^2$. Using the so-called *conformal coordinates* $(\tau, \chi, \theta, \varphi)$ with $\chi := \arctan r \in (0, \pi/2)$, one gets

$$g = \frac{\ell^2}{\cos^2 \chi} [-d\tau^2 + d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2)]. \quad (3.50)$$

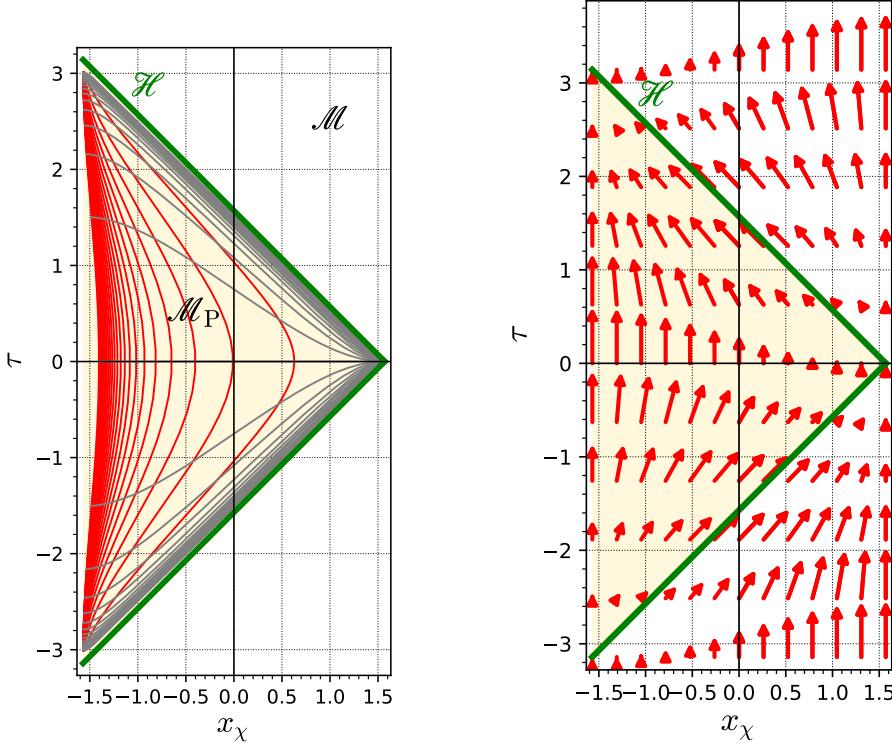


Figure 3.6: *Left:* 2-dimensional slice $(x, y) = (0, 0)$ of the Poincaré patch \mathcal{M}_P of anti-de Sitter spacetime plotted in terms of the (global) coordinates τ and $x_\chi := \chi \cos \varphi$ (pale yellow region). From Eq. (3.53), $(x, y) = (0, 0)$ implies $\theta = \pi/2$ and $\varphi = 0$ or π , so that $x_\chi = \chi$ in the right half of the plot ($\varphi = 0$) and $x_\chi = -\chi$ in the left half ($\varphi = \pi$). The plotted slice of \mathcal{M}_P is thus spanned by the Poincaré coordinates (t, u) . The red lines are curves of constant u , i.e. integral curves of the coordinate vector field $\partial_t = \xi$, with u increasing from 0 to $+\infty$ from the right to the left of the diagram. The grey lines are curves of curves of constant t , i.e. integral curves of the coordinate vector field ∂_u , with t increasing from $-\infty$ to $+\infty$ from the bottom to the top of the diagram. The Poincaré horizon \mathcal{H} is depicted in green. The two connected components \mathcal{H}_- and \mathcal{H}_+ of \mathcal{H} appear as straight line segments, since for $\theta = \pi/2$ and $\varphi \in \{0, \pi\}$, $u = 0 \iff \cos \tau = \pm \sin \chi$ (+ for $\varphi = 0$ and - for $\varphi = \pi$). Note that the curves of constant u tend to \mathcal{H} when $u \rightarrow 0$, in agreement with the characterization of \mathcal{H} by $u = 0$. The curves of constant t tend to \mathcal{H} when $t \rightarrow \pm\infty$, which is expected as well since the first line of Eq. (3.53) and Eq. (3.52) imply that $u \rightarrow 0$ for $t \rightarrow \pm\infty$. *Right:* Killing vector field ξ defined by Eq. (3.58) on AdS_4 ; ξ is timelike everywhere, except on the Poincaré horizon \mathcal{H} , where it is null (and normal, and thus tangent to \mathcal{H}). [Figures generated by the notebook D.3.1]

Note that (χ, θ, φ) spans one half⁵ of the hypersphere \mathbb{S}^3 . Yet another set of coordinates commonly used in AdS_4 is **Poincaré coordinates** (t, x, y, u) . They cover only a subpart \mathcal{M}_P of \mathcal{M} (cf. Fig. 3.6), usually called the **Poincaré patch** and defined by

$$\mathcal{M}_P : \quad u > 0 \quad \text{and} \quad -\pi < \tau < \pi, \quad (3.51)$$

where u is the following scalar field on \mathcal{M} :

$$u := \frac{\ell(\cos \tau - \sin \chi \sin \theta \cos \varphi)}{\cos \chi}. \quad (3.52)$$

On \mathcal{M}_P , the Poincaré coordinates (t, x, y, u) are related to the conformal coordinates $(\tau, \chi, \theta, \varphi)$ by⁶

$$\left\{ \begin{array}{l} t = \frac{\ell \sin \tau}{\cos \tau - \sin \chi \sin \theta \cos \varphi} \\ x = \frac{\ell \sin \chi \sin \theta \sin \varphi}{\cos \tau - \sin \chi \sin \theta \cos \varphi} \\ y = \frac{\ell \sin \chi \cos \theta}{\cos \tau - \sin \chi \sin \theta \cos \varphi} \\ u = \frac{\ell(\cos \tau - \sin \chi \sin \theta \cos \varphi)}{\cos \chi}. \end{array} \right. \quad (3.53)$$

The last line is simply (3.52) restricted to \mathcal{M}_P , the scalar field u being viewed there as one of the Poincaré coordinates. The metric components with respect to Poincaré coordinates⁷ are (cf. the notebook D.3.1 for the computation):

$$\mathbf{g} = \frac{u^2}{\ell^2} (-dt^2 + dx^2 + dy^2) + \frac{\ell^2}{u^2} du^2. \quad (3.54)$$

The **Poincaré horizon** is the hypersurface \mathcal{H} bounding the Poincaré patch \mathcal{M}_P in \mathcal{M} . In view of the definition (3.51) of the latter, \mathcal{H} appears to be the level set $u = 0$:

$$\mathcal{H} : \quad u = 0 \quad \text{and} \quad -\pi < \tau < \pi, \quad (3.55)$$

Note that \mathcal{H} is not included in \mathcal{M}_P , so that the Poincaré coordinates are not defined on \mathcal{H} (except for u , they actually diverge in the vicinity of \mathcal{H}). Note also that \mathcal{H} has two connected components: \mathcal{H}_- , where $\tau \in (-\pi, 0)$, and \mathcal{H}_+ , where $\tau \in (0, \pi)$, since $u = 0$ cannot be achieved for $\tau = 0$, as a consequence of formula (3.52) and $|\sin \chi| < 1$ on \mathcal{M} . The Poincaré horizon is depicted in Fig. 3.6. Its normal is given by the gradient of u , so that the vector field defined in all \mathcal{M} by $\mathbf{k} := \vec{\nabla}u$ is normal to \mathcal{H} on \mathcal{H} . Given expressions (3.50) and (3.52) for respectively \mathbf{g} and u , the components $k^\alpha = g^{\alpha\mu}\partial_\mu u$ of \mathbf{k} with respect to conformal coordinates are

$$\mathbf{k} = \frac{1}{\ell} \left[\sin \tau \cos \chi \partial_\tau + (\cos \tau \sin \chi - \sin \theta \cos \varphi) \partial_\chi - \frac{\cos \theta \cos \varphi}{\tan \chi} \partial_\theta + \frac{\sin \varphi}{\tan \chi \sin \theta} \partial_\varphi \right]. \quad (3.56)$$

The scalar square of \mathbf{k} is $\mathbf{k} \cdot \mathbf{k} = k_\mu k^\mu = k^\mu \partial_\mu u$; we obtain

$$\mathbf{k} \cdot \mathbf{k} = \frac{u^2}{\ell^2}. \quad (3.57)$$

⁵It would span the whole hypersphere if χ would run in all of $(0, \pi)$, instead of being limited to $(0, \pi/2)$.

⁶See e.g. Ref. [44].

⁷The name *Poincaré coordinates* stems from a variant of these coordinates obtained by using $z := \ell^2/u$ instead of u , so that $\mathbf{g} = \ell^2 (-dt^2 + dx^2 + dy^2 + dz^2)/z^2$, which is similar to the metric of the Poincaré half-space model of the hyperbolic space \mathbb{H}^4 , except for the signature $(-, +, +, +)$ instead of $(+, +, +, +)$.

Hence $\mathbf{k} \cdot \mathbf{k} \stackrel{\mathcal{H}}{=} 0$, which implies that \mathcal{H} is a null hypersurface.

Since the metric components (3.54) do not depend on t , the vector field $\xi := \partial_t$ is a Killing vector of (\mathcal{M}_P, g) . By inverting the Jacobian matrix associated with the change of coordinates (3.53), we get the expression of $\xi = \partial_t$ in terms of conformal coordinates:

$$\xi = \frac{1 - \cos \tau \sin \chi \sin \theta \cos \varphi}{\ell} \partial_\tau - \frac{\sin \tau \cos \chi \sin \theta \cos \varphi}{\ell} \partial_\chi - \frac{\sin \tau \cos \theta \cos \varphi}{\ell \sin \chi} \partial_\theta + \frac{\sin \tau \sin \varphi}{\ell \sin \chi \sin \theta} \partial_\varphi. \quad (3.58)$$

A priori, ξ is defined on \mathcal{M}_P only, but the right-hand side of the above expression is regular on all \mathcal{M} . Hence, we may use (3.58) to define ξ as a vector field on all \mathcal{M} . It is depicted in the right panel of Fig. 3.6. By analytical continuation, it is immediate that ξ obeys the Killing equation (3.19) everywhere and not only on \mathcal{M}_P (cf. the notebook D.3.1 for an explicit check). Hence ξ is a Killing vector of the entire anti-de Sitter spacetime (\mathcal{M}, g) . By comparing Eqs. (3.56) and (3.58), we get

$$\xi = \frac{\sin \tau}{\cos \chi} \mathbf{k} + \frac{u}{\ell^2} (\cos \tau \cos \chi \partial_\tau - \sin \tau \sin \chi \partial_\chi). \quad (3.59)$$

Since $u \stackrel{\mathcal{H}}{=} 0$ [Eq. (3.55)], it follows immediately that, on \mathcal{H} , ξ is collinear to the null normal to \mathcal{H} : $\xi \stackrel{\mathcal{H}}{=} (\sin \tau / \cos \chi) \mathbf{k}$, with $\sin \tau \neq 0$ (given that $\tau \neq 0$ on \mathcal{H}). Hence, ξ is normal to the null hypersurface \mathcal{H} . We therefore conclude that each of the connected components \mathcal{H}_- and \mathcal{H}_+ of the Poincaré horizon \mathcal{H} is a Killing horizon with respect to ξ . Let us evaluate the non-affinity coefficient κ of ξ on \mathcal{H}_\pm via formula (3.31). First of all, we compute the scalar square of ξ by noticing that on \mathcal{M}_P , $\xi = \partial_t$, so that $\xi \cdot \xi = g_{tt}$; with g_{tt} read on Eq. (3.54), we get

$$\xi \cdot \xi = -\frac{u^2}{\ell^2}. \quad (3.60)$$

By means of the global components (3.50) and (3.58) of respectively g and ξ , one checks that formula (3.60) holds in all \mathcal{M} . Given that the right-hand side is negative wherever $u \neq 0$, it follows that ξ is timelike everywhere on \mathcal{M} , except on \mathcal{H} . Furthermore, formula (3.60) results in $d(\xi \cdot \xi) = -2\ell^{-2}u du$. Since $u \stackrel{\mathcal{H}}{=} 0$, this implies $d(\xi \cdot \xi) \stackrel{\mathcal{H}}{=} 0$, so that formula (3.31) yields $\kappa = 0$. We conclude that the two connected components of the Poincaré horizon of AdS_4 are degenerate Killing horizons.

3.3.7 Interpretation of κ as a “surface gravity”

In this section, we assume that \mathcal{H} is a non-degenerate Killing horizon, i.e. that $\kappa \neq 0$. Let $p \in \mathcal{H}$ and $v \in T_p \mathcal{M}$ be a vector *transverse* to \mathcal{H} , i.e. not tangent to \mathcal{H} . According to Eq. (3.31), we have

$$\nabla_v(\xi \cdot \xi) = -2\kappa \xi \cdot v.$$

The right-hand side of this expression does not vanish, because $\kappa \neq 0$ and $\xi \cdot v \neq 0$ (since v is not tangent to \mathcal{H}). Hence we get $\nabla_v(\xi \cdot \xi) \neq 0$. In other words, the derivative of the scalar square $\xi \cdot \xi$ along any direction transverse to \mathcal{H} does not vanish. Since $\xi \cdot \xi = 0$ on \mathcal{H} , we conclude that, in the vicinity of \mathcal{H} , $\xi \cdot \xi < 0$ on one side of \mathcal{H} and $\xi \cdot \xi > 0$ on the other side:

Property 3.11

In the vicinity of a non-degenerate Killing horizon \mathcal{H} , the Killing vector field ξ is timelike on one side of \mathcal{H} , null on \mathcal{H} and spacelike on the other side.

Let us focus on the side of \mathcal{H} where ξ is timelike. There we define the “norm” of ξ by

$$V := \sqrt{-\xi \cdot \xi}. \quad (3.61)$$

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

$$\kappa^2 = \lim_{\mathcal{H}} \nabla_\mu V \nabla^\mu V, \quad (3.62)$$

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

Proof. Let us consider the **twist 3-form** ω defined by

$$\omega := \underline{\xi} \wedge d\underline{\xi} \quad (3.63)$$

or, using index notation,

$$\begin{aligned} \omega_{\alpha\beta\gamma} &:= \xi_\alpha(d\xi)_{\beta\gamma} + \xi_\beta(d\xi)_{\gamma\alpha} + \xi_\gamma(d\xi)_{\alpha\beta} \\ &= \xi_\alpha(\nabla_\beta\xi_\gamma - \nabla_\gamma\xi_\beta) + \xi_\beta(\nabla_\gamma\xi_\alpha - \nabla_\alpha\xi_\gamma) + \xi_\gamma(\nabla_\alpha\xi_\beta - \nabla_\beta\xi_\alpha). \end{aligned} \quad (3.64)$$

Killing equation (3.19) enables us to simplify each term inside parentheses in (3.64), yielding

$$\omega_{\alpha\beta\gamma} = 2(\xi_\alpha\nabla_\beta\xi_\gamma + \xi_\beta\nabla_\gamma\xi_\alpha + \xi_\gamma\nabla_\alpha\xi_\beta). \quad (3.65)$$

The “square” of ω is then

$$\begin{aligned} \omega_{\mu\nu\rho}\omega^{\mu\nu\rho} &= 4 \left(\xi_\mu\nabla_\nu\xi_\rho\xi^\mu\nabla^\nu\xi^\rho + \xi_\mu\nabla_\nu\xi_\rho\xi^\nu\nabla^\rho\xi^\mu + \xi_\mu\nabla_\nu\xi_\rho\xi^\rho\nabla^\mu\xi^\nu \right. \\ &\quad + \xi_\nu\nabla_\rho\xi_\mu\xi^\mu\nabla^\nu\xi^\rho + \xi_\nu\nabla_\rho\xi_\mu\xi^\nu\nabla^\rho\xi^\mu + \xi_\nu\nabla_\rho\xi_\mu\xi^\rho\nabla^\mu\xi^\nu \\ &\quad \left. + \xi_\rho\nabla_\mu\xi_\nu\xi^\mu\nabla^\nu\xi^\rho + \xi_\rho\nabla_\mu\xi_\nu\xi^\nu\nabla^\rho\xi^\mu + \xi_\rho\nabla_\mu\xi_\nu\xi^\rho\nabla^\mu\xi^\nu \right). \end{aligned}$$

Now in the first line,

$$\xi_\mu\nabla_\nu\xi_\rho\xi^\mu\nabla^\nu\xi^\rho = \xi_\mu\xi^\mu\nabla_\nu\xi_\rho\nabla^\nu\xi^\rho = -V^2\nabla_\nu\xi_\rho\nabla^\nu\xi^\rho = -V^2\nabla_\mu\xi_\nu\nabla^\mu\xi^\nu \quad (3.66)$$

and (using Killing equation (3.19))

$$\xi_\mu\nabla_\nu\xi_\rho\xi^\nu\nabla^\rho\xi^\mu = \xi_\mu\nabla^\rho\xi^\mu\xi^\nu\nabla_\nu\xi_\rho = -\xi_\mu\nabla^\rho\xi^\mu\xi^\nu\nabla_\rho\xi_\nu = -\frac{1}{4}\nabla^\rho V^2\nabla_\rho V^2 = -V^2\nabla^\rho V\nabla_\rho V. \quad (3.67)$$

Actually, we notice that each line is made of one term of type (3.66) and two terms of type (3.67). Hence

$$\omega_{\mu\nu\rho} \omega^{\mu\nu\rho} = -12V^2 (\nabla_\mu \xi_\nu \nabla^\mu \xi^\nu + 2\nabla_\mu V \nabla^\mu V). \quad (3.68)$$

On \mathcal{H} , each of the terms inside parentheses in Eq. (3.64) can be expressed thanks to the Frobenius identity (3.33):

$$\omega_{\alpha\beta\gamma} \stackrel{\mathcal{H}}{=} \xi_\alpha (a_\beta \xi_\gamma - a_\gamma \xi_\beta) + \xi_\beta (a_\gamma \xi_\alpha - a_\alpha \xi_\gamma) + \xi_\gamma (a_\alpha \xi_\beta - a_\beta \xi_\alpha).$$

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

$$\omega_{\alpha\beta\gamma} \stackrel{\mathcal{H}}{=} 0. \quad (3.69)$$

Equation (3.69) is actually nothing but a variant of Frobenius theorem, expressing the fact that the vector field ξ is hypersurface-orthogonal on \mathcal{H} (see e.g. Eq. (B.3.6) in Wald's textbook [498], taking into account that $\omega_{\alpha\beta\gamma} = 6\xi_{[\alpha} \nabla_{\beta} \xi_{\gamma]}$). Let us evaluate the gradient of the square (3.68) and take the limit on \mathcal{H} :

$$\begin{aligned} \nabla_\alpha \omega_{\mu\nu\rho} \underbrace{\omega^{\mu\nu\rho}}_{\rightarrow 0} + \underbrace{\omega_{\mu\nu\rho}}_{\rightarrow 0} \nabla_\alpha \omega^{\mu\nu\rho} &= -12 \underbrace{\nabla_\alpha V^2}_{\rightarrow 2\kappa \xi_\alpha} \left(\underbrace{\nabla_\mu \xi_\nu \nabla^\mu \xi^\nu}_{\rightarrow -2\kappa^2} + 2\nabla_\mu V \nabla^\mu V \right) \\ &\quad -12 \underbrace{V^2}_{\rightarrow 0} \nabla_\alpha (\nabla_\mu \xi_\nu \nabla^\mu \xi^\nu + 2\nabla_\mu V \nabla^\mu V), \end{aligned}$$

where we have used Eq. (3.31) in the form $\nabla_\alpha V^2 \stackrel{\mathcal{H}}{=} 2\kappa \xi_\alpha$, as well as expression (3.36) of κ^2 . Hence we are left with

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

Now, by the very definition of a Killing horizon, $\xi_\alpha \neq 0$ on \mathcal{H} . Moreover, \mathcal{H} being a non-degenerate Killing horizon, we have $\kappa \neq 0$ as well. The above limit is then equivalent to (3.62). \square

In the region where ξ is timelike, the vector field

$$\mathbf{u} := \frac{1}{V} \xi \quad (3.70)$$

is a future-directed unit timelike vector field. It is future-directed because by convention⁸ ξ is future-directed null on \mathcal{H} and by continuity this orientation must be preserved in the region where ξ is timelike. The unit vector field \mathbf{u} can be then considered as the 4-velocity of an observer \mathcal{O} , whose worldline is a field line of ξ , i.e. an orbit of the isometry group generated by ξ . One may call \mathcal{O} a **stationary observer** since the spacetime geometry is not changing along its worldline. The 4-acceleration of \mathcal{O} is

$$\begin{aligned} \mathbf{a} &:= \nabla_{\mathbf{u}} \mathbf{u} \\ &= \nabla_{V^{-1}\xi} (V^{-1}\xi) = V^{-1} \nabla_\xi (V^{-1}\xi) = V^{-1} [-V^{-2}(\nabla_\xi V) \xi + V^{-1} \nabla_\xi \xi]. \end{aligned}$$

⁸Were ξ past-directed, we could always consider the Killing field $-\xi$ instead.

Now, since ξ is a symmetry generator, $\nabla_\xi V = 0$. This can be shown explicitly by means of Killing equation (3.19):

$$\nabla_\xi V = \xi^\mu \nabla_\mu (\sqrt{-\xi_\nu \xi^\nu}) = -\frac{1}{2\sqrt{-\xi_\nu \xi^\nu}} \xi^\mu \nabla_\mu (\xi_\nu \xi^\nu) = -\frac{1}{V} \underbrace{\xi^\mu \xi^\nu \nabla_\mu \xi_\nu}_0 = 0.$$

We have thus

$$\mathbf{a} = \frac{1}{V^2} \nabla_\xi \xi. \quad (3.71)$$

Thanks to Killing equation (3.19), we may rewrite this relation as

$$a_\alpha = \frac{1}{V^2} \xi^\mu \nabla_\mu \xi_\alpha = -\frac{1}{V^2} \xi^\mu \nabla_\alpha \xi_\mu = -\frac{1}{2V^2} \nabla_\alpha (\xi_\mu \xi^\mu) = \frac{1}{2V^2} \nabla_\alpha V^2 = \nabla_\alpha \ln V,$$

hence

$$\mathbf{a} = \vec{\nabla} \ln V. \quad (3.72)$$

The norm of \mathbf{a} , which is always a spacelike vector (since $\mathbf{u} \cdot \mathbf{u} = -1$ implies $\mathbf{u} \cdot \mathbf{a} = 0$), is

$$a := \sqrt{\mathbf{a} \cdot \mathbf{a}} = \frac{1}{V} \sqrt{\nabla_\mu V \nabla^\mu V}. \quad (3.73)$$

Given the result (3.62), we get an expression of κ involving a :

$$\boxed{\kappa = \lim_{\mathcal{O} \rightarrow \mathcal{H}} V a}, \quad (3.74)$$

where $\mathcal{O} \rightarrow \mathcal{H}$ means that the limit is achieved by choosing the worldline of observer \mathcal{O} arbitrarily close to \mathcal{H} . Since $V \rightarrow 0$ as one approaches \mathcal{H} , it follows that

$$\lim_{\mathcal{O} \rightarrow \mathcal{H}} a = +\infty. \quad (3.75)$$

This means that the acceleration felt by observer \mathcal{O} (the “gravity”) diverges as \mathcal{O} approaches \mathcal{H} . In that sense, the *physical* surface gravity of \mathcal{H} is infinite. But Eq. (3.74) shows that the rescaled acceleration $V a$ remains finite as one approaches \mathcal{H} , and tends to κ . It is this quantity that is named the **surface gravity** of the Killing horizon \mathcal{H} .

Remark 7: 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 \mathcal{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 ξ is timelike in the entire region outside the Killing horizon⁹. In this case, one can identify κ with the magnitude of the force exerted by an observer “at infinity” to hold in place a particle of unit mass close to \mathcal{H} by means of an infinitely long massless string (see e.g. Sec. 5.2.4 of Poisson’s textbook [415]).

⁹This is not true for a rotating Kerr black hole: ξ becomes null at some “light-cylinder” outside \mathcal{H} and is then spacelike away from it, cf. Eq. (10.68), where ξ is denoted by χ .

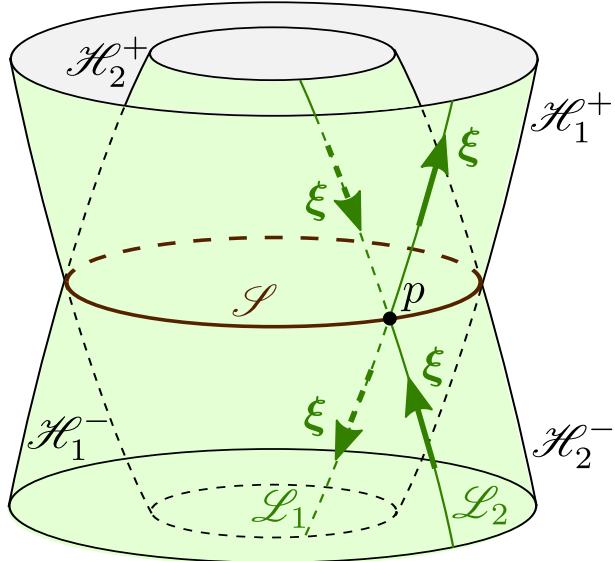


Figure 3.7: Bifurcate Killing horizon $\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2$ with respect to the Killing vector field ξ ; \mathcal{S} is the bifurcation surface. \mathcal{L}_1 and \mathcal{L}_2 are null geodesic generators of respectively \mathcal{H}_1 and \mathcal{H}_2 , which cross each other at the point $p \in \mathcal{S}$.

3.4 Bifurcate Killing horizons

We extend here the study of Killing horizons by introducing the concept of bifurcate Killing horizon, which is particularly important for black hole physics.

3.4.1 Definition and first properties

Let (\mathcal{M}, g) be a n -dimensional spacetime endowed with a Killing vector field ξ . A **bifurcate Killing horizon** is the union

$$\mathcal{H} = \mathcal{H}_1 \cup \mathcal{H}_2, \quad (3.76)$$

such that

- \mathcal{H}_1 and \mathcal{H}_2 are two null hypersurfaces;
- $\mathcal{S} := \mathcal{H}_1 \cap \mathcal{H}_2$ is a spacelike $(n - 2)$ -surface;
- each of the sets $\mathcal{H}_1 \setminus \mathcal{S}$ and $\mathcal{H}_2 \setminus \mathcal{S}$ has two connected components, which are Killing horizons with respect to ξ .

The $(n - 2)$ -dimensional submanifold \mathcal{S} is called the **bifurcation surface** of \mathcal{H} .

Hence we may say that a bifurcate Killing horizon is formed by four Killing horizons, \mathcal{H}_1^+ , \mathcal{H}_1^- , \mathcal{H}_2^+ and \mathcal{H}_2^- say, which are glued together at the bifurcation surface \mathcal{S} (cf. Fig. 3.7), in such a way that

$$\mathcal{H}_1 = \mathcal{H}_1^- \cup \mathcal{S} \cup \mathcal{H}_1^+ \quad \text{and} \quad \mathcal{H}_2 = \mathcal{H}_2^- \cup \mathcal{S} \cup \mathcal{H}_2^+$$

are null hypersurfaces.

A first property of bifurcate Killing horizons is

Property 3.12: vanishing of the Killing vector on the bifurcation surface

The Killing vector field vanishes on the bifurcation surface of a bifurcate Killing horizon:

$$\xi|_{\mathcal{S}} = 0. \quad (3.77)$$

Proof. Let $p \in \mathcal{S}$ and let us assume that $\xi|_p \neq 0$. Let \mathcal{L}_1 (resp. \mathcal{L}_2) be the null geodesic generator of \mathcal{H}_1 (resp. \mathcal{H}_2) that intersects \mathcal{S} at p (cf. Fig. 3.7). By definition of a Killing horizon, ξ is tangent to $\mathcal{L}_1 \cap \mathcal{H}_1^+$ and to $\mathcal{L}_1 \cap \mathcal{H}_1^-$, i.e. to $\mathcal{L}_1 \setminus \{p\}$. If $\xi|_p \neq 0$, then by continuity, ξ is a (non-vanishing) tangent vector field all along \mathcal{L}_1 . Similarly, ξ is tangent to all \mathcal{L}_2 . At their intersection point p , the geodesics \mathcal{L}_1 and \mathcal{L}_2 have thus a common tangent vector, namely $\xi|_p$. The geodesic uniqueness theorem (Property B.10 in Appendix B) then implies $\mathcal{L}_1 = \mathcal{L}_2$, so that $\mathcal{L}_1 \subset \mathcal{H}_1 \cap \mathcal{H}_2 = \mathcal{S}$. But since \mathcal{S} is spacelike and \mathcal{L}_1 is null, we reach a contradiction. Hence we must have $\xi|_p = 0$. \square

Remark 1: Having a Killing vector field that vanishes somewhere (here \mathcal{S}) is not the sign of any pathology: it simply means that the points of \mathcal{S} are fixed points of the isometries generated by ξ , since setting $\xi = 0$ in Eq. (3.14) leads to $dx = 0$, i.e. to $\Phi_{dt}(p) = p$.

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 \mathcal{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. (3.21)], while on a bifurcate Killing horizon, it is vanishing at the bifurcation surface.

Example 19 (Lorentz-boost bifurcate Killing horizon): Let us consider Minkowski spacetime and the Killing vector given by Eq. (3.23): $\xi := x\partial_t + t\partial_x$, namely the generator of Lorentz boosts in the (t, x) plane. Let us take for \mathcal{H}_1 the null hyperplane $t = x$ considered in Example 7 and denoted there by \mathcal{H} . The two half-hyperplanes defined by Eq. (3.24) are then the Killing horizons \mathcal{H}_1^+ and \mathcal{H}_1^- . The union $\mathcal{H}_1 \cup \mathcal{H}_2$, where \mathcal{H}_2 is the null hyperplane $t = -x$ is a bifurcate Killing horizon with respect to ξ , with the 2-plane $(t, x) = (0, 0)$ as bifurcation surface (cf. Fig. 3.8). Note that on \mathcal{H}_1 , the Killing vector ξ points away from \mathcal{S} , while on \mathcal{H}_2 , it points towards \mathcal{S} , as in the generic figure 3.7.

3.4.2 Non-degenerate Killing horizons and Boyer's theorem

Let us consider a Killing horizon \mathcal{H} with respect to some Killing vector ξ , such that the surface gravity κ is constant over \mathcal{H} . According to the zeroth law (Property 3.10), this is guaranteed if the null dominance condition is fulfilled. In what follows, we focus on the case where $\kappa \neq 0$, i.e. \mathcal{H} is a non-degenerate Killing horizon (cf. Sec. 3.3.6). In this case, the parameter t of the null geodesic generators of \mathcal{H} associated to ξ is not affine (for κ is the non-affinity coefficient of ξ). However, we may rescale ξ to get an affine parameter, at the price of loosing the Killing vector feature:

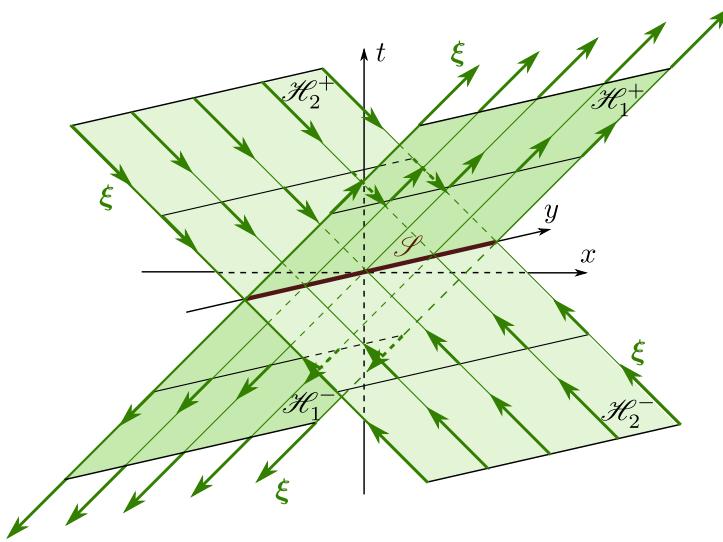


Figure 3.8: Bifurcate Killing horizon $\mathcal{H}_1 \cup \mathcal{H}_2$ with respect to the Killing vector $\xi = x\partial_t + t\partial_x$ generating Lorentz boosts in the plane (t, x) of Minkowski spacetime. The dimension along z having been suppressed, the bifurcation surface \mathcal{S} appears as a line, while it is actually a 2-plane.

Property 3.13: affine parametrization of a non-degenerate Killing horizon

Let \mathcal{H} be a Killing horizon (with respect to a Killing vector ξ) of constant nonzero surface gravity κ . Let t be the parameter of the null geodesic generators \mathcal{L} of \mathcal{H} associated to ξ ($\xi = dx/dt$ along \mathcal{L}). The null vector field ℓ defined on \mathcal{H} by

$$\ell = e^{-\kappa t} \xi \iff \xi = e^{\kappa t} \ell \quad (3.78)$$

is a geodesic vector field and the affine parameter associated to it is

$$\lambda = \frac{e^{\kappa t}}{\kappa} + \lambda_0, \quad (3.79)$$

where λ_0 is constant along a given geodesic generator.

Proof. We have

$$\nabla_\ell \ell = \nabla_{e^{-\kappa t} \xi} (e^{-\kappa t} \xi) = e^{-\kappa t} \nabla_\xi (e^{-\kappa t} \xi) = e^{-\kappa t} \left[\underbrace{(\nabla_\xi e^{-\kappa t})}_{de^{-\kappa t}/dt} \xi + e^{-\kappa t} \underbrace{\nabla_\xi \xi}_{\kappa \xi} \right] = 0.$$

Hence ℓ is a geodesic vector. Besides, along any null generator of \mathcal{H} , one has [cf. Eq. (A.8)]

$$\frac{d\lambda}{dt} = \xi(\lambda) = e^{\kappa t} \underbrace{\ell(\lambda)}_1 = e^{\kappa t},$$

which, once integrated, yields Eq. (3.79). □

Let us assume $\kappa > 0$ and consider a null geodesic generator \mathcal{L} of \mathcal{H} ; \mathcal{L} can be parameterized by t , the corresponding tangent vector being ξ . When t spans the whole interval $(-\infty, +\infty)$, Eq. (3.79) implies that λ spans the interval $(\lambda_0, +\infty)$ only. Since λ is an affine parameter of \mathcal{L} , this means that \mathcal{L} is an *incomplete* geodesic (cf. Sec. B.3.2). Moreover, Eq. (3.78) leads to

$$\xi \rightarrow 0 \quad \text{when} \quad t \rightarrow -\infty \quad (\kappa > 0). \quad (3.80)$$

In other words, the Killing vector field ξ vanishes and the null geodesic \mathcal{L} stops at the “edge” of \mathcal{H} corresponding to $t \rightarrow -\infty$. If there is no obstacle (spacetime singularity or spacetime edge¹⁰), \mathcal{L} can be extended to $\lambda \in (-\infty, \lambda_0]$, giving rise to a complete null geodesic $\tilde{\mathcal{L}}$. This operation can be performed for all the null geodesic generators of \mathcal{H} and we have the freedom to choose the same value of λ_0 in Eq. (3.79) for all of them. In this process, one gives birth to a null hypersurface, $\tilde{\mathcal{H}}$ say, which contains \mathcal{H} . Let $\mathcal{S} \subset \tilde{\mathcal{H}}$ be the set of points of affine parameter $\lambda = \lambda_0$ along all the extended null geodesics $\tilde{\mathcal{L}}$. \mathcal{S} is clearly a cross-section of $\tilde{\mathcal{H}}$ (cf. Sec. 2.3.4); it is then a spacelike $(n - 2)$ -dimensional surface. Assuming that ξ is future-directed on \mathcal{H} , \mathcal{S} constitutes the past boundary of \mathcal{H} , i.e. the boundary corresponding to $t \rightarrow -\infty$. Since ξ is a smooth vector field on \mathcal{M} , Eq. (3.80) implies that ξ vanishes on \mathcal{S} . In other words, \mathcal{S} is a set of fixed points for the isometry group generated by ξ (cf. Remark 1 above). Let us denote by \mathcal{H}^- the subset of $\tilde{\mathcal{H}}$ generated by the segments $\lambda < \lambda_0$ of the null geodesics $\tilde{\mathcal{L}}$: $\mathcal{H}^- = \tilde{\mathcal{H}} \setminus (\mathcal{H} \cup \mathcal{S})$. \mathcal{H}^- is clearly a null hypersurface. Since \mathcal{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 ℓ . 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 behavior of a Killing vector field around the set of its fixed points (here \mathcal{S}), Boyer [70] has shown that in the current setting (i.e. \mathcal{S} spacelike), ξ 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 19 (cf. Fig. 3.8). More precisely, Boyer proved the following:

Property 3.14: Boyer’s theorem (Boyer 1969 [70])

A Killing horizon \mathcal{H} with respect to a Killing vector ξ is contained in a bifurcate Killing horizon if, and only if, \mathcal{H} contains at least one null geodesic orbit of the isometry group that is complete as an orbit, i.e. such that the group parameter t associated to ξ takes all values in \mathbb{R} , but that is incomplete and extendable as a geodesic.

That \mathcal{H} contains an orbit that is an incomplete geodesic is guaranteed by $\kappa \neq 0$, as we have just 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. (3.79) that while t spans the whole interval $(-\infty, +\infty)$, the affine parameter λ spans the interval $(-\infty, \lambda_0)$ only. Moreover, Eq. (3.78) leads to

$$\xi \rightarrow 0 \quad \text{when} \quad t \rightarrow +\infty \quad (\kappa < 0). \quad (3.81)$$

¹⁰A spacetime edge is generally not a genuine obstacle for extending an incomplete geodesic: it simply indicates that the spacetime itself needs to be extended so that the geodesic becomes complete, cf. Remark 3 below.

The reasoning developed for $\kappa > 0$ can be 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:

Property 3.15: non-degenerate Killing horizons and bifurcate Killing horizons

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: It could be that the generators of \mathcal{H} cannot be extended simply because the spacetime (\mathcal{M}, g) is not been chosen “large enough”. I. Rácz and R.M. Wald [421] have shown that under some mild hypotheses (in particular that (\mathcal{M}, g) is globally hyperbolic), one can extend (\mathcal{M}, g) to a larger spacetime with a bifurcate Killing horizon containing \mathcal{H} .

Remark 4: 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 $\xi \rightarrow 0$ 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, p. 67 and 75 respectively (cf. Fig. 3.5): $\xi = 0$ 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 who had just been 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 a bifurcate Killing horizon and the proof of Property 3.14, have been turned into an article by Jürgen Ehlers and John Stachel, which has been published in 1969 [70].

3.4.3 Zeroth law for bifurcate Killing horizons

For a Killing horizon that is part of a bifurcate Killing horizon, one can establish the constancy of the surface gravity κ (the “zeroth law”) by means of the vanishing of the Killing vector ξ at the bifurcation surface, without requiring the null dominance condition as for a generic Killing horizon in Property 3.10:

Property 3.16: zeroth law for bifurcate Killing horizons

The surface gravity is a nonzero constant over any Killing horizon that is part of a bifurcate Killing horizon:

$$\kappa = \text{const} \neq 0. \quad (3.82)$$

The proof makes use of the following lemma:

Lemma 3.17: second derivative of a Killing vector

For any Killing vector ξ , the following identity, often called the **Kostant formula**, holds:

$$\nabla_\gamma \nabla_\beta \xi^\alpha = R^\alpha_{\beta\gamma\mu} \xi^\mu, \quad (3.83)$$

where $R^\alpha_{\beta\gamma\mu}$ is the Riemann curvature tensor of \mathbf{g} . The contraction on the indices α and γ lets the Ricci tensor appear:

$$\nabla_\mu \nabla_\alpha \xi^\mu = R_{\alpha\mu} \xi^\mu. \quad (3.84)$$

Proof. Consider the covariant version of the Ricci identity (A.98): $\nabla_\alpha \nabla_\beta \xi_\gamma - \nabla_\beta \nabla_\alpha \xi_\gamma = R_{\gamma\mu\alpha\beta} \xi^\mu$ and call it equation (1). Perform two successive cyclic permutations of the indices (α, β, γ) , leading to equations (2) and (3), say. Then consider the linear combination (1) – (2) – (3). The left-hand side of it reduces to $2\nabla_\gamma \nabla_\beta \xi_\alpha$ thanks to the Killing equation (3.19), while the right-hand side is $(R_{\gamma\mu\alpha\beta} - R_{\alpha\mu\beta\gamma} - R_{\beta\mu\gamma\alpha}) \xi^\mu$. Now, using the antisymmetries (A.100) and (A.103) of the Riemann tensor and then the cyclic symmetry (A.101), one has $-R_{\alpha\mu\beta\gamma} - R_{\beta\mu\gamma\alpha} = -R_{\mu\alpha\gamma\beta} - R_{\mu\beta\alpha\gamma} = R_{\mu\gamma\beta\alpha}$. This turns the right-hand side into $(R_{\gamma\mu\alpha\beta} + R_{\mu\gamma\beta\alpha}) \xi^\mu = 2R_{\gamma\mu\alpha\beta} \xi^\mu = 2R_{\alpha\beta\gamma\mu} \xi^\mu$, the last equality resulting from (A.104). This establishes Eq. (3.83). Finally, let us note that Eq. (3.84) can be proved directly, without appealing to the cyclic property of the Riemann tensor: it suffices to use the contracted Ricci identity (A.107): $\nabla_\mu \nabla_\alpha \xi^\mu - \nabla_\alpha \nabla_\mu \xi^\mu = R_{\mu\alpha} \xi^\mu = R_{\alpha\mu} \xi^\mu$ and the fact that $\nabla_\mu \xi^\mu = 0$ as a consequence of Killing equation (3.19). \square

Proof of Property 3.16. Let \mathcal{H} be a Killing horizon that is part of a bifurcate Killing horizon $\hat{\mathcal{H}}$ with respect to a Killing vector ξ . Let v be a generic non-vanishing vector field tangent to \mathcal{H} . Let us evaluate the derivative of κ along v from expression (3.36):

$$2\kappa \nabla_v \kappa = -\frac{1}{2} v^\rho \nabla_\rho (\nabla_\mu \xi_\nu \nabla^\mu \xi^\nu) = -v^\rho \nabla_\rho \nabla_\mu \xi_\nu \nabla^\mu \xi^\nu.$$

Using Eq. (3.83) and the antisymmetry property (A.103) to set $-R_{\nu\mu\rho\sigma} = R_{\mu\nu\rho\sigma}$, we get

$$2\kappa \nabla_v \kappa = v^\rho R_{\mu\nu\rho\sigma} \xi^\sigma \nabla^\mu \xi^\nu. \quad (3.85)$$

If the vector v is chosen to be the null normal $\ell \stackrel{\mathcal{H}}{=} \xi$, one gets immediately $2\kappa \nabla_\ell \kappa = 0$, given that $R_{\mu\nu\rho\sigma} \xi^\rho \xi^\sigma = 0$ [Eq. (A.100)]. Then $\kappa = 0$ or $\nabla_\ell \kappa = 0$. In both cases, this means that κ must be constant along a given null generator \mathcal{L} of \mathcal{H} . We thus recover the result (3.37). There remains to prove that κ does not vary from one generator to the other. For this, we shall use that \mathcal{H} belongs to the bifurcate Killing horizon $\hat{\mathcal{H}}$. By virtue of Boyer's theorem (Property 3.14), one has $\kappa \neq 0$, for if κ would be zero along \mathcal{L} , the latter would be a complete geodesic and \mathcal{H} could not be part of a bifurcate Killing horizon. The bifurcation surface \mathcal{S} of $\hat{\mathcal{H}}$ is part of the topological closure of \mathcal{H} (cf. Fig. 3.7, with e.g. \mathcal{H}_1^+ standing for \mathcal{H}), so it is meaningful to consider the limit of Eq. (3.85) to \mathcal{S} for a vector field v on \mathcal{H} that tends to a nonzero vector field tangent to \mathcal{S} . Given that $\xi \stackrel{\mathcal{S}}{=} 0$ and $\nabla \xi$ remains finite at \mathcal{S} for ξ is a smooth vector field on \mathcal{M} , the limit yields (taking into account $\kappa \neq 0$) $\nabla_v \kappa \stackrel{\mathcal{S}}{=} 0$. It follows that

κ (or more precisely the limit to \mathcal{S} of \mathcal{H} 's surface gravity κ) is uniform over \mathcal{S} . Each null generator \mathcal{L} of \mathcal{H} is an incomplete geodesic, which once extended to a complete geodesic, intersects \mathcal{S} at single point. The constancy of κ on \mathcal{S} implies then that κ cannot vary from one null generator to the other. κ is thus uniform over \mathcal{H} . \square

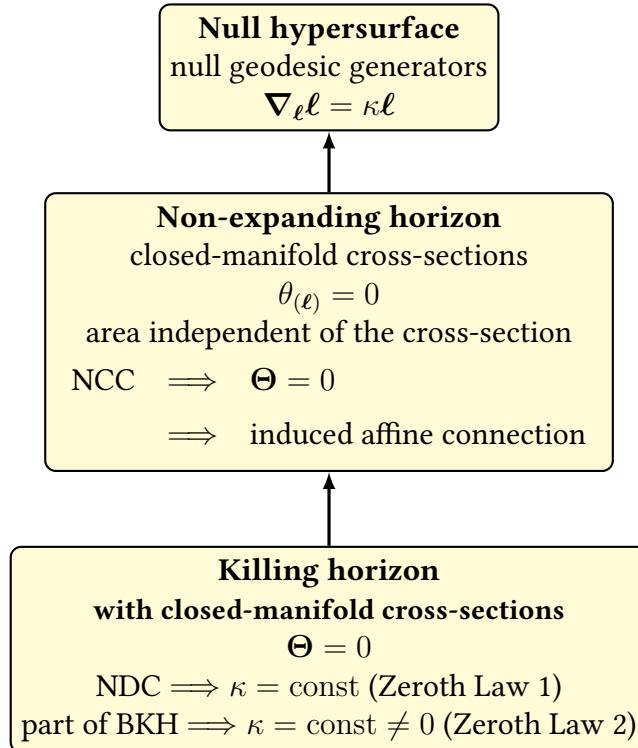
Example 20 (Lorentz-boost bifurcate Killing horizon): Let us consider the bifurcate Killing horizon $\hat{\mathcal{H}}$ of Minkowski spacetime discussed in Example 19 and illustrated in Fig. 3.8. The corresponding Killing vector is the Lorentz-boost generator $\xi := x\partial_t + t\partial_x$. An easy computation yields $\nabla_\xi \xi = t\partial_t + x\partial_x$. On the components \mathcal{H}_1^+ and \mathcal{H}_1^- of $\hat{\mathcal{H}}$ (cf. Fig. 3.8), one has $t = x$, so that $\xi = t(\partial_t + \partial_x)$ and $\nabla_\xi \xi = t(\partial_t + \partial_x)$. Hence $\nabla_\xi \xi = \xi$ and $\kappa = 1$ there. On the components \mathcal{H}_2^+ and \mathcal{H}_2^- of $\hat{\mathcal{H}}$ (cf. Fig. 3.8), one has $t = -x$, so that $\xi = t(-\partial_t + \partial_x)$ and $\nabla_\xi \xi = t(\partial_t - \partial_x)$. Hence $\nabla_\xi \xi = -\xi$ and $\kappa = -1$ there. We conclude that κ is constant over any of the four Killing horizons constituting $\hat{\mathcal{H}}$.

Remark 5: As the above example shows, the surface gravity is *not* uniform over all the bifurcate Killing horizon $\hat{\mathcal{H}}$; it keeps the same value only on the subsets $\mathcal{H}_1^- \cup \mathcal{H}_1^+$ and $\mathcal{H}_2^- \cup \mathcal{H}_2^+$.

Historical note : The proof of the zeroth law for Killing horizons that are part of a bifurcate Killing horizon (Property 3.16) has been first presented by Bernard Kay and Robert Wald in 1991 [309]; these authors mentioned that the proof had been suggested to them by Robert Geroch. The proof appears also in the notes from a lecture given by Wald at a school held in Erice in May 1991 [500]. It is also presented in Poisson's textbook [415].

3.5 Summary

Here is an inheritance diagram summarizing the main results of this chapter. The vertical arrow means “is a”, i.e. the item at the bottom of an arrow is a special case of the item at the top of the arrow. “NCC” stands for “Null Convergence Condition” [Eq. (2.94)], “NDC” for “Null Dominance Condition” [Eq. (3.43)] and “BKH” for “Bifurcate Killing Horizon” (Sec. 3.4.1).



Chapter 4

The concept of black hole 3: The global view

Contents

4.1	Introduction	91
4.2	Conformal completion of Minkowski spacetime	92
4.3	Conformal completions and asymptotic flatness	100
4.4	Black holes	105

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 particle has escaped from a given region, one has to wait until the “end of time” to make sure that the particle will never be back...

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 signal 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 [402, 403] (see Refs. [197] and [381] for a review). We start by investigating the conformal completion of the simplest of all spacetimes: Minkowski spacetime.

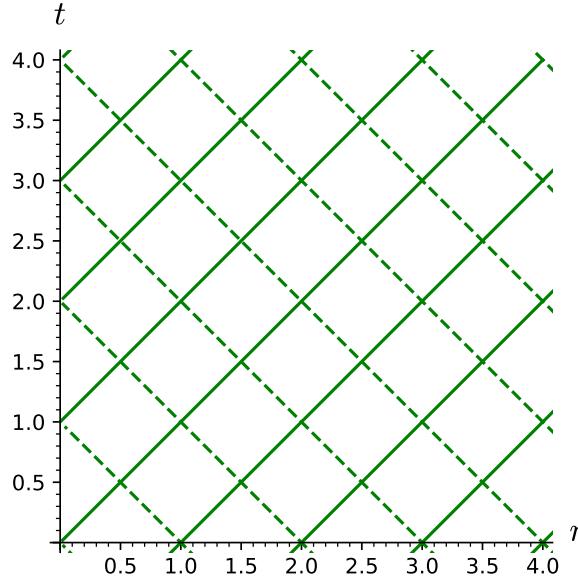


Figure 4.1: Lines of constant null coordinates u (solid) and v (dashed) in terms of the coordinates (t, r) .

4.2 Conformal completion of Minkowski spacetime

In this section (\mathcal{M}, g) is the 4-dimensional Minkowski spacetime, i.e. \mathcal{M} is a smooth manifold diffeomorphic to \mathbb{R}^4 and g is the metric tensor whose expression in terms of some global coordinates (t, x, y, z) implementing the diffeomorphism to \mathbb{R}^4 (i.e. *Minkowskian coordinates*) is

$$g = -dt^2 + dx^2 + dy^2 + dz^2. \quad (4.1)$$

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 (t, r, θ, φ) , which are related to the Minkowskian ones by

$$\begin{cases} x = r \sin \theta \cos \varphi \\ y = r \sin \theta \sin \varphi \\ z = r \cos \theta. \end{cases} \quad (4.2)$$

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 \mathcal{M} as a regular chart (cf. Sec. A.2.1 of Appendix A), but only $\mathcal{M} \setminus \Pi$, 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

$$g = -dt^2 + dr^2 + r^2 (\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\varphi^2). \quad (4.3)$$

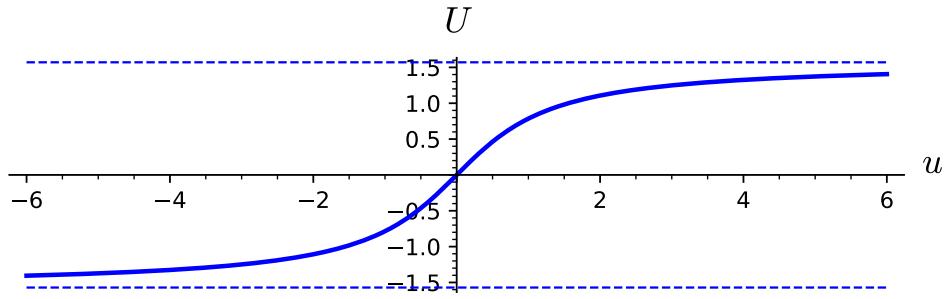


Figure 4.2: The arctangent function mapping \mathbb{R} to $(-\pi/2, \pi/2)$.

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)

$$\begin{cases} u = t - r \\ v = t + r \end{cases} \iff \begin{cases} t = \frac{1}{2}(v + u) \\ r = \frac{1}{2}(v - u). \end{cases} \quad (4.4)$$

One has then $du dv = dt^2 - dr^2$ and the metric tensor (4.3) takes the shape

$$g = -du dv + \frac{1}{4}(v - u)^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (4.5)$$

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):

$$\begin{cases} U = \arctan u \\ V = \arctan v \end{cases} \iff \begin{cases} u = \tan U \\ v = \tan V. \end{cases} \quad (4.6)$$

Given that \arctan is a monotonically increasing function (cf. Fig. 4.2), the coordinates (U, V) span the half part of $(-\pi/2, \pi/2) \times (-\pi/2, \pi/2)$ defined by $U < V$:

$$-\frac{\pi}{2} < U < \frac{\pi}{2}, \quad -\frac{\pi}{2} < V < \frac{\pi}{2}, \quad \text{and} \quad U < V. \quad (4.7)$$

Since

$$du = \frac{dU}{\cos^2 U}, \quad dv = \frac{dV}{\cos^2 V} \quad \text{and} \quad \tan V - \tan U = \frac{\sin(V - U)}{\cos U \cos V},$$

the Minkowski metric (4.5) is expressed in terms of the coordinates (U, V, θ, φ) as¹

$$g = \frac{1}{4 \cos^2 U \cos^2 V} [-4 dU dV + \sin^2(V - U) (d\theta^2 + \sin^2 \theta d\varphi^2)]. \quad (4.8)$$

¹See also the SageMath notebook D.2.1.

Remark 1: The retarded/advanced times u and v have the dimension of a time, or of a length in the $c = 1$ units used here. Therefore, one should introduce some length scale, ℓ_0 say, 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, so that, 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 $\tilde{\mathbf{g}}$ such that

$$\boxed{\tilde{\mathbf{g}} = \Omega^2 \mathbf{g}}, \quad (4.9)$$

where Ω is the scalar field $\mathcal{M} \rightarrow \mathbb{R}$ obeying

$$\Omega = 2 \cos U \cos V \quad (4.10a)$$

$$= \frac{2}{\sqrt{u^2 + 1} \sqrt{v^2 + 1}} \quad (4.10b)$$

$$= \frac{2}{\sqrt{(t - r)^2 + 1} \sqrt{(t + r)^2 + 1}}. \quad (4.10c)$$

We notice on (4.10b) and (4.10c) that the function Ω never vanishes on \mathcal{M} , so that the bilinear form $\tilde{\mathbf{g}}$ defined by (4.9) constitutes a well-behaved metric on \mathcal{M} . Moreover, since $\Omega^2 > 0$, $\tilde{\mathbf{g}}$ has the same signature as \mathbf{g} , i.e. $(-, +, +, +)$. The expression of $\tilde{\mathbf{g}}$ is deduced from (4.8) and (4.10a):

$$\tilde{\mathbf{g}} = -4 dU dV + \sin^2(V - U) (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (4.11)$$

In view of (4.9), one says that the metric $\tilde{\mathbf{g}}$ is **conformal to** the metric \mathbf{g} , or equivalently, that the metrics \mathbf{g} and $\tilde{\mathbf{g}}$ are **conformally related**, or that $\tilde{\mathbf{g}}$ arises from \mathbf{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 $p \in \mathcal{M}$,

$$\forall (\mathbf{u}, \mathbf{v}) \in T_p \mathcal{M} \times T_p \mathcal{M}, \quad \tilde{\mathbf{g}}(\mathbf{u}, \mathbf{v}) = 0 \iff \mathbf{g}(\mathbf{u}, \mathbf{v}) = 0.$$

In particular, null vectors for $\tilde{\mathbf{g}}$ coincide with null vectors for \mathbf{g} :

$$\forall \ell \in T_p \mathcal{M}, \quad \tilde{\mathbf{g}}(\ell, \ell) = 0 \iff \mathbf{g}(\ell, \ell) = 0.$$

Consequently the light cones of $(\mathcal{M}, \mathbf{g})$ and $(\mathcal{M}, \tilde{\mathbf{g}})$ are identical, which implies that $(\mathcal{M}, \mathbf{g})$ and $(\mathcal{M}, \tilde{\mathbf{g}})$ have the same causal structure. Moreover, since $\Omega^2 > 0$, the spacelike and timelike characters of vectors is preserved as well:

$$\begin{aligned} \forall \mathbf{v} \in T_p \mathcal{M}, \quad & \mathbf{v} \text{ spacelike for } \tilde{\mathbf{g}} \iff \mathbf{v} \text{ spacelike for } \mathbf{g} \\ & \mathbf{v} \text{ timelike for } \tilde{\mathbf{g}} \iff \mathbf{v} \text{ timelike for } \mathbf{g}. \end{aligned} \quad (4.12)$$

It follows that a curve \mathcal{L} is timelike (resp. null, spacelike) for $\tilde{\mathbf{g}}$ iff \mathcal{L} is timelike (resp. null, spacelike) for \mathbf{g} . Similarly, a hypersurface Σ is timelike (resp. null, spacelike) for $\tilde{\mathbf{g}}$ iff Σ is timelike (resp. null, spacelike) for \mathbf{g} .

What about geodesics? Let us first recall that a null curve is not necessarily a null geodesic (cf. Remark 2 on p. 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:

Property 4.1: null geodesics preserved by conformal transformations

Let \mathbf{g} and $\tilde{\mathbf{g}}$ be two Lorentzian metrics on a manifold \mathcal{M} that are conformally related: $\tilde{\mathbf{g}} = \Omega^2 \mathbf{g}$. A smooth curve \mathcal{L} in \mathcal{M} is a null geodesic for $\tilde{\mathbf{g}}$ iff \mathcal{L} is a null geodesic for \mathbf{g} ; any affine parameter $\tilde{\lambda}$ of \mathcal{L} as a $\tilde{\mathbf{g}}$ -geodesic is then related to any affine parameter λ of \mathcal{L} as a \mathbf{g} -geodesic by

$$\frac{d\tilde{\lambda}}{d\lambda} = a \Omega^2, \quad (4.13)$$

where a is a constant.

Sketch of proof. Write explicitly the geodesic equation [Eq. (B.10)] and express the Christoffel symbols of $\tilde{\mathbf{g}}$ in terms of those of \mathbf{g} and the derivatives of Ω (see e.g. Appendix D of Wald's textbook [498] for details). \square

On the contrary, conformal transformations preserve neither timelike geodesics nor space-like ones.

The coordinates (U, V) are of null type; let us consider instead the “time+space” coordinates (τ, χ) defined by²

$$\begin{cases} \tau = V + U \\ \chi = V - U \end{cases} \iff \begin{cases} U = \frac{1}{2}(\tau - \chi) \\ V = \frac{1}{2}(\tau + \chi). \end{cases} \quad (4.14)$$

Given (4.7), the range of these new coordinates is

$$0 < \chi < \pi \quad \text{and} \quad \chi - \pi < \tau < \pi - \chi. \quad (4.15)$$

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 (4.14), we get the link between (t, r) and (τ, χ) :

$$\begin{cases} \tau = \arctan(t+r) + \arctan(t-r) \\ \chi = \arctan(t+r) - \arctan(t-r) \end{cases} \iff \begin{cases} t = \frac{\sin \tau}{\cos \tau + \cos \chi} \\ r = \frac{\sin \chi}{\cos \tau + \cos \chi}. \end{cases} \quad (4.16)$$

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 (4.14):

$$\Omega = \cos \tau + \cos \chi. \quad (4.17)$$

²Notice the similarity with (4.4) up to some 1/2 factors.

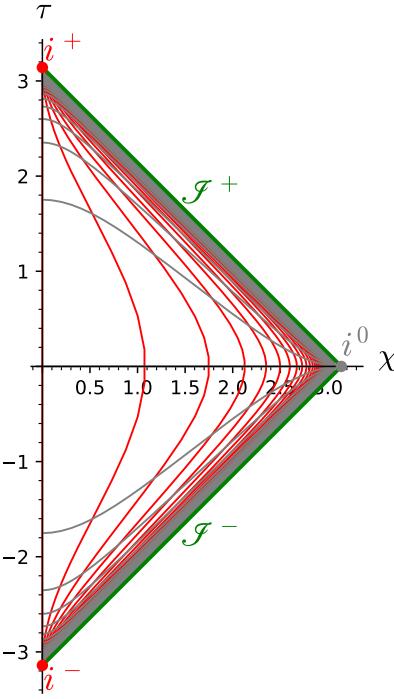


Figure 4.3: Conformal diagram of Minkowski spacetime. Constant- r curves are drawn in red, while constant- t ones are drawn in grey. [Figure generated by the notebook D.2.1]

4.2.3 Conformal completion

The expression of the conformal metric in terms of the coordinates $(\tau, \chi, \theta, \varphi)$ is easily deduced from that in terms of (U, V, θ, φ) as given by (4.11):

$$\tilde{g} = -d\tau^2 + d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (4.18)$$

Restricting to a $\tau = \text{const}$ hypersurface, i.e. setting $d\tau = 0$, 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 (4.18) is perfectly regular even if we relax the condition on τ in (4.15), i.e. if we let τ span the entire \mathbb{R} . We may then consider the manifold

$$\mathcal{E} = \mathbb{R} \times \mathbb{S}^3 \quad (4.19)$$

and \tilde{g} as the Lorentzian metric on \mathcal{E} given by (4.18). The Lorentzian manifold (\mathcal{E}, \tilde{g}) is nothing but the **Einstein static universe**, also called the **Einstein cylinder**, a static solution to the Einstein equation (1.40) with $\Lambda > 0$ and some pressureless matter of uniform density $\rho = \Lambda/(4\pi)$. We have thus an embedding³ of Minkowski spacetime into the Einstein cylinder (cf. Fig. 4.4):

$$\Phi : \mathcal{M} \longrightarrow \mathcal{E} \quad (4.20)$$

³Cf. Sec. A.2.7 of Appendix A

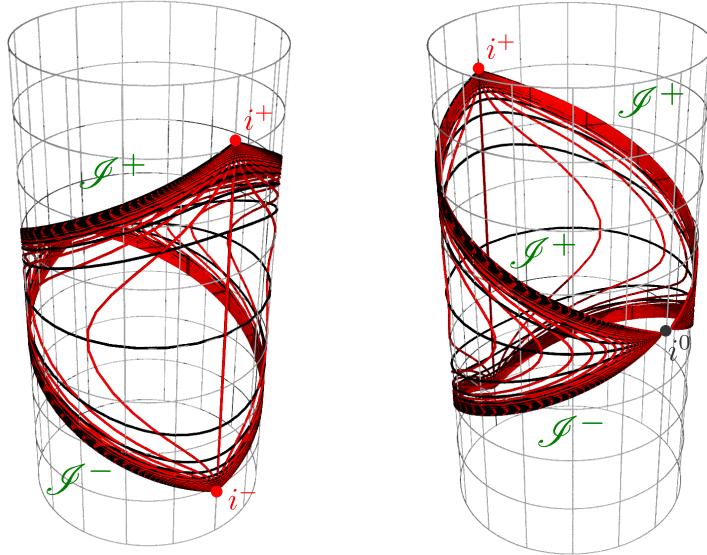


Figure 4.4: Two views of the Einstein cylinder \mathcal{E} , with the conformal embedding of Minkowski spacetime \mathcal{M} in it. Each view represents a 2-dimensional cut of \mathcal{E} at $\theta = \pi/2$ and $\varphi = 0$ (one half of the plotted cylinder) or $\varphi = \pi$ (the other half). The \mathbb{S}^3 sections of \mathcal{E} are then depicted as horizontal circles, which are made of two half-circles corresponding to $\varphi = 0$ and $\varphi = \pi$ respectively, with χ running from 0 to π on both on them. The plotted piece of \mathcal{M} is then the 2-dimensional cut $(y, z) = (0, 0)$ of \mathcal{M} , which is a timelike plane spanned by the coordinates (t, x) , with $x \geq 0$ on the $\varphi = 0$ half cylinder and $x \leq 0$ on the $\varphi = \pi$ half one. 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]

and this embedding is a conformal isometry from (\mathcal{M}, g) to $(\Phi(\mathcal{M}), \tilde{g})$. In the following, we shall identify $\Phi(\mathcal{M})$ and \mathcal{M} , i.e. use the same symbol \mathcal{M} to denote the subset of \mathcal{E} that is the image of \mathcal{M} via the embedding (4.20).

Since \mathcal{E} and \mathcal{M} have the same dimension, \mathcal{M} is an open subset of \mathcal{E} . Its (topological) closure $\overline{\mathcal{M}}$ in \mathcal{E} is (cf. Figs. 4.3 and 4.4)

$$\overline{\mathcal{M}} = \mathcal{M} \cup \mathcal{J}^+ \cup \mathcal{J}^- \cup \{i^0\} \cup \{i^+\} \cup \{i^-\}, \quad (4.21)$$

where

- \mathcal{J}^+ is the hypersurface of \mathcal{E} defined by $\tau = \pi - \chi$ and $0 < \tau < \pi \iff V = \pi/2$ and $-\pi/2 < U < \pi/2$;
- \mathcal{J}^- is the hypersurface of \mathcal{E} defined by $\tau = \chi - \pi$ and $-\pi < \tau < 0 \iff U = -\pi/2$ and $-\pi/2 < V < \pi/2$;
- i^0 is the point of \mathcal{E} defined by $\tau = 0$ and $\chi = \pi \iff U = -\pi/2$ and $V = \pi/2$;
- i^+ is the point of \mathcal{E} defined by $\tau = \pi$ and $\chi = 0 \iff U = \pi/2$ and $V = \pi/2$;
- i^- is the point of \mathcal{E} defined by $\tau = -\pi$ and $\chi = 0 \iff U = -\pi/2$ and $V = -\pi/2$.

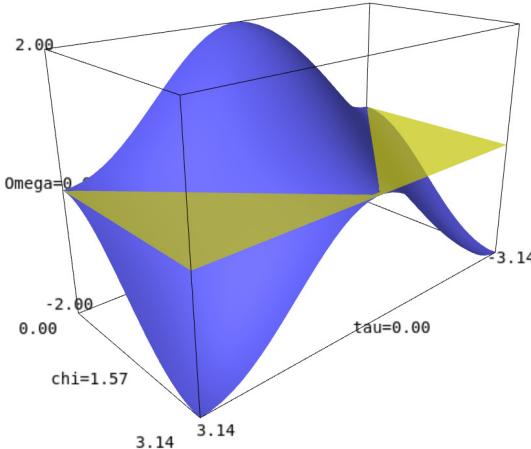


Figure 4.5: Conformal factor Ω as a function of (τ, χ) [cf. Eq. (4.17)]. Only the part above the yellow horizontal plane ($\Omega = 0$) is physical. [Figure generated by the notebook D.2.1]

It is customary to pronounce \mathcal{I} as “scri”, for *script i*. Note that in the above definitions, we have extended the coordinates (U, V) to \mathcal{E} by the transformations (4.14); (U, V, θ, φ) can be then considered as a chart on \mathcal{E} with (U, V) spanning the infinite strip $0 < V - U < \pi$ of \mathbb{R}^2 and the metric \tilde{g} on \mathcal{E} being given by expression (4.11).

Remark 2: It is precisely because Ω vanishes at the topological boundary of \mathcal{M} in \mathcal{E} ,

$$\overline{\mathcal{M}} \setminus \mathcal{M} = \mathcal{I}^+ \cup \mathcal{I}^- \cup \{i^0\} \cup \{i^+\} \cup \{i^-\}, \quad (4.22)$$

that the conformal transformation (4.9) brings the infinity of Minkowski spacetime to a finite distance (cf. Fig. 4.5).

Remark 3: 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 \mathcal{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 (4.16).

Let

$$\mathcal{I} := \mathcal{I}^+ \cup \mathcal{I}^- \quad (4.23)$$

and

$$\tilde{\mathcal{M}} := \mathcal{M} \cup \mathcal{I}. \quad (4.24)$$

$\tilde{\mathcal{M}}$ is naturally a smooth manifold with boundary⁴ and its boundary is \mathcal{I} :

$$\partial \tilde{\mathcal{M}} = \mathcal{I}. \quad (4.25)$$

⁴Cf. Sec. A.2.2 for the precise definition.

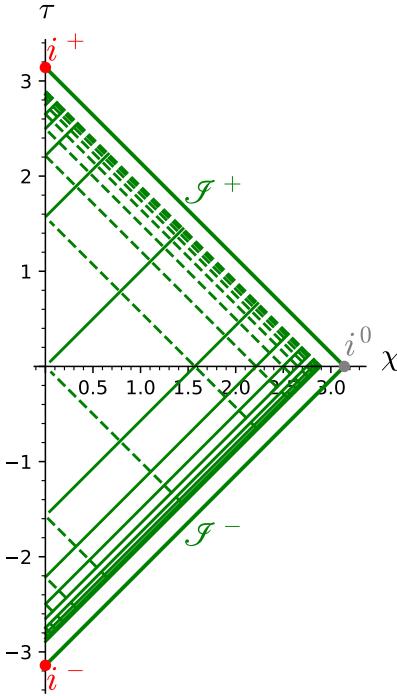


Figure 4.6: Null radial geodesics in the conformal diagram of Minkowski spacetime. The solid green lines are null geodesics $u = \text{const}$ for 17 values of u uniformly spanning $[-8, 8]$, while the dashed green lines are null geodesics $v = \text{const}$ for 17 values of v uniformly spanning $[-8, 8]$. [Figure generated by the notebook [D.2.1](#)]

Remark 4: Because the closure $\tilde{\mathcal{M}}$ is self-intersecting at the point i^0 (cf. Fig. 4.4), it is not a manifold with boundary: no open neighborhood of i^0 is homeomorphic to a neighborhood 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^- , $\tilde{\mathcal{M}}$ 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 $\tilde{\mathcal{M}}$.

The hypersurface \mathcal{I}^+ is the location of $\tilde{\mathcal{M}}$ where all radial null geodesics terminate, while \mathcal{I}^- is the location of $\tilde{\mathcal{M}}$ where all these geodesics originate (cf. Fig. 4.6). For this reason \mathcal{I}^+ is called the **future null infinity** of (\mathcal{M}, g) and \mathcal{I}^- the **past null infinity** of (\mathcal{M}, g) . On the other side, any timelike geodesic of (\mathcal{M}, g) originates at i^- and ends at i^+ (cf. Fig. 4.3), while any spacelike geodesic of (\mathcal{M}, 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 (\mathcal{M}, g) , i^- the **past timelike infinity** of (\mathcal{M}, g) and i^0 the **spacelike infinity** of (\mathcal{M}, g) .

Property 4.2

\mathcal{I}^+ and \mathcal{I}^- are null hypersurfaces of $(\tilde{\mathcal{M}}, \tilde{g})$.

Proof. Since \mathcal{I}^+ is defined by $V = \pi/2$, a valid coordinate system on \mathcal{I}^+ is (U, θ, φ) with U spanning $(-\pi/2, \pi/2)$. The metric induced by \tilde{g} on \mathcal{I}^+ is easily obtained by setting $V = \pi/2$

in Eq. (4.11):

$$\tilde{g}|_{\mathcal{I}^+} = \cos^2 U (\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\varphi^2). \quad (4.26)$$

It appears clearly that the signature of this metric is $(0, +, +)$, i.e. it is degenerate; hence \mathcal{I}^+ is a null hypersurface of $(\tilde{\mathcal{M}}, \tilde{g})$. Similarly, \mathcal{I}^- being defined by $U = -\pi/2$, a valid coordinate system on \mathcal{I}^- is (V, θ, φ) with V spanning $(-\pi/2, \pi/2)$ and the metric induced by \tilde{g} on \mathcal{I}^- is obtained by setting $U = -\pi/2$ in Eq. (4.11):

$$\tilde{g}|_{\mathcal{I}^-} = \cos^2 V (\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\varphi^2). \quad (4.27)$$

Again, it is clearly degenerate, so that \mathcal{I}^- is null hypersurface of $(\tilde{\mathcal{M}}, \tilde{g})$. \square

The null character of \mathcal{I}^+ and \mathcal{I}^- appears also clearly in the conformal diagrams of Figs. 4.3 and 4.6, since \mathcal{I}^+ and \mathcal{I}^- are straight lines of slope ± 1 in these diagrams.

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 [402] and expanded in 1964 in the seminal paper [403], where Penrose constructed the conformal completion of Minkowski spacetime as a part of the Einstein cylinder. In particular, Fig. 3 of Ref. [403] 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 (\mathcal{M}, g) admits a **conformal completion at infinity** iff there exists a Lorentzian manifold with boundary $(\tilde{\mathcal{M}}, \tilde{g})$ (cf. Sec. A.2.2 for the definition) equipped with a smooth non-negative scalar field $\Omega : \tilde{\mathcal{M}} \rightarrow \mathbb{R}^+$ such that

1. $\tilde{\mathcal{M}} = \mathcal{M} \cup \mathcal{I}$, with $\mathcal{I} := \partial \tilde{\mathcal{M}}$ (the manifold boundary^a of $\tilde{\mathcal{M}}$);
2. on \mathcal{M} , $\tilde{g} = \Omega^2 g$;
3. on \mathcal{I} , $\Omega = 0$;
4. on \mathcal{I} , $\mathrm{d}\Omega \neq 0$.

\mathcal{I} is called the **conformal boundary** of (\mathcal{M}, g) within the conformal completion $(\tilde{\mathcal{M}}, \tilde{g})$.

^aAs stressed in Remark 5 in Sec. A.2.2, the set $\partial \tilde{\mathcal{M}}$ is the boundary of $\tilde{\mathcal{M}}$ as a *manifold with boundary*; it is not the boundary of $\tilde{\mathcal{M}}$ as a *topological space*, the latter being \emptyset .

Condition 1 expresses that \mathcal{M} has been endowed with some boundary. A rigorous formulation of it would be via an embedding $\Phi : \mathcal{M} \rightarrow \tilde{\mathcal{M}}$, as in Eq. (4.20), so that $\tilde{\mathcal{M}} = \Phi(\mathcal{M}) \cup \mathcal{I}$.

However, as above, we identify $\Phi(\mathcal{M})$ with \mathcal{M} and therefore simply write $\tilde{\mathcal{M}} = \mathcal{M} \cup \mathcal{I}$. Conditions 2 and 3 express that the boundary of \mathcal{M} , which “lies at an infinite distance” with respect to \mathbf{g} , has been brought to a finite distance with respect to $\tilde{\mathbf{g}}$. Indeed, in terms of length elements [cf. Eq. (1.3)], condition 2 implies

$$ds^2 = \frac{1}{\Omega^2} d\tilde{s}^2$$

with $1/\Omega^2 \rightarrow +\infty$ as one approaches \mathcal{I} (condition 3). Finally, condition 4 ensures that \mathcal{I} is a regular hypersurface of $\tilde{\mathcal{M}}$. It is of course fulfilled by Minkowski spacetime, as we can check graphically on Fig. 4.5: the graph of Ω has no horizontal slope at \mathcal{I} .

Remark 1: The statement that $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ is a Lorentzian manifold with boundary implies that $\tilde{\mathbf{g}}$ is smooth everywhere on $\tilde{\mathcal{M}}$, including at the boundary \mathcal{I} .

Remark 2: Since $\tilde{\mathbf{g}}$ is a metric, it is by definition non-degenerate and condition 2 implies that Ω cannot vanish on \mathcal{M} . Being non-negative, we have necessarily $\Omega > 0$ on \mathcal{M} .

Remark 3: The conformal boundary \mathcal{I} is not part of the physical spacetime \mathcal{M} , but only of the conformal completion $\tilde{\mathcal{M}}$.

Remark 4: One often speaks about *conformal compactification* instead of *conformal completion*, but in general $\tilde{\mathcal{M}}$ is not a compact manifold. For instance, the completion $\tilde{\mathcal{M}}$ of Minkowski spacetime defined by Eq. (4.24) is not compact, for the points i^+ , i^- and i^0 have been omitted in the construction of $\tilde{\mathcal{M}}$.

Example 1 (conformal completion of AdS₄ spacetime): The 4-dimensional anti-de Sitter spacetime $(\mathcal{M}, \mathbf{g})$ has been introduced in Example 18 of Chap. 3. The metric tensor expressed in the conformal coordinates $(\tau, \chi, \theta, \varphi)$ is given by Eq. (3.50):

$$\mathbf{g} = \frac{\ell^2}{\cos^2 \chi} [-d\tau^2 + d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2)], \quad (4.28)$$

with $\tau \in \mathbb{R}$, $\chi \in (0, \pi/2)$, $\theta \in (0, \pi)$ and $\varphi \in (0, 2\pi)$. Defining $\Omega := \ell^{-1} \cos \chi$, we notice that a conformal completion of $(\mathcal{M}, \mathbf{g})$ is $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ where (i) $\tilde{\mathcal{M}}$ is the part $\chi \leq \pi/2$ of the Einstein cylinder⁵ introduced in Sec. 4.2.3 and (ii) $\tilde{\mathbf{g}}$ is the metric (4.18). The boundary $\mathcal{I} = \partial \tilde{\mathcal{M}}$ is then the hypersurface $\chi = \pi/2$ of the Einstein cylinder (cf. Fig. 4.7); \mathcal{I} is spanned by the coordinates (τ, θ, φ) and its topology is that of a 3-dimensional cylinder: $\mathcal{I} \simeq \mathbb{R} \times \mathbb{S}^2$. We notice that conditions 3 and 4 of the definition of a conformal completion are satisfied: $\Omega = \ell^{-1} \cos \chi = 0$ at \mathcal{I} and $d\Omega = -\ell^{-1} \sin \chi d\chi = -\ell^{-1} d\chi \neq 0$ at \mathcal{I} . The metric induced by $\tilde{\mathbf{g}}$ on \mathcal{I} is obtained by setting $\chi = \pi/2$ in (4.18): $-d\tau^2 + d\theta^2 + \sin^2 \theta d\varphi^2$. This 3-metric is clearly Lorentzian, which shows that \mathcal{I} is a timelike hypersurface of $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$.

The above example shows that \mathcal{I} is not necessarily a null hypersurface, as it is for Minkowski spacetime (cf. Sec. 4.2.3). Actually the causal type of \mathcal{I} is determined by the cosmological constant, as follows:

⁵Recall that on the Einstein cylinder the range of χ is $(0, \pi)$, cf. Eq. (4.15).

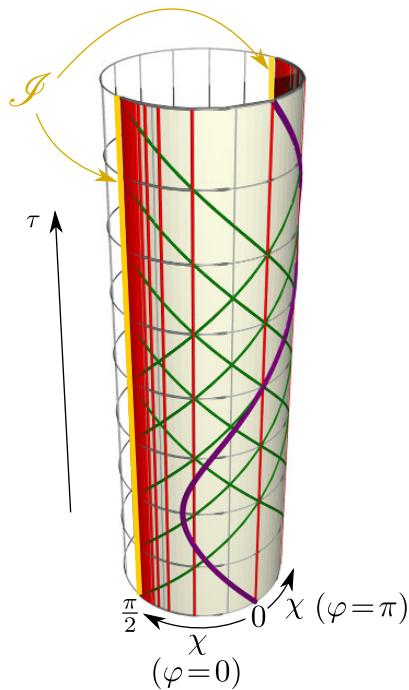


Figure 4.7: Conformal completion of AdS_4 spacetime, depicted on the Einstein cylinder. The conformal boundary \mathcal{I} is shown in yellow, red lines are lines $\chi = \text{const}$ (uniformly sampled in terms of $\tan \chi = \sinh \rho$), green curves are radial null geodesics and the purple curve is a radial timelike geodesic, bouncing back and forth around $\chi = 0$. [Figure generated by the notebook D.3.2]

Property 4.3: causal type of \mathcal{I} and sign of the cosmological constant

If the spacetime dimension obeys^a $n \geq 3$ and \mathbf{g} is a solution to the Einstein equation with a cosmological constant Λ [Eq. (1.40)] and the trace T of the energy-momentum tensor tends to zero in the vicinity of \mathcal{I} (i.e. when $\Omega \rightarrow 0$), then

- \mathcal{I} is a null hypersurface of $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ iff $\Lambda = 0$;
- \mathcal{I} is a spacelike hypersurface of $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ iff $\Lambda > 0$;
- \mathcal{I} is a timelike hypersurface of $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ iff $\Lambda < 0$.

^aCf. Remark 1 in Sec. 1.5.

Proof. It follows from $\tilde{\mathbf{g}} = \Omega^2 \mathbf{g}$ that the Ricci scalars \tilde{R} and R of respectively $\tilde{\mathbf{g}}$ and \mathbf{g} are related by⁶

$$\Omega^2 \tilde{R} = R - (n-1) \left(2\Omega \tilde{g}^{\mu\nu} \tilde{\nabla}_\mu \tilde{\nabla}_\nu \Omega - n \tilde{g}^{\mu\nu} \partial_\mu \Omega \partial_\nu \Omega \right), \quad (4.29)$$

where $n = \dim \mathcal{M}$ and $\tilde{\nabla}$ stands for the Levi-Civita connection of $\tilde{\mathbf{g}}$. Using the trace of the

⁶This relation is easily established by starting from Eq. (2.30) of Hawking & Ellis' textbook [265] or Eq. (2.19) on p. 645 of Choquet-Bruhat's one [107] and inverting the roles of $\tilde{\mathbf{g}}$ and \mathbf{g} , thereby substituting Ω^{-1} for Ω .

Einstein equation (1.42) to express R , we get

$$\Omega^2 \tilde{R} = \frac{2}{n-2} (n\Lambda - 8\pi T) - (n-1) \left(2\Omega \tilde{g}^{\mu\nu} \tilde{\nabla}_\mu \tilde{\nabla}_\nu \Omega - n \tilde{g}^{\mu\nu} \partial_\mu \Omega \partial_\nu \Omega \right)$$

This equation is a priori valid in $\mathcal{M} = \tilde{\mathcal{M}} \setminus \mathcal{I}$ only. Taking the limit $\Omega \rightarrow 0$ and assuming that $T \rightarrow 0$ in that limit, we get, by continuity, an identity on \mathcal{I} :

$$\tilde{g}^{\mu\nu} \partial_\mu \Omega \partial_\nu \Omega \stackrel{\mathcal{I}}{=} -\frac{2}{(n-1)(n-2)} \Lambda. \quad (4.30)$$

Since \mathcal{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 $\tilde{g}(\mathbf{n}, \mathbf{n})$ of the vector \mathbf{n} normal to \mathcal{I} defined as the dual with respect to \tilde{g} of the 1-form $d\Omega$: $n^\alpha = \tilde{g}^{\alpha\mu} \partial_\mu \Omega$ (remember that by hypothesis 4 in the definition of a conformal completion, $d\Omega$ is non-vanishing on \mathcal{I} , so that \mathbf{n} is a valid normal vector to \mathcal{I}). Equation (4.30) implies that the sign of $\tilde{g}(\mathbf{n}, \mathbf{n})$ is the opposite of that of Λ . Given the link between the causal type of a hypersurface and the causal type of its normal (cf. Sec. 2.2.2), this completes the proof. \square

One may distinguish two subparts of the conformal boundary:

Let (\mathcal{M}, g) be a time-orientable^a spacetime admitting a conformal completion at infinity $(\tilde{\mathcal{M}}, \tilde{g})$, with conformal boundary \mathcal{I} . One defines the **future infinity** of (\mathcal{M}, g) as the subset \mathcal{I}^+ of \mathcal{I} whose points can be reached from a point in \mathcal{M} by a future-directed causal curve in $\tilde{\mathcal{M}}$. Similarly the **past infinity** of (\mathcal{M}, g) is the subset \mathcal{I}^- of \mathcal{I} whose points can be reached from a point in \mathcal{M} by a past-directed causal curve in $\tilde{\mathcal{M}}$. If \mathcal{I}^+ (resp. \mathcal{I}^-) is a null hypersurface, it is called the **future null infinity** (resp. **past null infinity**) of (\mathcal{M}, g) . Furthermore if $\mathcal{I} = \mathcal{I}^+ \cup \mathcal{I}^-$, one says that $(\tilde{\mathcal{M}}, \tilde{g})$ is a **conformal completion at null infinity** of (\mathcal{M}, g) .

^aCf. Sec. 1.2.2.

Remark 5: The above definitions of \mathcal{I}^+ and \mathcal{I}^- generalize those given for the Minkowski spacetime in Sec. 4.2.3. Note that, contrary to the Minkowski case, \mathcal{I}^+ and \mathcal{I}^- are not null for spacetimes with a non-zero cosmological constant (cf. Property 4.3). In particular, the following examples exhibit respectively timelike and spacelike \mathcal{I}^+ and \mathcal{I}^- .

Example 2 (Future and past infinities of AdS₄ spacetime): Let us consider the conformal completion of AdS₄ discussed in Example 1. It is evident from the behavior of radial null geodesics (the green curves plotted in Fig. 4.7) that any point of \mathcal{I} can be connected to \mathcal{M} by a future-directed null geodesic as well as by a past-directed one. It follows that $\mathcal{I}^+ = \mathcal{I}^- = \mathcal{I}$ and both \mathcal{I}^+ and \mathcal{I}^- are timelike hypersurfaces.

Example 3 (Conformal completion of dS₄ spacetime): The 4-dimensional **de Sitter spacetime** is (\mathcal{M}, g) with $\mathcal{M} \simeq \mathbb{R} \times \mathbb{S}^3$ and g is the metric whose expression in the so-called *global coordinates* $(t, \chi, \theta, \varphi)$ is

$$g = \ell^2 \left[-dt^2 + \cosh^2 t \left(d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2) \right) \right], \quad (4.31)$$

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 (4.31) is a solution of the vacuum Einstein equation (1.43) with the positive cosmological constant $\Lambda = 3/\ell^2$. Using coordinates $(\tau, \chi, \theta, \varphi)$ with $\tau := 2 \arctan(\tanh(t/2)) \in (-\pi/2, \pi/2)$, one gets

$$\mathbf{g} = \frac{\ell^2}{\cos^2 \tau} [-d\tau^2 + d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2)]. \quad (4.32)$$

Defining $\Omega := \ell^{-1} \cos \tau = (\ell \cosh t)^{-1}$, we notice that a conformal completion of $(\mathcal{M}, \mathbf{g})$ is $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ where (i) $\tilde{\mathcal{M}}$ is the part $-\pi/2 \leq \tau \leq \pi/2$ of the Einstein cylinder introduced in Sec. 4.2.3 and (ii) $\tilde{\mathbf{g}}$ is the metric (4.18). The boundary $\mathcal{I} = \partial \tilde{\mathcal{M}}$ has two connected components: \mathcal{I}^+ , which is the hypersurface $\tau = \pi/2$ of $\tilde{\mathcal{M}}$, and \mathcal{I}^- , which is the hypersurface $\tau = -\pi/2$. Both \mathcal{I}^+ and \mathcal{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 \mathcal{I} and $d\Omega = -\ell^{-1} \sin \tau d\tau = \pm \ell^{-1} d\tau \neq 0$ at \mathcal{I} . The metric induced by $\tilde{\mathbf{g}}$ on \mathcal{I} is obtained by setting $\tau = \pm \pi/2$ in (4.18): $d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2)$. This 3-metric is clearly Riemannian (this is actually the standard round metric of \mathbb{S}^3), which shows that \mathcal{I} is a spacelike hypersurface of $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$. This of course agrees with Property 4.3, given that $\Lambda > 0$. Finally, it is clear that \mathcal{I}^+ (resp. \mathcal{I}^-) matches the definition of a future (resp. past) infinity given above. We conclude that $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ is a conformal completion at infinity of de Sitter spacetime.

4.3.2 Asymptotic flatness

Penrose [403, 406] has defined a spacetime $(\mathcal{M}, \mathbf{g})$ to be **asymptotically simple** iff there exists a conformal completion at infinity $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ of $(\mathcal{M}, \mathbf{g})$ such that every null geodesic in \mathcal{M} has two endpoints in \mathcal{I} .

The last condition, which is verified by Minkowski spacetime (cf. Fig. 4.6), 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 \mathcal{I}^+ , having only a past endpoint on \mathcal{I}^- , as far as \mathcal{I} is concerned. To cope with these spacetimes, Penrose has also introduced the following definition [406]: a spacetime $(\mathcal{M}, \mathbf{g})$ is **weakly asymptotically simple** iff there exists an open subset \mathcal{U} of \mathcal{M} and an asymptotically simple spacetime $(\mathcal{M}_0, \mathbf{g}_0)$ with an open neighborhood \mathcal{U}_0 of $\mathcal{I}_0 = \partial \tilde{\mathcal{M}}_0$ in $\tilde{\mathcal{M}}_0$ such that $(\mathcal{U}_0 \cap \mathcal{M}_0, \mathbf{g}_0)$ is isometric to $(\mathcal{U}, \mathbf{g})$.

Remark 6: For a given weakly asymptotically simple spacetime, there may be different (non overlapping) regions \mathcal{U} satisfying the above property. For instance we shall see in Chap. 10 that there are an infinite series of them in the (maximally extended) Kerr spacetime.

Finally one says that a spacetime $(\mathcal{M}, \mathbf{g})$ is **asymptotically flat** (or more precisely **weakly asymptotically simple and empty** [265]) iff $(\mathcal{M}, \mathbf{g})$ is weakly asymptotically simple and the Ricci tensor of \mathbf{g} vanishes in an open neighborhood of \mathcal{I} : $\mathbf{R} = 0$.

Example 4: The de Sitter and anti-de Sitter spacetimes are asymptotically simple but are not asymptotically flat.

Penrose [405] (see also [197]) has shown that if $(\mathcal{M}, \mathbf{g})$ is asymptotically simple and empty, the Weyl tensor of \mathbf{g} (cf. Sec. A.5.4) vanishes at \mathcal{I} . Since the Ricci tensor is zero, this implies

that the full Riemann curvature tensor vanishes at \mathcal{I} [cf. Eq. (A.114)], hence the qualifier *asymptotically flat*.

The following property holds:

Property 4.4: null \mathcal{I} for asymptotically flat spacetimes

The conformal boundary \mathcal{I} of an asymptotically flat spacetime $(\mathcal{M}, \mathbf{g})$ is a null hypersurface of the conformal completion $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$.

Proof. Consider Eq. (4.29). Near \mathcal{I} , we have $R = 0$ by the very definition of asymptotic flatness. The limit $\Omega \rightarrow 0$ results then in $\tilde{g}^{\mu\nu}\partial_\mu\Omega\partial_\nu\Omega \stackrel{\mathcal{I}}{=} 0$, which, following the argument in the proof on p. 102, implies that \mathcal{I} is a null hypersurface. \square

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 $(\mathcal{M}, \mathbf{g})$. For any subset S of \mathcal{M} , one defines

- the **chronological future of S** as the set $I^+(S)$ of all points of \mathcal{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 \mathcal{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 \mathcal{M} does not contain any closed timelike curve, one has $S \cap I^\pm(S) = \emptyset$ for $S = \{p\}$ with p being any point of \mathcal{M} .

Here are some topological properties of the future and past sets defined above (see e.g. § 6.2 of [265] or Chap. 14 of [389] for proofs):

- $I^\pm(S)$ is always an open subset⁷ of \mathcal{M} , while $J^\pm(S)$ is not necessarily a closed subset.
- The interior of $J^\pm(S)$ is $I^\pm(S)$:

$$\text{int } J^\pm(S) = I^\pm(S). \quad (4.33)$$

⁷This property is a direct consequence of Lemma 4.6 in Sec. 4.4.3 below.

- Both sets have the same closure:

$$\overline{J^\pm(S)} = \overline{I^\pm(S)}. \quad (4.34)$$

- It follows from (4.33) and (4.34) that both sets share the same (topological) boundary:

$$\partial J^\pm(S) = \partial I^\pm(S). \quad (4.35)$$

The subset of the causal future (resp. past) of S formed by points that cannot be connected to S by a timelike curve is called the ***future horismos*** (resp. ***past horismos***) of S and is denoted by $E^+(S)$ (resp. $E^-(S)$):

$$E^+(S) := J^+(S) \setminus I^+(S) \quad \text{and} \quad E^-(S) := J^-(S) \setminus I^-(S). \quad (4.36)$$

The horismos $E^\pm(S)$ is formed by null geodesics emanating from points in S (cf. Proposition 4.5.10 of Ref. [265]). One has $E^\pm(S) \subset \partial J^\pm(S)$. The spacetime (\mathcal{M}, g) is said to be ***causally simple*** iff for every compact set $K \subset \mathcal{M}$, $E^\pm(K) = \partial J^\pm(K)$. This is equivalent to saying that $J^+(K)$ and $J^-(K)$ are closed subsets of \mathcal{M} .

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 (\mathcal{M}, g) that admits a conformal completion at infinity as defined in Sec. 4.3.1 and such that the future infinity \mathcal{I}^+ is ***complete***: if \mathcal{I}^+ is a null hypersurface, which occurs if (\mathcal{M}, g) is asymptotically flat (cf. Property 4.4), this means that \mathcal{I}^+ is generated by complete⁸ null geodesics. The completeness condition is imposed to avoid “spurious” black holes, such as black holes in Minkowski space (cf. Remark 2 below). The neighborhood of \mathcal{I}^+ in $\tilde{\mathcal{M}}$ 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. More precisely:

Let (\mathcal{M}, g) be a time-oriented spacetime with a conformal completion at infinity such that the future infinity \mathcal{I}^+ is complete; the ***black hole region*** is the set of points of \mathcal{M} that do not belong to the causal past of \mathcal{I}^+ (cf. Fig. 4.8):

$$\mathcal{B} := \mathcal{M} \setminus (J^-(\mathcal{I}^+) \cap \mathcal{M}). \quad (4.37)$$

If $\mathcal{B} \neq \emptyset$, one says that the spacetime (\mathcal{M}, g) ***contains a black hole*** or that (\mathcal{M}, g) is a ***black hole spacetime***.

The black hole region is thus the set of points of \mathcal{M} from which no future-directed causal curve in $\tilde{\mathcal{M}}$ reaches \mathcal{I}^+ .

⁸Let us recall that a geodesic is *complete* iff its affine parameters range through the whole of \mathbb{R} , cf. Sec. B.3.2 in Appendix B. In particular, such a geodesic is inextendible.

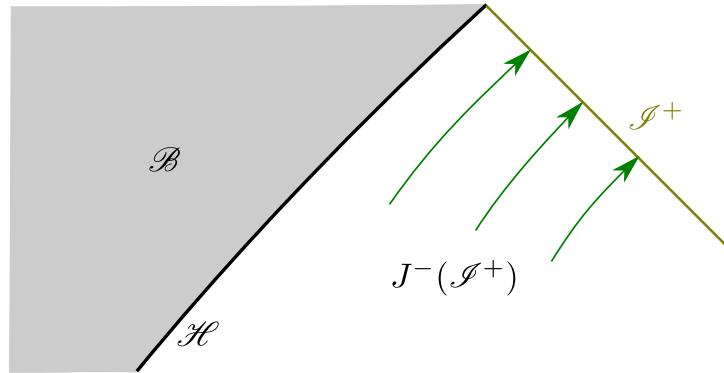


Figure 4.8: The black hole region \mathcal{B} defined as the complement of the causal past of the future infinity, $J^-(\mathcal{I}^+)$.

Example 5: The Minkowski spacetime contains no black hole, for all future-directed null geodesics terminate at \mathcal{I}^+ (cf. Fig. 4.6). More generally, any asymptotically simple spacetime contains no black hole (cf. Sec. 4.3.2).

Example 6: 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 \mathcal{B} that fulfills the above definition of a black hole region.

Remark 2: If we release the assumption of \mathcal{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 (\mathcal{M}, g) resulting from its embedding in the Einstein cylinder (\mathcal{E}, \tilde{g}) , as in Sec. 4.2.3, keeping the same \mathcal{I}^- but defining \mathcal{I}^+ as the hypersurface of \mathcal{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 $\tilde{\mathcal{M}} := \mathcal{M} \cup \mathcal{I}^+ \cup \mathcal{I}^-$, equipped with the Einstein cylinder metric \tilde{g} , is then a conformal completion of (\mathcal{M}, g) at null infinity. With such a \mathcal{I}^+ , the black hole region defined by (4.37) is non-empty, as shown in Fig. 4.9.

Remark 3: Some authors (in particular Hawking and Ellis [265]) define a *black hole* as a connected component of $\Sigma(\tau) \cap \mathcal{B}$, where $\Sigma(\tau)$ is a spacelike hypersurface that is a slice of the future development of a partial Cauchy surface⁹ $\Sigma(0)$ such that the closure in $\tilde{\mathcal{M}}$ of the domain of dependence of $\Sigma(0)$ contains \mathcal{I}^+ . According to such a definition, a black hole is a $(n-1)$ -dimensional object, while the black hole region \mathcal{B} defined above is a n -dimensional object.

If $\mathcal{B} \neq \emptyset$, the boundary \mathcal{H} of the black hole region is called the **future event horizon** (or simply the **event horizon** when no ambiguity may arise):

$$\mathcal{H} := \partial \mathcal{B}. \quad (4.38)$$

By plugging expression (4.37) for \mathcal{B} in the standard identity $\partial \mathcal{B} = \overline{\mathcal{B}} \cap \overline{\mathcal{M} \setminus \mathcal{B}}$, we get an equivalent expression for \mathcal{H} :

$$\mathcal{H} = \overline{\mathcal{M} \setminus (J^-(\mathcal{I}^+) \cap \mathcal{M})} \cap \overline{(J^-(\mathcal{I}^+) \cap \mathcal{M})} = \partial(J^-(\mathcal{I}^+) \cap \mathcal{M}).$$

⁹The concepts of *partial Cauchy surface* and *future development* are defined in Sec. 10.8.3.

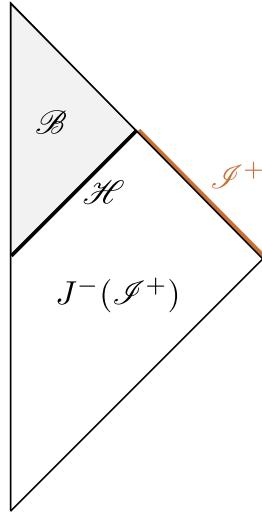


Figure 4.9: Spurious black hole region \mathcal{B} in Minkowski spacetime resulting from a conformal completion with an incomplete \mathcal{I}^+ . Compare with Fig. 4.6.

Now, the boundary of $J^-(\mathcal{I}^+)$ in $\tilde{\mathcal{M}}$ is $\partial J^-(\mathcal{I}^+) = \partial(J^-(\mathcal{I}^+) \cap \mathcal{M}) \cup \mathcal{I}^+$, so that $\partial J^-(\mathcal{I}^+) \cap \mathcal{M} = \partial(J^-(\mathcal{I}^+) \cap \mathcal{M})$; hence

$$\mathcal{H} = \partial J^-(\mathcal{I}^+) \cap \mathcal{M}. \quad (4.39)$$

In words: the future event horizon \mathcal{H} is the part of the boundary of the causal past of the future infinity \mathcal{I}^+ that lies in \mathcal{M} (cf. Fig. 4.8). Note that thanks to property (4.35), we can write as well

$$\mathcal{H} = \partial I^-(\mathcal{I}^+) \cap \mathcal{M}. \quad (4.40)$$

Example 7: The event horizon of the Schwarzschild black hole (Example 6) is nothing but the Schwarzschild horizon \mathcal{H} considered in the examples of Chaps. 2 and 3.

White hole

By inverting past and future in the black hole definition (4.37), one defines the **white hole region** of a spacetime (\mathcal{M}, g) with a conformal completion at infinity as the complement within \mathcal{M} of the causal future of the past infinity \mathcal{I}^- :

$$\mathcal{W} := \mathcal{M} \setminus (J^+(\mathcal{I}^-) \cap \mathcal{M}). \quad (4.41)$$

The white hole region is thus the set of points of \mathcal{M} from which no past-directed causal curve in $\tilde{\mathcal{M}}$ reaches \mathcal{I}^- . The boundary of white hole region is called the **past event horizon**:

$$\mathcal{H}^- := \partial \mathcal{W} = \partial J^+(\mathcal{I}^-) \cap \mathcal{M}. \quad (4.42)$$

Remark 4: The name **white fountain** is sometimes used instead of **white hole**. Actually, this name may seem better suited to describe the time symmetric of a black hole: one can fall into a hole, while one is expelled by a fountain.

Example 8: We shall encounter 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 $\langle\langle \mathcal{M} \rangle\rangle$ of \mathcal{M} that lies neither in the black hole region nor in the white hole one:

$$\langle\langle \mathcal{M} \rangle\rangle := \mathcal{M} \setminus (\mathcal{B} \cup \mathcal{W}) = (J^-(\mathcal{I}^+) \cap J^+(\mathcal{I}^-)) \cap \mathcal{M}. \quad (4.43)$$

The last equality, which is a direct consequence of the definitions of \mathcal{B} and \mathcal{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 arbitrarily 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:

$$\partial\langle\langle \mathcal{M} \rangle\rangle = \mathcal{H} \cup \mathcal{H}^-. \quad (4.44)$$

Historical note : The term *event horizon* has been introduced by Wolfgang Rindler in 1956 [434] in the context of a single observer moving in some cosmological spacetime. Regarding the name *black hole*, the standard story tells 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 conferences (cf. the account by Wheeler himself in Chap. 13 of Ref. [519]). However, a recent study [271] reveals that the expression *black hole* circulated as early as 1963 at 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. In any case, after have been put forward by Wheeler, the term *black hole* rapidly superseded the previous names *frozen star*, *collapsed star*, or *astre occlus* (the latter still appearing along *black holes* in the title of the proceedings of the famous 1972 Les Houches Summer School [163]). For instance, “*black hole*” was used abundantly, albeit with quotes, in a review article by Roger Penrose published in 1969 [407]. In the same article, Penrose defined the *absolute event horizon* as $\partial J^-(\mathcal{I}^+)$ (actually $\partial I^-(\mathcal{I}^+)$, but both coincide, cf. Eq. (4.35)), i.e. essentially the identity (4.39), but he did not provide any formal definition of the black hole region as $\mathcal{M} \setminus (J^-(\mathcal{I}^+) \cap \mathcal{M})$ [Eq. (4.37)]. It seems that the latter appears first in an “instant-of-time” version (cf. Remark 3) in a seminal article by Stephen Hawking published in 1972 [259]. The expression *domain of outer communications* has been introduced in 1971 by Brandon Carter [92].

4.4.3 Properties of the future event horizon

Having defined a black hole region in full generality, let us derive the main properties of its boundary – the future event horizon \mathcal{H} .

Property 4.5: the event horizon as an achronal set

A black hole event horizon \mathcal{H} is an **achronal set**, i.e. no pair of points of \mathcal{H} can be connected by a timelike curve of \mathcal{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 4.5 is: no timelike curve of \mathcal{M} encounters \mathcal{H} at more than one point.

The proof of Property 4.5 relies on the following lemma:

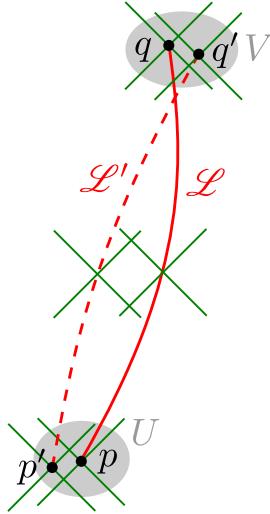


Figure 4.10: Lemma 4.6: moving slightly the ends p and q of a timelike curve \mathcal{L} necessarily results in another timelike curve \mathcal{L}' .

Lemma 4.6: stability of timelike curves with respect to their ends

One can “move the ends” of any timelike curve “a little bit” and still get a timelike curve. More precisely, if two points $p, q \in \mathcal{M}$ are connected by a timelike curve, there exists a neighborhood U of p and a neighborhood V of q such that any point $p' \in U$ can be connected to any point $q' \in V$ by a timelike curve.

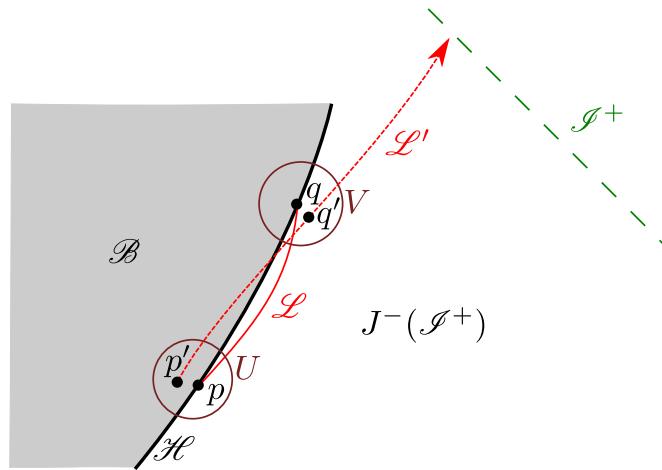
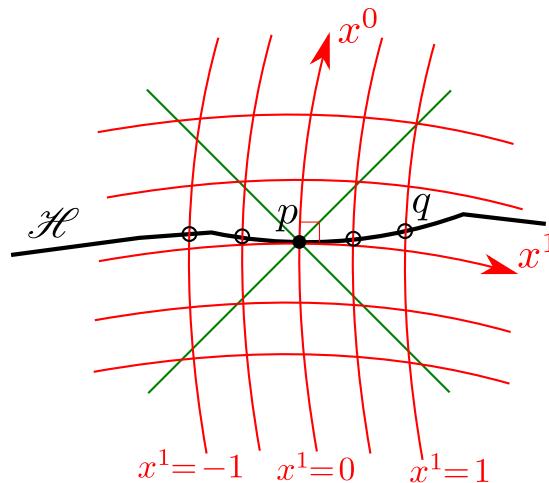
Proof. This is more or less evident on a spacetime diagram (cf. Fig. 4.10) and a formal proof can be found as Lemma 3 in Chap. 14 of O’Neill’s textbook [389]. \square

Proof of Property 4.5. Let us assume the negation of Property 4.5, i.e. that there exist two points p and q in \mathcal{H} that are connected by a timelike curve \mathcal{L} , with q lying in the future of p (cf. Fig. 4.11). Invoking Lemma 4.6, let U and V be the neighborhoods of respectively p and q within which one can deform \mathcal{L} to a timelike curve. Let us choose $p' \in U \cap \mathcal{B}$ (\mathcal{B} being the black hole region) and $q' \in V \cap J^-(\mathcal{I}^+)$. Such a choice is always possible since p and q lie on the boundary between \mathcal{B} and $J^-(\mathcal{I}^+)$ (cf. Fig. 4.11). Since $q' \in J^-(\mathcal{I}^+)$, the timelike curve linking p' and q' can then be extended to the future in a causal curve \mathcal{L}' reaching \mathcal{I}^+ . This implies $p' \in J^-(\mathcal{I}^+)$, which contradicts $p' \in \mathcal{B}$. \square

Property 4.7: the event horizon as manifold of codimension 1

\mathcal{H} is a topological manifold of dimension $n - 1$, n being the spacetime dimension.

Proof. Let $p \in \mathcal{H}$ and U some open neighborhood of p where one can define a normal coordinate system (x^α) . We have then ∂_0 timelike, ∂_i spacelike for $i \in \{1, \dots, n-1\}$ and

Figure 4.11: Proving that \mathcal{H} is achronal.Figure 4.12: Proving that \mathcal{H} is a topological manifold of dimension $n - 1$.

$g(\partial_0, \partial_i) = 0$. Let us consider a curve in U defined by $x^1 = a_1, \dots, x^{n-1} = a_{n-1}$, where a_1, \dots, a_{n-1} are $n - 1$ constants. This curve is timelike, since it has ∂_0 as a tangent vector (cf. Fig. 4.12). It therefore intersects \mathcal{H} at a single point q , for \mathcal{H} is achronal (Property 4.5). Let us then give the coordinates $(y^i) = (a_1, \dots, a_{n-1})$ to q . By varying (a_1, \dots, a_{n-1}) , we get a homeomorphism from $U \cap \mathcal{H}$ to an open subset of \mathbb{R}^{n-1} . \square

Remark 5: Generically, the topological manifold \mathcal{H} is not a smooth manifold, for it contains some points (the crossovers defined below) at which it is not differentiable. Actually \mathcal{H} is slightly more than a mere topological submanifold of \mathcal{M} : it is a *Lipschitz submanifold* of \mathcal{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 $U \cap \mathcal{H}$, the function x^0 is a Lipschitz function of the coordinates (y^i) : $|x^0(y^i) - x^0(y'^i)| < K \sqrt{\sum_i (y^i - y'^i)^2}$. This follows from the achronal character of \mathcal{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

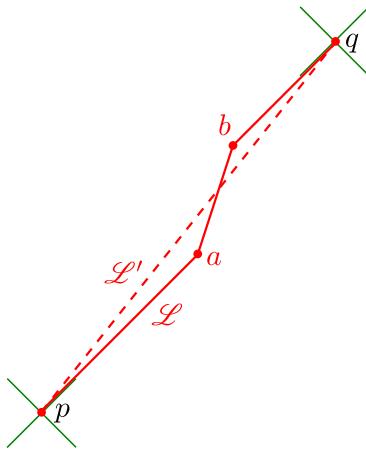


Figure 4.13: Lemma 4.9: A *causal* curve \mathcal{L} containing a timelike segment (between a and b on the figure) can be deformed into an entirely *timelike* curve \mathcal{L}' with the ends kept fixed (dashed curve).

timelike separated. Hence, one says that \mathcal{H} is a **Lipschitz submanifold** of \mathcal{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 4.8: the event horizon ruled by never leaving null geodesics (Penrose 1968 [406])

A black hole event horizon \mathcal{H} is ruled by a family of null geodesics, called the **generators of \mathcal{H}** , that (i) either lie entirely in \mathcal{H} or never leave \mathcal{H} when followed into the future from the point where they enter \mathcal{H} , and (ii) have no endpoint in the future. Moreover, there is exactly one generator through each point of \mathcal{H} , except at special points where null geodesics enter \mathcal{H} , which are called **crossovers**. A special case of crossover, called **caustic**, is a point where neighboring null geodesics focus and converge while arriving in \mathcal{H} .

In particular, once a null geodesic has merged with \mathcal{H} (at a point where it may intersect other null geodesics), it will stay forever on \mathcal{H} and will never intersect any other generator. The set of all crossovers is called the **crease set** [456, 457, 73].

The following proof of Property 4.8 is adapted from that given in Box 34.1 of MTW [370]. In addition of Lemma 4.6, it relies on the following lemma.

Lemma 4.9

Let \mathcal{L} be a causal curve connecting two points p and q of \mathcal{M} . If \mathcal{L} contains a timelike segment, then there exists an entirely timelike curve connecting p and q .

Proof. We shall only provide a graphical “proof”, based on the spacetime diagram of Fig. 4.13. The causal curve \mathcal{L} may have parts where it is null (segments pa and bq in Fig. 4.13); these parts are drawn with the angle of incline $\theta = \pm 45^\circ$. If \mathcal{L} contains a timelike segment (such as

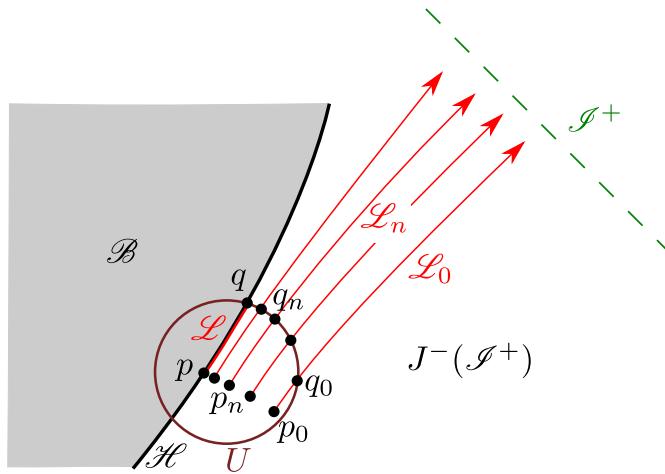


Figure 4.14: Causal curve \mathcal{L} connecting p to q obtained as a limit of causal curves in $J^-(\mathcal{I}^+)$.

ab in Fig. 4.13), i.e. a segment with $|\theta| > 45^\circ$, it can be deformed, while keeping the same ends, to a curve with $|\theta| > 45^\circ$ everywhere, i.e. to a timelike curve. \square

Proof of Property 4.8. Let $p \in \mathcal{H}$ and let U be some convex normal neighborhood¹⁰ of p . Since p lies in the boundary of $J^-(\mathcal{I}^+)$, it is always possible to consider a sequence of points $(p_n)_{n \in \mathbb{N}}$ converging toward p and such that $\forall n \in \mathbb{N}, p_n \in U \cap J^-(\mathcal{I}^+)$ (cf. Fig. 4.14). Since $p_n \in J^-(\mathcal{I}^+)$, there exists a future-directed causal curve \mathcal{L}_n from p_n to \mathcal{I}^+ for each $n \in \mathbb{N}$. The neighborhood U being convex, each \mathcal{L}_n intersects its boundary ∂U at a unique point, q_n say, in the future of p_n : $\{q_n\} = \mathcal{L}_n \cap \partial U$ (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 $\mathbb{N} \rightarrow \mathbb{N}$), that converges to some limit point q . Since from any point $p_{f(n)}$ arbitrarily close to p , there is the causal curve $\mathcal{L}_{f(n)}$ to the point $q_{f(n)}$ arbitrarily close to q , one can show that there exists a future-directed causal curve \mathcal{L} connecting p to q (cf. Fig. 4.14; see e.g. Lemma 6.2.1 of Hawking & Ellis' textbook [265] for a precise demonstration).

As the limit of points in $J^-(\mathcal{I}^+)$, q lies in the closure $\overline{J^-(\mathcal{I}^+)} = J^-(\mathcal{I}^+) \cup \mathcal{H}$, \mathcal{H} being the boundary of $J^-(\mathcal{I}^+)$. Let us show by contradiction that actually $q \in \mathcal{H}$. If we assume $q \notin \mathcal{H}$, then necessarily $q \in J^-(\mathcal{I}^+)$. There exists then an open neighborhood V of q such that $V \subset J^-(\mathcal{I}^+)$ (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 \mathcal{L} to a causal curve $\tilde{\mathcal{L}}$ from p to \mathcal{I}^+ via q and q' (cf. Fig. 4.15). Since $\tilde{\mathcal{L}}$ contains a timelike segment (between q and q'), we may invoke Lemma 4.9 to deform it into a timelike curve $\tilde{\mathcal{L}}'$ between p and \mathcal{I}^+ . Then, by Lemma 4.6, one can “move the past end” of $\tilde{\mathcal{L}}'$ to get a new timelike curve $\tilde{\mathcal{L}}''$ linking an event $p' \in \mathcal{B}$ close to p to \mathcal{I}^+ (dotted curve in Fig. 4.15), which is impossible by the very definition of the black hole region \mathcal{B} . Hence $q \in \mathcal{H}$.

¹⁰Basically, a convex normal neighborhood is an open subset U such that any two points of U can be connected by a geodesic lying entirely in U , cf. Sec. B.3.4. There always exists a convex normal neighborhood around any point of a pseudo-Riemannian manifold (cf. Proposition 7 p. 130 in O’Neill’s textbook [389]).

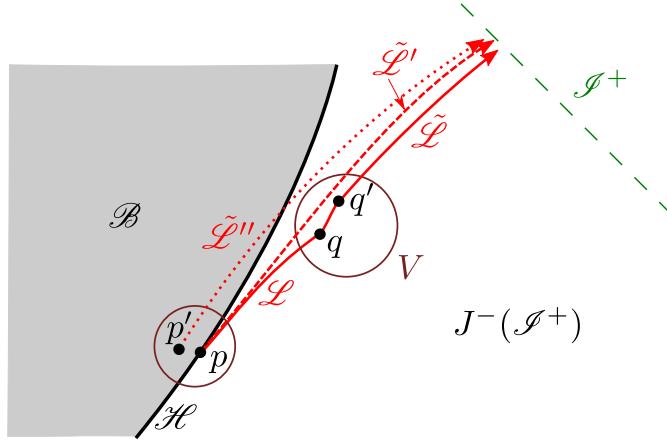


Figure 4.15: Proving by contradiction that q lies in \mathcal{H} .

The causal curve \mathcal{L} connecting p to q cannot be timelike since p and q are both in \mathcal{H} , which is achronal (Property 4.5). Actually, \mathcal{L} cannot even contain a timelike segment: if it would, then by Lemma 4.9, it could be deformed into a timelike curve between p and q , which again would contradict the achronal character of \mathcal{H} . Hence \mathcal{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 \mathcal{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 $P : [\lambda_a, \lambda_b] \rightarrow \mathcal{M}, \lambda \mapsto P(\lambda) \in \mathcal{L}$ is a stationary point of the action

$$E_{(a,b)}(P) := \int_{\lambda_a}^{\lambda_b} \mathbf{g}(\mathbf{v}, \mathbf{v}) d\lambda,$$

where $\mathbf{v} = d\mathbf{x}/d\lambda$ is the tangent vector associated with P . For the null segment ab of \mathcal{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 some timelike part. It follows that p and q can be connected by a causal curve with a timelike segment. But this feature has been excluded above. We conclude that \mathcal{L} is a null geodesic.

At this stage, we have shown that given $p \in \mathcal{H}$, there exists a future-directed null geodesic \mathcal{L} connecting p to another point $q \in \mathcal{H}$. There remains to show that \mathcal{L} lies entirely in \mathcal{H} . Let us start by showing that $\mathcal{L} \subset \overline{J^-(\mathcal{I}^+)}$. Let a be a generic point of \mathcal{L} between p and q . Since \mathcal{L} is null, there exists a point a' arbitrarily close to a such that a' is connected to q by a future-directed timelike curve (cf. Fig. 4.16). Thanks to Lemma 4.6 and the property $q \in \overline{J^-(\mathcal{I}^+)}$, we may find a point $q' \in J^-(\mathcal{I}^+)$ close to q such that a' is connected to q' by a future-directed timelike curve. Since $q' \in J^-(\mathcal{I}^+)$, such a curve can be extended to a causal curve to \mathcal{I}^+ (the dashed curve in Fig. 4.16); hence $a' \in J^-(\mathcal{I}^+)$. Since a' is arbitrarily close to a , we conclude that $a \in \overline{J^-(\mathcal{I}^+)}$. Then, by repeating the same reasoning as that employed above for proving $q \in \mathcal{H}$, simply replacing q by a , we get that $a \in \mathcal{H}$. Since a is a generic point of \mathcal{L} , we conclude that \mathcal{L} lies entirely in \mathcal{H} .

Given a point $p \in \mathcal{H}$, we have thus constructed a future-directed null geodesic \mathcal{L} lying entirely in \mathcal{H} and connecting p to another point $q \in \mathcal{H}$. One can repeat the construction

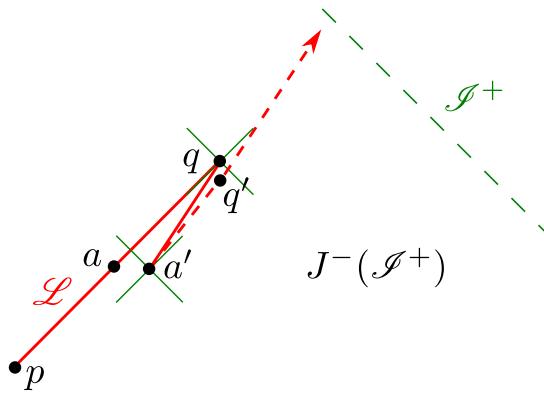
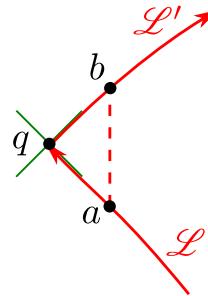
Figure 4.16: Proving that \mathcal{L} lies entirely in \mathcal{H} .

Figure 4.17: Proof of Lemma 4.10.

from the point q to get another future-directed null geodesic $\mathcal{L}' \subset \mathcal{H}$ connecting q to another point $q' \in \mathcal{H}$. Now \mathcal{L} and \mathcal{L}' must be two segments of the same null geodesic $\mathcal{L} \cup \mathcal{L}'$ by the following lemma.

Lemma 4.10

Let $q \in \mathcal{H}$. If $\mathcal{L} \subset \mathcal{H}$ is a null geodesic having q as future end point and $\mathcal{L}' \subset \mathcal{H}$ is a null geodesic having q as past end point, then \mathcal{L} and \mathcal{L}' have collinear tangent vectors at their common point q . It follows that \mathcal{L} and \mathcal{L}' are two segments of a same null geodesic through q .

Proof of Lemma 4.10. Assume that \mathcal{L} and \mathcal{L}' have non-collinear tangent vectors at q . Then, in the vicinity of q , one can find a point $a \in \mathcal{L}$ and a point $b \in \mathcal{L}'$ such that a and b can be connected by a timelike curve (cf. Fig. 4.17). Since $\mathcal{L} \subset \mathcal{H}$ and $\mathcal{L}' \subset \mathcal{H}$, we have $a \in \mathcal{H}$ and $b \in \mathcal{H}$ and therefore we get a contradiction with \mathcal{H} being achronal. \square

Thanks to Lemma 4.10, we conclude that \mathcal{L}' extends \mathcal{L} to the null geodesic $\mathcal{L} \cup \mathcal{L}'$ entirely lying in \mathcal{H} . By iterating, we conclude that the null geodesic \mathcal{L} through p can be extended indefinitely into the future. Moreover, it can never leave \mathcal{H} . Indeed, if \mathcal{L} would leave \mathcal{H} at some point q , by the same procedure used above for p , one could construct a future-directed

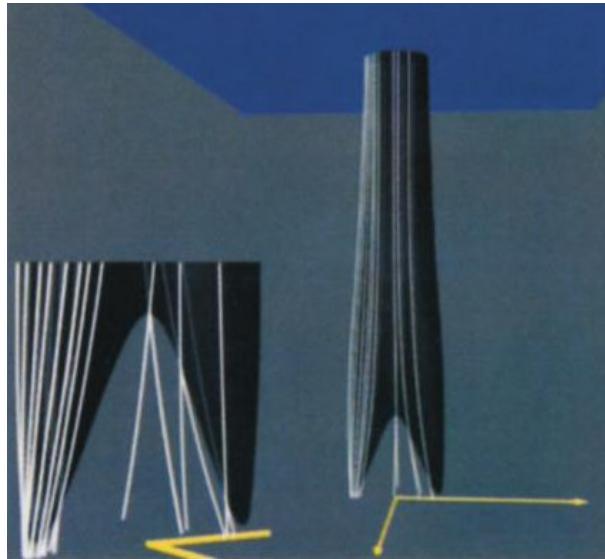


Figure 4.18: Spacetime diagram of the event horizon corresponding to the head-on merger of two black holes as computed by Matzner et al. (1995) [358]. The white curves are some null geodesic generators; the left picture is a zoom of the merger region, with the crease set (source: Fig. 4 of Ref. [358]; ©1995 American Association for the Advancement of Science).

null geodesic $\mathcal{L}' \subset \mathcal{H}$ starting from q ; then Lemma 4.10 would imply that \mathcal{L} and \mathcal{L}' would have the same tangent at q , which is not compatible with \mathcal{L} leaving \mathcal{H} at q .

Another direct consequence of Lemma 4.10 is that no two distinct null generators may intersect at a point $p \in \mathcal{H}$, except if their segments in the past of p lie outside \mathcal{H} . This completes the proof of Property 4.8. \square

Some features of Property 4.8 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) [358]. 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) [132], with an increased numerical accuracy (cf. Fig. 4.19). Cross-sections of the event horizon \mathcal{H} (cf. Sec. 2.3.4) are depicted in Fig. 4.20. The same figure also shows how some null geodesics are reaching \mathcal{H} to become null generators.

Finally, Fig. 4.21 shows a cross-section of the event horizon computed by Cohen et al. (2012) [131] in some inspiralling binary black hole merger. The black hole spacetime itself has been computed as a solution to the vacuum Einstein equation (1.44) by Scheel et al. [445]; it corresponds to 16 inspiralling orbits of an 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 \mathcal{H} and is bounded by the set of caustic points, which forms a 1-dimensional subset of \mathcal{H} [456, 457, 285, 131].

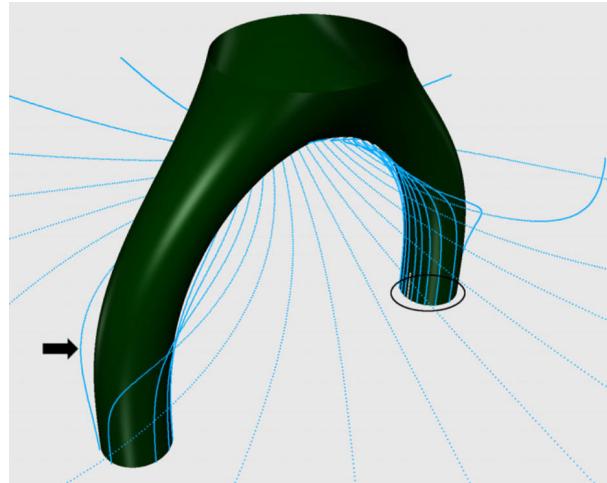


Figure 4.19: Spacetime diagram showing the event horizon in the head-on merger of two black holes, as computed by Cohen et al. (2009) [132]. The blue curves are null geodesics that will eventually become null generators of the event horizon; those arising from regions close to the event horizon are marked by the arrow and the black ellipse (source: Fig. 15 of Ref. [132]; ©2009 IOP Publishing Ltd).

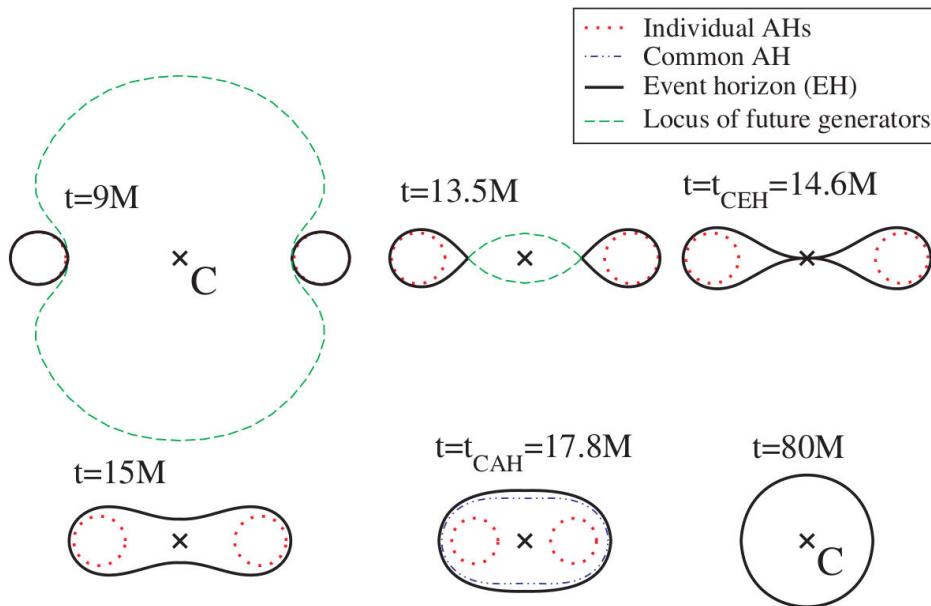


Figure 4.20: Cross-sections (at various coordinate times t) of the event horizon \mathcal{H} corresponding to the head-on merger of two black holes as computed by Cohen et al. (2009) [132] and displayed in Fig. 4.19. Each figure is a 2D cut of a hypersurface Σ_t defined by a constant value of the coordinate time t , expressed in units of the sum M of the initial irreducible masses of each black hole (to be discussed in Chap. 16). 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 \mathcal{H} becomes a connected 2-surface. The cross-sections of \mathcal{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 \mathcal{H} through the cusps in the “individual” event horizons. The red and blue dashed curves denotes *apparent horizons* (to be discussed in Chap. 18). (source: Fig. 1 of Ref. [132]; ©2009 IOP Publishing Ltd).

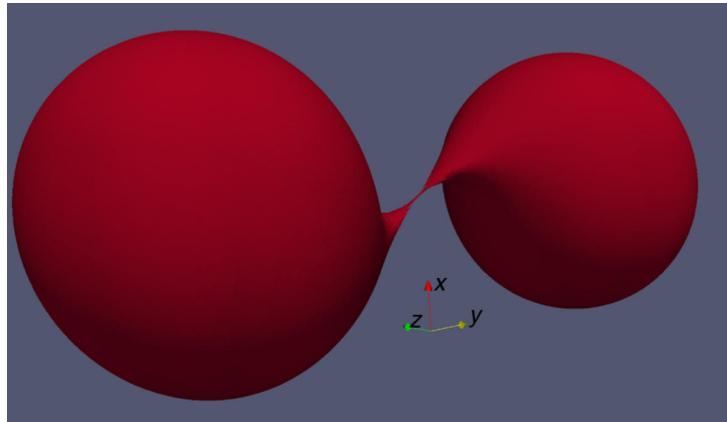


Figure 4.21: Cross-section of the event horizon \mathcal{H} of the inspiralling merger of two black holes as computed by Cohen et al. (2012) [131]. The x and y axes define the orbital plane. This cross-section is the first connected one in the slicing of \mathcal{H} by surfaces of constant coordinate time t (source: Fig. 2 of Ref. [131]; ©2012 American Physical Society).

Property 4.11: the event horizon as a null hypersurface

Wherever it is smooth, \mathcal{H} is a null hypersurface. Its generators, as defined by Property 4.8, are then nothing but the null-hypersurface generators as defined in Sec. 2.3.3.

Proof. Let us assume that \mathcal{H} is smooth in some open subset U . By Property 4.7, \mathcal{H} is then a smooth hypersurface in U . According to Property 4.8, there is a null geodesic lying in \mathcal{H} through any point of $\mathcal{H} \cap U$. This implies null tangent vectors at any point of $\mathcal{H} \cap U$, so that, in U , \mathcal{H} must be either a null hypersurface or a timelike one. But \mathcal{H} is achronal by Property 4.5 and therefore cannot be timelike. Hence, \mathcal{H} is a null hypersurface in U . Finally, there is only one congruence of null geodesics ruling a null hypersurface: the null generators defined in Sec. 2.3.3. The generators invoked in Property 4.8 have thus to belong to that congruence. \square

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 entering \mathcal{H} to become a null generator (cf. Fig. 4.18).

Remark 6: Properties 4.5 to 4.11 are not specific to black hole horizons: they are actually valid for the boundary $\partial J^-(S)$ of the causal past of any set $S \subset \mathcal{M}$, or $S \subset \tilde{\mathcal{M}}$ (such as \mathcal{I}^+) [406, 265, 208]. Indeed, none of the specific features of \mathcal{I}^+ (i.e. \mathcal{I}^+ lies at the boundary of $\tilde{\mathcal{M}}$ and cannot be intersected by past-directed causal curves) has been used in the proofs of Properties 4.5 to 4.11. These properties are also valid for the boundary $\partial J^+(S)$ of the causal future of any S , modulo the relevant changes future \leftrightarrow past in Property 4.8. A set of the type $\partial J^-(S)$ and $\partial J^+(S)$ is named an **achronal boundary** [265].

Historical note : The properties of generic achronal boundaries, which yield Properties 4.5 to 4.11 in the particular case of a black hole event horizon \mathcal{H} (cf. Remark 6), have been established by Roger Penrose in a seminal lecture at the Battelle-Seattle Center at the summer of 1967, which has been published in 1968 [406]. Penrose used the term *semispace-like boundary* instead of *achronal boundary*. It seems that the latter has been introduced by Stephen Hawking and George Ellis in 1973 in their famous textbook [265].

Chapter 5

Stationary black holes

Contents

5.1	Introduction	119
5.2	Stationary spacetimes	120
5.3	Mass and angular momentum	126
5.4	The event horizon as a Killing horizon	150
5.5	The generalized Smarr formula	160
5.6	The no-hair theorem	169

5.1 Introduction

Having defined black holes in all generality in Chap. 4, we focus here on the steady state case, i.e. black holes in stationary spacetimes. We have already discussed non-expanding horizons and Killing horizons in Chap. 3 as possible models for the event horizon of a steady state black hole. Actually, we shall see below that (each connected component of) the event horizon of a black hole in a stationary spacetime has to be a Killing horizon, at least in the framework of general relativity and within rather general hypotheses. We shall start by defining properly the concept of stationary spacetime and investigating some first properties of a black hole in such a spacetime (Sec. 5.2). Then, in Sec. 5.3, we discuss the concepts of mass and angular momentum in asymptotically flat spacetimes, which are useful to characterize black holes. In Sec. 5.4, we shall see that if the Killing vector ξ generating stationarity is null on a connected event horizon \mathcal{H} , the latter is a Killing horizon with respect to ξ . If \mathcal{H} is non-degenerate and the electrovacuum Einstein equation holds, this can only occur in a static spacetime (*staticity theorem*, Sec. 5.4.1). On the contrary, if ξ is spacelike¹ on some parts of \mathcal{H} , then, modulo the electrovacuum Einstein equation and some additional hypotheses, \mathcal{H} is still a Killing horizon, albeit with respect to a Killing vector distinct from ξ (*strong rigidity theorem*, Sec. 5.4.2). Section 5.5 is devoted to an important relation between various global quantities

¹The timelike case is excluded for an event horizon is a null hypersurface.

characterizing a stationary black hole: the *Smarr formula*. This is the opportunity to investigate electromagnetic fields on the horizon of a stationary black hole, in particular to define the black hole's electric charge and electric potential, both of them being involved in the Smarr formula. The culmination point of this chapter is Sec. 5.6, which presents the famous *no-hair theorem*. Modulo some hypotheses, this theorem stipulates that in 4-dimensional general relativity, all isolated stationary electrovacuum black holes are necessarily Kerr-Newman black holes; in the pure vacuum case (the most relevant one for astrophysics), they are Kerr black holes, to be explored in Part. III.

5.2 Stationary spacetimes

5.2.1 Definitions

A spacetime (\mathcal{M}, g) is called ***stationary*** iff (i) it is invariant under the action of the translation group $(\mathbb{R}, +)$ and (ii) the orbits of the group action (cf. Sec. 3.3.1) are everywhere timelike curves or (ii') (\mathcal{M}, 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 \mathcal{I} . It is equivalent to say that there exists a Killing vector field ξ (the generator of the translation group, cf. Sec. 3.3.1) that is timelike everywhere or at least in the vicinity of \mathcal{I} when there exists a conformal completion. We shall say that (\mathcal{M}, g) is ***strictly stationary*** iff the Killing vector field ξ is timelike in all \mathcal{M} , i.e. iff property (ii) above is fulfilled.

Remark 1: Some authors (e.g. Carter [95]) call *pseudo-stationary* the stationary spacetimes that obey (ii'), keeping the qualifier *stationary* for the strictly stationary case. As we are going to see, when \mathcal{M} contains a black hole, ξ cannot be timelike everywhere, so only *pseudo-stationarity* in Carter's sense is relevant for such spacetimes. Our terminology, namely keeping the qualifier *stationary* even for (ii'), follows that of Choquet-Bruhat [107], Chruściel, Lopes Costa & Heusler [122], Heusler [274] and Wald [504].

The Killing vector field ξ is a priori not unique: the translation group $(\mathbb{R}, +)$ admits the reparametrization $t \mapsto t' = \alpha t$, where α is a nonzero constant, which yields the rescaling $\xi \mapsto \xi' = \alpha^{-1} \xi$ of the Killing vector field [cf. Eq. (3.14)]. When the spacetime admits a conformal boundary \mathcal{I} — which is required if a black hole is present, given the definition (4.37) —, then one selects ξ by demanding that it is future-directed near \mathcal{I} and obeys

$$\xi \cdot \xi \rightarrow -1 \quad \text{near } \mathcal{I}. \quad (5.1)$$

This determines ξ uniquely, since this fixes the rescaling constant α to ± 1 , with only $+1$ being acceptable to keep the time orientation.

If the stationary spacetime is endowed with electromagnetic and/or matter fields, they must respect the stationarity as well. For the electromagnetic field \mathbf{F} (cf. Sec. 1.5.2), this is expressed by the vanishing of the Lie derivative along ξ :

$$\mathcal{L}_\xi \mathbf{F} = 0. \quad (5.2)$$

This condition is similar to Killing-vector property $\mathcal{L}_\xi g = 0$ [Eq. (3.18)].

A notion stronger than stationarity is that of *staticity*:

A spacetime (\mathcal{M}, g) is called **static** iff (i) it is stationary and (ii) the Killing vector field ξ generating the stationary action is orthogonal to a family of hypersurfaces (one says that ξ is **hypersurface-orthogonal**). The spacetime (\mathcal{M}, g) is called **strictly static** iff moreover ξ is timelike in all \mathcal{M} .

Remark 2: The same comment as in Remark 1 can be made: some authors would call *static* only spacetimes that are *strictly static* according to the above definition.

In loose terms, a spacetime is stationary if “nothing changes with time”, while it is static if, in addition, “nothing moves”. A prototype of a stationary spacetime that is not static is a spacetime containing a steadily rotating body, be it a star or a black hole.

5.2.2 Coordinates adapted to stationarity or staticity

In a n -dimensional stationary spacetime (\mathcal{M}, g) , a coordinate system $(x^\alpha) = (x^0, x^1, \dots, x^{n-1})$ is said **adapted to stationarity** iff the coordinate vector ∂_t , where $t := x^0$, coincides with the stationary Killing vector ξ . According to Property 3.7, t is then an ignorable coordinate, i.e. the components $g_{\alpha\beta}$ of the metric tensor with respect to (x^α) obey $\partial g_{\alpha\beta}/\partial t = 0$ [Eq. (3.20)]. It follows that

$$g_{\alpha\beta} = g_{\alpha\beta}(x^1, \dots, x^{n-1}). \quad (5.3)$$

If the spacetime (\mathcal{M}, g) is static, a coordinate system $(x^\alpha) = (x^0 = t, x^1, \dots, x^{n-1})$ is said **adapted to staticity** iff it is adapted to stationarity and the hypersurfaces $t = \text{const}$ are orthogonal to the static Killing vector ξ . Given that a normal 1-form to the hypersurfaces $t = \text{const}$ is dt , the coordinates (x^α) are adapted to staticity iff

$$\xi = \partial_t \quad \text{and} \quad \underline{\xi} = W dt, \quad (5.4)$$

where $\underline{\xi}$ is the metric dual of ξ (cf. Sec. A.3.3) and $W := g(\xi, \xi) = \langle \xi, \xi \rangle$. The orthogonality of ∂_t and the hypersurfaces $t = \text{const}$ translates to $g_{0i} = 0$ for $i \in \{1, \dots, n-1\}$, the $g_{\alpha\beta}$'s being the metric components with respect to the coordinates (x^α) . Hence one may write the metric of a static spacetime of dimension n as

$$g = W dt^2 + g_{ij} dx^i dx^j, \quad (5.5)$$

where the indices (i, j) range in $\{1, \dots, n-1\}$ and W and g_{ij} are functions of (x^1, \dots, x^{n-1}) only. It is clear that the metric (5.5) is invariant² in the transformation $t \mapsto -t$. One says that a static spacetime is **time-reflection symmetric**.

Example 1 (staticity of anti-de Sitter spacetime): The anti-de Sitter spacetime AdS_4 considered in Example 18 of Chap. 3 is strictly static, with the coordinates $(\tau, r, \theta, \varphi)$ being adapted to staticity (hence their name: *global static*). Indeed, the metric (3.49) is of the form (5.5) with $t = \tau$ and $W = -\ell^2(1 + r^2)$. The strict staticity follows from $W < 0$, which shows that the Killing vector $\xi = \partial_t$ is everywhere timelike.

²Would (5.5) have contained a non-vanishing $g_{0i} dt dx^i$ term, this would not have been the case.

Example 2 (staticity of Schwarzschild spacetime): The Schwarzschild spacetime, introduced in Example 3 of Chap. 2, is static, but not strictly static. The staticity is however not obvious from the metric components given by Eq. (2.5) because the coordinates (t, r, θ, φ) used there are not adapted to staticity: the Killing vector $\xi = \partial_t$ is not orthogonal to the hypersurfaces $t = \text{const}$, given that $g_{tr} \neq 0$. We shall introduce coordinates adapted to staticity in Chap. 6, namely the *Schwarzschild-Droste coordinates*. That the Schwarzschild spacetime is not strictly static can be read directly on Eq. (2.5): $\xi \cdot \xi = g_{tt} = -1 + 2m/r$, so that ξ is timelike only in the region $r > 2m$.

5.2.3 Black holes in stationary spacetimes

Let us consider a spacetime (\mathcal{M}, g) that *contains a black hole*, as defined in Sec. 4.4.2. In particular, (\mathcal{M}, g) admits a future infinity \mathcal{I}^+ . Furthermore, we assume that (\mathcal{M}, g) is *stationary*, as defined in Sec. 5.2.1. Since (\mathcal{M}, g) is invariant under the action of the isometry group $(\mathbb{R}, +)$, so is \mathcal{I}^+ (under some proper extension of ξ to the conformal completion $\tilde{\mathcal{M}}$) and therefore its causal past $J^-(\mathcal{I}^+)$. Since the event horizon \mathcal{H} is the boundary of $J^-(\mathcal{I}^+)$ inside \mathcal{M} [Eq. (4.39)], we get

Property 5.1: stationary event horizon

The event horizon \mathcal{H} of a black hole in a stationary spacetime is globally invariant under the action of the stationary group $(\mathbb{R}, +)$.

The word *globally* stresses that \mathcal{H} is invariant *as a whole*, not that each point of \mathcal{H} is a fixed point of the group action. Let us assume that \mathcal{H} is smooth (this sounds likely in a stationary context; a rigorous proof can be found in Ref. [119]); it is then a null hypersurface (Property 4.11 in Sec. 4.4.3). Now, \mathcal{H} is globally invariant if, and only if, the generator ξ of the isometry group is tangent to \mathcal{H} . Since a timelike vector cannot be tangent to a null hypersurface (cf. Lemma 2.3 in Sec. 2.3.4), we conclude:

Property 5.2: stationary Killing vector tangent to the event horizon

In a stationary spacetime containing a black hole, the stationary Killing vector field ξ is tangent to the event horizon \mathcal{H} , which implies that ξ is either null or spacelike on \mathcal{H} . Let \mathcal{H}_0 be a connected component of \mathcal{H} ($\mathcal{H}_0 = \mathcal{H}$ if \mathcal{H} is connected). If ξ is null on all \mathcal{H}_0 , one says that \mathcal{H}_0 is **non-rotating**, while if ξ is spacelike on some part of \mathcal{H}_0 , one says that \mathcal{H}_0 is **rotating**.

Since ξ cannot be timelike on \mathcal{H} , it follows immediately that a stationary spacetime containing a black hole cannot be strictly stationary, according to the definition given in Sec. 5.2.1. We shall discuss in detail the two cases – null or spacelike – allowed for ξ on \mathcal{H} in Sec. 5.4.

It is rather intuitive that the event horizon \mathcal{H} of a black hole in a stationary spacetime must have a vanishing expansion $\theta_{(\ell)}$ along its null normals ℓ . Indeed, if $\theta_{(\ell)}$ were nonzero, the area of cross-sections would vary when dragged along ℓ and this would define some “evolution”

along \mathcal{H} (e.g. from small areas to larger ones), which would break the invariance of \mathcal{H} under the action of the stationary group. This of course results from \mathcal{H} being part of the global spacetime structure, since not any null hypersurface in a stationary spacetime has a vanishing expansion: for instance, a future light cone in Minkowski spacetime (which is stationary!) has $\theta_{(\ell)} > 0$ (cf. Example 17 on p. 45), but the light cone is not invariant by any time translation isometry. The rigorous proof that $\theta_{(\ell)} = 0$ for stationary event horizons can be found in Hawking & Ellis' textbook (Proposition 9.3.1 of Ref. [265]). Here, we shall simply state:

Property 5.3: non-expanding horizons for stationary black holes

The event horizon \mathcal{H} of a black hole in a stationary spacetime is a null hypersurface of vanishing expansion:

$$\theta_{(\ell)} = 0. \quad (5.6)$$

Moreover, if the cross-sections \mathcal{S} of \mathcal{H} are closed manifolds such that \mathcal{H} has the topology $\mathbb{R} \times \mathcal{S}$, \mathcal{H} is a *non-expanding horizon*, according to the definition given in Sec. 3.2.

In dimension 4, one can strongly constrain the topology of the horizon cross-sections:

Property 5.4: topology theorem 1 (Hawking 1972 [259])

Let (\mathcal{M}, g) be a 4-dimensional stationary spacetime containing a black hole of event horizon \mathcal{H} . Let \mathcal{S} be a connected component of a complete cross-section of \mathcal{H} . Let us assume that (i) \mathcal{S} is closed (compact without boundary) and orientable; (ii) the null dominance condition (3.43) is fulfilled on \mathcal{H} for some scalar field $f \geq 0$ [for general relativity with $f = \Lambda$ this is equivalent to assuming $\Lambda \geq 0$ and the dominant null energy condition (3.46)] and (iii) when displaced into the black hole exterior along $-\mathbf{k}$ (the opposite of the ingoing null normal \mathbf{k} to \mathcal{S}), \mathcal{S} becomes a surface with $\theta_{(\ell)} > 0$. In particular, condition (iii) holds if $\theta_{(\mathbf{k})} < 0$ and there is no trapped surface (cf. Sec. 3.2.3) in the black hole exterior. Then the cross-section \mathcal{S} has *generically* the topology of the 2-sphere (i.e. \mathcal{S} is *homeomorphic* to \mathbb{S}^2). The non-generic case is that of \mathcal{S} having the topology of the 2-torus $\mathbb{T}^2 = \mathbb{S}^1 \times \mathbb{S}^1$; this can occur only under special circumstances, among which the metric induced by g on \mathcal{S} must be flat.

Proof. Let ℓ be a future-directed null normal to \mathcal{H} and \mathbf{k} a complementary future-directed null vector field normal to \mathcal{S} , normalized as in Eq. (2.33), i.e. $\mathbf{k} \cdot \ell = -1$. At each point $p \in \mathcal{S}$, the pair (\mathbf{k}, ℓ) is a basis of the 2-plane $T_p^\perp \mathcal{S}$ orthogonal to \mathcal{S} (cf. Fig. 2.10), with ℓ tangent to \mathcal{H} and \mathbf{k} transverse to it. Moreover \mathbf{k} points towards the black hole interior, otherwise null geodesics leaving \mathcal{H} along \mathbf{k} would enter $J^-(\mathcal{I}^+)$. As in Sec. 2.3, let us consider that \mathcal{H} is the level set $u = 0$ of 1-parameter family of hypersurfaces $(\mathcal{H}_u)_{u \in \mathbb{R}}$. This extends ℓ in the vicinity of \mathcal{H} via Eq. (2.11) as a null vector field normal to each \mathcal{H}_u . By Property 5.3, we have $\theta_{(\ell)} = 0$ on \mathcal{S} . Let us displace \mathcal{S} by a small parameter $\epsilon > 0$ along $-\mathbf{k}$. The expansion along ℓ of the obtained surface is positive by hypothesis (iii). By taking the limit $\epsilon \rightarrow 0$, we form the derivative $\mathcal{L}_{-\mathbf{k}} \theta_{(\ell)}$, which must obey $\mathcal{L}_{-\mathbf{k}} \theta_{(\ell)} \geq 0$ since $\theta_{(\ell)} = 0$ on \mathcal{S} and $\theta_{(\ell)} > 0$ on the

displaced surface. Given that $\mathcal{L}_{-\mathbf{k}} \theta_{(\ell)} = -\mathcal{L}_{\mathbf{k}} \theta_{(\ell)}$, we see that hypothesis (iii) implies

$$\mathcal{L}_{\mathbf{k}} \theta_{(\ell)} \leq 0. \quad (5.7)$$

A standard identity (cf. e.g. Eq. (3b) of [268], Eq. (3.1) of [67], Eq. (5.1) of [80] or Eq. (36) of [296]) expresses $\mathcal{L}_{\mathbf{k}} \theta_{(\ell)}$ as

$$\mathcal{L}_{\mathbf{k}} \theta_{(\ell)} = -\frac{1}{2} \mathcal{R} - \mathcal{D}_a \Omega^a + \Omega_a \Omega^a + \mathbf{G}(\ell, \mathbf{k}), \quad (5.8)$$

where \mathcal{R} is the Ricci scalar of the Riemannian metric \mathbf{q} on \mathcal{S} induced by the spacetime metric \mathbf{g} , \mathcal{D} is the Levi-civita connection associated to \mathbf{q} , \mathbf{G} is the Einstein tensor of the spacetime metric \mathbf{g} and Ω is the 1-form on \mathcal{S} that is the pullback of the 1-form ω defined by Eq. (2.75). We may also view Ω as a spacetime 1-form, defined as the composition of ω with the orthogonal projector $\overrightarrow{\mathbf{q}}$ onto \mathcal{S} : $\Omega = \omega \circ \overrightarrow{\mathbf{q}}$. In view of Eq. (2.77), we may then write $\langle \Omega, \mathbf{v} \rangle = -\mathbf{k} \cdot \nabla_{\overrightarrow{\mathbf{q}}(\mathbf{v})} \ell$, or in index notation, $\Omega_\alpha = -k_\mu \nabla_\nu \ell^\mu q^\nu_\alpha$. Besides, using Eq. (3.43), we have $\mathbf{G}(\ell, \mathbf{k}) = \overrightarrow{\mathbf{G}}(\ell) \cdot \mathbf{k} = -\mathbf{W} \cdot \mathbf{k} - f \ell \cdot \mathbf{k} = -\mathbf{W} \cdot \mathbf{k} + f$. Then, integrating (5.8) over the compact manifold \mathcal{S} and setting the integral of the divergence term $\mathcal{D}_a \Omega^a$ to zero since \mathcal{S} has no boundary, we get

$$\underbrace{\frac{1}{2} \int_{\mathcal{S}} \mathcal{R} \sqrt{q} d^2x}_{2\pi\chi} = \int_{\mathcal{S}} \left[\underbrace{\Omega_a \Omega^a}_{\geq 0} - \underbrace{\mathbf{W} \cdot \mathbf{k}}_{\geq 0} + \underbrace{f}_{\geq 0} - \underbrace{\mathcal{L}_{\mathbf{k}} \theta_{(\ell)}}_{\geq 0} \right] \sqrt{q} d^2x. \quad (5.9)$$

In the left-hand side, we have invoked the Gauss-Bonnet theorem to express the integral of \mathcal{R} in terms of the Euler characteristic χ of the surface \mathcal{S} , using the fact that in dimension 2, the Ricci scalar \mathcal{R} is twice the Gaussian curvature. The signs of the terms of the integrand in the right-hand side are justified as follows: $\Omega_a \Omega^a = q^{ab} \Omega_a \Omega_b \geq 0$ because \mathbf{q} is a Riemannian metric, $-\mathbf{W} \cdot \mathbf{k} \geq 0$ by Lemma 1.2 in Sec. 1.2.2, given that \mathbf{k} is a future-directed null vector and \mathbf{W} is a future-directed causal vector thanks to the null dominance condition (3.43), $f \geq 0$ by hypothesis and $-\mathcal{L}_{\mathbf{k}} \theta_{(\ell)} \geq 0$ follows from Eq. (5.7). Now, the Euler characteristic χ is a topological invariant, which is 2 for the sphere \mathbb{S}^2 , 0 for the torus \mathbb{T}^2 and $-2(g-1)$ for a connected orientable surface of genus g , i.e. with g “holes”. Equation (5.9) yields $\chi \geq 0$. The only possibilities for \mathcal{S} compact, connected and orientable are \mathbb{S}^2 ($\chi = 2$) and \mathbb{T}^2 ($\chi = 0$). However the torus case is very special. Indeed, setting $\chi = 0$ in Eq. (5.9) implies that each term in the integrand of the right-hand side vanishes separately: $\Omega_a \Omega^a = 0$, $\mathbf{W} \cdot \mathbf{k} = 0$, $f = 0$ and $\mathcal{L}_{\mathbf{k}} \theta_{(\ell)} = 0$. The last condition is the marginal case in the inequality (5.7). Moreover, since \mathbf{q} is positive definite, $\Omega_a \Omega^a = 0$ implies $\Omega = 0$, which in turn implies $\mathcal{D}_a \Omega^a = 0$. In addition, $\mathbf{G}(\ell, \mathbf{k}) = -\mathbf{W} \cdot \mathbf{k} + f = 0$. We then deduce immediately from Eq. (5.8) that $\mathcal{R} = 0$, i.e. the Ricci scalar of \mathbf{q} is identically zero. Given that the Riemann curvature tensor of a 2-dimensional metric is proportional to its Ricci scalar [cf. Eq. (A.112)], it follows that \mathbf{q} is a flat metric. \square

Remark 3: The topology theorem 1, as stated above, is slightly different from the original version given by Hawking [259, 260, 265], for it adds the “outermost” hypothesis (iii). In Hawking’s version, (iii) is deduced from the non-existence of trapped surfaces in the black hole exterior $J^-(\mathcal{I}^+)$, the latter property being proved by assumming that the spacetime is globally hyperbolic (Proposition 9.2.8 in

Ref. [265]). Another difference with Hawking's version, as stated in Proposition 9.3.2 of Ref. [265], is that the torus topology is not totally excluded by Property 5.4. However, as discussed in the historical note below, it seems that the exclusion of the torus in Hawking's proof requires additional hypotheses, which are not explicitly stated in Refs. [259, 265].

Remark 4: The topology theorem 1 is not specific to stationary black hole event horizons, for it relies only on quasilocal properties. Indeed, the proof does not rely on \mathcal{H} being the boundary of a black hole region, if one agrees to define the “interior” region as that pointed towards by k . The only required property is \mathcal{H} being a hypersurface (not even a null one) sliced by spacelike compact surfaces \mathcal{S} with $\theta_{(\ell)} = 0$ and $\mathcal{L}_k \theta_{(\ell)} \leq 0$. The theorem is therefore valid for *outer trapping horizons* [268] and (tubes of) *apparent horizons*, as noticed by Hawking himself [260] (p. 34), which are not necessarily null hypersurfaces and which exist in non-stationary spacetimes. Both concepts of *outer trapping horizon* and *apparent horizon* will be discussed in Chap. 18.

Another version of the topology theorem relies on the null convergence condition, which is weaker than the null dominance condition (hypothesis (ii) above), and replaces hypothesis (iii) by other ones, which are more global. It also fully excludes the 2-torus topology:

Property 5.5: topology theorem 2 (Chrúsciel & Wald 1994 [129])

Let (\mathcal{M}, g) be a 4-dimensional asymptotically flat stationary spacetime containing a black hole of event horizon \mathcal{H} . Let us assume that (i) the null convergence condition (2.94) is fulfilled, (ii) the domain of outer communications $\langle\langle \mathcal{M} \rangle\rangle$ [Eq. (4.43)] is globally hyperbolic, (iii) $\langle\langle \mathcal{M} \rangle\rangle$ contains an achronal asymptotically flat hypersurface Σ that intersects \mathcal{H} in a compact cross-section \mathcal{S} , and (iv) some technical condition is fulfilled (cf. [129] for the details). Then, $\langle\langle \mathcal{M} \rangle\rangle$ is simply connected and any connected component of \mathcal{S} is homeomorphic to the sphere S^2 .

A part \mathcal{U} of spacetime is said *globally hyperbolic* iff it admits a *Cauchy surface*, i.e. a spacelike hypersurface Σ such that every inextendible timelike curve of \mathcal{U} intersects Σ exactly once (cf. Sec. 8.3 of Wald's textbook [498] for more details). We shall not give the proof of Property 5.5 here; it can of course be found in the original article by Chrúsciel & Wald [129].

Remark 5: The topology theorems 1 and 2 regard only a *connected component* of a given horizon complete cross-section. Generally, the latter is a connected 2-manifold, but there exist 4-dimensional stationary (actually static) spacetimes containing a black hole, the event horizon of which has disconnected complete cross-sections: the *Majumdar-Papapetrou spacetimes* [352, 398, 257]. They are solutions of the electrovacuum Einstein equation (cf. Sec. 1.5.2) representing an arbitrary number of charged black holes, which form a static configuration thanks to an exact balance between the gravitational attraction and the electrostatic repulsion. The Majumdar-Papapetrou black holes will be discussed further in Sec. 5.6.1 [cf. Eq. (5.108)].

A generalization of the topology theorem 1 to spacetimes of dimension $n > 4$ has been obtained in 2006 by Galloway & Schoen [209] and Rácz provided a simplified proof of it in 2008 [420]. However, the theorem for $n > 4$ only says that some invariant of the smooth structure of

\mathcal{S} , called the *Yamabe invariant*, must be positive for $f \geq 0$ ⁽³⁾. This result implies that \mathcal{S} must admit metrics of positive scalar curvature; it is less stringent about the topology of \mathcal{S} than the theorem for $n = 4$. Actually, the higher n , the less constraints on the topology are provided by the Yamabe invariant. For instance, for $n = 5$, the cross-sections of the *Myers-Perry black holes* [375, 186, 424], which generalize Kerr black holes to $n > 4$, have the topology of the sphere \mathbb{S}^3 , but the topology $\mathbb{S}^1 \times \mathbb{S}^2$ is allowed as well, as demonstrated by the *black ring* solution found by Emparan & Reall [185, 186, 424] (see also Sec. 5.3 of Ref. [118]).

Historical note : The topology theorem for the spacetime dimension $n = 4$ has been formulated first by Stephen Hawking in 1992 [259] (see also p. 34 of Ref. [260] and Proposition 9.3.2 in Hawking & Ellis' textbook [265]). However, in 1993, Gregory Galloway [207] (p. 119) pointed out some limitation in Hawking's proof, namely that it cannot exclude the torus topology ($\chi = 0$) for the horizon's cross-sections without any extra hypothesis. For instance, for the proof given in Ref. [259], one shall require that the spacetime is *analytic* (cf. Remark 4 in Sec. A.2.1), which is a rather strong hypothesis. The proof presented above is based on that given by Sean Hayward in 1994 [268] for outer trapping horizons (see also the proof of Theorem 6.3 in Ref. [380]). The theorem obtained by Piotr Chruściel and Robert Wald in 1994 [129] (Property 5.5) relies on the *topological censorship theorem* established by John Friedman, Kristin Schleich and Donald Witt in 1993 [198]. It leads directly to the spherical topology, excluding the toroidal one.

5.3 Mass and angular momentum

For an asymptotically flat stationary spacetime, containing a black hole or not, there is a well-defined concept of mass: the *Komar mass*, which we introduce here (Secs. 5.3.3–5.3.5). For axisymmetric spacetimes, which are relevant for stationary rotating black holes, there is in addition the concept of *Komar angular momentum*, which we shall introduce in Sec. 5.3.6.

5.3.1 Mass and angular momentum of weakly relativistic stationary systems

We shall call **weakly relativistic** a n -dimensional spacetime (\mathcal{M}, g) such that \mathcal{M} is diffeomorphic to \mathbb{R}^n and the metric tensor can be expressed as

$$g = f + h, \quad (5.10)$$

where f is a flat Lorentzian metric on \mathcal{M} and h is small in the following sense: the components of h obey $|h_{\alpha\beta}| \ll 1$ in any **f -Minkowskian coordinates**, i.e. coordinates (x^α) on \mathcal{M} such that⁴ $(f_{\alpha\beta}) = \eta := \text{diag}(-1, 1, \dots, 1)$. In addition, we assume that g is ruled by the Einstein equation (1.40) with $\Lambda = 0$ and an energy momentum tensor T of compact support.

³In brief, Eq. (5.9) holds for $n > 4$ as well, except that the integral of the Ricci scalar \mathcal{R} is no longer proportional to the Euler characteristic of \mathcal{S} (no Gauss-Bonnet theorem for $\dim \mathcal{S} \neq 2$!) but is related to the Yamabe constant of the conformal class of the metric g .

⁴We keep the notation η for the matrix $\text{diag}(-1, 1, \dots, 1)$, so that $(\eta_{\alpha\beta})$ stands for the components of f in Minkowskian coordinates only.

Given the flat metric \mathbf{f} , \mathbf{f} -Minkowskian coordinates are not unique: any Poincaré transformation $\tilde{x}^\alpha = \Lambda_\mu^\alpha x^\mu + c^\alpha$, where (Λ_μ^α) is a Lorentz matrix and the c^α 's are constant, leads to coordinates (\tilde{x}^α) which are \mathbf{f} -Minkowskian as well. Furthermore, the flat metric \mathbf{f} itself, and hence the decomposition (5.10), is highly nonunique. Indeed any change of coordinates of the form $x'^\alpha = x^\alpha + \zeta^\alpha(x^0, \dots, x^{n-1})$ where (x^α) are \mathbf{f} -Minkowskian coordinates and ζ^α are infinitesimal functions, leads to the following components of the metric tensor with respect to (x'^α) :

$$g'_{\alpha\beta} = g_{\mu\nu} \frac{\partial x^\mu}{\partial x'^\alpha} \frac{\partial x^\nu}{\partial x'^\beta} = (\eta_{\mu\nu} + h_{\mu\nu}) (\delta_\alpha^\mu - \partial_\alpha \zeta^\mu) (\delta_\beta^\nu - \partial_\beta \zeta^\nu),$$

where we have used $x^\alpha = x'^\alpha - \zeta^\alpha$ and $\partial \zeta^\alpha / \partial x'^\beta = \partial \zeta^\alpha / \partial x^\sigma \times \partial x^\sigma / \partial x'^\beta \simeq \partial \zeta^\alpha / \partial x^\beta =: \partial_\beta \zeta^\alpha$ to the first order in ζ^α . Expanding to the first order in ζ^α and \mathbf{h} , we get

$$g'_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta} - \partial_\alpha \zeta_\beta - \partial_\beta \zeta_\alpha,$$

where $\zeta_\alpha := \eta_{\alpha\mu} \zeta^\mu$. We may recast the above expression as $\mathbf{g} = \mathbf{f}' + \mathbf{h}'$, where \mathbf{f}' is the metric whose components in the coordinates (x'^α) are $\eta_{\alpha\beta}$ (hence \mathbf{f}' is flat, as⁵ \mathbf{f}) and \mathbf{h}' has the following components with respect to the coordinates (x'^α) :

$$h'_{\alpha\beta} = h_{\alpha\beta} - \partial_\alpha \zeta_\beta - \partial_\beta \zeta_\alpha. \quad (5.11)$$

The above relation can be viewed as expressing some *gauge freedom* on \mathbf{h} . This freedom actually reflects the freedom in the choice of the flat background metric \mathbf{f} in the decomposition (5.10).

Property 5.6: Lorenz gauge for the metric perturbation

The gauge freedom (5.11) can be used to ensure that, in terms of \mathbf{f} -Minkowskian coordinates (x^α) , the metric perturbation \mathbf{h} fulfills

$$\boxed{\partial_\mu (\eta^{\mu\nu} \bar{h}_{\alpha\nu}) = 0}, \quad (5.12)$$

where

$$\bar{\mathbf{h}} := \mathbf{h} - \frac{1}{2} h \mathbf{f}, \quad (5.13)$$

h standing for the trace of \mathbf{h} with respect to \mathbf{f} : $h := \eta^{\mu\nu} h_{\mu\nu}$. The choice (5.12) is referred to as the *Lorenz gauge* (sometimes *Hilbert gauge* [463]).

Proof. From Eq. (5.11), we get $h' := \eta^{\mu\nu} h'_{\mu\nu} = h - 2\partial_\mu \zeta^\mu$, so that $\bar{h}'_{\alpha\beta} = \bar{h}_{\alpha\beta} - \partial_\alpha \zeta_\beta - \partial_\beta \zeta_\alpha + \partial_\mu \zeta^\mu \eta_{\alpha\beta}$. It follows then that $\partial_\mu (\eta^{\mu\nu} \bar{h}'_{\alpha\nu}) = \partial_\mu (\eta^{\mu\nu} \bar{h}_{\alpha\nu}) - \square_f \zeta_\alpha$, where $\square_f := \eta^{\mu\nu} \partial_\mu \partial_\nu$ is the d'Alembertian operator with respect to \mathbf{f} . Accordingly, if the Lorenz gauge (5.12) is not fulfilled, it suffices to solve the d'Alembert equation

$$\square_f \zeta_\alpha = \partial_\mu (\eta^{\mu\nu} \bar{h}_{\alpha\nu}) \quad (5.14)$$

and plug the solution ζ_α into Eq. (5.11) to get a metric perturbation \mathbf{h}' that obeys the Lorenz gauge. \square

⁵Note that the components of \mathbf{f} with respect to the coordinates (x'^α) are not $\eta_{\alpha\beta}$ but $\eta_{\alpha\beta} - \partial_\alpha \zeta_\beta - \partial_\beta \zeta_\alpha$.

Remark 1: The Lorenz gauge (5.12) is equivalent to the first-order expansion in \hbar of the relation defining **harmonic coordinates** on \mathcal{M} , i.e.

$$\square_g x^\alpha = 0 \iff \partial_\mu (\sqrt{-g} g^{\mu\alpha}) = 0. \quad (5.15)$$

Remark 2: The Lorenz gauge does not fully specify the pair (f, \hbar) . Indeed, the solutions of the d'Alembert equation (5.14) are nonunique: they depend on initial and boundary data.

A standard computation (see e.g. Chap. 18 of [370] or Chap. 5 of [463], noticing that the computation is independent of the spacetime dimension n) shows that, at first order in \hbar and in the Lorenz gauge, the Einstein tensor of g is

$$G = -\frac{1}{2} \square_f \bar{h}, \quad (5.16)$$

where \square_f stands for the d'Alembertian operator relative to the metric f : in f -Minkowskian coordinates (x^α) , $\square_f \bar{h}_{\alpha\beta} = \eta^{\mu\nu} \partial_\mu \partial_\nu \bar{h}_{\alpha\beta}$.

Let us now assume that (\mathcal{M}, g) is a stationary spacetime, with Killing vector ξ . We may then choose f so that the f -Minkowskian coordinates (x^α) are adapted to stationarity, i.e. x^0 is an ignorable coordinate, or equivalently $\partial_0 = \xi$. Let then Σ be a hypersurface $x^0 = \text{const}$ and γ the metric induced by f on Σ . The coordinates⁶ $(x^i)_{1 \leq i \leq n-1}$ form a Cartesian coordinate system of (Σ, γ) : $\gamma = \delta_{ij} dx^i \otimes dx^j = (dx^1)^2 + \dots + (dx^{n-1})^2$. The operator \square_f reduces to the Laplace operator of γ , Δ say, and, thanks to property (5.16), the Einstein equation (1.40) becomes a system of $n(n+1)/2$ independent Poisson equations:

$$\Delta \bar{h}_{\alpha\beta} = -16\pi T_{\alpha\beta}. \quad (5.17)$$

The solutions are obtained via the Green function of the $(n-1)$ -dimensional Laplace operator:

$$\bar{h}_{\alpha\beta}(x) = \frac{16\pi}{(n-3)\Omega_{n-2}} \int_{\Sigma} \frac{T_{\alpha\beta}(x')}{|x - x'|^{n-3}} d^{n-1}x', \quad (5.18)$$

where $x := (x^1, \dots, x^{n-1})$, $x' := (x'^1, \dots, x'^{n-1})$, $|x - x'|^2 := \sum_{i=1}^{n-1} (x^i - x'^i)^2$ and Ω_{n-2} is the area of the unit sphere \mathbb{S}^{n-2} ; the latter is given by the formula

$$\Omega_p = \frac{2\pi^{(p+1)/2}}{\Gamma((p+1)/2)}, \quad \text{with} \quad \Gamma(u) := \int_0^{+\infty} t^{u-1} e^{-t} dt, \quad (5.19)$$

so that

$$\Omega_2 = 4\pi, \quad \Omega_3 = 2\pi^2, \quad \Omega_4 = \frac{8}{3}\pi^2, \quad \Omega_5 = \pi^3, \dots \quad (5.20)$$

Property 5.7: asymptotic metric of a weakly relativistic system

Let (\mathcal{M}, g) be a weakly relativistic stationary spacetime of dimension $n \geq 4$, with g obeying the Einstein equation (1.40) with $\Lambda = 0$ and an energy momentum tensor T of compact support (the “source”), where the energy-density dominates over the spatial stresses (weakly relativistic matter). Within the Lorenz gauge, there exists a coordinate system (x^α) such that

⁶Latin indices i, j, k, \dots range in $\{1, \dots, n-1\}$, while Greek ones range in $\{0, \dots, n-1\}$.

the metric tensor has the following behavior when $r := \sqrt{(x^1)^2 + \dots + (x^{n-1})^2} \rightarrow +\infty$:

$$g_{00} = -1 + \alpha_n \frac{M}{r^{n-3}} + \mathcal{O}\left(\frac{1}{r^{n-1}}\right) \quad (5.21a)$$

$$g_{0i} = \frac{n-2}{2} \alpha_n \frac{J_{ij} x^j}{r^{n-1}} + \mathcal{O}\left(\frac{1}{r^{n-1}}\right) \quad (5.21b)$$

$$g_{ij} = \left(1 + \frac{\alpha_n}{n-3} \frac{M}{r^{n-3}}\right) \delta_{ij} + \mathcal{O}\left(\frac{1}{r^{n-2}}\right), \quad (5.21c)$$

where $\alpha_n := 16\pi/((n-2)\Omega_{n-2})$,

$$M := \int_{\Sigma} T_{00}(x) d^{n-1}x \quad (5.22)$$

and

$$J_{ij} := \int_{\Sigma} (x^j T_{0i}(x) - x^i T_{0j}(x)) d^{n-1}x, \quad (5.23)$$

Σ being any hypersurface $x^0 = \text{const}$. The quantities M and J_{ij} are independent of Σ (i.e. of x^0) and are called respectively the **mass** and the **angular momentum** of the central source. The coordinates (x^α) correspond to the *central source rest-frame* and take their origin at the *center of mass*, in the sense that

$$\int_{\Sigma} T_{0i}(x) d^{n-1}x = 0 \quad \text{and} \quad \int_{\Sigma} x^i T_{00}(x) d^{n-1}x = 0 \quad (5.24)$$

The dominance of the energy-density over the spatial stresses is expressed in terms of the components of \mathbf{T} with respect to the coordinates (x^α) as $T_{00} \gg |T_{ij}|$.

Proof. Let (f, h) obeys the Lorenz gauge and (x^α) be some corresponding f -Minkowskian coordinates. By a Poincaré transformation (cf. p. 127), one can enforce (5.24). Far from the source, one may expand the term $1/|x - x'|^{n-3}$ in Eq. (5.18) in powers of x'^i/r :

$$\frac{1}{|x - x'|^{n-3}} = \frac{1}{r^{n-3}} \left(1 + (n-3) \frac{x^j}{r} \frac{x'^j}{r} + \mathcal{O}\left(\frac{|x'|^2}{r^2}\right)\right),$$

where Einstein's summation convention is assumed on the repeated index $j \in \{1, \dots, n-1\}$. Accordingly, Eq. (5.18) yields

$$\bar{h}_{\alpha\beta}(x) = \frac{16\pi}{\Omega_{n-2}} \left(\frac{1}{(n-3)r^{n-3}} \int_{\Sigma} T_{\alpha\beta}(x') d^{n-1}x' + \frac{x^j}{r^{n-1}} \int_{\Sigma} x'^j T_{\alpha\beta}(x') d^{n-1}x' \right) + \mathcal{O}\left(\frac{1}{r^{n-1}}\right). \quad (5.25)$$

For the component 00, the first integral in the right-hand side is nothing but M , as defined by (5.22), while the second integral vanishes due to the second equation in (5.24). Hence, we get

$$\bar{h}_{00}(x) = \frac{16\pi}{(n-3)\Omega_{n-2}} \frac{M}{r^{n-3}} + \mathcal{O}\left(\frac{1}{r^{n-1}}\right). \quad (5.26)$$

Regarding the component $0i$ of Eq. (5.25), the first integral in the right-hand side vanishes due to the first equation in (5.24). There remains then

$$\bar{h}_{0i}(x) = \frac{16\pi}{\Omega_{n-2} r^{n-1}} \int_{\Sigma} x'^j T_{0i}(x') d^{n-1}x' + \mathcal{O}\left(\frac{1}{r^{n-1}}\right) = \frac{8\pi}{\Omega_{n-2}} \frac{J_{ij} x^j}{r^{n-1}} + \mathcal{O}\left(\frac{1}{r^{n-1}}\right), \quad (5.27)$$

where the second equality follows from the identity

$$J_{ij} = 2 \int_{\Sigma} x^j T_{0i}(x) d^{n-1}x. \quad (5.28)$$

To prove it, consider

$$\partial_k(x^i x^j T^{0k}) = \delta^i_k x^j T^{0k} + x^i \delta^j_k T^{0k} + x^i x^j \underbrace{\partial_k T^{0k}}_0 = x^j T^{0i} + x^i T^{0j},$$

where $\partial_k T^{0k} = 0$ results from the equation of energy-momentum conservation (1.45) specialized to stationary spacetimes and expressed at 0th order in \mathbf{h} . Integrating over Σ and invoking the Gauss-Ostrogradsky theorem to set the integral of the divergence term $\partial_k(x^i x^j T^{0k})$ to zero (since \mathbf{T} has compact support), we get

$$\int_{\Sigma} x^j T^{0i}(x) d^{n-1}x + \int_{\Sigma} x^i T^{0j}(x) d^{n-1}x = 0.$$

In view of the definition (5.23) of J_{ij} , the identity (5.28) follows, since $T^{0i} = -T_{0i}$ at the 0th order in \mathbf{h} .

Let us now reconstruct \mathbf{h} from $\bar{\mathbf{h}}$. Taking the trace of Eq. (5.13) with respect to \mathbf{f} yields $\bar{\mathbf{h}} := \eta^{\mu\nu} \bar{h}_{\mu\nu} = h - (h/2) \times n = (2-n)h/2$. Hence we may invert Eq. (5.13) to

$$\mathbf{h} = \bar{\mathbf{h}} - \frac{\bar{h}}{n-2} \mathbf{f} \quad \text{with} \quad \bar{h} = -\bar{h}_{00} + \bar{h}_{ii}. \quad (5.29)$$

The part⁷ \bar{h}_{ii} of the trace \bar{h} is deduced from Eq. (5.25):

$$\bar{h}_{ii}(x) = \frac{16\pi}{\Omega_{n-2}} \left(\frac{1}{(n-3)r^{n-3}} \int_{\Sigma} T_{ii}(x') d^{n-1}x' + \frac{x^j}{r^{n-1}} \int_{\Sigma} x'^j T_{ii}(x') d^{n-1}x' \right) + \mathcal{O}\left(\frac{1}{r^{n-1}}\right).$$

The second integral vanishes identically, as it can be seen from the identity

$$\begin{aligned} \partial_i \left(x^k x^j T^{ik} - \frac{1}{2} r^2 T^{ij} \right) &= \delta^k_i x^j T^{ik} + x^k \delta^j_i T^{ik} + x^k x^j \underbrace{\partial_i T^{ik}}_0 - r \underbrace{\partial_i r}_r T^{ij} - \frac{1}{2} r^2 \underbrace{\partial_i T^{ij}}_0 \\ &= x^j T^{ii} = x^j T_{ii}, \end{aligned}$$

where $\partial_i T^{ij} = 0$ follows from the energy-momentum conservation law (1.45). Hence the integral of $x^j T_{ii}$ over Σ is that of the divergence of a vector field that vanishes outside the source and the Gauss-Ostrogradsky theorem allows one to set it to zero. We are thus left with

$$\bar{h}_{ii}(x) = \frac{16\pi}{(n-3)\Omega_{n-2}} \frac{1}{r^{n-3}} \int_{\Sigma} T_{ii}(x') d^{n-1}x' + \mathcal{O}\left(\frac{1}{r^{n-1}}\right). \quad (5.30)$$

⁷Note that Einstein's summation convention is used: $\bar{h}_{ii} = \bar{h}_{11} + \cdots + \bar{h}_{n-1,n-1}$.

Gathering Eqs. (5.26) and (5.30), we get

$$\bar{h} = \frac{16\pi}{(n-3)\Omega_{n-2}} \frac{1}{r^{n-3}} \left(-M + \int_{\Sigma} T_{ii}(x') d^{n-1}x' \right) + \mathcal{O}\left(\frac{1}{r^{n-1}}\right).$$

However, given expression (5.22) for M , the weakly relativistic matter condition $T_{00} \gg |T_{ij}|$ implies that the integral of T_{ii} over Σ is negligible in front of M . We conclude that

$$\bar{h} = -\bar{h}_{00} + \mathcal{O}\left(\frac{1}{r^{n-1}}\right).$$

It follows then from Eq. (5.29) that, up to terms decaying at least as $1/r^{n-1}$,

$$\begin{aligned} h_{00} &= \bar{h}_{00} + \frac{\bar{h}_{00}}{n-2} \times (-1) = \frac{n-3}{n-2} \bar{h}_{00} \\ h_{0i} &= \bar{h}_{0i} \\ h_{ij} &= \bar{h}_{ij} + \frac{\bar{h}_{00}}{n-2} \delta_{ij} = \frac{\bar{h}_{00}}{n-2} \delta_{ij} + \mathcal{O}\left(\frac{1}{r^{n-2}}\right). \end{aligned}$$

The last equality holds because the $1/r^{n-3}$ term in \bar{h}_{ij} , which is proportional to the integral of T_{ij} over Σ according to Eq. (5.25), is negligible in front of the $1/r^{n-3}$ term in \bar{h}_{00} , given that $T_{00} \gg |T_{ij}|$. The above three equations, along with the values (5.26) and (5.27) for \bar{h}_{00} and \bar{h}_{0i} , establish formulas (5.21) for the components $g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}$ of the metric tensor. \square

5.3.2 Asymptotic expression of the metric in stationary asymptotically flat spacetimes

Let us now consider a generic (not necessarily weakly relativistic) n -dimensional spacetime (\mathcal{M}, g) that is stationary and asymptotically flat, with g being ruled by the Einstein equation with $\Lambda = 0$ and with \mathbf{T} vanishing in the asymptotic region. The decomposition $g = f + h$, with f flat and h “small” [Eq. (5.10)] still holds in some neighborhood \mathcal{U} of infinity. Moreover, one can still choose the pair (f, h) such that the Lorenz gauge (5.12) is fulfilled and the linearized Einstein equation in \mathcal{U} reduces to $\Delta \bar{h} = 0$ [Eq. (5.17) with $\mathbf{T} = 0$]. There exists then a f -Minkowskian coordinate system (x^α) on \mathcal{U} such that the solution for g is given by Eq. (5.21) to the lowest order in $1/r$, the difference being that the constants M and J_{ij} can no longer be expressed by integrals of the energy-momentum tensor, as in Eqs. (5.22) and (5.23). This can be shown rigorously by using the expansion in powers of $1/r$ of the generic solution to $\Delta f = 0$ and the Lorenz gauge (to set some terms to zero), cf. Exercice 19.3 of MTW [370] or Sec. 5.7 of Straumann’s textbook [463] for details. For a configuration that is not assumed to be weakly relativistic, one should not limit oneself to the linearized Einstein equation. However, going beyond the first order expansion in h does not spoil the first terms in the expansions (5.21) of g ’s components. This simply adds terms with a higher power of $1/r$. For instance, for $n = 4$, the second order expansion in h introduces the term $-2M^2/r^2$ in g_{00} (see the above references for details, as well as Refs. [64, 417] for the nonlinear expansion within a post-Newtonian framework). We may then assert:

Property 5.8: asymptotic metric of a stationary asymptotically flat spacetime

Let (\mathcal{M}, g) be a stationary asymptotically flat spacetime of dimension $n \geq 4$, with g obeying the Einstein equation (1.40) with $\Lambda = 0$ and with the energy-momentum tensor T vanishing in the asymptotic region. Then there exists a coordinate system (x^α) in the asymptotic region such that the metric tensor has the following behavior when $r := \sqrt{(x^1)^2 + \cdots + (x^{n-1})^2} \rightarrow +\infty$:

$$g_{00} = -1 + \alpha_n \frac{M}{r^{n-3}} + \mathcal{O}\left(\frac{1}{r^{n-2}}\right) \quad (5.31a)$$

$$g_{0i} = \frac{n-2}{2} \alpha_n \frac{J_{ij} x^j}{r^{n-1}} + \mathcal{O}\left(\frac{1}{r^{n-1}}\right) \quad (5.31b)$$

$$g_{ij} = \left(1 + \frac{\alpha_n}{n-3} \frac{M}{r^{n-3}}\right) \delta_{ij} + \mathcal{O}\left(\frac{1}{r^{n-2}}\right), \quad (5.31c)$$

where

$$\alpha_n := \frac{16\pi}{(n-2)\Omega_{n-2}}, \quad (5.32)$$

(Ω_{n-2} being the area of the unit sphere \mathbb{S}^{n-2} , cf. Eqs. (5.19)-(5.20)), so that

$$\alpha_4 = 2, \quad \alpha_5 = \frac{8}{3\pi}, \quad \alpha_6 = \frac{3}{2\pi}, \quad \alpha_7 = \frac{16}{5\pi^2}, \quad (5.33)$$

M is a constant and (J_{ij}) is a constant antisymmetric $(n-1) \times (n-1)$ matrix. Furthermore, by a suitable $\text{SO}(n-1)$ transformation of the coordinates $(x^i)_{1 \leq i \leq n-1}$, (J_{ij}) can be brought to the following block diagonal form, depending only on $p := [(n-1)/2]$ numbers $J_{(1)}, \dots, J_{(p)}$ (possibly equal to zero):

$$J_{ij} = \begin{pmatrix} 0 & J_{(1)} & & & \\ -J_{(1)} & 0 & & & \\ & & 0 & J_{(2)} & \\ & & -J_{(2)} & 0 & \\ & & & & \ddots \end{pmatrix}, \quad (5.34)$$

with the last row and the last column containing only zeros if n is even.

The last assertion follows from the fact that any real antisymmetric matrix J is similar to a matrix J' of the type (5.34) via $J' = PJP^{-1}$, where P is a special orthogonal matrix.

Note that the difference between (5.31) and the asymptotic expansion (5.21) for a weakly relativistic system is $\mathcal{O}(1/r^{n-2})$ in Eq. (5.31a) versus $\mathcal{O}(1/r^{n-1})$ in Eq. (5.21a). As discussed above, this results from nonlinear terms in the expansion of the Einstein equation.

5.3.3 Komar mass

In Newtonian gravity, the mass M of an isolated body is defined similarly to Eq. (5.22), namely by the integral of the mass density $\rho \sim T_{00}$ over the body. An alternative formula for M is provided by **Gauss's law**: M is $-(4\pi)^{-1}$ times the flux of the gravitational field $\mathbf{g} = -\vec{\nabla}\Phi$ through any closed surface \mathcal{S} surrounding the body, namely

$$M = \frac{1}{4\pi} \int_{\mathcal{S}} \vec{\nabla}\Phi \cdot d\mathbf{S}, \quad (5.35)$$

where Φ is the Newtonian gravitational and $d\mathbf{S}$ is the area element vector normal to \mathcal{S} . The above formula is easily derived when \mathcal{S} is a sphere of large radius, using spherical coordinates (r, θ, φ) . Indeed, for large r , one has $\Phi \sim -M/r$, so that $\vec{\nabla}\Phi \sim (M/r^2) \partial_r$. Given that $d\mathbf{S} = r^2 \sin \theta d\theta d\varphi \partial_r$ and $\partial_r \cdot \partial_r = 1$, formula (5.35) follows. This formula is actually not restricted to a remote sphere, but is valid for any closed surface \mathcal{S} surrounding the body (this follows from the Gauss-Ostrogradsky theorem and Laplace's equation, $\Delta\Phi = 0$, which holds outside the central body). Gauss's law (5.35) reflects the “gravitating aspect” of the mass, while the volume integral (5.22) identifies the mass with the “amount of matter” constituting the central body. The latter definition would yield $M = 0$ for any vacuum spacetime (set $\mathbf{T} = 0$ in Eq. (5.22)) and therefore cannot be used to generalize the concept of mass to strongly relativistic systems. In particular, one would expect $M > 0$ for the gravitating mass of Schwarzschild and Kerr black holes, which are vacuum solutions of the Einstein equation.

Accordingly, Gauss's law (5.35), rather than the volume integral (5.22), is the good basis for any attempt to generalize the concept of mass to strongly relativistic spacetimes. Taking a look at the asymptotic expression (5.21) of the metric tensor of a weakly relativistic stationary system, we notice that the mass M appears as the coefficient of the dominant $1/r^{n-3}$ term in the expansion of g_{00} , so that one could recover it by considering the flux integral of $\partial g_{00}/\partial r$ over a $(n-2)$ -dimensional sphere at large r . In order to have a coordinate-invariant definition, it is more appropriate to consider the 1-form $\underline{\xi}$ metric-dual to the stationary Killing vector ξ . Indeed, in any coordinate system adapted to stationarity, i.e. such that the components of ξ are $\xi^\alpha = (1, 0, \dots, 0)$, the components of $\underline{\xi}$ are $\xi_\alpha = g_{0\alpha}$, so that in particular $\xi_0 = g_{00}$. Instead of the partial derivative $\partial g_{00}/\partial r$, a natural coordinate-independent quantity is then the exterior derivative $d\underline{\xi}$ (cf. Sec. A.4.3). Since the concept of flux is naturally conveyed by the integral of the Hodge dual (see e.g. Sec. 16.4.7 of Ref. [227]), one arrives at the definition of mass given below [Eq. (5.36)], named *Komar mass*. In addition to be coordinate-invariant, it shares the same property as the Newtonian mass (5.35), namely to be independent from the integration surface outside the central body, at least within general relativity (Property 5.15 below).

Let $(\mathcal{M}, \mathbf{g})$ be an asymptotically flat spacetime of dimension $n \geq 4$ that is stationary, with stationary Killing vector ξ , normalized such that $\xi \cdot \xi \rightarrow -1$ near the asymptotic boundary [Eq. (5.1)]. Given a spacelike closed $(n-2)$ -surface $\mathcal{S} \subset \mathcal{M}$, the **Komar mass over \mathcal{S}** is defined by

$$M_{\mathcal{S}} := -\frac{n-2}{16\pi(n-3)} \int_{\mathcal{S}} \star(d\underline{\xi}), \quad (5.36)$$

where

- (i) $\underline{\xi}$ is the 1-form associated to ξ by metric duality (cf. Sec. A.3.3), i.e. the 1-form of components $\xi_\alpha = g_{\alpha\mu}\xi^\mu$;
- (ii) $d\underline{\xi}$ is the exterior derivative of $\underline{\xi}$ (cf. Sec. A.4.3, especially Eqs. (A.90b) and (A.91b)), namely the 2-form whose components are

$$(d\underline{\xi})_{\alpha\beta} = \partial_\alpha\xi_\beta - \partial_\beta\xi_\alpha = \nabla_\alpha\xi_\beta - \nabla_\beta\xi_\alpha = 2\nabla_\alpha\xi_\beta, \quad (5.37)$$

the last equality following from the Killing equation (3.19);

- (iii) $\star(d\underline{\xi})$ is the $(n-2)$ -form that is the Hodge dual of the 2-form $d\underline{\xi}$. The **Hodge dual** of any p -form A is defined^a as the $(n-p)$ -form $\star A$ given by

$$\star A_{\alpha_1\dots\alpha_{n-p}} := \frac{1}{p!} A^{\mu_1\dots\mu_p} \epsilon_{\mu_1\dots\mu_p\alpha_1\dots\alpha_{n-p}}, \quad (5.38)$$

where ϵ is the Levi-Civita tensor associated with the metric g (cf. Sec. A.3.4);

- (iv) the orientation of \mathcal{S} (which is required to define the integral of $\star(d\underline{\xi})$ over \mathcal{S}) is set by the $(n-2)$ -form^b

$$\mathcal{S}\epsilon := \epsilon(\mathbf{n}, \mathbf{s}, \underbrace{\dots}_{n-2 \text{ slots}}), \quad (5.39)$$

where \mathbf{n} is any future-directed unit timelike vector field normal to \mathcal{S} and \mathbf{s} is a unit spacelike vector field normal to \mathbf{n} , orthogonal to \mathbf{n} and directed towards the “exterior” of \mathcal{S} in the following sense: assume that \mathcal{S} can be embedded into a spacelike hypersurface Σ that extends towards the asymptotic flat end of (\mathcal{M}, g) ; then for any point $p \in \mathcal{S}$, there exists a curve entirely lying in Σ , starting from p with tangent vector $\mathbf{s}|_p$ and extending to the asymptotic flat end without crossing \mathcal{S} again.

^aSee e.g. Sec. 14.6 of Ref. [463] or Sec. 14.5 of Ref. [227] for an introduction to Hodge duality.

^bSee Sec. A.3.4 for the definition of an oriented manifold.

We shall check below (Property 5.11) that the Komar mass (5.36) gives the coefficient M that appears in the asymptotic expansion (5.31) of the metric tensor.

Remark 3: As the integral of a $(n-2)$ -form over an oriented $(n-2)$ -dimensional manifold, formula (5.36) is well posed. More precisely, on the $(n-2)$ -dimensional manifold \mathcal{S} , the theory of integration is defined for $(n-2)$ -forms *on* \mathcal{S} , while $\star(d\underline{\xi})$ is a $(n-2)$ -form *on* \mathcal{M} . However, any $(n-2)$ -form ω on \mathcal{M} yields canonically to a unique $(n-2)$ -form $\iota^*\omega$ on the submanifold \mathcal{S} , by restricting the action of ω at each point $p \in \mathcal{S}$ to vectors tangent to \mathcal{S} . One says that $\iota^*\omega$ is the *pullback* of ω to \mathcal{S} via the embedding ι of \mathcal{S} in \mathcal{M} (cf. Sec. A.2.8). The pullback operator is thus implicit in Eq. (5.36).

Remark 4: The numerical prefactor $-(n-2)/(16\pi(n-3))$ in the definition (5.36) is adjusted so that for weakly relativistic isolated objects, the Komar mass coincides with the volume integral of the matter

energy density, as it will be shown in Property 5.11. For $n = 4$ (the standard spacetime dimension), this prefactor is $-1/(8\pi)$, while for $n = 5$, it becomes $-3/(32\pi)$.

Remark 5: The Komar mass is not defined for $n \leq 3$. In particular, formula (5.36) is ill-posed for $n = 3$. Already in Newtonian gravity, the very concept of gravitational mass is well defined only for $n \geq 4$. Indeed, in spherical symmetry, a solution to the Poisson equation outside the sources ($\Delta\Phi = 0$) that decays with the radius r exists only for $n \geq 4$: this solution is⁸ $\Phi = -K/r^{n-3}$ where the constant K is proportional to the mass M of the central source. For $n = 3$, the solutions of $\Delta\Phi = 0$ are $\Phi = a \ln r + b$ (a and b being constant), while for $n = 2$, they are $\Phi = ar + b$, none of them decaying as $r \rightarrow +\infty$.

Remark 6: We have given above a heuristic motivation for defining the Komar mass by Eq. (5.36). We shall see in Chap. 17 a more fundamental approach, which makes the Komar mass appear as the *Noether charge* associated with the invariance of the Einstein-Hilbert Lagrangian under the diffeomorphisms generated by the vector field ξ (cf. Property 17.9).

In view of Eq. (5.38) with $p = 2$ and the last equality in Eq. (5.37), the Komar mass formula (5.36) can be written as

$$M_{\mathcal{S}} = -\frac{n-2}{16\pi(n-3)} \int_{\mathcal{S}} \nabla^{\mu} \xi^{\nu} \epsilon_{\mu\nu\alpha_1\dots\alpha_{n-2}}, \quad (5.40)$$

where $\nabla^{\mu} \xi^{\nu} \epsilon_{\mu\nu\alpha_1\dots\alpha_{n-2}}$ stands for the $(n-2)$ -form ω defined by

$$\omega(\mathbf{u}_1, \dots, \mathbf{u}_{n-2}) = \nabla^{\mu} \xi^{\nu} \epsilon_{\mu\nu\alpha_1\dots\alpha_{n-2}} u_1^{\alpha_1} \cdots u_{n-2}^{\alpha_{n-2}}$$

for any $(n-2)$ -tuple of vector fields $(\mathbf{u}_1, \dots, \mathbf{u}_{n-2})$ tangent to \mathcal{S} (cf. Remark 3 above).

Instead of integrals of $(n-2)$ -forms *along* the $(n-2)$ -surface \mathcal{S} , as in (5.36) and (5.40), one may express the Komar mass as a *flux integral*, i.e. the integral of a 2-form contracted with some “area element”, which is *normal* to \mathcal{S} . Let us introduce the latter first.

Property 5.9: area element normal bivector to a codimension-2 surface

Given a spacelike $(n-2)$ -dimensional surface $\mathcal{S} \subset \mathcal{M}$, the **area element normal bivector to \mathcal{S}** is the infinitesimal antisymmetric type-(2,0) tensor field defined on \mathcal{S} by

$$d\mathbf{S} := (\mathbf{s} \wedge \mathbf{n}) \sqrt{q} d^{n-2}x, \quad (5.41)$$

or, in index notation,

$$dS^{\alpha\beta} := (s^\alpha n^\beta - n^\alpha s^\beta) \sqrt{q} d^{n-2}x, \quad (5.42)$$

where

- (i) \mathbf{n} and \mathbf{s} are unit timelike and spacelike vector fields normal to \mathcal{S} obeying the same properties as in item (iv) of the above definition of the Komar mass; in particular at each point $p \in \mathcal{S}$, $(\mathbf{n}|_p, \mathbf{s}|_p)$ is an orthonormal basis of the timelike plane $T_p^{\perp}\mathcal{S}$ normal to \mathcal{S} (cf. Fig. ??);

⁸This is easy to get since in the Euclidean space of dimension $n-1$ the Laplace operator is $\Delta\Phi = \frac{1}{r^{n-2}} \frac{d}{dr} (r^{n-2} \frac{d\Phi}{dr})$ for spherically symmetric fields $\Phi(r)$.

- (ii) $\mathbf{s} \wedge \mathbf{n}$ stands for the **exterior product** of \mathbf{s} by \mathbf{n} : $\mathbf{s} \wedge \mathbf{n} := \mathbf{s} \otimes \mathbf{n} - \mathbf{n} \otimes \mathbf{s}$ (hence formula (5.42));
- (iii) $d^{n-2}x := dx^1 \cdots dx^{n-2}$, where $(x^a) = (x^1, \dots, x^{n-2})$ is a coordinate system on \mathcal{S} such that $(\partial_1, \dots, \partial_{n-2})$ is a right-handed basis for the orientation of \mathcal{S} defined by $\mathcal{S}\epsilon$ [Eq. (5.39)], i.e. $\epsilon(\mathbf{n}, \mathbf{s}, \partial_1, \dots, \partial_{n-2}) > 0$;
- (iv) $q := \det(q_{ab})$ is the determinant w.r.t. (x^a) of the metric \mathbf{q} induced on \mathcal{S} by the spacetime metric \mathbf{g} .

The bivector $d\mathbf{S}$ is independent of the choice of the coordinates (x^a) on \mathcal{S} and of the orthonormal basis (\mathbf{n}, \mathbf{s}) of $T_p^\perp \mathcal{S}$.

Proof. The combination $\sqrt{q} d^{n-2}x = \sqrt{q} dx^1 \cdots dx^{n-2}$ with $q := \det(q_{ab})$ is the volume (area) element of the Riemannian manifold $(\mathcal{S}, \mathbf{q})$; it is thus invariant in any change of coordinates $(x^a) \mapsto (x'^a)$. Besides, any two orthonormal bases (\mathbf{n}, \mathbf{s}) and $(\mathbf{n}', \mathbf{s}')$ of $T_p^\perp \mathcal{S}$ having the same orientation and with \mathbf{n} and \mathbf{n}' both future-directed are necessarily related by a 2-dimensional Lorentz boost (cf. Fig. ??):

$$\begin{cases} \mathbf{n}' = \cosh \psi \mathbf{n} + \sinh \psi \mathbf{s} \\ \mathbf{s}' = \sinh \psi \mathbf{n} + \cosh \psi \mathbf{s}, \end{cases}$$

where $\psi \in \mathbb{R}$ is the boost rapidity. It follows immediately that $\mathbf{s}' \wedge \mathbf{n}' = \sinh^2 \psi \mathbf{n} \wedge \mathbf{s} + \cosh^2 \psi \mathbf{s} \wedge \mathbf{n} = (\cosh^2 \psi - \sinh^2 \psi) \mathbf{s} \wedge \mathbf{n} = \mathbf{s} \wedge \mathbf{n}$, which shows that the bivector $\mathbf{s} \wedge \mathbf{n}$, and hence $d\mathbf{S}$, does not depend on the choice of the orthonormal basis (\mathbf{n}, \mathbf{s}) . \square

The area element normal bivector $d\mathbf{S}$ appears in the following useful lemma.

Lemma 5.10: flux integral of a 2-form

For any 2-form \mathbf{A} defined in the vicinity of a spacelike $(n-2)$ -dimensional surface \mathcal{S} , one has

$$\int_{\mathcal{S}} \star \mathbf{A} = \frac{1}{2} \int_{\mathcal{S}} A_{\mu\nu} dS^{\mu\nu}, \quad (5.43)$$

where the $(n-2)$ -form $\star \mathbf{A}$ is the Hodge dual of \mathbf{A} , as defined by Eq. (5.38) with $p = 2$.

Proof. Using the definition (5.38) with $p = 2$, we have

$$\int_{\mathcal{S}} \star \mathbf{A} = \frac{1}{2} \int_{\mathcal{S}} A^{\mu\nu} \epsilon_{\mu\nu\alpha_1 \dots \alpha_{n-2}} = \frac{1}{2} \int_{\mathcal{S}} \mathbf{A}^\sharp(e^{(\mu)}, e^{(\nu)}) \epsilon(e_{(\mu)}, e_{(\nu)}, dx_1, \dots, dx_{n-2}),$$

where \mathbf{A}^\sharp is the tensor field of components $A^{\alpha\beta}$, $(e_{(\alpha)})$ is an orthonormal tetrad such that $e_{(0)} = \mathbf{n}$ and $e_{(1)} = \mathbf{s}$, $(e^{(\alpha)})$ is its dual cobasis and dx_1, \dots, dx_{n-2} are displacement vectors forming elementary parallelograms on \mathcal{S} ; for instance $dx_a = dx^a \partial_a$ (no summation on a) for $a \in \{1, \dots, n-2\}$. Note that the last equality in the above expression results from the very definition of the integral of a $(n-2)$ -form over a $(n-2)$ -surface. Given the definition of

the tetrad $(\mathbf{e}_{(\alpha)}), (\mathbf{e}_{(2)}, \dots, \mathbf{e}_{(n-1)})$ is a basis of the tangent space $T_p\mathcal{S}$; consequently $d\mathbf{x}_1, \dots, d\mathbf{x}_{n-2}$ are linear combinations of $\mathbf{e}_{(2)}, \dots, \mathbf{e}_{(n-1)}$. Thanks to the alternate character of ϵ , we may then restrict the sum over the indices μ and ν to $(\mu, \nu) = (0, 1)$ and $(\mu, \nu) = (1, 0)$. Hence

$$\int_{\mathcal{S}} \star \mathbf{A} = \int_{\mathcal{S}} \mathbf{A}^\sharp(\mathbf{e}^{(0)}, \mathbf{e}^{(1)}) \epsilon(\mathbf{e}_{(0)}, \mathbf{e}_{(1)}, d\mathbf{x}_1, \dots, d\mathbf{x}_{n-2}) = \int_{\mathcal{S}} A^{(0)(1)} \epsilon(\mathbf{n}, \mathbf{s}, d\mathbf{x}_1, \dots, d\mathbf{x}_{n-2}).$$

Now, since $(\mathbf{e}_{(\alpha)})$ 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

$$A_{(0)(1)} = \mathbf{A}(\mathbf{e}_{(0)}, \mathbf{e}_{(1)}) = \mathbf{A}(\mathbf{n}, \mathbf{s}) = A_{\mu\nu} n^\mu s^\nu = -A_{\mu\nu} s^\mu n^\nu = -\frac{1}{2} A_{\mu\nu} (s^\mu n^\nu - n^\mu s^\nu).$$

On the other side, $\epsilon(\mathbf{n}, \mathbf{s}, \dots)$, which appears above, is the $(n-2)$ -form \mathcal{E} defined by Eq. (5.39). Actually, it is nothing but the volume (area) form of $(\mathcal{S}, \mathbf{q})$, i.e. the Levi-Civita tensor of the metric \mathbf{q} (see e.g. Sec. 16.4.3 of Ref. [227]). In particular, \mathcal{E} is independent of the choice of the orthonormal basis (\mathbf{n}, \mathbf{s}) of $T_p^\perp \mathcal{S}$. Hence we may write, for $d\mathbf{x}_a = dx^a \partial_a$,

$$\epsilon(\mathbf{n}, \mathbf{s}, d\mathbf{x}_1, \dots, d\mathbf{x}_{n-2}) = \mathcal{E}_{a_1 \dots a_{n-2}} dx_1^{a_1} \dots dx_{n-2}^{a_{n-2}} = \sqrt{q} dx^1 \dots dx^{n-2}.$$

Gathering the above results and using Eq. (5.42) establishes Eq. (5.43). \square

Thanks to Lemma 5.10, we may re-express the Komar mass (5.36) as

$$M_{\mathcal{S}} = -\frac{n-2}{32\pi(n-3)} \int_{\mathcal{S}} (\mathbf{d}\underline{\xi})_{\mu\nu} dS^{\mu\nu}. \quad (5.44)$$

Using Eq. (5.37), this becomes

$$\boxed{M_{\mathcal{S}} = -\frac{n-2}{16\pi(n-3)} \int_{\mathcal{S}} \nabla_\mu \xi_\nu dS^{\mu\nu}}. \quad (5.45)$$

Alternatively, we may express the exterior derivative in terms of partial derivatives (first equality in Eq. (5.37)) and get

$$(\mathbf{d}\underline{\xi})_{\mu\nu} dS^{\mu\nu} = (\partial_\mu \xi_\nu - \partial_\nu \xi_\mu)(s^\mu n^\nu - n^\mu s^\nu) \sqrt{q} d^{n-2}x = 2\partial_\mu \xi_\nu (s^\mu n^\nu - n^\mu s^\nu) \sqrt{q} d^{n-2}x.$$

Equation (5.44) yields then an expression of $M_{\mathcal{S}}$ that can be used for explicit computations:

$$\boxed{M_{\mathcal{S}} = -\frac{n-2}{16\pi(n-3)} \int_{\mathcal{S}} \partial_\mu \xi_\nu (s^\mu n^\nu - n^\mu s^\nu) \sqrt{q} d^{n-2}x}. \quad (5.46)$$

Example 3 (Komar mass in Schwarzschild spacetime): Let us consider the Schwarzschild spacetime $(\mathcal{M}, \mathbf{g})$ introduced in Example 3 of Sec. 2.3. In terms of the ingoing Eddington-Finkelstein coordinates (t, r, θ, φ) used there, the stationary Killing vector is simply $\xi = \partial_t$. We have $n = 4$ and let us choose

\mathcal{S} to be a 2-sphere defined by $(t, r) = \text{const}$. The coordinates $(x^\alpha) = (x^1, x^2)$ spanning \mathcal{S} are then (θ, φ) . For the pair of normal vectors to \mathcal{S} , we choose $\mathbf{n} = N^{-1}\partial_t - 2mN/r\partial_r$ and $\mathbf{s} = N\partial_r$, where $N := (1 + 2m/r)^{-1/2}$. Given the metric (2.5), it is easy to check that $\mathbf{g}(\mathbf{n}, \mathbf{n}) = -1$, $\mathbf{g}(\mathbf{n}, \mathbf{s}) = 0$ and $\mathbf{g}(\mathbf{s}, \mathbf{s}) = 1$. It is also immediate that $\mathbf{g}(\mathbf{n}, \partial_\theta) = \mathbf{g}(\mathbf{n}, \partial_\varphi) = 0$ and $\mathbf{g}(\mathbf{s}, \partial_\theta) = \mathbf{g}(\mathbf{s}, \partial_\varphi) = 0$, which proves that \mathbf{n} and \mathbf{s} are normal to \mathcal{S} . Moreover, \mathbf{s} points towards the exterior of \mathcal{S} (since $N > 0$) and $(\mathbf{n}, \mathbf{s}, \partial_\theta, \partial_\varphi)$ is a right-handed basis. All the hypotheses stated below Eq. (5.42) are thus fulfilled and we may use Eq. (5.46) to evaluate the Komar mass over \mathcal{S} . The components ξ_α of ξ are easily evaluated via the components (2.5) of \mathbf{g} : $\xi_\alpha = g_{\alpha\mu}\xi^\nu = g_{\alpha\mu}(\partial_t)^\mu = g_{\alpha t} = (-1 + 2m/r, 2m/r, 0, 0)$. The only non-zero derivatives $\partial_\mu\xi_\nu$ are then $\partial_r\xi_t$ and $\partial_r\xi_r$; they are both equal to $-2m/r^2$. It follows that the integrand in Eq. (5.46) is reduced to only two terms:

$$\partial_\mu\xi_\nu(s^\mu n^\nu - n^\mu s^\nu) = \partial_r\xi_t(s^r n^t - n^r s^t) = -\frac{2m}{r^2}(N \times N^{-1} + 2mN \times 0) = -\frac{2m}{r^2}.$$

We read on Eq. (2.5) that the metric induced on \mathcal{S} is $\mathbf{q} = r^2 \mathbf{d}\theta^2 + r^2 \sin^2 \theta \mathbf{d}\varphi^2$, so that $\sqrt{q} = r^2 \sin \theta$. Accordingly, Eq. (5.46) leads to

$$M_{\mathcal{S}} = -\frac{1}{8\pi} \int_{\mathcal{S}} \left(-\frac{2m}{r^2} \right) r^2 \sin \theta \, d\theta \, d\varphi = \frac{m}{4\pi} \int_{\mathcal{S}} d\theta \, d\varphi.$$

Since the remaining integral is nothing but 4π , we get simply

$$M_{\mathcal{S}} = m. \quad (5.47)$$

We note that $M_{\mathcal{S}}$ does not depend on r , i.e. does not depend on the choice of the sphere \mathcal{S} .

More generally, we have:

Property 5.11: Komar mass in the asymptotic expression of the metric

The coefficient M that appears in expression (5.31) of the metric tensor of a stationary asymptotically flat n -dimensional spacetime in some asymptotically Minkowskian coordinates (x^α) coincides in the limit $r := \sqrt{(x^1)^2 + \dots + (x^{n-1})^2} \rightarrow +\infty$ with the Komar mass over the sphere \mathcal{S} of constant (x^0, r) . In particular, for a weakly relativistic system, the Komar mass over \mathcal{S} is equal to the volume integral of the matter energy density T_{00} , as given by formula (5.22).

Proof. Without any loss of generality, the coordinate system (x^α) leading to the expansion (5.31) can be chosen to be adapted to stationarity, i.e. such that $\partial_0 = \xi$, where ξ is the stationary Killing vector. Let us evaluate the Komar mass $M_{\mathcal{S}}$ via the integral (5.46), denoting the surface element of \mathcal{S} by $\sqrt{q} d^{n-2}y$ instead of $\sqrt{q} d^{n-2}x$ to avoid any confusion with the asymptotically Minkowskian coordinates (x^α) . We have $\xi_\alpha = g_{\alpha\mu}\xi^\mu = g_{\alpha\mu}\delta^\mu_0 = g_{\alpha 0}$, with $g_{\alpha 0} = g_{0\alpha}$ given by formulas (5.31) for large r , so that

$$\partial_0\xi_\alpha = 0, \quad \partial_i\xi_0 \simeq -(n-3)\alpha_n \frac{M}{r^{n-2}} \frac{x^i}{r} \quad \text{and} \quad \partial_i\xi_j \simeq \frac{n-2}{2}\alpha_n \frac{J_{jk}}{r^{n-1}} \left(\delta_{ki} - (n-1)\frac{x^k x^i}{r^2} \right).$$

On the other hand, for $r \rightarrow +\infty$, we have $s^\alpha \simeq (0, \frac{x^1}{r}, \dots, \frac{x^{n-1}}{r})$, $n^\alpha \simeq (1, 0, \dots, 0)$ and $\sqrt{q} d^{n-2}y = r^{n-2} \sqrt{\bar{q}} d^{n-2}y$, where \bar{q} stands for the determinant of the round metric of the unit

sphere \mathbb{S}^{n-2} with respect to the coordinates (y^1, \dots, y^{n-2}) . Given that $\partial_i \xi_j$ decays as $1/r^{n-1}$, it follows that only the terms involving $\partial_i \xi_0$ contribute to the integral (5.46), which becomes

$$\begin{aligned} M_{\mathcal{S}} &= -\frac{n-2}{16\pi(n-3)} \int_{\mathcal{S}} \partial_i \xi_0 \left(\frac{x^i}{r} \times 1 - 0 \times 0 \right) r^{n-2} \sqrt{\bar{q}} d^{n-2}y \\ &= \frac{n-2}{16\pi} \alpha_n M \underbrace{\int_{\mathcal{S}} \frac{x^i}{r} \frac{x^i}{r} \sqrt{\bar{q}} d^{n-2}y}_{1} = \frac{n-2}{16\pi} \alpha_n M \underbrace{\int_{\mathcal{S}} \sqrt{\bar{q}} d^{n-2}y}_{\Omega_{n-2}}. \end{aligned}$$

Since $\alpha_n \Omega_{n-2} = 16\pi/(n-2)$ [cf. Eq. (5.32)], we get $M_{\mathcal{S}} = M$. The statement about a weakly relativistic system is an immediate consequence of expansion (5.21) being a particular case of (5.31), given that M in (5.21) has the integral form (5.22). \square

Volume integral formula for the Komar mass

We are going to derive a volume integral formula for $M_{\mathcal{S}}$, from which one can show the independence of $M_{\mathcal{S}}$ from the choice of \mathcal{S} in the vacuum case, thereby generalizing the result of Example 3. To this aim, we shall make use of the following lemma:

Lemma 5.12: flux integral of the divergence of a 2-form

In a n -dimensional asymptotically flat spacetime (\mathcal{M}, g) , let Σ be a compact spacelike hypersurface bounded by two closed $(n-2)$ -surfaces: an “internal” one, \mathcal{S}_{int} , and an “external” one, \mathcal{S}_{ext} , the exterior direction being that of the asymptotic flat end of (\mathcal{M}, g) . Let \mathbf{n} be the future-directed unit normal to Σ and dV the **normal volume element vector** of Σ defined by [compare Eq. (5.42)]:

$$dV^\alpha := -n^\alpha \sqrt{\gamma} dx^1 \cdots dx^{n-1} =: -n^\alpha \sqrt{\gamma} d^{n-1}x, \quad (5.48)$$

where (x^1, \dots, x^{n-1}) stands for a coordinate system on Σ and $\gamma := \det(\gamma_{ij})$ is the determinant w.r.t. these coordinates of the metric γ induced on Σ by the spacetime metric g , so that $\sqrt{\gamma} dx^1 \cdots dx^{n-1}$ is the volume element on Σ . Then, for any 2-form A , we have

$$2 \int_{\Sigma} \nabla^\nu A_{\mu\nu} dV^\mu = \int_{\mathcal{S}_{\text{ext}}} A_{\mu\nu} dS^{\mu\nu} - \int_{\mathcal{S}_{\text{int}}} A_{\mu\nu} dS^{\mu\nu}, \quad (5.49)$$

where $dS^{\mu\nu}$ is defined by Eq. (5.42) with s pointing to the asymptotic flat end of (\mathcal{M}, g) for both \mathcal{S}_{ext} and \mathcal{S}_{int} .

Proof. Let (t, x^1, \dots, x^{n-1}) be a local coordinate system adapted to Σ , i.e. such that Σ is the hypersurface $t = 0$, and let \mathbf{V} be the vector defined by $\underline{\mathbf{V}} := \mathbf{A}(., \mathbf{n})$, i.e. $V_\alpha := A_{\alpha\mu} n^\mu$. Thanks to the antisymmetry of \mathbf{A} , \mathbf{V} is tangent to Σ : $\mathbf{n} \cdot \mathbf{V} = \mathbf{A}(\mathbf{n}, \mathbf{n}) = 0$. We have then, using

again the antisymmetry of \mathbf{A} ,

$$\begin{aligned}\nabla^\nu A_{\mu\nu} dV^\mu &= \nabla^\nu A_{\nu\mu} n^\mu \sqrt{\gamma} d^{n-1}x = (\nabla^\nu V_\nu - A_{\nu\mu} \nabla^\nu n^\mu) \sqrt{\gamma} d^{n-1}x \\ &= (\nabla_\mu V^\mu + \underbrace{A_{\nu\mu} K^{\nu\mu}}_0 + \underbrace{A_{\nu\mu} n^\nu D^\mu}_{-V_\mu} \ln N) \sqrt{\gamma} d^{n-1}x \\ &= (\nabla_\mu V^\mu - V^i \partial_i \ln N) \sqrt{\gamma} d^{n-1}x.\end{aligned}$$

In the second line, we have expressed the derivative of \mathbf{n} in terms of the extrinsic curvature tensor \mathbf{K} of Σ and of the lapse function $N := -\mathbf{n} \cdot \partial_t$, via the identity $\nabla^\alpha n^\beta = -K^{\alpha\beta} - D^\beta \ln N n^\alpha$ (see e.g. Eq. (4.24) of Ref. [226]), where $D^\beta \ln N = \nabla^\beta \ln N + n^\beta n^\mu \nabla_\mu \ln N$. To get the third line, we have used the symmetry of \mathbf{K} to set $A_{\nu\mu} K^{\nu\mu} = 0$ and the property $V_\mu n^\mu = 0$ to write $V_\mu D^\mu \ln N = V_\mu \nabla^\mu \ln N = V^\mu \partial_\mu \ln N = V^i \partial_i \ln N$ ($i \in \{1, \dots, n-1\}$). Let us now use formula (A.74) for the divergence of \mathbf{V} , along with $V^0 = 0$ (\mathbf{V} is tangent to Σ) and the relation $\sqrt{-g} = N \sqrt{\gamma}$ between the determinants of the metrics \mathbf{g} and γ (see e.g. Eq. (5.55) of Ref. [226])

$$\begin{aligned}\nabla_\mu V^\mu &= \frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} V^\mu) = \frac{1}{N \sqrt{\gamma}} \partial_i (N \sqrt{\gamma} V^i) \\ &= \frac{1}{\sqrt{\gamma}} \partial_i (\sqrt{\gamma} V^i) + V^i \partial_i \ln N = D_i V^i + V^i \partial_i \ln N,\end{aligned}$$

where $D_i V^i$ is the divergence of \mathbf{V} considered as a vector field on (Σ, γ) , \mathbf{D} standing for the Levi-Civita connection of γ . Hence we get

$$\nabla^\nu A_{\mu\nu} dV^\mu = D_i V^i \sqrt{\gamma} d^{n-1}x.$$

We may then apply the Gauss-Ostrogradsky theorem to \mathbf{V} and write the integral of the divergence of \mathbf{A} as

$$\begin{aligned}\int_\Sigma \nabla^\nu A_{\mu\nu} dV^\mu &= \int_\Sigma D_i V^i \sqrt{\gamma} d^{n-1}x = \int_{\partial\Sigma} \sigma^i V_i \sqrt{q} d^{n-2}x \\ &= \int_{\mathcal{S}_{\text{ext}}} s^i V_i \sqrt{q} d^{n-2}x - \int_{\mathcal{S}_{\text{int}}} s^i V_i \sqrt{q} d^{n-2}x,\end{aligned}$$

where $\boldsymbol{\sigma}$ is the unit vector normal to the boundary $\partial\Sigma$ and outward-pointing with respect to Σ . In the present case, $\partial\Sigma$ has two connected components: \mathcal{S}_{int} and \mathcal{S}_{ext} , with $\boldsymbol{\sigma} = -\mathbf{s}$ on \mathcal{S}_{int} and $\boldsymbol{\sigma} = \mathbf{s}$ on \mathcal{S}_{ext} , hence the last equality. Now, using again the antisymmetry of \mathbf{A} ,

$$s^i V_i = s^\mu V_\mu = s^\mu A_{\mu\nu} n^\nu = \frac{1}{2} A_{\mu\nu} (s^\mu n^\nu - n^\mu s^\nu).$$

In view of expression (5.42) for $dS^{\mu\nu}$, we thus get $2s^i V_i \sqrt{q} d^{n-2}x = A_{\mu\nu} dS^{\mu\nu}$, so that Eq. (5.49) follows. \square

Let us apply Lemma 5.12 to the 2-form $\mathbf{A} := \mathbf{d}\underline{\xi}$, i.e. to the 2-form of components $A_{\alpha\beta} := \nabla_\alpha \xi_\beta - \nabla_\beta \xi_\alpha = 2\nabla_\alpha \xi_\beta$ [cf. Eq. (5.37)]. The 1-form $\nabla^\nu A_{\mu\nu}$ which appears in the left-hand

side of Eq. (5.49) is then, up to a sign, the metric dual of a vector field $\mathcal{J}(\xi)$ called the **Komar current of ξ** : $\nabla^\nu A_{\mu\nu} = -\mathcal{J}(\xi)_\mu$, where

$$\boxed{\mathcal{J}(\xi)^\alpha := \nabla_\mu (\nabla^\mu \xi^\alpha - \nabla^\alpha \xi^\mu).} \quad (5.50)$$

The Komar current is actually a conserved current, independently of the Killing nature of ξ :

Property 5.13: Conservation of the Komar current

For any vector field ξ defined on the spacetime (\mathcal{M}, g) (not necessarily a Killing vector field), the Komar current defined by Eq. (5.50) is conserved:

$$\nabla_\mu \mathcal{J}(\xi)^\mu = 0. \quad (5.51)$$

Proof. We have, from the definition (5.50), $\mathcal{J}(\xi)^\alpha = -\nabla_\nu A^{\alpha\nu}$ with $A^{\alpha\beta} := \nabla^\alpha \xi^\beta - \nabla^\beta \xi^\alpha$. Let (x^α) be a coordinate system on \mathcal{M} . Since A is antisymmetric, its divergence can be expressed in terms of partial derivatives as $\nabla_\nu A^{\alpha\nu} = \partial_\nu (\sqrt{-g} A^{\alpha\nu}) / \sqrt{-g}$, where g is the determinant of g with respect to (x^α) . Similarly the divergence of $\mathcal{J}(\xi)$ can be written as $\nabla_\mu \mathcal{J}(\xi)^\mu = \partial_\mu (\sqrt{-g} \mathcal{J}(\xi)^\mu) / \sqrt{-g}$. Combining the above formulas leads to $\nabla_\mu \mathcal{J}(\xi)^\mu = -\partial_\mu \partial_\nu (\sqrt{-g} A^{\mu\nu}) / \sqrt{-g}$. Since $\partial_\mu \partial_\nu = \partial_\nu \partial_\mu$ and $A^{\mu\nu} = -A^{\nu\mu}$, we get immediately $\nabla_\mu \mathcal{J}(\xi)^\mu = 0$. \square

Thanks to the identities $A_{\alpha\beta} = 2\nabla_\alpha \xi_\beta$ and $\mathcal{J}(\xi)_\alpha = -\nabla^\mu A_{\alpha\mu}$, Lemma 5.12 yields

$$-\int_{\Sigma} \mathcal{J}(\xi)_\mu dV^\mu = \int_{\mathcal{S}_{\text{ext}}} \nabla_\mu \xi_\nu dS^{\mu\nu} - \int_{\mathcal{S}_{\text{int}}} \nabla_\mu \xi_\nu dS^{\mu\nu}.$$

Using expression (5.45) of the Komar mass, we get

$$\boxed{M_{\mathcal{S}_{\text{ext}}} = M_{\mathcal{S}_{\text{int}}} + \frac{n-2}{16\pi(n-3)} \int_{\Sigma} \mathcal{J}(\xi)_\mu dV^\mu.} \quad (5.52)$$

Thanks to the Killing equation (3.19), expression (5.50) for the Komar current $\mathcal{J}(\xi)$ simplifies to $\mathcal{J}(\xi)^\alpha = -2\nabla_\mu \nabla^\alpha \xi^\mu$. Now, since ξ is a Killing vector, Eq. (3.84) holds: $\nabla_\mu \nabla^\alpha \xi^\mu = R^\alpha_\mu \xi^\mu$. Hence we get

$$\boxed{\mathcal{J}(\xi)^\alpha = -2R^\alpha_\mu \xi^\mu.} \quad (5.53)$$

Plugging this relation into formula (5.52) and using expression (5.48) for dV^μ , we arrive at the following property.

Property 5.14: volume integral formula for the Komar mass

Let (\mathcal{M}, g) be an asymptotically flat stationary spacetime of dimension $n \geq 4$, with stationary Killing vector ξ . Let Σ be a compact spacelike hypersurface bounded by two closed $(n-2)$ -surfaces: an “internal” one, \mathcal{S}_{int} , and an “external” one, \mathcal{S}_{ext} , the exterior direction being that of the asymptotic flat end of (\mathcal{M}, g) . The Komar masses over \mathcal{S}_{ext}

and \mathcal{S}_{int} are then related by

$$M_{\mathcal{S}_{\text{ext}}} = M_{\mathcal{S}_{\text{int}}} + \frac{n-2}{8\pi(n-3)} \int_{\Sigma} \mathbf{R}(\boldsymbol{\xi}, \mathbf{n}) \sqrt{\gamma} d^{n-1}x, \quad (5.54)$$

where \mathbf{R} is the Ricci tensor of \mathbf{g} , \mathbf{n} is the future-directed unit normal to Σ and $\sqrt{\gamma} d^{n-1}x$ is the volume element on Σ induced by \mathbf{g} . If \mathbf{g} obeys the Einstein equation (1.42) with some energy-momentum tensor \mathbf{T} and $\Lambda = 0$, then the above formula may be rewritten as

$$M_{\mathcal{S}_{\text{ext}}} = M_{\mathcal{S}_{\text{int}}} + \frac{n-2}{n-3} \int_{\Sigma} \left(\mathbf{T}(\boldsymbol{\xi}, \mathbf{n}) - \frac{T}{n-2} \boldsymbol{\xi} \cdot \mathbf{n} \right) \sqrt{\gamma} d^{n-1}x, \quad (5.55)$$

where $T := \text{tr}_{\mathbf{g}} \mathbf{T} = g^{\mu\nu} T_{\mu\nu}$.

Remark 7: If Σ is a compact spacelike hypersurface bounded by a single closed $(n-2)$ -surface, \mathcal{S}_{ext} say, formulas (5.54) and (5.55) still hold, provided that $M_{\mathcal{S}_{\text{int}}}$ is replaced by zero. This follows directly from Eq. (5.49) with the integral on \mathcal{S}_{int} set to zero, since the boundary of Σ , which appears in the Gauss-Ostrogradsky theorem, is only made of \mathcal{S}_{ext} .

Example 4 (Komar mass of a static star): We may use formula (5.55) to check that at the Newtonian limit, the Komar mass over a surface \mathcal{S}_{ext} surrounding a star reduces to the volume integral of the matter mass density, i.e. to the classical expression of the mass in Newtonian gravity. To this aim, we set $n = 4$ and assume that the spacetime is strictly static, i.e. that $\boldsymbol{\xi}$ is timelike and hypersurface-orthogonal in all \mathcal{M} . A static fluid star is then defined by the energy-momentum tensor \mathbf{T} being a perfect fluid one: $\mathbf{T} = (\rho + p)\mathbf{u} \otimes \mathbf{u} + pg$, where ρ is the fluid-comoving energy density, p is the fluid pressure and the fluid 4-velocity \mathbf{u} is collinear to the static Killing field $\boldsymbol{\xi}$:

$$\mathbf{u} = \frac{1}{V} \boldsymbol{\xi} \quad \text{with} \quad V := \sqrt{-\boldsymbol{\xi} \cdot \boldsymbol{\xi}},$$

the value of V being determined to ensure $\mathbf{u} \cdot \mathbf{u} = -1$. Let us choose for Σ a hypersurface orthogonal to $\boldsymbol{\xi}$ with an outer boundary \mathcal{S}_{ext} but no inner boundary \mathcal{S}_{int} . Since Σ is orthogonal to $\boldsymbol{\xi}$, we have necessarily $\mathbf{n} = \mathbf{u}$, so that $\mathbf{T}(\boldsymbol{\xi}, \mathbf{n}) = VT(\mathbf{u}, \mathbf{u}) = V\rho$ and $\boldsymbol{\xi} \cdot \mathbf{n} = V\mathbf{u} \cdot \mathbf{u} = -V$. Given that $T = 3p - \rho$, formula (5.55) with $n = 4$ and $M_{\mathcal{S}_{\text{int}}} = 0$ (cf. Remark 7 above) results in

$$M_{\mathcal{S}_{\text{ext}}} = 2 \int_{\Sigma} \left(V\rho - \frac{3p - \rho}{2}(-V) \right) \sqrt{\gamma} d^3x = \int_{\Sigma} V(\rho + 3p) \sqrt{\gamma} d^3x.$$

Now, for a weakly relativistic star, $|p| \ll \rho$ and in coordinates (t, x^i) such that $\boldsymbol{\xi} = \partial_t$, we have $V = \sqrt{-\boldsymbol{\xi} \cdot \boldsymbol{\xi}} = \sqrt{-g_{tt}} \simeq 1 + \Phi$, where Φ is the Newtonian gravitational potential, i.e. the solution to the Poisson equation $\Delta\Phi = 4\pi\rho$ that vanishes at infinity. We may then write

$$M_{\mathcal{S}_{\text{ext}}} \simeq \int_{\Sigma} \rho \sqrt{\gamma} d^3x + \int_{\Sigma} \rho \Phi \sqrt{\gamma} d^3x. \quad (5.56)$$

The first term is the volume integral of the fluid mass density; the second term is negative (since $\Phi < 0$) and represents the *total gravitational potential energy* of the star, or equivalently the *gravitational binding energy* of the star. Its presence should not be a surprise given the mass-energy equivalence in relativity. However, at the Newtonian limit, this term can be neglected in front of the first one and formula (5.56) reduces to the classical expression of mass of a Newtonian body.

Since in vacuum $\mathbf{T} = 0$, an immediate corollary of Property 5.14 is:

Property 5.15: independence of the Komar mass from the integration surface

Two $(n - 2)$ -surfaces \mathcal{S}_{ext} and \mathcal{S}_{int} that are connected by a spacelike hypersurface Σ in a *vacuum* region of an asymptotically flat stationary spacetime ruled by the Einstein equation with $\Lambda = 0$ share the same Komar mass:

$$M_{\mathcal{S}_{\text{ext}}} = M_{\mathcal{S}_{\text{int}}}. \quad (5.57)$$

The concept of “total mass” of a stationary and asymptotically flat spacetime is well captured by the Komar mass at infinity:

Property 5.16: Komar mass at infinity

Let (\mathcal{M}, g) be a stationary and asymptotically flat spacetime of dimension $n \geq 4$. Let (x^α) be an asymptotically Minkowskian coordinate system (cf. Sec. 5.3.2) such that $\partial_0 = \xi$ (the stationary Killing vector). Let \mathcal{S}_r be the $(n - 2)$ -surface defined by $(x^0, r) = \text{const}$, where $r := \sqrt{(x^1)^2 + \dots + (x^{n-1})^2}$. If the Komar mass $M_{\mathcal{S}_r}$ converges when $r \rightarrow +\infty$, we call

$$M_\infty := \lim_{r \rightarrow +\infty} M_{\mathcal{S}_r} \quad (5.58)$$

the **Komar mass at infinity** or **total mass** of the spacetime (\mathcal{M}, g) . Moreover, if g obeys the Einstein equation with $\Lambda = 0$ and if the energy-momentum tensor T vanishes for r larger than some threshold r_0 , one has $M_\infty = M_{\mathcal{S}_r}$ for any $r \geq r_0$.

Proof. Since for $r \geq r_0$, \mathcal{S}_r is located in a vacuum region of spacetime, the value of $M_{\mathcal{S}_r}$ is independent of r by Property 5.15, hence $M_\infty = M_{\mathcal{S}_r}$. \square

Example 5 (Total mass of Schwarzschild spacetime): We have already noticed in Example 3 that in the Schwarzschild spacetime, the Komar mass $M_{\mathcal{S}}$ is equal to the mass parameter m of Schwarzschild’s metric, whatever the 2-surface \mathcal{S} [Eq. (5.47)]. We thus conclude that $M_\infty = m$.

Example 6 (Vanishing total mass of cubic Galileon black holes): In a stationary and asymptotically flat 4-dimensional spacetime, if $d\xi$ decays faster than $1/r^2$, then $M_\infty = 0$. This happens for black holes in a scalar-tensor theory of gravity known as the *cubic Galileon*, as shown in Sec. 5.1 of Ref. [489]. Actually, for these black holes, the spacetime metric tends to Minkowsky metric as fast as $1/r^4$ when $r \rightarrow +\infty$, which implies that $d\xi$ decays as $1/r^5$. This rapid convergence to flat space can be interpreted as a screening of the black hole’s gravity by the surrounding scalar field.

Historical note : The Komar current (5.50) associated to any vector field of spacetime has been introduced by Arthur Komar in 1959 [318]; he showed that this current is conserved (divergence-free) [Eq. (5.51)]. In a subsequent study published in 1962 [319], Komar showed that the Komar current $\mathcal{J}(\xi)$ associated to a Killing vector ξ is, up to some factor, the Ricci tensor applied to ξ [Eq. (5.53)] and that the converved integral representing the flux of $\mathcal{J}(\xi)$ through a one-ended hypersurface Σ can be transformed into a surface integral on the boundary \mathcal{S} of Σ , which is nothing but Eq. (5.45) with $n = 4$. He proved that such an integral is independent of the surface \mathcal{S} as soon as \mathcal{S} is located outside

the sources of the Einstein equation (Property 5.15). The generalization to dimensions $n \geq 4$ has been performed by Robert Myers and Malcolm Perry in 1986 [375] [cf. their Eq. (4.1), where $N = n - 1$].

5.3.4 N+1 decomposition

Let $\mathcal{N} := n - 1$, n being the spacetime dimension. Given a spacelike hypersurface Σ , with future-directed unit normal \mathbf{n} , the **$\mathcal{N}+1$ decomposition** of ξ with respect to Σ is the pair (N, β) where N is a scalar field on Σ and β is a vector field tangent to Σ such that

$$\xi = N\mathbf{n} + \beta. \quad (5.59)$$

In other words, β is the orthogonal projection of ξ on Σ and $N\mathbf{n}$ is the part of ξ along \mathbf{n} . N and β are uniquely determined by $N = -\mathbf{n} \cdot \xi$ and $\beta = \xi + (n \cdot \xi)\mathbf{n}$. In the language of the $\mathcal{N}+1$ formalism of general relativity, N is called the **lapse function** and β the **shift vector** (cf. e.g. Ref. [226]).

Similarly, the **$\mathcal{N}+1$ decomposition** of the energy-momentum tensor \mathbf{T} with respect to the hypersurface Σ is the triplet $(E, \mathbf{p}, \mathbf{S})$, where E is a scalar field, \mathbf{p} is a 1-form satisfying $\langle \mathbf{p}, \mathbf{n} \rangle = 0$ and \mathbf{S} is a symmetric type-(0, 2) tensor field satisfying $\mathbf{S}(\mathbf{n}, .) = 0$, such that⁹

$$\mathbf{T} = E\mathbf{n} \otimes \mathbf{n} + \underline{\mathbf{n}} \otimes \mathbf{p} + \mathbf{p} \otimes \underline{\mathbf{n}} + \mathbf{S}. \quad (5.60)$$

E , \mathbf{p} and \mathbf{S} are uniquely determined by

$$E = T_{\mu\nu}n^\mu n^\nu, \quad p_\alpha = -n^\mu T_{\mu\nu}\gamma^\nu_\alpha, \quad S_{\alpha\beta} = T_{\mu\nu}\gamma^\mu_\alpha\gamma^\nu_\beta, \quad (5.61)$$

where $\gamma^\alpha_\beta = \delta^\alpha_\beta + n^\alpha n_\beta$ stands for the **orthogonal projector onto** Σ . Physically, E is the matter energy density measured by the observer \mathcal{O} of 4-velocity \mathbf{n} , \mathbf{p} is the matter momentum density measured by \mathcal{O} and \mathbf{S} is the matter stress tensor measured by \mathcal{O} .

Property 5.17: $\mathcal{N}+1$ expression of the Komar mass

Under the same hypotheses as Property 5.14, one can rewrite Eq. (5.55) as

$$M_{\mathcal{S}_{\text{ext}}} = M_{\mathcal{S}_{\text{int}}} + \int_{\Sigma} \left[N \left(E + \frac{1}{n-3} S \right) - \frac{n-2}{n-3} \langle \mathbf{p}, \beta \rangle \right] \sqrt{\gamma} d^{n-1}x, \quad (5.62)$$

where $S := g^{\mu\nu}S_{\mu\nu} = \gamma^{ij}S_{ij}$ is the trace of the stress tensor \mathbf{S} .

Proof. From Eq. (5.60), the trace of the energy-momentum tensor is $T := g^{\mu\nu}T_{\mu\nu} = En_\mu n^\mu + n_\mu p^\mu + p_\mu n^\mu + S = -E + S$, since $n_\mu n^\mu = -1$ and $p_\mu n^\mu = 0$. In view of Eqs. (5.59) and (5.60), we can then rewrite the integrand in the r.h.s. of Eq. (5.55) as

$$\mathbf{T}(\mathbf{n}, \xi) - \frac{T}{n-2} \mathbf{n} \cdot \xi = N \underbrace{\mathbf{T}(\mathbf{n}, \mathbf{n})}_E + \underbrace{\mathbf{T}(\mathbf{n}, \beta)}_{-\langle \mathbf{p}, \beta \rangle} - \frac{S-E}{n-2} \underbrace{\mathbf{n} \cdot \xi}_{-N} = \frac{N}{n-2} [(n-3)E + S] - \langle \mathbf{p}, \beta \rangle.$$

This establishes Eq. (5.62). □

⁹See e.g. Sec. 5.1.2 of Ref. [226].

Example 7 (Komar mass of a weakly relativistic star): Let us check that formula (5.62) reduces to (5.22) for a weakly relativistic star in a n -dimensional spacetime. In this case, we have $N \simeq 1$ and $\beta \rightarrow 0$. Furthermore, the matter constituting the star shall be weakly relativistic as well, which implies that the trace S of the stress tensor is negligible in front of the energy density E . Under these hypotheses, Eq. (5.62) with $M_{\mathcal{S}_{\text{int}}} = 0$ yields $M_{\mathcal{S}_{\text{ext}}} = M$, where M is given by (5.22), given that $T_{00} \simeq E$.

5.3.5 Link between the Komar mass and the ADM mass

For generic (i.e. not necessarily stationary) asymptotically flat spacetimes, there exists the concept of *ADM energy-momentum*, *ADM* standing for Arnowitt, Deser and Misner [20]. It arises from the Hamiltonian formulation of general relativity developed by these authors. For a spacelike hypersurface Σ that is asymptotically Euclidean, the **ADM energy** of Σ is defined as some flux over a closed $(n - 2)$ -sphere \mathcal{S} of quantities involving derivatives of the metric γ induced by \mathbf{g} on Σ , in the limit where the coordinate radius of \mathcal{S} tends to infinity, while the 3 components of the **ADM momentum** of Σ are defined by similar integrals with quantities involving the extrinsic curvature tensor \mathbf{K} of Σ (see e.g. [297], Chap. 8 of [226], Sec. 4.3 of [415], Sec. 3.7 of [463] or Sec. 3.2.1 of [470] for details). The **ADM mass** of Σ is then defined as the Minkowskian norm of the ADM energy-momentum. For a vanishing ADM momentum, the ADM mass coincides with the ADM energy. One may then ask the relation with the Komar mass if the spacetime is stationary. The answer is very simple:

Property 5.18: equality of Komar and ADM masses (Beig 1978)

If $(\mathcal{M}, \mathbf{g})$ is a stationary and asymptotically flat n -dimensional spacetime, its Komar mass at infinity equals the ADM mass of any spacelike hypersurface Σ that is asymptotically orthogonal to the stationary Killing vector ξ .

For $n = 4$, the proof can be found in R. Beig's article [48] (see also Ref. [27] for a different proof), while the proof for $n > 4$ has been obtained by H. Barzegar, P.T. Chruściel and M. Hörzinger [42].

5.3.6 Komar angular momentum

As a stationary Killing vector gave birth to the Komar mass, an axisymmetric Killing vector gives birth to an invariant integral quantity: the Komar angular momentum. Before discussing it, let us first introduce the concept of axisymmetry:

A spacetime $(\mathcal{M}, \mathbf{g})$ is called **axisymmetric** iff (i) it is invariant under the action of the rotation group $\text{SO}(2)$ (or equivalently $\text{U}(1)$) and (ii) the orbits of the group action^a are everywhere spacelike curves or (ii') $(\mathcal{M}, \mathbf{g})$ admits a conformal completion (cf. Sec. 4.3) and the orbits of the group action are spacelike in the vicinity of the conformal boundary \mathcal{I} . Since $\text{SO}(2)$ is a one-parameter Lie group, its action is generated by a single Killing vector η , which is unique up to some scaling constant. The orbits of the $\text{SO}(2)$ action being closed curves, we shall fix the scaling constant so that the group parameter φ associated

with η has the standard periodicity 2π . If non-empty, the set of points of \mathcal{M} that are left invariant by the $\text{SO}(2)$ action is called the ***rotation axis***. Equivalently, the rotation axis is the set of points where the Killing vector η vanishes. Usually, if the dimension of \mathcal{M} is n , the rotation axis is a submanifold of \mathcal{M} of dimension $n - 2$. Moreover, it is a timelike submanifold in the region where the $\text{SO}(2)$ orbits are spacelike.

^aCf. Sec. 3.3.1 for the concepts of *group action* and *orbit*.

Let (\mathcal{M}, g) be an axisymmetric spacetime of dimension $n \geq 3$, with axisymmetric Killing vector η . Given a spacelike closed $(n - 2)$ -surface $\mathcal{S} \subset \mathcal{M}$, the ***Komar angular momentum over \mathcal{S}*** is defined by

$$J_{\mathcal{S}} := \frac{1}{16\pi} \int_{\mathcal{S}} \star(\underline{d}\eta), \quad (5.63)$$

where $\star(\underline{d}\eta)$ is the Hodge dual of the 2-form $\underline{d}\eta$ given by

$$(\underline{d}\eta)_{\alpha\beta} = \partial_\alpha \eta_\beta - \partial_\beta \eta_\alpha = \nabla_\alpha \eta_\beta - \nabla_\beta \eta_\alpha. \quad (5.64)$$

Remark 8: Besides the change $\xi \leftrightarrow \eta$, we notice a difference by a factor $-(n - 2)/(n - 3)$ between the r.h.s. of (5.36) and (5.63). For $n = 4$, this factor is -2 and is known as *Komar's anomalous factor* (see Ref. [308] for a discussion). Fundamentally, the independence of formula (5.63) from the spacetime dimension n stems from the fact that the $\text{SO}(2)$ action generated by η is effective only in a 2-dimensional space: the orthogonal complement of the rotation axis.

Remark 9: The Komar angular momentum is well defined for any spacetime dimension $n \geq 3$, while the definition of the Komar mass requires $n \geq 4$ (cf. Sec. 5.3.3).

Formulas (5.45) and (5.46) for the Komar mass rely only on the fact that ξ is a Killing vector. In an axisymmetric spacetime, we may therefore replace ξ by η and the numerical prefactor by $1/(16\pi)$ to get the following expressions for the Komar angular momentum:

$$J_{\mathcal{S}} = \frac{1}{16\pi} \int_{\mathcal{S}} \nabla_\mu \eta_\nu dS^{\mu\nu} \quad (5.65)$$

and

$$J_{\mathcal{S}} = \frac{1}{16\pi} \int_{\mathcal{S}} \partial_\mu \eta_\nu (s^\mu n^\nu - n^\mu s^\nu) \sqrt{q} d^{n-2}x. \quad (5.66)$$

As for the Komar mass (cf. Property 5.11), one can read the Komar angular momentum in the asymptotic expansion of the metric tensor in appropriate coordinates:

Property 5.19: Komar angular momentum in the asymptotic expression of the metric

Let (\mathcal{M}, g) be an asymptotically flat spacetime of dimension $n \geq 4$ that is axisymmetric and stationary, with axisymmetric Killing vector η . Let (x^α) be an asymptotically Minkowskian coordinate system in which g admits the expansion (5.31) and that is adapted to the

axisymmetry in the sense that $\eta = -x^2 \partial_{x^1} + x^1 \partial_{x^2}$ (i.e. the rotation axis of the $\text{SO}(2)$ action is the $(n-2)$ -dimensional surface $(x^1, x^2) = (0, 0)$). Then the coefficient $J_{(1)} = J_{12}$ in the expansion (5.31)-(5.34) of \mathbf{g} coincides in the limit $r := \sqrt{(x^1)^2 + \dots + (x^{n-1})^2} \rightarrow +\infty$ with the Komar angular momentum $J_{\mathcal{S}}$ over the sphere \mathcal{S} of constant (x^0, r) . In particular, for a weakly relativistic system, $J_{\mathcal{S}} = J_{12}$, where J_{12} is given by the volume integral (5.23).

Proof. Let us use formula (5.66) to evaluate $J_{\mathcal{S}}$. From $\eta = -x^2 \partial_{x^1} + x^1 \partial_{x^2}$ and the asymptotic expression (5.31) of \mathbf{g} , we get the following components of η with respect to the coordinates (x^α) in the limit $r \rightarrow +\infty$: $\eta_\alpha = g_{\alpha\mu}\eta^\mu = g_{\alpha 1}(-x^2) + g_{\alpha 2}x^1 \simeq (\eta_0, -x^2, x^1, 0, \dots, 0)$, with

$$\eta_0 = -g_{01}x^2 + g_{02}x^1 = \frac{n-2}{2}\alpha_n \left[-\frac{J_{(1)}x^2}{r^{n-1}}x^2 + \frac{(-J_{(1)})x^1}{r^{n-1}}x^1 \right] = -\frac{8\pi J_{(1)}}{\Omega_{n-2}} \frac{(x^1)^2 + (x^2)^2}{r^{n-1}},$$

where we have used Eq. (5.32) to express α_n . The integrand in Eq. (5.66) is then, taking into account that $\partial_0 = 0$ (stationarity) and $s^0 = 0$ (\mathbf{s} tangent to the hypersurface $x^0 = \text{const}$):

$$\begin{aligned} \partial_\mu \eta_\nu (s^\mu n^\nu - n^\mu s^\nu) &= \partial_i \eta_0 s^i n^0 + \underbrace{\partial_i (-x^2)}_{-\delta^2_i} (s^i n^1 - n^i s^1) + \underbrace{\partial_i x^1}_{\delta^1_i} (s^i n^2 - n^i s^2) \\ &= n^0 s^i \partial_i \eta_0 + 2(s^1 n^2 - s^2 n^1). \end{aligned}$$

Since $s^i \simeq x^i/r$ and $\partial_i r = x^i/r$, we deduce from the above expression of η_0 that

$$s^i \partial_i \eta_0 = \frac{8\pi(n-3)J_{(1)}}{\Omega_{n-2}} \frac{(x^1)^2 + (x^2)^2}{r^n}.$$

In the proof of Property 5.11 for the Komar mass, it was sufficient to consider only $\partial_i \xi_0$ and to express the components of \mathbf{n} to the lowest order in $1/r$, namely $n^\alpha \simeq (1, 0, \dots, 0)$. Here we have $\partial_i \eta_j = O(1)$, so that we must consider the next order in the spatial components n^i of \mathbf{n} . They are given by $n^i = -\beta^i/N$ (e.g. Eq. (5.36) in Ref. [226]) where N is the lapse function and the β^i 's are the components of the shift vector (cf. Sec. 5.3.4). For r large, we may take $N \simeq 1$ and $\beta^i \simeq \beta_i = g_{0i}$ (Eq. (5.49) in Ref. [226]). Hence $n^1 \simeq -g_{01} = -8\pi J_{(1)}x^2/(\Omega_{n-2}r^{n-1})$ and $n^2 \simeq -g_{02} = 8\pi J_{(1)}x^1/(\Omega_{n-2}r^{n-1})$. On the other side, $n^0 = 1/N \simeq 1$. Hence the integrand becomes

$$\begin{aligned} \partial_\mu \eta_\nu (s^\mu n^\nu - n^\mu s^\nu) &= \frac{8\pi J_{(1)}}{\Omega_{n-2}} \left[(n-3) \frac{(x^1)^2 + (x^2)^2}{r^n} + 2 \left(\frac{x^1}{r} \times \frac{x^1}{r^{n-1}} - \frac{x^2}{r} \times \left(-\frac{x^2}{r^{n-1}} \right) \right) \right] \\ &= \frac{8\pi(n-1)J_{(1)}}{\Omega_{n-2}} \frac{(x^1)^2 + (x^2)^2}{r^n}. \end{aligned}$$

Accordingly, formula (5.66) yields

$$J_{\mathcal{S}} = \frac{(n-1)J_{(1)}}{2\Omega_{n-2}} \int_{\mathcal{S}} \frac{(x^1)^2 + (x^2)^2}{r^n} \sqrt{q} d^{n-2}y,$$

where we have denoted the surface element of \mathcal{S} by $\sqrt{q} d^{n-2}y$ to avoid any confusion with the coordinates (x^α) . Let us use spherical coordinates $(y^a)_{1 \leq a \leq n-2} = (\theta_1, \theta_2, \dots, \theta_{n-3}, \varphi)$ on

\mathcal{S} , i.e. coordinates such that all the θ_k 's span $(0, \pi)$, φ spans $(0, 2\pi)$ and the asymptotically Minkowskian coordinates $(x^i)_{1 \leq i \leq n-1}$ are expressible as¹⁰

$$\left\{ \begin{array}{lcl} x^1 & = & r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{n-4} \sin \theta_{n-3} \cos \varphi \\ x^2 & = & r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{n-4} \sin \theta_{n-3} \sin \varphi \\ x^3 & = & r \sin \theta_1 \sin \theta_2 \cdots \sin \theta_{n-4} \cos \theta_{n-3} \\ x^4 & = & r \sin \theta_1 \sin \theta_2 \cdots \cos \theta_{n-4} \\ & \vdots & \\ x^{n-2} & = & r \sin \theta_1 \cos \theta_2 \\ x^{n-1} & = & r \cos \theta_1 \end{array} \right.$$

Then $(x^1)^2 + (x^2)^2 = r^2 \sin^2 \theta_1 \cdots \sin^2 \theta_{n-3}$. Given that \mathcal{S} is the $(n-2)$ -sphere $(x^0, r) = \text{const}$, one has, for r large, $\sqrt{\bar{q}} \simeq r^{n-2} \sqrt{\bar{q}}$, where \bar{q} is the determinant with respect to the coordinates (y^a) of the round metric on \mathbb{S}^{n-2} :

$$\sqrt{\bar{q}} = \sin^{n-3} \theta_1 \sin^{n-4} \theta_2 \cdots \sin^2 \theta_{n-4} \sin \theta_{n-3}. \quad (5.67)$$

It follows that

$$\begin{aligned} J_{\mathcal{S}} &= \frac{(n-1)J_{(1)}}{2\Omega_{n-2}} \int_{\mathcal{S}} \sin^{n-1} \theta_1 \sin^{n-2} \theta_2 \cdots \sin^4 \theta_{n-4} \sin^3 \theta_{n-3} d\theta_1 \cdots d\theta_{n-3} d\varphi \\ &= \frac{(n-1)J_{(1)}}{2\Omega_{n-2}} \times 2\pi I_{n-1} I_{n-2} \cdots I_3, \end{aligned} \quad (5.68)$$

where we have introduced

$$I_k := \int_0^\pi \sin^k \theta d\theta.$$

Writing $\sin^k \theta = \sin^{k-1} \theta \sin \theta$ and integrating by part, we get $I_k = (k-1)I_{k-2}/k$ for $k \geq 2$. Hence

$$I_{n-1} I_{n-2} \cdots I_3 = \frac{n-2}{n-1} I_{n-3} \times \frac{n-3}{n-2} I_{n-4} \times \cdots \times \frac{3}{4} I_2 \times \frac{2}{3} I_1 = \frac{2}{n-1} I_{n-3} I_{n-4} \cdots I_1.$$

Now, from its definition as the area of \mathbb{S}^{n-2} , Ω_{n-2} is the integral of $\sqrt{\bar{q}}$ over the spherical coordinates $(\theta_1, \dots, \theta_{n-3}, \varphi)$. In view of expression (5.67) for $\sqrt{\bar{q}}$, we get $\Omega_{n-2} = 2\pi I_{n-3} I_{n-4} \cdots I_1$. Equation (5.68) reduces then to $J_{\mathcal{S}} = J_{(1)}$. \square

The volume integral formula (5.54) for the Komar mass relies only on ξ being a Killing vector. We may therefore transpose it to the Komar angular momentum:

¹⁰Note that these formulas generalize the standard expression (4.2) with $(x, y, z) = (x^1, x^2, x^3)$ to $n \geq 4$.

Property 5.20: volume integral formula for the Komar angular momentum

Let $(\mathcal{M}, \mathbf{g})$ be an axisymmetric spacetime of dimension $n \geq 3$, with axisymmetric Killing vector $\boldsymbol{\eta}$. Let Σ be a compact spacelike hypersurface bounded by two closed $(n-2)$ -surfaces: an “internal” one, \mathcal{S}_{int} , and an “external” one, \mathcal{S}_{ext} , the exterior direction being that of the asymptotic flat end of $(\mathcal{M}, \mathbf{g})$. The Komar angular momenta over \mathcal{S}_{ext} and \mathcal{S}_{int} are then related by

$$J_{\mathcal{S}_{\text{ext}}} = J_{\mathcal{S}_{\text{int}}} + \frac{1}{16\pi} \int_{\Sigma} \mathcal{J}(\boldsymbol{\eta}) \cdot \mathbf{n} \sqrt{\gamma} d^{n-1}x, \quad \text{with } \mathcal{J}(\boldsymbol{\eta})^{\alpha} = -2R_{\mu}^{\alpha}\eta^{\mu}, \quad (5.69)$$

where \mathbf{n} is the future-directed unit normal to Σ and $\sqrt{\gamma} d^{n-1}x$ is the volume element on Σ induced by \mathbf{g} . The vector $\mathcal{J}(\boldsymbol{\eta})$ is the Komar current of $\boldsymbol{\eta}$ [compare Eq. (5.53)]. If \mathbf{g} obeys the Einstein equation (1.42) with some energy-momentum tensor \mathbf{T} and $\Lambda = 0$, then the above formula may be rewritten as

$$J_{\mathcal{S}_{\text{ext}}} = J_{\mathcal{S}_{\text{int}}} - \int_{\Sigma} \left(\mathbf{T}(\boldsymbol{\eta}, \mathbf{n}) - \frac{T}{n-2} \boldsymbol{\eta} \cdot \mathbf{n} \right) \sqrt{\gamma} d^{n-1}x, \quad (5.70)$$

where $T := \text{tr}_g \mathbf{T} = g^{\mu\nu} T_{\mu\nu}$.

Remark 10: Since the axisymmetry generator $\boldsymbol{\eta}$ is spacelike, the spacelike hypersurface Σ can be chosen so that $\boldsymbol{\eta}$ is tangent to it; one has then $\boldsymbol{\eta} \cdot \mathbf{n} = 0$ and the second term in the integrand of Eq. (5.70) vanishes identically. Moreover, if one expresses \mathbf{T} via the $\mathcal{N}+1$ decomposition (5.60), one can write the first term as $\mathbf{T}(\boldsymbol{\eta}, \mathbf{n}) = -\langle \mathbf{p}, \boldsymbol{\eta} \rangle$, so that formula (5.70) simplifies to

$$J_{\mathcal{S}_{\text{ext}}} = J_{\mathcal{S}_{\text{int}}} + \int_{\Sigma} \langle \mathbf{p}, \boldsymbol{\eta} \rangle \sqrt{\gamma} d^{n-1}x. \quad (5.71)$$

As for the Komar mass (Property 5.15), it follows immediately from Eq. (5.70) that

Property 5.21: independence of the Komar angular momentum from the integration surface

Two $(n-2)$ -surfaces \mathcal{S}_{ext} and \mathcal{S}_{int} that are connected by a spacelike hypersurface Σ in a *vacuum* region of an axisymmetric spacetime ruled by the Einstein equation share the same Komar angular momentum:

$$J_{\mathcal{S}_{\text{ext}}} = J_{\mathcal{S}_{\text{int}}}. \quad (5.72)$$

As for the total mass (cf. Property 5.16), the concept of “total angular momentum” of an axisymmetric and asymptotically flat spacetime is well captured by the Komar angular momentum at infinity:

Property 5.22: Komar angular momentum at infinity

Let (\mathcal{M}, g) be an axisymmetric and asymptotically flat spacetime of dimension $n \geq 3$. Let (x^α) be an asymptotically Minkowskian coordinate system (cf. Sec. 5.3.2) such that the axisymmetric Killing vector η is tangent to the hypersurfaces $x^0 = \text{const}$. Let \mathcal{S}_r be the $(n-2)$ -surface defined by $(x^0, r) = \text{const}$, where $r := \sqrt{(x^1)^2 + \dots + (x^{n-1})^2}$. If the Komar angular momentum $J_{\mathcal{S}_r}$ converges when $r \rightarrow +\infty$, we call

$$J_\infty := \lim_{r \rightarrow +\infty} J_{\mathcal{S}_r} \quad (5.73)$$

the **Komar angular momentum at infinity** or **total angular momentum** of the spacetime (\mathcal{M}, g) . Moreover, if g obeys the Einstein equation with $\Lambda = 0$ and if the energy-momentum tensor T vanishes for r larger than some threshold r_0 , one has $J_\infty = J_{\mathcal{S}_r}$ for any $r \geq r_0$.

Remark 11: There is no concept of “ADM angular momentum” (due to the so-called *supertranslation ambiguity* at spatial infinity, cf. e.g. Sec. 6 of Ref. [522] or Sec. 8.5 of Ref. [226]). Consequently, there is no point in comparing the Komar angular momentum with an “ADM” counterpart, contrary to what was done in Sec. 5.3.5 for the mass.

5.4 The event horizon as a Killing horizon

In view of Property 5.2, let us examine successively the only two allowed cases for a connected component¹¹ \mathcal{H} of the event horizon of a stationary black hole: either the Killing vector ξ is null on all \mathcal{H} (Sec. 5.4.1) or ξ is spacelike in some part (possibly all) of \mathcal{H} (Sec. 5.4.2). In both cases, we shall see that, under some hypotheses, \mathcal{H} is a Killing horizon; this is immediate when ξ is null on \mathcal{H} , but not trivial otherwise.

5.4.1 Null stationary Killing field on \mathcal{H} : the staticity theorem

By Lemma 2.3 (Sec. 2.3.4), if the Killing vector field ξ is null on \mathcal{H} , it is necessarily tangent to the null geodesic generators of \mathcal{H} and therefore collinear to the null normals ℓ of \mathcal{H} . From the definition of a Killing horizon (cf. Sec. 3.3.2), we get immediately:

Property 5.23: event horizon as a Killing horizon for ξ null on it

In a stationary spacetime (\mathcal{M}, g) containing a black hole, if the stationary Killing vector ξ is null on the whole of a connected component \mathcal{H} of the event horizon, then \mathcal{H} is a

¹¹In the most astrophysically relevant case, namely that of a stationary black hole in a 4-dimensional vacuum spacetime, the event horizon is a connected hypersurface. However, we shall see that disconnected event horizons exist for some stationary black holes in 5-dimensional spacetimes; they exist as well in 4 dimensions if one allows for an electric charge (Majumdar-Papapetrou black holes mentioned in Remark 5 in Sec. 5.2.3).

Killing horizon with respect to ξ .

It follows that, on \mathcal{H} , the Killing vector field ξ is hypersurface-orthogonal, being normal to \mathcal{H} . If this holds for all the connected components of the event horizon, then, under certain hypotheses (especially that the electrovacuum Einstein equation is fulfilled), one can go further and prove that ξ must actually be hypersurface-orthogonal everywhere in \mathcal{M} , i.e. that (\mathcal{M}, g) is *static*, according to the definition given in Sec. 5.2.1:

Property 5.24: staticity theorem (Sudarsky & Wald 1993 [467])

Let (\mathcal{M}, g) be an asymptotically flat stationary spacetime of dimension $n \geq 4$ that contains a black hole of event horizon \mathcal{H} so that the stationary Killing vector ξ is null on all \mathcal{H} . By Property 5.23, \mathcal{H} is a Killing horizon. If \mathcal{H} is non-degenerate (i.e. has a non-vanishing surface gravity κ , cf. Sec. 3.3.6) and g obeys the vacuum or electrovacuum Einstein equation (cf. Sec. 1.5.2), then the spacetime is static and the domain of outer communications (cf. Sec. 4.4.2) is strictly static.

Proof. For simplicity, we give only the proof for the case of the vacuum Einstein equation (1.44), referring to Sudarsky & Wald's article [467] for the electrovacuum case. Since \mathcal{H} is a non-degenerate Killing horizon, it is contained in a bifurcate Killing horizon (cf. Property 3.15 and Remark 3 on p. 87). Let \mathcal{S} be the corresponding bifurcation surface. Thanks to a theorem by Chruściel & Wald (Theorem 4.2 in Ref. [128]), the domain of outer communications can be foliated by a family of spacelike hypersurfaces $(\Sigma_t)_{t \in \mathbb{R}}$ that are asymptotically flat, are asymptotically orthogonal to ξ , have \mathcal{S} as inner boundary, are Lie-dragged¹² to each other by ξ and are **maximal**. The last property means that the extrinsic curvature tensor K of each hypersurface Σ_t has a vanishing trace: $K^i{}_i = 0$, or equivalently, that the unit normal n to Σ_t is divergence-free: $\nabla_\mu n^\mu = 0$ (cf. Sec. 10.2.2 of Ref. [226]). Let us consider the $\mathcal{N} + 1$ decomposition with respect to Σ_t introduced in Sec. 5.3.4. The maximality property $K^i{}_i = 0$ allows one to simplify one of the Einstein equations in the $\mathcal{N} + 1$ decomposition, namely the momentum constraint (see e.g. Chap. 5 of Ref. [226]), which reduces to $D^j K_{ij} = 8\pi p_i$, where D is the Levi-Civita connection associated to the metric γ induced by g on Σ_t and p is the matter momentum density in the hypersurface Σ_t (cf. Sec. 5.3.4). Since we consider the vacuum case, we have $p = 0$. Hence the momentum constraint reduces to $D^j K_{ij} = 0$. Contracting with the shift vector β arising from the orthogonal split (5.59) of ξ , we get

$$0 = D^j K_{ij} \beta^i = D^j (K_{ij} \beta^i) - K_{ij} D^j \beta^i = D^j (K_{ij} \beta^i) - N K_{ij} K^{ij},$$

where the last equality results from the symmetry of K_{ij} and the identity $D^j \beta^i + D^i \beta^j = 2N K^{ij}$, which holds for the shift vector β associated with the stationary Killing vector ξ (see e.g. Eqs. (5.66) and (5.67) in Ref. [226] with $\partial \gamma_{ij} / \partial t = 0$). Let us integrate the above equation over Σ_t and use the Gauss-Ostrogradsky theorem for the divergence term $D^j (K_{ij} \beta^i)$, taking into account that the boundary of Σ_t is made of the bifurcation sphere \mathcal{S} (inner boundary) and the

¹² Σ_{t+dt} is the image of Σ_t by the infinitesimal displacement $dt \xi$.

sphere “at infinity” \mathcal{S}_∞ (outer boundary), the latter being a $(n - 2)$ -sphere in Σ_t of coordinate radius $r \rightarrow +\infty$ (r is defined from the asymptotic flatness property); we get

$$\int_{\mathcal{S}_\infty} s^j K_{ij} \beta^i \sqrt{q} d^{n-2}x - \int_{\mathcal{S}} s^j K_{ij} \beta^i \sqrt{q} d^{n-2}x - \int_{\Sigma_t} N K_{ij} K^{ij} \sqrt{\gamma} d^{n-1}x = 0,$$

where s is the unit normal to \mathcal{S}_∞ and \mathcal{S} in Σ_t , oriented towards the asymptotic flat end of Σ_t . Now, each of the two surface integrals vanishes identically: the first one because on the asymptotically flat slice Σ_t , \mathbf{K} decays as $O(r^{-2})$ for $r \rightarrow +\infty$ and, still in this limit, $\beta \rightarrow 0$ since ξ is asymptotically orthogonal to Σ_t [cf. the orthogonal decomposition (5.59)]. As for the integral over \mathcal{S} , it is zero because $\beta|_{\mathcal{S}} = 0$. Indeed, since \mathcal{S} is the bifurcation surface of a bifurcate Killing horizon with respect to ξ (Property 5.23), one has $\xi|_{\mathcal{S}} = 0$ (Property 3.12), which implies via Eq. (5.59) both $N|_{\mathcal{S}} = 0$ and $\beta|_{\mathcal{S}} = 0$. We therefore get

$$\int_{\Sigma_t} N K_{ij} K^{ij} \sqrt{\gamma} d^{n-1}x = 0.$$

Given that $N > 0$ on $\Sigma_t \setminus \mathcal{S}$ and $K_{ij} K^{ij} \geq 0$ on Σ_t (for $K_{ij} K^{ij} = \gamma^{ik} \gamma^{jl} K_{ij} K_{kl}$ with γ being a Riemannian metric), the vanishing of the above integral implies $K_{ij} = 0$. Now, $\mathbf{K} = -(2N)^{-1} \mathcal{L}_{\mathbf{m}} \gamma$ (cf. Eq. (4.30) in Ref. [226]), where $\mathbf{m} := N\mathbf{n}$ is the evolution vector along the normal to Σ_t (cf. Sec. 4.3 of Ref. [226]). The vanishing of \mathbf{K} implies then

$$\dot{\gamma} := \mathcal{L}_{\mathbf{m}} \gamma = 0 \quad \text{and} \quad \dot{\pi} := \mathcal{L}_{\mathbf{m}} \pi = 0, \quad (5.74)$$

where π is the momentum canonically conjugate to γ in the ADM Hamiltonian formulation of general relativity [20, 195]. π is a tensor density that is related to the extrinsic curvature tensor \mathbf{K} by $\pi^{ij} = \sqrt{\gamma} (K^k{}_k \gamma^{ij} - K^{ij})$. Hence, $\mathbf{K} = 0$ implies $\pi = 0$ so that $\dot{\pi} = 0$ in Eq. (5.74) is trivially fulfilled. In the Hamiltonian formulation, the spacetime dynamics is entirely described by the “time development” (i.e. the variation across the Σ_t foliation) of (γ, π) , which are the (q, p) pairs of Hamiltonian mechanics. Hamilton’s equations (5.74) shows that (γ, π) do not evolve at all. This means that \mathbf{m} is a Killing vector of the spacetime metric \mathbf{g} . Since $\mathbf{m} = N\mathbf{n}$ is by construction timelike and hypersurface-orthogonal, this proves that the part of spacetime foliated by the $(\Sigma_t)_{t \in \mathbb{R}}$, i.e. the domain of outer communications, is strictly static. The connection between \mathbf{m} and ξ is given by Eq. (5.59): $\xi = \mathbf{m} + \beta$. Hence the Killing vectors ξ and \mathbf{m} coincide iff the shift vector β vanishes. If β would not vanish identically, it would be a Killing vector itself, as the difference $\xi - \mathbf{m}$ of two Killing vectors. However, as seen above, β vanishes at the inner boundary \mathcal{S} of Σ_t and at the “outer boundary”, i.e. for $r \rightarrow +\infty$. In most cases, these constraints will result in $\beta = 0$ (for instance, they exclude β being a rotational Killing vector), so that we conclude that $\mathbf{m} = \xi$. \square

Remark 1: In the original formulation by Sudarsky & Wald (Theorem 2 of Ref. [467]), the non-degeneracy of \mathcal{H} is replaced by \mathcal{H} being part of a bifurcate Killing horizon. However, as mentioned in the above proof, these two properties are essentially equivalent.

Remark 2: Sudarsky & Wald stated the staticity theorem for the dimension $n = 4$ but, as pointed out in Refs. [117, 122, 279], the theorem straightforwardly generalizes to $n \geq 4$. Actually, the proof given above¹³ does not rely on $n = 4$.

¹³The first part of the proof, leading to $\mathbf{K} = 0$, differs from Sudarsky & Wald’s one, which is based on integral mass formulas.

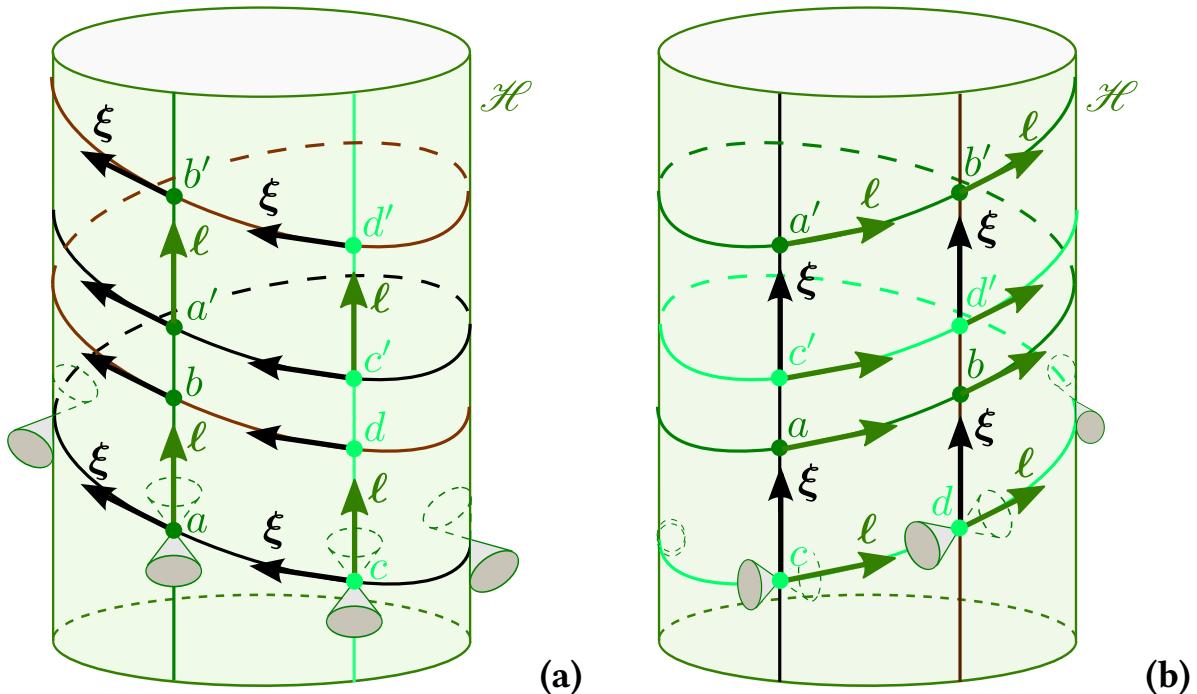


Figure 5.1: Two equivalent representations of an event horizon \mathcal{H} with cross-sections of \mathbb{S}^n topology and with a stationary Killing vector field ξ that is spacelike on \mathcal{H} : **(a)** Representation with the null geodesic generators of \mathcal{H} drawn as vertical lines; two of them are actually depicted, in dark green and light green respectively, with a null normal ℓ along them; besides, two field lines of ξ (orbits of the stationary group) are depicted, in black and brown respectively. **(b)** Representation with the field lines of ξ 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 ξ , being spacelike, lies outside of them, while the null normal ℓ is tangent to them. The strong rigidity theorem (Property 5.25) states that ℓ is collinear to a Killing vector field χ , making \mathcal{H} a Killing horizon.

Remark 3: Given the definition of a non-rotating horizon given in Property 5.2, the staticity theorem can be summarized as follows: an electrovacuum spacetime containing a non-rotating and non-degenerate black hole is static.

5.4.2 Spacelike stationary Killing field on \mathcal{H} : the strong rigidity theorem

When the stationary Killing vector ξ is spacelike on some part of a connected component \mathcal{H} of the event horizon, it obviously cannot be collinear to a null normal ℓ of \mathcal{H} . Assuming that \mathcal{H} has cross-sections of spherical topology (as it must in dimension $n = 4$ according to Property 5.4), we observe that, with respect to the null geodesic generators of \mathcal{H} , the field lines of ξ form helices, as depicted in Fig. 5.1a. By reciprocity, with respect to the field lines of ξ , the null geodesic generators form 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 ξ . Since asymptotically the field lines of ξ are worldlines of inertial observers, Fig. 5.1b leads us to say that \mathcal{H} is rotating, thereby justifying the terminology introduced in Sec. 5.2.3 (Property 5.2).

When the Killing field ξ is not null on \mathcal{H} , we cannot say a priori that \mathcal{H} is a Killing horizon.

However, modulo some additional hypotheses, it turns out that \mathcal{H} is still a Killing horizon, albeit with respect to a Killing vector distinct from ξ . This result is due to S.W. Hawking (1972) [259, 265] and is known as the **(strong) rigidity theorem**. We give below a modern version of this theorem, due to Hollands, Ishibashi & Wald (2007) [281] and Moncrief & Isenberg (2008) [371] (see also Theorem 8.1 p. 470 of Choquet-Bruhat's textbook [107]).

Property 5.25: strong rigidity theorem

Let (\mathcal{M}, g) be a stationary spacetime of dimension $n \geq 4$ containing a black hole. Let \mathcal{H} be a connected component of the black hole event horizon. Let us assume that the stationary Killing vector ξ is spacelike on some parts of \mathcal{H} . If

1. \mathcal{M} and \mathcal{H} are (real) analytic manifolds (cf. Remark 4 in Sec. A.2.1), with g being an analytic field,
2. g fulfills the vacuum Einstein equation (1.44) or the electrovacuum Einstein equation (1.54),
3. $n = 4$ or \mathcal{H} has a null geodesic generator that is incomplete,
4. \mathcal{H} has compact cross-sections,
5. ξ is transverse to some cross-sections,

then the spacetime (\mathcal{M}, g) admits a second Killing vector, χ say, such that ξ and χ commute: $[\xi, \chi] = 0$ and \mathcal{H} is a Killing horizon with respect to χ .

Sketch of proof. The proof of the strong rigidity theorem is pretty long and we give here only a sketch for $n = 4$, referring to Hawking & Ellis' textbook [265] (Proposition 9.3.6) for full details and to Refs. [281, 279, 371] for $n > 4$. Let us thus assume $n = 4$ and that the (electro)vacuum Einstein equation holds. The null dominant energy condition is then satisfied and, since \mathcal{H} is connected, the topology theorem 1 (Property 5.4) implies that the complete cross-sections of \mathcal{H} have the topology of \mathbb{S}^2 , or equivalently that the topology of \mathcal{H} is $\mathbb{R} \times \mathbb{S}^2$.

Let us first show that the spacetime stationary action $\Phi : \mathbb{R} \times \mathcal{M} \rightarrow \mathcal{M}$, $(t, p) \mapsto \Phi_t(p)$ induces a $\text{SO}(2)$ action $\hat{\Phi}$ on the set \mathcal{G} of null geodesic generators of \mathcal{H} . Each Φ_t being an isometry of (\mathcal{M}, g) , it maps any null geodesic \mathcal{L} of \mathcal{H} to a null geodesic $\hat{\Phi}_t(\mathcal{L}) := \Phi_t(\mathcal{L})$. Moreover, since Φ_t leaves \mathcal{H} globally invariant (Property 5.1), $\hat{\Phi}_t(\mathcal{L}) \subset \mathcal{H}$, i.e. $\hat{\Phi}_t(\mathcal{L}) \in \mathcal{G}$. Let check that $\hat{\Phi}$ obeys the laws of a group action: obviously $\hat{\Phi}_0(\mathcal{L}) = \mathcal{L}$ and for any $(t_1, t_2) \in \mathbb{R}^2$ and $p \in \mathcal{L}$, we have $\Phi_{t_2+t_1}(p) \in \hat{\Phi}_{t_2+t_1}(\mathcal{L})$ but $\Phi_{t_2+t_1}(p) = \Phi_{t_2}(\Phi_{t_1}(p)) \in \hat{\Phi}_{t_2}(\hat{\Phi}_{t_1}(\mathcal{L}))$ as well, from which we conclude that $\hat{\Phi}_{t_2+t_1}(\mathcal{L}) = \hat{\Phi}_{t_2}(\hat{\Phi}_{t_1}(\mathcal{L}))$, given that there is only one null geodesic generator through the point $\Phi_{t_2+t_1}(p)$. Since the stationary generator ξ is assumed to be spacelike on some parts of \mathcal{H} , the group action is not trivial: there exist some null generators \mathcal{L} for which $\hat{\Phi}_t(\mathcal{L}) \neq \mathcal{L}$ (when ξ is null on a part of \mathcal{H} , it is necessarily tangent to a null generator \mathcal{L} , which is then a fixed point of the group action: $\forall t \in \mathbb{R}, \hat{\Phi}_t(\mathcal{L}) = \mathcal{L}$). Moreover, the group action $\hat{\Phi}$ is periodic: due to the topology $\mathbb{R} \times \mathbb{S}^2$ of \mathcal{H} , there exists a smallest $T > 0$

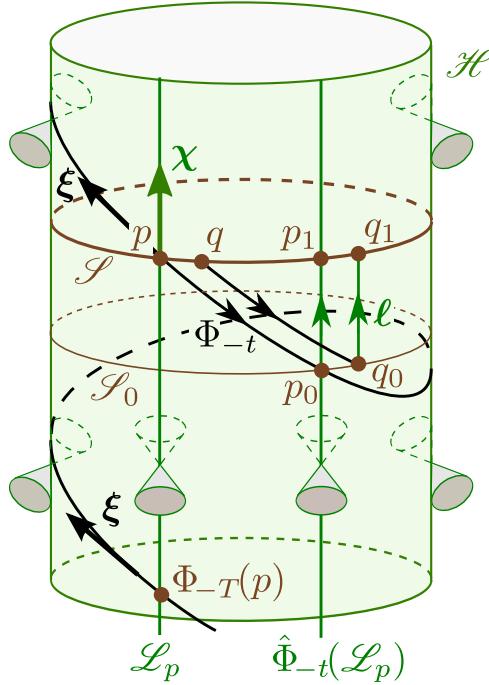


Figure 5.2: Action Ψ induced by the spacetime stationary action Φ (Killing vector ξ) on a cross-section \mathcal{S} of a rotating event horizon \mathcal{H} : for $p \in \mathcal{S}$ and $t \in \mathbb{R}$, $\Psi_t(p)$ is denoted by p_1 on the figure, while $p_0 = \Phi_{-t}(p)$. Similarly, for a point $q \in \mathcal{S}$ close to p , $q_0 = \Phi_{-t}(q)$ and $q_1 = \Psi_t(q)$. Note that this figure is drawn following the same convention as Fig. 5.1a, namely the null geodesic generators of \mathcal{H} appear as vertical lines.

such that for any null generator \mathcal{L} , $\hat{\Phi}_T(\mathcal{L}) = \mathcal{L}$ (cf. Fig. 5.1, where $a' = \Phi_T(a)$, $b' = \Phi_T(b)$, $c' = \Phi_T(c)$ and $d' = \Phi_T(d)$, noticing that each pair (p, p') , where p stands for a, b, c or d , lie on the same null generator). Hence $\hat{\Phi}$ is a $\text{SO}(2)$ action on the set \mathcal{G} of null generators of \mathcal{H} , the standard angle parameter $\varphi \in [0, 2\pi]$ of $\text{SO}(2)$ being related to $t \in [0, T]$ by $\varphi = 2\pi t/T$. Moreover, the $\text{SO}(2)$ action must have exactly two fixed points, \mathcal{L}_1 and \mathcal{L}_2 say, for \mathcal{G} has the topology of \mathbb{S}^2 : one can identify \mathcal{G} with an arbitrary complete cross-section $\hat{\mathcal{S}}$ of \mathcal{H} (each $\mathcal{L} \in \mathcal{G}$ intersects $\hat{\mathcal{S}}$ at a single point and through each $p \in \hat{\mathcal{S}}$ there is a single $\mathcal{L} \in \mathcal{G}$) and, as recalled above, all cross-sections have the topology of \mathbb{S}^2 .

Let \mathcal{S} be a complete cross-section of \mathcal{H} such that the stationary Killing vector ξ is nowhere tangent to \mathcal{S} . Let q be the metric induced by g on \mathcal{S} . We are going to show that the spacetime stationary action Φ induces a $\text{SO}(2)$ action $\Psi : [0, T] \times \mathcal{S} \rightarrow \mathcal{S}$, $(t, p) \mapsto \Psi_t(p)$ such that each Ψ_t is an isometry of (\mathcal{S}, q) . For $t \in \mathbb{R}$ and $p \in \mathcal{S}$, the point $\Phi_{-t}(p)$ lies in \mathcal{H} since \mathcal{H} is globally invariant by Φ_t (Property 5.1). Let then \mathcal{L} be the unique null generator of \mathcal{H} through $\Phi_{-t}(p)$. Since \mathcal{S} is a complete cross-section of \mathcal{H} , \mathcal{L} intersects \mathcal{S} at a unique point, which we define as $\Psi_t(p)$ (cf. Fig. 5.2, where \mathcal{L} is denoted $\hat{\Phi}_{-t}(\mathcal{L}_p)$). In other words, $\Psi_t(p)$ is obtained by displacing p by a parameter $-t$ along the field lines of ξ and then going back to \mathcal{S} along a null generator of \mathcal{H} . Equivalently, we may write

$$\Psi_t(p) = \hat{\Phi}_{-t}(\mathcal{L}_p) \cap \mathcal{S}, \quad (5.75)$$

where \mathcal{L}_p is the null geodesic generator through p (cf. Fig. 5.2). Let us check that Ψ is a 1-parameter group action on \mathcal{S} , i.e. fulfills $\Psi_0(p) = p$ and $\Psi_{t_2+t_1}(p) = \Psi_{t_2}(\Psi_{t_1}(p))$ for

any $p \in \mathcal{S}$ (cf. Sec. 3.3.1) and $(t_1, t_2) \in \mathbb{R}^2$. The first property is trivial. Regarding the second one, $\Psi_{t_1}(p)$ lies on the same null generator \mathcal{L} as $\Phi_{-t_1}(p)$ (by definition of Ψ_{t_1}). It follows that both $\Phi_{-t_2}(\Psi_{t_1}(p))$ and $\Phi_{-t_2}(\Phi_{-t_1}(p))$ lie on the null generator $\hat{\Phi}_{-t_2}(\mathcal{L})$. But $\Phi_{-t_2}(\Phi_{-1}(p)) = \Phi_{-t_2-t_1}(p)$, so that $\Psi_{t_2+t_1}(p) = \hat{\Phi}_{-t_2}(\mathcal{L}) \cap \mathcal{S} = \Psi_{t_2}(\Psi_{t_1}(p))$. Hence $\Psi : \mathbb{R} \times \mathcal{S} \rightarrow \mathcal{S}$, $(t, p) \mapsto \Psi_t(p)$ is 1-parameter group action on \mathcal{S} . Moreover this group action inherits the T -periodicity from $\hat{\Phi}$ and thus is a $\text{SO}(2)$ action. Indeed, Eq. (5.75) implies $\Psi_T(p) = \hat{\Phi}_{-T}(\mathcal{L}_p) \cap \mathcal{S} = \mathcal{L}_p \cap \mathcal{S} = p$ (see also Fig. 5.2, where $\Phi_{-T}(p)$ lies on the same null generator as p).

Let us now show that for any $t \in [0, T]$, Ψ_t is an isometry of $(\mathcal{S}, \mathbf{g})$. Let p and q be infinitely close points of \mathcal{S} . Because Φ_{-t} is a spacetime isometry, we have

$$\mathbf{q}(\vec{pq}, \vec{pq}) = \mathbf{g}|_p(\vec{pq}, \vec{pq}) = \mathbf{g}|_{p_0}(\vec{p_0q_0}, \vec{p_0q_0}), \quad (5.76)$$

where $p_0 := \Phi_{-t}(p)$ and $q_0 := \Phi_{-t}(q)$ (cf. Fig. 5.2). Now, p_0 and q_0 are two points of $\mathcal{S}_0 := \Phi_{-t}(\mathcal{S})$, which is a complete cross-section of \mathcal{H} that has no intersection with \mathcal{S} for $t \neq 0$, given that ξ is transverse to \mathcal{S} . One can then choose a parametrization λ of the null generators \mathcal{L} of \mathcal{H} such that $\lambda = 0$ on \mathcal{S}_0 and $\lambda = 1$ on \mathcal{S} . Let $\ell = dx/d\lambda|_{\mathcal{L}}$ be the associated tangent vector. One may consider that \mathcal{S}_0 and \mathcal{S} are the two ends of a foliation of \mathcal{H} by a family $(\mathcal{S}_\lambda)_{\lambda \in [0,1]}$ of complete cross-sections (same construction as in the proof of Property 3.1). By denoting by \mathbf{q} the metric induced by \mathbf{g} on each of the surfaces \mathcal{S}_λ , we have $\mathbf{g}|_{p_0}(\vec{p_0q_0}, \vec{p_0q_0}) = \mathbf{q}(\vec{p_0q_0}, \vec{p_0q_0})$. Now, since \mathcal{H} is a non-expanding horizon (Property 5.3) and the null energy condition is fulfilled (since we are considering the (electro)vacuum Einstein equation), the metric \mathbf{q} is preserved along ℓ : $\vec{q}^* \mathcal{L}_\ell \mathbf{q} = 0$ [Property 3.2]. Since $p_1 := \Psi_t(p)$ and $q_1 := \Psi_t(q)$ are obtained from p_0 and q_0 by the same displacement $\delta\lambda = 1$ along ℓ (cf. Fig. 5.2), it follows that $\mathbf{q}(\vec{p_0q_0}, \vec{p_0q_0}) = \mathbf{q}(\vec{p_1q_1}, \vec{p_1q_1})$. By combining with Eq. (5.76), we get $\mathbf{q}(\vec{pq}, \vec{pq}) = \mathbf{q}(\vec{p_1q_1}, \vec{p_1q_1})$, which proves that Ψ_t is an isometry of $(\mathcal{S}, \mathbf{q})$.

At this stage, we have shown that each cross-section of \mathcal{H} admits an isometry group isomorphic to $\text{SO}(2)$, i.e. is axisymmetric. One thus may say that, in addition to be stationary, \mathcal{H} is axisymmetric, although the concept of axisymmetry as the result of the metric invariance under a group action (cf. Sec. 3.3.1) is ill-defined for \mathcal{H} , given that, as a null manifold, \mathcal{H} is not endowed with a proper (non-degenerate) metric tensor.

The major step is to show that there exists a Killing vector field χ of $(\mathcal{M}, \mathbf{g})$ that, on \mathcal{H} , is tangent to the null generators; it will then follow that the whole spacetime $(\mathcal{M}, \mathbf{g})$ is axisymmetric, the generator of the $\text{SO}(2)$ isometry being $T/(2\pi)(\chi - \xi)$. We first define χ on \mathcal{H} only, by demanding that χ is the tangent vector to the null generators \mathcal{L} of \mathcal{H} corresponding to a parameter v of \mathcal{L} that is connected to the parameter t of the stationary action Φ , in the sense that for any point p of a given null generator¹⁴ \mathcal{L} ,

$$v(\Phi_T(p)) = v(p) + T. \quad (5.77)$$

The above condition is not sufficient to fully specify the parameter v , and hence χ , along \mathcal{L} . We require in addition that the non-affinity coefficient κ of χ is constant along \mathcal{L} : $\nabla_\chi \kappa = 0$. In view of Eq. (2.21), this is equivalent to $\chi = -e^\rho \vec{\nabla} u$ with $d^2\rho/dv^2 = 0$ along \mathcal{L} , where u is

¹⁴Recall that on \mathcal{H} , $\Phi_T(p)$ and p always lie on the same null generator.

a scalar field such that \mathcal{H} is the hypersurface $u = 0$. By construction, the vector field χ on \mathcal{H} is invariant under the stationary action, i.e. $\mathcal{L}_\xi \chi = 0$, so that ξ and χ commute: $[\xi, \chi] = 0$. Let then define the following vector field on \mathcal{H} :

$$\eta := \frac{1}{\Omega_{\mathcal{H}}} (\chi - \xi), \quad \text{with} \quad \Omega_{\mathcal{H}} := \frac{2\pi}{T}. \quad (5.78)$$

By construction, η has closed field lines of period 2π . Indeed, since ξ and χ commute, moving by a parameter 2π along η is equivalent to moving by a parameter $t = -T$ along the field lines of ξ and then by a parameter $v = T$ along the field lines of χ ; the property (5.77) ensures that one is back to the starting point. The vector field η vanishes at points where $\xi = \chi$, i.e. at points where ξ is tangent to some null generator of \mathcal{H} . This occurs along null generators that are fixed points of the group action $\hat{\Phi}$ on \mathcal{G} . As discussed above, there are only two such fixed points, \mathcal{L}_1 and \mathcal{L}_2 . These two null generators of \mathcal{H} define thus the common “axis” of all the rotations generated by η . Let \mathcal{C} be a curve in \mathcal{H} from \mathcal{L}_1 and \mathcal{L}_2 , orthogonal to η and such that the field lines of η from \mathcal{C} form a smooth cross-section \mathcal{S} of \mathcal{H} . Let us then denote by \mathcal{S}_v the cross-section obtained by displacing \mathcal{S} by a parameter v along χ . Finally, let \mathbf{k} be the future-directed null vector orthogonal to \mathcal{S}_v and complementary to χ , normalized such that $\chi \cdot \mathbf{k} = -1$ (\mathbf{k} is transverse to \mathcal{H} and is pointing towards the black hole region). In a neighborhood of \mathcal{H} , we introduce a spacetime coordinate system (v, r, θ, φ) , named **Gaussian null coordinates**, as follows. First of all, we choose spherical coordinates (θ, φ) on the cross-section \mathcal{S} , such that $\eta = \partial_\varphi$ and θ is a parameter along the “meridional” curve \mathcal{C} . To any point $p \in \mathcal{H}$, we assign then the coordinates (v, θ, φ) where v is such that $p \in \mathcal{S}_v$ and (θ, φ) are the coordinates of the intersection of \mathcal{S} and the null generator \mathcal{L} through p . Let us then consider the null geodesics \mathcal{L}' of tangent vector \mathbf{k} , which are transverse to \mathcal{H} and an affine parameter r along each of them such that, on \mathcal{H} , $r = 0$ and $\mathbf{k} = -dx/dr$. To any point $p \in \mathcal{L}'$ in the vicinity of \mathcal{H} , we assign the coordinates (v, r, θ, φ) such that r is the affine parameter of p along \mathcal{L}' and (v, θ, φ) are the coordinates of the intersection of \mathcal{L}' and \mathcal{H} . We may then extend the vector fields χ , η and \mathbf{k} away from \mathcal{H} by setting

$$\chi := \partial_v, \quad \eta := \partial_\varphi, \quad \mathbf{k} := -\partial_r.$$

By means of the (electro)vacuum Einstein equation, one can then show by induction that,

$$\forall N \in \mathbb{N}, \quad \underbrace{\mathcal{L}_k \cdots \mathcal{L}_k}_{N \text{ times}} \mathcal{L}_\chi g^{\mathcal{H}} = 0. \quad (5.79)$$

This is the lengthy part of the computation (see p. 343 of Ref. [265]). Once (5.79) is established, one can use the analyticity hypothesis to conclude that $\mathcal{L}_\chi g = 0$ in all \mathcal{M} , i.e. that χ is a Killing vector of (\mathcal{M}, g) . Since, on \mathcal{H} , χ is normal to \mathcal{H} , it follows that \mathcal{H} is a Killing horizon with respect to χ . \square

Remark 4: For $n \neq 4$, the Killing horizon \mathcal{H} is necessarily non-degenerate, given that \mathcal{H} must have an incomplete null geodesic generator in that case (third hypothesis of the theorem).

Remark 5: For $n = 4$, the stationary Killing vector ξ cannot be spacelike on the whole of \mathcal{H} : it is indeed null on the rotation axis, i.e. at the points where $\eta = 0$, given that Eq. (5.78) implies $\xi = \chi$ for $\eta = 0$.

Remark 6: In some formulations [281, 265], the hypothesis of ξ being transverse to some cross-sections is replaced by \mathcal{H} lying in the future of the past infinity \mathcal{I}^- . Actually it can be shown that the latter implies the former (see e.g. the proof of Proposition 9.3.1 in Ref. [265]).

Remark 7: The strong rigidity theorem may not hold in theories of gravity distinct from general relativity (hypothesis 2 of Property 5.25). For instance, the so-called *disformed Kerr solution* [16, 53] describes a $n = 4$ stationary black hole in a higher-order scalar-tensor theory (belonging to the so-called DHOST family, for *Degenerate Higher-Order Scalar-Tensor* [337]), the event horizon of which does not seem to be a Killing horizon [16].

Historical note : Stephen Hawking first established the strong rigidity theorem in 1972 [259] for a spacetime of dimension $n = 4$ and assuming that there exists a past event horizon (i.e. a white hole region) in addition to the future event horizon \mathcal{H} . The theorem was restated and proved without the white hole hypothesis in Hawking & Ellis' famous textbook published in 1973 [265] (Proposition 9.3.6). An account by Hawking himself about this reformulation can be found in Sec. 8 of Ref. [260]. The demonstration presented in Ref. [265] relies on the vacuum Einstein equation, but it is noted there that “*similar arguments hold in the presence of matter fields, like the electromagnetic or scalar fields, which obey well-behaved hyperbolic equations*”. Some details in Hawking's proof have been fixed by Piotr T. Chruściel in 1997 [115]. The extension to spacetimes of dimension $n \geq 4$ has been performed independently by Stefan Hollands, Akihiro Ishibashi and Robert M. Wald in 2007 [281] and by Vincent Moncrief and James Isenberg in 2008 [371]. The treatment of the electrovacuum case is done explicitly in Hollands, Ishibashi & Wald's work [281].

The strong rigidity theorem relies on the rather strong assumption that the manifolds and fields are analytic. On physical grounds, it would be desirable to assume only *smooth* manifolds and fields. In 1999, H. Friedrich, I. Rácz, and R.M. Wald [199] could remove the analyticity hypothesis but only towards the *interior* of the black hole, i.e. they could prove that the vector χ normal to \mathcal{H} can be extended to a Killing vector of the black hole interior. Another partial success has been achieved in 2014 by S. Alexakis, A.D. Ionescu and S. Klainerman [11], who proved the strong rigidity theorem without the analyticity assumption, but only for slowly rotating black holes. See Ref. [287] for a review of the progresses in removing the analyticity hypothesis.

The strong rigidity theorem has been extended beyond general relativity (assumption 2 in Property 5.25) by Hollands, Ishibashi and Reall in 2023 [280]. These authors have shown that, under some assumptions, the theorem holds in gravity theories with actions involving quadratic and higher orders combinations of the curvature tensor.

It turns out that, for each connected component of the even horizon, the Killing vector χ is expressible as a linear combination of the stationary Killing vector ξ and some Killing vectors generating axisymmetries:

Property 5.26: axisymmetry of stationary black holes

Under the same hypotheses as in Property 5.25, there exist L Killing vectors $\eta_{(1)}, \dots, \eta_{(L)}$ with $1 \leq L \leq [(n - 1)/2]$, which have closed orbits of period 2π , such that

$$[\xi, \eta_{(i)}] = 0 \quad \text{and} \quad [\eta_{(i)}, \eta_{(j)}] = 0 \quad (1 \leq i, j \leq L) \quad (5.80)$$

and the Killing vector χ normal to \mathcal{H} writes

$$\chi = \xi + \Omega^{(1)}\eta_{(1)} + \cdots + \Omega^{(L)}\eta_{(L)}, \quad (5.81)$$

where $\Omega^{(1)}, \dots, \Omega^{(L)}$ are L real constants, all of whose ratios are irrational.

For $n = 4$, one has $[(n - 1)/2] = 1$, so that necessarily $L = 1$. Denoting $\eta_{(1)}$ simply by η and $\Omega^{(1)}$ by $\Omega_{\mathcal{H}}$, Eq. (5.81) reduces to

$$\chi = \xi + \Omega_{\mathcal{H}}\eta. \quad (5.82)$$

The constant $\Omega_{\mathcal{H}}$ is then called the *black hole rotation velocity* if \mathcal{H} coincides with the whole event horizon (i.e. if the event horizon is connected).

Since it has closed orbits, the isometry group generated by each Killing vector $\eta_{(i)}$ is $\text{SO}(2)$. Each isometry is thus an axisymmetry, according to the definition given in Sec. 5.3.6. For $n = 4$, Eq. (5.82) is equivalent to Eq. (5.78). The constancy of $\Omega_{\mathcal{H}}$ in Eq. (5.82) can be interpreted by saying that the event horizon is rotating rigidly (no differential rotation) and justifies the name *rigidity theorem*. For $n > 4$, we refer to the original articles [281, 371] for the proof of Property 5.26.

Remark 8: Property 5.26 is often considered as the second part of the strong rigidity theorem. In some statements, it is even incorporated into the strong rigidity theorem (e.g. Proposition 9.3.6 of Ref. [265]).

Remark 9: If the event horizon has various connected components, the Killing vectors $\eta_{(i)}$ and the constants $\Omega^{(i)}$ may vary from one connected component to the other. In particular, the horizon-generating Killing vector χ may not be the same for each connected component. For instance, in dimension $n = 5$, the *black saturn* found by H. Elvang and P. Figueras in 2007 [183] is a stationary black hole, the event horizon of which has two connected components: \mathcal{H}_p (the “planet”) and \mathcal{H}_r (the “ring”), with topology $\mathbb{R} \times \mathbb{S}^3$ and $\mathbb{R} \times \mathbb{S}^1 \times \mathbb{S}^2$ respectively. For both \mathcal{H}_p and \mathcal{H}_r , $L = 1$ and the two components share the same axisymmetry Killing vector $\eta_{(1)}$, but they may have different angular velocities: $\Omega_{\mathcal{H}_p}^{(1)} \neq \Omega_{\mathcal{H}_r}^{(1)}$, and hence different normal Killing vectors: $\chi_{\mathcal{H}_p} \neq \chi_{\mathcal{H}_r}$.

In the literature, the strong rigidity theorem is sometimes called simply the *rigidity theorem*, without the qualifier *strong* (e.g. Refs. [107, 279]). The terminology *strong rigidity*, employed in Refs. [102, 122, 274], aims to distinguish from the so-called *weak rigidity theorem* proved by Carter in 1969 (corollary to Theorem 1 in Ref. [91]; see also [93], Sec. 4.5 of Ref. [100] and Sec. 6.3 of Ref. [274]). The latter asserts that the event horizon of a black hole in a stationary, axisymmetric and circular¹⁵ spacetime must be a Killing horizon. The *strong rigidity theorem* is stronger in that it assumes only the stationarity of the spacetime, the axisymmetry becoming a consequence (Property 5.26). On the other hand, the weak rigidity theorem is stronger in so far as it does not rely on the Einstein equation, nor on the assumption of analyticity.

By the very definition of stationarity, the Killing vector field ξ is timelike in the vicinity of \mathcal{I}^+ and \mathcal{I}^- . If ξ is spacelike on some parts of \mathcal{H} , as assumed in this section, by continuity it must be spacelike in some part of the domain of outer communications $\langle\langle \mathcal{M} \rangle\rangle$ near \mathcal{H} . The

¹⁵A stationary and axisymmetric spacetime is said to be *circular* iff the $\mathbb{R} \times \text{SO}(2)$ group action is *orthogonally transitive*, i.e. the 2-surfaces generated by the two commuting Killing vectors ξ and η are orthogonal to a family of $(n - 2)$ -dimensional surfaces [91].

simplest configuration is when ξ is spacelike in some connected region $\mathcal{G} \subset \langle\langle \mathcal{M} \rangle\rangle$ around \mathcal{H} , null at the boundary of \mathcal{G} and timelike outside \mathcal{G} up to \mathcal{I}^+ and \mathcal{I}^- . The subset \mathcal{G} is called the **ergoregion** and its boundary $\mathcal{E} := \partial\mathcal{G}$ the **ergosphere**. We shall discuss it further in Chaps. 10 and 11, especially in connection with the so-called *Penrose process* (Sec. 11.3.2).

5.4.3 Extremal black holes

Properties 5.23 and 5.25, show that, under some rather generic hypotheses (at least as far as general relativity is concerned, cf. Remark 7), the event horizon of a stationary black hole, if connected, is a Killing horizon. Now, in Sec. 3.3.6 we have classified Killing horizons in two categories: the degenerate ones (zero surface gravity) and the non-degenerate ones (nonzero surface gravity). This justifies the following definition:

A black hole in a stationary spacetime is said to be **extremal** (resp. **non-extremal**) if, and only if, its event horizon \mathcal{H} is a degenerate (resp. non-degenerate) Killing horizon.

The surface gravity κ of an extremal black hole is identically zero and the null generators of its horizon are *complete* geodesics: the parameter t associated to the stationary Killing ξ is an affine parameter ranging in the whole of \mathbb{R} . Standard examples are the extremal Reissner-Norström black hole (presented briefly in Sec. 5.6.1) and the extremal Kerr black hole (to be discussed in Chap. 13).

5.5 The generalized Smarr formula

As shown in the previous section, under some hypotheses, the connected components of the event horizon of a stationary black hole are Killing horizons; this gives rise to a nice formula connecting the mass, angular momentum and area of these objects. After having established the formula in the most general case (Sec. 5.5.1), we shall specialize it to electrovacuum black holes (Sec. 5.5.2).

5.5.1 General form

Property 5.27: generalized Smarr formula

Let (\mathcal{M}, g) be a stationary spacetime of dimension $n \geq 4$ (stationary Killing vector ξ) that contains a black hole, the event horizon of which has $K \geq 1$ connected components $\mathcal{H}_1, \dots, \mathcal{H}_K$. We shall assume that each \mathcal{H}_k is a Killing horizon with respect to a Killing vector χ_k . This is guaranteed with $\chi_k = \xi$ if \mathcal{H}_k is non-rotating (ξ null on all \mathcal{H}_k ; Property 5.23), while if \mathcal{H}_k is rotating (ξ spacelike on some parts of \mathcal{H}_k), this holds under the hypotheses of the strong rigidity theorem (Property 5.25), with $\chi_k = \xi + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \eta_{\mathcal{H}_k(i)}$, where the $\Omega_{\mathcal{H}_k}^{(i)}$ are constants and the $\eta_{\mathcal{H}_k(i)}$ are axisymmetric Killing vectors (Property 5.26). Being a non-expanding horizon, each \mathcal{H}_k has a well defined area A_k (Property 3.1). Let κ_k be the surface gravity of \mathcal{H}_k , i.e. the coefficient such that $\nabla_{\chi_k} \chi_k = \kappa_k \chi_k$ on \mathcal{H}_k [cf. Eq. (3.29)].

Let us assume that the null dominance condition (3.43) is fulfilled on \mathcal{H}_k or that \mathcal{H}_k is part of a bifurcate Killing horizon; by the zeroth law of black hole dynamics (Property 3.10 or Property 3.16), this implies that κ_k is constant. Then the Komar mass $M_{\mathcal{H}_k}$ over any cross-section of \mathcal{H}_k , as defined by Eq. (5.36), obeys

$$M_{\mathcal{H}_k} = \frac{n-2}{2(n-3)} \left(\frac{\kappa_k A_k}{4\pi} + 2 \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} J_{\mathcal{H}_k}^{(i)} \right), \quad (5.83)$$

where $J_{\mathcal{H}_k}^{(i)}$ is the Komar angular momentum with respect to the axisymmetric Killing vector $\eta_{\mathcal{H}_k(i)}$ over any cross-section of \mathcal{H}_k as given by Eq. (5.63). In Eq. (5.83), it is intended that $L_k = 0$ if $\chi_k = \xi$, so that there is no sum over i in that case. Furthermore, the Komar mass at infinity (cf. Property 5.16) obeys the **generalized Smarr formula**:

$$\boxed{\frac{2(n-3)}{n-2} M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k A_k}{4\pi} + 2 \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} J_{\mathcal{H}_k}^{(i)} \right) + \frac{1}{4\pi} \int_{\Sigma} \mathbf{R}(\xi, n) \sqrt{\gamma} d^{n-1}x}, \quad (5.84)$$

where Σ is any asymptotically flat spacelike hypersurface, the inner boundary of which is a cross-section of the event horizon, i.e. some union of cross-sections of the connected components \mathcal{H}_k , \mathbf{R} is the Ricci tensor of \mathbf{g} and n is the future-directed unit normal to Σ . If \mathbf{g} obeys the Einstein equation (1.42) with $\Lambda = 0$, this formula can be rewritten as

$$\begin{aligned} \frac{2(n-3)}{n-2} M_\infty &= \sum_{k=1}^K \left(\frac{\kappa_k A_k}{4\pi} + 2 \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} J_{\mathcal{H}_k}^{(i)} \right) \\ &\quad + 2 \int_{\Sigma} \left(\mathbf{T}(\xi, n) - \frac{T\xi \cdot n}{n-2} \right) \sqrt{\gamma} d^{n-1}x, \end{aligned} \quad (5.85)$$

where \mathbf{T} is the matter energy-momentum tensor.

Proof. Let us evaluate the Komar mass $M_{\mathcal{H}_k}$ via formula (5.45). If we denote the cross-section of \mathcal{H}_k by \mathcal{S}_k and we substitute ξ by its expression arising from Eq. (5.81), we get¹⁶

$$\begin{aligned} M_{\mathcal{H}_k} &= -\frac{n-2}{16\pi(n-3)} \int_{\mathcal{S}_k} \nabla_\mu \left(\chi_\nu - \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \eta_{(i)\nu} \right) dS^{\mu\nu} \\ &= -\frac{n-2}{16\pi(n-3)} \left[\underbrace{\int_{\mathcal{S}_k} \nabla_\mu \chi_\nu dS^{\mu\nu}}_{-2\kappa_k A_k} - \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \underbrace{\int_{\mathcal{S}_k} \nabla_\mu \eta_{(i)\nu} dS^{\mu\nu}}_{16\pi J_{\mathcal{H}_k}^{(i)}} \right]. \end{aligned}$$

The second surface integral being $16\pi J_{\mathcal{H}_k}^{(i)}$ follows directly from Eq. (5.65). As for the first integral, the value $-2\kappa_k A_k$ is obtained by introducing a null vector \mathbf{k} normal to \mathcal{S}_k such that

¹⁶Here we drop the index k on χ and the label \mathcal{H}_k on $\Omega^{(i)}$ and $\eta_{(i)}$ for clarity, given that the integral regards a single connected component \mathcal{H}_k .

$\chi \cdot k = -1$. At each point $p \in \mathcal{S}_k$, the pair (χ, k) is then a null basis of the 2-plane $T_p^\perp \mathcal{S}_k$ (cf. Fig. 2.10 where ℓ stands for χ) and we may rewrite the area element normal bivector $dS^{\alpha\beta}$ as¹⁷

$$dS^{\alpha\beta} = (\chi^\alpha k^\beta - k^\alpha \chi^\beta) \sqrt{q} d^{n-2}x. \quad (5.86)$$

We have then

$$\nabla_\mu \chi_\nu dS^{\mu\nu} = \left(\chi^\mu \nabla_\mu \chi_\nu k^\nu - k^\mu \chi^\nu \underbrace{\nabla_\mu \chi_\nu}_{-\nabla_\nu \chi_\mu} \right) \sqrt{q} d^{n-2}x = 2 \underbrace{\chi^\mu \nabla_\mu \chi_\nu}_{\kappa_k \chi_\nu} k^\nu \sqrt{q} d^{n-2}x.$$

Since $\chi_\nu k^\nu = -1$, we get

$$\nabla_\mu \chi_\nu dS^{\mu\nu} = -2\kappa_k \sqrt{q} d^{n-2}x. \quad (5.87)$$

Given that κ_k is constant (Property 3.10 or Property 3.16), the integral of the above expression over \mathcal{S}_k reduces to $-2\kappa_k A_k$, by the very definition of the area A_k [cf. Eq. (3.3)]. This establishes Eq. (5.83). Finally, Eq. (5.84) follows from Eqs. (5.58), (5.54) and (5.83), once one has noticed that $\mathcal{S}_{\text{int}} = \bigcup_{k=1}^K \mathcal{S}_k$, so that $M_{\mathcal{S}_{\text{int}}} = \sum_{k=1}^K M_{\mathcal{H}_k}$. \square

For vacuum black holes of general relativity, $\mathbf{T} = 0$, the null dominance condition is trivially fulfilled and the generalized Smarr formula (5.85) simplifies since the integral term disappears.

Example 8 (black saturn): For the *black saturn* solution discussed in Remark 9, one has $n = 5$, $K = 2$, $L_1 = L_2 = 1$ and formula (5.83) reduces to

$$M_{\mathcal{H}_k} = \frac{3}{32\pi} \kappa_k A_k + \frac{3}{2} \Omega_{\mathcal{H}_k} J_{\mathcal{H}_k}, \quad k \in \{1, 2\}, \quad (5.88)$$

where $k = 1$ corresponds to the “planet” component of the event horizon and $k = 2$ corresponds to the “ring” component. Equation (5.88) coincides with the two formulas given in Eq. (3.43) of the discovery article [183], which were obtained from the explicit expressions of $M_{\mathcal{H}_k}$, A_k , $J_{\mathcal{H}_k}$, κ_k and $\Omega_{\mathcal{H}_k}$ in terms of the solution parameters. For the same solution, the Smarr formula (5.85) with $\mathbf{T} = 0$ is recovered by combining Eqs. (3.36) and (3.43) of Ref. [183].

It is worth to specialize the generalized Smarr formula to a 4-dimensional spacetime and to a connected event horizon, by setting $n = 4$, $K = 1$ and $L_1 = 1$ (cf. Property 5.26) in Eqs. (5.83), (5.84) and (5.85):

Property 5.28: 4-dimensional generalized Smarr formula

Let (\mathcal{M}, g) be a 4-dimensional stationary spacetime (stationary Killing vector ξ) containing a black hole with a connected event horizon \mathcal{H} . If \mathcal{H} is non-rotating (ξ null on all \mathcal{H}), it is necessarily a Killing horizon with respect to ξ . If \mathcal{H} is rotating (ξ partly spacelike on \mathcal{H}), we shall assume that the hypotheses of the strong rigidity theorem (Property 5.25) are fulfilled, so that \mathcal{H} is a Killing horizon with respect to the Killing vector $\chi = \xi + \Omega_{\mathcal{H}} \eta$ [Eq. (5.82)], where $\Omega_{\mathcal{H}}$ is the angular velocity of \mathcal{H} and η is an axisymmetric Killing vector. We shall combine these two cases by stating that the former corresponds to $\Omega_{\mathcal{H}} = 0$, so that $\chi = \xi$. The surface gravity κ of \mathcal{H} is defined by the identity $\nabla_\chi \chi \stackrel{\mathcal{H}}{=} \kappa \chi$ [cf. Eq. (3.29)]. Assuming that the null dominance condition (3.43) is fulfilled on \mathcal{H} or that \mathcal{H}

¹⁷This follows readily by setting $n = (\chi + k)/\sqrt{2}$ and $s = (\chi - k)/\sqrt{2}$ in formula (5.42).

is part of a bifurcate Killing horizon, the zeroth law of black hole dynamics (Property 3.10 or 3.16) leads to $\kappa = \text{const}$. Then the Komar mass $M_{\mathcal{H}}$ over any cross-section of \mathcal{H} [cf. Eq. (5.36)] obeys

$$M_{\mathcal{H}} = \frac{\kappa}{4\pi} A + 2\Omega_{\mathcal{H}} J_{\mathcal{H}}, \quad (5.89)$$

where A is the area of \mathcal{H} (cf. Property 3.1) and $J_{\mathcal{H}}$ is the Komar angular momentum over any cross-section of \mathcal{H} [cf. Eq. (5.63)]. Furthermore, the Komar mass at infinity (cf. Property 5.16) obeys

$$M_{\infty} = \frac{\kappa}{4\pi} A + 2\Omega_{\mathcal{H}} J_{\mathcal{H}} + \frac{1}{4\pi} \int_{\Sigma} \mathbf{R}(\boldsymbol{\xi}, \mathbf{n}) \sqrt{\gamma} d^3x, \quad (5.90)$$

where Σ is an asymptotically flat spacelike hypersurface, the inner boundary of which is a cross-section of \mathcal{H} , \mathbf{R} is the Ricci tensor of \mathbf{g} and \mathbf{n} is the future-directed unit normal to Σ . If \mathbf{g} obeys the Einstein equation (1.42) with $\Lambda = 0$, this formula can be rewritten as

$$M_{\infty} = \frac{\kappa}{4\pi} A + 2\Omega_{\mathcal{H}} J_{\mathcal{H}} + \int_{\Sigma} (2\mathbf{T}(\boldsymbol{\xi}, \mathbf{n}) - T \boldsymbol{\xi} \cdot \mathbf{n}) \sqrt{\gamma} d^3x, \quad (5.91)$$

where \mathbf{T} is the matter energy-momentum tensor.

Example 9 (Schwarzschild black hole): The Schwarzschild black hole will be discussed in details in Chap. 6; however, we have sufficiently studied its event horizon in various examples of the preceding chapters to check that the Smarr formula holds for it. From Example 10 of Chap. 2, we have $\kappa = 1/(4m)$ [Eq. (2.29)], while from Example 3 of Chap. 3 we have $A = 16\pi m^2$ [Eq. (3.4)]. Hence $\kappa A/(4\pi) = m$. Since we have seen in Example 5 of the current chapter that the parameter m is nothing but the Komar mass at infinity M_{∞} , we get

$$M_{\infty} = \frac{\kappa}{4\pi} A. \quad (5.92)$$

This is nothing but the Smarr formula (5.91) for a static ($\Omega_{\mathcal{H}} = 0$) black hole in vacuum ($\mathbf{T} = 0$).

5.5.2 Smarr formula for charged black holes

Beside the vacuum case, the generalized Smarr formulas (5.85) and (5.91) simplifies significantly for electrovacuum spacetimes, i.e. spacetimes ruled by general relativity and for which \mathbf{T} is the energy-momentum tensor of a source-free electromagnetic field (cf. Sec. 1.5.2). To perform the computation of the integral involving \mathbf{T} , we need first to characterize a stationary electromagnetic field on a Killing horizon:

Property 5.29: electromagnetic field on a Killing horizon

Let $(\mathcal{M}, \mathbf{g})$ be a n -dimensional asymptotically flat spacetime endowed with some Killing vector field $\boldsymbol{\chi}$ and containing a Killing horizon \mathcal{H} with respect to $\boldsymbol{\chi}$. Let us assume that $(\mathcal{M}, \mathbf{g})$ is endowed with an electromagnetic field \mathbf{F} that respects the symmetry generated by $\boldsymbol{\chi}$, i.e. that obeys $\mathcal{L}_{\boldsymbol{\chi}} \mathbf{F} = 0$. Furthermore let us assume that the electrovacuum Einstein

equation is fulfilled by (g, \mathbf{F}) (cf. Sec. 1.5.2). Then the pseudo-electric field^a \mathbf{E} defined by

$$\mathbf{E} := \mathbf{F}(., \chi) \iff E_\alpha := F_{\alpha\mu}\chi^\mu \quad (5.93)$$

is an exact 1-form:

$$\mathbf{E} = -d\Phi, \quad (5.94)$$

where the scalar potential Φ is expressible in terms of any electromagnetic potential \mathbf{A} that obeys the χ -symmetry (i.e. any 1-form \mathbf{A} such that $\mathbf{F} = d\mathbf{A}$ and $\mathcal{L}_\chi \mathbf{A} = 0$) by

$$\Phi = -\langle \mathbf{A}, \chi \rangle + \text{const} \iff \Phi = -A_\mu \chi^\mu + \text{const.} \quad (5.95)$$

We shall choose the additive constant so that $\Phi \rightarrow 0$ in the asymptotic flat end of (\mathcal{M}, g) . This determines Φ uniquely and we shall call it the **Killing electric potential associated to χ** . Furthermore, on \mathcal{H} , the vector field $\vec{\mathbf{E}}$ is collinear to the null normal χ :

$$\vec{\mathbf{E}} \stackrel{\mathcal{H}}{=} \mu_0 \sigma \chi \iff E^\alpha \stackrel{\mathcal{H}}{=} \mu_0 \sigma \chi^\alpha, \quad (5.96)$$

where the scalar field σ defined on \mathcal{H} can be interpreted as the *effective electric charge density* of any cross-section \mathcal{S} of \mathcal{H} (Eq. (5.98) below). Finally, the Killing electric potential Φ is constant on \mathcal{H} :

$$\Phi \stackrel{\mathcal{H}}{=} \Phi_{\mathcal{H}} = \text{const.} \quad (5.97)$$

The constant $\Phi_{\mathcal{H}}$ is called the **electric potential of the Killing horizon \mathcal{H}** .

^aIf χ were a unit timelike vector, \mathbf{E} would be a genuine electric field: the one measured by the observer of 4-velocity χ .

Proof. Let \mathbf{A} be some electromagnetic potential associated to \mathbf{F} ($\mathbf{F} = d\mathbf{A}$) such that $\mathcal{L}_\chi \mathbf{A} = 0$. Thanks to the Cartan identity (A.95), we get $0 = \mathcal{L}_\chi \mathbf{A} = \chi \cdot d\mathbf{A} + d(\chi \cdot \mathbf{A}) = \chi \cdot \mathbf{F} + d\langle \mathbf{A}, \chi \rangle = -\mathbf{E} + d\langle \mathbf{A}, \chi \rangle$. This proves Eq. (5.94) with Φ given by Eq. (5.95). Let \mathbf{A}' be another electromagnetic potential associated to \mathbf{F} such that $\mathcal{L}_\chi \mathbf{A}' = 0$. We have the change-of-gauge relation $\mathbf{A}' = \mathbf{A} + d\Psi$, where Ψ is a scalar field. Then $\mathcal{L}_\chi d\Psi = \mathcal{L}_\chi \mathbf{A}' - \mathcal{L}_\chi \mathbf{A} = 0 - 0 = 0$. Invoking the commutation relation (A.96), we get $d\mathcal{L}_\chi \Psi = 0$. Hence, $\mathcal{L}_\chi \Psi = \text{const}$. Given the identity $\langle d\Psi, \chi \rangle = \mathcal{L}_\chi \Psi$, it follows that $\langle \mathbf{A}', \chi \rangle = \langle \mathbf{A}, \chi \rangle + \langle d\Psi, \chi \rangle = \langle \mathbf{A}, \chi \rangle + \text{const}$. This proves that, up to some additive constant, the scalar field Φ defined by Eq. (5.95) does not depend on the choice of the electromagnetic potential \mathbf{A} . Let us now turn to relation (5.96). First, it is trivial that $\vec{\mathbf{E}}$ is tangent to \mathcal{H} , given that $\chi \cdot \vec{\mathbf{E}} = \langle \mathbf{E}, \chi \rangle = \mathbf{F}(\chi, \chi) = 0$ by antisymmetry of \mathbf{F} . Second, given that the Killing horizon \mathcal{H} is a non-expanding horizon and that any electromagnetic field obeys the null energy condition (cf. Sec. 2.4.2), property (3.8) holds: $\mathbf{R}(\chi, \chi) \stackrel{\mathcal{H}}{=} 0$. Since χ is null on \mathcal{H} , we can invoke the Einstein equation (1.40) to transform this relation into $\mathbf{T}(\chi, \chi) \stackrel{\mathcal{H}}{=} 0$, where \mathbf{T} is the energy-momentum tensor of the electromagnetic field. Using expression (1.52) for \mathbf{T} , we get

$$\underbrace{F_{\mu\rho}\chi^\rho}_{E_\mu} \underbrace{F_{\sigma}^{\mu}\chi^\sigma}_{E^\mu} - \frac{1}{4} F_{\mu\nu} F^{\mu\nu} \underbrace{g_{\rho\sigma}\chi^\rho\chi^\sigma}_0 \stackrel{\mathcal{H}}{=} 0,$$

i.e. $\vec{E} \cdot \vec{E} \stackrel{\mathcal{H}}{=} 0$. Thus \vec{E} is null vector on \mathcal{H} . Being tangent to \mathcal{H} , it is then necessarily collinear to the null normal to \mathcal{H} , namely χ , hence there exists a scalar field σ on \mathcal{H} such that Eq. (5.96) holds. Finally, since on \mathcal{H} , \vec{E} is collinear to \mathcal{H} 's null normal χ , we have $\vec{E} \cdot v = 0$ for any vector field v tangent to \mathcal{H} . This identity can be rewritten as $\langle E, v \rangle = 0$ or, in view of Eq. (5.94), as $\langle d\Phi, v \rangle = 0$. It follows that the scalar field Φ is constant on \mathcal{H} , hence Eq. (5.97). \square

The fact that σ is an effective¹⁸ surface charge density follows from:

Property 5.30: electric charge of a Killing horizon

Under the same assumptions as in Property 5.29, the electric charge within any cross-section \mathcal{S} of the Killing horizon \mathcal{H} is

$$Q_{\mathcal{H}} = \frac{1}{\mu_0} \int_{\mathcal{S}} \star F = \frac{1}{2\mu_0} \int_{\mathcal{S}} F_{\mu\nu} dS^{\mu\nu} = \int_{\mathcal{S}} \sigma \sqrt{q} d^{n-2}x. \quad (5.98)$$

The quantity $Q_{\mathcal{H}}$ is independent of the choice of the cross-section \mathcal{S} and is called the **electric charge of the Killing horizon \mathcal{H}** .

Proof. The first equality in Eq. (5.98) is the Gauss-law expression of the electric charge (see e.g. Eq. (8.201) of Ref. [463] or Eq. (18.40) of Ref. [227]), while the second equality results from Lemma 5.10. Now, using expression (5.86) for $dS^{\mu\nu}$, we have

$$\begin{aligned} F_{\mu\nu} dS^{\mu\nu} &= F_{\mu\nu} (\chi^\mu k^\nu - k^\mu \chi^\nu) \sqrt{q} d^{n-2}x = (-E_\nu k^\nu - k^\mu E_\mu) \sqrt{q} d^{n-2}x \\ &= -2\mu_0 \sigma \underbrace{\chi_\mu k^\mu}_{-1} \sqrt{q} d^{n-2}x = 2\mu_0 \sigma \sqrt{q} d^{n-2}x, \end{aligned} \quad (5.99)$$

where we have used successively Eqs. (5.93) and (5.96). This establishes the last equality in Eq. (5.98). To show that $Q_{\mathcal{H}}$ is independent of the choice of \mathcal{S} , it suffices to consider the part \mathcal{W} of \mathcal{H} that is delineated by two cross-sections \mathcal{S} and \mathcal{S}' , as in Fig. 2.9. \mathcal{W} is then a $(n-1)$ -dimensional manifold with boundary $\partial\mathcal{W} = \mathcal{S} \cup \mathcal{S}'$ and we may apply Stokes' theorem (A.94) on \mathcal{W} to the $(n-2)$ -form $\star F$; taking into account the outward orientation of $\partial\mathcal{W}$ involved in Stokes formula (A.94), we get

$$\underbrace{\int_{\mathcal{S}'} \star F}_{\mu_0 Q'_{\mathcal{H}}} - \underbrace{\int_{\mathcal{S}} \star F}_{\mu_0 Q_{\mathcal{H}}} = \int_{\mathcal{W}} d\star F.$$

Now, thanks to the second source-free Maxwell equation, $d \star F = 0$ [Eq. (1.50) with $j = 0$], so that the right-hand side of the above equation identically vanishes. Hence $Q'_{\mathcal{H}} = Q_{\mathcal{H}}$. \square

We are now in position to state:

¹⁸ σ is not a genuine surface charge density since there are no electrically charged particles on \mathcal{H} , given the electrovacuum hypothesis (vanishing of the electric n -current density j , cf. Sec. 1.5.2); see Damour's work [149] for interpreting σ as the component along χ of an effective surface current 4-vector on \mathcal{H} .

Property 5.31: generalized Smarr formula for charged black holes

Let $(\mathcal{M}, \mathbf{g})$ be a stationary spacetime of dimension $n \geq 4$, with stationary Killing vector ξ , endowed with a source-free electromagnetic field \mathbf{F} such that (\mathbf{g}, \mathbf{F}) obeys the electrovacuum Einstein equation (1.54). Let us assume that $(\mathcal{M}, \mathbf{g})$ contains a black hole, the event horizon of which has $K \geq 1$ connected components $\mathcal{H}_1, \dots, \mathcal{H}_K$. Each \mathcal{H}_k ($1 \leq k \leq K$) is assumed to be a Killing horizon with respect to the Killing vector $\chi = \xi + \sum_{i=1}^L \Omega^{(i)} \eta_{(i)}$, where $0 \leq L \leq [(n-1)/2]$, the $\Omega^{(i)}$ are constants and the $\eta_{(i)}$ are axisymmetric Killing vectors^a. One may have $L = 0$, i.e. $\chi = \xi$ (static configuration). The Smarr formula is then

$$\frac{2(n-3)}{n-2} M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k A_k}{4\pi} + \frac{2(n-3)}{n-2} \Phi_{\mathcal{H}_k} Q_{\mathcal{H}_k} \right) + 2 \sum_{i=1}^L \Omega^{(i)} J_\infty^{(i)}, \quad (5.100)$$

where M_∞ is the Komar mass at infinity, $J_\infty^{(i)}$ is the Komar angular momentum at infinity with respect to the axisymmetric Killing vector $\eta_{(i)}$, κ_k is the surface gravity of \mathcal{H}_k , A_k is the area of \mathcal{H}_k , $\Phi_{\mathcal{H}_k}$ is the electric potential of \mathcal{H}_k , as defined by Property 5.29, and $Q_{\mathcal{H}_k}$ is the electric charge of \mathcal{H}_k , as defined by Property 5.30.

^aNote that contrary to the more general setting of Property 5.27, all the connected components \mathcal{H}_k are assumed to be Killing horizons with respect to the same Killing vector χ ; in other words, the rotation velocities $\Omega^{(i)}$ and axisymmetric Killing vectors $\eta_{(i)}$ are the same for all the \mathcal{H}_k 's.

Proof. The electromagnetic field energy-momentum tensor (1.52) obeys the null dominant energy condition (3.46) [320], so that the surface gravities κ_k are constant over each Killing horizon \mathcal{H}_k (cf. Property 3.10). All the hypotheses of Property 5.27 are thus fulfilled and the Smarr formula (5.84) holds. Thanks to the independence of $\Omega^{(i)}$ and $\eta_{(i)}$ from \mathcal{H}_k , we may permute the sums over k and i in it and write

$$\frac{2(n-3)}{n-2} M_\infty = \sum_{k=1}^K \frac{\kappa_k A_k}{4\pi} + 2 \sum_{i=1}^L \Omega^{(i)} \sum_{k=1}^K J_{\mathcal{H}_k}^{(i)} + \frac{1}{4\pi} \int_{\Sigma} \mathbf{R}(\xi, \mathbf{n}) \sqrt{\gamma} d^{n-1}x,$$

where Σ is an asymptotically flat spacelike hypersurface, of future-directed unit normal \mathbf{n} to Σ and of inner boundary a cross-section \mathcal{S}_{int} of the event horizon, i.e. $\mathcal{S}_{\text{int}} = \bigcup_{k=1}^K \mathcal{S}_k$, where \mathcal{S}_k is a cross-section of \mathcal{H}_k . Now, according to Eq. (5.69) with \mathcal{S}_{ext} going to infinity, we have

$$\sum_{k=1}^K J_{\mathcal{H}_k}^{(i)} = J_{\mathcal{S}_{\text{int}}}^{(i)} = J_\infty^{(i)} + \frac{1}{8\pi} \int_{\Sigma} \mathbf{R}(\eta_{(i)}, \mathbf{n}) \sqrt{\gamma} d^{n-1}x.$$

Hence

$$\frac{2(n-3)}{n-2} M_\infty = \sum_{k=1}^K \frac{\kappa_k A_k}{4\pi} + 2 \sum_{i=1}^L \Omega^{(i)} J_\infty^{(i)} + \frac{1}{4\pi} \int_{\Sigma} \mathbf{R}(\chi, \mathbf{n}) \sqrt{\gamma} d^{n-1}x, \quad (5.101)$$

where we have used the identity $\xi + \sum_{i=1}^L \Omega^{(i)} \eta_{(i)} = \chi$ to collect all the integrals on Σ . Let us evaluate the integral by using the electrovacuum Einstein equation (1.54), which yields

$$\frac{1}{4\pi} \mathbf{R}(\chi, \mathbf{n}) = \frac{2n^\mu}{\mu_0} \left(F_{\sigma\mu} F_\nu^\sigma \chi^\nu - \frac{1}{2(n-2)} F_{\rho\sigma} F^{\rho\sigma} \chi_\mu \right). \quad (5.102)$$

We recognize the pseudo-electric field $E^\sigma = F_\nu^\sigma \chi^\nu$ [cf. Eq. (5.93)] in the first term in the right-hand side; thanks to Eq. (5.94), we may re-express it in terms of the Killing electric potential Φ :

$$F_{\sigma\mu} F_\nu^\sigma \chi^\nu = F_{\sigma\mu} E^\sigma = -F_{\sigma\mu} \nabla^\sigma \Phi = -\nabla_\sigma \Phi F_\mu^\sigma = -\nabla_\sigma (\Phi F_\mu^\sigma) = \nabla^\sigma (\Phi F_{\mu\sigma}), \quad (5.103)$$

where the last but one equality stems from the source-free Maxwell equation $\nabla_\sigma F^{\alpha\sigma} = 0$ [Eq. (1.48)]. As for the second term in the integrand (5.102), let us rewrite it in terms of some electromagnetic potential 1-form \mathbf{A} that obeys $\mathcal{L}_\chi \mathbf{A} = 0$ and such that $\Phi = -\langle \mathbf{A}, \chi \rangle$ (i.e. such that the constant in Eq. (5.95) is zero). Then $F_{\rho\sigma} = \nabla_\rho A_\sigma - \nabla_\sigma A_\rho$, so that, using the antisymmetry of $F^{\rho\sigma}$ and again the source-free Maxwell equation $\nabla_\sigma F^{\sigma\rho} = 0$,

$$F_{\rho\sigma} F^{\rho\sigma} \chi_\mu = 2\nabla_\sigma A_\rho F^{\sigma\rho} \chi_\mu = 2[\nabla_\sigma (A_\rho F^{\sigma\rho} \chi_\mu) - A_\rho F^{\sigma\rho} \nabla_\sigma \chi_\mu].$$

Now, expressing the property $\mathcal{L}_\chi \mathbf{F} = 0$ via formula (A.86), we have $\mathcal{L}_\chi F^{\mu\rho} = \chi^\sigma \nabla_\sigma F^{\mu\rho} - F^{\sigma\rho} \nabla_\sigma \chi^\mu - F^{\mu\sigma} \nabla_\sigma \chi^\rho = 0$, from which

$$\begin{aligned} A_\rho F^{\sigma\rho} \nabla_\sigma \chi_\mu &= A_\rho F_\mu^\sigma \nabla_\sigma \chi^\rho - A_\rho \chi^\sigma \nabla_\sigma F_\mu^\rho = A_\rho F_\mu^\sigma \nabla_\sigma \chi^\rho - \nabla_\sigma (\chi^\sigma A_\rho F_\mu^\rho) + F_\mu^\rho \chi^\sigma \nabla_\sigma A_\rho \\ &= -\nabla_\sigma (\chi^\sigma A_\rho F_\mu^\rho) + F_\mu^\rho \underbrace{(\chi^\sigma \nabla_\sigma A_\rho + A_\sigma \nabla_\rho \chi^\sigma)}_0 = -\nabla_\sigma (A_\rho F_\mu^\rho \chi^\sigma), \end{aligned}$$

where we have used $\nabla_\sigma \chi^\sigma = 0$, as implied by the Killing equation for χ , in the first line and $\mathcal{L}_\chi A_\rho = 0$ in the second line. Hence, we get

$$F_{\rho\sigma} F^{\rho\sigma} \chi_\mu = 2\nabla_\sigma (A_\rho F^{\sigma\rho} \chi_\mu + A_\rho F_\mu^\rho \chi^\sigma) = 2\nabla^\sigma (A_\rho F_\mu^\rho \chi_\sigma - A_\rho F_\sigma^\rho \chi_\mu).$$

Together with Eq. (5.103), this expression allows us to rewrite Eq. (5.102) as

$$\frac{1}{4\pi} \mathbf{R}(\chi, \mathbf{n}) = \frac{2n^\mu}{\mu_0} \nabla^\nu \Omega_{\mu\nu} \quad \text{with} \quad \Omega_{\mu\nu} := \Phi F_{\mu\nu} + \frac{1}{n-2} (\chi_\mu A_\nu F_\nu^\rho - \chi_\nu A_\mu F_\nu^\rho).$$

Note that $\Omega_{\mu\nu}$ defines a 2-form, since $\Omega_{\mu\nu} = -\Omega_{\nu\mu}$. It follows then from Lemma 5.12 that

$$\begin{aligned} I &:= \frac{1}{4\pi} \int_\Sigma \mathbf{R}(\chi, \mathbf{n}) \sqrt{\gamma} d^{n-1}x = \frac{2}{\mu_0} \int_\Sigma \nabla^\nu \Omega_{\mu\nu} n^\mu \sqrt{\gamma} d^{n-1}x = -\frac{2}{\mu_0} \int_\Sigma \nabla^\nu \Omega_{\mu\nu} dV^\mu \\ &= -\frac{1}{\mu_0} \left(\underbrace{\int_{\mathcal{S}_\infty} \Omega_{\mu\nu} dS^{\mu\nu}}_0 - \int_{\mathcal{S}_{\text{int}}} \Omega_{\mu\nu} dS^{\mu\nu} \right) \end{aligned}$$

We have set the integral on \mathcal{S}_∞ to zero because \mathbf{F} decays to zero sufficiently fast at the asymptotically flat end of Σ (otherwise $(\mathcal{M}, \mathbf{g})$ could not be asymptotically flat). Since $\mathcal{S}_{\text{int}} = \bigcup_{k=1}^K \mathcal{S}_k$, we may write

$$I = \frac{1}{\mu_0} \sum_{k=1}^K \int_{\mathcal{S}_k} \Omega_{\mu\nu} dS^{\mu\nu}.$$

Now, thanks to the antisymmetry of $dS^{\mu\nu}$,

$$\begin{aligned}\Omega_{\mu\nu} dS^{\mu\nu} &= \Phi F_{\mu\nu} dS^{\mu\nu} + \frac{2}{n-2} A_\rho F^\rho_\nu \chi_\mu dS^{\mu\nu} \\ &= \Phi F_{\mu\nu} dS^{\mu\nu} + \frac{2}{n-2} A_\rho F^\rho_\nu (\underbrace{\chi_\mu \chi^\mu}_0 k^\nu - \underbrace{\chi_\mu k^\mu}_{-1} \chi^\nu) \sqrt{q} d^{n-2}x \\ &= \Phi F_{\mu\nu} dS^{\mu\nu} + \frac{2}{n-2} A_\rho E^\rho \sqrt{q} d^{n-2}x,\end{aligned}$$

where use has been made of expression (5.86) for $dS^{\mu\nu}$, of the null character of χ on \mathcal{H}_k and of the definition (5.93) of E . Since the latter obeys (5.96), we have $A_\rho E^\rho = \mu_0 \sigma A_\rho \chi^\rho = -\mu_0 \sigma \Phi$. Hence

$$\Omega_{\mu\nu} dS^{\mu\nu} = \Phi \left(F_{\mu\nu} dS^{\mu\nu} - \frac{2\mu_0}{n-2} \sigma \sqrt{q} d^{n-2}x \right).$$

Given that Φ is a constant $\Phi_{\mathcal{H}_k}$ on each \mathcal{H}_k [Eq. (5.97)], the above expression yields

$$I = \sum_{k=1}^K \Phi_{\mathcal{H}_k} \left(\underbrace{\frac{1}{\mu_0} \int_{\mathcal{H}_k} F_{\mu\nu} dS^{\mu\nu}}_{2Q_{\mathcal{H}_k}} - \underbrace{\frac{2}{n-2} \int_{\mathcal{H}_k} \sigma \sqrt{q} d^{n-2}x}_{Q_{\mathcal{H}_k}} \right) = \frac{2(n-3)}{n-2} \sum_{k=1}^K \Phi_{\mathcal{H}_k} Q_{\mathcal{H}_k},$$

where we have used two expressions of the electric charge of \mathcal{H}_k given by Eq. (5.98). In view of Eq. (5.101), the above value for I proves the Smarr formula (5.100). \square

Property 5.32: 4-dimensional Smarr formula for a connected charged black hole

For $n = 4$ (which implies $L \leq 1$) and for a connected black hole event horizon \mathcal{H} ($K = 1$) of surface gravity κ , area A , angular velocity $\Omega_{\mathcal{H}}$, electric charge Q and electric potential $\Phi_{\mathcal{H}}$, the electrovacuum Smarr formula (5.100) simplifies to

$$M_\infty = \frac{\kappa}{4\pi} A + 2\Omega_{\mathcal{H}} J_\infty + \Phi_{\mathcal{H}} Q.$$

(5.104)

Historical note : Formula (5.104) has been first derived in 1972 (published: 1973) by Larry Smarr [458] in the case of the Kerr-Newman black hole. Actually, Smarr did not obtain it from integral formulas for M_∞ and J_∞ , as presented here; rather he used an explicit expression of M_∞ in terms of A , J_∞ and Q , which holds for the Kerr-Newman solution. After noticing that this expression is a homogeneous function of degree 1/2 of (A, J_∞, Q^2) , he applied Euler's homogeneous function theorem to get Eq. (5.104). The derivation of Smarr formula from generic Komar-type integral formulas for mass and angular momentum, possibly with some non-vacuum exterior, is due to James Bardeen, Brandon Carter and Stephen Hawking in 1973 [39]. They obtained Eq. (5.91) [their Eq. (13)]. The general (i.e. not assuming the Kerr-Newmann metric) $n = 4$ electrovacuum Smarr formula (5.104) has been derived by Brandon Carter in 1973 [95] [his Eq. (9.29)]. Actually, the formula obtained by Carter is more general than Eq. (5.104) since it allows for electric sources (non-vanishing electric 4-current j) and matter in the black

hole exterior (see also Carter's review articles [98] [Eq. (6.323)] and [100] [Eq. (4.68)]). The generalization of the Smarr formula to dimensions $n \geq 4$ has been obtained by Robert Myers and Malcolm Perry in 1986 [375] for connected event horizons in vacuum: they obtained Eq. (5.85) for $K = 1$ and $\mathbf{T} = 0$ [their Eq. (4.6), where $N := n - 1$]. The $n \geq 4$ generalization of the electrovacuum Smarr formula has been obtained by Rabin Banerjee, Bibhas Majhi, Sujoy Modak, and Saurav Samanta in 2010 [31], but only for electrically charged Myers-Perry black holes, which are approximate solutions for small rotation velocities and a single angular momentum parameter, i.e. they obtained Eq. (5.100) for $K = 1$ (connected horizon) and $L = 1$ (single angular momentum) [their Eq. (66)].

5.6 The no-hair theorem

Arguably, the most beautiful achievement in general relativity is the *no-hair theorem*. This is a uniqueness theorem, which basically states that all stationary black holes in our (4-dimensional) Universe are described by the same rather simple solution to the Einstein equation: the Kerr metric (or the Kerr-Newman metric if one allows for an electric charge), which depends on only two numbers: the mass and the angular momentum (plus the electric charge for Kerr-Newman). The absence of any functional degree of freedom justifies the “no-hair” qualifier. Actually, there are various uniqueness theorems, depending on various assumptions on the equilibrium state (static versus rotating), the matter/field content (vacuum, electrovacuum, scalar field, etc.), the Killing-horizon type of the event horizon (non-degenerate versus degenerate) and the spacetime dimension. We shall discuss first uniqueness theorems that regard static black holes in any spacetime dimension (Sec. 5.6.1), before theorems about axisymmetric rotating black holes in dimension 4 (Sec. 5.6.2) and finally “the” no-hair theorem itself, which regards rotating black holes in dimension 4 (Sec. 5.6.3).

5.6.1 Uniqueness theorems for static black holes

The uniqueness theorems for static (i.e. non-rotating) black holes originate from two famous theorems by Werner Israel at the end of the sixties, regarding respectively vacuum [288] and electrovacuum configurations [289]. Both were derived in dimension 4 and for non-degenerate horizons. They have been generalized to higher dimensions and to degenerate horizons; we present these generalized versions here.

Property 5.33: generalized Israel uniqueness theorem (vacuum)

Let $(\mathcal{M}, \mathbf{g})$ be a static and asymptotically flat spacetime of dimension $n \geq 4$ containing a black hole of event horizon \mathcal{H} . If

- \mathbf{g} fulfills the vacuum Einstein equation (1.44),
- $n = 4$ or all the connected components of \mathcal{H} are non-degenerate (i.e. have non-zero surface gravity),

then the domain of outer communications $\langle\langle \mathcal{M} \rangle\rangle$ of $(\mathcal{M}, \mathbf{g})$ is isometric to the domain of

outer communications of a ***n-dimensional Schwarzschild spacetime***, also known as a ***Schwarzschild-Tangherlini spacetime***. This means that there exists a coordinate system $(t, r, \theta_1, \dots, \theta_{n-3}, \varphi)$ on $\langle\langle\mathcal{M}\rangle\rangle$ such that $t \in \mathbb{R}$, $r \in (\mu, +\infty)$, $\theta_i \in (0, \pi)$ ($1 \leq i \leq n-3$), $\varphi \in (0, 2\pi)$ and

$$\mathbf{g} = -\left(1 - \frac{\mu}{r^{n-3}}\right) dt^2 + \left(1 - \frac{\mu}{r^{n-3}}\right)^{-1} dr^2 + r^2 \overset{\circ}{\mathbf{q}}, \quad (5.105)$$

where $\overset{\circ}{\mathbf{q}}$ is the standard round metric^a on the unit sphere \mathbb{S}^{n-2} , which is spanned by $(\theta_1, \dots, \theta_{n-3}, \varphi)$, and the constant $\mu > 0$ is related to the spacetime Komar mass (or ADM mass, cf. Property 5.18) M by

$$\mu = \frac{16\pi M}{(n-2)\Omega_{n-2}}. \quad (5.106)$$

Ω_{n-2} is the area of \mathbb{S}^{n-2} , as given by Eqs. (5.19)-(5.20), so that $\mu = 2M$ for $n = 4$, $\mu = 8M/(3\pi)$ for $n = 5$ and $\mu = 3M/(2\pi)$ for $n = 6$. In particular, the event horizon \mathcal{H} is connected and is a non-degenerate Killing horizon.

^afor $n = 4$, $\overset{\circ}{\mathbf{q}} = d\theta_1^2 + \sin^2 \theta_1 d\varphi^2$; for $n = 5$, $\overset{\circ}{\mathbf{q}} = d\theta_1^2 + \sin^2 \theta_1 (d\theta_2^2 + \sin^2 \theta_2 d\varphi^2)$; etc.

The proof can be found in Ref. [218], which relies on the assumption that the connected components of \mathcal{H} are non-degenerate Killing horizons. The fact that this assumption is not necessary for $n = 4$ has been proven in Ref. [126]. The proof in Ref. [218] relies on the positive mass theorem. An elementary proof specific to the case $n = 4$ can be found in Sec. 8.2 of Straumann's textbook [463] and well as in Chap. 9 of Heusler's textbook [274] and in Robinson's articles [437] and [438] (appendix). See also Ref. [388] for a recent alternative proof.

Remark 1: The fact that \mathcal{H} is connected means that there does not exist any static solution in vacuum general relativity with “multiple black holes”¹⁹. Intuitively, two or more black holes in vacuum would attract each other, making a static configuration impossible.

Historical note : The uniqueness theorem 5.33 has been first proven in the case $n = 4$ by Werner Israel in 1967 [288] under the additional hypotheses that (i) the spacetime is analytic (cf. Remark 4 in Sec. A.2.1), (ii) the isosurfaces $t = \text{const}$, $V := \sqrt{-\xi \cdot \xi} = \text{const}$ are regular and topologically 2-spheres in $\langle\langle\mathcal{M}\rangle\rangle$ (in particular, $dV \neq 0$ in $\langle\langle\mathcal{M}\rangle\rangle$), (iii) the event horizon \mathcal{H} is connected, (iv) \mathcal{H} is non-degenerate and (v) \mathcal{H} 's cross-sections are topologically 2-spheres. Property (v) actually follows from the topology theorem for $n = 4$ (Properties 5.4 and 5.5), which has been established 5 years later by Stephen Hawking [259]; it can therefore be removed from the hypotheses of the theorem. The hypothesis (ii) has been shown unnecessary by David Robinson in 1977 [437] (see also the work [372]). The requirement (iii) (\mathcal{H} connected) has been removed as unnecessary by Gary Bunting and Abul Masood-ul-Alam in 1987 [76], while the requirement (iv) (\mathcal{H} non-degenerate) has been removed by Piotr Chruściel, Harvey Reall and Paul Tod in 2006 [126]. Finally the analyticity hypothesis (i), which was implicit in Israel's study, has been removed in 2010 by Piotr Chruściel and Gregory Galloway [120]. The generalization to dimensions $n \geq 4$ has been achieved by Seungsu Hwang in 1998 [286] and by Gary Gibbons, Daisuke Ida and Tetsuya Shiromizu in 2002 [218]. The generalization (5.105) of Schwarzschild solution to dimensions $n \geq 4$ has been found by Frank Tangherlini in 1963 [472].

¹⁹The quotes indicate that even when \mathcal{H} has various connected components, there is formally a single black hole region in spacetime, albeit not connected.

For the static electrovacuum uniqueness theorem, we shall distinguish the dimension $n = 4$ from $n \geq 5$. Starting by the former, we have:

Property 5.34: generalized Israel uniqueness theorem ($n = 4$ electrovacuum)

Let (\mathcal{M}, g) be a 4-dimensional static and asymptotically flat spacetime endowed with an electromagnetic field \mathbf{F} such that (g, \mathbf{F}) fulfills the electrovacuum Einstein equation (1.54). If (\mathcal{M}, g) contains a black hole, then the domain of outer communications $\langle\langle\mathcal{M}\rangle\rangle$ is isometric to the domain of outer communications of either a Reissner-Nordström black hole or a Majumdar-Papapetrou black hole.

The **Reissner-Nordström black hole** is defined by the event horizon being connected and the existence of a coordinate system (t, r, θ, φ) on $\langle\langle\mathcal{M}\rangle\rangle$ such that $t \in \mathbb{R}$, $r \in (r_{\mathcal{H}}, +\infty)$, $\theta \in (0, \pi)$, $\varphi \in (0, 2\pi)$ and

$$\begin{aligned} g = & - \left(1 - \frac{2M}{r} + \frac{\mu_0}{4\pi} \frac{Q^2 + P^2}{r^2} \right) dt^2 + \left(1 - \frac{2M}{r} + \frac{\mu_0}{4\pi} \frac{Q^2 + P^2}{r^2} \right)^{-1} dr^2 \\ & + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \end{aligned} \quad (5.107a)$$

$$\mathbf{F} = - \frac{\mu_0}{4\pi} \left(\frac{Q}{r^2} dt \wedge dr - P \sin \theta d\theta \wedge d\varphi \right), \quad (5.107b)$$

where M is the Komar mass at infinity (or ADM mass), Q is the black hole electric charge (cf. Property 5.30) and P is its magnetic monopole charge (cf. Remark 2 below). The three constants M , Q and P must fulfill $\sqrt{Q^2 + P^2} \leq \sqrt{\frac{4\pi}{\mu_0}} M$ and the lower bound of r on $\langle\langle\mathcal{M}\rangle\rangle$ is $r_{\mathcal{H}} = M + \sqrt{M^2 - \mu_0/(4\pi)(Q^2 + P^2)}$.

The **Majumdar-Papapetrou black hole** is defined by (i) the event horizon being disconnected, with $K \geq 2$ connected components $(\mathcal{H}_k)_{1 \leq k \leq K}$, which are all degenerate Killing horizons, and (ii) the existence of a coordinate system (t, x, y, z) on $\langle\langle\mathcal{M}\rangle\rangle$ such that

$$g = -U^{-2} dt^2 + U^2 (dx^2 + dy^2 + dz^2), \quad \mathbf{F} = \pm \sqrt{\frac{\mu_0}{4\pi}} U^{-2} dU \wedge dt, \quad (5.108a)$$

$$U := 1 + \sum_{k=1}^K \frac{\mu_k}{\sqrt{(x - x_k)^2 + (y - y_k)^2 + (z - z_k)^2}}. \quad (5.108b)$$

This solution is characterized by K positive constants μ_k , which can be interpreted as the masses of each component \mathcal{H}_k for widely separated configurations, and $3K$ constants (x_k, y_k, z_k) , which are the coordinate locations of the K degenerate Killing horizons \mathcal{H}_k . Note that each (x_k, y_k, z_k) corresponds to a coordinate singularity of the (t, x, y, z) system, so that \mathcal{H}_k appears as the curve $(t, x, y, z) = (t, x_k, y_k, z_k)$, while it is actually a

hypersurface. The electric charge Q_k of \mathcal{H}_k is proportional to μ_k :

$$Q_k = \pm \sqrt{\frac{4\pi}{\mu_0}} \mu_k, \quad (5.109)$$

where the \pm sign is the same as in expression (5.108a) for F ; in particular it must be the same for all the components \mathcal{H}_k , i.e. the charges Q_k are either all positive or all negative.

For a proof, see Secs. 9.3 and 9.4 of Heusler's textbook [274] for the case where all connected components of the event horizon are assumed to be non-degenerate and Ref. [127] for the complementary case.

Remark 2: At the level of elementary particles, a **magnetic monopole** is a particle that generates the radial magnetic field $B = \mu_0 P / (4\pi r^2) \partial_r$ in its inertial rest frame, where P is the magnetic charge. Although predicted by some grand unified theories and string theories, no magnetic monopoles have ever been detected and standard electromagnetism postulates that magnetic monopoles do not exist (see Refs. [359, 422] for reviews). From a formal point of view, magnetic monopoles would render the Maxwell equations more symmetric, since instead of Eq. (1.50), they would write

$$\mathbf{d}F = \mu_0 \star \underline{j}_m \quad \text{and} \quad \mathbf{d} \star F = \mu_0 \star \underline{j}, \quad (5.110)$$

where \underline{j}_m is the 1-form associated to the **magnetic current density** vector j_m , which describes the distribution of elementary magnetic monopoles. In the electrovacuum black hole framework, we are considering the source-free Maxwell equations $\mathbf{d}F = 0$ and $\mathbf{d} \star F = 0$, which are perfectly symmetric in terms of F and $\star F$. As the Reissner-Nordström solution (5.107) shows, a black hole with a non-vanishing magnetic monopole charge can exist even if magnetic monopoles are excluded as elementary particles, i.e. even if $j_m = 0$.

Remark 3: The constraint $\sqrt{Q^2 + P^2} \leq \sqrt{\frac{4\pi}{\mu_0}} M$ in the Reissner-Nordström case enforces the existence of a black hole. If it is not fulfilled, Eqs. (5.107) still define a valid solution to the electrovacuum Einstein equation, but it corresponds to a naked singularity, not to a black hole.

Remark 4: The Majumdar-Papapetrou solution (5.108) with $K = 1$ (connected event horizon) is still a valid static solution to the electrovacuum Einstein equation. It is actually a member of the first family invoked in the uniqueness theorem, namely the Reissner-Nordström one. More precisely, it is an **extremal Reissner-Nordström spacetime**, i.e. a solution (5.107) for which $|Q| = \sqrt{4\pi/\mu_0} M$ and $P = 0$, or equivalently a Reissner-Nordström solution with $P = 0$ and a degenerate event horizon. To see this, it suffices to choose $(x_1, y_1, z_1) = (0, 0, 0)$, to move from coordinates (x, y, z) to spherical coordinates $(\bar{r}, \theta, \varphi)$ via the standard formulas (in particular $\bar{r}^2 = x^2 + y^2 + z^2$) and to introduce a new radial coordinate $r := \bar{r} + \mu_1$. Then $U = r/(r - \mu_1)$ and the Majumdar-Papapetrou solution (5.108) coincides with the Reissner-Nordström solution (5.107) with $M = \mu_1$, $Q = \pm \sqrt{4\pi/\mu_0} \mu_1$ and $P = 0$.

In higher dimensions, the static electrovacuum uniqueness theorem is

Property 5.35: generalized Israel uniqueness theorem ($n \geq 5$ electrovacuum)

Let (\mathcal{M}, g) be a static and asymptotically flat spacetime of dimension $n \geq 5$. Let us assume that (\mathcal{M}, g) is endowed with an electromagnetic field F and contains a black hole of event horizon \mathcal{H} . If

- (g, F) fulfills the electrovacuum Einstein equation (1.54),
- all the connected components of \mathcal{H} are non-degenerate (i.e. have non-zero surface gravity),

then the domain of outer communications $\langle\langle \mathcal{M} \rangle\rangle$ of (\mathcal{M}, g) is isometric to the domain of outer communications of a ***n-dimensional Reissner-Nordström spacetime*** with vanishing magnetic monopole, also known as a ***Reissner-Nordström-Tangherlini spacetime***. This means that there exists a coordinate system $(t, r, \theta_1, \dots, \theta_{n-3}, \varphi)$ on $\langle\langle \mathcal{M} \rangle\rangle$ such that $t \in \mathbb{R}$, $r \in (r_{\mathcal{H}}, +\infty)$, $\theta_i \in (0, \pi)$ ($1 \leq i \leq n-3$), $\varphi \in (0, 2\pi)$ and

$$g = - \left(1 - \frac{\mu}{r^{n-3}} + \frac{q^2}{r^{2(n-3)}} \right) dt^2 + \left(1 - \frac{\mu}{r^{n-3}} + \frac{q^2}{r^{2(n-3)}} \right)^{-1} dr^2 + r^2 \overset{\circ}{q}, \quad (5.111a)$$

$$F = -\frac{\mu_0}{\Omega_{n-2}} \frac{Q}{r^{n-2}} dt \wedge dr, \quad (5.111b)$$

where $\overset{\circ}{q}$ is the standard round metric on the unit sphere \mathbb{S}^{n-2} , which is spanned by $(\theta_1, \dots, \theta_{n-3}, \varphi)$, Q is the black hole electric charge (cf. Property 5.30) and the constants μ and q are related to the Komar mass at infinity (or ADM mass) M and to the electric charge Q by

$$\mu = \frac{16\pi M}{(n-2)\Omega_{n-2}} \quad \text{and} \quad q^2 = \frac{8\pi\mu_0 Q^2}{(n-2)(n-3)\Omega_{n-2}^2}, \quad (5.112)$$

where Ω_{n-2} is the area of \mathbb{S}^{n-2} [Eqs. (5.19)-(5.20)]. The constants μ and q must fulfill the constraint $\mu > 2q$ and the lower bound of r on $\langle\langle \mathcal{M} \rangle\rangle$ is $r_{\mathcal{H}} = (\mu/2 + \sqrt{\mu^2/4 - q^2})^{1/(n-3)}$.

For a proof, see Ref. [217] under the additional assumption that the magnetic field vanishes (i.e. that $\xi \cdot \star F = 0$) and Ref. [327] for relaxing this assumption.

Example 10: For $n = 5$, $\Omega_{n-2} = \Omega_3 = 2\pi^2$ [Eq. (5.20)] and Eqs. (5.111)-(5.112) become

$$\begin{aligned} g &= - \left(1 - \frac{\mu}{r^2} + \frac{q^2}{r^4} \right) dt^2 + \left(1 - \frac{\mu}{r^2} + \frac{q^2}{r^4} \right)^{-1} dr^2 + r^2 [\mathbf{d}\theta_1^2 + \sin^2 \theta_1 (\mathbf{d}\theta_2^2 + \sin^2 \theta_2 \mathbf{d}\varphi^2)] \\ F &= -\frac{\mu_0}{2\pi^2} \frac{Q}{r^3} dt \wedge dr \\ \mu &= \frac{8M}{3\pi} \quad \text{and} \quad q^2 = \frac{\mu_0 Q^2}{3\pi^3}. \end{aligned}$$

Remark 5: A static and asymptotically flat black hole spacetime cannot carry a nonzero magnetic monopole P for $n \geq 5$ [184]. Hence Property 5.35 is not a direct generalization of Property 5.34 to $n > 4$, even if \mathcal{H} is assumed to be made of non-degenerate components.

Historical note (electrovacuum static solutions): The Reissner-Nordström solution (5.107) of the Einstein-Maxwell system with a vanishing magnetic monopole charge ($P = 0$) has been found independently by Hans Reissner in 1916 [428], Hermann Weyl in 1917 [516] and Gunnar Nordström in

1918 [384]. The generalization to $P \neq 0$ can be found in Brandon Carter's lecture at Les Houches (1972) [94]. The generalization to dimensions $n > 4$ with $P = 0$, i.e. the solution (5.111), has been obtained by Frank Tangherlini in 1963 [472]. The Majumdar-Papapetrou solution (5.108) has been found in 1947 independently by Sudhansu Datta Majumdar [352] and Achilles Papapetrou [398], as a solution to the electrovacuum Einstein equation describing K charged point masses in gravito-electrostatic equilibrium. The extremal Reissner-Nordström solution seems to have been discussed in details first by Papapetrou in his 1947 article [398]. It was actually the starting point that lead him to the Majumdar-Papapetrou solution (cf. Remark 4 above). The black hole character of the Majumdar-Papapetrou solution has been truly recognized and analyzed by James Hartle and Stephen Hawking in 1972 [257]. The generalization of the Majumdar-Papapetrou solution to spacetime dimensions $n > 4$ is due to Robert Myers in 1987 [374].

Historical note (uniqueness theorems for static electrovacuum black holes): The Reissner-Nordström part (with $P = 0$) of the 4-dimensional uniqueness theorem 5.34 has been first proven by Werner Israel in 1968 [289], under the same additional hypotheses as the ones used in his proof of the vacuum theorem (cf. historical note on p. 170), notably that the event horizon \mathcal{H} is connected, has the topology $\mathbb{R} \times \mathbb{S}^2$ and is non-degenerate. The connectedness hypothesis has been removed by Abul Masood-ul-Alam in 1992 [357]. The allowance for a magnetic charge ($P \neq 0$) has been achieved by Markus Heusler in 1994 [273] (see also Sec. 9.4 of Ref. [274]). Three years later, Heusler obtained the Majumdar-Papapetrou part of the theorem by assuming that all the connected components of \mathcal{H} are degenerate [275]. This assumption has been relaxed by Piotr Chruściel and Paul Tod in 2007 [127], who could exclude configurations with both degenerate and non-degenerate components of \mathcal{H} . The higher-dimensional uniqueness theorem 5.35 has been established by Gary Gibbons, Daisuke Ida and Tetsuya Shiromizu in 2002 [217] with the extra assumption of a vanishing magnetic field. This assumption has been relaxed by Hari Kunduri and James Lucietti in 2018 [327].

5.6.2 Uniqueness theorems for stationary and axisymmetric black holes

In this section, we restrict ourselves to the standard spacetime dimension $n = 4$. The strong rigidity theorem discussed in Sec. 5.4 (Properties 5.25 and 5.26) basically states that any stationary black hole spacetime obeying the (electro)vacuum Einstein equation has to be axisymmetric. A crucial step towards the no-hair theorem is then the following uniqueness theorem for axisymmetric stationary black holes:

Property 5.36: Carter-Robinson theorem (Carter 1971 [92], Robinson 1975 [436])

Let (\mathcal{M}, g) be a 4-dimensional asymptotically flat spacetime containing a black hole of event horizon \mathcal{H} . If

- (\mathcal{M}, g) is stationary and axisymmetric,
- g fulfills the vacuum Einstein equation (1.44),
- \mathcal{H} is connected,

- there is no closed causal curve in the domain of outer communications $\langle\langle \mathcal{M} \rangle\rangle$,

then $\langle\langle \mathcal{M} \rangle\rangle$ is isometric to the domain of outer communications of the *Kerr spacetime*^a.

^aWe shall not define the Kerr spacetime here; this will be done in Chap. 10.

Remark 6: In their original works, Carter and Robinson assumed that \mathcal{H} is a *non-degenerate* Killing horizon (i.e. has non-zero surface gravity). However, this hypothesis can be relaxed [124] (see [122] for an extended discussion).

The proof of Property 5.36 can be found in Chap. 10 of Heusler's textbook [274] (see also Sec. 5 of Carter's review article [100]) with the additional assumption that the horizon is non-degenerate; see Ref. [124] for the degenerate case.

Remark 7: The causality condition (absence of closed causal (i.e. null or timelike) curves in the black hole exterior), which is one of the assumptions of Carter-Robinson's theorem (cf. [102] for a discussion), does not appear in Israel's theorem (Property 5.33) because a static spacetime, which by definition has hypersurface-orthogonal timelike curves, cannot contain any closed causal curve.

The electrovacuum version of the Carter-Robinson theorem is

Property 5.37: Bunting-Mazur theorem (Bunting 1983 [75], Mazur 1982 [360])

Let (\mathcal{M}, g) be a 4-dimensional asymptotically flat spacetime containing a black hole of event horizon \mathcal{H} . If

- (\mathcal{M}, g) is stationary and axisymmetric,
- g fulfills the electrovacuum Einstein equation (1.54),
- \mathcal{H} is connected,
- there is no closed causal curve in the domain of outer communications $\langle\langle \mathcal{M} \rangle\rangle$,

then $\langle\langle \mathcal{M} \rangle\rangle$ is isometric to the domain of outer communications of the *Kerr-Newman spacetime*^a.

^aThe Kerr-Newman solution extends the Reissner-Nordström solution (5.107) to the rotating case; it depends on four parameters: M, Q, P (as for Reissner-Nordström) and J , the latter being the total angular momentum. One may say as well that the Kerr-Newman solution extends the Kerr solution, which depends only on (M, J) , to the electromagnetic case. See e.g. Eq. (5.54) of Ref. [94] for the explicit form of the Kerr-Newman metric.

The proof of Property 5.37 with the additional assumption that the horizon is non-degenerate can be found in Chap. 10 of Heusler's textbook [274] as well as in Carter's articles [99, 100] and Mazur's review article [363]; the degenerate case is treated in Ref. [124].

Remark 8: Some of the Majumdar-Papapetrou solutions (5.108) are axisymmetric (those for which all the points (x_k, y_k, z_k) are aligned). However, they are excluded from the conclusion of the above theorem because they do not fulfill the hypothesis of \mathcal{H} being connected.

Historical note : The first stationary-axisymmetric uniqueness result has been obtained by Brandon Carter in 1971 [92]. Under the assumptions of a non-degenerate horizon with spherical cross-sections, Carter reduced the 4-dimensional stationary-axisymmetric vacuum Einstein equation to a 2-dimensional non-linear elliptic system of two partial differential equations, with boundary conditions depending on only two parameters: the mass M and $c := \kappa A/(4\pi)$, where κ is the surface gravity and A is the horizon area ($\kappa \neq 0$ since the horizon was assumed non-degenerate). Then Carter could show that the solutions of this system form disjoint 2-parameter families and that within a given family, the members are fully specified by the pair (M, c) . The Kerr family is such a family and is the only one containing the Schwarzschild solution, but Carter could not exclude that there exist other families. This step has been achieved by David Robinson in 1975 [436], who, via a *tour de force* [102], showed that the solution has to belong to the Kerr family. The electrovacuum case turned out to be too complicated to be dealt with the same techniques; only a result similar to Carter's one (i.e. reduction to disjoint 4-parameter families of solutions) could be achieved by Robinson in 1974 [435]. New techniques have been introduced independently by Paweł Mazur in 1982 [360] and Gary Bunting in 1983 [75], which enabled them to get the uniqueness theorem 5.37 under the non-degeneracy assumption. The case of a degenerate horizon has been dealt with only in 2010, when Piotr Chruściel and Luc Nguyen [124] showed that electrovacuum rotating black holes with a degenerate horizon necessarily belong to the Kerr-Newman family (they are the so-called *extremal* members of that family). For further details, see the historical accounts [102, 363, 438] by the actors themselves. As for the Kerr-Newman solution itself, it has been discovered by Ezra Newman and his collaborators in 1965 [379], two years after Roy Kerr found his famous solution [311]. A historical note about the discovery of the Kerr metric is to be found in Chap. 10, p. 324.

5.6.3 The 4-dimensional no-hair theorem

We are now in position to state the famous theorem:

Property 5.38: no-hair theorem

Let (\mathcal{M}, g) be a spacetime that

- is 4-dimensional,
- is analytic,
- is asymptotically flat,
- is stationary,
- fulfills the electrovacuum Einstein equation (1.54) (with the vacuum Einstein equation (1.44) considered as a subcase),
- contains a black hole with a connected event horizon that is either (i) rotating^a or (ii) non-rotating and non-degenerate,
- does not contain any closed causal curve in the domain of outer communications $\langle\langle\mathcal{M}\rangle\rangle$.

Then $\langle\langle \mathcal{M} \rangle\rangle$ is isometric to the domain of outer communications of a Kerr-Newman spacetime. The latter depends on four parameters: the total mass M , the total angular momentum J , the electric charge Q and the magnetic monopole charge P , with the following subcases:

- $J = 0$: Reissner-Nordström spacetime;
- $Q = 0, P = 0$: Kerr spacetime;
- $Q = 0, P = 0, J = 0$: Schwarzschild spacetime.

In the last two cases, the electromagnetic field vanishes and \mathbf{g} fulfills the vacuum Einstein equation (1.44).

^aLet us recall that the concepts of *rotating* and *non-rotating* horizons have been defined in Property 5.2.

Proof. Let \mathcal{H} be the black hole event horizon. In case (i) (\mathcal{H} rotating), we may invoke the strong rigidity theorem 5.25 to get that $(\mathcal{M}, \mathbf{g})$ is axisymmetric, in addition to being stationary. Then the Bunting-Mazur theorem 5.37 leads to $\langle\langle \mathcal{M} \rangle\rangle$ being isometric to the domain of outer communications of a Kerr-Newman spacetime. In case (ii) (\mathcal{H} is non-rotating and non-degenerate), we may invoke the staticity theorem 5.24 to assert that $(\mathcal{M}, \mathbf{g})$ is static. Then the generalized Israel theorem 5.34 leads to $\langle\langle \mathcal{M} \rangle\rangle$ being isometric to the domain of outer communications of a Reissner-Nordström spacetime (the Majumdar-Papapetrou case in theorem 5.34 is excluded since \mathcal{H} is assumed non-degenerate). Since the Reissner-Nordström family is a subfamily of the Kerr-Newman one, this completes the proof. \square

Remark 9: A stationary black hole spacetime with a non-rotating degenerate event horizon and not belonging to the Kerr-Newman family is not excluded by the above no-hair theorem. Such a spacetime cannot be static, otherwise the generalized Israel theorem 5.34 would imply that it belongs to the Reissner-Nordström subfamily. The existence of such a black hole is not excluded by the staticity theorem 5.24 due to the degeneracy hypothesis and is an open question.

Remark 10: The no-hair theorem does not hold for dimensions $n > 4$, even if one restricts oneself to the vacuum Einstein equation. Indeed, for $n > 4$, the Kerr solution is generalized to the Myers-Perry solution discussed in Sec. 5.2.3, but there exist other families of black holes solutions. For $n = 5$, some of these families are the black rings (cf. Sec. 5.2.3) and the black saturns (cf. Remark 9 on p. 159).

From an astrophysical perspective, the weaknesses of the no-hair theorem are the hypotheses of analyticity and of absence of any closed causal curve in the domain of outer communications. The analyticity hypothesis is inherited from the strong rigidity theorem (Property 5.25). In physics, one usually assumes that the fields are smooth, not necessarily analytic. As discussed in Sec. 5.4.2, the only successful attempt to date to get rid of the analyticity requirement regards slowly rotating black holes [11]. As for the hypothesis of non-existence of closed causal curves in $\langle\langle \mathcal{M} \rangle\rangle$, which holds if $\langle\langle \mathcal{M} \rangle\rangle$ is assumed globally hyperbolic (cf. p. 125), it would be much more satisfactory if this would be a consequence²⁰ of the theorem and not one of its premises. In the current state, we cannot exclude that there is somewhere

²⁰The Kerr-Newmann spacetime has a globally hyperbolic domain of outer communications.

in our Universe a stationary black hole that is distinct from a Kerr-Newman one and around which one can travel backward in time...

Historical note : The first hints towards the “hairlessness” of black holes arised in 1964-65 from studies by Vitaly Ginzburg [221] and by Andrei Doroshkevich, Yakov Zeldovich and Igor Novikov [171] (cf. Chap. 7 of Thorne’s book [478] for some historical details). Ginzburg showed that the gravitational collapse of a (electrically neutral) magnetized star gives birth to an unmagnetized black hole, while Doroshkevich, Zeldovich and Novikov showed that the collapse of a non-rotating body slightly departing from spherical symmetry leads to a perfectly spherical (i.e. Schwarzschild) black hole; the deformations away from spherical symmetry (the “hairs”) in the external metric are decaying to zero as the black hole forms. These authors have also shown in the same article [171] that any static quadrupolar deformation of the Schwarzschild metric leads to a curvature singularity at the event horizon and hence cannot provide a regular black hole solution to the vacuum Einstein equation. This result is quoted as a motivation for proving the uniqueness of the Schwarzschild solution by Werner Israel in the famous 1967 article [288] presenting the theorem that bears his name (Property 5.33). In the same article, Israel raised the question about the uniqueness of the Kerr solution. In 1968, Brandon Carter [89] conjectured that the Kerr-Newman family – which contains the Schwarzschild one – may represent all stationary black holes. This conjecture became known as the *Carter-Israel conjecture* [438, 265]. It has been popularized by the phrase “*a black hole has no hair*” by John Wheeler in a review article published in 1971 [442], as well as in the famous 1973 MTW textbook [370] (Box 33.1). In a 1996 interview, Wheeler [518] stated that Jacob Bekenstein (then his PhD student) coined that phrase. The conjecture made most of its way in becoming a theorem in 1972 when Stephen Hawking [259] established the first version of the strong rigidity theorem (Property 5.25). Indeed the latter basically states that all stationary black holes are necessarily axisymmetric and a year before, Brandon Carter [92] had shown that all axisymmetric stationary and 4-dimensional black holes belong either to the Kerr family or to a disconnected 2-parameter family, the latter possibility being eventually excluded by David Robinson in 1975 [436] (cf. the historical note on p. 176). The no-hair theorem has been strengthened in the following years by making explicit and/or relaxing certain hypotheses (cf. the historical notes on p. 170, 174 and 176). A critical review of the theorem as of 1994, distinguishing the folklore around it from what has been mathematically established, has been performed by Piotr Chruściel [114]. Further details about the theorem history can be found in the accounts by Carter [102] and Robinson [438].

5.6.4 To go further

See Heusler’s textbook (1996) [274] and the review articles by Chruściel, Lopes Costa & Heusler (2012) [122], Hollands & Ishibashi (2012) [279] and Ionescu & Klainerman (2015) [287]. The no-hair theorem has been extended from general relativity to scalar-tensor theories, under some assumptions (among which the scalar field obeys the same symmetries as the spacetime ones), see the article by Capuano, Santoni & Barausse (2023) [81] for details.

Part II

Schwarzschild black hole

Chapter 6

Schwarzschild black hole

Contents

6.1	Introduction	181
6.2	The Schwarzschild-(anti-)de Sitter solution	181
6.3	Radial null geodesics and Eddington-Finkelstein coordinates	187
6.4	Black hole character	199

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 Israel uniqueness theorem (Property 5.33, Sec. 5.6), which states that any static black hole in an asymptotically flat 4-dimensional vacuum spacetime ruled by general relativity must be a Schwarzschild black hole.

In this chapter, we derive the Schwarzschild metric as a solution to the Einstein equation, possibly with a non-vanishing cosmological constant Λ (Sec. 6.2); we then focus on the case $\Lambda = 0$ (the true Schwarzschild solution) and 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 the timelike and null geodesics in 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 to the Einstein equation (1.43) in a vacuum 4-dimensional spacetime (\mathcal{M}, g) with some arbitrary cosmological constant Λ . For

$n = 4$, Eq. (1.43) becomes

$$\boxed{\mathbf{R} = \Lambda \mathbf{g}}. \quad (6.1)$$

An equivalent form is obtained by setting $\mathbf{T} = 0$ in the original Einstein equation (1.40):

$$\boxed{\mathbf{R} + \left(\Lambda - \frac{1}{2} R \right) \mathbf{g} = 0}. \quad (6.2)$$

6.2.2 Static and spherically symmetric metric

Let us assume that the spacetime $(\mathcal{M}, \mathbf{g})$ is *static*, in the sense defined in Sec. 5.2.1: the translation group $(\mathbb{R}, +)$ is a isometry group of $(\mathcal{M}, \mathbf{g})$ (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 ξ 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 \mathcal{M} by a 1-parameter family of hypersurfaces $(\Sigma_t)_{t \in \mathbb{R}}$, such that ξ is normal to all Σ_t 's and t is a parameter associated to ξ :

$$\xi(t) = 1 \quad (6.3)$$

or equivalently, $\langle dt, \xi \rangle = 1$ [cf. Eq. (5.4)].

In addition to being static, we assume that $(\mathcal{M}, \mathbf{g})$ is *spherically symmetric*, i.e. that it is invariant under the action of the rotation group $\text{SO}(3)$, whose orbits are spacelike 2-spheres (cf. Sec. 3.3.1). Let \mathcal{S} be some generic 2-sphere orbit. The static Killing vector field ξ must be orthogonal to \mathcal{S} , otherwise the orthogonal projection of ξ onto \mathcal{S} would define some privileged direction on \mathcal{S} , which is incompatible with spherical symmetry. The orthogonality of ξ and \mathcal{S} implies that $\mathcal{S} \subset \Sigma_t$. Let $(x^a) = (\theta, \varphi)$ be spherical coordinates on \mathcal{S} . The (Riemannian) metric \mathbf{q} induced by \mathbf{g} on \mathcal{S} is given by

$$\mathbf{q} = r^2 (\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\varphi^2). \quad (6.4)$$

The positive coefficient r^2 in front of the standard spherical element must be constant over \mathcal{S} , by virtue of spherical symmetry. The area of \mathcal{S} is then $A = 4\pi r^2$. For this reason, r is called the *areal radius* of \mathcal{S} . Letting \mathcal{S} vary, r can be considered as a scalar field on \mathcal{M} . If $\mathrm{d}r \neq 0$, we may use it as a coordinate¹. Since $\mathcal{S} \subset \Sigma_t$, (r, θ, φ) is a coordinate system on each hypersurface Σ_t . The set (t, r, θ, φ) , where t is adapted to ξ thanks to (6.3), is then a spacetime coordinate system and, by construction, the expression of the metric tensor with respect to this system is

$$\mathbf{g} = -A(r) \mathrm{d}t^2 + B(r) \mathrm{d}r^2 + r^2 (\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\varphi^2). \quad (6.5)$$

Note that this is a special case of the general static metric element (5.5) and that Eq. (5.4) holds:

$$\xi = \partial_t. \quad (6.6)$$

¹An example of spherically symmetric spacetime with $\mathrm{d}r = 0$ somewhere is the maximally extended Schwarzschild spacetime, to be studied in Chap. 9: $\mathrm{d}r = 0$ on the so-called *bifurcation sphere* \mathcal{S} (cf. Sec. 9.3.3). Hence r cannot be used as a coordinate in the vicinity of \mathcal{S} ; one shall use Kruskal-Szekeres coordinates instead.

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 ξ 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 the Einstein equation

The Christoffel symbols of the metric (6.5) with respect to the areal coordinates are (cf. Sec. D.4.2 for the computation):

$$\begin{aligned}\Gamma^t_{tr} &= \Gamma^t_{rt} = \frac{1}{2A} \frac{dA}{dr} & \Gamma^r_{tt} &= \frac{1}{2B} \frac{dB}{dr} & \Gamma^r_{rr} &= \frac{1}{2B} \frac{dB}{dr} & \Gamma^r_{\theta\theta} &= -\frac{r}{B} \\ \Gamma^r_{\varphi\varphi} &= -\frac{r \sin^2 \theta}{B} & \Gamma^\theta_{r\theta} &= \Gamma^\theta_{\theta r} = \frac{1}{r} & \Gamma^\theta_{\varphi\varphi} &= -\sin \theta \cos \theta \\ \Gamma^\varphi_{r\varphi} &= \Gamma^\varphi_{\varphi r} = \frac{1}{r} & \Gamma^\varphi_{\theta\varphi} &= \Gamma^\varphi_{\varphi\theta} = \frac{1}{\tan \theta},\end{aligned}\tag{6.7}$$

the Christoffel symbols not listed above being zero.

The tt component of the Einstein equation (6.2) leads to (cf. Sec. D.4.2 for the computation)

$$r \frac{dB}{dr} - B + (1 - \Lambda r^2) B^2 = 0,\tag{6.8}$$

while the rr component leads to

$$r \frac{dA}{dr} + A - (1 - \Lambda r^2) AB = 0.\tag{6.9}$$

Finally, the $\theta\theta$ and $\varphi\varphi$ components lead to the same equation:

$$2 \frac{d^2 A}{dr^2} + \frac{2}{r} \frac{dA}{dr} - \frac{1}{B} \left(\frac{dA}{dr} + \frac{2A}{r} \right) \frac{dB}{dr} - \frac{1}{A} \left(\frac{dA}{dr} \right)^2 + 4\Lambda AB = 0.\tag{6.10}$$

All the other components of the Einstein equation (6.2) are identically zero.

Adding Eq. (6.8) multiplied by A to Eq. (6.9) multiplied by B yields

$$B \frac{dA}{dr} + A \frac{dB}{dr} = \frac{d}{dr}(AB) = 0.$$

The solution to 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. (6.5) results in

$$\mathbf{g} = -\frac{C}{B(r)} \mathbf{dt}^2 + B(r) \mathbf{dr}^2 + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2).$$

Assuming $C > 0$, the change of variable $t' = \sqrt{C}t$, which is equivalent to changing the stationary Killing vector from ξ to $\xi' = 1/\sqrt{C}\xi$, yields

$$\mathbf{g} = -\frac{1}{B(r)} \mathbf{dt}'^2 + B(r) \mathbf{dr}^2 + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2),$$

which is exactly the solution corresponding to $C = 1$. Hence from now on, we set $C = 1$, i.e.

$$B(r) = \frac{1}{A(r)}. \quad (6.11)$$

Substituting this expression in Eq. (6.9) yields an ordinary differential equation for $A(r)$:

$$r \frac{dA}{dr} + A - 1 + \Lambda r^2 = 0,$$

the solution to which is

$$A(r) = 1 - \frac{2m}{r} - \frac{\Lambda}{3} r^2, \quad (6.12)$$

where m is a constant. The general static and spherically symmetric solution to the vacuum Einstein equation (6.1) is therefore

$$\boxed{\mathbf{g} = - \left(1 - \frac{2m}{r} - \frac{\Lambda}{3} r^2\right) dt^2 + \left(1 - \frac{2m}{r} - \frac{\Lambda}{3} r^2\right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)}. \quad (6.13)$$

It is called the **Kottler metric** (cf. the historical note below). The **Schwarzschild metric** is the particular case $\Lambda = 0$. If $\Lambda > 0$, (6.13) 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. (6.13):

$$\boxed{\mathbf{g} = - \left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)}. \quad (6.14)$$

The areal coordinates (t, r, θ, φ) are then called the **Schwarzschild-Droste coordinates**².

Since $A(r) = 1 - 2m/r$ and $B(r) = (1 - 2m/r)^{-1}$ for the Schwarzschild metric, the non-vanishing Christoffel symbols (6.7) become³

$$\begin{aligned} \Gamma^t_{tr} = \Gamma^t_{rt} &= \frac{m}{r(r-2m)} & \Gamma^r_{tt} &= \frac{m(r-2m)}{r^3} & \Gamma^r_{rr} &= -\frac{m}{r(r-2m)} \\ \Gamma^r_{\theta\theta} &= 2m - r & \Gamma^r_{\varphi\varphi} &= (2m - r) \sin^2 \theta & \Gamma^\theta_{r\theta} = \Gamma^\theta_{\theta r} &= \frac{1}{r} \\ \Gamma^\theta_{\varphi\varphi} &= -\sin \theta \cos \theta & \Gamma^\varphi_{r\varphi} = \Gamma^\varphi_{\varphi r} &= \frac{1}{r} & \Gamma^\varphi_{\theta\varphi} = \Gamma^\varphi_{\varphi\theta} &= \frac{1}{\tan \theta}. \end{aligned} \quad (6.15)$$

²In the literature they are often referred to as simply **Schwarzschild coordinates**; we follow here Deruelle & Uzan [161, 162].

³See also the notebook D.4.3 for a check.

6.2.4 The mass parameter

The Schwarzschild metric (6.14) 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 metric (6.14) takes the standard weak-field form (see e.g. [86, 370]):

$$g \simeq - (1 + 2\Phi(r)) dt^2 + (1 - 2\Phi(r)) dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (6.16)$$

where $\Phi(r) := -m/r$ is the Newtonian gravitational potential outside a spherically symmetric body of mass m .

Another argument for identifying m with a mass is provided by Example 3 of Sec. 5.3.3, where it has been shown that m is nothing but the *Komar mass* associated with the stationarity of Schwarzschild spacetime.

Historical note : The Schwarzschild metric (6.14) is actually the first non-trivial (i.e. different from Minkowski metric) exact solution to the Einstein equation ever found. It has been obtained by the astrophysicist Karl Schwarzschild in the end of 1915 [448], 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 metric (6.14), 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 (6.14) in May 1916 [174]. 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 metric (6.14). The generalization to a non-vanishing cosmological constant, i.e. Eq. (6.13), has been obtained by Friedrich Kottler in 1918 [321] and, independently, by Hermann Weyl in 1919 [517]. We refer to Eisenstaedt's article [179] for a detailed account of the early history of the Schwarzschild solution.

6.2.5 The Schwarzschild-Droste domain

We immediately notice on (6.14) 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 \mathcal{M} ,

which we call the ***Schwarzschild-Droste domain***:

$$\mathcal{M}_{\text{SD}} := \mathcal{M}_I \cup \mathcal{M}_{\text{II}}, \quad (6.17\text{a})$$

$$\mathcal{M}_I := \mathbb{R} \times (2m, +\infty) \times \mathbb{S}^2, \quad (6.17\text{b})$$

$$\mathcal{M}_{\text{II}} := \mathbb{R} \times (0, 2m) \times \mathbb{S}^2, \quad (6.17\text{c})$$

with the coordinate t spanning \mathbb{R} , the coordinate r spanning $(2m, +\infty)$ on \mathcal{M}_I and $(0, 2m)$ on \mathcal{M}_{II} , and the coordinates (θ, φ) constituting a standard spherical chart of \mathbb{S}^2 . Note that \mathcal{M}_{SD} is a disconnected open subset of the full spacetime manifold \mathcal{M} (to be specified later), whose connected components are \mathcal{M}_I and \mathcal{M}_{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 $\theta = \pi$. It is fully understood that the metric \mathbf{g} , as expressed by (6.14), 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 $\theta = \pi$ reflects merely the coordinate singularity of the (θ, φ) chart there. We shall not discuss this coordinate singularity any further.

The boundary value $r_S := 2m$ of r between \mathcal{M}_I and \mathcal{M}_{II} is conventionally 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.

Immediate properties of the Schwarzschild-Droste domain are:

Property 6.1: the Schwarzschild exterior \mathcal{M}_I

Region $(\mathcal{M}_I, \mathbf{g})$ is strictly static, in the sense defined in Sec. 5.2.1: the Killing vector $\xi = \partial_t$ is timelike on all \mathcal{M}_I and is orthogonal to the hypersurfaces $t = \text{const}$. Moreover, $(\mathcal{M}_I, \mathbf{g})$ is asymptotically flat: the metric \mathbf{g} tends to Minkowski metric when $r \rightarrow +\infty$.

Proof. From expression (6.14), we see that $g_{tt} < 0$ for $r > 2m$. Since $g_{tt} = \mathbf{g}(\partial_t, \partial_t)$, this implies that ∂_t is timelike. Moreover ∂_t is orthogonal to any hypersurface Σ_t defined by a constant value of t , since we read on (6.14) that $\mathbf{g}(\partial_t, \partial_i) = g_{ti} = 0$ for $i \in \{1, 2, 3\}$ and by definition, the vectors $(\partial_i) = (\partial_r, \partial_\theta, \partial_\varphi)$ form a basis of the tangent planes to Σ_t . Finally, for $r \rightarrow +\infty$, all the metric components given by (6.14) tend to those of the Minkowski metric (4.3) [see also Eq. (6.16)]. \square

Property 6.2: the Schwarzschild interior \mathcal{M}_{II}

In region \mathcal{M}_{II} , the Killing vector $\xi = \partial_t$ is spacelike. It follows that $(\mathcal{M}_{\text{II}}, \mathbf{g})$ is *not* static: the translation group $(\mathbb{R}, +)$ is still an isometry group of $(\mathcal{M}_{\text{II}}, \mathbf{g})$, but its orbits are spacelike curves (cf. Sec. 5.2.1). Besides, in \mathcal{M}_{II} , t is a spacelike coordinate (i.e. the hypersurfaces $t = \text{const}$ are timelike, cf. Sec. A.3.2) and r is a timelike one (i.e. the hypersurfaces $r = \text{const}$ are spacelike).

Proof. We see from (6.14) that $g_{tt} > 0$ for $r < 2m$, which implies that ∂_t is spacelike there. Moreover, we have $g^{tt} > 0$ and $g^{rr} < 0$ in \mathcal{M}_{II} , which, according to the criterion (A.56), implies that t is a spacelike coordinate and r a timelike one. \square

Remark 2: In the region \mathcal{M}_{II} , $g_{rr} < 0$, so that the metric g keeps a Lorentzian signature, as it should!

Remark 3: As a consequence of Property 6.2 and of the diagonal character of the metric components (6.14), the axes of the light cones are horizontal lines for $r < 2m$ in Fig. 6.1 below.

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 (6.14) that are radial, i.e. along which $\theta = \text{const}$ and $\varphi = \text{const}$. They are found by setting $d\theta = 0$ and $d\varphi = 0$ in the line element (1.3) and searching for $ds^2 = g_{\mu\nu} dx^\mu dx^\nu = 0$, with the $g_{\mu\nu}$'s read on (6.14):

$$ds^2 = 0 \iff dt^2 = \frac{dr^2}{\left(1 - \frac{2m}{r}\right)^2}. \quad (6.18)$$

Hence the radial null geodesics are governed by

$$dt = \pm \frac{dr}{1 - \frac{2m}{r}}. \quad (6.19)$$

This equation is easily integrated:

$$t = \pm \left(r + 2m \ln \left| \frac{r}{2m} - 1 \right| \right) + \text{const.} \quad (6.20)$$

We have thus two families of curves, one for each choice of the sign \pm :

- the **outgoing radial null geodesics** $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$, whose equation is

$$\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}} : \boxed{t = r + 2m \ln \left| \frac{r}{2m} - 1 \right| + u}, \quad (6.21)$$

where $u \in \mathbb{R}$ is a constant along $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$;

- the **ingoing radial null geodesics** $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$, whose equation is

$$\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}} : \boxed{t = -r - 2m \ln \left| \frac{r}{2m} - 1 \right| + v}, \quad (6.22)$$

where $v \in \mathbb{R}$ is a constant along $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$.

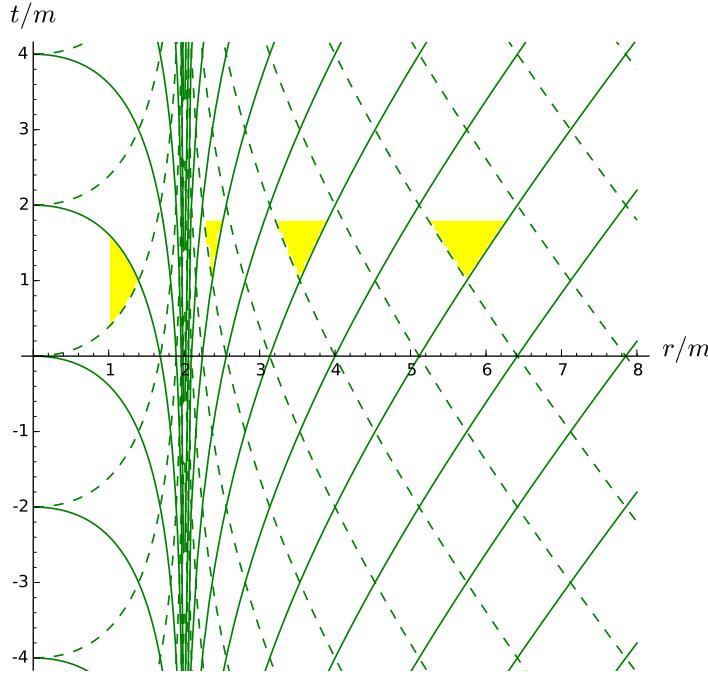


Figure 6.1: Radial null geodesics of Schwarzschild spacetime, plotted in terms of Schwarzschild-Droste coordinates (t, r) : the solid (resp. dashed) lines correspond to outgoing (resp. ingoing) geodesics $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ (resp. $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$), as given by Eq. (6.21) (resp. Eq. (6.22)). The interiors of some future light cones are depicted in yellow.

Note that along a given outgoing radial null geodesic, (u, θ, φ) are constant and that two outgoing radial null geodesics that have distinct (u, θ, φ) are distinct. We can thus use the triplet (u, θ, φ) to label the outgoing radial null geodesics, leading to the notation $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$. Moreover, the family $(\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}})$, with $u \in \mathbb{R}$, $\theta \in [0, \pi]$ and $\varphi \in [0, 2\pi]$, forms a **congruence**: there is one, and only one, such curve through every point of \mathcal{M}_{SD} . Similar considerations apply to the ingoing radial null geodesics $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$.

By introducing the **tortoise coordinate**

$$r_* := r + 2m \ln \left| \frac{r}{2m} - 1 \right|, \quad (6.23)$$

one may rewrite Eqs. (6.21)-(6.22) as respectively

$$\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}} : \quad t = r_* + u \quad (6.24)$$

$$\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}} : \quad t = -r_* + v. \quad (6.25)$$

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. (6.18). Since not all null curves are null geodesics⁴, there remains to prove that the curves defined by (6.21)

⁴A famous counterexample is the null helix in Minkowski spacetime, cf. Remark 2 on p. 35.

and (6.22) obey the geodesic equation [Eq. (B.10) in Appendix B]:

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma_{\mu\nu}^\alpha \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = 0, \quad (6.26)$$

where λ is an affine parameter (cf. Sec. B.2.1). Let us check that (6.26) is satisfied by choosing $\lambda = r$. For the curves defined by (6.21), we have

$$(x^\alpha(r)) = \left(r + 2m \ln \left| \frac{r}{2m} - 1 \right| + u, r, \theta, \varphi \right).$$

Hence

$$\left(\frac{dx^\alpha}{dr} \right) = \left(\frac{r}{r-2m}, 1, 0, 0 \right) \quad \text{and} \quad \left(\frac{d^2x^\alpha}{dr^2} \right) = \left(-\frac{2m}{(r-2m)^2}, 0, 0, 0 \right).$$

Given the Christoffel symbols (6.15), it is then a simple exercise to show that Eq. (6.26) is satisfied. The same property holds for the family (6.22). Hence we conclude

Property 6.3: radial null geodesics

The radial null geodesics in the Schwarzschild-Droste domain form two congruences, $(\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}})$ and $(\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}})$, obeying Eqs. (6.21)-(6.22). Moreover, the areal radius r is an affine parameter along them.

The two congruences 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 $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ 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 vector and we shall see in Sec. 6.3.6 that $-\partial_r$ is oriented towards the future (cf. the “tilted” light cone in Fig. 6.1).

6.3.2 Eddington-Finkelstein coordinates

The parameter v is one of the three parameters labelling the ingoing radial null geodesics $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$. Let us promote it to a spacetime coordinate, instead of t , i.e. let us consider the coordinate system (v, r, θ, φ) that is related to the Schwarzschild-Droste coordinates (t, r, θ, φ) by Eq. (6.22):

$$v = t + r + 2m \ln \left| \frac{r}{2m} - 1 \right|. \quad (6.27)$$

By differentiation, it follows immediately that

$$dt = dv - \frac{dr}{1 - 2m/r}, \quad (6.28)$$

the tensor square of which is [cf. Eq. (A.39)]

$$\mathrm{d}t^2 = \mathrm{d}v^2 - \frac{2}{1-2m/r} \mathrm{d}v \mathrm{d}r + \frac{1}{(1-2m/r)^2} \mathrm{d}r^2.$$

Substituting this expression for $\mathrm{d}t^2$ in Eq. (6.14) yields the metric components with respect to the coordinates $(x^{\hat{\alpha}}) := (v, r, \theta, \varphi)$:

$$\boxed{\mathbf{g} = -\left(1 - \frac{2m}{r}\right) \mathrm{d}v^2 + 2 \mathrm{d}v \mathrm{d}r + r^2 (\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\varphi^2).} \quad (6.29)$$

Remark 2: Since $\mathrm{d}v \mathrm{d}r := 1/2 (\mathrm{d}v \otimes \mathrm{d}r + \mathrm{d}r \otimes \mathrm{d}v)$ [cf. Eq. (A.39)], the metric component g_{vr} is *one half* of the coefficient of the $\mathrm{d}v \mathrm{d}r$ term in Eq. (6.29), i.e. $g_{vr} = 1$.

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 v is a null coordinate, i.e. the level sets $v = \text{const}$ are null hypersurfaces (cf. Sec. A.3.2). This can be seen from $g^{vv} = 0$ [cf. Eq. (A.56)].

To deal with a timelike coordinate instead of a null one, let us set

$$\boxed{\tilde{t} := v - r \iff v = \tilde{t} + r} \quad (6.30)$$

and define the **ingoing Eddington-Finkelstein (IEF) coordinates** to be

$$(x^{\hat{\alpha}}) := (\tilde{t}, r, \theta, \varphi). \quad (6.31)$$

Remark 3: From (6.30), v appears as the “time” \tilde{t} “advanced” by r , while from (6.25), 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. (6.27) and (6.30):

$$\boxed{\tilde{t} = t + 2m \ln \left| \frac{r}{2m} - 1 \right|.} \quad (6.32)$$

The hypersurfaces $t = \text{const}$ are plotted in Fig. 6.2, in terms of the IEF coordinates.

From (6.30), we have $\mathrm{d}v = \mathrm{d}\tilde{t} + \mathrm{d}r$. Substituting into (6.29) yields

$$\boxed{\mathbf{g} = -\left(1 - \frac{2m}{r}\right) \mathrm{d}\tilde{t}^2 + \frac{4m}{r} \mathrm{d}\tilde{t} \mathrm{d}r + \left(1 + \frac{2m}{r}\right) \mathrm{d}r^2 + r^2 (\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\varphi^2).} \quad (6.33)$$

To avoid any ambiguity, we shall denote by $\partial_{\tilde{r}}$ the coordinate vector of the IEF frame and by ∂_r the coordinate vector of the Schwarzschild-Droste frame:

$$\partial_{\tilde{r}} := \frac{\partial}{\partial r} \Big|_{\tilde{t}, \theta, \varphi} \quad \text{and} \quad \partial_r := \frac{\partial}{\partial r} \Big|_{t, \theta, \varphi}. \quad (6.34)$$

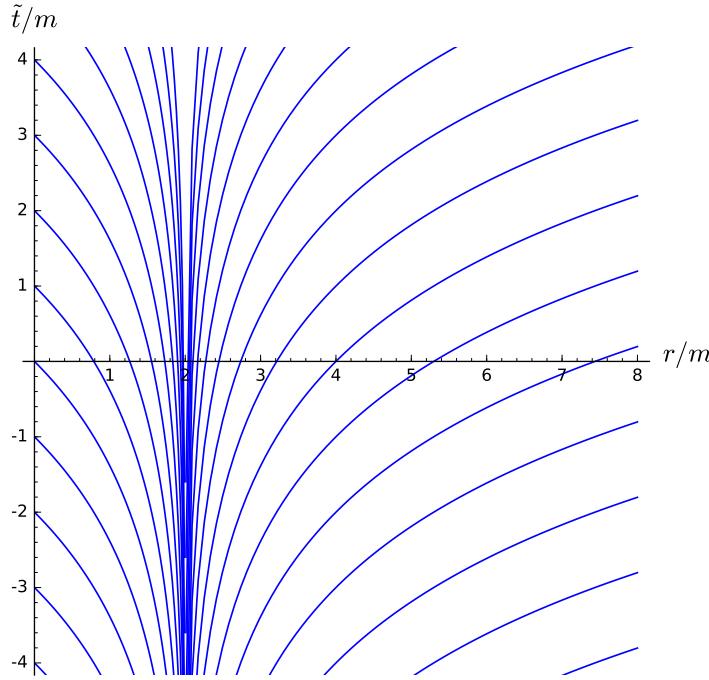


Figure 6.2: Hypersurfaces of constant Schwarzschild-Droste coordinate t , drawn in terms of the ingoing Eddington-Finkelstein coordinates (\tilde{t}, r) . Since the dimensions along θ and φ are not represented, these 3-dimensional surfaces appear as curves.

The relation between the two vectors is given by the chain rule:

$$\frac{\partial}{\partial r} \Big|_{\tilde{t}, \theta, \varphi} = \frac{\partial}{\partial t} \Big|_{r, \theta, \varphi} \underbrace{\frac{\partial t}{\partial r} \Big|_{\tilde{t}, \theta, \varphi}}_{\left(1 - \frac{r}{2m}\right)^{-1}} + \frac{\partial}{\partial r} \Big|_{t, \theta, \varphi} \underbrace{\frac{\partial r}{\partial r} \Big|_{\tilde{t}, \theta, \varphi}}_1 + \frac{\partial}{\partial \theta} \Big|_{t, r, \varphi} \underbrace{\frac{\partial \theta}{\partial r} \Big|_{\tilde{t}, \theta, \varphi}}_0 + \frac{\partial}{\partial \varphi} \Big|_{t, r, \theta} \underbrace{\frac{\partial \varphi}{\partial r} \Big|_{\tilde{t}, \theta, \varphi}}_0,$$

where (6.32) has been used to evaluate $\partial t / \partial r|_{\tilde{t}, \theta, \varphi}$. Hence

$$\partial_{\tilde{r}} = \partial_r + \left(1 - \frac{r}{2m}\right)^{-1} \partial_t. \quad (6.35)$$

On the other hand, we deduce from (6.32) that

$$\frac{\partial}{\partial \tilde{t}} \Big|_{r, \theta, \varphi} = \frac{\partial}{\partial t} \Big|_{r, \theta, \varphi},$$

which implies:

$$\partial_{\tilde{t}} = \partial_t. \quad (6.36)$$

In particular, the vector $\partial_{\tilde{t}}$ of the IEF frame coincides with the Killing vector ξ :

$$\boxed{\partial_{\tilde{t}} = \xi}. \quad (6.37)$$

Remark 4: The result (6.37) is not surprising since the metric components (6.33) are independent from \tilde{t} . This implies that $\partial_{\tilde{t}}$ is a Killing vector. \tilde{t} being a timelike coordinate in \mathcal{M}_I , it follows that $\partial_{\tilde{t}} = \alpha \xi$, where α is a constant. Since $\tilde{t} \sim t$ when $r \rightarrow +\infty$, we get $\alpha = 1$.

Remark 5: In region \mathcal{M}_{II} , the four vectors $(\partial_{\tilde{t}}, \partial_{\tilde{r}}, \partial_{\theta}, \partial_{\varphi})$ are spacelike⁵. There is nothing wrong about that; in particular this does not contradict the signature $(-, +, +, +)$ of the metric. The latter is related to the type of the basis vectors only when the metric components take a *diagonal* form, which is not the case here, since $g_{\tilde{t}\tilde{r}} \neq 0$ [Eq. (6.33)]. All that is demanded to $(\tilde{t}, r, \theta, \varphi)$ for being (locally) a regular coordinate system is that $(\partial_{\tilde{t}}, \partial_{\tilde{r}}, \partial_{\theta}, \partial_{\varphi})$ form a basis of the tangent space $T_p \mathcal{M}$ at each point p . This can be achieved with any causal type of the vectors $(\partial_{\tilde{t}}, \partial_{\tilde{r}}, \partial_{\theta}, \partial_{\varphi})$.

Remark 6: The IEF expression (6.33) for the metric can be recast in the following remarkable form:

$$g = \underbrace{-d\tilde{t}^2 + dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2)}_f + \frac{2m}{r} \underbrace{(d\tilde{t} + dr)^2}_{\underline{k} \otimes \underline{k}}, \quad (6.38)$$

where the f is the (flat) Minkowski metric (expressed above in terms of the spherical coordinates $(\tilde{t}, r, \theta, \varphi)$) and $\underline{k} = -d(\tilde{t} + r) = -dv$. The 1-form \underline{k} is dual to a vector k , which is null, as it can be seen from $g^{\tilde{u}\tilde{v}} k_{\tilde{u}} k_{\tilde{v}} = 0$. The latter property is easily deduced from $k_{\tilde{u}} = (-1, -1, 0, 0)$ and expression (6.40) for $g^{\tilde{u}\tilde{v}}$ below. A metric of the type (6.38) is said to be a **Kerr-Schild metric**; these peculiar metrics are discussed in Appendix C.

Historical note : Eddington-Finkelstein coordinates have been introduced by Arthur Eddington in 1924 [176]. 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. (6.32) by the minus sign in front of the logarithm⁶, which means that Eddington's time coordinate is actually $\tilde{t} = u + r$, instead of $\tilde{t} = v - r$ (our Eq. (6.30)). Eddington used his transformation to get the Kerr-Schild form (6.38) of Schwarzschild metric, with $(d\tilde{t} + dr)^2$ replaced by $(d\tilde{t} - dr)^2$ 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 [176] was elsewhere, in the comparison of general relativity to an alternative theory proposed in 1922 by the mathematician Alfred N. Whitehead (see e.g. [220]). Only in 1958 did David Finkelstein reintroduce the Eddington transformation and conclude that the Schwarzschild metric is analytic over the whole domain $r \in (0, +\infty)$ [194]. Meanwhile the regularity of Schwarzschild metric at $r = 2m$ had been proven by Georges Lemaître in 1932 [345], via another coordinate system, which we shall introduce in Sec. 14.2.6 (see also [181] for a detailed discussion), as well as by John Synge in 1950 [468], by means of yet another coordinate system (cf. the historical note on p. 292).

Remark 7: 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 considering the

⁵This follows from the diagonal components $g_{\alpha\alpha}$ read on (6.33) being positive for $r < 2m$, but this can also be seen graphically on Fig. 6.3 below: for $r < 2m$, both the \tilde{t} and r coordinate lines, i.e. the vertical and horizontal lines, are outside the light cones.

⁶The other differences with (6.32) are a constant additive term and a misprint in Eddington's formula: the term $\ln(r - m)$ should be replaced by $\ln(r - 2m)$.

components (6.29). However, neither Eddington [176] nor Finkelstein [194] considered this null version: they used coordinates $(\tilde{t}, r, \theta, \varphi)$ and they exhibited (the outgoing version of) the metric components (6.33). Hence our terminology is more faithful to history.

6.3.3 The Schwarzschild horizon

Contrary to the Schwarzschild-Droste components (6.14), the metric components (6.33) are regular as $r \rightarrow 2m$. Hence (6.33) defines a regular non-degenerate metric on the whole *ingoing Eddington-Finkelstein domain*

$$\mathcal{M}_{\text{IEF}} := \mathbb{R} \times (0, +\infty) \times \mathbb{S}^2, \quad (6.39)$$

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 IEF coordinates are

$$g^{\tilde{\alpha}\tilde{\beta}} = \begin{pmatrix} -\left(1 + \frac{2m}{r}\right) & \frac{2m}{r} & 0 & 0 \\ \frac{2m}{r} & 1 - \frac{2m}{r} & 0 & 0 \\ 0 & 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & 0 & \frac{1}{r^2 \sin^2 \theta} \end{pmatrix}. \quad (6.40)$$

In particular, the components $g^{\tilde{\alpha}\tilde{\beta}}$ are regular at $r = 2m$. Moreover we notice that $g^{\tilde{t}\tilde{t}} < 0$ for all $r \in (0, +\infty)$. In view of the criterion (A.56), we may assert

Property 6.4: Timelike character of the IEF coordinate \tilde{t}

The coordinate \tilde{t} is timelike in all \mathcal{M}_{IEF} .

This is in contrast with the Schwarzschild-Droste coordinate t , which is timelike in \mathcal{M}_I but spacelike in \mathcal{M}_{II} (Property 6.2).

The IEF domain is an extension of the Schwarzschild-Droste domain introduced in Sec. 6.2.5:

$$\mathcal{M}_{\text{IEF}} = \mathcal{M}_{\text{SD}} \cup \mathcal{H} = \mathcal{M}_I \cup \mathcal{M}_{II} \cup \mathcal{H}, \quad (6.41)$$

where \mathcal{H} is the subset of \mathcal{M}_{IEF} defined by $r = 2m$. Note that \mathcal{H} has the topology

$$\mathcal{H} \simeq \mathbb{R} \times \mathbb{S}^2 \quad (6.42)$$

and that $(\tilde{t}, \theta, \varphi)$ is a coordinate system on \mathcal{H} . Actually \mathcal{H} is nothing but what has been called the **Schwarzschild horizon** in the examples of Chaps. 2 and 3. Indeed, the metric (6.33) 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 (6.40) as well). We have thus the fundamental result, the proof of which is given in Example 10 of Chap. 3 (p. 68):

Property 6.5

The hypersurface \mathcal{H} defined by $r = 2m$ is a Killing horizon, the null normal of which is ξ .

In particular, \mathcal{H} is a null hypersurface, whose null geodesic generators admit $\xi = \partial_{\tilde{t}}$ 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)

$$A = 16\pi m^2. \quad (6.43)$$

\mathcal{H} is depicted in Fig. 2.7. We shall see in Sec. 6.4 that \mathcal{H} is actually a black hole event horizon in Schwarzschild spacetime.

Historical note : The first author to recognize that the hypersurface $r = 2m$ in Schwarzschild spacetime is a one-way membrane, i.e. a horizon, is David Finkelstein in 1958 [194]. Amazingly, Finkelstein stressed rather $r = 2m$ as a *white hole* boundary in a time-reversed version of $(\mathcal{M}_{\text{IEF}}, g)$. In particular, he wrote “*causal influences propagating into the “future” can cross the Schwarzschild surface only in an outward direction*” (see also his Fig. 1). The reason is that Finkelstein considered the extension of the \mathcal{M}_1 region to $r \leq 2m$ constructed with the *outgoing* coordinate $\tilde{t} = t - 2m \ln(r/(2m) - 1)$ [his Eq. (2.3) in our notations] instead of \tilde{t} as given by Eq. (6.32). However, he noticed that another extension of \mathcal{M}_1 can be obtained by time inversion, corresponding to a “*surface that is permeable inwards*”. As we shall see in Chap. 9, both black hole and white hole regions actually exist in the maximal extension of Schwarzschild spacetime.

6.3.4 Coordinate singularity vs. curvature singularity

The above considerations show that the divergence of the metric component g_{rr} in (6.14) when $r \rightarrow 2m$ reflects a pathology of Schwarzschild-Droste coordinates and not a singularity in the metric tensor g by itself: $(\mathcal{M}_{\text{IEF}}, g)$ is perfectly regular spacetime, including at the Schwarzschild radius $r = 2m$. The bad behavior of Schwarzschild-Droste coordinates is obvious in Fig. 6.2: the hypersurfaces $t = \text{const}$ 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 $r \rightarrow 0$. 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

$$K := R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}, \quad (6.44)$$

is (cf. Sec. D.4.3 for the computation)

$$K = \frac{48m^2}{r^6}. \quad (6.45)$$

Hence $K \rightarrow +\infty$ when $r \rightarrow 0$. 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

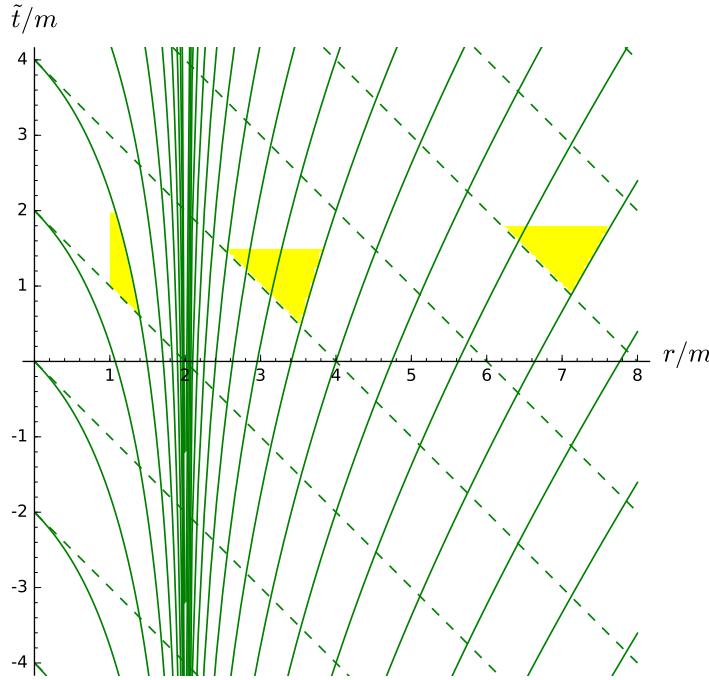


Figure 6.3: Radial null geodesics of Schwarzschild spacetime, plotted in terms of ingoing Eddington-Finkelstein coordinates (\tilde{t}, r) : the solid (resp. dashed) lines correspond to outgoing (resp. ingoing) geodesics $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ (resp. $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$), as given by Eq. (6.51) (resp. Eq. (6.47)). The interiors of some future light cones are depicted in yellow. Note that the hypersurfaces of constant \tilde{t} (horizontal lines) always lie outside the light cones, i.e. are spacelike, in agreement with \tilde{t} being everywhere a timelike coordinate (Property 6.4). Similarly the hypersurfaces of constant r (vertical lines) lie outside the light cones for $r < 2m$, in agreement with r being a timelike coordinate in \mathcal{M}_{II} (Property 6.2).

Riemann tensor per se: it is called a **curvature singularity**. Physically, this means that unbounded tidal forces are felt by any system approaching $r = 0$. This interpretation holds because tidal forces correspond to the acceleration of the separation vector between neighboring timelike geodesics and that acceleration is governed by the Riemann tensor, via the *geodesic deviation equation*, cf. Property B.22 (see Chap. 11 of MTW [370] for more details).

6.3.5 Radial null geodesics in terms of the Eddington-Finkelstein coordinates

Let us search directly for the radial null geodesics on the IEF domain \mathcal{M}_{IEF} by looking for the radial null curves of the metric (6.33). The vanishing of the line element $ds^2 = g(dx, dx)$ for the radial displacement $dx = d\tilde{t} \partial_{\tilde{t}} + dr \partial_r$ leads to

$$-\left(\frac{dr}{d\tilde{t}}\right)^2 + \frac{4m}{r+2m} \frac{dr}{d\tilde{t}} - \frac{r-2m}{r+2m} = 0.$$

The two solutions of this quadratic equation in $dr/d\tilde{t}$ are

$$\frac{dr}{d\tilde{t}} = \frac{\pm r - 2m}{r + 2m}. \quad (6.46)$$

Ingoing radial null geodesics

Choosing $-\pm$ in Eq. (6.46), we get $dr/d\tilde{t} = -1$, so that the integration is immediate and leads to the *ingoing radial null geodesics*:

$$\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}} : \quad \boxed{\tilde{t} = -r + v}, \quad (\theta, \varphi) = \text{const}, \quad (6.47)$$

where the constant $v \in \mathbb{R}$ labels the null curve, along with (θ, φ) . The simplicity of Eq. (6.47) reflects the construction of the IEF coordinates on the ingoing radial null geodesics. Note that in \mathcal{M}_{SD} , Eq. (6.47) is equivalent to Eq. (6.22), given relation (6.32) between \tilde{t} and t .

We have seen in Sec. 6.3.1 that r is an affine parameter along $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$. It follows that $\lambda = -r$ is an affine parameter as well, in terms of which the equation of $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$ deduced from Eq. (6.47) is

$$\tilde{t}(\lambda) = \lambda + v, \quad r(\lambda) = -\lambda, \quad \theta(\lambda) = \theta = \text{const}, \quad \varphi(\lambda) = \varphi = \text{const}.$$

The tangent vector \mathbf{k} associated with this parametrization of $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$ has components $k^\alpha = dx^\alpha/d\lambda = (1, -1, 0, 0)$ with respect to the IEF coordinates. We have therefore

$$\boxed{\mathbf{k} = \partial_{\tilde{t}} - \partial_r}. \quad (6.48)$$

The IEF components of the 1-form $\underline{\mathbf{k}}$ metric-dual to \mathbf{k} are $k_\alpha = g_{\alpha\mu}k^\mu$ with $g_{\alpha\mu}$ given by Eq. (6.33). We get $k_\alpha = (-1, -1, 0, 0)$, hence $\underline{\mathbf{k}} = -d\tilde{t} - dr$, i.e.

$$\underline{\mathbf{k}} = -d\mathbf{v}. \quad (6.49)$$

Outgoing radial null geodesics

For \pm equal to $+$ in Eq. (6.46), we get

$$\frac{dr}{d\tilde{t}} = \frac{r - 2m}{r + 2m}. \quad (6.50)$$

To proceed, we have to distinguish two cases. First, if $r \neq 2m$, Eq. (6.50) can be rewritten as

$$d\tilde{t} = \frac{r + 2m}{r - 2m} dr = \left(1 + \frac{4m}{r - 2m}\right) dr,$$

the integration of which leads to the *outgoing radial null geodesics*:

$$\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}} : \quad \boxed{\tilde{t} = r + 4m \ln \left| \frac{r}{2m} - 1 \right| + u}, \quad (\theta, \varphi) = \text{const}, \quad (6.51)$$

where the integration constant $u \in \mathbb{R}$ labels the null curve, along with (θ, φ) . Since $r \neq 2m$, Eq. (6.51) regards \mathcal{M}_{SD} and we actually recover Eq. (6.21), given relation (6.32) between \tilde{t} and t . One can easily show that a curve defined by (6.51) never crosses \mathcal{H} , i.e. it either lies entirely in \mathcal{M}_I or lies entirely in \mathcal{M}_{II} (cf. Fig. 6.3). We have thus two families of outgoing radial null geodesics, each labelled by (u, θ, φ) : $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out},I}$ in \mathcal{M}_I and $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out},II}$ in \mathcal{M}_{II} .

Let us now consider the case $r = 2m$ in Eq. (6.50); the equation reduces then to $dr/d\tilde{t} = 0$, so that the equation of the radial null curve is $r = \text{const}$. The constant being necessarily $2m$, we get the family of curves $\mathcal{L}_{(\theta,\varphi)}^{\text{out},\mathcal{H}}$ defined by

$$\mathcal{L}_{(\theta,\varphi)}^{\text{out},\mathcal{H}} : [r = 2m], (\theta, \varphi) = \text{const.} \quad (6.52)$$

These null curves, which form a family labelled by (θ, φ) , are actually the null geodesic generators of the Killing horizon \mathcal{H} (cf. Secs. 6.3.3 and 2.3.3). Indeed, it follows from Eq. (6.52) that a tangent vector to them is $\partial_{\tilde{t}} = \xi$, which, for $r = 2m$, is the null normal of \mathcal{H} . The null geodesics $\mathcal{L}_{(\theta,\varphi)}^{\text{out},\mathcal{H}}$ had not been found in Sec. 6.3.1 for they don't belong to \mathcal{M}_{SD} . They extend the outgoing family $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ since they obey the same differential equation (6.50) as the geodesics $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$. Moreover, they correspond to the limiting case $r = \text{const}$ separating the two subfamilies of outgoing geodesics: $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out},\text{I}}$, which have r increasing towards the future (and therefore are truly “outgoing”) and $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out},\text{II}}$, which have r decreasing towards the future (cf. Fig. 6.3).

Given that both $\mathcal{L}_{(\theta,\varphi)}^{\text{out},\mathcal{H}}$ and $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ obey the differential equation (6.50), the vector field tangent to them when parametrized by \tilde{t} is $\hat{\ell} := \partial_{\tilde{t}} + (r - 2m)/(r + 2m)\partial_{\tilde{r}}$. For future convenience, we shall actually consider the vector field $\ell := (1/2 + m/r)\hat{\ell}$; we have then by construction:

Property 6.6: outgoing radial null vector field

The vector field

$$\ell = \frac{1}{2} \left[\left(1 + \frac{2m}{r} \right) \partial_{\tilde{t}} + \left(1 - \frac{2m}{r} \right) \partial_{\tilde{r}} \right] \quad (6.53)$$

is a null regular vector field in all \mathcal{M}_{IEF} , which is tangent to the outgoing radial null geodesics $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out},\text{I}}$ in \mathcal{M}_1 , to $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out},\text{II}}$ in \mathcal{M}_{II} and to $\mathcal{L}_{(\theta,\varphi)}^{\text{out},\mathcal{H}}$ on \mathcal{H} .

Note that, on \mathcal{H} , $\ell^{\mathcal{H}} = \partial_{\tilde{t}} = \xi$.

6.3.6 Time orientation of the spacetime manifold

From now on, we consider as **Schwarzschild spacetime** (\mathcal{M}, g) the spacetime whose manifold is the largest one considered so far, i.e. the ingoing Eddington-Finkelstein domain:

$$\mathcal{M} := \mathcal{M}_{\text{IEF}} = \mathcal{M}_1 \cup \mathcal{H} \cup \mathcal{M}_{\text{II}}. \quad (6.54)$$

We have then $\mathcal{M} = \mathbb{R} \times (0, +\infty) \times \mathbb{S}^2$ [Eq. (6.39)]. Note that we shall extend this spacetime in Chap. 9.

The vector k defined by Eq. (6.48) is a nonzero null vector field defined on the whole manifold \mathcal{M} . It may therefore be used to set the time orientation of the Schwarzschild spacetime (\mathcal{M}, g) (cf. Sec. 1.2.2). Since for $r \rightarrow +\infty$, k clearly points towards increasing \tilde{t} , we declare that k defines the *future* direction:

Property 6.7: time orientation of Schwarzschild spacetime

The time orientation of the Schwarzschild spacetime $(\mathcal{M}, \mathbf{g})$ is such that the null vector \mathbf{k} defined by Eq. (6.48) is everywhere future-directed.

The above choice induces a time orientation of the subdomains \mathcal{M}_I and \mathcal{M}_{II} of \mathcal{M} . We read on the metric components (6.14) that $g_{rr} < 0$ for $r < 2m$, so that the coordinate vector ∂_r of Schwarzschild-Droste coordinates is timelike in \mathcal{M}_{II} . According to Lemma 1.2 (Sec. 1.2.2) with $\mathbf{u} = \mathbf{k}$, its time orientation is given by the scalar product $\mathbf{k} \cdot \partial_r$. Given Eqs. (6.35) and (6.36), we have

$$\partial_r = - \left(1 - \frac{2m}{r} \right)^{-1} \partial_{\tilde{t}} + \partial_{\tilde{r}}.$$

Via (6.48) and (6.33), we deduce then that

$$\mathbf{k} \cdot \partial_r = \frac{r}{2m} \left(1 - \frac{r}{2m} \right)^{-1} > 0 \quad \text{in } \mathcal{M}_{II}.$$

In view of Eq. (1.6c) in Lemma 1.2, we conclude:

Property 6.8: ∂_r past-directed timelike in the black hole region

In region \mathcal{M}_{II} , the vector ∂_r of Schwarzschild-Droste coordinates is a past-directed timelike vector.

This explains why $-\partial_r$ lies within the future null cones in Fig. 6.1 for $r < 2m$.

A corollary is:

Property 6.9: decreasing of r in the black hole region

In region \mathcal{M}_{II} , r must decrease towards the future along any null or timelike worldline.

Proof. Let \mathcal{L} be a causal curve in region \mathcal{M}_{II} and λ a parameter along \mathcal{L} increasing towards the future. The associated tangent vector $\mathbf{v} = d\mathbf{x}/d\lambda$ is then future-directed. According to Property 6.8, $-\partial_r$ is a future-directed timelike vector in \mathcal{M}_{II} , so that we can apply Lemma 1.1 (Sec. 1.2.2) with $\mathbf{u} = -\partial_r$ and get $\mathbf{g}(-\partial_r, \mathbf{v}) < 0$. Now, using the Schwarzschild-Droste components (6.14), we have

$$\mathbf{g}(-\partial_r, \mathbf{v}) = -g_{r\mu} v^\mu = -g_{rr} v^r = -g_{rr} \frac{dr}{d\lambda} = \left(\frac{2m}{r} - 1 \right)^{-1} \frac{dr}{d\lambda}.$$

Since $2m/r - 1 > 0$ in \mathcal{M}_{II} , $\mathbf{g}(-\partial_r, \mathbf{v}) < 0$ is thus equivalent to $dr/d\lambda < 0$, which proves that r is decreasing along \mathcal{L} as λ increases. \square

Thus not only an observer in \mathcal{M}_{II} cannot cross \mathcal{M}_{II} 's outer boundary \mathcal{H} to visit \mathcal{M}_I , \mathcal{H} being a null hypersurface, but he is forced to move to decreasing r until he reaches the curvature singularity at $r \rightarrow 0$. We shall study this motion in detail in Sec. 7.3.2.

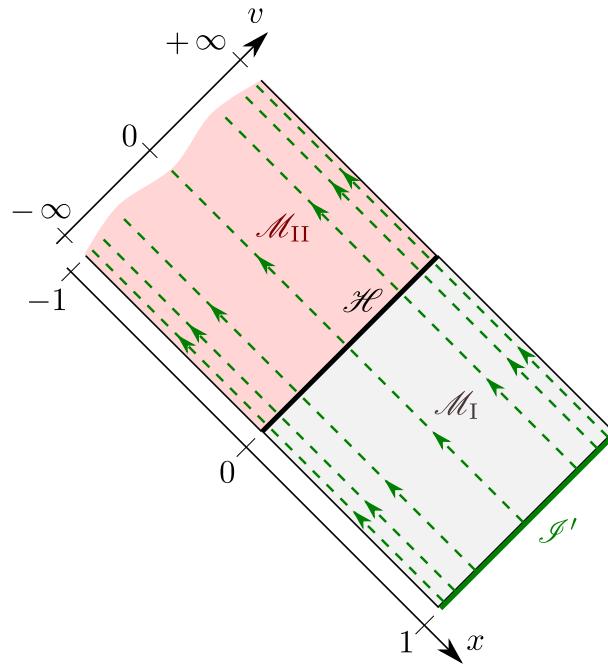


Figure 6.4: Manifold with boundary $\mathcal{M}' = \mathcal{M}_{\text{II}} \cup \mathcal{H} \cup \mathcal{M}_{\text{I}} \cup \mathcal{I}'$, drawn in terms of the coordinates x and (a compactified version of) v . The dashed lines are the ingoing radial null geodesics (as in Fig. 6.3), the arrows marking the future orientation.

6.4 Black hole character

We have already seen in Sec. 6.3.3 that \mathcal{H} is a Killing horizon. In particular, it is a null hypersurface, and thereby a one-way membrane (cf. Sec. 2.2.2). Since \mathcal{H} is the boundary of \mathcal{M}_{II} , we conclude that no particle nor electromagnetic signal may emerge from \mathcal{M}_{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 \mathcal{M}_1 , \mathcal{M}_{II} is a black hole.

It would be satisfactory though to check that \mathcal{M}_{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 $(\tilde{\mathcal{M}}, \tilde{\mathbf{g}})$ of the Schwarzschild spacetime $(\mathcal{M}, \mathbf{g})$, as defined by Eq. (6.54). 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 \mathcal{M} and the metric tensor \mathbf{g} is expressed in terms of them by Eq. (6.29). Performing the change of coordinates $(x^{\hat{\alpha}}) = (v, r, \theta, \varphi) \mapsto (x^{\alpha'}) = (v, x, \theta, \varphi)$ with

$$x = 1 - \frac{2m}{r} \iff r = \frac{2m}{1-x}, \quad x \in (-\infty, 1), \quad (6.55)$$

we deduce from (6.29) that

$$\mathbf{g} = -x \, dv^2 + \frac{4m}{(1-x)^2} \, dv \, dx + \frac{4m^2}{(1-x)^2} (\, d\theta^2 + \sin^2 \theta \, d\varphi^2 \,). \quad (6.56)$$

Defining

$$\Omega := 1 - x = \frac{2m}{r}, \quad (6.57)$$

we may rewrite the metric tensor as

$$\mathbf{g} = \Omega^{-2} \tilde{\mathbf{g}}, \quad (6.58)$$

with

$$\tilde{\mathbf{g}} = -x(1-x)^2 \mathbf{d}v^2 + 4m \mathbf{d}v \mathbf{d}x + 4m^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2). \quad (6.59)$$

Since (v, x, θ, φ) is a global coordinate system on \mathcal{M} (up to the trivial coordinate singularities of (θ, φ)), we can identify \mathcal{M} with the following open subset of $\mathbb{R}^2 \times \mathbb{S}^2$:

$$\mathcal{M} = \mathbb{R} \times (-\infty, 1) \times \mathbb{S}^2, \quad (6.60)$$

with v spanning \mathbb{R} , x spanning $(-\infty, 1)$ and (θ, φ) spanning \mathbb{S}^2 . We can then extend \mathcal{M} to the manifold with boundary⁷

$$\mathcal{M}' := \mathbb{R} \times (-\infty, 1] \times \mathbb{S}^2. \quad (6.61)$$

Notice the change $(-\infty, 1) \rightarrow (-\infty, 1]$ with respect to (6.60), which means that $x = 1$ is an allowed value on \mathcal{M}' ; it actually defines the boundary of \mathcal{M}' , \mathcal{I}' say. According to (6.55), \mathcal{I}' corresponds to $r \rightarrow +\infty$. A view of the manifold \mathcal{M}' is provided in Fig. 6.4. We note that the conformal metric (6.59) can be extended to the boundary \mathcal{I}' , yielding a regular metric. Indeed, the determinant of the metric components (6.59) is

$$\det(\tilde{g}_{\alpha'\beta'}) = -64m^6 \sin^2 \theta,$$

which does not vanish at $x = 1$ (except at the trivial coordinate singularity $\theta = 0$ or $\theta = \pi$), showing that $\tilde{\mathbf{g}}$ is a non-degenerate symmetric bilinear form at \mathcal{I}' and hence a well-defined metric on all \mathcal{M}' . Furthermore we have $\Omega > 0$ on \mathcal{M} and $\Omega = 0$ at \mathcal{I}' [set $x = 1$ in Eq. (6.57)], as well as

$$\mathbf{d}\Omega = -\mathbf{d}x \neq 0. \quad (6.62)$$

Hence $(\mathcal{M}', \tilde{\mathbf{g}})$ obeys all the conditions listed in Sec. 4.3 to be a *conformal completion at infinity* of $(\mathcal{M}, \mathbf{g})$. However, it is not adapted to the black hole definition given in Sec. 4.4.2. Indeed \mathcal{I}' does not include any future infinity (\mathcal{I}^+). Actually, \mathcal{I}' is entirely a past infinity: a generic point of \mathcal{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 $(v, \theta, \varphi) = (v_0, \theta_0, \varphi_0)$. Therefore, we shall extend \mathcal{M}' to include some \mathcal{I}^+ part. To achieve this, we shall construct \mathcal{I}^+ as the set of endpoints of the *outgoing* radial null geodesics in \mathcal{M}_1 . In terms of the null ingoing Eddington-Finkelstein coordinates (v, r, θ, φ) , the equation of these geodesics is obtained by combining (6.51) and (6.47):

$$v = 2r + 4m \ln \left| \frac{r}{2m} - 1 \right| + u, \quad (6.63)$$

where $u \in \mathbb{R}$ is a constant parameter along a given geodesic. We notice that on \mathcal{M}_1 , we may use $(x^\alpha) = (u, r, \theta, \varphi)$ as a coordinate system, naturally called the **null outgoing Eddington-Finkelstein coordinates**. Since (6.63) implies

$$\mathbf{d}v = \mathbf{d}u + \frac{2}{1 - 2m/r} \mathbf{d}r,$$

⁷Cf. Sec. A.2.2 for the definition of a *manifold with boundary*.

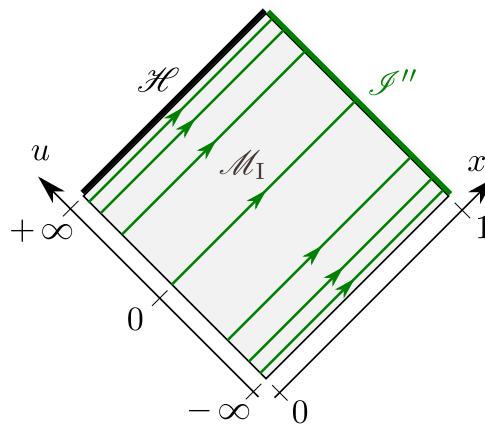


Figure 6.5: Manifold with boundary $\mathcal{M}_I'' = \mathcal{M}_I \cup \mathcal{I}''$, drawn in terms of the coordinates x and (a compactified version of) u . The green solid lines are the outgoing radial null geodesics (as in Fig. 6.3), the arrows marking the future orientation. Note that \mathcal{H} , which is drawn on this figure, is not part of \mathcal{M}_I'' .

we easily deduce from (6.29) the metric components in these coordinates:

$$g = - \left(1 - \frac{2m}{r} \right) du^2 - 2 du dr + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (6.64)$$

Remark 1: Contrary to (v, r, θ, φ) , the coordinates (u, r, θ, φ) do not cover all $\mathcal{M} = \mathcal{M}_{\text{IEF}}$, but only \mathcal{M}_I . This is graphically evident from Fig. 6.3, where the outgoing radial null geodesics $\mathcal{L}_{(u, \theta, \varphi)}^{\text{out}, I}$ accumulate on \mathcal{H} as $u \rightarrow +\infty$ from the \mathcal{M}_I side.

On \mathcal{M}_I , let us perform the change of coordinates $(x^\alpha) = (u, r, \theta, \varphi) \rightarrow (x^{\alpha''}) = (u, x, \theta, \varphi)$, where x is related to r by the same formula as (6.55), except that on \mathcal{M}_I , x 's range is $(0, 1)$ only. We deduce from (6.64) and (6.55) the expression of g in terms of the coordinates (u, x, θ, φ) :

$$g = -x du^2 - \frac{4m}{(1-x)^2} du dx + \frac{4m^2}{(1-x)^2} (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (6.65)$$

Let us identify \mathcal{M}_I with the following open subset of $\mathbb{R}^2 \times \mathbb{S}^2$:

$$\mathcal{M}_I = \mathbb{R} \times (0, 1) \times \mathbb{S}^2, \quad (6.66)$$

with u spanning \mathbb{R} , x spanning $(0, 1)$ and (θ, φ) spanning \mathbb{S}^2 . Similarly to what we did above for \mathcal{M} , we may then extend \mathcal{M}_I to the manifold with boundary

$$\mathcal{M}_I' := \mathbb{R} \times (0, 1] \times \mathbb{S}^2. \quad (6.67)$$

The boundary of \mathcal{M}_I' , \mathcal{I}'' say, lies at $x = 1$ (cf. Fig. 6.5). It shall not be confused with the boundary of \mathcal{M}_I as a submanifold of \mathcal{M}' , which is \mathcal{I}' . The difference arises from the fact that u diverges (to $-\infty$) when one approaches \mathcal{I}' in \mathcal{M}' , so that u cannot be used as a coordinate on \mathcal{M}' . This is clear on the relation (6.63) between u , v and r , which, once re-expressed in terms of x , becomes

$$u = v - 4m \left[\frac{1}{1-x} + \ln \left(\frac{x}{1-x} \right) \right]. \quad (6.68)$$

For a fixed value of v in \mathcal{M}' , this relation yields indeed diverging values of u at two places:

- $x \rightarrow 0^+$ (the horizon \mathcal{H}): $u \rightarrow +\infty$;
- $x \rightarrow 1^-$ (the boundary \mathcal{I}'): $u \rightarrow -\infty$.

Reciprocally, for a fixed value of u , relation (6.68) implies that v diverges (to $+\infty$) when $x \rightarrow 1^-$, which shows that \mathcal{I}'' is not included in \mathcal{M}' .

The conformal metric \tilde{g} on \mathcal{M}_I'' is given by

$$\tilde{g} = -x(1-x)^2 du^2 - 4m du dx + 4m^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (6.69)$$

We notice that it is regular and non-degenerate in all \mathcal{M}_I'' , including on \mathcal{I}'' ($x = 1$), and that on the submanifold \mathcal{M}_I , it is related to the physical metric g by $\tilde{g} = \Omega^2 g$, with the scalar field Ω taking the same expression in terms of x as that introduced in Eq. (6.57): $\Omega = 1 - x$.

The conformal completion of (\mathcal{M}, g) including both \mathcal{I}' (as \mathcal{I}^-) and \mathcal{I}'' (as \mathcal{I}^+) is constructed as follows. Let

$$\tilde{\mathcal{M}} = \mathcal{M}' \cup \mathcal{M}_I''. \quad (6.70)$$

We endow $\tilde{\mathcal{M}}$ with two coordinate charts:

$$\begin{aligned} \Phi_1 : \mathcal{M}' &\longrightarrow \mathbb{R} \times (-\infty, 1] \times \mathbb{S}^2 \\ p &\longmapsto (v, x, \theta, \varphi) \end{aligned} \quad \text{and} \quad \begin{aligned} \Phi_2 : \mathcal{M}_I'' &\longrightarrow \mathbb{R} \times (0, 1] \times \mathbb{S}^2 \\ p &\longmapsto (u, x, \theta, \varphi) \end{aligned} \quad (6.71)$$

and define the intersection of the two chart codomains:

$$\mathcal{M}' \cap \mathcal{M}_I'' = \{p \in \mathcal{M}', x(p) \in (0, 1)\} = \{p \in \mathcal{M}_I'', x(p) \in (0, 1)\}, \quad (6.72)$$

along with the transition map implementing (6.68):

$$\begin{aligned} \Phi_2 \circ \Phi_1^{-1} : \mathbb{R} \times (0, 1) \times \mathbb{S}^2 &\longrightarrow \mathbb{R} \times (0, 1) \times \mathbb{S}^2 \\ (v, x, \theta, \varphi) &\longmapsto (u = v - 4m \left[\frac{1}{1-x} + \ln \left(\frac{x}{1-x} \right) \right], x, \theta, \varphi), \end{aligned} \quad (6.73)$$

The above construction makes $\tilde{\mathcal{M}}$ a manifold with boundary (cf. Fig. 6.6), the boundary being

$$\mathcal{I} = \mathcal{I}^+ \cup \mathcal{I}^-, \quad (6.74)$$

with

$$\mathcal{I}^+ := \{p \in \mathcal{M}_I'', x(p) = 1\} \quad \text{and} \quad \mathcal{I}^- := \{p \in \mathcal{M}', x(p) = 1\}. \quad (6.75)$$

We then endow $\tilde{\mathcal{M}}$ with a Lorentzian metric \tilde{g} , whose expression is given by (6.59) on \mathcal{M}' and by (6.69) on \mathcal{M}_I'' . By construction, $(\tilde{\mathcal{M}}, \tilde{g})$ is then a conformal completion at null infinity of the Schwarzschild spacetime (\mathcal{M}, g) , the conformal factor Ω being given by (6.57) in both charts (\mathcal{M}', Φ_1) and $(\mathcal{M}_I'', \Phi_2)$: $\Omega = 1 - x$. In particular, it is clear that no past-directed causal curve originating in \mathcal{M} intersects \mathcal{I}^+ and that no future-directed causal curve originating in \mathcal{M} intersects \mathcal{I}^- . We also check immediately that \mathcal{I}^+ and \mathcal{I}^- are null hypersurfaces with respect to the metric \tilde{g} : both hypersurfaces are defined by $x = 1$, so that the induced metric on them, as deduced from (6.59) and (6.69), is

$$\tilde{g}|_{\mathcal{I}^\pm} = 4m^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (6.76)$$

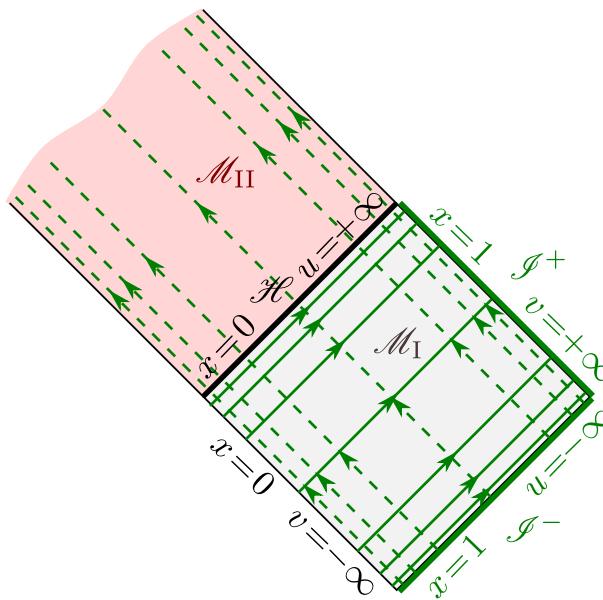


Figure 6.6: Schematic view of the manifold with boundary $\tilde{\mathcal{M}}$, which defines a conformal completion at null infinity of Schwarzschild spacetime (\mathcal{M}, g) . NB: contrary to Figs. 6.4 and 6.5, this figure is not drawn on some specific coordinate system. As in Figs. 6.3, 6.4 and 6.5, the green solid (resp. dashed) lines are the outgoing (resp. ingoing) radial null geodesics, the arrows marking the future orientation.

which is clearly degenerate (along the u direction for \mathcal{I}^+ and along the v direction for \mathcal{I}^-).

As it is clear from Fig. 6.6, \mathcal{M}_I is the interior of the causal past of \mathcal{I}^+ within \mathcal{M} :

$$\mathcal{M}_I = \text{int}(J^-(\mathcal{I}^+) \cap \mathcal{M}). \quad (6.77)$$

In view of the formal definition (4.37), we conclude:

Property 6.10: black hole region in Schwarzschild spacetime

The Schwarzschild spacetime $(\mathcal{M} = \mathcal{M}_{\text{IEF}}, g)$ has a black hole region \mathcal{B} , the interior of which is \mathcal{M}_{II} ; the event horizon is nothing but the Schwarzschild horizon \mathcal{H} discussed in Sec. 6.3.3.

Remark 2: As stated at the beginning of this section, the null character of the boundary \mathcal{H} between \mathcal{M}_I and \mathcal{M}_{II} and the fact that \mathcal{M}_{II} never intersect the asymptotically flat region $r \rightarrow +\infty$, was sufficient to claim that \mathcal{M}_{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 $(\tilde{\mathcal{M}}, \tilde{g})$ involves two coordinate charts, (v, x, θ, φ) and (u, x, θ, φ) , with two different domains, \mathcal{M}' and \mathcal{M}''_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	205
7.2	Geodesic motion	205
7.3	Timelike geodesics	212

7.1 Introduction

We have already investigated some geodesics of Schwarzschild spacetime in Chap. 6, namely the radial null geodesics (Secs. 6.3.1 and 6.3.5). 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. 7.3. 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 \mathcal{L} be a geodesic¹ of Schwarzschild spacetime (\mathcal{M}, g) . We shall assume that \mathcal{L} is causal, i.e. either timelike or null². It therefore can be considered as the worldline of some particle \mathcal{P} ,

¹The definition and basic properties of geodesics are recalled in Appendix B; see also Sec. 1.3.2.

²As shown in Sec. B.2.1, a geodesic cannot be partly timelike and partly null.

either massive (\mathcal{L} timelike) or massless (\mathcal{L} null). As recalled in Sec. 1.3, the worldline of a particle \mathcal{P} is a geodesic if, and only if, \mathcal{P} is submitted only to gravitation, i.e. \mathcal{P} is in free fall.

7.2.1 First integrals of motion

The Schwarzschild spacetime (\mathcal{M}, g) is static and spherically symmetric; the Killing vector ξ associated with the staticity (cf. Sec. 6.2.2) and the Killing vector η associated with the rotation symmetry along some axis, give birth to two conserved quantities along \mathcal{L} :

Property 7.1: conserved quantities along causal geodesics

Denoting by \mathbf{p} the 4-momentum of particle \mathcal{P} that follows the geodesic \mathcal{L} (cf. Sec. 1.3), the scalar products

$$E := -\xi \cdot \mathbf{p} = -g(\xi, \mathbf{p}) \quad (7.1a)$$

$$L := \eta \cdot \mathbf{p} = g(\eta, \mathbf{p}), \quad (7.1b)$$

are constant along \mathcal{L} . The scalar E is called \mathcal{P} 's **conserved energy** or **energy at infinity**, while L is called \mathcal{P} 's **conserved angular momentum** or **angular momentum at infinity**.

Proof. The 4-momentum \mathbf{p} is a tangent vector associated with an affine parameter of \mathcal{L} , i.e. it obeys the geodesic equation (1.11). The constancy of E and L follow then from Property B.20 (Sec. B.5). \square

In coordinates (t, r, θ, φ) adapted to the spacetime symmetries, i.e. coordinates such that $\xi = \partial_t$ and $\eta = \partial_\varphi$, 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 $\underline{\mathbf{p}}$ associated to \mathbf{p} by metric duality:

$$E = -p_t \quad (7.2a)$$

$$L = p_\varphi \quad (7.2b)$$

Indeed, in such a coordinate system, $\xi^\mu = \delta^\mu_t$ and $\eta^\mu = \delta^\mu_\varphi$, so that $E = -g_{\mu\nu} \xi^\mu p^\nu = -g_{tt} p^t = -p_t$ and $L = g_{\mu\nu} \eta^\mu p^\nu = g_{\varphi\mu} p^\mu = p_\varphi$.

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. (1.23), which resembles Eq. (7.1a) but differs from it by ξ not being a unit vector in general: $\xi \cdot \xi \neq -1$. In other words, ξ 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 $\xi \cdot \xi = g_{tt} \rightarrow -1$, that ξ is eligible as a 4-velocity, hence the name *energy at infinity*. Note that this name is commonly used, even in the particle \mathcal{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

$$L \simeq g_{\varphi\varphi} p^\varphi \simeq r^2 \sin^2 \theta p^\varphi \simeq r^2 \sin^2 \theta P^\varphi \simeq r \sin \theta P^{(\varphi)}, \quad (7.3)$$

where $P^{(\varphi)}$ is the azimuthal component of the momentum \mathbf{P} of particle \mathcal{P} as measured by an asymptotic inertial observer \mathcal{O} (cf. Sec. 1.4), i.e. the component of \mathbf{P} along $e_{(\varphi)}$ in the orthonormal basis $(e_{(r)}, e_{(\theta)}, e_{(\varphi)})$, with $e_{(\varphi)} = (r \sin \theta)^{-1} \partial_\varphi$. In view of (7.3), we may say that L is the angular momentum about the symmetry axis $\theta = 0$ that \mathcal{O} would attribute to particle \mathcal{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_{tot}^z of the total angular momentum of \mathcal{P} as measured by \mathcal{O} :

$$\mathbf{L}_{\text{tot}} := \mathbf{r} \times \mathbf{P} = L_{\text{tot}}^x \partial_x + L_{\text{tot}}^y \partial_y + L \partial_z. \quad (7.4)$$

Property 7.2: positivity of the conserved energy

The conserved energy E is a positive quantity as soon as the geodesic \mathcal{L} has some part in \mathcal{M}_I , i.e. some part with $r > 2m$:

$$\mathcal{L} \cap \mathcal{M}_I \neq \emptyset \implies E > 0. \quad (7.5)$$

Proof. In \mathcal{M}_I , the Killing vector ξ is timelike and future-directed. The 4-momentum \mathbf{p} is either timelike or null and always future-directed. By Eq. (1.5a) in Lemma 1.1 (Sec. 1.2.2), one has then necessarily $\xi \cdot \mathbf{p} < 0$; hence Eq. (7.1a) implies $E > 0$ in \mathcal{M}_I . Since E is constant along \mathcal{L} , it follows that $E > 0$ everywhere. \square

Remark 1: If the geodesic \mathcal{L} is confined to \mathcal{M}_{II} , i.e. to the black hole region (cf. Sec. 6.4), where ξ is spacelike (cf. Sec. 6.2.5), it is possible to have $E \leq 0$, since the scalar product of \mathbf{p} with a spacelike vector can take any value.

Remark 2: The Killing vector η being always spacelike, the scalar product $\mathbf{g}(\eta, \mathbf{p})$ 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 \mathcal{P} is located in the equatorial plane $\theta = \pi/2$ and the spatial projection of the worldline \mathcal{L} lies in that plane, i.e. \mathbf{p} has no component along ∂_θ :

$$\mathbf{p} \stackrel{t=0}{=} p^t \partial_t + p^r \partial_r + p^\varphi \partial_\varphi. \quad (7.6)$$

Now, for $t > 0$, if the geodesic \mathcal{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. Hence³ \mathcal{L} must stay at $\theta = \pi/2$, which implies

$$\boxed{p^\theta = 0}. \quad (7.7)$$

We conclude:

Property 7.3: planar character of geodesics

A geodesic \mathcal{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 \mathcal{L} as worldline vanishes identically [Eq. (7.7)].

Let us denote by μ the mass of particle \mathcal{P} , with possibly $\mu = 0$ if \mathcal{P} is a photon. The scalar square of the 4-momentum \mathbf{p} is then [cf. Eq. (1.10)]

$$\mathbf{g}(\mathbf{p}, \mathbf{p}) = -\mu^2. \quad (7.8)$$

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 \mathcal{L} associated with the 4-momentum \mathbf{p} [cf. Eq. (B.2)]:

$$\boxed{\mathbf{p} = \frac{d\mathbf{x}}{d\lambda}}, \quad (7.9)$$

where $d\mathbf{x}$ is the infinitesimal displacement along \mathcal{L} corresponding to the parameter change $d\lambda$. Note that λ is dimensionless and necessarily increases towards the future⁴, since \mathbf{p} is by definition future-directed (cf. Sec. 1.3.1). In terms of the components with respect to Schwarzschild-Droste coordinates, this yields

$$\dot{t} := \frac{dt}{d\lambda} = p^t, \quad \dot{r} := \frac{dr}{d\lambda} = p^r, \quad \dot{\theta} := \frac{d\theta}{d\lambda} = p^\theta, \quad \dot{\varphi} := \frac{d\varphi}{d\lambda} = p^\varphi. \quad (7.10)$$

³More rigorously, Eq. (7.7) can be derived from the geodesic equation (1.11): given the expression of the Christoffel symbols of g in Schwarzschild-Droste coordinates (cf. Sec. D.4.2), Eq. (1.11) yields

$$\frac{dp^\theta}{d\lambda} + \frac{2}{r} p^r p^\theta - \sin \theta \cos \theta (p^\varphi)^2 = 0,$$

where λ is the affine parameter of \mathcal{L} associated with \mathbf{p} , so that $p^\theta = d\theta/d\lambda$. Whatever the values of $r(\lambda)$, $p^r(\lambda)$ and $p^\varphi(\lambda)$, the solution to this ordinary differential equation with the initial conditions $p^\theta = 0$ and $\cos \theta = 0$ is $p^\theta = 0$ for all values of λ .

⁴Let us recall that Schwarzschild spacetime is time-oriented, cf. Sec. 6.3.6.

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 (6.14) 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

$$\boxed{\frac{dt}{d\lambda} = E \left(1 - \frac{2m}{r}\right)^{-1}}. \quad (7.11)$$

Similarly, Eq. (7.1b) becomes

$$L = g_{\varphi\mu}p^\mu = g_{\varphi\varphi}p^\varphi = g_{\varphi\varphi}\dot{\varphi} = r^2 \sin^2 \theta \dot{\varphi}.$$

Since $\theta = \pi/2$, we get

$$\boxed{\frac{d\varphi}{d\lambda} = \frac{L}{r^2}}. \quad (7.12)$$

We have already noticed that the sign of L is unconstrained (Remark 2 on p. 207). The above equation shows that it corresponds to the increase ($L > 0$) or decrease ($L < 0$) of φ along the geodesic \mathcal{L} . In other words, we deduce from Eq. (7.12) that

Property 7.4: monotonic behavior of φ along geodesics

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

$$\boxed{\left(\frac{dr}{d\lambda}\right)^2 - \frac{2\mu^2 m}{r} + \frac{L^2}{r^2} \left(1 - \frac{2m}{r}\right) = E^2 - \mu^2}. \quad (7.13)$$

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 \mathbf{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 \mathcal{M}_I , as we have seen above, this is guaranteed by choosing $E > 0$ [cf. Eq. (7.5)]. In \mathcal{M}_{II} , a future-directed timelike vector is $-\partial_r$ (cf. Sec. 6.3.6). According to Eq. (1.5a) in Lemma 1.1 (Sec. 1.2.2), we have then \mathbf{p} future-directed iff $-\partial_r \cdot \mathbf{p} < 0$, i.e. iff

$$\left(\frac{2m}{r} - 1 \right)^{-1} p^r < 0.$$

Since $2m/r - 1 > 0$ in \mathcal{M}_{II} , this is equivalent to $p^r < 0$, i.e. to $dr/d\lambda < 0$. Hence

Property 7.5: decreasing of r in the black hole region

In the black hole region \mathcal{M}_{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 Property 6.9, which has been derived 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 \mathcal{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 \mathcal{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 \mathcal{L} for $L > 0$ (resp. $L < 0$). Consequently,

Property 7.6: φ as a parameter along geodesics

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 \mathcal{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 $dr/d\lambda = dr/d\varphi \times d\varphi/d\lambda = L/r^2 dr/d\varphi$ and substitute this expression into the equation of motion (7.13). We get

$$\frac{1}{r^4} \left(\frac{dr}{d\varphi} \right)^2 = \frac{2m}{r^3} - \frac{1}{r^2} + 2 \left(\frac{\mu}{L} \right)^2 \frac{m}{r} + \left(\frac{E}{L} \right)^2 - \left(\frac{\mu}{L} \right)^2. \quad (7.15)$$

To simplify this equation, it is natural to introduce the dimensionless variable

$$u := \frac{m}{r} \quad (7.16)$$

instead of r . We then get

$$\left(\frac{du}{d\varphi} \right)^2 = 2u^3 - u^2 + 2 \left(\frac{m\mu}{L} \right)^2 u + \left(\frac{mE}{L} \right)^2 - \left(\frac{m\mu}{L} \right)^2. \quad (7.17)$$

This differential equation determines entirely the (r, φ) -part of the geodesic \mathcal{L} , which we shall call the ***trajectory of \mathcal{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 ($u = \text{const}$, 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 to the differential equation

$$(\wp'(z))^2 = 4\wp(z)^3 - g_2\wp(z) - g_3, \quad (7.18)$$

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

$$\left(\frac{dv}{d\tilde{\varphi}} \right)^2 = 4v^3 - \left[\frac{1}{3} - 4 \left(\frac{m\mu}{L} \right)^2 \right] v + 2 \left(\frac{mE}{L} \right)^2 - \frac{4}{3} \left(\frac{m\mu}{L} \right)^2 - \frac{1}{27}, \quad (7.19)$$

which is obviously of type (7.18) (no square in the cubic polynomial). The solution is thus

$$u = \frac{m}{r} = \wp \left(\frac{\varphi}{\sqrt{2}} + C; \omega_1, \omega_2 \right) + \frac{1}{6}, \quad (7.20)$$

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. [219] for more details regarding this method applied to null geodesics).

⁵The character \wp is a kind of calligraphic lowercase p, which is standard to denote this function.

- The second approach consists in noticing that the method of separation of variables can easily be applied to Eq. (7.17), leading to

$$\varphi = \pm \int_{u_0}^u \frac{d\bar{u}}{\sqrt{2\bar{u}^3 - \bar{u}^2 + 2\left(\frac{m\mu}{L}\right)^2 \bar{u} + \left(\frac{mE}{L}\right)^2 - \left(\frac{m\mu}{L}\right)^2}} + \varphi_0, \quad (7.21)$$

where u_0 and φ_0 are two constants, and the \pm sign can be $+$ on some parts of the geodesic \mathcal{L} and $-$ on some other parts of \mathcal{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*.

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):

$$\frac{dt}{d\varphi} = \frac{E}{L} \frac{r(\varphi)^3}{r(\varphi) - 2m}. \quad (7.22)$$

This is an ordinary differential equation for $t = t(\varphi)$, the solution to 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. 7.3) and for null geodesics (Chap. 8).

7.3 Timelike geodesics

7.3.1 Effective potential

When the geodesic \mathcal{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 p . Since the tangent vector associated with τ is the 4-velocity u (cf. Sec. 1.3.3) and p and u are related by Eq. (1.18): $p = \mu u$, we get $dx/d\lambda = \mu dx/d\tau$, from which we infer the relation between τ and λ :

$$\tau = \mu \lambda, \quad (7.23)$$

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

$$\frac{1}{2} \left(\frac{dr}{d\tau} \right)^2 + V_\ell(r) = \frac{\varepsilon^2 - 1}{2},$$

(7.24)

where

$$V_\ell(r) := -\frac{m}{r} + \frac{\ell^2}{2r^2} \left(1 - \frac{2m}{r} \right)$$

(7.25)

and ε and ℓ are respectively the *specific conserved energy* and *specific conserved angular momentum* of particle \mathcal{P} :

$$\boxed{\varepsilon := \frac{E}{\mu} = -\xi \cdot u} \quad \text{and} \quad \boxed{\ell := \frac{L}{\mu} = \eta \cdot u}, \quad (7.26)$$

where u is the 4-velocity of \mathcal{P} and the second equalities result from definitions (7.1) and the relation $p = \mu u$ [Eq. (1.18)]. Note that ε is dimensionless (in units $c = 1$) and that it shares the same positiveness property (7.5) as E :

Property 7.7: positivity of the specific conserved energy

The specific conserved energy ε is positive as soon as the timelike geodesic \mathcal{L} has some part in \mathcal{M}_I , i.e. some part with $r > 2m$:

$$\mathcal{L} \cap \mathcal{M}_I \neq \emptyset \implies \varepsilon > 0. \quad (7.27)$$

On the contrary, ℓ can be either positive, zero or negative, depending on the variation of φ along \mathcal{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 $1/2 (dr/d\tau)^2$ 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 \mathcal{M}_I , where the Killing vector ξ is timelike, we may introduce the *static observer* \mathcal{O} , whose 4-velocity $u_{\mathcal{O}}$ is collinear to ξ :

$$u_{\mathcal{O}} = \left(1 - \frac{2m}{r}\right)^{-1/2} \xi, \quad (7.28)$$

the proportionality coefficient ensuring that $u_{\mathcal{O}} \cdot u_{\mathcal{O}} = -1$ given that $\xi \cdot \xi = g_{tt} = -(1 - 2m/r)$. We have then, from (7.26),

$$\varepsilon = -\left(1 - \frac{2m}{r}\right)^{1/2} u_{\mathcal{O}} \cdot u = \Gamma \left(1 - \frac{2m}{r}\right)^{1/2}, \quad (7.29)$$

where $\Gamma = -u_{\mathcal{O}} \cdot u$ is the Lorentz factor of \mathcal{P} with respect to \mathcal{O} (cf. Sec. 1.4; in particular Eq. (1.33)). We may express Γ in terms of the norm v of the velocity of \mathcal{P} with respect to \mathcal{O} , according to Eq. (1.35): $\Gamma = (1 - v^2)^{-1/2}$ and get

$$\varepsilon = (1 - v^2)^{-1/2} \left(1 - \frac{2m}{r}\right)^{1/2}. \quad (7.30)$$

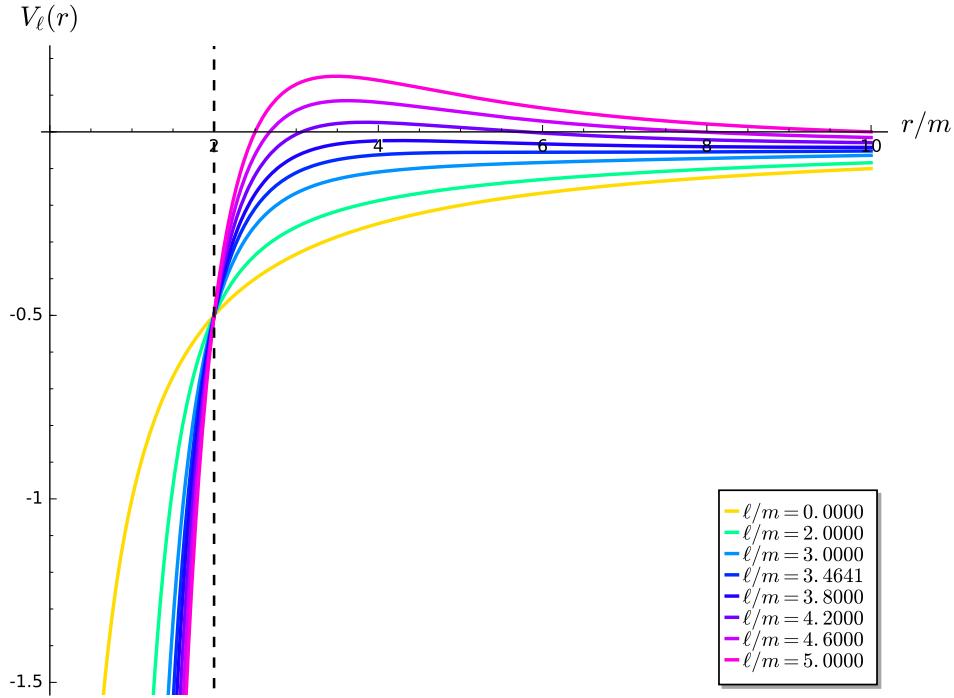


Figure 7.1: Effective potential $V_\ell(r)$ governing the r -part of the motion along a timelike geodesic in Schwarzschild spacetime via Eq. (7.24). The vertical dashed line marks $r = 2m$, i.e. the location of the event horizon. The numerical value $\ell/m = 3.4641$ is that of the critical specific angular momentum (7.32).

In the region $r \gg m$, we may perform a first order expansion, assuming that \mathcal{P} moves at nonrelativistic velocity with respect to \mathcal{O} ($v \ll 1$), thereby obtaining:

$$\varepsilon - 1 \simeq \frac{1}{2}v^2 - \frac{m}{r} \quad (r \gg m \text{ and } v \ll 1). \quad (7.31)$$

We recognize in the right-hand side the *Newtonian mechanical energy per unit mass* of particle \mathcal{P} with respect to observer \mathcal{O} , who can then be 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 $dV_\ell/dr = 0$, which is equivalent to

$$mr^2 - \ell^2 r + 3\ell^2 m = 0.$$

This quadratic equation admits real roots iff $|\ell| \geq \ell_{\text{crit}}$, with

$$\ell_{\text{crit}} = 2\sqrt{3}m \simeq 3.464102m. \quad (7.32)$$

For $|\ell| \geq \ell_{\text{crit}}$, the two roots are

$$r_{\max} = \frac{\ell}{2m} \left(\ell - \sqrt{\ell^2 - \ell_{\text{crit}}^2} \right) \quad \text{and} \quad r_{\min} = \frac{\ell}{2m} \left(\ell + \sqrt{\ell^2 - \ell_{\text{crit}}^2} \right), \quad (7.33)$$

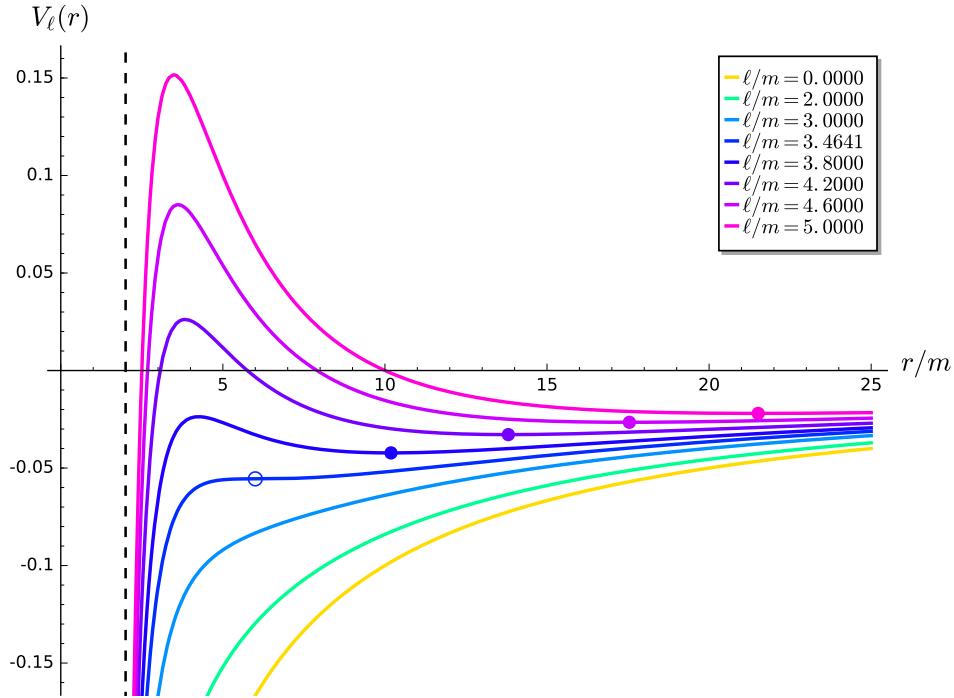


Figure 7.2: Same as Fig. 7.1, but with a zoom in along the y -axis and a zoom out along the x -axis. The dots mark the minima of V_ℓ , locating stable circular orbits.

corresponding respectively to a maximum of V_ℓ and a minimum of V_ℓ , hence the indices ‘‘max’’ and ‘‘min’’. Note that $r_{\text{max}} \leq r_{\text{min}}$. In the marginal case $|\ell| = \ell_{\text{crit}}$, the two roots coincide and correspond to an inflection point of V_ℓ (the circled dot in Fig. 7.2).

For $|\ell| < \ell_{\text{crit}}$, 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

$$\frac{dt}{d\tau} = \varepsilon \left(1 - \frac{2m}{r(\tau)}\right)^{-1}, \quad (7.34)$$

$$\frac{d\varphi}{d\tau} = \frac{\ell}{r(\tau)^2}. \quad (7.35)$$

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

$$\frac{1}{2} \left(\frac{dr}{d\tau} \right)^2 - \frac{m}{r} = \frac{\varepsilon^2 - 1}{2}. \quad (7.36)$$

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

$$\begin{cases} \tau = \frac{m}{(\varepsilon^2 - 1)^{3/2}} (\sinh \eta - \eta) + \tau_0 \\ r = \frac{m}{\varepsilon^2 - 1} (\cosh \eta - 1), \end{cases} \quad (7.37)$$

- if $\varepsilon = 1$, the solution is

$$r(\tau) = \left(\frac{9m}{2} (\tau - \tau_0)^2 \right)^{1/3}, \quad (7.38)$$

- if $\varepsilon < 1$, the solution is given in parameterized form (parameter η) by

$$\begin{cases} \tau = \frac{m}{(1 - \varepsilon^2)^{3/2}} (\eta + \sin \eta) + \tau_0 \\ r = \frac{m}{1 - \varepsilon^2} (1 + \cos \eta), \end{cases} \quad (7.39)$$

In the above formulas, τ_0 is a constant; for $\varepsilon \geq \mu$, τ_0 is the value of τ for which $r \rightarrow 0$, 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 $dr/d\tau = 0$ at $\tau = 0$. The equation of radial motion (7.36) leads then to $-m/r_0 = (\varepsilon^2 - 1)/2$, or equivalently

$$\varepsilon^2 = 1 - \frac{2m}{r_0}. \quad (7.40)$$

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 \mathcal{M}_{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

$$\begin{cases} \tau = \sqrt{\frac{r_0^3}{8m}} (\eta + \sin \eta) \\ r = \frac{r_0}{2} (1 + \cos \eta) \end{cases}$$

$0 \leq \eta \leq \pi, \quad (7.41)$

where the range of η is such that $r = r_0$ for $\tau = 0$ ($\eta = 0$) and r decays to 0 when $\eta \rightarrow \pi$. The function $r(\tau)$ resulting from (7.41) is depicted in Fig. 7.3.

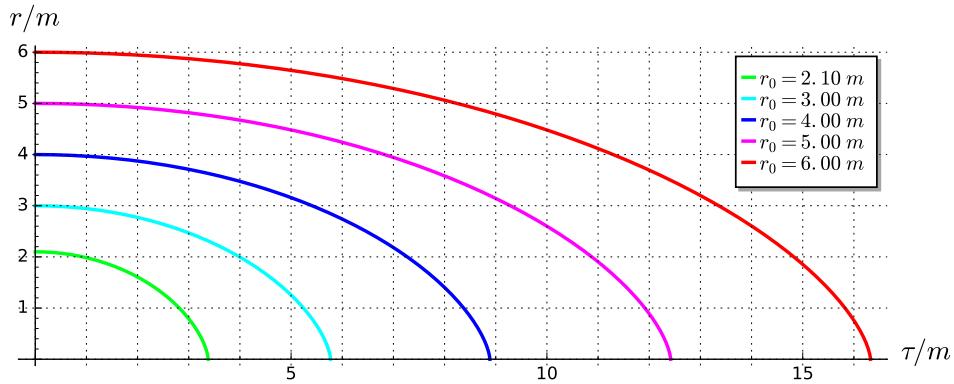


Figure 7.3: Coordinate r as a function of the proper time τ for the radial free fall from rest, for various initial values r_0 of r .

The solution for $t = t(\tau)$ is obtained by combining $dt/d\tau$ as expressed by (7.34) and $d\tau/d\eta$ deduced from (7.41):

$$\frac{d\tau}{d\eta} = \sqrt{\frac{r_0^3}{8m}} (1 + \cos \eta) = \sqrt{\frac{r_0}{2m}} r.$$

We get

$$\frac{dt}{d\eta} = \frac{dt}{d\tau} \frac{d\tau}{d\eta} = \varepsilon \sqrt{\frac{r_0}{2m}} r \left(1 - \frac{2m}{r}\right)^{-1} = \sqrt{\frac{r_0}{2m} - 1} \frac{r^2}{r - 2m},$$

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

$$\frac{dt}{d\eta} = \frac{r_0}{2} \sqrt{\frac{r_0}{2m} - 1} \frac{(1 + \cos \eta)^2}{1 + \cos \eta - 4m/r_0}.$$

This equation can be integrated to (cf. the SageMath computation in Sec. D.4.5)

$$t = 2m \left\{ \sqrt{\frac{r_0}{2m} - 1} \left[\eta + \frac{r_0}{4m} (\eta + \sin \eta) \right] + \ln \left| \frac{\sqrt{\frac{r_0}{2m} - 1} + \tan \frac{\eta}{2}}{\sqrt{\frac{r_0}{2m} - 1} - \tan \frac{\eta}{2}} \right| \right\}, \quad (7.42)$$

where we have assumed $t = 0$ at $\tau = 0$ ($\eta = 0$).

The solution for 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 \mathcal{H} . So, one might wonder if such a solution can describe the full infall, with the crossing of \mathcal{H} . In particular, we notice that the differential equation for $t(\tau)$, Eq. (7.34), is singular at $r(\tau) = 2m$, i.e. on \mathcal{H} . The solution $t(\eta)$, as given by Eq. (7.42), is singular at $\eta = \eta_h$, where

$$\eta_h := 2 \arctan \sqrt{\frac{r_0}{2m} - 1} \quad (7.43)$$

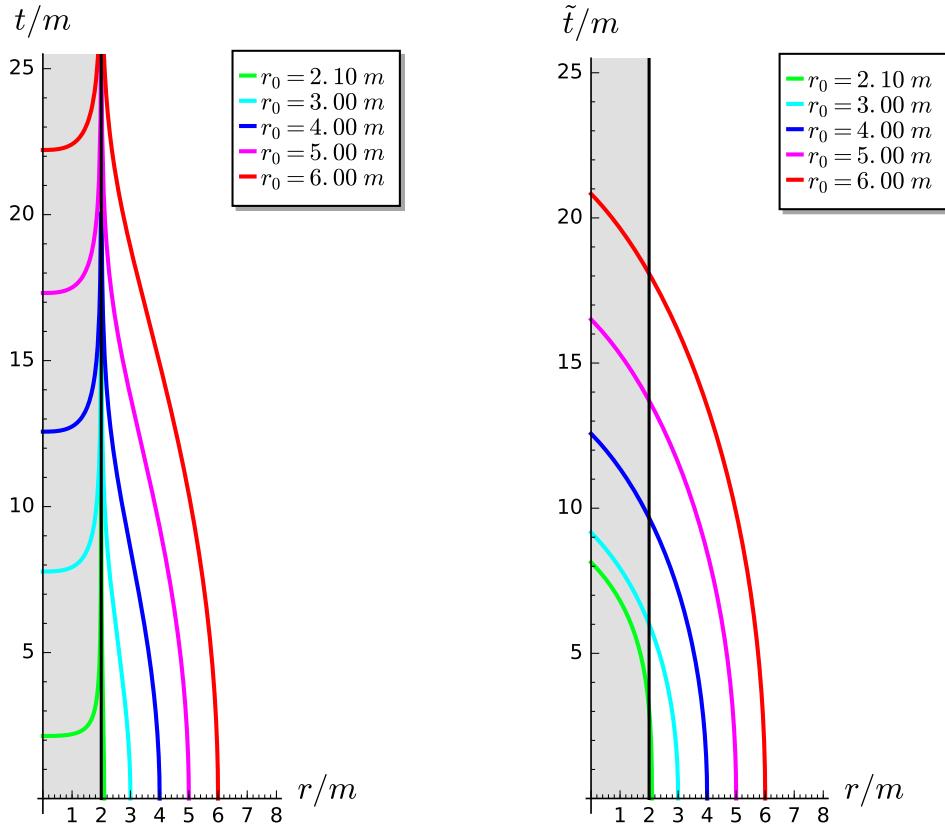


Figure 7.4: Radial free fall from rest, viewed in Schwarzschild-Droste coordinates (t, r) (left) and in the ingoing Eddington-Finkelstein coordinates (\tilde{t}, r) (right), for various values r_0 of the coordinate r at $\tau = 0$. The grey area is the black hole region \mathcal{M}_{II} .

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 found⁶ exactly the same solution for $r(\tau)$ (which is not surprising since r , considered as a scalar field on \mathcal{M} , is perfectly regular at \mathcal{H}). The solution for $\tilde{t}(\tau)$ can be deduced from that for $t(\tau)$ by the coordinate transformation law (6.32). Noticing that and $r/(2m) = \cos^2(\eta/2)/\cos^2(\eta_h/2)$, we get

$$\begin{aligned} \frac{r}{2m} - 1 &= \frac{\cos^2(\eta/2)}{\cos^2(\eta_h/2)} - 1 = \cos^2(\eta/2) \left(\frac{1}{\cos^2(\eta_h/2)} - \frac{1}{\cos^2(\eta/2)} \right) \\ &= \cos^2(\eta/2) (\tan^2(\eta_h/2) - \tan^2(\eta/2)). \end{aligned} \quad (7.44)$$

Using this identity, as well as (7.43) to express $\sqrt{r_0/(2m) - 1}$ in the logarithm term of Eq. (7.42),

⁶Exercise: do it!

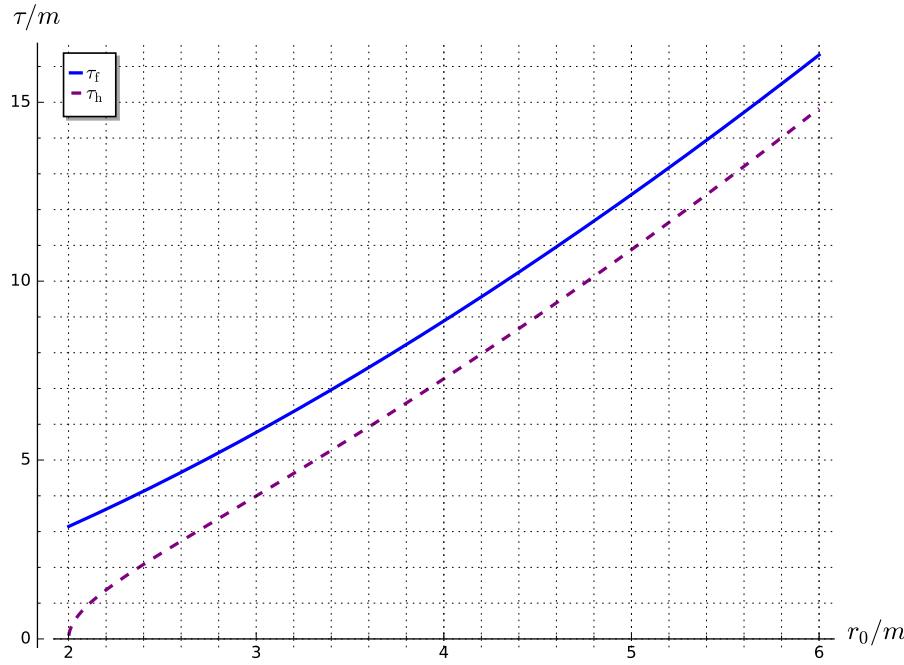


Figure 7.5: Elapsed proper time to reach the event horizon (τ_h , dashed curve) and the central singularity (τ_f , solid curve), as a function of the initial value of r for a radial free fall from rest.

the transformation law (6.32) yields

$$\begin{aligned} \tilde{t} &= 2m \left\{ \sqrt{\frac{r_0}{2m} - 1} \left[\eta + \frac{r_0}{4m}(\eta + \sin \eta) \right] + \ln \left| \frac{\tan \frac{\eta_h}{2} + \tan \frac{\eta}{2}}{\tan \frac{\eta_h}{2} - \tan \frac{\eta}{2}} \cos^2 \frac{\eta}{2} \left(\tan^2 \frac{\eta_h}{2} - \tan^2 \frac{\eta}{2} \right) \right| \right\} \\ &= 2m \left\{ \sqrt{\frac{r_0}{2m} - 1} \left[\eta + \frac{r_0}{4m}(\eta + \sin \eta) \right] + \ln \left| \cos^2 \frac{\eta}{2} \left(\tan \frac{\eta_h}{2} + \tan \frac{\eta}{2} \right)^2 \right| \right\} \\ &= 2m \left\{ \sqrt{\frac{r_0}{2m} - 1} \left[\eta + \frac{r_0}{4m}(\eta + \sin \eta) \right] + 2 \ln \left(\cos \frac{\eta}{2} \tan \frac{\eta_h}{2} + \sin \frac{\eta}{2} \right) \right\}. \end{aligned}$$

From this expression, we have $\tilde{t} = 4m \ln \tan(\eta_h/2)$ 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 \mathcal{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 \mathcal{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

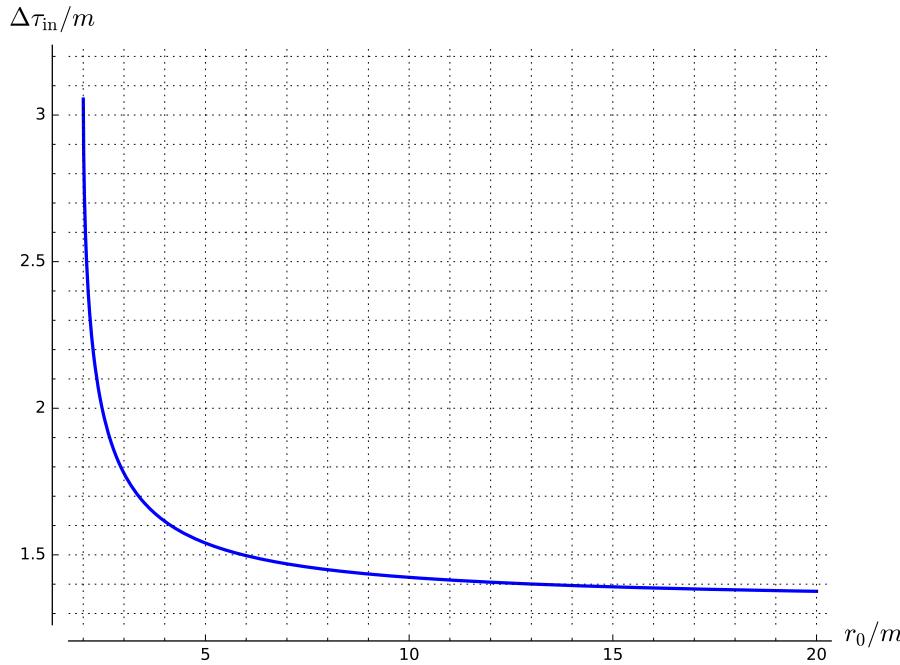


Figure 7.6: Proper time spent inside the black hole region as a function of the initial value of r for a radial free fall from rest. Note that $r_0 = 2m$ does not correspond to any asymptote but to the finite value $\Delta\tau_{\text{in}} = \pi m$ with a vertical tangent. On the other side, there is an horizontal asymptote $\Delta\tau_{\text{in}} \rightarrow 4m/3$ for $r_0 \rightarrow +\infty$.

setting $\eta = \pi$ in Eq. (7.41):

$$\boxed{\tau_f = \frac{\pi}{2} \sqrt{\frac{r_0^3}{2m}}}. \quad (7.46)$$

Similarly, the final value of \tilde{t} is obtained by setting $\eta = \pi$ in (7.45):

$$\boxed{\tilde{t}_f = 2m \left[\pi \sqrt{\frac{r_0}{2m} - 1} \left(\frac{r_0}{4m} + 1 \right) - \ln \left(\frac{r_0}{2m} - 1 \right) \right]}. \quad (7.47)$$

As noticed above, the event horizon \mathcal{H} is crossed at $\eta = \eta_h$; via (7.41) and (7.43), this corresponds to the following value of the proper time:

$$\boxed{\tau_h = \sqrt{\frac{r_0^3}{2m}} \left[\arctan \sqrt{\frac{r_0}{2m} - 1} + \sqrt{\frac{2m}{r_0} \left(1 - \frac{2m}{r_0} \right)} \right]}, \quad (7.48)$$

while (7.45) leads to the following value of the IEF coordinate \tilde{t} :

$$\boxed{\tilde{t}_h = 2m \left[2 \left(1 + \frac{r_0}{4m} \right) \sqrt{\frac{r_0}{2m} - 1} \arctan \sqrt{\frac{r_0}{2m} - 1} + \frac{r_0}{2m} - 1 - \ln \frac{r_0}{2m} \right]}. \quad (7.49)$$

The variation of τ_h and τ_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.

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
$4.3 \cdot 10^6 M_\odot$ (Sgr A*)	$12.7 \cdot 10^6$ km = 0.085 au	5 min 14 s	5 min 46 s	32 s
$6 \cdot 10^9 M_\odot$ (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 [392, 224], the supermassive black hole at the center of our galaxy (Sagittarius A*) [214, 300] and the supermassive black hole M87* in the nucleus of the galaxy Messier 87 [213, 6].

The proper time spent inside the black hole is

$$\Delta\tau_{in} = \tau_f - \tau_h = \sqrt{\frac{r_0^3}{2m}} \left[\frac{\pi}{2} - \arctan \sqrt{\frac{r_0}{2m} - 1} - \sqrt{\frac{2m}{r_0} \left(1 - \frac{2m}{r_0} \right)} \right]. \quad (7.50)$$

It varies between πm ($r_0 \rightarrow 2m$) and $4m/3$ ($r_0 \rightarrow +\infty$) (cf. Fig. 7.6 and Sec. D.4.5 for the computation of $\lim_{r_0 \rightarrow +\infty} \Delta\tau_{in}$). Numerical values for astrophysical black holes are provided in Table 7.1.

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

$$V_\ell(r) = \frac{\varepsilon^2 - 1}{2} \quad (7.51a)$$

$$\frac{dV_\ell}{dr} = 0. \quad (7.51b)$$

Given the expression (7.25) of V_ℓ , Eq. (7.51b) is equivalent to

$$mr^2 - \ell^2 r + 3\ell^2 m = 0. \quad (7.52)$$

As already noticed in Sec. 7.3.1, this quadratic equation in r admits two real roots iff $|\ell| \geq \ell_{\text{crit}}$, with $\ell_{\text{crit}} = 2\sqrt{3}m$ [Eq. (7.32)], which are

$$r_{\text{circ}}^\pm(\ell) = \frac{\ell}{2m} \left(\ell \pm \sqrt{\ell^2 - \ell_{\text{crit}}^2} \right). \quad (7.53)$$

$r_{\text{circ}}^+(\ell)$ corresponds to a minimum of the effective potential V_ℓ and thus to a stable orbit (see the dots in Fig. 7.2), while $r_{\text{circ}}^-(\ell)$ corresponds to a maximum of V_ℓ and thus to an unstable orbit. When ℓ varies from ℓ_{crit} to $+\infty$, $r_{\text{circ}}^+(\ell)$ increases from $6m$ to $+\infty$, while $r_{\text{circ}}^-(\ell)$ decreases from $6m$ to $3m$ (cf. Fig. 7.7). We conclude that

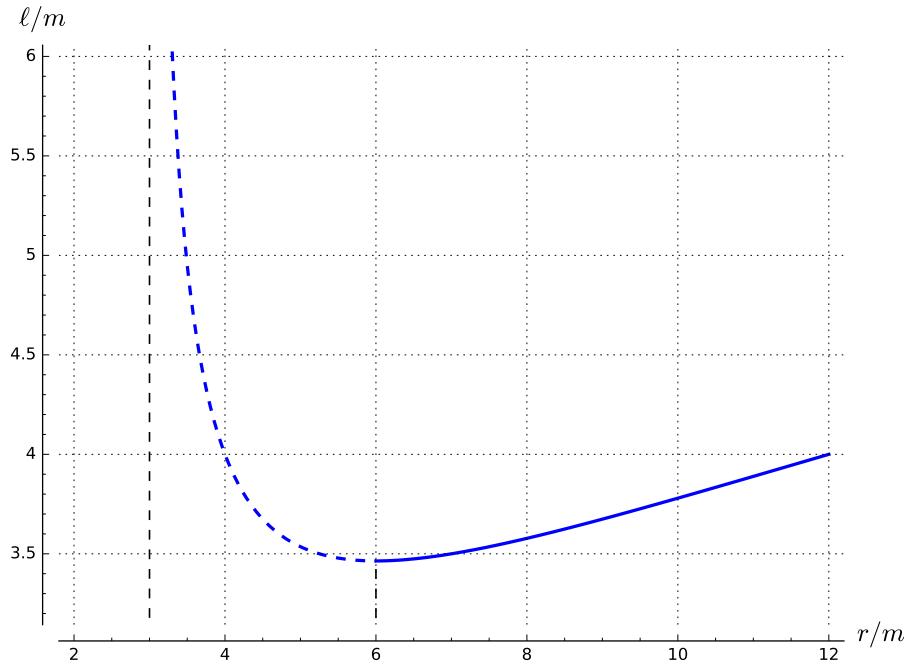


Figure 7.7: Specific conserved angular momentum $\ell = L/\mu$ on circular orbits as a function of the orbit circumferential radius r . The dashed part of the curve corresponds to unstable orbits ($r = r_{\text{circ}}^-(\ell)$, as given by Eq. (7.53)), while the solid part corresponds to stable orbits ($r = r_{\text{circ}}^+(\ell)$). The minimal value of ℓ is $\ell_{\text{crit}} = 2\sqrt{3}m \simeq 3.46 m$.

Property 7.8: circular orbits

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. 213), 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. $|\ell| < \ell_{\text{crit}}$).

From Eq. (7.52), we can easily express ℓ as a function of r on a circular orbit:

$$|\ell| = r \sqrt{\frac{m}{r - 3m}}. \quad (7.54)$$

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 :

$$\boxed{\varepsilon = \frac{r - 2m}{\sqrt{r(r - 3m)}}}. \quad (7.55)$$

This function is represented in Fig. 7.8. The minimal value of ε is achieved for $r = 6m$, i.e. at the ISCO:

$$\boxed{\min \varepsilon = \varepsilon_{\text{ISCO}} = \frac{2\sqrt{2}}{3} \simeq 0.9428}. \quad (7.56)$$

From Fig. 7.8, we notice that

$$r > 4m \iff \varepsilon < 1 \iff E < \mu. \quad (7.57)$$

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 $r \rightarrow +\infty$, the Killing vector ξ 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 (1.32) is then $E = \Gamma\mu$, where Γ is the Lorentz factor of the particle with respect to the observer. Since $\Gamma \geq 1$ [Eq. (1.35)], we have obviously⁷ $E \geq \mu$. 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 \mathcal{L} is defined by

$$\Omega := \left. \frac{d\varphi}{dt} \right|_{\mathcal{L}} = \frac{u^\varphi}{u^t}, \quad (7.58)$$

where $u^\varphi = d\varphi/d\tau$ and $u^t = dt/d\tau$ are the only nonzero components w.r.t. Schwarzschild-Droste coordinates of the 4-velocity \mathbf{u} along the worldline \mathcal{L} . It follows from (7.58) that Ω enters into the linear combination of the two Killing vectors ξ and η expressing the 4-velocity on a circular orbit according to

$$\mathbf{u} = u^t (\xi + \Omega \eta). \quad (7.59)$$

We have the following nice physical interpretation:

Property 7.9: angular velocity measured at infinity

The quantity Ω defined by Eq. (7.58) is nothing but the angular velocity of the orbiting particle \mathcal{P} monitored by a infinitely distant static observer \mathcal{O} .

⁷Similarly, the radial-motion solutions (7.37)-(7.38), which allow for $r \rightarrow +\infty$, have $E \geq \mu$, while the solution (7.39), which is relevant for a free fall from rest, has $E < \mu$.

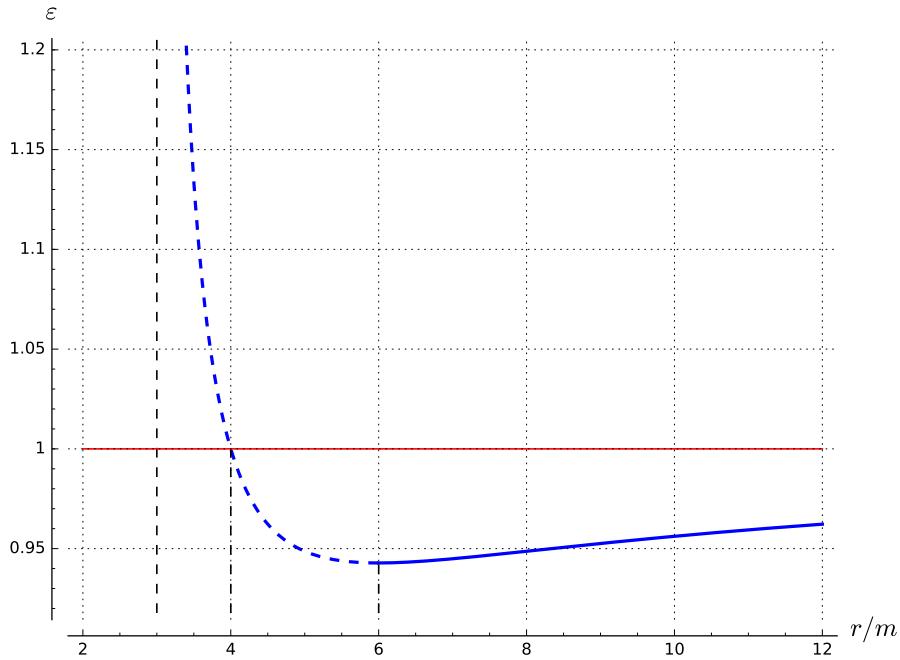


Figure 7.8: Specific conserved energy $\varepsilon = E/\mu$ on circular orbits as a function of the orbit circumferential radius r . The dashed part of the curve corresponds to unstable orbits, while the solid part corresponds to stable ones. The horizontal red line $\varepsilon = 1$ marks the limit of bound orbits.

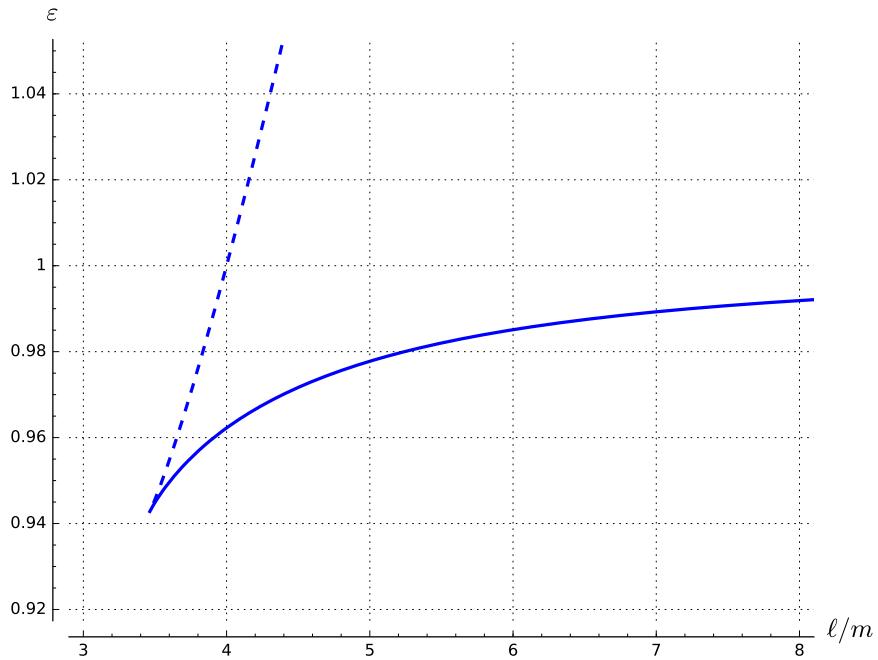


Figure 7.9: Circular orbits in the (ℓ, ε) plane. The solid (resp. dashed) curve corresponds to stable (resp. unstable) orbits. The ISCO is located at the cusp point.

m	$r_{\text{ISCO}} = 6m$	$\frac{\Omega_{\text{ISCO}}}{2\pi}$	T_{ISCO}	$T_{\mathcal{P},\text{ISCO}}$
$15 M_{\odot}$ (Cyg X-1)	133 km	147 Hz	6.80 ms	4.81 ms
$4.3 \cdot 10^6 M_{\odot}$ (Sgr A*)	$38.1 \cdot 10^6$ km (0.255 au)	$5.11 \cdot 10^{-4}$ Hz	32 min 37 s	23 min 4 s
$6 \cdot 10^9 M_{\odot}$ (M87*)	355 au	$3.66 \cdot 10^{-7}$ 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 $\Omega_{\text{ISCO}}/(2\pi)$, orbital period seen from infinity T_{ISCO} and orbital period measured by the orbiting observer/particle $T_{\mathcal{P},\text{ISCO}}$ (proper time).

Proof. Suppose that \mathcal{O} is located at fixed coordinates $(r, \theta, \varphi) = (r_{\mathcal{O}}, \pi/2, 0)$ with $r_{\mathcal{O}} \gg m$ and that \mathcal{P} emits a photon at the event $(t_1, r, \pi/2, 0)$ along a *radial* null geodesic. This photon is received by \mathcal{O} at $t = t'_1$. After one orbit, at the event $(t_2, r, \pi/2, 2\pi)$, \mathcal{P} emits a second photon in the radial direction, which is received at $t = t'_2$ by \mathcal{O} . According to the definition (7.58) of Ω , we have

$$2\pi = \Omega(t_2 - t_1).$$

On the other hand, since $r_{\mathcal{O}} \gg m$, the proper time of \mathcal{O} is t , so that the angular velocity measured by \mathcal{O} is

$$\Omega_{\mathcal{O}} = \frac{2\pi}{t'_2 - t'_1}.$$

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 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 $\Omega_{\mathcal{O}} = \Omega$. \square

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

$$\Omega = \pm \sqrt{\frac{m}{r^3}},$$

(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 \mathcal{P} on the circular orbit \mathcal{L} , because the proper time of \mathcal{P} is τ and not t . The actual orbital frequency measured by \mathcal{P} is

$$\Omega_{\mathcal{P}} = \frac{dt}{d\tau} \Omega = u^t \Omega,$$

with $u^t = dt/d\tau$ obtained from (7.34) and (7.55): $u^t = \sqrt{r/(r-3m)}$. Hence

$$\Omega_{\mathcal{P}} = \sqrt{\frac{r}{r-3m}} \Omega = \pm \frac{1}{r} \sqrt{\frac{m}{r-3m}}. \quad (7.61)$$

Note that $|\Omega_{\mathcal{P}}| > |\Omega|$; in particular, at the ISCO ($r = 6m$), $\Omega_{\mathcal{P}} = \sqrt{2}\Omega$. The orbital period measured by \mathcal{P} is $T_{\mathcal{P}} = 2\pi/|\Omega_{\mathcal{P}}|$. Some ISCO values of $T_{\mathcal{P}}$ for astrophysical black holes are provided in Table 7.2.

At the ISCO, $r = 6m$ and formula (7.60) yields (for orbits with positive ℓ)

$$\Omega_{\text{ISCO}} = \frac{1}{6\sqrt{6}m}.$$

(7.62)

Numerical values of Ω_{ISCO} (actually the frequency $\Omega_{\text{ISCO}}/(2\pi)$, 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 \mathcal{L} obeying

$$|\ell| > \ell_{\text{crit}} \quad \text{and} \quad 0 < \varepsilon < 1. \quad (7.63)$$

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 \mathcal{P} is trapped in this well. Indeed, $0 < \varepsilon < 1$ makes the right-hand of Eq. (7.24) negative, so that the region $r \rightarrow +\infty$, where $V_{\ell}(r) \rightarrow 0$, cannot be reached. We have also argued in Sec. 7.3.3 that $\varepsilon < 1$ is forbidden in the region $r \rightarrow +\infty$ on physical grounds [cf. the discussion below Eq. (7.57)].

In the potential well, the r -coordinate along \mathcal{L} varies between two extrema: a minimum r_{per} , for **periastron** (or **pericenter** or **periapsis**), and a maximum r_{apo} , for **apoastron** (or **apocenter** or **apoapsis**) (cf. Figs. 7.10 and 7.11). Being extrema of $r(\tau)$, the values of r_{per} and r_{apo} are obtained by setting $dr/d\tau = 0$ in Eq. (7.24), which leads to

$$\left(1 - \frac{2m}{r}\right) \left(1 + \frac{\ell^2}{r^2}\right) = \varepsilon^2. \quad (7.64)$$

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),

$$\frac{dr}{d\tau} = \pm \sqrt{\varepsilon^2 - \left(1 - \frac{2m}{r}\right) \left(1 + \frac{\ell^2}{r^2}\right)}. \quad (7.65)$$

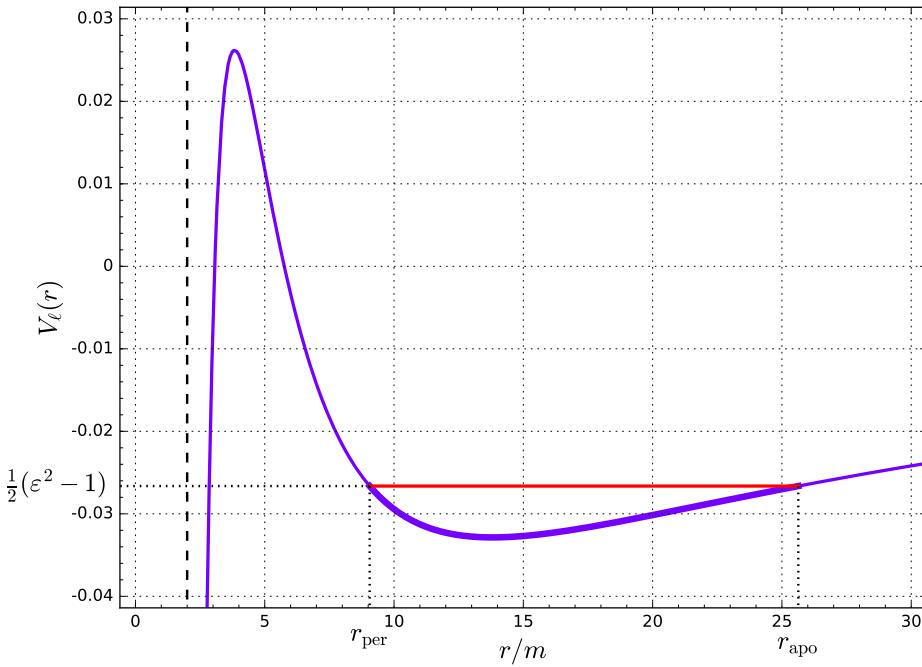


Figure 7.10: Effective potential $V_\ell(r)$ for $\ell = 4.2m$ (one of the values displayed in Figs. 7.1 and 7.2). The horizontal red line marks $V_\ell(r) = (\varepsilon^2 - 1)/2$ with $\varepsilon = 0.973$, leading to $r_{\text{per}} = 9.058 \text{ m}$ and $r_{\text{apo}} = 25.634 \text{ m}$. The corresponding orbit is shown in Fig. 7.11.

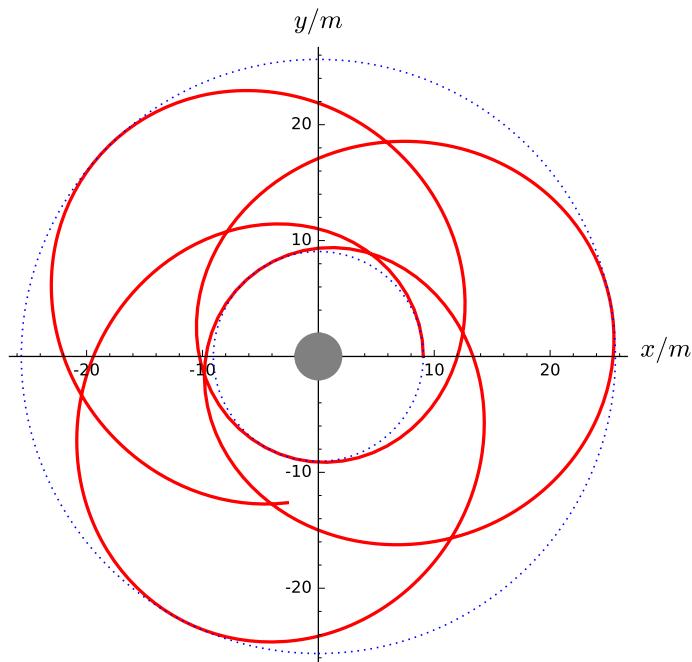


Figure 7.11: Timelike geodesic with $\varepsilon = 0.973$ and $\ell = 4.2m$ (same values as in Fig. 7.10), plotted in terms of the coordinates $(x, y) := (r \cos \varphi, r \sin \varphi)$. The dotted circles correspond to $r = r_{\text{per}}$ (periastron) and $r = r_{\text{apo}}$ (apoastron). The grey disk indicates the black hole region $r < 2m$. [Figure produced with the notebook [D.4.6](#)]

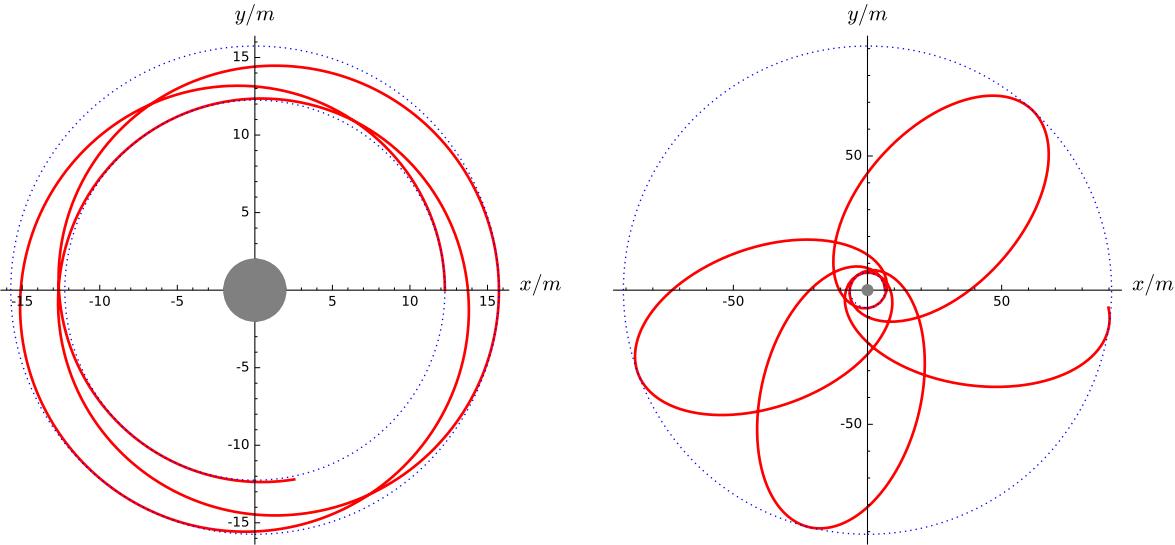


Figure 7.12: Timelike geodesics with the same value of ℓ as in Fig. 7.11 ($\ell = 4.2m$), but for different values of ε : $\varepsilon = 0.967$ (left) and $\varepsilon = 0.990$ (right). Note that the left and right figures have different scales. [Figure produced with the notebook D.4.6]

The equation governing the trajectory of \mathcal{L} in the orbital plane is Eq. (7.17), which can be recast as

$$\frac{du}{d\varphi} = \pm \sqrt{2u^3 - u^2 + 2\left(\frac{m}{\ell}\right)^2 u + \left(\frac{m\varepsilon}{\ell}\right)^2 - \left(\frac{m}{\ell}\right)^2}. \quad (7.66)$$

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

$$\frac{du}{d\varphi} \simeq \pm \sqrt{2\left(\frac{m}{\ell}\right)^2 \varepsilon_0 - u^2 + 2\left(\frac{m}{\ell}\right)^2 u}. \quad (7.67)$$

Let us introduce the constants

$$p := \frac{\ell^2}{m} \quad \text{and} \quad e := \sqrt{1 + 2\frac{\varepsilon_0 \ell^2}{m^2}}. \quad (7.68)$$

Then Eq. (7.67) can be rewritten as

$$\frac{d\varphi}{du} = \pm \frac{\frac{p}{me}}{\sqrt{1 - \left(\frac{p}{m}u - 1\right)^2}},$$

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,

$$r = \frac{p}{1 + e \cos(\varphi - \varphi_0)} \quad (7.69)$$

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 \mathcal{P} for which $r = r_{\text{per}}$, 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 [448] in which he presented his famous solution⁸. 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 spacetime⁹ has been given as early as May 1916 by Johannes Droste, in the same article [174] in which he derived the Schwarzschild solution, independently of Karl Schwarzschild (cf. historical note on p. 185). 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 plane¹⁰. 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 Carlo De Jans in 1923 [159] and other one by Yusuke Hagihara in 1931 [248]. For a detailed account about the history of geodesic motion in Schwarzschild spacetime see Ref. [180].

⁸Equation (7.17) for the trajectories is Eq. (18) in the article [448], the link between Schwarzschild's notations and ours being $R = r$, $x = u/m$, $c = L$, $1 = E$, $h = \mu^2$.

⁹More precisely: in the \mathcal{M}_1 region of Schwarzschild spacetime.

¹⁰The 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	231
8.2	Main properties of null geodesics	232
8.3	Trajectories of null geodesics in the equatorial plane	239
8.4	Asymptotic direction from some emission point	258
8.5	Images	265

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* [6]. 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: (\mathcal{M}, g) is Schwarzschild spacetime, (t, r, θ, φ) are Schwarzschild-Droste coordinates and \mathcal{P} is a particle, the worldline of which is a geodesic \mathcal{L} . The particle \mathcal{P} is characterized by its 4-momentum \mathbf{p} and E and L stand for the conserved energy and conserved angular momentum along \mathcal{L} [cf. Eq. (7.1)].

In all this chapter, we consider that \mathcal{L} is a *null* geodesic, or equivalently that \mathcal{P} is a *massless* particle, typically a photon. As in Chap. 7, we assume that the coordinates (θ, φ) are chosen so that \mathcal{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 \mathcal{P} is governed by Eqs. (7.11), (7.12) and (7.13) with $\mu = 0$ (massless particle):

$$\boxed{\frac{dt}{d\lambda} = E \left(1 - \frac{2m}{r}\right)^{-1}}, \quad (8.1)$$

$$\boxed{\frac{d\varphi}{d\lambda} = \frac{L}{r^2}}, \quad (8.2)$$

$$\boxed{\left(\frac{dr}{d\lambda}\right)^2 + \frac{L^2}{r^2} \left(1 - \frac{2m}{r}\right) = E^2}. \quad (8.3)$$

Let us recall that the variable λ with respect to which these differential equations hold is the affine parameter associated with the 4-momentum \mathbf{p} of \mathcal{P} : $\mathbf{p} = d\mathbf{x}/d\lambda$ [Eq. (7.9)] and that λ increases towards the future, \mathbf{p} being always future-directed (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:

$$\frac{dr}{d\lambda} = \pm E, \quad (8.4)$$

the solution to which is immediate:

$$r = \pm E\lambda + r_0, \quad (8.5)$$

where r_0 is some constant. Moreover, writing $dt/d\lambda = dt/dr \times dr/d\lambda$ and combining Eqs. (8.1) and (8.4), we get

$$\frac{dt}{dr} = \pm \left(1 - \frac{2m}{r}\right)^{-1}. \quad (8.6)$$

We recognize the equation governing the radial null geodesics $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ and $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$ obtained in Sec. 6.3.1 [Eq. (6.19)], the solution to which is given by Eq. (6.20):

$$t = \pm r \pm 2m \ln \left| \frac{r}{2m} - 1 \right| + \text{const.} \quad (8.7)$$

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.

Remark 2: The radial null geodesics $\mathcal{L}_{(\theta,\varphi)}^{\text{out},\mathcal{H}}$ discussed in Sec. 6.3.5 have $L = 0$, but they are not recovered from Eq. (8.6), since the latter is based on Schwarzschild-Droste coordinates (t, r, θ, φ) and the $\mathcal{L}_{(\theta,\varphi)}^{\text{out},\mathcal{H}}$'s are the null geodesic generators of the event horizon \mathcal{H} , which does not lie in the Schwarzschild-Droste domain.

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 \mathcal{L} the affine parameter

$$\tilde{\lambda} := |L|\lambda, \quad (8.8)$$

instead of λ , the latter being the affine parameter associated with the 4-momentum 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

$$\frac{dt}{d\tilde{\lambda}} = b^{-1} \left(1 - \frac{2m}{r} \right)^{-1}, \quad (8.9)$$

$$\frac{d\varphi}{d\tilde{\lambda}} = \frac{\epsilon_L}{r^2}, \quad (8.10)$$

$$\left(\frac{dr}{d\tilde{\lambda}} \right)^2 + U(r) = \frac{1}{b^2}, \quad (8.11)$$

where

$$b := \frac{|L|}{E}, \quad (8.12)$$

$$\epsilon_L := \frac{L}{|L|} = \text{sgn } L \quad (8.13)$$

and

$$U(r) := \frac{1}{r^2} \left(1 - \frac{2m}{r} \right). \quad (8.14)$$

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

Property 8.1: characterization of null geodesics by the impact parameter

A null geodesic \mathcal{L} is entirely characterized by the constants b and $\epsilon_L = \pm 1$. As soon as \mathcal{L} has some part in region \mathcal{M}_I , then $E > 0$ [cf. Eq. (7.5)], so that b , as defined by Eq. (8.12), is finite and positive. It 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 \mathcal{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 \mathcal{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

$$\varphi \simeq \frac{y_0}{r} \quad (r \gg m).$$

Deriving with respect to t leads to

$$\frac{d\varphi}{dt} \simeq -\frac{y_0}{r^2} \frac{dr}{dt} \simeq \frac{y_0}{r^2} \quad (r \gg m), \quad (8.15)$$

where we have used $dr/dt = -1$ 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 dr/dt . Now by combining Eqs. (8.9) and (8.10), we have

$$\frac{d\varphi}{dt} = \epsilon_L \frac{b}{r^2} \left(1 - \frac{2m}{r} \right). \quad (8.16)$$

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 \mathcal{P} with respect to the central black hole. \square

The effective potential $U(r)$ is plotted in Fig. 8.1. It has no minimum and a maximum at $r = 3m$, which is

$$U_{\max} = \frac{1}{27m^2}. \quad (8.17)$$

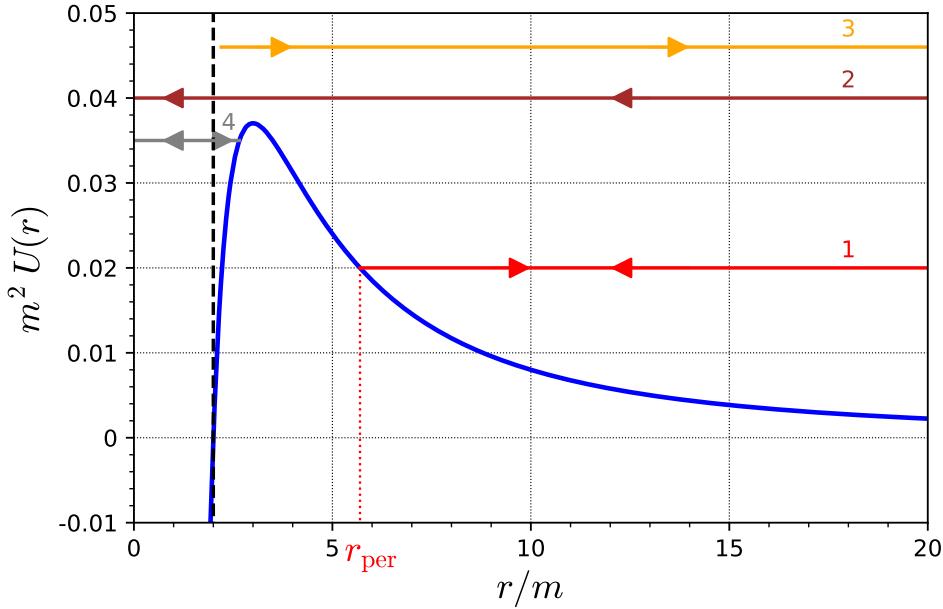


Figure 8.1: Effective potential $U(r)$ (rescaled by m^2 to make it dimensionless) governing the r -part of the motion along a null geodesic in Schwarzschild spacetime [Eq. (8.14)]. The vertical dashed line marks $r = 2m$, i.e. the location of the event horizon. The horizontal lines marked “1” to “4” 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.4.7]

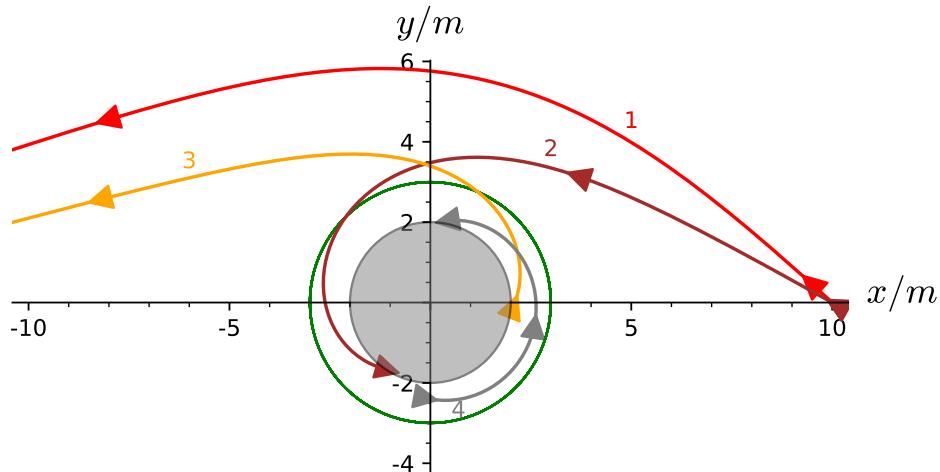


Figure 8.2: Selected null geodesics in the equatorial plane of Schwarzschild spacetime, plotted in terms of the coordinates $(x, y) := (r \cos \varphi, r \sin \varphi)$, with the black hole region $r < 2m$ depicted as a grey disk. The green geodesic is the photon circular orbit at $r = 3m$. The red geodesic (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 geodesics is depicted in Fig. 8.1. [Figure produced with the notebook D.4.8]

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...

8.2.4 Radial behavior of null geodesics

For the sake of concreteness, in this section and the remaining ones, we refer to the massless particle \mathcal{P} as a “photon”. In most applications, in particular the astrophysical ones, \mathcal{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 photons¹.

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 $(dr/d\tilde{\lambda})^2$. The region under the curve $U(r)$ is thus excluded.

For an initially inward photon, i.e. a photon emitted with $dr/d\tilde{\lambda} < 0$ from a position $r = r_{\text{em}}$, there are two possibilities, depending on the values of r_{em} and of the impact parameter b :

- if $r_{\text{em}} > 3m$ and b is large enough to fulfill $b^{-2} < U_{\max}$ (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_p of r – the **periastron**, which is given by $U(r_p) = b^{-2}$, or equivalently by

$$r_p^3 - b^2 r_p + 2mb^2 = 0, \quad r_p > 3m. \quad (8.18)$$

Equation (8.11) implies then

$$\left. \frac{dr}{d\tilde{\lambda}} \right|_{r=r_p} = 0. \quad (8.19)$$

Actually, $dr/d\tilde{\lambda}$ changes sign at $r = r_p$ 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 $r_{\text{em}} < 3m$ (e.g. trajectory no. 4 in Fig. 8.1) or b is small enough to fulfill $b^{-2} > U_{\max}$ (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 arbitrarily 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 $dr/d\tilde{\lambda} > 0$, one has necessarily $r_{\text{em}} > 2m$ according to the result obtained in Sec. 7.2.2 and there are then two possible outcomes:

¹unless one counts as “bound” a worldline that terminates in the black hole region

- if $r_{\text{em}} > 3m$ (e.g. trajectory no. 1 in Fig. 8.1) or b is small enough to fulfill $b^{-2} > U_{\max}$ (e.g. trajectory no. 3 in Figs. 8.1), the photon escapes to infinity (cf. geodesic no. 3 in Fig. 8.2);
- if $2m < r_{\text{em}} < 3m$ 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_a of r – the ***apoastron***, which is given by $U(r_a) = b^{-2}$, or equivalently by

$$r_a^3 - b^2 r_a + 2mb^2 = 0, \quad r_a < 3m. \quad (8.20)$$

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 $b_c^{-2} = U_{\max}$. Given the value (8.17) of U_{\max} , we get

$$b_c = 3\sqrt{3} m \simeq 5.196152 m. \quad (8.21)$$

The above discussion leads us to

Property 8.2: Behaviour of r along a null geodesics

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 $b > b_c$, the explicit expression of the periastron radius r_p (or the apoastron radius r_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 depressed² cubic equation, which makes it simpler to solve. For $b > b_c$, 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_p and r_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:

$$x_k = 2\sqrt{-\frac{p}{3}} \cos \left[\frac{1}{3} \arccos \left(\frac{3q}{2p} \sqrt{-\frac{3}{p}} \right) + \frac{2k\pi}{3} \right], \quad k \in \{0, 1, 2\}. \quad (8.22)$$

²A depressed cubic equation is a polynomial equation of the type $x^3 + px + q = 0$.

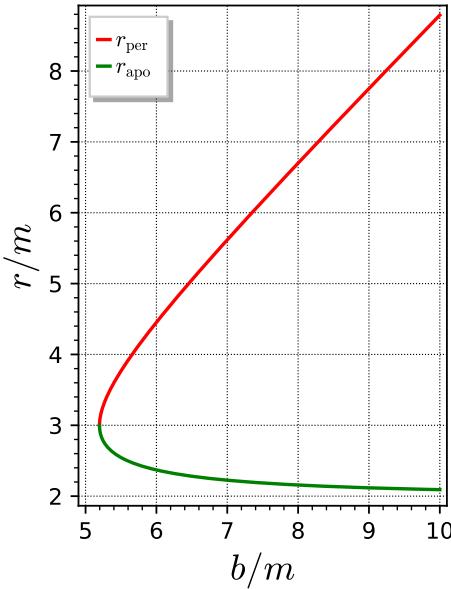


Figure 8.3: Radial coordinate of the periastron, r_p , and of the apoastrom, r_a , along null geodesics with $b > b_c \simeq 5.196 m$. Note that a given null geodesic has either a periastron or an apoastrom, but not both. [Figure produced with the notebook D.4.9]

In the present case, $p = -b^2$, $q = 2mb^2$ and r_p (resp. r_a) corresponds to $k = 0$ (resp. $k = 2$), so that we obtain

$$r_p = \frac{2b}{\sqrt{3}} \cos \left[\frac{\pi}{3} - \frac{1}{3} \arccos \left(\frac{b_c}{b} \right) \right]. \quad (8.23)$$

$$r_a = \frac{2b}{\sqrt{3}} \cos \left[\frac{5\pi}{3} - \frac{1}{3} \arccos \left(\frac{b_c}{b} \right) \right]. \quad (8.24)$$

As a check, we note that for $b \gg b_c$, Eq. (8.23) yields $r_p \simeq 2b/\sqrt{3} \cos(\pi/3 - 1/3 \times \pi/2) \simeq 2b/\sqrt{3} \cos(\pi/6) \simeq b$, as expected. Indeed, $b \gg b_c$ 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: $b \simeq r_p$.

The variation of r_p and r_a with b is plotted in Fig. 8.3. Note that $r_p = r_a = 3m$ (the photon orbit) at the limit $b = b_c$.

Remark 3: It is clear on Fig. 8.3 that, for the same value of b , $r_a \leq r_p$, which may seem contradictory with the apoastrom 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 apoastrom and the periastron always refer to different null geodesics: geodesics with an apoastrom 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 apoastrom 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): $d\varphi/d\tilde{\lambda} = \epsilon_L/r^2$, given that $\epsilon_L = \text{sgn } L$ 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)]:

$$\left(\frac{du}{d\varphi} \right)^2 = P_b(u) \quad (8.25)$$

where $u := m/r$ [Eq. (7.16)] and $P_b(u)$ stands for the cubic polynomial

$$P_b(u) := 2u^3 - u^2 + \frac{m^2}{b^2}. \quad (8.26)$$

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 $du/d\varphi = 0$. They are thus stationary points for u , and hence for r . This can only occur at the circular photon orbit $r = \text{const} = 3m$ (then $b = b_c$) or at the periastron or apoastron discussed in Sec. 8.2.4 (then $b > b_c$). The zeros of P_b are governed by its discriminant, which is

$$\Delta = 4 \frac{m^4}{b^4} \left(\frac{b^2}{m^2} - 27 \right). \quad (8.27)$$

One may distinguish three cases:

- $\Delta > 0 \iff b > \sqrt{27}m = b_c$: 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

$$u_p = \frac{m}{r_p} \quad \text{and} \quad u_a = \frac{m}{r_a}, \quad (8.28)$$

where r_p and r_a are the periastron and apoastron radii given by Eqs. (8.23)-(8.24).

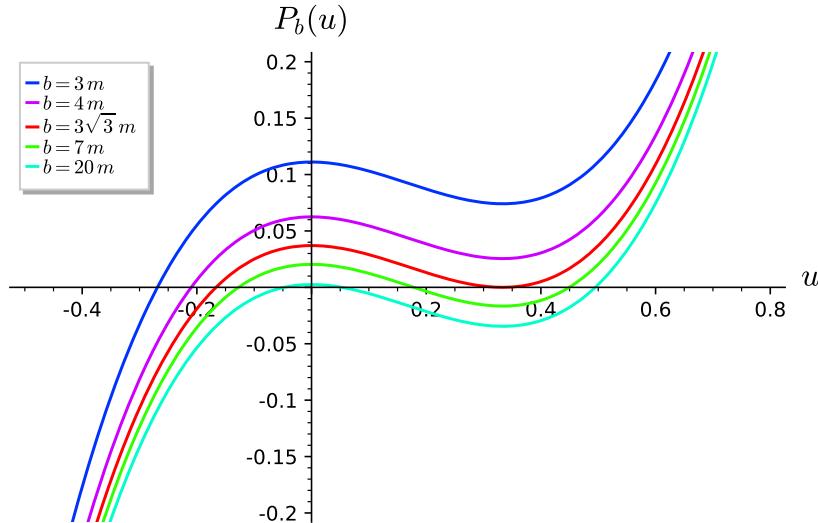


Figure 8.4: Graph of the cubic polynomial $P_b(u) = 2u^3 - u^2 + m^2/b^2$ [Eq. (8.26)]. [Figure produced with the notebook [D.4.9](#)]

- $\Delta = 0 \iff b = b_c$: P_b has a double zero: $u_a = u_p = 1/3$ and a negative zero: $u_n = -1/6$ (cf. red curve in Fig. 8.4); only the first one is physical and correspond to $r = 3m$.
- $\Delta < 0 \iff b < b_c$: 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

$$x^3 - \frac{1}{12}x + \frac{m^2}{2b^2} - \frac{1}{108} = 0. \quad (8.29)$$

We shall discuss separately the cases $b > b_c$ and $b < b_c$.

Zeros of P_b for $b \geq b_c$

For $b > b_c$, 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

$$u_k = \frac{1}{3} \cos \left[\frac{1}{3} \arccos \left(1 - 2 \frac{b_c^2}{b^2} \right) + \frac{2k\pi}{3} \right] + \frac{1}{6}, \quad k \in \{0, 1, 2\}.$$

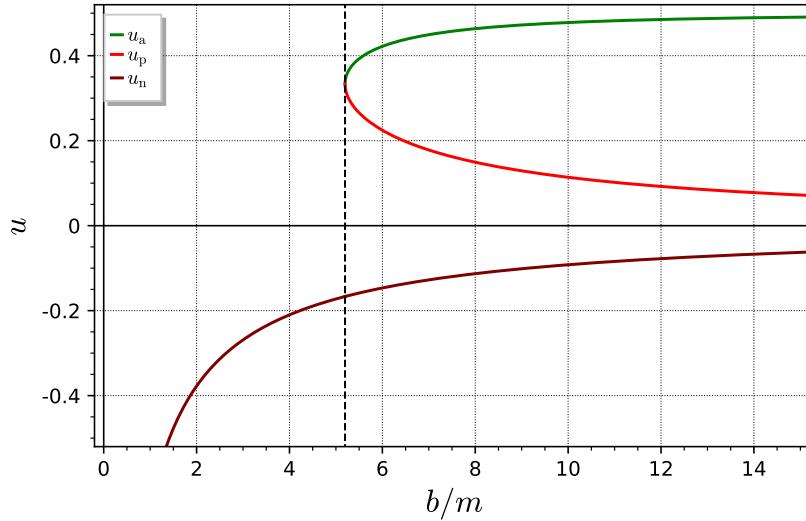


Figure 8.5: Real zeros u_n , u_p and u_a of the cubic polynomial $P_b(u) = 2u^3 - u^2 + m^2/b^2$ as functions of b . The vertical dashed line marks the critical value $b = b_c \simeq 5.196 m$. [Figure produced with the notebook D.4.9]

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

$$u_n = \frac{1}{3} \cos \left[\frac{2}{3} \arcsin \left(\frac{b_c}{b} \right) + \frac{2\pi}{3} \right] + \frac{1}{6} \quad (b \geq b_c) \quad (8.30a)$$

$$u_p = \frac{1}{3} \cos \left[\frac{2}{3} \arcsin \left(\frac{b_c}{b} \right) + \frac{4\pi}{3} \right] + \frac{1}{6} \quad (b \geq b_c) \quad (8.30b)$$

$$u_a = \frac{1}{3} \cos \left[\frac{2}{3} \arcsin \left(\frac{b_c}{b} \right) \right] + \frac{1}{6} \quad (b \geq b_c). \quad (8.30c)$$

These zeros are plotted in terms of b on Fig. 8.5. Note the ordering

$$-\frac{1}{6} \leq u_n < 0 < u_p \leq \frac{1}{3} \leq u_a < \frac{1}{2}, \quad (8.31)$$

with the inequalities \leq being saturated for $b = b_c$. Note also that

$$\lim_{b \rightarrow +\infty} u_n = 0, \quad \lim_{b \rightarrow +\infty} u_p = 0, \quad \text{and} \quad \lim_{b \rightarrow +\infty} u_a = \frac{1}{2}. \quad (8.32)$$

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

$$\begin{aligned} P_b(u) = 0 &\iff 2u^3 - u^2 - 2u_p^3 + u_p^2 = 0 \\ &\iff (u - u_p) [2(u^2 + u_p u + u_p^2) - (u + u_p)] = 0 \\ &\iff (u - u_p) [2u^2 + (2u_p - 1)u + u_p(2u_p - 1)] = 0. \end{aligned}$$

The two zeros different from u_p of this equation, namely u_n and u_a , must then obey

$$2u^2 + (2u_p - 1)u + u_p(2u_p - 1) = 0.$$

Solving this quadratic equation leads to the sought expressions:

$$u_n = \frac{1}{4} \left(1 - 2u_p - \sqrt{(1 - 2u_p)(1 + 6u_p)} \right) \quad (8.33a)$$

$$u_a = \frac{1}{4} \left(1 - 2u_p + \sqrt{(1 - 2u_p)(1 + 6u_p)} \right). \quad (8.33b)$$

Real zero of P_b for $b < b_c$

As mentioned above, for $b < b_c$, 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

$$u_n = \frac{1}{6} \left[1 - \left(\frac{b_c}{b} - \sqrt{\frac{b_c^2}{b^2} - 1} \right)^{2/3} - \left(\frac{b_c}{b} - \sqrt{\frac{b_c^2}{b^2} - 1} \right)^{-2/3} \right] \quad (b < b_c). \quad (8.34)$$

u_n is plotted as a function of b in the left part of Fig. 8.5. Note that

$$u_n < -\frac{1}{6}, \quad \lim_{b \rightarrow b_c^-} u_n = -\frac{1}{6} \quad \text{and} \quad \lim_{b \rightarrow 0} u_n = -\infty. \quad (8.35)$$

The last limit results from the expansion of Eq. (8.34), which yields $u_n \sim -(2b_c/b)^{2/3}/6$ when $b \rightarrow 0$. Observe also from Fig. 8.5 the smooth transition with the value of u_n for $b \geq b_c$, which is given by Eq. (8.30a).

Historical note : As mentioned in the historical note on p. 229, the equations of geodesic motion in Schwarzschild metric have been first derived by Karl Schwarzschild in January 1916 [448] and Johannes Droste in May 1916 [174], 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 [196]. In particular, Flamm derived³ 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. [180].

8.3.2 Critical null geodesics

For $b = b_c = 3\sqrt{3}m$, 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) is equivalent to

$$\frac{d\varphi}{du} = \pm \frac{1}{|u - \frac{1}{3}| \sqrt{2u + \frac{1}{3}}}. \quad (8.36)$$

³The link between Flamm’s notations and ours is $\alpha = 2m$, $R = r$, $x = u/m$ and $\Delta = b$.

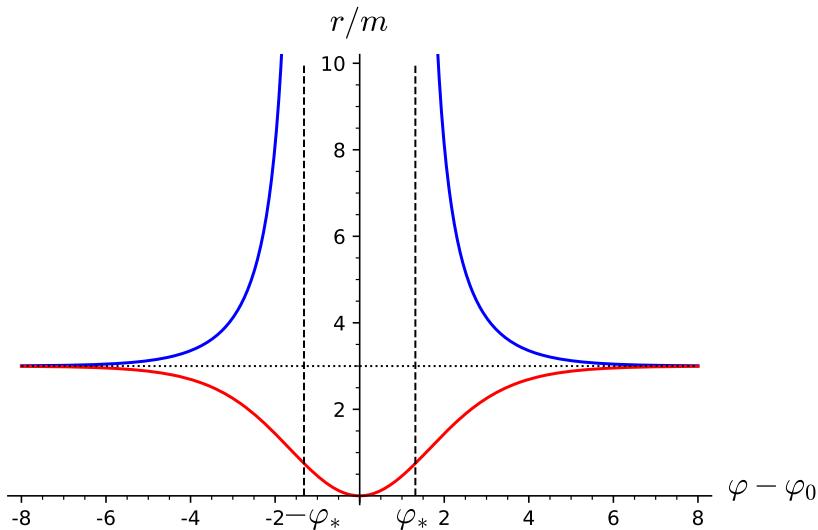


Figure 8.6: r as a function of φ along a null geodesic with an impact parameter equal to the critical one: $b = b_c = 3\sqrt{3} m$; the blue curve is for a critical null geodesic with $r > 3m$ [Eq. (8.40)], while the red one regards $r < 3m$ [Eq. (8.42)]. [Figure produced with the notebook [D.4.10](#)]

This equation can be easily integrated by noticing that

$$\frac{1}{(1-x)\sqrt{x}} = \begin{cases} 2 \frac{d}{dx} (\operatorname{artanh} \sqrt{x}) & \text{for } x \in (0, 1) \\ 2 \frac{d}{dx} (\operatorname{arcoth} \sqrt{x}) & \text{for } x \in (1, +\infty). \end{cases} \quad (8.37)$$

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

$$\varphi = \pm 2 \operatorname{artanh} \sqrt{2u + \frac{1}{3}} + \varphi_0, \quad (8.38)$$

where φ_0 is some integration constant. This relation is easily inverted to

$$u = \frac{1}{2} \tanh^2 \left(\frac{\varphi - \varphi_0}{2} \right) - \frac{1}{6}. \quad (8.39)$$

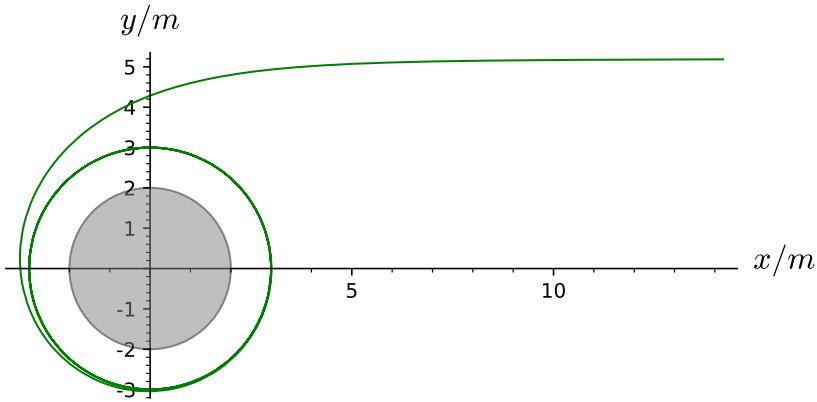


Figure 8.7: Trace in the equatorial plane spanned by the coordinates $x := r \cos \varphi$, $y := r \sin \varphi$ of an external critical null geodesic ($b = 3\sqrt{3} m \simeq 5.196 m$ and $r > 3m$) with $\varphi_\infty = 0$. It obeys Eq. (8.40) with $\varphi_0 = -\varphi_*$ and $\varphi \in (0, +\infty)$ (right branch of the blue curve in Fig. 8.6). [Figure produced with the notebook D.4.10]

Moving back to $r = m/u$, we obtain the equation of the external critical null geodesic in polar form:

$$r = \frac{2m}{\tanh^2\left(\frac{\varphi-\varphi_0}{2}\right) - \frac{1}{3}}. \quad (8.40)$$

The constant φ_0 can be related to the asymptotic value φ_∞ of φ when $r \rightarrow +\infty$ by setting $u = 0$ in Eq. (8.39); we get, using the identity $\text{artanh } x = 1/2 \ln[(1+x)/(1-x)]$,

$$\varphi_0 = \varphi_\infty \pm \varphi_*, \quad \text{with} \quad \varphi_* := \ln\left(\frac{\sqrt{3}+1}{\sqrt{3}-1}\right) \simeq 1.316958. \quad (8.41)$$

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 $\varphi \rightarrow +\infty$, the geodesic rolls up indefinitely onto the photon orbit discussed in Sec. 8.2.3; this behavior 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 $r \sim 3m(1 + 6e^{-\varphi})$ when $\varphi \rightarrow +\infty$.

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 $r \rightarrow +\infty$ 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 $r = r_{\text{em}} > 3m$, $\varphi = \varphi_{\text{em}} > 0$, spiralling outwards and escaping to $r \rightarrow +\infty$ as $\varphi \rightarrow 0$. 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

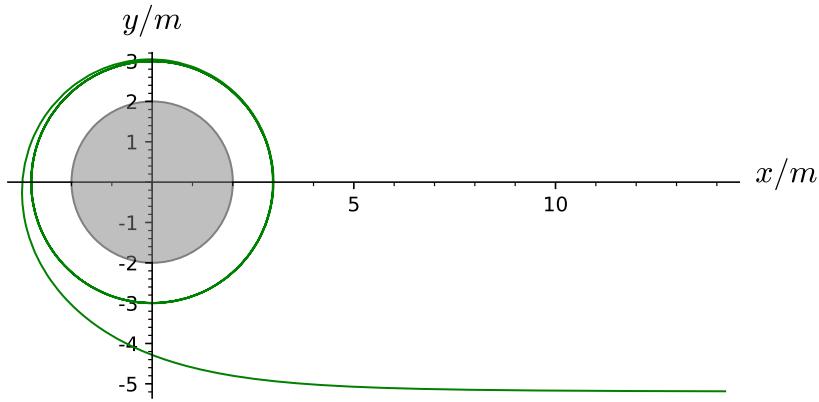


Figure 8.8: Same as Fig. 8.7, but for $\varphi_0 = \varphi_*$ and $\varphi \in (-\infty, 0)$ (left branch of the blue curve in Fig. 8.6). [Figure produced with the notebook D.4.10]

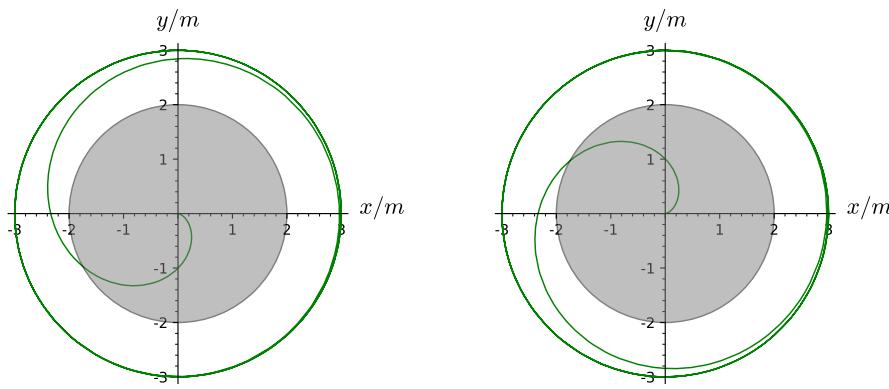


Figure 8.9: Trace in the equatorial plane of two internal critical null geodesics ($b = 3\sqrt{3}m \simeq 5.196m$ and $r < 3m$). They both obey Eq. (8.42) with $\varphi_0 = 0$. The left panel corresponds to $\varphi \in (-\infty, 0)$ (left part of the red curve in Fig. 8.6), while the right one is for $\varphi \in (0, +\infty)$ (right part of the red curve in Fig. 8.6). [Figure produced with the notebook D.4.10]

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 $r \rightarrow +\infty$ and spiralling inwards to the photon orbit when $\varphi \rightarrow -\infty$ and another one with $L > 0$ arising from $r = r_{\text{em}} > 3m$, $\varphi = \varphi_{\text{em}} < 0$, spiralling outwards and escaping to $r \rightarrow +\infty$ as $\varphi \rightarrow 0$.

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

$$\boxed{r = \frac{2m}{\coth^2\left(\frac{\varphi-\varphi_0}{2}\right) - \frac{1}{3}}}. \quad (8.42)$$

Again, φ_0 is an integration constant. But this time, it cannot be determined by the value of φ when $r \rightarrow +\infty$ 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 $\varphi \rightarrow -\infty$, 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 $r \sim 3m(1 - 6e^\varphi)$ when $\varphi \rightarrow -\infty$. The curve plotted in Fig. 8.9 actually represents the trace in the equatorial plane of two distinct geodesics: (i) a geodesic with $L > 0$ arising from $r = r_{\text{em}} < 3m$, $\varphi = \varphi_{\text{em}} < 0$ and inspiralling to the central singularity as $\varphi \rightarrow 0$ and (ii) a geodesic with $L < 0$ arising from $r = r_{\text{em}} \in (2m, 3m)$, $\varphi = \varphi_{\text{em}} < 0$ and spiralling outward to the photon orbit as $\varphi \rightarrow -\infty$. Note that for (ii), the emission point must fulfill $r_{\text{em}} > 2m$, 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 $\varphi \rightarrow +\infty$, 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 $r = r_{\text{em}} < 3m$, $\varphi = \varphi_{\text{em}} > 0$ and inspiralling to the central singularity as $\varphi \rightarrow 0$ and (ii) a geodesic with $L > 0$ arising from $r = r_{\text{em}} \in (2m, 3m)$, $\varphi = \varphi_{\text{em}} > 0$ and spiralling outward to the photon orbit as $\varphi \rightarrow +\infty$.

Historical note : The photon circular orbit at $r = 3m$ has been exhibited by David Hilbert in December 1916 [276, 277]. 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 ingoing null geodesics with an impact parameter $b > b_c$ are deflected (possibly looping around the photon orbit if b is close to b_c) and escape to infinity, while those with $b < b_c$ cross the photon orbit and terminate on “the circle $r = 2m$ ”. Hilbert claimed that null geodesics stop there, because their (coordinate!) velocity vanishes⁴. 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 note on p. 229). 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 [339].

⁴The coordinate velocity for ingoing radial null geodesics is given by Eq. (8.6): $dr/dt = -1 + 2m/r$, which clearly vanishes at $r = 2m$.

8.3.3 Null geodesics with $b > b_c$ and $r > 3m$

For $b \neq b_c$, 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

$$\frac{d\varphi}{du} = \epsilon_L \epsilon_{in} \frac{1}{\sqrt{P_b(u)}}, \quad (8.43)$$

where $\epsilon_L = \pm 1$ is the sign of the conserved angular momentum L [cf. Eq. (8.13)] and $\epsilon_{in} = \pm 1$ is defined by

$$\epsilon_{in} := \text{sgn} \left(\frac{du}{d\tilde{\lambda}} \right), \quad (8.44)$$

i.e. $\epsilon_{in} = +1$ if u increases towards the future along \mathcal{L} , or equivalently if r decreases towards the future along \mathcal{L} (*inward* motion, hence the index “in”), and $\epsilon_{in} = -1$ otherwise. Let us recall that by virtue of the equation of motion (8.10), ϵ_L gives the sign of $d\varphi/d\tilde{\lambda}$, so that the sign of $d\varphi/du = d\varphi/d\tilde{\lambda} \times (du/d\tilde{\lambda})^{-1}$ is $\epsilon_L \epsilon_{in}$, which justifies the factor $\epsilon_L \epsilon_{in}$ 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 $b > b_c$ 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 $u = u_p$, and obeys $u \rightarrow 0$ ($r \rightarrow +\infty$) for both $\tilde{\lambda} \rightarrow -\infty$ and $\tilde{\lambda} \rightarrow +\infty$. The general solution to the differential equation (8.43) can be written as

$$\varphi = \varphi_p + \epsilon_L \epsilon_{in} \int_{u_p}^u \frac{d\bar{u}}{\sqrt{P_b(\bar{u})}}, \quad (8.45)$$

where φ_p is the value of φ at the periastron. We shall rewrite it as

$$\boxed{\varphi = \varphi_p - \epsilon_L \epsilon_{in} \Phi_b(u)}, \quad (8.46)$$

where we have introduced the function

$$\boxed{\Phi_b(u) := \int_u^{u_p} \frac{d\bar{u}}{\sqrt{P_b(\bar{u})}} = \int_u^{u_p} \frac{d\bar{u}}{\sqrt{2\bar{u}^3 - \bar{u}^2 + (m/b)^2}}} \quad (b > b_c). \quad (8.47)$$

Since u_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 $u \leq u_p$ (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_n, u_p, u_a) of P_b :

$$\Phi_b(u) = \int_u^{u_p} \frac{d\bar{u}}{\sqrt{2(\bar{u} - u_n)(u_p - \bar{u})(u_a - \bar{u})}}. \quad (8.48)$$

Let us perform the change of variable⁵

$$t := \frac{\bar{u} - u_n}{u_p - u_n} \iff \bar{u} = (u_p - u_n)t + u_n.$$

⁵It should be clear that the variable t introduced here has nothing to do with the Schwarzschild-Droste coordinate t .

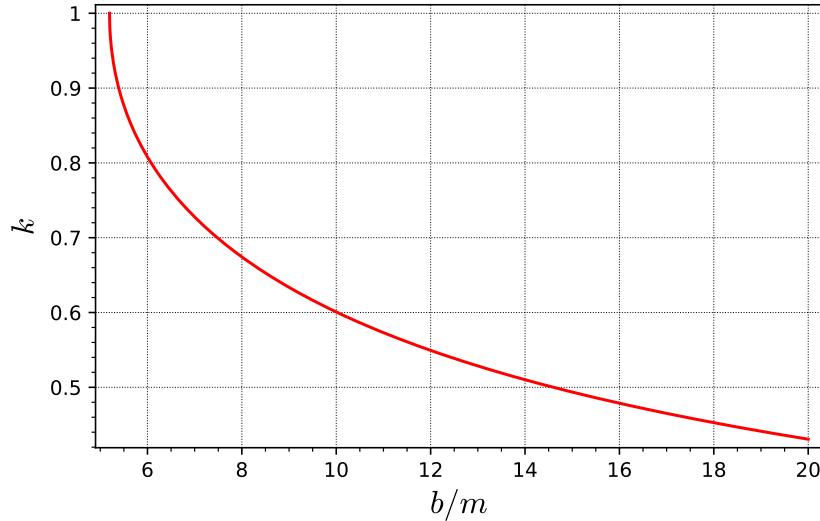


Figure 8.10: Modulus k of the elliptic integrals F and K that are involved in expression (8.53) for $\Phi_b(u)$. [Figure produced with the notebook [D.4.9](#)]

We get

$$\Phi_b(u) = \frac{1}{\sqrt{2(u_a - u_n)}} \int_{\frac{u-u_n}{u_p-u_n}}^1 \frac{dt}{\sqrt{t(1-t)(1-k^2t)}}, \quad (8.49)$$

where k is the following constant:

$$k := \sqrt{\frac{u_p - u_n}{u_a - u_n}} = \frac{\sqrt{2}}{\sqrt{\sqrt{3} \cot\left(\frac{2}{3} \arcsin\left(\frac{b_c}{b}\right)\right) + 1}} \quad (b > b_c), \quad (8.50)$$

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 $0 < u \leq u_p$ implies $0 < t \leq 1$, so that $0 < \vartheta \leq \pi/2$. Since $dt = 2 \sin \vartheta \cos \vartheta d\vartheta = 2\sqrt{t(1-t)} d\vartheta$, we arrive immediately at

$$\Phi_b(u) = \frac{\sqrt{2}}{\sqrt{u_a - u_n}} \int_{\phi_b(u)}^{\pi/2} \frac{d\vartheta}{\sqrt{1 - k^2 \sin^2 \vartheta}}, \quad (8.51)$$

where⁶

$$\phi_b(u) := \arcsin \sqrt{\frac{u - u_n}{u_p - u_n}} \quad (b > b_c). \quad (8.52)$$

⁶Do not confuse the letter ϕ with the symbol used for the coordinate φ .

By splitting the integral according to $\int_{\phi}^{\pi/2} = \int_0^{\pi/2} - \int_0^{\phi}$, we rewrite (8.51) as

$$\boxed{\Phi_b(u) = \frac{\sqrt{2}}{\sqrt{u_a - u_n}} [K(k) - F(\phi_b(u), k)]} \quad (b > b_c), \quad (8.53)$$

where $F(\phi, k)$ is the *incomplete elliptic integral of the first kind* [78, 231, 3]:

$$\boxed{F(\phi, k) := \int_0^{\phi} \frac{d\vartheta}{\sqrt{1 - k^2 \sin^2 \vartheta}}} \quad (8.54)$$

and $K(k)$ is the *complete elliptic integral of the first kind*:

$$\boxed{K(k) := \int_0^{\frac{\pi}{2}} \frac{d\vartheta}{\sqrt{1 - k^2 \sin^2 \vartheta}} = F\left(\frac{\pi}{2}, k\right).} \quad (8.55)$$

The notation $F(\phi, k)$ is the most common one in the literature [78, 231], but one may encounter as well $F(\phi|m)$ for $F(\phi, k)$ with $m = k^2$ [3]. 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

$$0 < k < 1, \quad (8.56)$$

with

$$\lim_{b \rightarrow b_c^+} k = 1 \quad \text{and} \quad \lim_{b \rightarrow +\infty} k = 0. \quad (8.57)$$

Given the range $(0, u_p)$ of u , we deduce from expression (8.52) that

$$0 < \arcsin \sqrt{\frac{u_n}{u_n - u_p}} < \phi_b(u) \leq \frac{\pi}{2}, \quad (8.58)$$

with $\phi_b(u_p) = \pi/2$.

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 $\Phi_b(u) = 0 \iff u = u_p$. Note also that the closer b is from $b_c \simeq 5.1961452$, 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 kind⁷:

$$\tan \psi \tan \phi = \frac{1}{\sqrt{1 - k^2}} \implies F(\psi, k) = K(k) - F(\phi, k). \quad (8.59)$$

Then by setting

$$\psi_b(u) := \arcsin \left(\frac{1}{k} \sqrt{\frac{u_p - u}{u_a - u}} \right), \quad (8.60)$$

⁷See e.g. Eq. (117.01) of Ref. [78] with $k' = \sqrt{1 - k^2}$ or Eq. (17.4.13) of Ref. [3] with $\sin \alpha = k$.

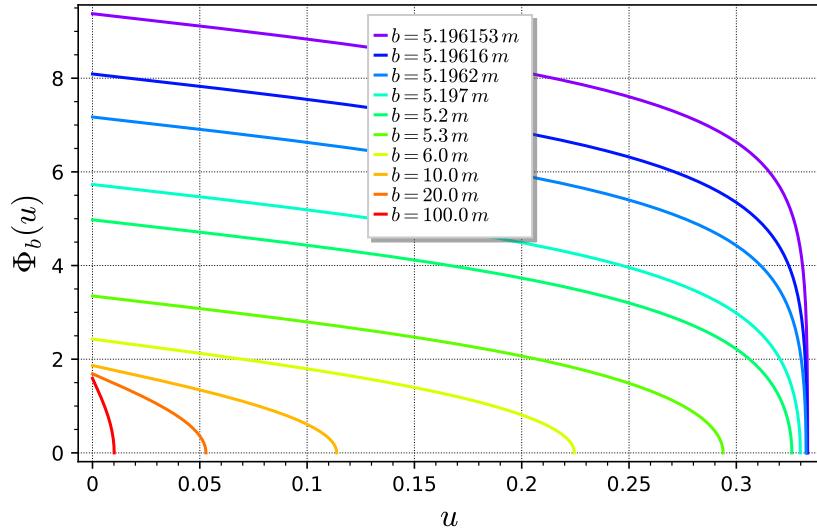


Figure 8.11: Function $\Phi_b(u)$ defined by Eq. (8.47) and evaluated via the elliptic integral expression (8.52)-(8.53), for selected values of $b > b_c$. For each value of b , the range of u is $(0, u_p]$, with the inverse periastron radius $u_p = u_p(b)$ given by Eq. (8.30b). [Figure produced with the notebook D.4.11]

we can rewrite Eq. (8.53) as

$$\Phi_b(u) = \frac{\sqrt{2}}{\sqrt{u_a - u_n}} F(\psi_b(u), k). \quad (8.61)$$

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 $b \rightarrow b_c$, 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** $\text{sn}(x, k)$, defined by $\sin \phi = \text{sn}(F(\phi, k), k)$. We deduce then from Eqs. (8.60)-(8.61) that

$$\frac{u_p - u}{u_a - u} = k^2 \text{sn}\left(\frac{\sqrt{u_a - u_n}}{\sqrt{2}} \Phi_b(u), k\right). \quad (8.62)$$

We refer the reader to Refs. [79, 373] 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 $u = u_a$, where u_a is the function of b given by Eq. (8.30c). In other words, i.e. one has $u \geq u_a$ along the geodesic.

The treatment is similar to that of Sec. 8.3.3, but with u_a playing the role of u_p .

To be detailed later...

8.3.5 Null geodesics with $b < b_c$

A null geodesic \mathcal{L} with an impact parameter $b < b_c$ 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 $u_n < 0$, 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 monotonic function along \mathcal{L} . We shall then distinguish two cases:

- \mathcal{L} is *ingoing*: r decreases all along \mathcal{L} , from $+\infty$ to 0 as $\tilde{\lambda}$ varies from $-\infty$ to some value $\tilde{\lambda}_0$ where \mathcal{L} hits the curvature singularity at $r = 0$:

$$\lim_{\tilde{\lambda} \rightarrow -\infty} r = +\infty \text{ and } \lim_{\tilde{\lambda} \rightarrow \tilde{\lambda}_0} r = 0 \iff \lim_{\tilde{\lambda} \rightarrow -\infty} u = 0 \text{ and } \lim_{\tilde{\lambda} \rightarrow \tilde{\lambda}_0} u = +\infty. \quad (8.63)$$

- \mathcal{L} is *outgoing*: r increases all along \mathcal{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 $+\infty$:

$$\lim_{\tilde{\lambda} \rightarrow \tilde{\lambda}_1} r = 2m \text{ and } \lim_{\tilde{\lambda} \rightarrow +\infty} r = +\infty \iff \lim_{\tilde{\lambda} \rightarrow \tilde{\lambda}_1} u = \frac{1}{2} \text{ and } \lim_{\tilde{\lambda} \rightarrow +\infty} u = 0. \quad (8.64)$$

That the range of $\tilde{\lambda}$ for outgoing geodesics is $(\tilde{\lambda}_1, +\infty)$ and not the whole real line can be understood by considering the limit $b \rightarrow 0$. 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 located in a part of the spacetime that is not covered by the Schwarzschild-Droste coordinates considered here.

Since $P_b(u_n) = 0$, we have $m^2/b^2 = -2u_n^3 + u_n^2$, so that we may write

$$P_b(u) = 2(u^3 - u_n^3) - (u^2 - u_n^2) = (u - u_n) [2(u^2 + uu_n + u_n^2) - (u + u_n)].$$

After a slight rearrangement of the term in square brackets, we arrive at

$$P_b(u) = 2(u - u_n) [(u - u_0)^2 + (u_* - u_n)^2 - (u_0 - u_n)^2], \quad (8.65)$$

with

$$u_0 := \frac{1}{4} - \frac{u_n}{2} \quad \text{and} \quad u_* := \sqrt{u_n(3u_n - 1)} + u_n. \quad (8.66)$$

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 $b < b_c$. Furthermore, according to the limits (8.35), we have

$$\lim_{b \rightarrow 0} u_* = +\infty \quad \text{and} \quad \lim_{b \rightarrow b_c^-} u_* = \frac{1}{3}. \quad (8.67)$$

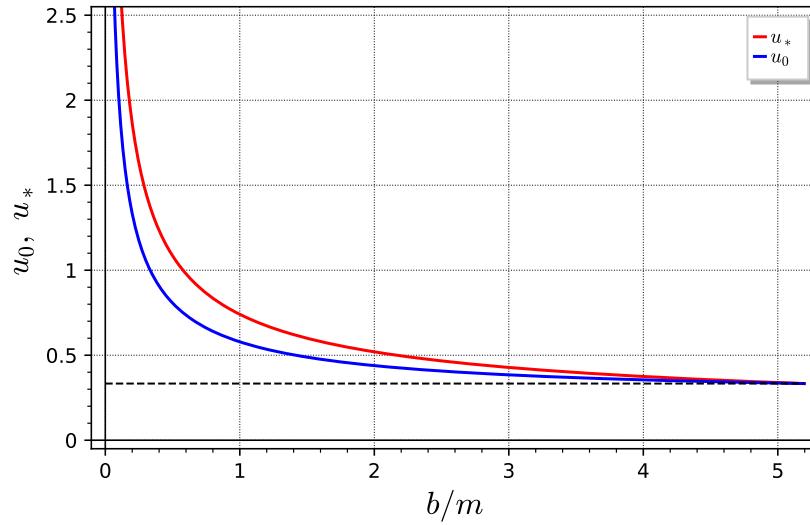


Figure 8.12: Parameters u_0 and u_* , defined by Eq. (8.66), as functions of b . The horizontal dashed line marks $u = 1/3$. [Figure produced with the notebook D.4.12]

The general solution to the differential equation (8.43) can be written as

$$\boxed{\varphi = \varphi_* - \epsilon_L \epsilon_{\text{in}} \Phi_b(u)}, \quad (8.68)$$

where

$$\boxed{\Phi_b(u) := \int_u^{u_*} \frac{d\bar{u}}{\sqrt{P_b(\bar{u})}} = \int_u^{u_*} \frac{d\bar{u}}{\sqrt{2\bar{u}^3 - \bar{u}^2 + (m/b)^2}} \quad (b < b_c)} \quad (8.69)$$

and φ_* is the value of φ at $u = u_*$. Since

$$\lim_{b \rightarrow b_c^-} u_* = \frac{1}{3} = \lim_{b \rightarrow b_c^+} u_p,$$

there is a kind of continuity of (8.69) with the definition (8.47) of $\Phi_b(u)$ for $b > b_c$, despite the fact that $\Phi_b(u)$ is not defined for $b = b_c$ (more precisely, as we shall see later, $\lim_{b \rightarrow b_c^-} \Phi_b(u) = +\infty$ for $u < u_*$ and $\lim_{b \rightarrow b_c^+} \Phi_b(u) = +\infty$ for $u < u_p$).

To evaluate $\Phi_b(u)$, we shall use expression (8.65) for $P_b(u)$:

$$\Phi_b(u) = \frac{1}{\sqrt{2}} \int_u^{u_*} \frac{d\bar{u}}{\sqrt{(\bar{u} - u_n)[(\bar{u} - u_0)^2 + (u_* - u_n)^2 - (u_0 - u_n)^2]}}.$$

Performing the change of variable

$$t := \frac{u_* - \bar{u}}{u_* + \bar{u} - 2u_n} \iff \bar{u} = (u_* - u_n) \frac{1-t}{1+t} + u_n \quad (8.70)$$

yields

$$\Phi_b(u) = \int_0^{\frac{u_* - u}{u_* + u - 2u_n}} \frac{dt}{\sqrt{1-t^2} \sqrt{(u_* - u_n)(1+t^2) - (u_0 - u_n)(1-t^2)}}. \quad (8.71)$$

Given the range $(0, +\infty)$ for u , the range of t is

$$-1 < t \leq \frac{u_*}{u_* - 2u_n} < 1.$$

To proceed, we shall distinguish the cases $u \leq u_*$ and $u > u_*$. The first case corresponds to $0 \leq t < \frac{u_*}{u_* - 2u_n} < 1$ and we perform the change of variable $t = \cos \vartheta$ with $\vartheta \in (0, \pi/2)$ in the integral (8.71), yielding

$$\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_n)}} \int_{\phi_b(u)}^{\pi/2} \frac{d\vartheta}{\sqrt{1 - k^2 \sin^2 \vartheta}},$$

with

$$\boxed{\phi_b(u) := \arccos \left(\frac{|u_* - u|}{u_* + u - 2u_n} \right)} \quad (b < b_c). \quad (8.72)$$

and⁸

$$\boxed{k := \sqrt{\frac{u_* - 5u_n/2 + 1/4}{2(u_* - u_n)}}} \quad (b < b_c). \quad (8.73)$$

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

$$\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_n)}} \left[\int_0^{\pi/2} \frac{d\vartheta}{\sqrt{1 - k^2 \sin^2 \vartheta}} - \int_0^{\phi_b(u)} \frac{d\vartheta}{\sqrt{1 - k^2 \sin^2 \vartheta}} \right].$$

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

$$\boxed{\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_n)}} [K(k) - F(\phi_b(u), k)]} \quad (u \leq u_*, b < b_c). \quad (8.74)$$

As for the case $b > b_c$, 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:

$$\underbrace{\frac{1}{2}\sqrt{2 + \sqrt{3}}}_{\simeq 0.96592} < k < 1. \quad (8.75)$$

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

$$\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_n)}} \int_{\pi/2}^{\phi_b(u)} \frac{d\vartheta}{\sqrt{1 - k^2 \sin^2 \vartheta}},$$

⁸To get expression (8.73), we have used Eq. (8.66) to get rid of u_0 .

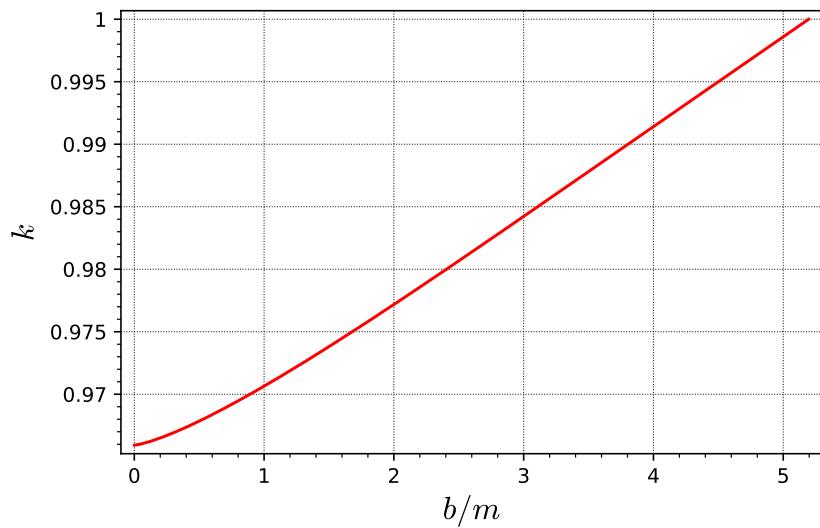


Figure 8.13: Modulus k of the elliptic integrals F and K that are involved in expressions (8.74) and (8.76) for $\Phi_b(u)$. [Figure produced with the notebook D.4.12]

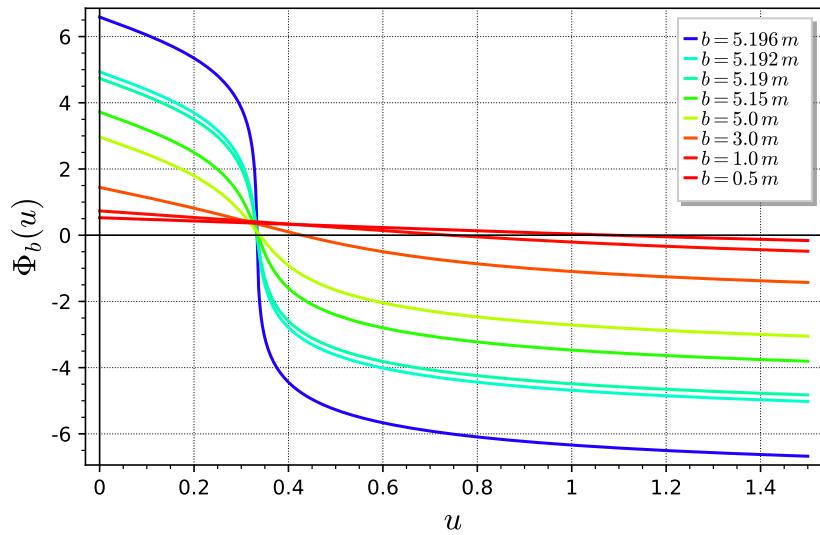


Figure 8.14: Function $\Phi_b(u)$ defined by Eq. (8.69) and evaluated via the elliptic integral expressions (8.74) and (8.76), for selected values of $b < b_c$. [Figure produced with the notebook D.4.12]

where $\phi_b(u)$ and k are given by Eqs. (8.72) and (8.73). We conclude that

$$\boxed{\Phi_b(u) = \frac{1}{\sqrt{2(u_* - u_n)}} [F(\phi_b(u), k) - K(k)]} \quad (u > u_*, b < b_c). \quad (8.76)$$

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 $b > b_c$ (Sec. 8.3.3), the closer b is from $b_c \simeq 5.1961452$, the larger the amplitude of $\Phi_b(u)$.

8.3.6 Deflection angle and winding number

Let us consider a null geodesic \mathcal{L} with $b > b_c$ arising from $r \rightarrow +\infty$, i.e. belonging to the family studied in Sec. 8.3.3. \mathcal{L} reaches some periastron and departs to $r \rightarrow +\infty$. The total change of φ along the geodesic history is

$$\Delta\varphi = \varphi_\infty - \varphi_{-\infty}, \quad (8.77)$$

with $\varphi_\infty := \lim_{\tilde{\lambda} \rightarrow +\infty} \varphi$ and $\varphi_{-\infty} := \lim_{\tilde{\lambda} \rightarrow -\infty} \varphi$. Given that $\lim_{\tilde{\lambda} \rightarrow \pm\infty} u = 0$, Eq. (8.46) with respectively $\epsilon_{\text{in}} = +1$ and $\epsilon_{\text{in}} = -1$ leads to

$$\varphi_{-\infty} = \varphi_p - \epsilon_L \Phi_b(0) \quad \text{and} \quad \varphi_\infty = \varphi_p + \epsilon_L \Phi_b(0). \quad (8.78)$$

Hence

$$\Delta\varphi = 2\epsilon_L \Phi_b(0). \quad (8.79)$$

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

$$\Delta\varphi =: \overline{\Delta\varphi} + 2\pi n \quad \text{with} \quad \begin{cases} \overline{\Delta\varphi} \in [0, 2\pi) & \text{if } L > 0 \\ \overline{\Delta\varphi} \in (-2\pi, 0] & \text{if } L < 0. \end{cases} \quad (8.80)$$

Note that $n \geq 0$ for $L > 0$ and $n \leq 0$ for $L < 0$. We then define the **deflection angle** Θ by

$$\Theta := \overline{\Delta\varphi} - \epsilon_L \pi. \quad (8.81)$$

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

$$-\pi \leq \Theta \leq \pi. \quad (8.82)$$

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_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

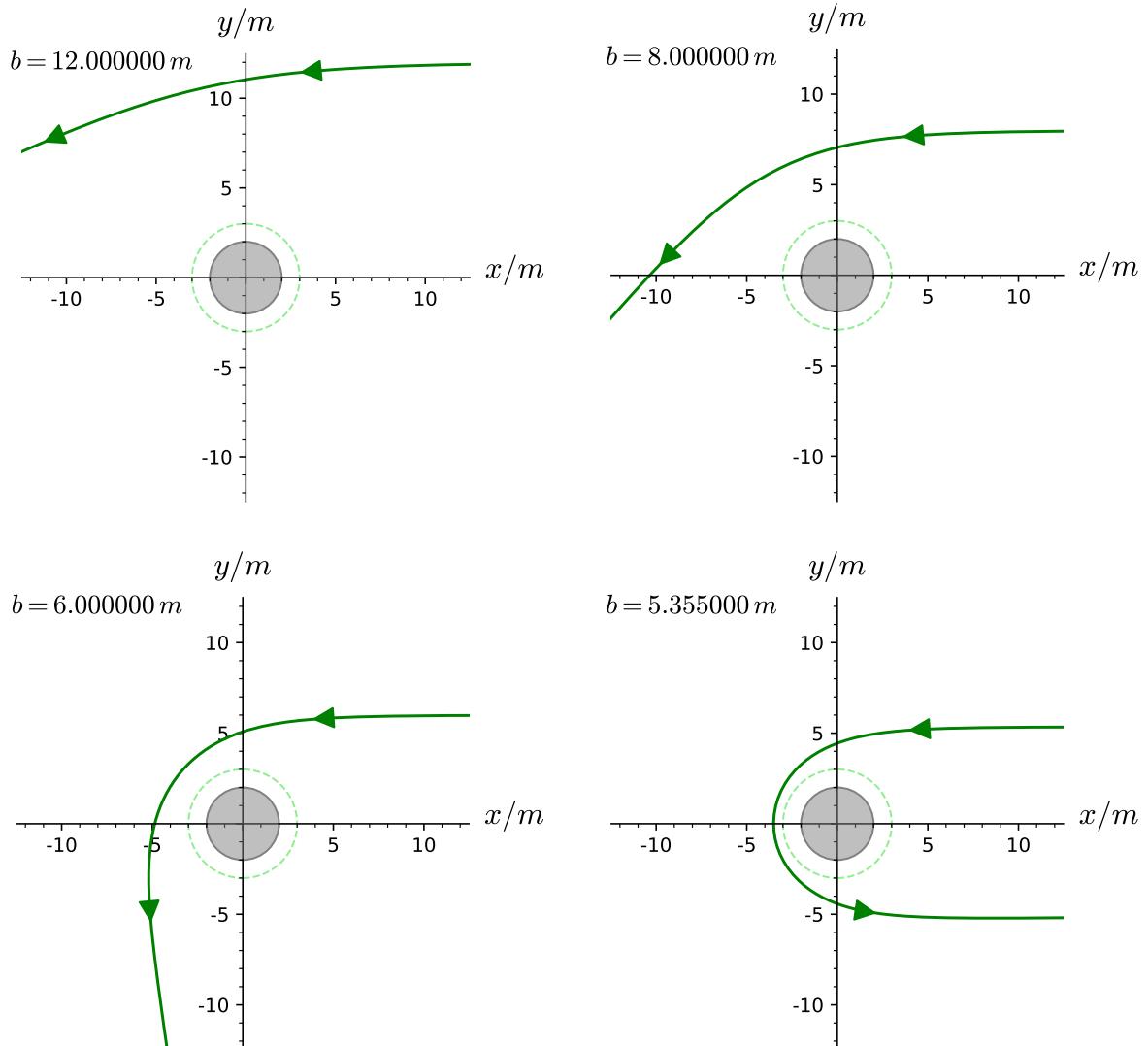


Figure 8.15: Null geodesics in the plane $\theta = \pi/2$ of Schwarzschild spacetime, plotted in terms of the coordinates $(x, y) := (r \cos \varphi, r \sin \varphi)$. All geodesics arise from $x \gg m$ with trajectories initially parallel to the x -axis; they differ by the value of the impact parameter 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.4.8]

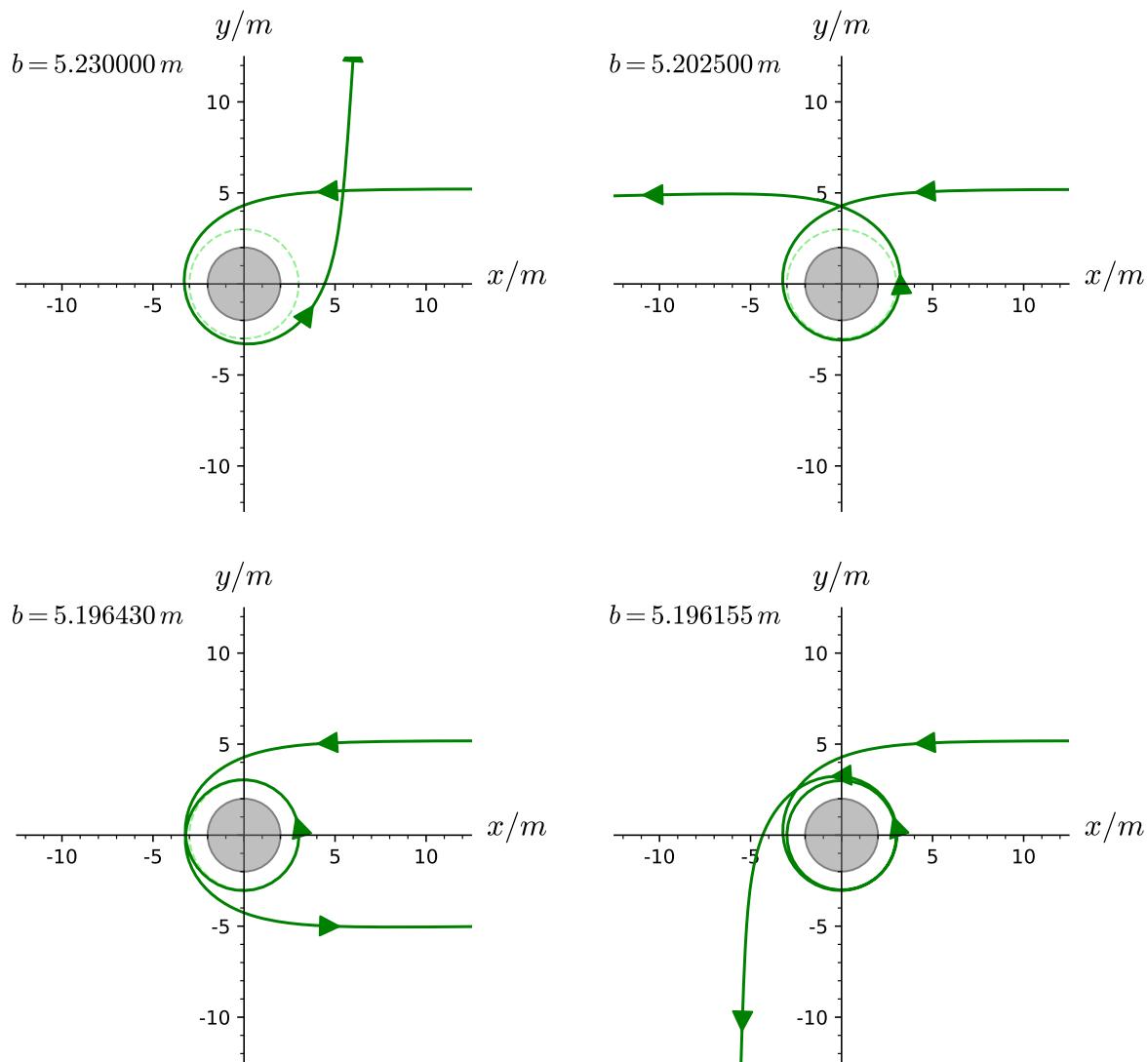


Figure 8.16: Same as Fig. 8.15 but for values of b closer to $b_c \simeq 5.196152 \text{ m}$. [Figure produced with the notebook D.4.8]

$b = 12m$ (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 $b = 5.355\text{ m}$ (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 $b_c \simeq 5.196152\text{ m}$ [Eq. (8.21)], the null geodesic starts to wind around the black hole before escaping to infinity (Fig. 8.16). For $b = 5.230\text{ m}$ (upper left panel of Fig. 8.16), the winding number is $n = 1$ and the deflection angle is $\Theta \simeq -\pi/2$. For $b = 5.2025\text{ m}$ (upper right panel of Fig. 8.16), one has $n = 1$ and $\Theta = 0$. For $b = 5.19643\text{ m}$ (lower left panel of Fig. 8.16), one has $n = 1$ and $\Theta \simeq \pi$ and for $b = 5.196155\text{ m}$ (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_c without investigating the properties of elliptic integrals. Indeed, by combining Eqs. (8.79) and (8.48), we get

$$\Delta\varphi = \epsilon_L \sqrt{2} \int_0^{u_p} \frac{du}{\sqrt{(u_p - u)(u_a - u)(u - u_n)}}. \quad (8.83)$$

Given the ordering (8.31), each of the three factors under the square root is positive on the integration range $(0, u_p)$. For $b \neq b_c$, one has $u_a \neq u_p$ (cf. Fig. 8.5) and the only diverging term in the integrand of (8.83) is $1/\sqrt{u_p - u}$, which diverges at the integral boundary $u = u_p$. However, the integral

$$\int_0^{u_p} \frac{du}{\sqrt{u_p - u}}$$

is finite, being equal to $2\sqrt{u_p}$, so that $\Delta\varphi$ remains finite. When $b \rightarrow b_c$, $u_a \rightarrow u_p$ and the integral (8.83) has a behavior similar to

$$\int_0^{u_p} \frac{du}{\sqrt{(u_p - u)^2}} = \int_0^{u_p} \frac{du}{u_p - u}.$$

Since the latter is a diverging integral, we conclude that

$$\boxed{\Delta\varphi \rightarrow \pm\infty \quad \text{when} \quad b \rightarrow b_c}. \quad (8.84)$$

We recover the behavior observed for $b = b_c$ 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)].

8.4 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 \mathcal{L} with $b > b_c$ and an event of (finite) affine parameter $\tilde{\lambda} = \tilde{\lambda}_{\text{em}}$ on \mathcal{L} , which we shall call the *emission point*. We are interested in null geodesics which reach the asymptotic flat region $r \rightarrow +\infty$ for $\tilde{\lambda} \rightarrow +\infty$. This implies that the emission point is located outside the photon sphere, i.e. obeys

$$r_{\text{em}} := r(\tilde{\lambda}_{\text{em}}) > 3m. \quad (8.85)$$

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 latter (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 $\tilde{\lambda} \rightarrow +\infty$ (the asymptotic direction) to its value φ_{em} at $\tilde{\lambda} = \tilde{\lambda}_{\text{em}}$.

If the emission point is past \mathcal{L} 's periastron, i.e. if $\tilde{\lambda}_{\text{em}} > \tilde{\lambda}_p$, φ_{em} is related to φ_p by Eq. (8.46) with $\epsilon_{\text{in}} = -1$:

$$\varphi_{\text{em}} = \varphi_p + \epsilon_L \Phi_b(u_{\text{em}}), \quad (8.86)$$

where $u_{\text{em}} := m/r_{\text{em}}$. 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

$$\varphi_\infty = \varphi_{\text{em}} + \epsilon_L [\Phi_b(0) - \Phi_b(u_{\text{em}})]. \quad (8.87)$$

Substituting expression (8.53) for Φ_b yields

$$\varphi_\infty = \varphi_{\text{em}} + \epsilon_L \frac{\sqrt{2}}{\sqrt{u_a - u_n}} [F(\phi_b(u_{\text{em}}), k) - F(\phi_b(0), k)]$$

$(b > b_c)$
 $(\tilde{\lambda}_{\text{em}} > \tilde{\lambda}_p)$

$$(8.88)$$

where, according to Eq. (8.52),

$$\phi_b(u_{\text{em}}) = \arcsin \sqrt{\frac{u_{\text{em}} - u_n}{u_p - u_n}} \quad \text{and} \quad \phi_b(0) = \arcsin \sqrt{\frac{|u_n|}{u_p - u_n}}. \quad (8.89)$$

If the emission point is located prior to the periastron, i.e. if $\tilde{\lambda}_{\text{em}} < \tilde{\lambda}_p$, φ_{em} is related to φ_p by Eq. (8.46) with $\epsilon_{\text{in}} = +1$:

$$\varphi_{\text{em}} = \varphi_p - \epsilon_L \Phi_b(u_{\text{em}}) \quad (8.90)$$

and we get, by combining with Eq. (8.78),

$$\varphi_\infty = \varphi_{\text{em}} + \epsilon_L [\Phi_b(0) + \Phi_b(u_{\text{em}})]. \quad (8.91)$$

Equation (8.53) then yields

$$\varphi_\infty = \varphi_{\text{em}} + \epsilon_L \frac{\sqrt{2}}{\sqrt{u_a - u_n}} [2K(k) - F(\phi_b(u_{\text{em}}), k) - F(\phi_b(0), k)]$$

$(b > b_c)$
 $(\tilde{\lambda}_{\text{em}} < \tilde{\lambda}_p)$

$$(8.92)$$

$\varphi_\infty - \varphi_{\text{em}}$ is plotted as a function of b in Fig. 8.17 for various values of $r_{\text{em}} = m/u_{\text{em}}$, with the dashed curves corresponding to the case $\tilde{\lambda}_{\text{em}} > \tilde{\lambda}_p$. We notice that for $b \gg m$ (cf. the red curve near $b = 12m$), one has either $\varphi_\infty - \varphi_{\text{em}} \simeq 0$ (dashed red curve: $\tilde{\lambda}_{\text{em}} > \tilde{\lambda}_p$) or $\varphi_\infty - \varphi_{\text{em}} \simeq \pi$ (solid red curve: $\tilde{\lambda}_{\text{em}} < \tilde{\lambda}_p$): null geodesics with large impact parameters suffer almost no deflection, as expected. Another striking feature of Fig. 8.17 is the divergence of $\varphi_\infty - \varphi_{\text{em}}$ when b tends to b_c . Let us examine this in details.

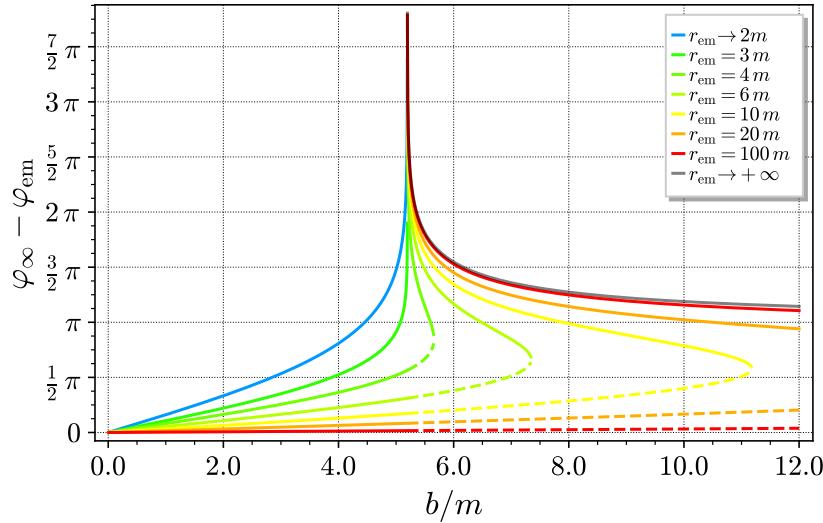


Figure 8.17: Total change $\varphi_\infty - \varphi_{\text{em}}$ in φ from the emission point at $(r, \varphi) = (r_{\text{em}}, \varphi_{\text{em}})$ as a function of the impact parameter b of the null geodesic, assuming $\epsilon_L = +1$. The dashed curves correspond to trajectories with $b > b_c$ that do not pass through the periastron. For each value of $r_{\text{em}} > 3m$, the maximum of b is given by $r_p(b) = r_{\text{em}}$, i.e. by $b^2 = r_{\text{em}}^3 / (r_{\text{em}} - 2m)$ [cf. Eq. (8.18) and Fig. 8.3]. [Figure produced with the notebook D.4.11]

Limit $b \rightarrow b_c^+$

From Eqs. (8.30), we have the following limits:

$$\lim_{b \rightarrow b_c^+} u_n = -\frac{1}{6} \quad \text{and} \quad \lim_{b \rightarrow b_c^+} u_p = \lim_{b \rightarrow b_c^+} u_a = \frac{1}{3}. \quad (8.93)$$

It follows then from expressions (8.89) that

$$\lim_{b \rightarrow b_c^+} \phi_b(u_{\text{em}}) = \arcsin \sqrt{2u_{\text{em}} + \frac{1}{3}} \quad \text{and} \quad \lim_{b \rightarrow b_c^+} \phi_b(0) = \arcsin \left(\frac{1}{\sqrt{3}} \right). \quad (8.94)$$

Besides, $\lim_{b \rightarrow b_c^+} k = 1$ [cf. Eq. (8.57)] and, for any $\phi \in [0, \pi/2)$, the definition (8.54) of F leads to

$$F(\phi, 1) = \int_0^\phi \frac{d\vartheta}{\sqrt{1 - \sin^2 \vartheta}} = \int_0^\phi \frac{d\vartheta}{\cos \vartheta} = \ln \left(\frac{1 + \sin \phi}{\cos \phi} \right). \quad (8.95)$$

The limits (8.94) result then in

$$\lim_{b \rightarrow b_c^+} F(\phi_b(u_{\text{em}}), k) = \ln \left(\frac{\sqrt{3} + \sqrt{6u_{\text{em}} + 1}}{\sqrt{2(1 - 3u_{\text{em}})}} \right) \quad \text{and} \quad \lim_{b \rightarrow b_c^+} F(\phi_b(0), k) = \ln \left(\frac{\sqrt{3} + 1}{\sqrt{2}} \right). \quad (8.96)$$

For $\tilde{\lambda}_{\text{em}} > \tilde{\lambda}_p$, inserting these formulas into (8.88) leads to

$$\varphi_\infty \underset{b \rightarrow b_c^+}{\sim} \varphi_{\text{em}} + \epsilon_L \ln \left(\frac{2 + 3u_{\text{em}} + \sqrt{3(6u_{\text{em}} + 1)}}{(2 + \sqrt{3})(1 - 3u_{\text{em}})} \right) \quad (\tilde{\lambda}_{\text{em}} > \tilde{\lambda}_p). \quad (8.97)$$

Note that the constraint (8.85) is equivalent to $1 - 3u_{\text{em}} > 0$, so that the above formula is well posed.

For the case $\tilde{\lambda}_{\text{em}} < \tilde{\lambda}_p$, φ_∞ is given by Eq. (8.92) and we need to determine the behavior of $K(k)$ to conclude. Actually, when $k \rightarrow 1$, $K(k)$ is diverging with the following behavior⁹:

$$\lim_{k \rightarrow 1} \left[K(k) - \ln \left(\frac{4}{\sqrt{1 - k^2}} \right) \right] = 0. \quad (8.98)$$

To relate $\sqrt{1 - k^2}$ to $b - b_c$, let us introduce the small parameter $\varepsilon > 0$ such that

$$u_p =: \frac{1}{3} - \varepsilon. \quad (8.99)$$

The relation $P_b(u_p) = 0$, once expanded to second order in ε^2 , leads to the following relation between $b - b_c$ and ε :

$$\frac{b - b_c}{m} = \frac{81\sqrt{3}}{2}\varepsilon^2 + O(\varepsilon^3). \quad (8.100)$$

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 ,

$$u_n = -\frac{1}{6} + 2\varepsilon^2 + O(\varepsilon^3) \quad \text{and} \quad u_a = \frac{1}{3} + \varepsilon - 2\varepsilon^2 + O(\varepsilon^3).$$

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

$$\sqrt{1 - k^2} \underset{\varepsilon \rightarrow 0}{\sim} \sqrt{1 - (1 - 4\varepsilon)} \underset{\varepsilon \rightarrow 0}{\sim} 2\sqrt{\varepsilon}.$$

The property (8.98) then leads to

$$2K(k) \underset{\varepsilon \rightarrow 0}{\sim} 2 \ln \left(\frac{2}{\sqrt{\varepsilon}} \right) \underset{\varepsilon \rightarrow 0}{\sim} \ln \left(\frac{4}{\varepsilon} \right).$$

Using this result, as well as (8.96), in Eq. (8.92) yields

$$\begin{aligned} \varphi_\infty &\underset{b \rightarrow b_c^+}{\sim} \varphi_{\text{em}} + \epsilon_L \times 2 \left[\ln \left(\frac{4}{\varepsilon} \right) - \ln \left(\frac{\sqrt{3} + \sqrt{6u_{\text{em}} + 1}}{\sqrt{2(1 - 3u_{\text{em}})}} \right) - \ln \left(\frac{\sqrt{3} + 1}{\sqrt{2}} \right) \right] \\ &\underset{b \rightarrow b_c^+}{\sim} \varphi_{\text{em}} + \epsilon_L \ln \left(\frac{16}{\varepsilon^2} \frac{2(1 - 3u_{\text{em}})}{(\sqrt{3} + \sqrt{6u_{\text{em}} + 1})^2} \frac{2}{(\sqrt{3} + 1)^2} \right). \end{aligned}$$

Note that we have used the limits (8.93) to evaluate the prefactor in Eq. (8.92) as $\sqrt{2/(u_a - u_n)} \sim \sqrt{2/(1/3 - (-1/6))} = 2$. Expressing ε^2 in terms of $b - b_c$ via Eq. (8.100) leads to the final formula:

$$\varphi_\infty \underset{b \rightarrow b_c^+}{\sim} \varphi_{\text{em}} + \epsilon_L \ln \left(\frac{648(2\sqrt{3} - 3)(1 - 3u_{\text{em}})}{2 + 3u_{\text{em}} + \sqrt{3(1 + 6u_{\text{em}})}} \times \frac{m}{b - b_c} \right) \quad (\tilde{\lambda}_{\text{em}} < \tilde{\lambda}_p). \quad (8.101)$$

⁹See e.g. Eq. (112.1) of Ref. [78] with $k' = \sqrt{1 - k^2}$ or Eq. (17.3.26) of Ref. [3] with $m = k^2$ and $m_1 = 1 - k^2$.

The prefactor of $m/(b - b_c)$ in the logarithm is plotted as a function of $r_{\text{em}} = m/u_{\text{em}}$ 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 $u_{\text{em}} \rightarrow 0$ (i.e. $r_{\text{em}} \rightarrow +\infty$) in this formula. We get

$$\boxed{\Delta\varphi \underset{b \rightarrow b_c^+}{\sim} \epsilon_L \ln \left(\frac{648(7\sqrt{3} - 12)m}{b - b_c} \right)}. \quad (8.102)$$

We recover the result (8.84): $\Delta\varphi$ is diverging when $b \rightarrow b_c^+$. Moreover, Eq. (8.102) specifies this divergence as logarithmic in $b - b_c$. More generally, Eq. (8.101) shows that $\varphi_\infty - \varphi_{\text{em}}$ diverges logarithmically in $b - b_c$ when $b \rightarrow b_c^+$, whatever the position of the emission point.

Let us express $\Delta\varphi$ in terms of the deflection angle Θ and the winding number n through Eqs. (8.80) and (8.81):

$$\Delta\varphi = \Theta - \epsilon_L \pi + 2\pi n. \quad (8.103)$$

We can then deduce from (8.102) that

$$\boxed{b - b_c \underset{b \rightarrow b_c^+}{\sim} \underbrace{648(7\sqrt{3} - 12)e^{-\pi}}_{\simeq 3.482284} m e^{-\epsilon_L \Theta - 2\pi|n|}.} \quad (8.104)$$

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. [106], where $\delta D = b - b_c$ and Θ is the same as ours, or with Eq. (7.4.54) of Ref. [203], where $\Delta\ell = (b - b_c)/(2m)$.

Historical note : Equation (8.104) has been first derived by Charles Galton Darwin – the grandson of the famous naturalist Charles Robert Darwin – in 1959 [154] (cf. Eqs. (31) and (32) of Ref. [154], where $\mu = \Theta$).

8.4.2 Asymptotic direction for $b < b_c$

When the impact parameter b is lower than the critical value, the null geodesic \mathcal{L} has no periastrion and φ_{em} and φ_∞ are given by Eq. (8.68) with $\epsilon_{\text{in}} = -1$ (outgoing motion):

$$\varphi_{\text{em}} = \varphi_* + \epsilon_L \Phi_b(u_{\text{em}}) \quad \text{and} \quad \varphi_\infty = \varphi_* + \epsilon_L \Phi_b(0). \quad (8.105)$$

We have then

$$\varphi_\infty = \varphi_{\text{em}} + \epsilon_L [\Phi_b(0) - \Phi_b(u_{\text{em}})]. \quad (8.106)$$

$\Phi_b(0)$ is given by Eq. (8.74):

$$\Phi_b(0) = \frac{1}{\sqrt{2(u_* - u_n)}} [K(k) - F(\phi_b(0), k)]. \quad (8.107)$$

If $u_{\text{em}} < u_*$, $\Phi_b(u_{\text{em}})$ is given by Eq. (8.74) as well, so that Eq. (8.106) becomes

$$\boxed{\varphi_\infty = \varphi_{\text{em}} + \frac{\epsilon_L}{\sqrt{2(u_* - u_n)}} [F(\phi_b(u_{\text{em}}), k) - F(\phi_b(0), k)]} \quad \begin{array}{l} (b < b_c) \\ (u_{\text{em}} < u_*) \end{array}, \quad (8.108)$$

with $\phi_b(u_{\text{em}})$ and $\phi_b(0)$ given by Eq. (8.72):

$$\phi_b(u_{\text{em}}) = \arccos \left(\frac{|u_* - u_{\text{em}}|}{u_* + u_{\text{em}} - 2u_n} \right) \quad \text{and} \quad \phi_b(0) = \arccos \left(\frac{u_*}{u_* - 2u_n} \right). \quad (8.109)$$

If $u_{\text{em}} > u_*$, then $\Phi_b(u_{\text{em}})$ is given by Eq. (8.76). Combining with Eq. (8.107), we can then write Eq. (8.106) as

$$\boxed{\varphi_\infty = \varphi_{\text{em}} + \frac{\epsilon_L}{\sqrt{2(u_* - u_n)}} [2K(k) - F(\phi_b(u_{\text{em}}), k) - F(\phi_b(0), k)]} \quad \begin{array}{l} (b < b_c) \\ (u_{\text{em}} > u_*) \end{array}, \quad (8.110)$$

where $\phi_b(u_{\text{em}})$ is still given by Eq. (8.109). $\varphi_\infty - \varphi_{\text{em}}$ is plotted as a function of b in Fig. 8.17 for various values of $r_{\text{em}} = m/u_{\text{em}}$. For $r_{\text{em}} \gg m$ (cf. the red curve $r_{\text{em}} = 100$ m for $b < b_c$), we note that $\varphi_\infty - \varphi_{\text{em}} \simeq 0$, as expected. As for the case $b > b_c$, we also note that $\varphi_\infty - \varphi_{\text{em}}$ is diverging when b tends to b_c . Let us quantify this diverging behavior:

Limit $b \rightarrow b_c^-$

When $b \rightarrow b_c^-$, we have the following limits [cf. Eqs. (8.35) and (8.67)]:

$$\lim_{b \rightarrow b_c^-} u_n = -\frac{1}{6} \quad \text{and} \quad \lim_{b \rightarrow b_c^-} u_* = \frac{1}{3}.$$

Consequently, Eq. (8.109) yields

$$\lim_{b \rightarrow b_c^-} \phi_b(u_{\text{em}}) = \arccos \left(\frac{|1 - 3u_{\text{em}}|}{2 + 3u_{\text{em}}} \right) \quad \text{and} \quad \lim_{b \rightarrow b_c^-} \phi_b(0) = \arccos \left(\frac{1}{2} \right) = \frac{\pi}{3}. \quad (8.111)$$

Moreover, from expression (8.73) for k and the above limits for u_* and u_n , we have (see also Fig. 8.13)

$$\lim_{b \rightarrow b_c^-} k = 1.$$

Given the values (8.111) and expression (8.95) for the elliptic integral F when $k = 1$, we get

$$\lim_{b \rightarrow b_c^-} F(\phi_b(u_{\text{em}}), k) = \ln \left(\frac{2 + 3u_{\text{em}} + \sqrt{3(1 + 6u_{\text{em}})}}{|1 - 3u_{\text{em}}|} \right) \quad (8.112a)$$

$$\lim_{b \rightarrow b_c^-} F(\phi_b(0), k) = \ln(2 + \sqrt{3}). \quad (8.112b)$$

When $u_{\text{em}} < u_*$, inserting these formulas into Eq. (8.108) leads to

$$\boxed{\varphi_\infty \underset{b \rightarrow b_c^-}{\sim} \varphi_{\text{em}} + \epsilon_L \ln \left(\frac{2 + 3u_{\text{em}} + \sqrt{3(6u_{\text{em}} + 1)}}{(2 + \sqrt{3})(1 - 3u_{\text{em}})} \right)} \quad \left(u_{\text{em}} < \frac{1}{3} \right). \quad (8.113)$$

We have written $u_{\text{em}} < 1/3$ because $u_* \rightarrow 1/3$ when $b \rightarrow b_c^-$. For this reason, we also get rid of the absolute value around $1 - 3u_{\text{em}}$. Note that the value of φ_∞ given by Eq. (8.113) is

identical to that given by Eq. (8.97), which was obtained for $b \rightarrow b_c^+$ and an emission point beyond the periastron. This is not surprising if one invokes the continuity around $b = b_c$ of null geodesics as regards their part beyond the periastron for $b > b_c$ and outside $r = 3m$ for $b < b_c$. Indeed, the geodesics with $b \rightarrow b_c^+$ differ significantly from those with $b \rightarrow b_c^-$ only on parts including the periastron. This continuity appears clearly on Fig. 8.17: the curves with $b < b_c$ and $r_{\text{em}} > 3m$ have a continuous prolongation with the dashed curves, which are the curves with $b > b_c$ without any periastron on the path from the emission point to infinity.

For $u_{\text{em}} > u_*$, we shall use formula (8.110), with $K(k)$ having the diverging behavior (8.98) since $\lim_{b \rightarrow b_c^-} k = 1$. To express $K(k)$ in terms of $|b - b_c|$, let us introduce the small parameter $\varepsilon > 0$ such that

$$u_n = -\frac{1}{6} - \varepsilon. \quad (8.114)$$

The relation $P_b(u_n) = 0$, once expanded to first order in ε , leads then to the following relation between $b_c - b$ and ε :

$$\frac{b_c - b}{m} = \frac{81\sqrt{3}}{4}\varepsilon. \quad (8.115)$$

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

$$\sqrt{1 - k^2} \underset{\varepsilon \rightarrow 0}{\sim} \sqrt{1 - \left(1 - \frac{\varepsilon}{2}\right)} \underset{\varepsilon \rightarrow 0}{\sim} \sqrt{\frac{\varepsilon}{2}}$$

The property (8.98) then leads to

$$2K(k) \underset{\varepsilon \rightarrow 0}{\sim} 2 \ln \left(\frac{4\sqrt{2}}{\sqrt{\varepsilon}} \right) \underset{\varepsilon \rightarrow 0}{\sim} \ln \left(\frac{32}{\varepsilon} \right).$$

Substituting this expression for $K(k)$, as well as Eq. (8.112) for $F(\phi_b(u_{\text{em}}), k)$ and $F(\phi_b(0), k)$, into Eq. (8.110) results in

$$\begin{aligned} \varphi_\infty &\underset{b \rightarrow b_c^-}{\sim} \varphi_{\text{em}} + \frac{\epsilon_L}{\sqrt{2\left(\frac{1}{3} + \frac{1}{6}\right)}} \left[\ln \left(\frac{32}{\varepsilon} \right) - \ln \left(\frac{2 + 3u_{\text{em}} + \sqrt{3(1 + 6u_{\text{em}})}}{3u_{\text{em}} - 1} \right) - \ln(2 + \sqrt{3}) \right] \\ &\underset{b \rightarrow b_c^-}{\sim} \varphi_{\text{em}} + \epsilon_L \ln \left(\frac{32}{\varepsilon} \frac{(2 - \sqrt{3})(3u_{\text{em}} - 1)}{2 + 3u_{\text{em}} + \sqrt{3(1 + 6u_{\text{em}})}} \right). \end{aligned}$$

Finally, using (8.115) to let appear $b_c - b$ instead of ε , we get

$$\varphi_\infty \underset{b \rightarrow b_c^-}{\sim} \varphi_{\text{em}} + \epsilon_L \ln \left(\frac{648(2\sqrt{3} - 3)(3u_{\text{em}} - 1)}{2 + 3u_{\text{em}} + \sqrt{3(1 + 6u_{\text{em}})}} \times \frac{m}{b_c - b} \right) \quad \left(u_{\text{em}} > \frac{1}{3} \right). \quad (8.116)$$

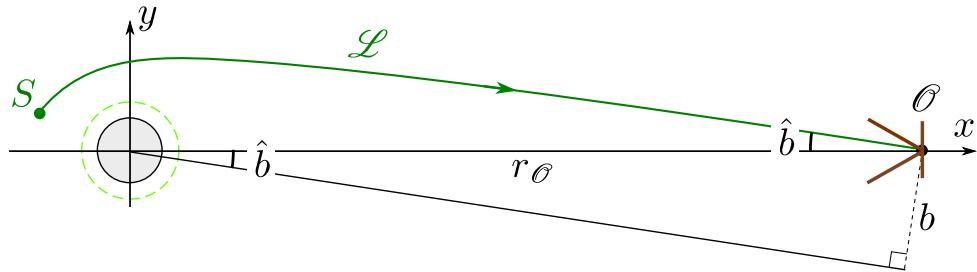


Figure 8.18: Link between the observation angle \hat{b} and the impact parameter b for an asymptotic observer \mathcal{O} located at $r_{\mathcal{O}} \gg m$.

Hence we recover a logarithmic divergence of $\varphi_{\infty} - \varphi_{\text{em}}$ when b tends to b_c by lower values.

Remark 2: The prefactor of $m/(b_c - b)$ in formula (8.116) is exactly the opposite of the prefactor of $m/(b - b_c)$ in formula (8.101). It is plotted as a function of $r_{\text{em}} = m/u_{\text{em}}$ in Fig. 8.22 below.

The upper bound on u_{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 $u_{\text{em}} \rightarrow 1/2$ in Eq. (8.116) results in, after simplification,

$$\varphi_{\infty} \underset{b \rightarrow b_c^-}{\sim} \varphi_{\text{em}} + \epsilon_L \ln \left(\frac{648(26\sqrt{3} - 45)m}{b_c - b} \right) \quad \left(u_{\text{em}} \rightarrow \frac{1}{2} \right). \quad (8.117)$$

Remark 3: As a check, Eq. (8.117) agrees with Eq. (4) in Ref. [233].

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 [6, 82]. As stated in Sec. 8.1, we shall defer the discussion of that image to the chapter dealing with rotating black holes (Sec. 12.5.3).

8.5.1 The asymptotic observer

Let us consider some “far-away” static observer, i.e. an observer \mathcal{O} located at $r = r_{\mathcal{O}} \gg m$. Without any loss of generality, we may assume that \mathcal{O} is located at $\theta = \pi/2$ and $\varphi = 0$ (cf. Fig. 8.18). Furthermore, we suppose that \mathcal{O} is equipped with an optical device (telescope) pointing in the direction from \mathcal{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 \mathcal{O} 's screen with a angle \hat{b} with respect to the telescope axis (in \mathcal{O} 's frame) within the telescope aperture. The angle \hat{b} is actually related to the geodesic impact parameter b by $\sin \hat{b} = b/r_{\mathcal{O}}$ (cf. Fig. 8.18), which, for large $r_{\mathcal{O}}$, can be rewritten as

$$\hat{b} = \frac{b}{r_{\mathcal{O}}}. \quad (8.118)$$

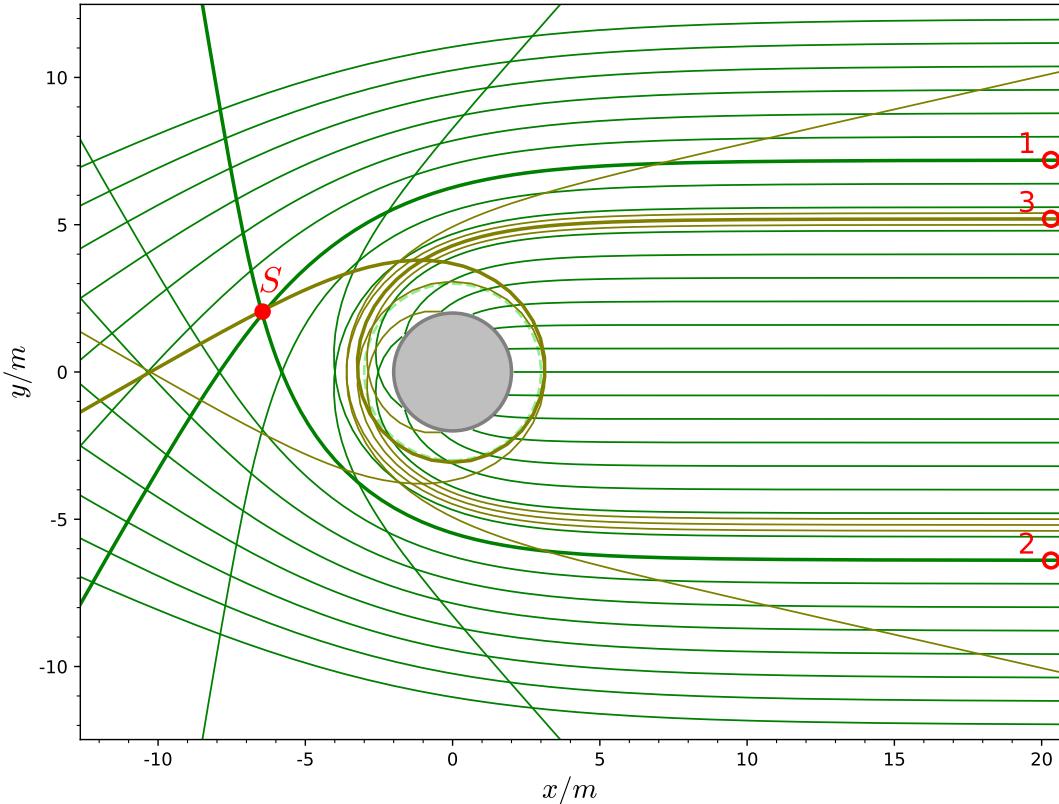


Figure 8.19: Null geodesics in Schwarzschild spacetime with $\varphi_\infty = 0$ and various values of the impact parameter b . The green ones have b ranging from 0 to $12m$, by steps of $0.8m$, while the olive ones have b close to b_c , namely $b/m \in \{5.0, 5.2, 5.4\}$. S is a luminous point source and three of its images for an observer at $y = 0$ and $x \rightarrow +\infty$ are obtained by following the drawn null geodesics through S . [Figure produced with the notebook [D.4.13](#)]

A second parameter characterizing any incoming null geodesic is a polar angle $\alpha \in [0, 2\pi)$ so that (\hat{b}, α) are spherical coordinates on \mathcal{O} 's celestial sphere, the North pole $(\hat{b} = 0)$ of which coinciding with the direction to the black hole. The image on \mathcal{O} 's screen at the telescope output may be then parameterized by polar coordinates (\hat{b}, α) , 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 $1/r_\mathcal{O}$.

Moreover, we shall consider the asymptotic observer limit: $r_\mathcal{O} \rightarrow \infty$. In that limit, null geodesics that reach \mathcal{O} must have

$$\varphi_\infty = 0 \mod 2\pi. \quad (8.119)$$

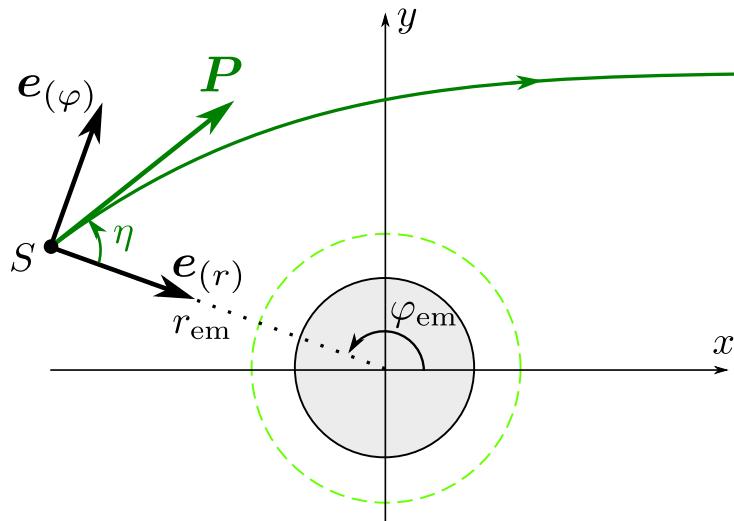


Figure 8.20: Part $(\mathbf{e}_{(r)}, \mathbf{e}_{(\varphi)})$ of the vector frame of the emitter \mathcal{O}_{em} and emission angle η . The coordinates (x, y) are defined in terms of the Schwarzschild-Droste coordinates by $x = r \cos \varphi$ and $y = r \sin \varphi$. \mathbf{P} is the photon momentum as measured by \mathcal{O}_{em} .

8.5.2 Images of a point source

Qualitative analysis from a concrete example

To understand the formation of images on \mathcal{O} 's screen, a bunch of null geodesics with $\varphi_\infty = 0$ is drawn in Fig. 8.19. \mathcal{O} is located in the far right of this figure and all the drawn geodesics eventually hit \mathcal{O} 's screen. Let us consider a luminous point source S . From the 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 \mathcal{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 \mathcal{O} ; its 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 \mathcal{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 $x_S > x_{\mathcal{O}}$. This is clear if we consider the lower-right panel of Fig. 8.15 (case $b = 5.355 \text{ m}$): a source with $x_S > x_{\mathcal{O}} \gg m$ and $y = 5.355 \text{ m}$ gives an image on \mathcal{O} 's screen at $b = 5.355 \text{ m}$.

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 \mathcal{O}_{em} . Furthermore, we suppose that the rest frame of \mathcal{O}_{em} is an orthonormal tetrad $(\mathbf{u}_{\text{em}}, \mathbf{e}_{(r)}, \mathbf{e}_{(\theta)}, \mathbf{e}_{(\varphi)})$ such that the first spacelike vector, $\mathbf{e}_{(r)}$, is always pointing towards the black hole (cf. Fig. 8.20). Note that the first vector of the tetrad is necessarily the 4-velocity \mathbf{u}_{em} of \mathcal{O}_{em} . More precisely, we set

$$\mathbf{u}_{\text{em}} = \left(1 - \frac{2m}{r_{\text{em}}}\right)^{-1/2} \partial_t \quad (8.120\text{a})$$

$$\mathbf{e}_{(r)} = -\left(1 - \frac{2m}{r_{\text{em}}}\right)^{1/2} \partial_r \quad (8.120\text{b})$$

$$\mathbf{e}_{(\theta)} = -r_{\text{em}}^{-1} \partial_\theta \quad (8.120\text{c})$$

$$\mathbf{e}_{(\varphi)} = -r_{\text{em}}^{-1} \partial_\varphi, \quad (8.120\text{d})$$

where $(\partial_t, \partial_r, \partial_\theta, \partial_\varphi)$ is the natural basis associated with Schwarzschild-Droste coordinates and r_{em} is the (constant) r -coordinate of \mathcal{O}_{em} , and hence of \mathcal{O}_{em} . It is immediate that $(\mathbf{u}_{\text{em}}, \mathbf{e}_{(r)}, \mathbf{e}_{(\theta)}, \mathbf{e}_{(\varphi)})$ is an orthonormal tetrad with respect to the Schwarzschild metric (6.14). Moreover, having a 4-velocity \mathbf{u}_{em} collinear to the Killing vector ∂_t makes the observer \mathcal{O}_{em} static [cf. Eq. (7.28)].

The generic 4-momentum \mathbf{p} of a photon evolving in the equatorial plane $\theta = \pi/2$ is given by Eqs. (7.10) and (8.1)-(8.3):

$$\mathbf{p} = E \left[\left(1 - \frac{2m}{r}\right)^{-1} \partial_t \pm \sqrt{1 - b^2 U(r)} \partial_r + \epsilon_L \frac{b}{r^2} \partial_\varphi \right], \quad (8.121)$$

where we have used Eq. (8.14) to let appear $U(r)$ and Eqs. (8.12)-(8.13) to express¹⁰ L in terms of E and b . Taking the value of \mathbf{p} at the emission point S and using (8.120) to express ∂_t , ∂_r and ∂_φ in terms of \mathcal{O}_{em} 's orthonormal tetrad, we get

$$\mathbf{p} = \varepsilon_{\text{em}} (\mathbf{u}_{\text{em}} + \mathbf{n}), \quad (8.122)$$

where

$$\varepsilon_{\text{em}} := -\mathbf{u}_{\text{em}} \cdot \mathbf{p} = E \left(1 - \frac{2m}{r_{\text{em}}}\right)^{-1/2} \quad (8.123)$$

is the energy of the photon as measured by \mathcal{O}_{em} (cf. Sec. 1.4) and

$$\mathbf{n} := \pm \sqrt{1 - b^2 U(r_{\text{em}})} \mathbf{e}_{(r)} - \epsilon_L b \sqrt{U(r_{\text{em}})} \mathbf{e}_{(\varphi)} \quad (8.124)$$

is a unit spacelike vector orthogonal to \mathbf{u}_{em} : $\mathbf{n} \cdot \mathbf{n} = 1$ and $\mathbf{u}_{\text{em}} \cdot \mathbf{n} = 0$.

¹⁰Note that formula (8.121) assumes $E \neq 0$, otherwise $b := |L|/E$ would be ill-defined. As discussed in Sec. 8.2.3, the property $E \neq 0$, and even $E > 0$, holds as soon as the photon travels in the black hole's exterior.

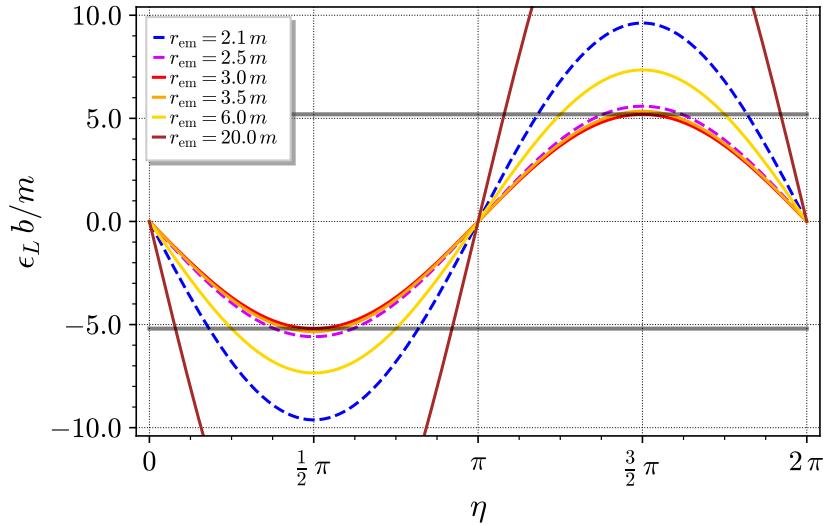


Figure 8.21: Impact parameter b of null geodesics (multiplied by $\epsilon_L = \pm 1$) as a function of the emission angle η in the emitter's rest frame, for various values of the r -coordinate r_{em} of the static emitter. Dashed curves correspond to values of r_{em} lower than 3 m . The two horizontal grey lines marks $b = b_c \simeq 5.196\text{ m}$. [Figure produced with the notebook [D.4.14](#)]

In view of (8.122) and $\mathbf{u}_{\text{em}} \cdot \mathbf{n} = 0$, the linear momentum of the photon as measured by \mathcal{O}_{em} is $\mathbf{P} = \varepsilon_{\text{em}} \mathbf{n}$ [cf. Eq. (1.22) and Fig. 8.20]. We conclude that the unit vector \mathbf{n} gives the direction of emission of the photon in \mathcal{O}_{em} 's rest frame. Let us then define the **emission angle** η by

$$\mathbf{n} = \cos \eta \mathbf{e}_{(r)} + \sin \eta \mathbf{e}_{(\varphi)}. \quad (8.125)$$

Equation (8.124) yields immediately

$$\cos \eta = \pm \sqrt{1 - b^2 U(r_{\text{em}})} \quad \text{and} \quad \sin \eta = -\epsilon_L b \sqrt{U(r_{\text{em}})}, \quad (8.126)$$

from which we can express the impact parameter b and in terms of the emission angle η :

$$b = -\epsilon_L \frac{\sin \eta}{\sqrt{U(r_{\text{em}})}}.$$

(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 $b^{-2} - U(r_{\text{em}}) \geq 0$, from which we get $0 \leq b^2 U(r_{\text{em}}) \leq 1$.

The variation of b in terms of η and r_{em} , as given by formula (8.127), is depicted in Fig. 8.21. We notice that the smallest range of b is $[0, b_c]$ and is reached for $r_{\text{em}} = 3\text{ m}$, which is not surprising in view of (8.127) since the maximum of U precisely occurs at $r = 3\text{ m}$ (cf. Fig. 8.1).

Infinite sequence of images

It appears clearly from Fig. 8.21 that, whatever the location r_{em} of the emitter, including locations inside the photon sphere ($r_{\text{em}} < 3\text{ m}$), there always exist four emission angles η giving

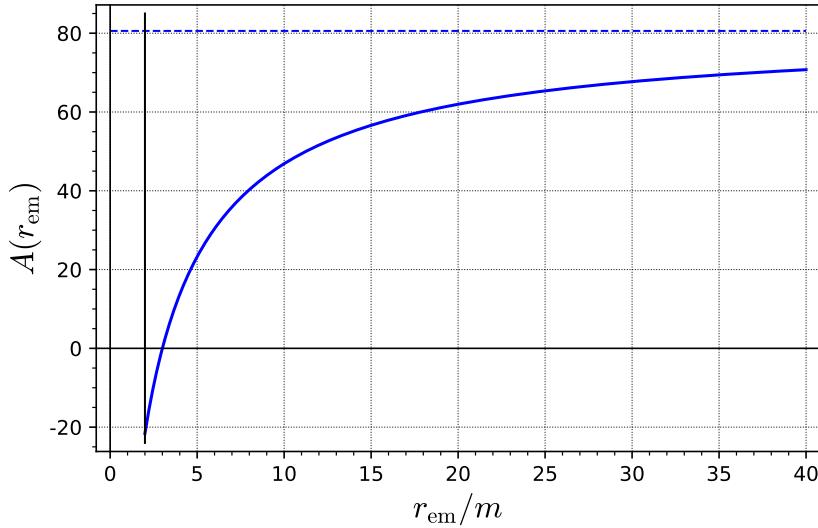


Figure 8.22: Function $A(r_{\text{em}})$ defined by Eq. (8.130). The dashed line marks the horizontal asymptote $A = 648(7\sqrt{3} - 12) \simeq 80.58246$. [Figure produced with the notebook D.4.14]

birth to impact parameters b in the vicinity of the critical value b_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 $b > b_c$ and only a single emission direction is outgoing and reaches infinity for $b < b_c$. In other words, on the four values of η , only two of them define geodesics that may reach the asymptotic observer \mathcal{O} .

Let $\varphi_{\text{em}} \in [0, 2\pi)$ be the φ coordinate of the emitter \mathcal{O}_{em} . A null geodesic emitted by \mathcal{O}_{em} reaches the asymptotic observer \mathcal{O} , who is located at $\varphi = 0$, iff $\varphi_\infty = 0 \pmod{2\pi}$ [Eq. (8.119)], i.e. iff

$$\varphi_\infty = 2\pi\epsilon_L n, \quad n \in \mathbb{N}. \quad (8.128)$$

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) $r_{\text{em}} > 3m$, b is close to b_c from above and $\tilde{\lambda}_{\text{em}} < \tilde{\lambda}_p$ or (ii) $r_{\text{em}} < 3m$ and b is close to b_c from below. In the first case, we may use the approximate formula (8.101) with $u_{\text{em}} = m/r_{\text{em}}$, while in the second case formula (8.116) is relevant. Inserting (8.128) in either of these formulas, we get

$$2\pi\epsilon_L n = \varphi_{\text{em}} + \epsilon_L \ln \left(A(r_{\text{em}}) \frac{m}{b - b_c} \right), \quad (8.129)$$

with (cf. Remark 2, page 265)

$$A(r_{\text{em}}) := \frac{648(2\sqrt{3} - 3)(r_{\text{em}} - 3m)}{2r_{\text{em}} + 3m + \sqrt{3r_{\text{em}}(r_{\text{em}} + 6m)}}.$$

(8.130)

Note that $A(r_{\text{em}}) > 0$ for $r_{\text{em}} > 3m$ and $A(r_{\text{em}}) < 0$ for $r_{\text{em}} < 3m$. Taking into account that $b - b_c > 0$ for case (i) and $b - b_c < 0$ 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 :

$$b = b_c + mA(r_{\text{em}})e^{\epsilon_L \varphi_{\text{em}} - 2\pi n},$$

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^- :

$$b_n^+ = b_c + mA(r_{\text{em}})e^{\varphi_{\text{em}} - 2\pi n}, \quad n \in \mathbb{N}^* \quad (8.131a)$$

$$b_n^- := b_c + mA(r_{\text{em}})e^{-\varphi_{\text{em}} - 2\pi n}, \quad n \in \mathbb{N}^*. \quad (8.131b)$$

The function $A(r_{\text{em}})$ is plotted in Fig. 8.22. It varies monotonically between

$$\lim_{r_{\text{em}} \rightarrow 2m} A(r_{\text{em}}) = \underbrace{-648(26\sqrt{3} - 45)}_{\simeq -21.59201} \quad \text{and} \quad \lim_{r_{\text{em}} \rightarrow +\infty} A(r_{\text{em}}) = \underbrace{648(7\sqrt{3} - 12)}_{\simeq 80.58246}.$$

One shall recall that formulas (8.131) have been derived under the assumption $|b - b_c| \ll m$. Given the amplitude of $A(r_{\text{em}})$ shown in Fig. 8.22 and the value $e^{-2\pi} \simeq 1.9 \times 10^{-3}$, this is valid for any $n \geq 1$, except maybe for b_n^+ with $n = 1$ and φ_{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 :

$$\sin \eta_n^+ := -\sqrt{U(r_{\text{em}})} (b_c + mA(r_{\text{em}})e^{\varphi_{\text{em}} - 2\pi n}), \quad n \in \mathbb{N}^* \quad (8.132a)$$

$$\sin \eta_n^- := \sqrt{U(r_{\text{em}})} (b_c + mA(r_{\text{em}})e^{-\varphi_{\text{em}} - 2\pi n}), \quad n \in \mathbb{N}^*. \quad (8.132b)$$

Note that $0 \leq b_c \sqrt{U(r_{\text{em}})} \leq 1$ and $b_c \sqrt{U(r_{\text{em}})} = 1 \iff r_{\text{em}} = 3m$. As discussed above (cf. Fig. 8.21), for $r_{\text{em}} \neq 3m$, there are always some solutions for η_n^+ and η_n^- for n sufficiently large (in practice, $n \geq 1$). We therefore conclude:

Property 8.3: images of a point source

Any static point source S give birth to two infinite sequences of images on the screen of the asymptotic observer \mathcal{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 , \mathcal{O} and the black hole are not aligned, i.e. if $\varphi_{\text{em}} \notin \{0, \pi\}$, all the images of S are aligned on \mathcal{O} 's screen: they all lie in the plane defined by S , \mathcal{O} and the black hole (plane of Fig. 8.19). If S is located outside the photon sphere ($r_{\text{em}} > 3m$), all images lie outside the circle $b = b_c$ on \mathcal{O} 's screen, while if S is located inside the photon sphere ($r_{\text{em}} < 3m$), 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 $b = b_c$. 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 $e^{-2\pi} \simeq 1.9 \times 10^{-3}$.

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.

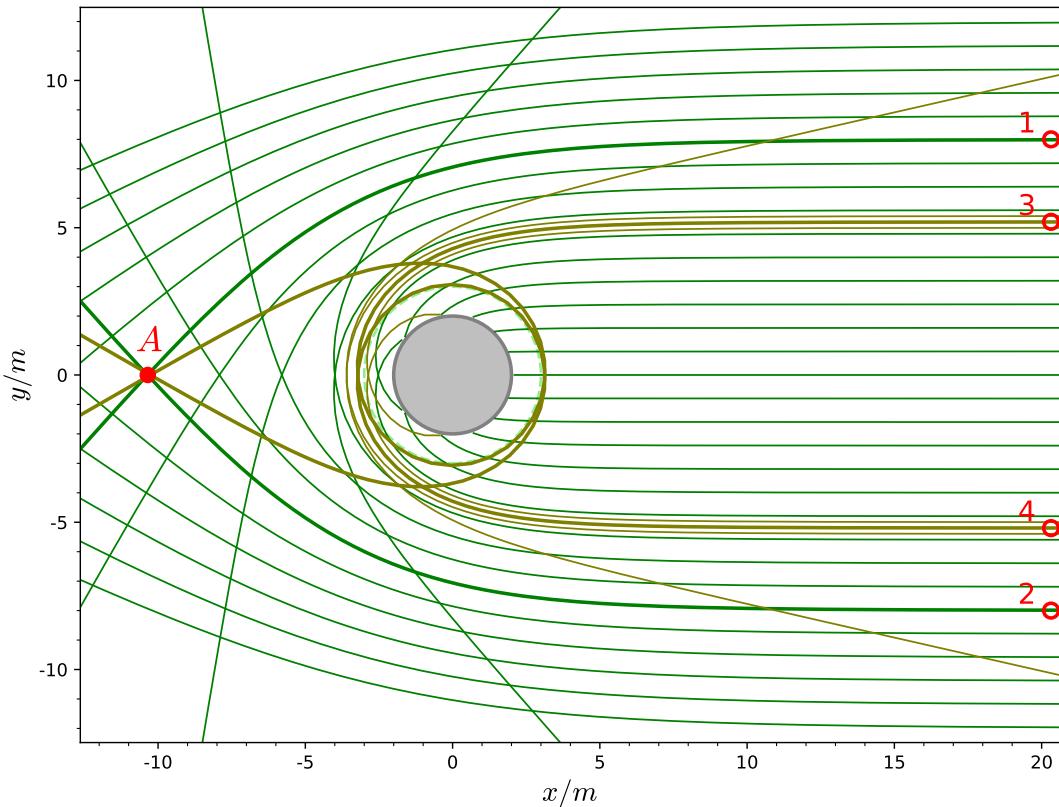


Figure 8.23: Same as Fig. 8.19 but underlining four null geodesics from a point source A located on the x -axis, i.e. aligned with the black hole and the observer on the far right. [Figure produced with the notebook D.4.13]

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 $b = b_c$. 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 , \mathcal{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 \mathcal{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 \mathcal{O} 's screen is composed by a infinite sequence $(\mathcal{C}_n)_{n \in \mathbb{N}}$ of concentric circles that accumulate exponentially fast onto the circle of radius $b = b_c$. Each circle \mathcal{C}_n has a radius b_n that is given by Eq. (8.131) with $\varphi_{\text{em}} = 0$ or π (case of Fig. 8.23), so that $b_n^+ = b_n^- =: b_n$ (case $\varphi_{\text{em}} = 0$) or $b_n^+ = b_{n-1}^- =: b_n$ (case $\varphi_{\text{em}} = \pi$). The outermost circle, \mathcal{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 \mathcal{C}_n the **$n^{\text{th}} \text{ Einstein ring of } A$** . For $n \geq 1$, \mathcal{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 \mathcal{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 $b = b_c$, has been predicted in 1959 by Charles Galton Darwin [154] (cf. historical note on p. 262). Darwin has also derived the sequence of relativistic Einstein rings for aligned sources.

8.5.4 Black hole shadow

Let us determine the image on observer \mathcal{O} 's screen when all the sources of light are very far, both from the black hole and from \mathcal{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 \mathcal{S} that is uniformly bright. In such a setting, all values of $b > b_c$ on \mathcal{O} 's screen, whatever the polar angle α , correspond to a null geodesic that originates from the shining sphere \mathcal{S} , while this is not possible for $b \leq b_c$. This is clear on Fig. 8.19: when traced backward from the right (\mathcal{O} position), (i) null geodesics with $b > b_c$ eventually end up far away from the black hole, i.e. necessarily on \mathcal{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 \mathcal{O} is located at $r = 15m$ and that the shining sphere \mathcal{S} has a coordinate radius $r = 20m$. These values allow us to put \mathcal{O} and \mathcal{S} in Fig. 8.1 according to our settings (\mathcal{S} is encompassing both \mathcal{O} and the black hole), but one shall keep in mind that both \mathcal{O} and \mathcal{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 $r = 15m$ (\mathcal{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 \mathcal{O} 's location, i.e. a region that does not intersect \mathcal{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 \mathcal{S} inwards and have bounced on the potential barrier before reaching \mathcal{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

Property 8.4: shadow of a Schwarzschild black hole

For a distant observer, the image of a uniformly bright sphere \mathcal{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 [339], 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,

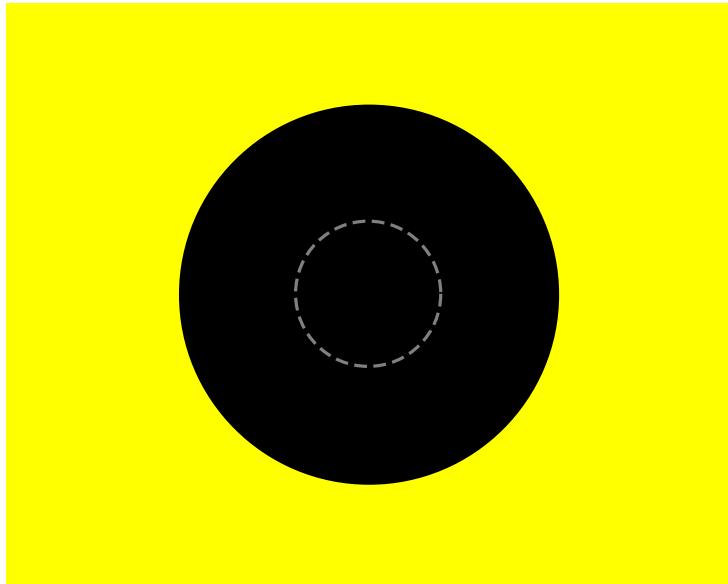


Figure 8.24: Shadow of a Schwarzschild black hole. The dashed grey circle delimits the dark disk that would appear on the screen of an observer looking at a black sphere of radius $r = 2m$ in Minkowski spacetime. The ratio between the radii of the two disks is $b_c/2m = 3\sqrt{3}/2 \simeq 2.60$.

independently of the value of r_0 . Maybe the first mention of a shadow lies in the article by Charles Galton Darwin in 1959 [154]. 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 1972 Les Houches Summer School [37]; 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 [143].

8.5.5 Image of an accretion disk

A realistic “source of light” in the vicinity of a black hole is an accretion disk [2, 315]. 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 [6]. 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 [387] and Page & Thorne [395], lying in some plane around the black hole. It is geometrically

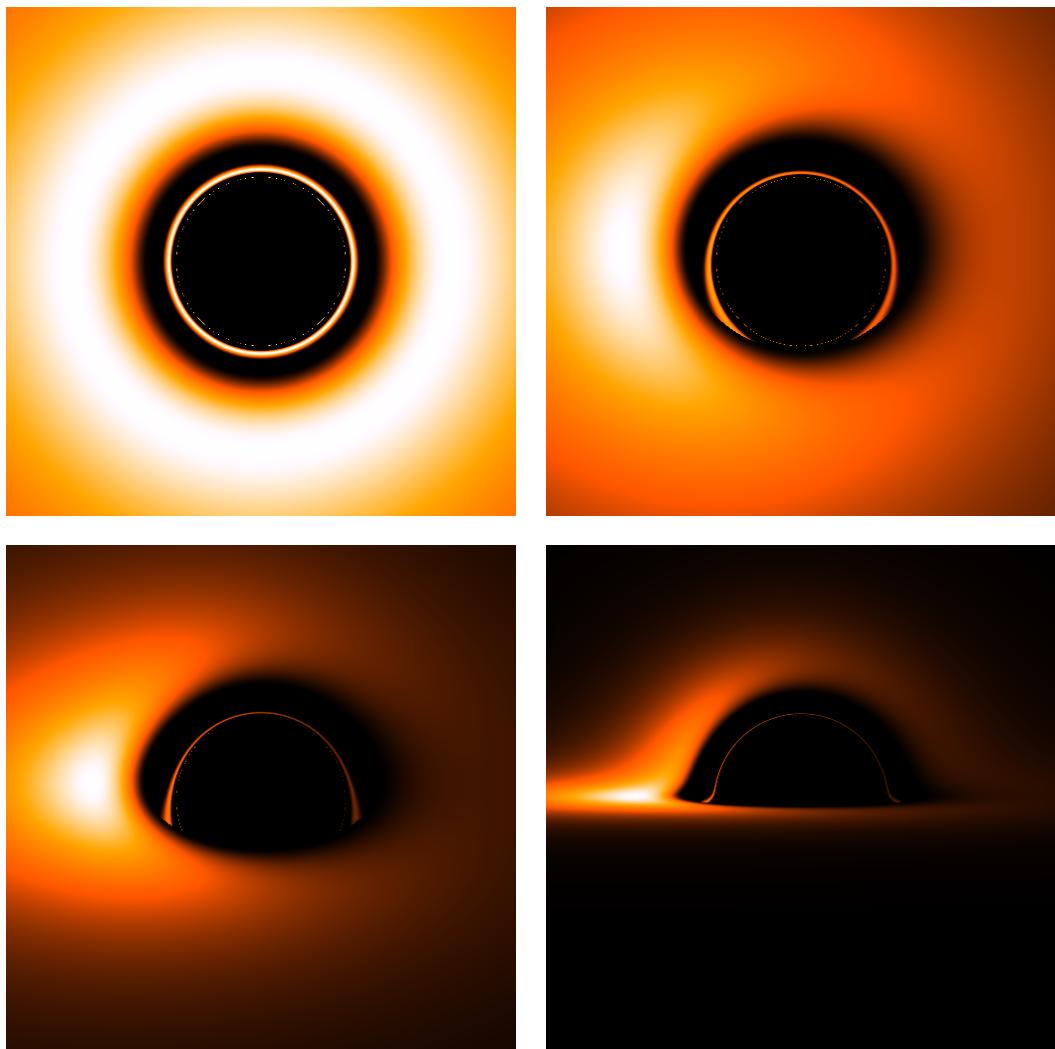


Figure 8.25: Image of an accretion disk around a Schwarzschild black hole, as seen by a distant observer, with various inclination angles: $\iota = 0$ (upper left), $\iota = \pi/6$ (upper right), $\iota = \pi/3$ (lower left) and $\iota = 3/2$ (lower right). [Figure produced by *Gyoto* with the input files given in Sec. E.2.1]

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 [492] (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 $r = 1000\text{ m}$ 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.

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 $b = 6.932\text{ m}$ 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 $b = 5.479\text{ m}$ 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 $b = 5.208\text{ m}$ 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 $b = b_c$. 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)¹¹, so that Eqs. (8.131) with $\varphi_{\text{em}} = \pi/2$ ($b = b_1^-$) and $\varphi_{\text{em}} = 3\pi/2$ ($b = b_2^+$) yield $(b - b_c)/b_c = A(r_{\text{em}})e^{-5\pi/2}/(3\sqrt{3})$. With $r_{\text{em}} = 6m$, we have $A(r_{\text{em}}) \simeq 30.3$, so that $(b - b_c)/b_c \simeq 0.0013$ 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

¹¹This asymmetry of n with respect to ϵ_L is due to the convention $\varphi_{\text{em}} \in [0, 2\pi)$; had we chosen $\varphi_{\text{em}} \in (-\pi, \pi]$, it would not be present.

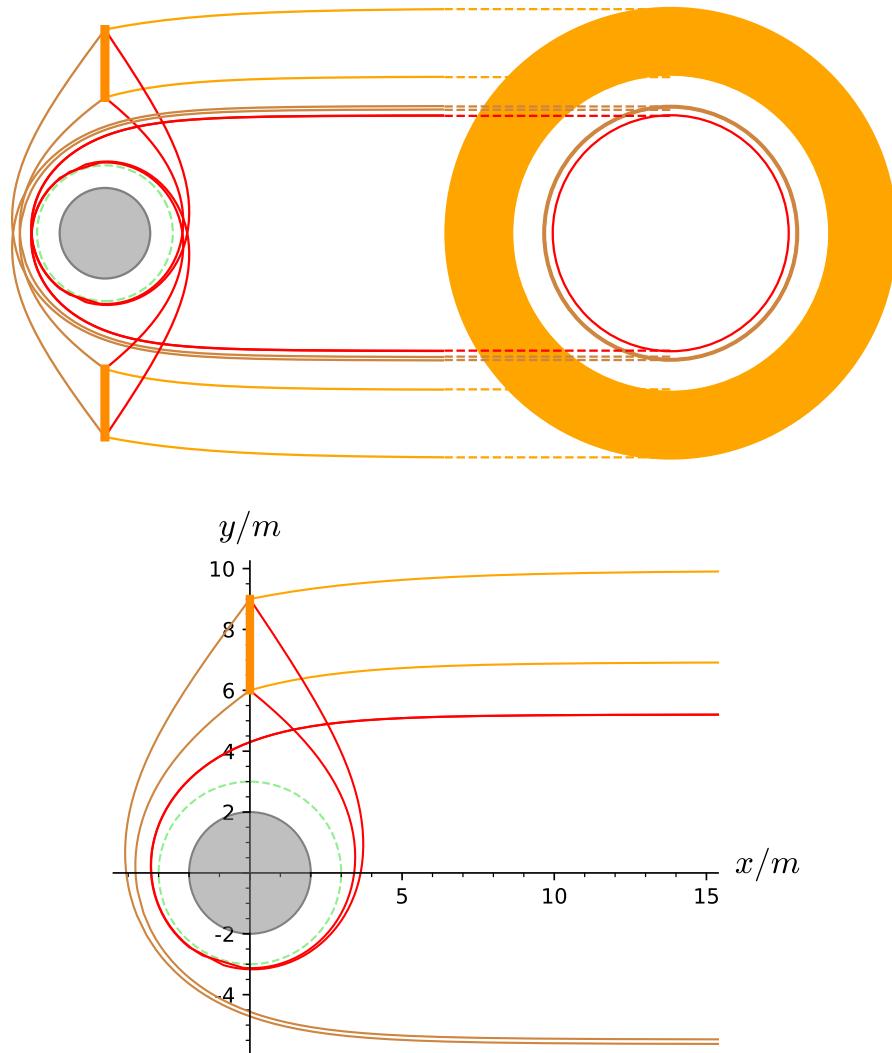


Figure 8.26: Formation of the primary (orange), secondary (brown) and tertiary (red) images of an accretion disk when viewed face-on. The figure plane is orthogonal to the disk plane, the intersection between the two planes being the y -axis, so that disk is depicted by the thick segments located at $\varphi = \pi/2$ and $\varphi = 3\pi/2$. As in Fig. 8.25, the inner boundary of the disk is $r_{\text{in}} = 6 \text{ m}$ (the ISCO), but the outer boundary has been truncated at $r_{\text{out}} = 9 \text{ m}$ for pedagogical purposes. The observer is located at $x \gg m$ and $y = 0$. The dashed green circle is the photon orbit at $r = 3 \text{ m}$. 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.4.15](#)]

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.

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 [347], (see Ref. [348] 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 [354, 355]. In 2007, Alain Riazuelo computed very precise images of a Schwarzschild black hole in front of a realistic stellar field [431, 432], 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	279
9.2	Kruskal-Szekeres coordinates	280
9.3	Maximal extension	288
9.4	Carter-Penrose diagram	295
9.5	Einstein-Rosen bridge	305
9.6	Physical relevance of the maximal extension	316

9.1 Introduction

In the preceding chapters, the Schwarzschild spacetime was considered to be $(\mathcal{M}_{\text{IEF}}, \mathbf{g})$, where the manifold \mathcal{M}_{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, $(\mathcal{M}, \mathbf{g})$ say, which on its turn cannot be extended, i.e. it is maximal. To construct $(\mathcal{M}, \mathbf{g})$, we first introduce the so-called *Kruskal-Szekeres coordinates* on \mathcal{M}_{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 halted¹ at some past boundary of \mathcal{M}_{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

¹Technically, one says that they are *incomplete* geodesics, cf. Sec. B.3.2.

of the maximal Schwarzschild spacetime $(\mathcal{M}, \mathbf{g})$, via two compactified versions of the Kruskal-Szekeres coordinates. We show that the simplest one, which is standard in the literature, does not lead to a regular conformal completion of $(\mathcal{M}, \mathbf{g})$, as defined in Sec. 4.3. The second version, built on Penrose-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 $(\mathcal{M}, \mathbf{g})$ 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 \mathcal{M}_1 , 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. (6.21)-(6.22):

$$\begin{cases} u = t - r - 2m \ln \left| \frac{r}{2m} - 1 \right| \\ v = t + r + 2m \ln \left| \frac{r}{2m} - 1 \right| \end{cases} \iff \begin{cases} t = \frac{1}{2}(u + v) \\ r + 2m \ln \left| \frac{r}{2m} - 1 \right| = \frac{1}{2}(v - u). \end{cases} \quad (9.1)$$

Although r cannot be explicitly expressed 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 \mathcal{M}_1 . The range of (u, v) is \mathbb{R}^2 .

The above relations imply

$$du = dt - \frac{dr}{1 - \frac{2m}{r}}, \quad dv = dt + \frac{dr}{1 - \frac{2m}{r}} \quad \text{and} \quad du \, dv = dt^2 - \frac{dr^2}{\left(1 - \frac{2m}{r}\right)^2}.$$

The Schwarzschild metric (6.14) can be then rewritten as

$$\mathbf{g} = - \left(1 - \frac{2m}{r}\right) du \, dv + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \, \mathbf{d}\varphi^2).$$

(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 \mathcal{M}_1 . Having a look at Fig. 9.1, we realize that we cannot extend this coordinate system to include the Schwarzschild horizon \mathcal{H} , since $r \rightarrow 2m$ is equivalent to $v - u \rightarrow -\infty$: if u (resp. v) were taking a finite value on \mathcal{H} , we would have $v \rightarrow -\infty$ (resp. $u \rightarrow +\infty$). This impossibility of extending to \mathcal{H} is also reflected by the fact that

$$\det(g_{\hat{\alpha}\hat{\beta}}) = -\frac{1}{4} \left(1 - \frac{2m}{r}\right)^2 r^4 \sin^2 \theta$$

vanishes for $r \rightarrow 2m$, which would make \mathbf{g} a degenerate bilinear form at $r = 2m$, while it is not of course.

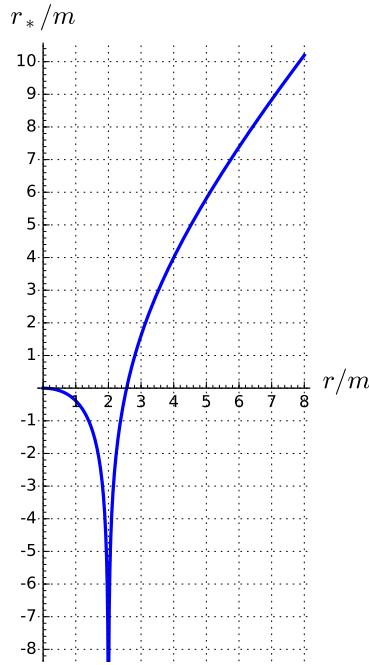


Figure 9.1: Function $r_*(r) = r + 2m \ln \left| \frac{r}{2m} - 1 \right|$ (the tortoise coordinate, cf. Eq. (6.23)). It relates r to (u, v) via $r_*(r) = (u - v)/2$ [Eq. (9.1)].

On \mathcal{M}_I , $r > 2m$ and (9.1) yields

$$r + 2m \ln \left(\frac{r}{2m} - 1 \right) = \frac{1}{2}(v - u) \implies e^{r/2m} \left(\frac{r}{2m} - 1 \right) = e^{(v-u)/4m}, \quad (9.3)$$

This motivates to introduce the coordinates (U, V) defined on \mathcal{M}_I by

$$\begin{cases} U := -e^{-u/4m} \\ V := e^{v/4m}. \end{cases} \quad (9.4)$$

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

$$dU = \frac{1}{4m} e^{-u/4m} du, \quad dV = \frac{1}{4m} e^{v/4m} dv \quad \text{and} \quad du dv = 16m^2 e^{(u-v)/4m} dU dV.$$

In view of Eq. (9.3), we get

$$du dv = 16m^2 e^{-r/2m} \left(\frac{r}{2m} - 1 \right)^{-1} dU dV = \frac{32m^3}{r} e^{-r/2m} \left(1 - \frac{2m}{r} \right)^{-1} dU dV.$$

Substituting this expression in (9.2) yields the expression of the metric with respect to the coordinates $X^\alpha := (U, V, \theta, \varphi)$:

$$g = -\frac{32m^3}{r} e^{-r/2m} dU dV + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2).$$

(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):

$$\boxed{e^{r/2m} \left(\frac{r}{2m} - 1 \right) = -UV}. \quad (9.6)$$

This relation takes a very simple form in terms of the tortoise coordinate (cf. Eq. (6.23)):

$$e^{r_*/2m} = -UV. \quad (9.7)$$

We notice that the factor $(1 - 2m/r)$ has disappeared in the metric components (9.5), which are thus perfectly regular as $r \rightarrow 2m$.

The coordinates U and V are null coordinates, since u and v are (cf. Sec. 6.3.2) and Eq. (9.4) shows that the hypersurfaces of constant U (resp. V) obviously coincide with those of constant u (resp. v). To cope with a timelike-spacelike coordinate system instead, let us introduce on \mathcal{M}_I the pair (T, X) such that U is T retarded by X and V is T advanced by X :

$$\begin{cases} U = T - X \\ V = T + X \end{cases} \iff \begin{cases} T = \frac{1}{2}(U + V) \\ X = \frac{1}{2}(V - U) \end{cases} \quad (9.8)$$

Since the range of U on \mathcal{M}_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):

$$\mathcal{M}_I : \quad X > 0 \quad \text{and} \quad -X < T < X. \quad (9.9)$$

The coordinates $X^\alpha := (T, X, \theta, \varphi)$ are called the **Kruskal-Szekeres coordinates**.

We have $dU dV = (dT - dX)(dT + dX) = dT^2 - dX^2$, so that the metric components with respect to the Kruskal-Szekeres coordinates are easily deduced from Eq. (9.5):

$$\boxed{g = \frac{32m^3}{r} e^{-r/2m} (-dT^2 + dX^2) + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)}. \quad (9.10)$$

Here r is to be considered as a function of (T, X) , which is implicitly defined by

$$\boxed{e^{r/2m} \left(\frac{r}{2m} - 1 \right) = X^2 - T^2}. \quad (9.11)$$

This relation is a direct consequence of (9.6) and (9.8). We may rewrite it as $F(r/2m) = X^2 - T^2$, where F is the function defined by

$$\begin{aligned} F : \quad (0, +\infty) &\longrightarrow (-1, +\infty) \\ x &\longmapsto e^x(x - 1). \end{aligned} \quad (9.12)$$

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 \mathcal{M}_I) and $(0, +\infty)$ (the range of

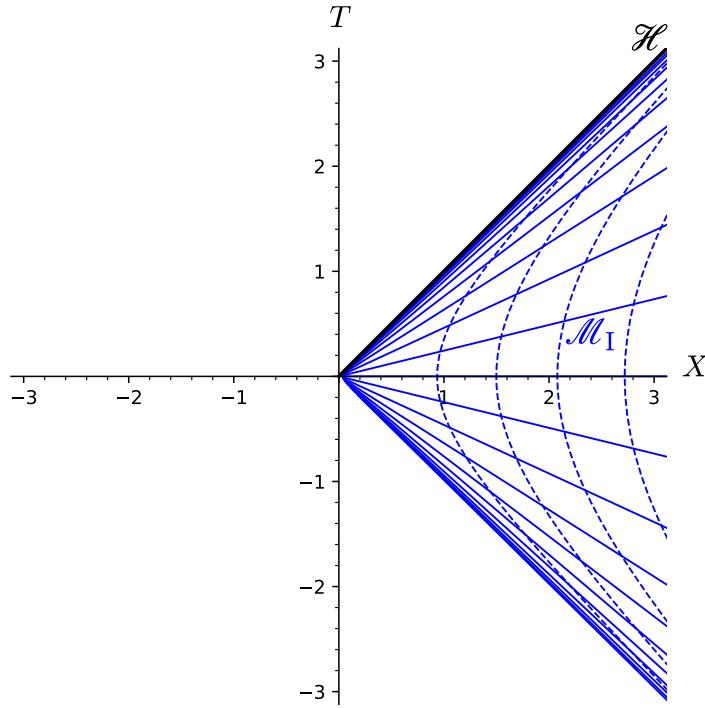


Figure 9.2: Submanifold \mathcal{M}_I in the Kruskal-Szekeres coordinates (T, X) : \mathcal{M}_I is covered by the Schwarzschild-Droste grid (in blue): the solid lines have $t = \text{const}$ (spaced apart by $\delta t = m$), while the dashed curves have $r = \text{const}$ (spaced apart by $\delta r = m/2$). [Figure generated by the notebook D.4.16]

$X^2 - T^2$ on \mathcal{M}_I , according to (9.9)). The inverse of F can be expressed in terms of the **Lambert function** W_0 [3, 141], which is defined as the inverse of $x \mapsto xe^x$:

$$\begin{aligned} W_0 : (-1/e, +\infty) &\longrightarrow (-1, +\infty) \\ x &\longmapsto y \text{ such that } ye^y = x. \end{aligned} \tag{9.13}$$

Noticing that $F(x) = e^x(x - 1) = e \times (x - 1)e^{x-1}$, we may write $F^{-1} = W$, where W is the rescaled Lambert function defined by

$$W(x) := W_0\left(\frac{x}{e}\right) + 1. \tag{9.14}$$

By construction, W is a bijection $(-1, +\infty) \rightarrow (0, +\infty)$, which obeys

$$e^{W(x)}(W(x) - 1) = x. \tag{9.15}$$

Its graph is shown in Fig. 9.4.

Since $F^{-1} = W$, we may invert relation (9.11) to $r = 2mW(X^2 - T^2)$. This allows us to eliminate r from the metric expression (9.10); using Eq. (9.15) to substitute $W(X^2 - T^2)e^{W(X^2 - T^2)}$ by $X^2 - T^2 + e^{W(X^2 - T^2)}$, we arrive at a fully explicit expression of Schwarzschild

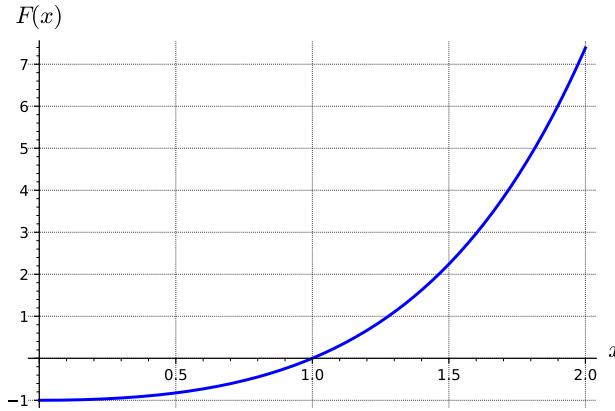


Figure 9.3: Function $F : x \mapsto e^x(x - 1)$, yielding $X^2 - T^2 = F(r/2m)$, cf. Eq. (9.11).

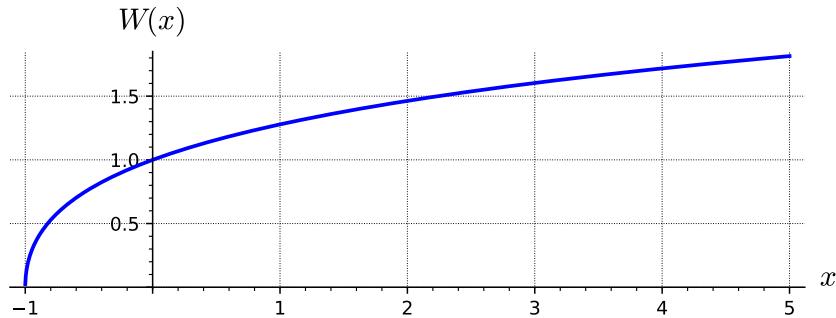


Figure 9.4: Rescaled Lambert function $W = F^{-1}$, defined by (9.14) and obeying $e^{W(x)}(W(x) - 1) = x$.

metric in terms of the Kruskal-Szekeres coordinates:

$$\boxed{\mathbf{g} = 4m^2 \left[\frac{4}{X^2 - T^2 + e^{W(X^2 - T^2)}} (-dT^2 + dX^2) + W(X^2 - T^2)^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right].} \quad (9.16)$$

The relation between the Kruskal-Szekeres coordinates and the Schwarzschild-Droste ones is obtained by combining (9.8), (9.4) and (9.1):

$$T = \frac{1}{2}(U + V) = \frac{1}{2}(e^{v/4m} - e^{-u/4m}) = \frac{1}{2}(e^{(t+r_*)/4m} - e^{(r_*-t)/4m}) = e^{r_*/4m} \sinh\left(\frac{t}{4m}\right),$$

where r_* is related to r by (6.23). Similarly $X = e^{r_*/4m} \cosh(t/4m)$. In particular, we have

$$\frac{T}{X} = \tanh\left(\frac{t}{4m}\right). \quad (9.17)$$

Since Eq. (6.23) yields $e^{r_*/4m} = e^{r/4m} \sqrt{r/(2m) - 1}$, the above relations can be written as

$$\mathcal{M}_1 : \boxed{\begin{cases} T = e^{r/4m} \sqrt{\frac{r}{2m} - 1} \sinh\left(\frac{t}{4m}\right) \\ X = e^{r/4m} \sqrt{\frac{r}{2m} - 1} \cosh\left(\frac{t}{4m}\right) \end{cases}} \iff \boxed{\begin{cases} t = 2m \ln\left(\frac{X+T}{X-T}\right) \\ r = 2m W(X^2 - T^2). \end{cases}} \quad (9.18)$$

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. (9.17).

Remark 1: Given the properties of the cosh and sinh functions, it is clear on these expressions that the constraints (9.9) are satisfied.

Remark 2: In expression (9.16) the metric components g_{TT} and g_{XX} depend on both X and T ; this shows that neither ∂_T nor ∂_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 \mathcal{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 \mathcal{M}_{IEF} introduced in Sec. 6.3.3: $\mathcal{M}_{\text{IEF}} = \mathcal{M}_I \cup \mathcal{H} \cup \mathcal{M}_{II}$. Let us show this in detail. Back on \mathcal{M}_I , we can express the IEF coordinate \tilde{t} in terms of (T, X) by combining $\tilde{t} = v - r$ [Eq. (6.30)], $v = 4m \ln V$ [Eq. (9.4)] and $V = T + X$ [Eq. (9.8)]:

$$\tilde{t} = 4m \ln(T + X) - r. \quad (9.19)$$

The above relation is a valid expression as long as $T + X > 0$. Besides, we already noticed that the function F defined by (9.12) is a bijection from the range of $r/2m$ on \mathcal{M}_{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 \mathcal{M}_I . We may use these properties to extend the Kruskal-Szekeres coordinates to all \mathcal{M}_{IEF} by requiring

$$\tilde{t} = 4m \ln(T + X) - r \quad (9.20a)$$

$$\underbrace{e^{r/2m} \left(\frac{r}{2m} - 1 \right)}_{F(r/2m)} = X^2 - T^2. \quad (9.20b)$$

The range of the coordinates (T, X) on \mathcal{M}_{IEF} is then ruled by

$$\mathcal{M}_{\text{IEF}} : \quad T + X > 0 \quad \text{and} \quad X^2 - T^2 > -1,$$

which can be rewritten as

$$\mathcal{M}_{\text{IEF}} : \quad -X < T < \sqrt{X^2 + 1}. \quad (9.21)$$

We deduce from (9.20) that

$$\begin{cases} X + T &= e^{(\tilde{t}+r)/4m} \\ X - T &= e^{(r-\tilde{t})/4m} \left(\frac{r}{2m} - 1 \right). \end{cases} \quad (9.22)$$

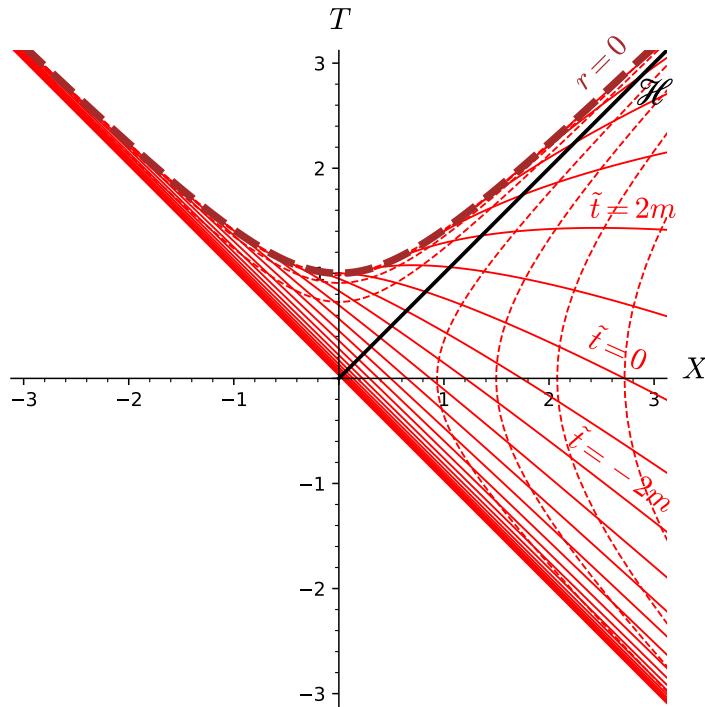


Figure 9.5: Domain of ingoing Eddington-Finkelstein coordinates, $\mathcal{M}_{\text{IEF}} = \mathcal{M}_I \cup \mathcal{H} \cup \mathcal{M}_{\text{II}}$, depicted in terms of the Kruskal-Szekeres coordinates (T, X): the solid red curves have $\tilde{t} = \text{const}$ (spaced apart by $\delta\tilde{t} = m$), while the dashed red curves have $r = \text{const}$ (spaced apart by $\delta r = m/2$). [Figure generated by the notebook [D.4.16](#)]

Hence the relation between the ingoing Eddington-Finkelstein coordinates and the Kruskal-Szekeres ones on \mathcal{M}_{IEF} :

$$\left\{ \begin{array}{l} T = e^{r/4m} \left[\cosh \left(\frac{\tilde{t}}{4m} \right) - \frac{r}{4m} e^{-\tilde{t}/4m} \right] \\ X = e^{r/4m} \left[\sinh \left(\frac{\tilde{t}}{4m} \right) + \frac{r}{4m} e^{-\tilde{t}/4m} \right] \end{array} \right. \iff \left\{ \begin{array}{l} \tilde{t} = 2m [2 \ln(T + X) - W(X^2 - T^2)] \\ r = 2m W(X^2 - T^2) \end{array} \right. \quad (9.23)$$

The various subsets of \mathcal{M}_{IEF} correspond then to the following coordinate ranges (cf. Fig. 9.5):

$$\mathcal{M}_I : \quad X > 0 \quad \text{and} \quad -X < T < X \quad (9.24a)$$

$$\mathcal{H} : \quad X > 0 \quad \text{and} \quad T = X \quad (9.24b)$$

$$\mathcal{M}_{\text{II}} : \quad |X| < T < \sqrt{X^2 + 1}. \quad (9.24c)$$

Since the relation between IEF coordinates and Kruskal-Szekeres ones is the same in \mathcal{M}_{II} as in \mathcal{M}_I (being given by (9.23) in both cases), we conclude that expressions (9.10) and (9.16) for the metric components with respect to Kruskal-Szekeres coordinates are valid in all \mathcal{M}_{IEF} .

Let us determine the relation between the Kruskal-Szekeres coordinates and the Schwarzschild-Droste ones in \mathcal{M}_{II} . Since $r < 2m$ in \mathcal{M}_{II} , Eq. (6.32) gives

$$\mathcal{M}_{\text{II}} : \quad e^{\tilde{t}/4m} = e^{t/4m} \sqrt{1 - \frac{r}{2m}},$$

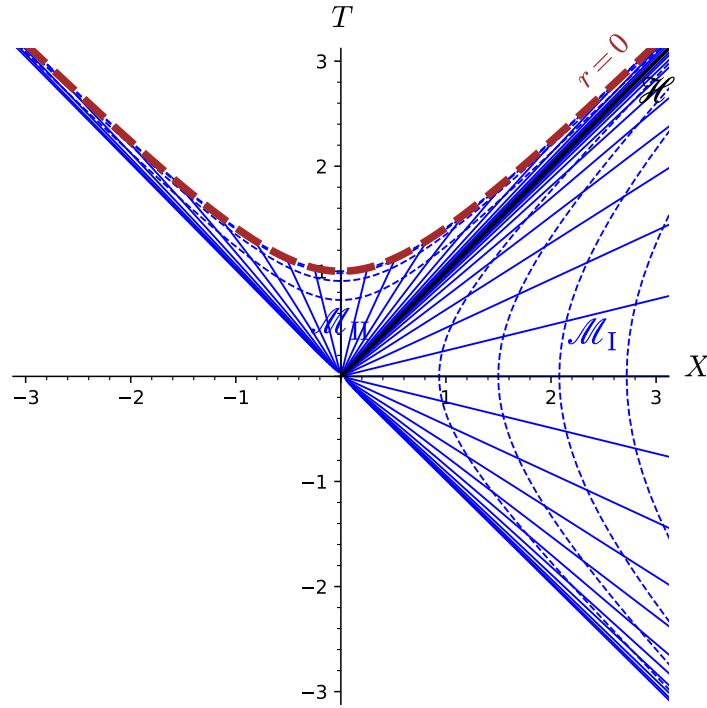


Figure 9.6: Schwarzschild-Droste coordinates in $\mathcal{M}_{\text{SD}} = \mathcal{M}_I \cup \mathcal{M}_{\text{II}}$ depicted in terms of the Kruskal-Szekeres coordinates (T, X) : the solid blue curves have $t = \text{const}$ (spaced apart by $\delta t = m$), while the dashed blue curves have $r = \text{const}$ (spaced apart by $\delta r = m/2$). [Figure generated by the notebook D.4.16]

so that (9.22) can be rewritten as

$$\mathcal{M}_{\text{II}} : \begin{cases} X + T &= e^{(t+r)/4m} \sqrt{1 - \frac{r}{2m}} \\ X - T &= -e^{(r-t)/4m} \sqrt{1 - \frac{r}{2m}}. \end{cases}$$

We obtain then

$$\mathcal{M}_{\text{II}} : \begin{cases} T = e^{r/4m} \sqrt{1 - \frac{r}{2m}} \cosh \left(\frac{t}{4m} \right) \\ X = e^{r/4m} \sqrt{1 - \frac{r}{2m}} \sinh \left(\frac{t}{4m} \right) \end{cases} \iff \begin{cases} t = 2m \ln \left(\frac{T+X}{T-X} \right) \\ r = 2m W(X^2 - T^2). \end{cases} \quad (9.25)$$

This is to be compared with (9.18). 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 \mathcal{M}_{II} .

As discussed in Sec. 6.3.4, one approaches a curvature singularity as $r \rightarrow 0$. According to (9.23) or (9.25), this corresponds to $X^2 - T^2 \rightarrow -1$ (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 $-dT^2 + dX^2$. Consequently the radial null geodesics are straight lines of slope $\pm 45^\circ$ in the (T, X) plane (cf. Fig. 9.7):

- the ingoing radial null geodesics obey

$$T = -X + V, \quad (9.26)$$

where V is a positive constant (the constraint $V > 0$ following from (9.21)), so that each geodesic of this family can be labelled by (V, θ, φ) ;

- the outgoing radial null geodesics obey

$$T = X + U, \quad (9.27)$$

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 \mathcal{H} is generated by the outgoing radial null geodesics having $U = 0$: Eqs. (9.27) and (9.11) clearly imply $r = 2m$ for $U = 0$, i.e. $X = T$. The outgoing radial null geodesics not lying on \mathcal{H} , which are denoted $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ in Sec. 6.3.5, have an equation in terms of the IEF coordinates given by Eq. (6.51): $\tilde{t} = r + 4m \ln |r/2m - 1| + u$, where the constant u is related to U by

$$U = -e^{-u/4m} \quad \text{on } \mathcal{M}_I \quad (9.28a)$$

$$U = 0 \quad \text{on } \mathcal{H} \quad (9.28b)$$

$$U = e^{-u/4m} \quad \text{on } \mathcal{M}_{II}. \quad (9.28c)$$

These relations are easily established by combining (6.51) and (9.23).

Remark 3: The relation $U = -e^{-u/4m}$ introduced in Sec. 9.2.1 by Eq. (9.4) is thus valid only in \mathcal{M}_I . On the contrary the relation $V = e^{v/4m}$ is valid in all \mathcal{M}_{IEF} .

9.3 Maximal extension

9.3.1 Construction

The spacetime $(\mathcal{M}_{\text{IEF}}, g)$ 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 \mathcal{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 to negative values of the

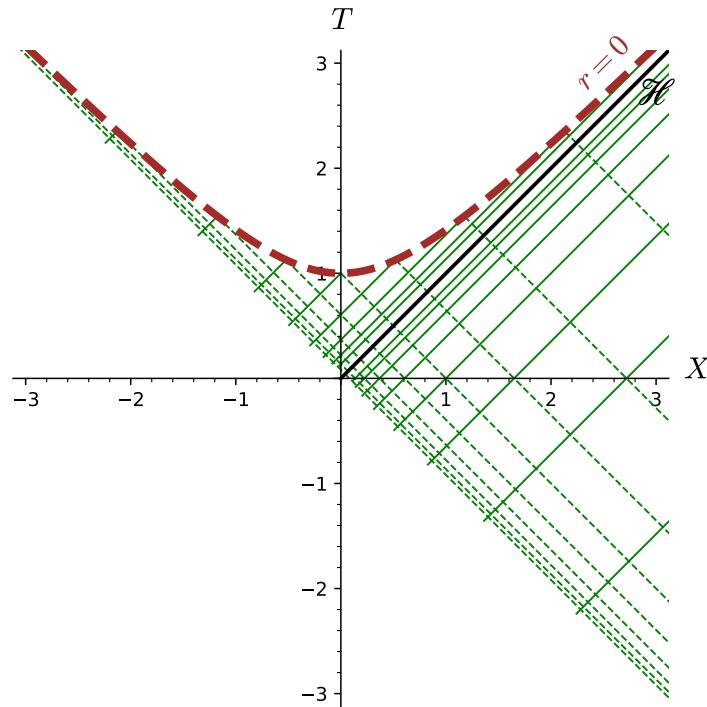


Figure 9.7: Radial null geodesics in $\mathcal{M}_{\text{IEF}} = \mathcal{M}_I \cup \mathcal{H} \cup \mathcal{M}_{II}$ depicted in terms of the Kruskal-Szekeres coordinates (T, X) : the solid lines correspond to the outgoing family, with u spanning $[-6m, 8m]$ (with steps $\delta u = 2m$), from the left to the right in \mathcal{M}_{II} and from the right to the left in \mathcal{M}_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.4.16](#)]

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 \mathcal{M}_I , with r increasing towards the future, and to $r < 2m$ in \mathcal{M}_{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 \mathcal{H} , since for the latter $T = X$, but rather to $\tilde{t} \rightarrow -\infty$, as it is clear when comparing Fig. 9.7 with Fig. 9.5.

Another hint regarding the extendability of $(\mathcal{M}_{\text{IEF}}, g)$ is the fact that the Killing horizon \mathcal{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. 3.4 that non-degenerate Killing horizons have incomplete null generators and, if they can be extended, they must be part of a bifurcate Killing horizon (Property 3.15). In the present case, the null generators of \mathcal{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 \mathcal{M}_{IEF} is clear on the metric element (9.16): it is invariant by the transformation

$$\begin{aligned}\Phi : \quad \mathbb{R}^2 &\longrightarrow \quad \mathbb{R}^2 \\ (T, X) &\longmapsto (-T, -X).\end{aligned}\tag{9.29}$$

Thus we may include the part $T + X < 0$ by adding a copy of \mathcal{M}_{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 $\mathbb{R}^2 \times \mathbb{S}^2$:

$$\mathcal{M} := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad T^2(p) - X^2(p) < 1 \right\},\tag{9.30}$$

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 g on the whole \mathcal{M} is then defined by (9.16):

$$g = 4m^2 \left[\frac{4}{X^2 - T^2 + e^{W(X^2 - T^2)}} (-dT^2 + dX^2) + W(X^2 - T^2)^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right],\tag{9.31}$$

where W is the rescaled Lambert function defined by (9.14) (cf. Fig. 9.4); it is the inverse of the function $x \mapsto e^x(x - 1)$, which establishes a bijection from $(0, +\infty)$ to $(-1, +\infty)$. Note that the metric components $g_{\alpha\beta}$ appearing in Eq. (9.31) are regular functions of (T, X) in all \mathcal{M} . In particular, the denominator $X^2 - T^2 + e^{W(X^2 - T^2)}$ of g_{TT} and g_{XX} never vanishes for $X^2 - T^2 > -1$, which is the range of variation of (T, X) on \mathcal{M} according to Eq. (9.30). For $|T| = |X|$, and in particular at $(T, X) = (0, 0)$, one has $g_{XX} = -g_{TT} = 16m^2/e$, given that $W(0) = 1$ (cf. Fig. 9.4),

Let us define the following open subsets of \mathcal{M} , which are respectively the images of \mathcal{M}_{I} and \mathcal{M}_{II} by the reflection through the origin (9.29):

$$\mathcal{M}_{\text{III}} : \quad X < 0 \quad \text{and} \quad X < T < -X\tag{9.32a}$$

$$\mathcal{M}_{\text{IV}} : \quad -\sqrt{X^2 + 1} < T < -|X|.\tag{9.32b}$$

On $\mathcal{M}_{\text{III}} \cup \mathcal{M}_{\text{IV}}$, one may introduce coordinates $(t', r', \theta, \varphi)$ of Schwarzschild-Droste type; they are related to the Kruskal-Szekeres coordinates by formulas analogous to (9.18) and (9.25), simply changing T to $-T$ and X to $-X$:

$$\mathcal{M}_{\text{III}} : \quad \begin{cases} T = -e^{r'/4m} \sqrt{\frac{r'}{2m} - 1} \sinh \left(\frac{t'}{4m} \right) \\ X = -e^{r'/4m} \sqrt{\frac{r'}{2m} - 1} \cosh \left(\frac{t'}{4m} \right) \end{cases} \iff \begin{cases} t' = 2m \ln \left(\frac{X+T}{X-T} \right) \\ r' = 2m W(X^2 - T^2). \end{cases}\tag{9.33}$$

$$\mathcal{M}_{\text{IV}} : \quad \begin{cases} T = -e^{r'/4m} \sqrt{1 - \frac{r'}{2m}} \cosh \left(\frac{t'}{4m} \right) \\ X = -e^{r'/4m} \sqrt{1 - \frac{r'}{2m}} \sinh \left(\frac{t'}{4m} \right) \end{cases} \iff \begin{cases} t' = 2m \ln \left(\frac{T+X}{T-X} \right) \\ r' = 2m W(X^2 - T^2). \end{cases}\tag{9.34}$$

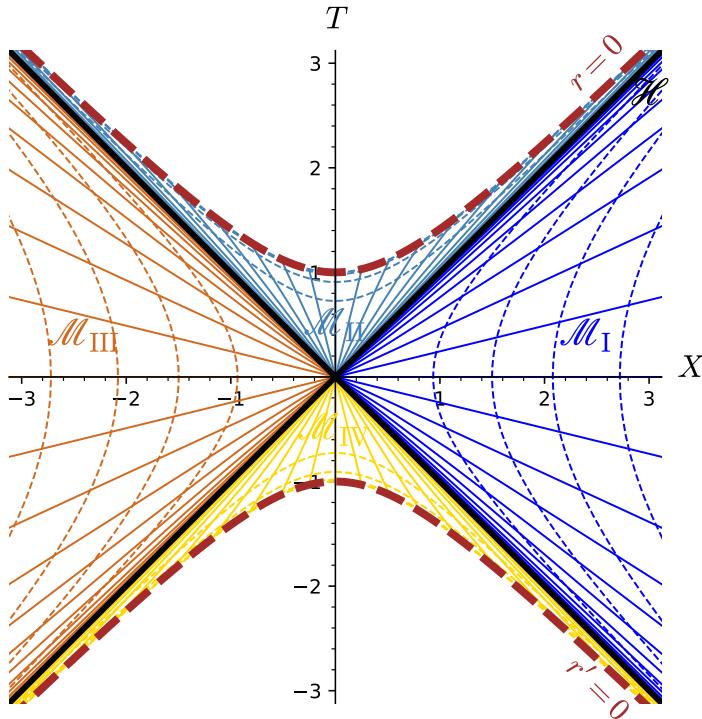


Figure 9.8: *Kruskal diagram*: Schwarzschild spacetime \mathcal{M} depicted in terms of Kruskal-Szekeres coordinates (T, X) . Each point in this diagram, including the one at $(T, X) = (0, 0)$, is actually a sphere \mathbb{S}^2 , spanned by the coordinates (θ, φ) . Solid lines denote the hypersurfaces $t = \text{const}$ in \mathcal{M}_I and \mathcal{M}_{II} and the hypersurfaces $t' = \text{const}$ in \mathcal{M}_{III} and \mathcal{M}_{IV} , while dashed curves denote the hypersurfaces $r = \text{const}$ in \mathcal{M}_I and \mathcal{M}_{II} and the hypersurfaces $r' = \text{const}$ in \mathcal{M}_{III} and \mathcal{M}_{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.4.16](#)]

The extended Schwarzschild spacetime (\mathcal{M}, g) is depicted in Fig. 9.8, which is usually called a **Kruskal diagram**. There are two curvature singularities, which formally are not part of \mathcal{M} : the hypersurfaces $r = 0$ and $r' = 0$, which are the two branches of the hyperbola $T^2 - X^2 = 1$. As discussed in Sec. 9.2.3, the radial null geodesics appear as straight lines of slope $\pm 45^\circ$ (+ for the outgoing family, and - for the ingoing one). As in $(\mathcal{M}_{\text{IEF}}, g)$, 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, (\mathcal{M}, g) is called the **maximal extension** of Schwarzschild spacetime.

Remark 1: The extended manifold \mathcal{M} is not just the union $\mathcal{M}_I \cup \mathcal{M}_{II} \cup \mathcal{M}_{III} \cup \mathcal{M}_{IV}$, since the latter does not contain the hypersurfaces $T = \pm X$ (cf. the strict inequalities in Eqs. (9.24) and (9.32)), which are parts of \mathcal{M} according to the definition (9.30). Actually, we have

$$\mathcal{M} = \mathcal{M}_I \cup \mathcal{M}_{II} \cup \mathcal{M}_{III} \cup \mathcal{M}_{IV} \cup \hat{\mathcal{H}}, \quad (9.35)$$

where $\hat{\mathcal{H}}$ 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 \mathcal{M}_{IEF} the null coordinates (U, V) ; they are related to the coordinates (T, X) by Eq. (9.8) (or equivalently Eqs. (9.26)-(9.27)), which we can use to define (U, V) in all the maximal extension \mathcal{M} :

$$\begin{cases} U = T - X \\ V = T + X \end{cases} \iff \begin{cases} T = \frac{1}{2}(U + V) \\ X = \frac{1}{2}(V - U) \end{cases} \quad (9.36)$$

The range of (U, V) of \mathcal{M} is deduced from the constraint $T^2 - X^2 < 1$ [cf. Eq. (9.30)]: since $T^2 - X^2 = UV$, we get:

$$\mathcal{M} : \quad (U, V) \in \mathbb{R}^2 \quad \text{and} \quad UV < 1. \quad (9.37)$$

The expression of the metric with respect to the null coordinates $X^{\hat{\alpha}} = (U, V, \theta, \varphi)$ is deduced from (9.31):

$$\boxed{\mathbf{g} = 4m^2 \left[\frac{4}{UV - e^{W(-UV)}} dU dV + W(-UV)^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right].} \quad (9.38)$$

We can also rewrite it as (9.5):

$$\boxed{\mathbf{g} = -\frac{32m^3}{r} e^{-r/2m} dU dV + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2),} \quad (9.39)$$

where r is the function of (U, V) given by

$$r = 2m W(-UV). \quad (9.40)$$

Note that relation (9.6) between r and (U, V) holds in all \mathcal{M} :

$$\boxed{e^{r/2m} \left(\frac{r}{2m} - 1 \right) = -UV}. \quad (9.41)$$

Remark 2: In Sec. 9.3.1, we have distinguished the coordinate r in $\mathcal{M}_I \cup \mathcal{M}_{II}$ from the coordinate r' in $\mathcal{M}_{III} \cup \mathcal{M}_{IV}$. Here, r is the function (9.40) of (U, V) , which has the same expression in $\mathcal{M}_I \cup \mathcal{M}_{II}$ and $\mathcal{M}_{III} \cup \mathcal{M}_{IV}$. There is no need to make any distinction. Hence there is no mention of r' in (9.39).

Historical note : The Kruskal-Szekeres coordinates have been introduced in 1960 independently by Martin Kruskal [325] and George Szekeres [471]. Actually the coordinates introduced by Szekeres were² $(2T/\sqrt{e}, 2X/\sqrt{e})$. 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. [325]). However, the first author to obtain the maximal extension of Schwarzschild spacetime is John Synge in 1950 [468], i.e. 10

²They are denoted by (v, u) in Szekeres' article [471].

years before Kruskal and Szekeres. Synge used coordinates (T', X') whose relation to Schwarzschild-Droste coordinates is more complicated than the Kruskal-Szekeres one: $T' = R(r) \sinh(\frac{t}{4m})$ and $X' = R(r) \cosh(\frac{t}{4m})$, with $R(r) := 2m \left[\operatorname{arcosh} \sqrt{\frac{r}{2m}} + \sqrt{\frac{r}{2m} (\frac{r}{2m} - 1)} \right]$; compare with (9.18). Albeit looking complicated, $R(r)$ is nothing but the primitive vanishing at $r = 2m$ of $r \mapsto (\frac{r}{2m} - 1)^{-1/2}$. Interestingly, in his article [471], Szekeres says that the transformations (9.18) “are essentially due to Synge”, probably because they differ only in the choice of the function $R(r)$, the latter being $R_{\text{KS}}(r) = e^{r/4m} \sqrt{r/2m - 1}$ for Kruskal-Szekeres coordinates. For this reason, both coordinate systems share some similarities: in Synge diagram (Figs. 8 and 9 in Ref. [468]), 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 $\pm 45^\circ$ slope. Even, in some regions, the coordinate X' ceases to be spacelike³. The maximal extension of Schwarzschild spacetime has also been found by Christian Fronsdal [204] 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. 9.3.1, the Schwarzschild horizon \mathcal{H} is a non-degenerate Killing horizon and therefore shall be part of a bifurcate Killing horizon (cf. Sec. 3.4) in the extended spacetime. The bifurcate Killing horizon, $\hat{\mathcal{H}}$ say, is easily found by considering the Killing vector field ξ in the maximal extension of Schwarzschild spacetime. The components of ξ w.r.t. to the Kruskal-Szekeres coordinates are obtained from the property $\xi = \partial_t$:

$$\xi^T = \frac{\partial T}{\partial t}, \quad \xi^X = \frac{\partial X}{\partial t}, \quad \xi^\theta = \frac{\partial \theta}{\partial t} = 0, \quad \xi^\varphi = \frac{\partial \varphi}{\partial t} = 0.$$

Given the coordinate transformation laws (9.18) and (9.25), we get in \mathcal{M}_I and \mathcal{M}_{II} :

$$\xi^T = \frac{1}{4m} X, \quad \xi^X = \frac{1}{4m} T, \quad \xi^\theta = \xi^\varphi = 0.$$

Hence in $\mathcal{M}_I \cup \mathcal{M}_{II}$,

$$\boxed{\xi = \frac{1}{4m} (X \partial_T + T \partial_X)}.$$

(9.42)

Now, this formula defines a smooth vector field in all \mathcal{M} . Moreover, in $\mathcal{M}_{III} \cup \mathcal{M}_{IV}$, this vector coincides with $\partial_{t'}$ since $\xi^T = \partial T / \partial t'$ and $\xi^X = \partial X / \partial t'$, with the partial derivatives with respect to t' evaluated from (9.33)-(9.34). Hence the vector field ξ defined by (9.42) is a Killing vector field of maximal extension (\mathcal{M}, g) . This vector field is depicted in Fig. 9.9.

The bifurcate Killing horizon with respect to ξ that extends \mathcal{H} is $\hat{\mathcal{H}} = \mathcal{H}_1 \cup \mathcal{H}_2$, where

- \mathcal{H}_1 is the null hypersurface $T = X$ (or equivalently $U = 0$);
- \mathcal{H}_2 is the null hypersurface $T = -X$ (or equivalently $V = 0$).

³We refer the reader to Fig. 2 of Ref. [484] for a plot of Synge coordinates in terms of Kruskal-Szekeres ones

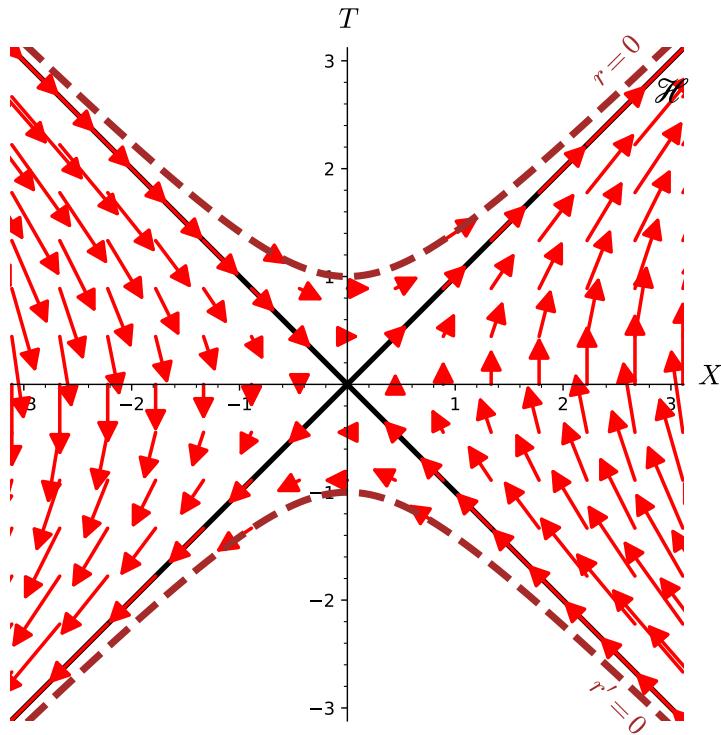


Figure 9.9: Killing vector field ξ on the extended Schwarzschild manifold. [Figure generated by the notebook D.4.16]

The bifurcate Killing horizon $\hat{\mathcal{H}}$ is depicted in black in Fig. 9.9. The Schwarzschild horizon \mathcal{H} is the part of $\hat{\mathcal{H}}$ defined by $X > 0$. In terms of the null coordinates (U, V) introduced in Sec. 9.3.2, we have, given (9.36),

$$\hat{\mathcal{H}} : \quad U = 0 \quad \text{or} \quad V = 0 \quad (9.43a)$$

$$\mathcal{H} : \quad U = 0 \quad \text{and} \quad V > 0. \quad (9.43b)$$

The bifurcation surface is $\mathcal{S} = \mathcal{H}_1 \cap \mathcal{H}_2$, 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 \mathcal{M} as a part of $\mathbb{R}^2 \times \mathbb{S}^2$ [cf. Eq. (9.30)]. Accordingly, \mathcal{S} is called the **bifurcation sphere**. It is located at the center of Fig. 9.9. The areal radius of \mathcal{S} is found by setting $dT = 0$, $dX = 0$ and $(T, X) = (0, 0)$ in the metric expression (9.31): $r_{\mathcal{S}}^2 = 4m^2W(0)^2$. Since $W(0) = 1$ (cf. Fig. 9.4), we get

$$r_{\mathcal{S}} = 2m. \quad (9.44)$$

Moreover, setting $(T, X) = (0, 0)$ in Eq. (9.42), we recover Property 3.12 (Sec. 3.4.1): the Killing vector field vanishes at the bifurcation sphere:

$$\xi|_{\mathcal{S}} = 0. \quad (9.45)$$

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}) :

$$\begin{cases} \hat{U} = \arctan U \\ \hat{V} = \arctan V \end{cases} \iff \begin{cases} U = \tan \hat{U} \\ V = \tan \hat{V}. \end{cases} \quad (9.46)$$

The range of (\hat{U}, \hat{V}) is deduced from (9.37):

$$UV < 1 \iff \tan \hat{U} \tan \hat{V} < 1.$$

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:

$$\mathcal{M} : -\frac{\pi}{2} < \hat{U} < \frac{\pi}{2}, \quad -\frac{\pi}{2} < \hat{V} < \frac{\pi}{2} \quad \text{and} \quad -\frac{\pi}{2} < \hat{U} + \hat{V} < \frac{\pi}{2}. \quad (9.47)$$

Since (9.46) yields $dU = d\hat{U}/\cos^2 \hat{U}$ and $dV = d\hat{V}/\cos^2 \hat{V}$, we deduce immediately from (9.39) the expression of the metric in terms of the coordinates $x^\alpha = (\hat{U}, \hat{V}, \theta, \varphi)$:

$$g = -\frac{32m^3}{r} e^{-r/2m} \frac{d\hat{U}}{\cos^2 \hat{U}} \frac{d\hat{V}}{\cos^2 \hat{V}} + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2), \quad (9.48)$$

where [cf. Eq. (9.40)]

$$r = 2m W(-\tan \hat{U} \tan \hat{V}). \quad (9.49)$$

To depict \mathcal{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. (4.14)]:

$$\begin{cases} \hat{T} = \hat{U} + \hat{V} \\ \hat{X} = \hat{V} - \hat{U} \end{cases} \iff \begin{cases} \hat{U} = \frac{1}{2}(\hat{T} - \hat{X}) \\ \hat{V} = \frac{1}{2}(\hat{T} + \hat{X}). \end{cases} \quad (9.50)$$

The range of (\hat{T}, \hat{X}) is deduced from (9.47):

$$\mathcal{M} : -\frac{\pi}{2} < \hat{T} < \frac{\pi}{2}, \quad \hat{T} - \pi < \hat{X} < \hat{T} + \pi \quad \text{and} \quad -\hat{T} - \pi < \hat{X} < -\hat{T} + \pi. \quad (9.51)$$

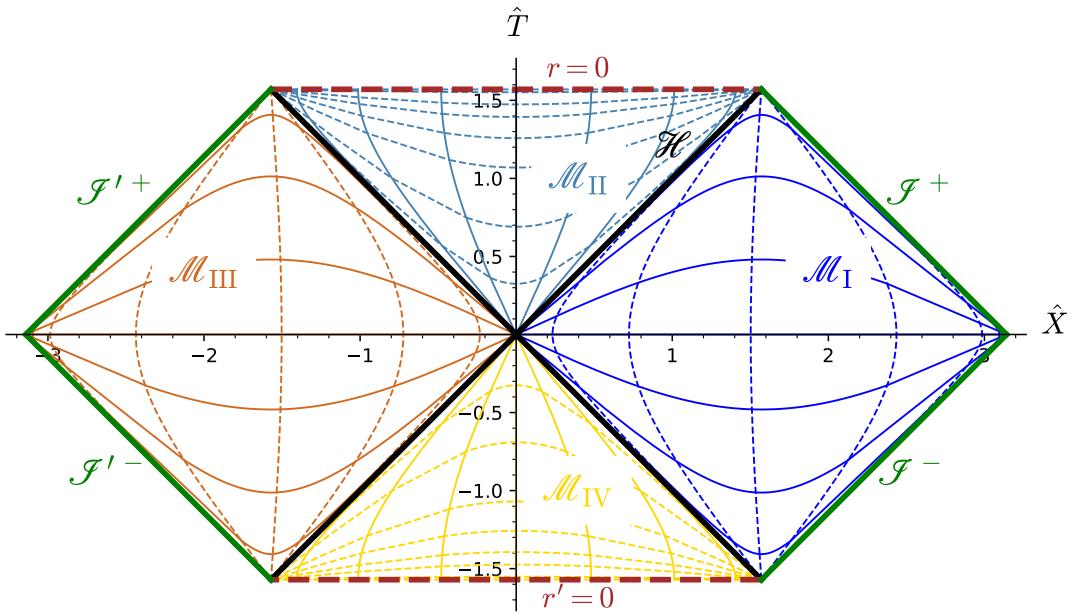


Figure 9.10: Carter-Penrose diagram of the Schwarzschild spacetime constructed with the compactified coordinates (\hat{T}, \hat{X}) . Solid curves denote hypersurfaces of constant Schwarzschild-Droste coordinate t : in region \mathcal{M}_I , from the \hat{X} -axis to the top: $t = 0, 2m, 5m, 10m, 20m$ and $50m$, the last two being barely visible; in region \mathcal{M}_{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 \mathcal{M}_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 \mathcal{M}_{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.4.17]

Via (9.46) and (9.36), 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 Minkowski spacetime [Eq. (4.16)]:

$$\left\{ \begin{array}{l} \hat{T} = \arctan(T + X) + \arctan(T - X) \\ \hat{X} = \arctan(T + X) - \arctan(T - X) \end{array} \right. \iff \left\{ \begin{array}{l} T = \frac{\sin \hat{T}}{\cos \hat{T} + \cos \hat{X}} \\ X = \frac{\sin \hat{X}}{\cos \hat{T} + \cos \hat{X}}. \end{array} \right. \quad (9.52)$$

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. 298). As the Kruskal diagram (Fig. 9.8), it has the feature of displaying radial null geodesics as straight lines with slope $\pm 45^\circ$. This holds since \hat{U} (resp. \hat{V}) is a function of U only (resp. V only), cf. Eq. (9.46), 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} :

$$\hat{\mathcal{H}} : \hat{U} = 0 \quad \text{or} \quad \hat{V} = 0 \quad (9.53a)$$

$$\hat{\mathcal{H}} : \hat{U} = 0 \quad \text{and} \quad \hat{V} > 0. \quad (9.53b)$$

These relations follow immediately from (9.43) and (9.46).

We have seen in Sec. 6.4 that the future null infinity \mathcal{I}^+ corresponds to $v \rightarrow +\infty$ and that the past null infinity \mathcal{I}^- to $u \rightarrow -\infty$ (cf. Fig. 6.6). Since on \mathcal{M}_I , $U = -e^{-u/4m}$ and $V = e^{v/4m}$ [cf. Eq. (9.4)], we may write equivalently:

$$\mathcal{I}^+ : V \rightarrow +\infty \quad \text{and} \quad U \in (-\infty, 0) \quad (9.54a)$$

$$\mathcal{I}^- : U \rightarrow -\infty \quad \text{and} \quad V \in (0, +\infty). \quad (9.54b)$$

In view of (9.46), we get then:

$$\mathcal{I}^+ : \hat{V} \rightarrow \frac{\pi}{2} \quad \text{and} \quad \hat{U} \in \left(-\frac{\pi}{2}, 0\right) \quad (9.55a)$$

$$\mathcal{I}^- : \hat{U} \rightarrow -\frac{\pi}{2} \quad \text{and} \quad \hat{V} \in \left(0, \frac{\pi}{2}\right). \quad (9.55b)$$

By symmetry, the extension $\mathcal{M}_{III} \cup \mathcal{M}_{IV}$ of Schwarzschild spacetime has the following null infinity:

$$\mathcal{I}'^+ : \hat{U} \rightarrow \frac{\pi}{2} \quad \text{and} \quad \hat{V} \in \left(-\frac{\pi}{2}, 0\right) \quad (9.56a)$$

$$\mathcal{I}'^- : \hat{V} \rightarrow -\frac{\pi}{2} \quad \text{and} \quad \hat{U} \in \left(0, \frac{\pi}{2}\right). \quad (9.56b)$$

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 (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 (9.48) in a form that makes clear their behavior near null infinity. Given (9.49) and (9.15), we have

$$e^{r/2m} \left(\frac{r}{2m} - 1 \right) = -\tan \hat{U} \tan \hat{V}, \quad (9.57)$$

from which we get

$$\frac{2m}{r} e^{-r/2m} = -\frac{1 - 2m/r}{\tan \hat{U} \tan \hat{V}}.$$

Hence

$$\frac{2m}{r} \frac{e^{-r/2m}}{\cos^2 \hat{U} \cos^2 \hat{V}} = -\frac{1 - 2m/r}{\sin \hat{U} \cos \hat{U} \sin \hat{V} \cos \hat{V}} = -\frac{4(1 - 2m/r)}{\sin 2\hat{U} \sin 2\hat{V}}.$$

Therefore, we may rewrite expression (9.48) for the metric tensor as

$$\mathbf{g} = 64m^2 \left(1 - \frac{2m}{r}\right) \frac{d\hat{U} d\hat{V}}{\sin 2\hat{U} \sin 2\hat{V}} + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2), \quad (9.58)$$

with r given by (9.49). To get a conformal completion, we should write (cf. Sec. 4.3)

$$\mathbf{g} = \Omega^{-2} \tilde{\mathbf{g}}, \quad (9.59)$$

where $\Omega = 0$ and $d\Omega \neq 0$ on the spacetime boundary \mathcal{I} and $\tilde{\mathbf{g}}$ is a regular metric on the completion $\mathcal{M} \cup \mathcal{I}$. Since in Eq. (9.58), the term $\sin 2\hat{U} \sin 2\hat{V}$ vanishes at $\mathcal{I} = \mathcal{I}^+ \cup \mathcal{I}^- \cup \mathcal{I}'^+ \cup \mathcal{I}'^-$ [cf. Eqs. (9.55)-(9.56)], we would have, up to some constant factor,

$$\Omega = \sqrt{-\sin 2\hat{U} \sin 2\hat{V}}, \quad (9.60)$$

the minus sign taking into account that $\sin 2\hat{U} \sin 2\hat{V}$ approaches zero via negative values near \mathcal{I} . A first issue is that the square root in (9.60) makes Ω not differentiable on \mathcal{I} , where either $\sin 2\hat{U} = 0$ or $\sin 2\hat{V} = 0$. In other words, $d\Omega$ is diverging on \mathcal{I} . Suppose we accept this and are ready to introduce a slight deviation (given that Ω^2 , which is involved in (9.59), is smooth) from the definition given in Sec. 4.3. Then the conformal metric should be

$$\tilde{\mathbf{g}} = -64m^2 \left(1 - \frac{2m}{r}\right) d\hat{U} d\hat{V} - r^2 \sin 2\hat{U} \sin 2\hat{V} (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (9.61)$$

Near \mathcal{I} , $r \rightarrow +\infty$ and we have $\tilde{g}_{\hat{U}\hat{V}} \rightarrow -32m^2$. 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 (9.57):

$$\cos \hat{U} \cos \hat{V} = -\sin \hat{U} \sin \hat{V} \frac{e^{-r/2m}}{r/2m - 1}.$$

Hence

$$\tilde{g}_{\theta\theta} = 8m \sin^2 \hat{U} \sin^2 \hat{V} \frac{re^{-r/2m}}{1 - 2m/r}$$

and (9.61) becomes

$$\tilde{\mathbf{g}} = -64m^2 \left(1 - \frac{2m}{r}\right) d\hat{U} d\hat{V} + 8m \sin^2 \hat{U} \sin^2 \hat{V} \frac{re^{-r/2m}}{1 - 2m/r} (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (9.62)$$

On this expression, we can read directly the value of the conformal metric at \mathcal{I} , where $r \rightarrow +\infty$, $2m/r \rightarrow 0$, $re^{-r/2m} \rightarrow 0$ and $\sin^2 \hat{U} \rightarrow 1$ or $\sin^2 \hat{V} \rightarrow 1$:

$$\tilde{\mathbf{g}} \stackrel{\mathcal{I}}{=} -64m^2 d\hat{U} d\hat{V}. \quad (9.63)$$

This bilinear form is clearly degenerate (cf. Sec. A.3.1). Therefore, $\tilde{\mathbf{g}}$ is not a regular metric on the whole manifold $\mathcal{M} \cup \mathcal{I}$. We conclude that (9.60)-(9.61) does not define a conformal completion of $(\mathcal{M}, \mathbf{g})$.

Historical note : The first compactified conformal diagram of the (maximal extension of) Schwarzschild spacetime has been constructed by Brandon Carter in 1966 [87], using the same coordinates (\hat{T}, \hat{X}) as here⁴: compare Fig. 9.10 with Fig. 1c of Ref. [87]. In his article, Carter notes that “the manner in which

⁴ (\hat{T}, \hat{X}) are denoted (ψ, ξ) by Carter [87].

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 [403], the diagrams in Penrose's article [403] regarding Minkowski and de Sitter spacetimes only. This justifies the name *Carter-Penrose diagram* used in the literature (e.g. [203]) for the graphical representation shown in Fig. 9.10, while the mere *Penrose diagram* or *conformal diagram* is quite common; one encounters as well the name *Penrose-Carter diagram* (e.g. [415]).

9.4.3 A regular conformal completion based on Penrose-Frolov-Novikov coordinates

In order to get a regular conformal completion of the maximally extended Schwarzschild spacetime (\mathcal{M}, g) , let us introduce the finite-range coordinates (\tilde{U}, \tilde{V}) that are related to the null Kruskal-Szekeres coordinates (U, V) by

$$\begin{cases} \tilde{U} = \arctan(\text{arsinh } U) \\ \tilde{V} = \arctan(\text{arsinh } V) \end{cases} \iff \begin{cases} U = \sinh(\tan \tilde{U}) \\ V = \sinh(\tan \tilde{V}). \end{cases} \quad (9.64)$$

The range of (\tilde{U}, \tilde{V}) is deduced from (9.37):

$$\mathcal{M} : -\frac{\pi}{2} < \tilde{U} < \frac{\pi}{2}, \quad -\frac{\pi}{2} < \tilde{V} < \frac{\pi}{2} \quad \text{and} \quad \sinh(\tan \tilde{U}) \sinh(\tan \tilde{V}) < 1. \quad (9.65)$$

Note that contrary to what happened for (\hat{U}, \hat{V}) , these conditions do not yield to a simple polygonal region in the (\tilde{U}, \tilde{V}) plane. The presence of the sinh function in the expression (9.64) 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. (9.55)-(9.56)]:

$$\mathcal{I}^+ : \tilde{V} \rightarrow \frac{\pi}{2} \quad \text{and} \quad \tilde{U} \in \left(-\frac{\pi}{2}, 0\right) \quad (9.66a)$$

$$\mathcal{I}^- : \tilde{U} \rightarrow -\frac{\pi}{2} \quad \text{and} \quad \tilde{V} \in \left(0, \frac{\pi}{2}\right) \quad (9.66b)$$

$$\mathcal{I}'^+ : \tilde{U} \rightarrow \frac{\pi}{2} \quad \text{and} \quad \tilde{V} \in \left(-\frac{\pi}{2}, 0\right) \quad (9.66c)$$

$$\mathcal{I}'^- : \tilde{V} \rightarrow -\frac{\pi}{2} \quad \text{and} \quad \tilde{U} \in \left(0, \frac{\pi}{2}\right). \quad (9.66d)$$

We shall call $(\tilde{U}, \tilde{V}, \theta, \varphi)$ the **Penrose-Frolov-Novikov coordinates** (cf. the historical note on p. 304). From (9.64), we get

$$dU = \frac{\cosh(\tan \tilde{U})}{\cos^2 \tilde{U}} d\tilde{U} \quad \text{and} \quad dV = \frac{\cosh(\tan \tilde{V})}{\cos^2 \tilde{V}} d\tilde{V},$$

so that the metric expression in terms of the coordinates $x^\alpha = (\tilde{U}, \tilde{V}, \theta, \varphi)$ is easily deduced from (9.39)-(9.40):

$$g = -\frac{32m^3}{r} e^{-r/2m} \frac{\cosh(\tan \tilde{U}) \cosh(\tan \tilde{V})}{\cos^2 \tilde{U} \cos^2 \tilde{V}} d\tilde{U} d\tilde{V} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (9.67)$$

where r is the function of (\tilde{U}, \tilde{V}) given by

$$r = 2m W \left(-\sinh(\tan \tilde{U}) \sinh(\tan \tilde{V}) \right). \quad (9.68)$$

As we did for (\hat{U}, \hat{V}) , let us rewrite (9.67) in a form that is better adapted to the null asymptotics. Given (9.68) and (9.15), we have

$$e^{r/2m} \left(\frac{r}{2m} - 1 \right) = -\sinh(\tan \tilde{U}) \sinh(\tan \tilde{V}), \quad (9.69)$$

from which we get

$$\frac{2m}{r} e^{-r/2m} = - \left(1 - \frac{2m}{r} \right) \frac{1}{\sinh(\tan \tilde{U}) \sinh(\tan \tilde{V})}.$$

Hence (9.67) becomes

$$\begin{aligned} g &= 16m^2 \left(1 - \frac{2m}{r} \right) \frac{d\tilde{U} d\tilde{V}}{\tanh(\tan \tilde{U}) \tanh(\tan \tilde{V}) \cos^2 \tilde{U} \cos^2 \tilde{V}} \\ &\quad + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \end{aligned} \quad (9.70)$$

Given the values (9.66) of \tilde{U} and \tilde{V} near \mathcal{I} , $\tanh(\tan \tilde{U}) \tanh(\tan \tilde{V})$ does not vanish there. A natural choice of conformal factor is then

$$\boxed{\Omega := \cos \tilde{U} \cos \tilde{V}}. \quad (9.71)$$

The corresponding conformal metric is

$$\boxed{\begin{aligned} \tilde{g} &= 16m^2 \left(1 - \frac{2m}{r} \right) \frac{d\tilde{U} d\tilde{V}}{\tanh(\tan \tilde{U}) \tanh(\tan \tilde{V})} \\ &\quad + r^2 \cos^2 \tilde{U} \cos^2 \tilde{V} (d\theta^2 + \sin^2 \theta d\varphi^2) \end{aligned}}. \quad (9.72)$$

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

$$\begin{aligned} \tilde{\mathcal{M}} &:= \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, (\tilde{U}(p), \tilde{V}(p)) \in \left(-\frac{\pi}{2}, \frac{\pi}{2} \right)^2 \text{ and } \sinh(\tan \tilde{U}(p)) \sinh(\tan \tilde{V}(p)) < 1 \right\} \\ &\quad \cup \mathcal{I}^+ \cup \mathcal{I}^- \cup \mathcal{I}'^+ \cup \mathcal{I}'^-, \end{aligned} \quad (9.73)$$

with

$$\mathcal{I}^+ := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \tilde{V}(p) = \frac{\pi}{2} \text{ and } \tilde{U}(p) \in \left(-\frac{\pi}{2}, 0 \right) \right\} \quad (9.74a)$$

$$\mathcal{I}^- := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \tilde{U}(p) = -\frac{\pi}{2} \text{ and } \tilde{V}(p) \in \left(0, \frac{\pi}{2} \right) \right\} \quad (9.74b)$$

$$\mathcal{I}'^+ := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \tilde{U}(p) = \frac{\pi}{2} \text{ and } \tilde{V}(p) \in \left(-\frac{\pi}{2}, 0 \right) \right\} \quad (9.74c)$$

$$\mathcal{I}'^- := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \tilde{V}(p) = -\frac{\pi}{2} \text{ and } \tilde{U}(p) \in \left(0, \frac{\pi}{2} \right) \right\}. \quad (9.74d)$$

Note that the first line in (9.73) corresponds to \mathcal{M} , identified as a subset of $\mathbb{R}^2 \times \mathbb{S}^2$ [cf. Eq. (9.65)] and that the definitions of \mathcal{I}^+ , \mathcal{I}^- , \mathcal{I}'^+ and \mathcal{I}'^- are in agreement with (9.66). It is clear that $\tilde{\mathcal{M}}$ is a manifold with boundary and that

$$\partial \tilde{\mathcal{M}} = \mathcal{I} := \mathcal{I}^+ \cup \mathcal{I}^- \cup \mathcal{I}'^+ \cup \mathcal{I}'^-.$$
 (9.75)

Moreover, the scalar field Ω defined by (9.71) satisfies $\Omega \geq 0$ on $\tilde{\mathcal{M}}$, along with $\Omega = 0$ on \mathcal{I} and $d\Omega \neq 0$ on \mathcal{I} . The last property follows from

$$d\Omega = -\sin \tilde{U} \cos \tilde{V} d\tilde{U} - \cos \tilde{U} \sin \tilde{V} d\tilde{V},$$

which implies $d\Omega|_{\mathcal{I}^+} = -\cos \tilde{U} d\tilde{V} \neq 0$, $d\Omega|_{\mathcal{I}^-} = \cos \tilde{V} d\tilde{U} \neq 0$, $d\Omega|_{\mathcal{I}'^+} = -\cos \tilde{V} d\tilde{U} \neq 0$ and $d\Omega|_{\mathcal{I}'^-} = \cos \tilde{U} d\tilde{V} \neq 0$. 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{g} defined by (9.72) is a regular metric on the whole $\tilde{\mathcal{M}}$. This was the main failing point in the attempt of Sec. 9.4.2. Since $\Omega^2 > 0$ on \mathcal{M} , \tilde{g} is well-behaved on \mathcal{M} . Let us thus examine its behavior on \mathcal{I} . We shall focus on \mathcal{I}^+ , the behavior on the other parts of \mathcal{I} being obtained by some trivial symmetry. As one approaches \mathcal{I}^+ , $r \rightarrow +\infty$, $\tilde{V} \rightarrow \pi/2$ and $\tanh(\tan(\tilde{V})) \rightarrow 1$; accordingly we read from (9.72) that

$$\tilde{g}_{\tilde{U}\tilde{V}} \stackrel{\mathcal{I}^+}{=} \frac{8m^2}{\tanh(\tan \tilde{U})}.$$

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 \mathcal{I}^+ . Noticing that $\tan \tilde{V} \sim 1/\cos \tilde{V}$ when $\tilde{V} \rightarrow \pi/2$, we get from (9.69)

$$\sinh\left(\frac{1}{\cos \tilde{V}}\right) \sim -\frac{re^{r/2m}}{2m \sinh(\tan \tilde{U})} \quad \text{when } \tilde{V} \rightarrow \frac{\pi}{2}.$$

Since $\text{arsinh } x = \ln(x + \sqrt{x^2 + 1}) \sim \ln(2x)$ when $x \rightarrow +\infty$, we obtain

$$\begin{aligned} \frac{1}{\cos \tilde{V}} &\sim \ln\left(-\frac{re^{r/2m}}{m \sinh(\tan \tilde{U})}\right) = \frac{r}{2m} + \ln\left(\frac{r}{m}\right) - \ln\left(-\sinh(\tan \tilde{U})\right) \\ &\sim \frac{r}{2m} \quad \text{when } \tilde{V} \rightarrow \frac{\pi}{2}. \end{aligned}$$

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

$$\tilde{g} \stackrel{\mathcal{I}^+}{=} 4m^2 \left[\frac{4}{\tanh(\tan \tilde{U})} d\tilde{U} d\tilde{V} + \cos^2 \tilde{U} (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2) \right].$$
 (9.76)

Since $\cos^2 \tilde{U} \neq 0$ on \mathcal{I}^+ [cf. Eq. (9.74a)], this bilinear form is non-degenerate. Moreover, since $\tanh(\tan \tilde{U}) < 0$ on \mathcal{I}^+ [again by (9.74a)], its signature is $(-, +, +, +)$. We conclude that \tilde{g} is a well-behaved metric on the whole manifold $\tilde{\mathcal{M}}$. This completes the demonstration of the following result:

Property 9.1: conformal completion of Schwarzschild spacetime

The pair $(\tilde{\mathcal{M}}, \tilde{g})$, with $\tilde{\mathcal{M}}$ defined by (9.73)-(9.74) and \tilde{g} defined by (9.72) is a conformal completion of the maximally extended Schwarzschild spacetime (\mathcal{M}, g) , the conformal factor being given by (9.71). The Penrose-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 (9.64).

Remark 1: At first sight, the metric \tilde{g} given by (9.72) looks degenerate at the bifurcate Killing horizon \mathcal{H} , since $1 - 2m/r = 0$ there. But one shall not forget that on \mathcal{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{g} near \mathcal{H} , it is more appropriate to use the form that is deduced from (9.67) and (9.71):

$$\tilde{g} = -\frac{32m^3}{r} e^{-r/2m} \cosh(\tan \tilde{U}) \cosh(\tan \tilde{V}) d\tilde{U} d\tilde{V} + r^2 \cos^2 \tilde{U} \cos^2 \tilde{V} (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (9.77)$$

Remark 2: The conformal completion constructed above cannot be analytically extended “beyond” \mathcal{I} , because the function $\tilde{V} \mapsto 1/\tanh(\tan \tilde{V})$, which appears in (9.72), 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 [253] for details.

To depict $\tilde{\mathcal{M}}$, let us introduce “time+space” coordinates (\tilde{T}, \tilde{X}) , which are related to (\hat{U}, \hat{V}) in exactly the same way as (\hat{T}, \hat{X}) are related to (\hat{U}, \hat{V}) [cf. Eq. (9.50)]:

$$\begin{cases} \tilde{T} = \tilde{U} + \tilde{V} \\ \tilde{X} = \tilde{V} - \tilde{U} \end{cases} \iff \begin{cases} \tilde{U} = \frac{1}{2}(\tilde{T} - \tilde{X}) \\ \tilde{V} = \frac{1}{2}(\tilde{T} + \tilde{X}). \end{cases} \quad (9.78)$$

The range of (\tilde{T}, \tilde{X}) is deduced from (9.65):

$$\begin{aligned} \mathcal{M} : \quad & -\pi < \tilde{T} - \tilde{X} < \pi, \quad -\pi < \tilde{T} + \tilde{X} < \pi \\ & \sinh[\tan((\tilde{T} - \tilde{X})/2)] \sinh[\tan((\tilde{T} + \tilde{X})/2)] < 1. \end{aligned} \quad (9.79)$$

The picture of (\mathcal{M}, 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 feature of displaying radial null geodesics as straight lines with slope $\pm 45^\circ$, since \tilde{U} (resp. \tilde{V}) is a function of U only (resp. V only) [cf. Eq. (9.64)]. In particular, the bifurcate Killing horizon and the Schwarzschild horizon are defined by:

$$\hat{\mathcal{H}} : \quad \tilde{U} = 0 \quad \text{or} \quad \tilde{V} = 0 \iff \tilde{T} = \tilde{X} \quad \text{or} \quad \tilde{T} = -\tilde{X} \quad (9.80a)$$

$$\mathcal{H} : \quad \tilde{U} = 0 \quad \text{and} \quad \tilde{V} > 0 \iff \tilde{T} = \tilde{X} \quad \text{and} \quad \tilde{T} > 0. \quad (9.80b)$$

These relations follow immediately from (9.43), (9.64) and (9.78).

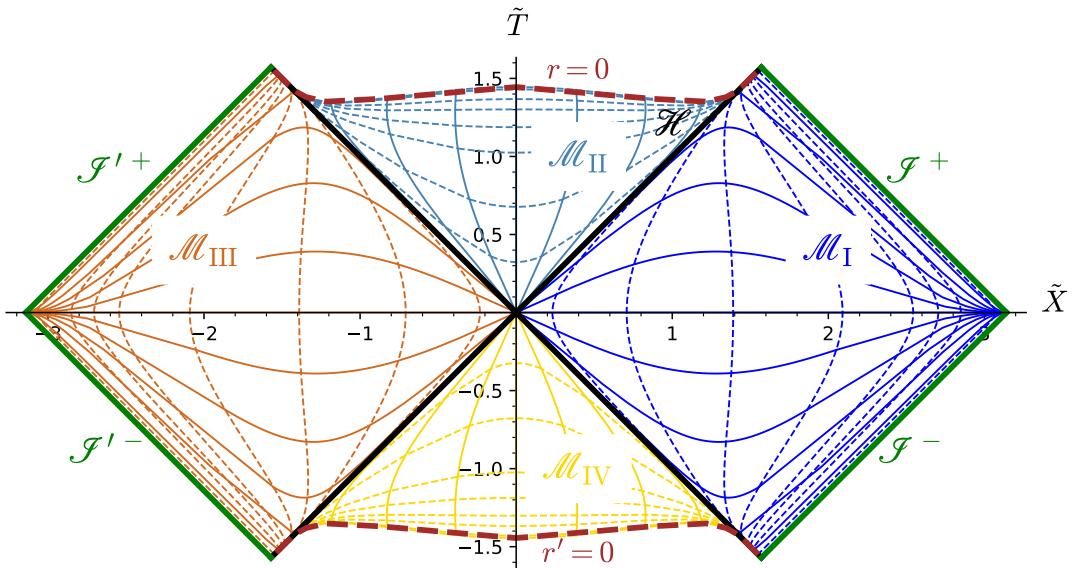


Figure 9.11: Carter-Penrose diagram of the Schwarzschild spacetime based on the Penrose-Frolov-Novikov coordinates. Solid curves denote the same hypersurfaces of constant Schwarzschild-Droste coordinate t as in Fig. 9.10: in region \mathcal{M}_I , from the \hat{X} -axis to the top: $t = 0, 2m, 5m, 10m, 20m$ and $50m$; in region \mathcal{M}_{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 \mathcal{M}_I , from the left to the right: $r = 2.01m, 2.1m, 2.5m, 4m, 8m, 12m, 20m$ and $100m$; in region \mathcal{M}_{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.4.18](#)]

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 \hat{T} -axis. This follows from the third condition in (9.79), which is more involved than the third condition in (9.51). Actually this boundary corresponds to the curvature singularity limit $r \rightarrow 0$ or $r' \rightarrow 0$. Indeed, from Eq. (9.41), we have

$$r = 0 \iff UV = 1. \quad (9.81)$$

In terms of the coordinates (\hat{U}, \hat{V}) , we have then [cf. Eq. (9.46)]

$$\begin{aligned} r = 0 &\iff \tan \hat{U} \tan \hat{V} = 1 \iff \sin \hat{U} \sin \hat{V} = \cos \hat{U} \cos \hat{V} \\ &\iff \cos \hat{U} \cos \hat{V} - \sin \hat{U} \sin \hat{V} = 0 \iff \cos(\hat{U} + \hat{V}) = 0 \iff \hat{U} + \hat{V} = \pm \frac{\pi}{2}. \end{aligned}$$

Since $\hat{T} = \hat{U} + \hat{V}$, we get the simple relation

$$r = 0 \iff \hat{T} = \pm \frac{\pi}{2}. \quad (9.82)$$

On the contrary, in terms of the coordinates (\tilde{U}, \tilde{V}) , Eq. (9.81) becomes [cf. Eq. (9.64)]

$$r = 0 \iff \sinh(\tan \tilde{U}) \sinh(\tan \tilde{V}) = 1,$$

which yields to the complicated formula

$$r = 0 \iff \sinh[\tan((\tilde{T} - \tilde{X})/2)] \sinh[\tan((\tilde{T} + \tilde{X})/2)] = 1. \quad (9.83)$$

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. [201]; see also Fig. 10.6 of Ref. [203]) 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. 292). Accordingly, in Frolov & Novikov's version, one shall replace the 1 in the right-hand side of Eq. (9.83) by $1/e$, yielding to a different shape of the boundary $r = 0$.

Remark 4: As noticed by Frolov and Novikov [201] (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 Penrose-Frolov-Novikov coordinates (Fig. 9.11) and the “standard” diagram of Fig. 9.10 is that the $t = \text{const}$ hypersurfaces of the former (solid curves in Fig. 9.11) are all tangent to the horizontal axis when $\tilde{X} \rightarrow \pm\pi$. 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.

Historical note : The transformation (9.64) from (U, V) to (\tilde{U}, \tilde{V}) has been suggested by Roger Penrose in 1967 (p. 209 of Ref. [406]) to get a “version of the Kruskal diagram in which conformal infinity is represented”. The figure in Penrose's article (Fig. 37 in Ref. [406]) resembles Fig. 9.11, except that the curve $r = 0$ does not look tangent to the lines $\tilde{T} = \pm\tilde{X}$, as it is in Fig. 9.11. The same coordinate change (9.64) has been discussed in details by Valeri P. Frolov and Igor D. Novikov in their 1998 textbook [201] (Sec. 5.1.3), stressing that it leads to a regular conformal completion, contrary to the “standard” coordinate change (9.46) introduced by Brandon Carter [87] and used in Hawking & Ellis [265] and MTW [370] textbooks.

9.4.4 Black hole and white hole regions

$(\tilde{\mathcal{M}}, \tilde{g})$ is not only a conformal completion of the maximally extended Schwarzschild spacetime (\mathcal{M}, 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 \mathcal{I} of $\tilde{\mathcal{M}}$ expressed by Eq. (9.75) as

$$\mathcal{I} = \mathcal{I}_*^+ \cup \mathcal{I}_*^-, \quad (9.84)$$

with

$$\mathcal{I}_*^+ := \mathcal{I}^+ \cup \mathcal{I}'^+ \quad \text{and} \quad \mathcal{I}_*^- := \mathcal{I}^- \cup \mathcal{I}'^-. \quad (9.85)$$

It is clear that any point of \mathcal{I}_*^+ is the future end point of a future-directed null curve from \mathcal{M} , while any point of \mathcal{I}_*^- is the past end point of a future-directed null curve to \mathcal{M} (cf. Fig. 9.11).

Hence the writing (9.84) matches the definition of a conformal completion at null infinity given in Sec. 4.3.1. \mathcal{I}_*^+ is then the future null infinity of $(\mathcal{M}, \mathbf{g})$ and \mathcal{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 \mathcal{I}_*^+ is $J^-(\mathcal{I}_*^+) = \mathcal{M}_I \cup \mathcal{M}_{III} \cup \mathcal{M}_{IV}$ (cf. Fig. 9.11). In view of the definition (4.37), we get

Property 9.2: black hole in the maximal Schwarzschild spacetime

The maximal extension of Schwarzschild spacetime admits a black hole region, the interior of which is \mathcal{M}_{II} . The black hole event horizon is the part of the bifurcate Killing horizon $\hat{\mathcal{H}}$ (cf. Sec. 9.3.3) that has $\tilde{T} > 0$.

The black hole region has thus the same interior \mathcal{M}_{II} as the black hole region of the Schwarzschild spacetime $(\mathcal{M}_{IEF}, \mathbf{g})$ considered in Chap. 6. It differs only a larger boundary: it is not reduced to the Schwarzschild horizon \mathcal{H} , but contains the part $\tilde{X} < 0$ and $\tilde{T} > 0$ of $\hat{\mathcal{H}}$.

The novelty with respect to the original Schwarzschild spacetime $(\mathcal{M}_{IEF}, \mathbf{g})$ is the existence of a white hole region. Indeed, the causal future of \mathcal{I}_*^- is $J^+(\mathcal{I}_*^-) = \mathcal{M}_I \cup \mathcal{M}_{II} \cup \mathcal{M}_{III}$ (cf. Fig. 9.11). In view of the definition (4.41), we get

Property 9.3: white hole in the maximal Schwarzschild spacetime

The maximal extension of Schwarzschild spacetime admits a white hole region, the interior of which is \mathcal{M}_{IV} . The corresponding past event horizon is the part of the bifurcate Killing horizon $\hat{\mathcal{H}}$ (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 $(\mathcal{M}, \mathbf{g})$, 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 \mathcal{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 (9.30):

$$X^2 > T_0^2 - 1. \quad (9.86)$$

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 .

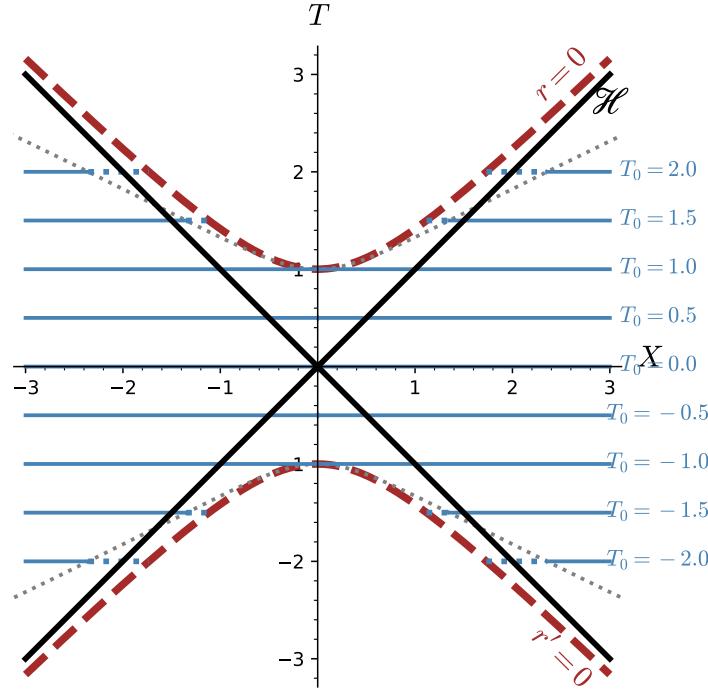


Figure 9.12: Kruskal diagram with the hypersurfaces Σ_{T_0} (defined by $T = T_0 = \text{const}$) as blue horizontal lines. For $|T_0| > 1$, the dotted part of Σ_{T_0} corresponds to a region that cannot be embedded isometrically in the Euclidean space. When T_0 varies, the limit of these regions form the grey dotted curve. [Figure generated by the notebook D.4.19]

- 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 $\mathbb{R} \times \mathbb{S}^2$.

For future convenience, we split Σ_{T_0} in two disjoint parts, according to the sign of X :

$$\Sigma_{T_0}^+ = \{p \in \Sigma_{T_0}, \quad X(p) \geq 0\} \quad \text{and} \quad \Sigma_{T_0}^- = \{p \in \Sigma_{T_0}, \quad X(p) < 0\}. \quad (9.87)$$

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 γ induced on it by g :

$$\gamma = \frac{32m^3}{r} e^{-r/2m} dX^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (9.88)$$

where r is the function of X defined by

$$r = r(X) = 2mW(X^2 - T_0^2). \quad (9.89)$$

The metric (9.88) is obtained by setting $T = T_0$ and $dT = 0$ in (9.31). Since $r > 0$, the metric (9.88) is clearly positive definite, i.e. γ is a Riemannian metric and Σ_{T_0} is a spacelike hypersurface.

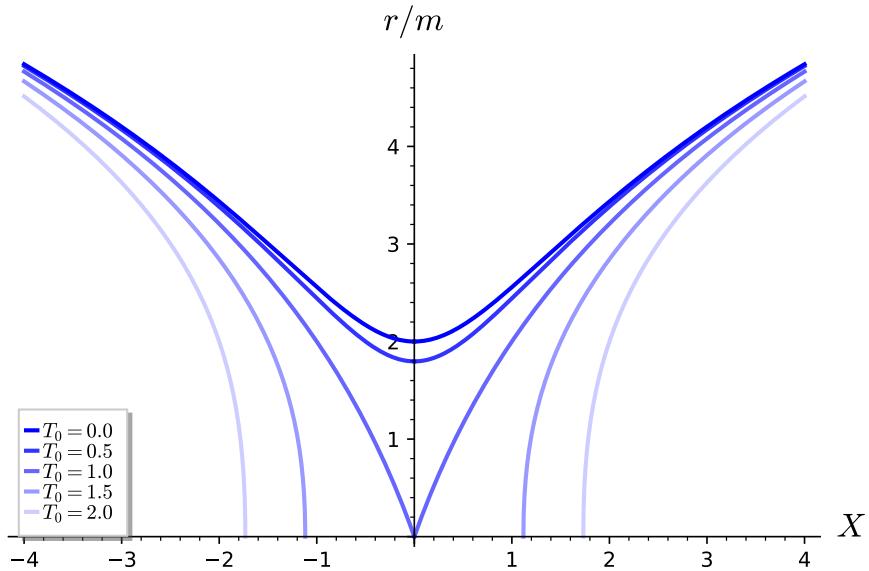


Figure 9.13: Function $r = r(X)$ on the hypersurface Σ_{T_0} , for the same values of T_0 as in Fig. 9.12. [Figure generated by the notebook D.4.19]

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 $(X_0, +\infty) \rightarrow (r_0, +\infty)$ (resp. $(-\infty, -X_0) \rightarrow (r_0, +\infty)$), where

$$X_0 = \begin{cases} 0 & \text{if } |T_0| < 1 \\ \sqrt{T_0^2 - 1} & \text{if } |T_0| \geq 1. \end{cases} \quad \text{and} \quad r_0 = \begin{cases} 2m W(-T_0^2) & \text{if } |T_0| < 1 \\ 0 & \text{if } |T_0| \geq 1. \end{cases} \quad (9.90)$$

The inverse of this bijection is⁵

$$X = X(r) = \pm \sqrt{e^{r/2m} \left(\frac{r}{2m} - 1 \right) + T_0^2}, \quad (9.91)$$

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 (9.91) leads to

$$dX = \pm \frac{re^{r/2m}}{8m^2 \sqrt{e^{r/2m} \left(\frac{r}{2m} - 1 \right) + T_0^2}} dr.$$

Substituting in (9.88) we get the expression of the metric on Σ_{T_0} in terms of the coordinates $(x^i) = (r, \theta, \varphi)$:

$$\gamma = \left(1 - \frac{2m}{r} (1 - T_0^2 e^{-r/2m}) \right)^{-1} dr^2 + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2). \quad (9.92)$$

⁵Let us recall that $W^{-1}(x) = e^x(x - 1)$.

Remark 1: As a check of the above formula, we notice that for $T_0 = 0$ it reduces to the metric of a slice $t = \text{const}$ in Schwarzschild-Droste coordinates [set $dt = 0$ in Eq. (6.14)]. 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. (9.18) (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 $(\mathbb{R}^3, \mathbf{f})$, \mathbf{f} being the standard flat (Euclidean) metric. By *isometric embedding* of a 2-dimensional Riemannian manifold $(\mathcal{S}, \mathbf{g})$ in $(\mathbb{R}^3, \mathbf{f})$, it is meant a *smooth embedding* $\Phi : \mathcal{S} \rightarrow \mathbb{R}^3$, 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 \mathbf{g} on \mathcal{S} :

$$\forall p \in \mathcal{S}, \quad \forall (\mathbf{u}, \mathbf{v}) \in (T_p \mathcal{S})^2, \quad \mathbf{f}(\Phi_* \mathbf{u}, \Phi_* \mathbf{v}) = \mathbf{g}(\mathbf{u}, \mathbf{v}), \quad (9.93)$$

where $\Phi_* \mathbf{u}$ is the vector of \mathbb{R}^3 that is the pushforward of \mathbf{u} by Φ , as defined in Sec. A.2.8. Another phrasing of the isometry property (9.93) is: the pullback of \mathbf{f} on \mathcal{S} by Φ coincides with \mathbf{g} : $\Phi^* \mathbf{f} = \mathbf{g}$ [cf. Eq. (A.33)].

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 $\Sigma_{T_0}^{\text{eq}}$. Coordinates on $\Sigma_{T_0}^{\text{eq}}$ are $(x^a) = (X, \varphi)$, or on each of the two parts $\Sigma_{T_0}^{+, \text{eq}}$ ($X \geq 0$) and $\Sigma_{T_0}^{-, \text{eq}}$ ($X < 0$), $(x^a) = (r, \varphi)$. If $|T_0| < 1$, the topology of $\Sigma_{T_0}^{\text{eq}}$ is $\mathbb{R} \times \mathbb{S}^1$, i.e. that of a cylinder, while for $|T_0| \geq 1$, it has two connected components, $\Sigma_{T_0}^{+, \text{eq}}$ and $\Sigma_{T_0}^{-, \text{eq}}$, each of them having the topology of a cylinder.

The metric induced by \mathbf{g} on $\Sigma_{T_0}^{\text{eq}}$, \mathbf{q} say, is obtained by setting $\theta = \pi/2$ and $d\theta = 0$ in Eq. (9.92):

$$\mathbf{q} = \left(1 - \frac{2m}{r} (1 - T_0^2 e^{-r/2m}) \right)^{-1} dr^2 + r^2 d\varphi^2. \quad (9.94)$$

Given the invariance in φ , it is quite natural to embed $(\Sigma_{T_0}^{\text{eq}}, \mathbf{q})$ as a surface of revolution in the Euclidean space $(\mathbb{R}^3, \mathbf{f})$. Describing \mathbb{R}^3 with cylindrical coordinates $(x^i) = (r, z, \varphi)$, the Euclidean metric \mathbf{f} is

$$\mathbf{f} = dr^2 + dz^2 + r^2 d\varphi^2. \quad (9.95)$$

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 $dz = Z'(r) dr$, so that the metric \mathbf{h} induced by \mathbf{f} on it is

$$\mathbf{h} = (1 + Z'(r)^2) dr^2 + r^2 d\varphi^2. \quad (9.96)$$

Comparing (9.96) with (9.94), we see that a possible isometric embedding of $(\Sigma_{T_0}^{\text{eq}}, \mathbf{q})$ into $(\mathbb{R}^3, \mathbf{f})$ is

$$\begin{aligned} \Phi : \Sigma_{T_0}^{\text{eq}} &\longrightarrow \mathbb{R}^3 \\ (X, \varphi) &\longmapsto (r, z, \varphi) = (r(X), \pm Z(r(X)), \varphi), \end{aligned} \quad (9.97)$$

with the function $r(X)$ is given by Eq. (9.89), the sign \pm is $+$ on $\Sigma_{T_0}^{+, \text{eq}}$ and $-$ on $\Sigma_{T_0}^{-, \text{eq}}$ and the function $Z(r)$ obeys

$$1 + Z'(r)^2 = \left(1 - \frac{2m}{r} (1 - T_0^2 e^{-r/2m}) \right)^{-1}.$$

Thanks to Eq. (9.91), this expression can be recast as

$$Z'(r)^2 = \frac{1 - T_0^2 e^{-r/2m}}{T_0^2 e^{-r/2m} + \frac{r}{2m} - 1} = \frac{e^{r/2m} - T_0^2}{X(r)^2}. \quad (9.98)$$

For $|T_0| < 1$, i.e. when $\Sigma_{T_0}^{\text{eq}}$ is connected, the map (9.97) defines a smooth embedding if, and only if, at the boundary $X = 0$ between $\Sigma_{T_0}^{+, \text{eq}}$ and $\Sigma_{T_0}^{-, \text{eq}}$, the following holds:

$$Z(r(0)) = 0 \quad \text{and} \quad Z'(r(0)) = \infty. \quad (9.99)$$

The condition $Z(r(0)) = 0$ insures the continuity of the embedded surface $\Phi(\Sigma_{T_0}^{\text{eq}})$, while $Z'(r(0)) = +\infty$ insures that it has a vertical tangent at the junction between $\Phi(\Sigma_{T_0}^{+, \text{eq}})$ and $\Phi(\Sigma_{T_0}^{-, \text{eq}})$, 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 (9.98): $Z'(r)^2$ clearly diverges at $X = 0$. Moreover, in order for the isometric embedding Φ to be well-defined, the right-hand side of (9.98) 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 $e^{r/2m} \geq T_0^2$, or equivalently,

$$r \geq 4m \ln |T_0|. \quad (9.100)$$

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 ,

$$r_{\text{emb}}(T_0) := 4m \ln |T_0|, \quad (9.101)$$

such that the part of $\Sigma_{T_0}^{\text{eq}}$ with $r < r_{\text{emb}}(T_0)$ 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* [378], which states that any smooth Riemannian manifold of dimension n can be isometrically embedded in the Euclidean space (\mathbb{R}^m, 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 (9.89) and the fact that the rescaled Lambert function W is an increasing function (cf. Fig. 9.4), of inverse $F(x) = e^x(x-1)$, the condition (9.100) can be turned into a condition on X :

$$|X| \geq X_{\text{emb}}(T_0) := |T_0| \sqrt{2 \ln |T_0|}. \quad (9.102)$$

This limit is shown as the grey dotted curve in Fig. 9.12.

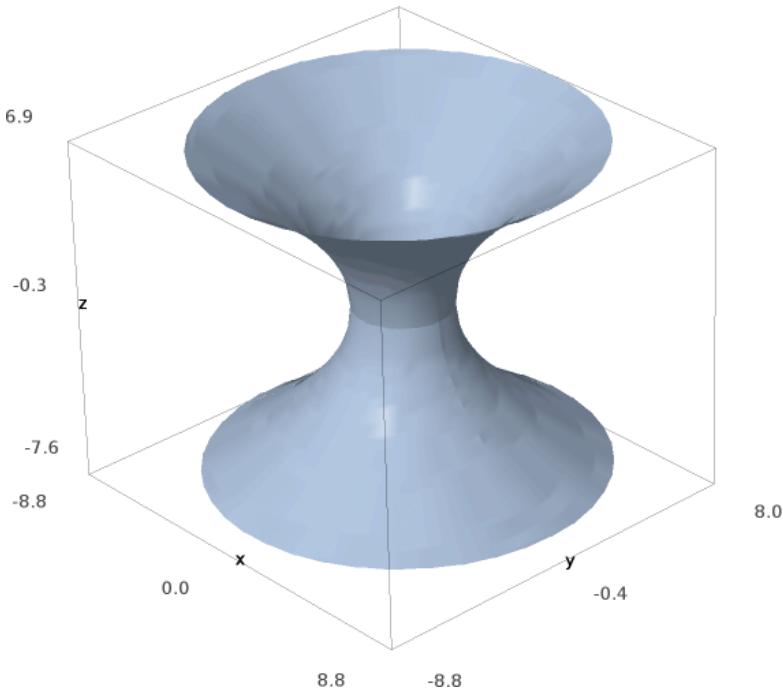


Figure 9.14: Flamm paraboloid: isometric embedding in the Euclidean \mathbb{R}^3 of the spacelike slice $T = 0$ and $\theta = \pi/2$ of Schwarzschild spacetime. The (x, y) coordinates are the standard Cartesian coordinates of \mathbb{R}^3 related to (r, φ) via $x = r \cos \varphi$ and $y = r \sin \varphi$. The labels are in units of m . [Figure generated by the notebook D.4.19]

Summarizing, the minimal value of r on the embedded surface is r_0 [cf. Eq. (9.90)] for $|T_0| \leq 1$ or $r_{\text{emb}}(T_0)$ for $|T_0| > 1$:

$$r_{\min}(T_0) = \begin{cases} 2m W(-T_0^2) & \text{if } |T_0| \leq 1 \\ 4m \ln |T_0| & \text{if } |T_0| > 1. \end{cases} \quad (9.103)$$

Note that $r_{\min}(T_0)$ is a continuous function, with the peculiar values $r_{\min}(0) = 2m$ and $r_{\min}(1) = 0$. The embedding function $Z(r)$ is found by integration of $Z'(r)$, as given by (9.98), from $r_{\min}(T_0)$ to r :

$$Z(r) = 2m \int_{\frac{r_{\min}(T_0)}{2m}}^{\frac{r}{2m}} \sqrt{\frac{1 - T_0^2 e^{-x}}{T_0^2 e^{-x} + x - 1}} dx. \quad (9.104)$$

The integral cannot be computed exactly in terms of elementary functions, except for $T_0 = 0$, where it reduces to

$$Z(r) = 2m \int_1^{\frac{r}{2m}} \frac{dx}{\sqrt{x-1}} \quad (T_0 = 0).$$

Hence

$$Z(r) = 4m \sqrt{\frac{r}{2m} - 1} \quad (T_0 = 0).$$

(9.105)

According to (9.97), the whole surface $\Phi(\Sigma_{T_0}^{\text{eq}})$ 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

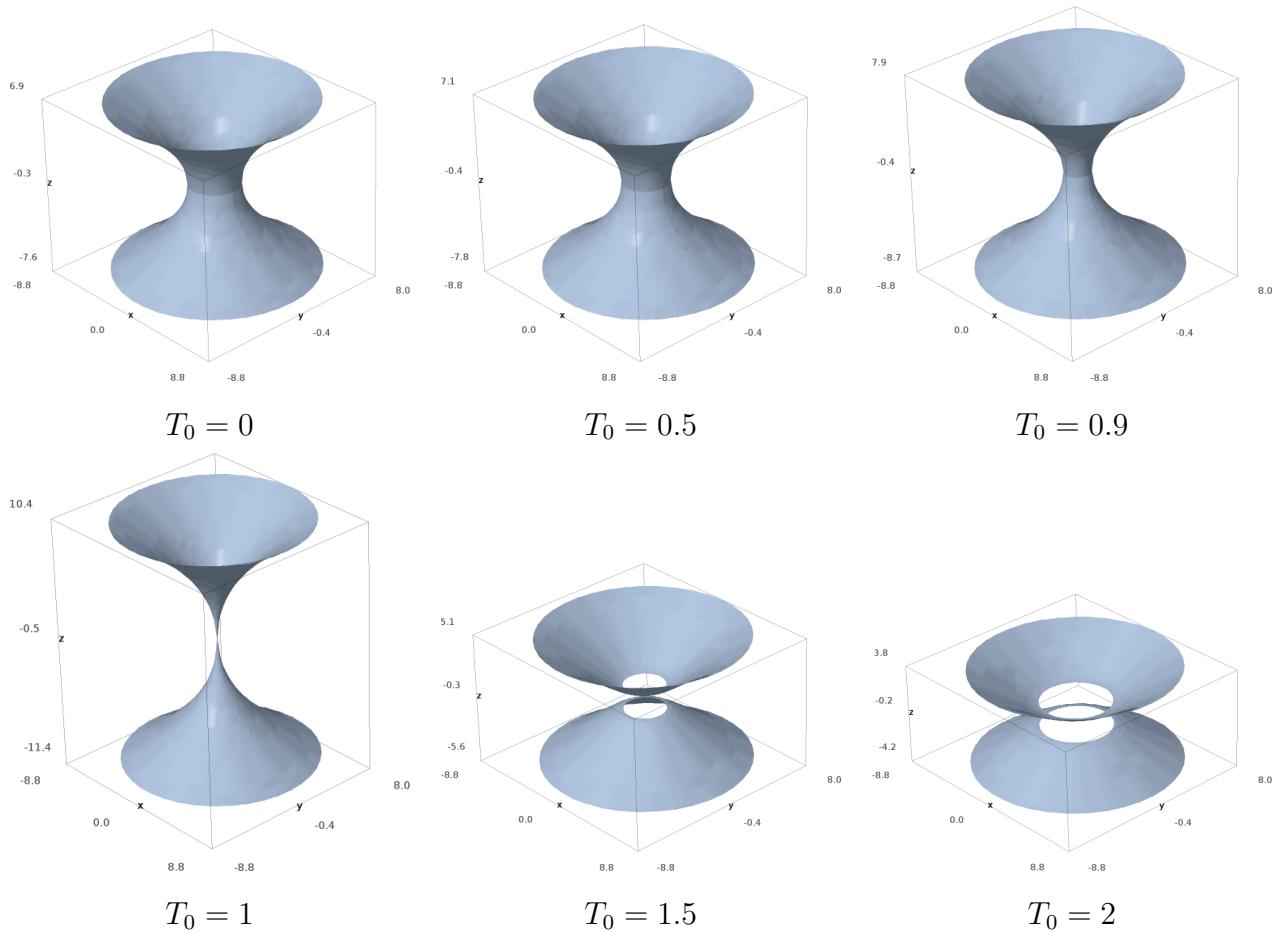


Figure 9.15: Sequence of isometric embeddings in the Euclidean space of spacelike slices of Schwarzschild spacetime defined by $T = T_0$ and $\theta = \pi/2$. The slices are those shown in the Kruskal diagram of Fig. 9.12 (except for $T_0 = 0.9$). The first embedding ($T_0 = 0$) is the Flamm paraboloid 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.4.19]

$$z^2 = 16m^2(r/2m - 1), \text{ or}$$

$$\boxed{z^2 = 8m(r - 2m)}. \quad (9.106)$$

We recognize the equation of a paraboloid of revolution around the z -axis. It is known as **Flamm paraboloid** [196] 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, $r_{\min} = 2m$, and then increases. The “neck” around $r = r_{\min}$, or equivalently $X = 0$, is called the **Einstein-Rosen bridge** [178]. Contemplating the slice $T = 0$ in the Kruskal diagram of Fig. 9.12, we realize that this “bridge” connects the two asymptotically flat regions \mathcal{M}_I and \mathcal{M}_{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 \mathcal{M}_I to \mathcal{M}_{III} .

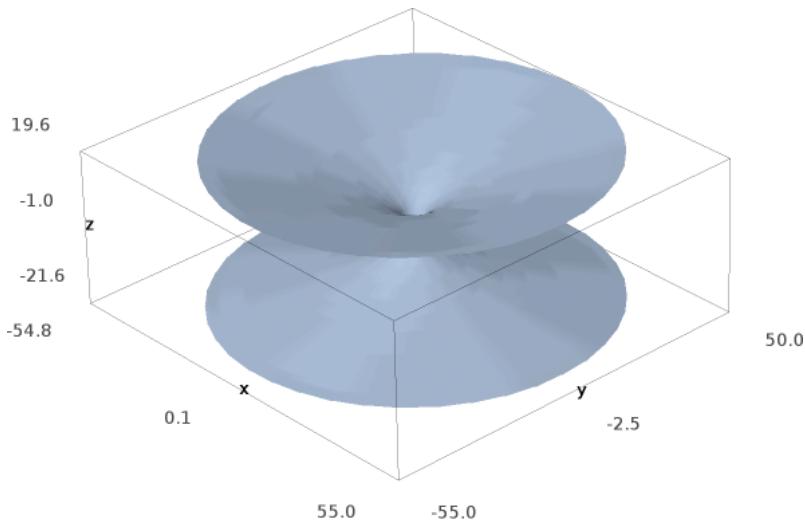


Figure 9.16: Same Flamm paraboloid as in Fig. 9.14 but seen from farther away. Despite being more and more flat, none of the two sheets is asymptotic to a plane. [Figure generated by the notebook D.4.19]

When $T_0 \neq 0$, the integral in (9.104) has to be computed numerically (see Sec. D.4.19 for the computation with SageMath). The resulting embedded surfaces $\Phi(\Sigma_{T_0}^{\text{eq}})$ 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 $\Sigma_{T_0}^{\text{eq}}$ 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. (9.101). 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 $T = \text{const}$ of Schwarzschild spacetime, with the isometric embedding limitation for $|T| > 1$. We refer the reader to Ref. [133] 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. (9.105) 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 (9.94) tends to a flat metric when $r \rightarrow +\infty$, 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 $T = \text{const}$ cannot reach the region near $r = 0$ for $|T_0| > 1$.

Historical note : In 1916, very soon after the publication of Schwarzschild solution [448], the Austrian physicist Ludwig Flamm (1885-1964) showed that the slice $t = \text{const}$ 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. (9.106) [196]. 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. 308). Although he draw the whole paraboloid (actually a parabola in a 2-dimensional plot – Fig. 2 of Ref. [196]), Flamm did not seem to have considered the negative- z part as physically relevant. In other words, he limited his considerations to \mathcal{M}_I and did not contemplate any bridge to the extension \mathcal{M}_{III} .

9.5.3 Isotropic coordinates

Let us consider a hypersurface of constant Schwarzschild-Droste time t in \mathcal{M}_I . According to Eq. (9.17), this hypersurface obeys

$$T = \tanh\left(\frac{t}{4m}\right) X, \quad (9.107)$$

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 \mathcal{M}_{III} obeys an equation identical to (9.107), except for t replaced by t' and $X < 0$ [cf. Eq. (9.33)]. Accordingly, for $t' = t$, the union of these two hypersurfaces forms a hypersurface of \mathcal{M} ruled by Eq. (9.107), with $X < 0$ or $X > 0$. If we add the points $(T, X) = (0, 0)$ to it (i.e. the bifurcation sphere \mathcal{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 \mathcal{M} defined by Eq. (9.107) 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 $\mathcal{S}_t \cap \mathcal{M}_I$ and (r', θ, φ) on $\mathcal{S}_t \cap \mathcal{M}_{III}$, with both r and r' ranging $(2m, +\infty)$. Let us introduce on \mathcal{S}_t a third coordinate system $(x^i) = (\bar{r}, \theta, \varphi)$ as follows:

$$\text{on } \mathcal{S}_t \cap \mathcal{M}_I : \quad \bar{r} \in \left(\frac{m}{2}, +\infty\right), \quad r = \bar{r} \left(1 + \frac{m}{2\bar{r}}\right)^2 \quad (9.108a)$$

$$\iff \bar{r} = \frac{1}{2} \left(r - m + \sqrt{r(r - 2m)}\right) \quad (9.108b)$$

$$\text{on } \mathcal{S}_t \cap \mathcal{M}_{III} : \quad \bar{r} \in \left(0, \frac{m}{2}\right), \quad r' = \bar{r} \left(1 + \frac{m}{2\bar{r}}\right)^2 \quad (9.108c)$$

$$\iff \bar{r} = \frac{1}{2} \left(r' - m - \sqrt{r'(r' - 2m)}\right) \quad (9.108d)$$

$$\text{on } \mathcal{S}_t \cap \mathcal{S} : \quad \bar{r} = \frac{m}{2}. \quad (9.108e)$$

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 \mathcal{M}_{III} part) and $\bar{r} \in (m/2, +\infty)$ (the \mathcal{M}_I part). In each of these part, there is a one-to-one correspondence

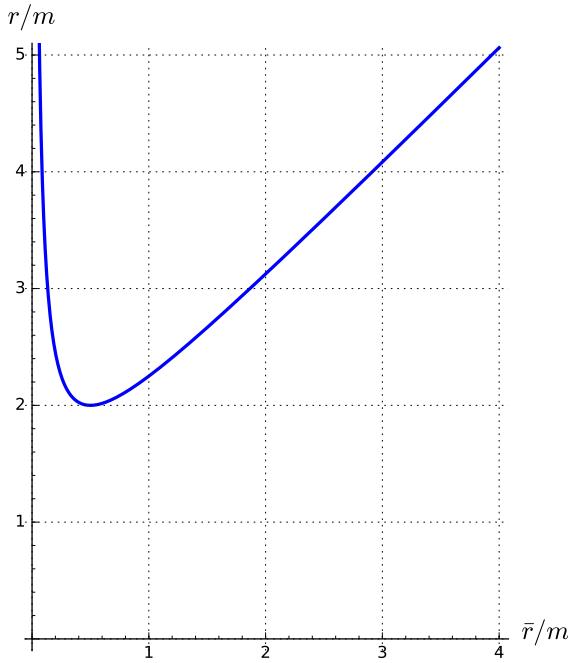


Figure 9.17: Areal radius r as a function of the isotropic coordinate \bar{r} .

between \bar{r} and r (or r'). Note that

$$\text{when } r \rightarrow +\infty, \quad \bar{r} \sim r \quad (9.109a)$$

$$\text{when } r' \rightarrow +\infty, \quad \bar{r} \sim \frac{m^2}{4r'}. \quad (9.109b)$$

When t varies, \mathcal{S}_t constitute a foliation of

$$\mathcal{M}_{\text{iso}} := \mathcal{M}_I \cup \mathcal{S} \cup \mathcal{M}_{III}. \quad (9.110)$$

This foliation is regular in both \mathcal{M}_I and \mathcal{M}_{III} , but is singular at the bifurcation sphere \mathcal{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 \mathcal{M}_{iso} , which is regular on \mathcal{M}_I and \mathcal{M}_{III} , but is singular at \mathcal{S} , i.e. at $\bar{r} = m/2$. This system is called **isotropic coordinates**.

From (9.108a), we get

$$dr = \left(1 + \frac{m}{2\bar{r}}\right) \left(1 - \frac{m}{2\bar{r}}\right) d\bar{r} \quad \text{and} \quad 1 - \frac{2m}{r} = \left(\frac{1 - \frac{m}{2\bar{r}}}{1 + \frac{m}{2\bar{r}}}\right)^2.$$

It is then immediate to deduce from (6.14) the expression of the metric tensor in terms of the isotropic coordinates $(x^{\bar{\alpha}}) = (t, \bar{r}, \theta, \varphi)$:

$$\mathbf{g} = - \left(\frac{1 - \frac{m}{2\bar{r}}}{1 + \frac{m}{2\bar{r}}}\right)^2 dt^2 + \left(1 + \frac{m}{2\bar{r}}\right)^4 [d\bar{r}^2 + \bar{r}^2 (d\theta^2 + \sin^2 \theta d\varphi^2)]. \quad (9.111)$$

Since the relation between r' and \bar{r} is identical to that between r and \bar{r} [cf. Eqs. (9.108a) and (9.108c)], the above expression of \mathbf{g} is valid on \mathcal{M}_I and \mathcal{M}_{III} . Note that all metric coefficients are regular on $\mathcal{M}_I \cup \mathcal{M}_{III}$ (except for the standard coordinate singularity of the spherical coordinates (θ, φ) for $\theta \in \{0, \pi\}$). On the contrary Eq. (9.111) 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 \mathcal{S} .

A remarkable feature of expression (9.111) is that the spatial part is proportional to the flat metric \mathbf{f} of the Euclidean 3-space:

$$\mathbf{f} = d\bar{r}^2 + \bar{r}^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2). \quad (9.112)$$

In other words, the metric γ induced by \mathbf{g} on \mathcal{S}_t is *conformal* to the flat metric \mathbf{f} (cf. Sec. 4.2.2):

$$\gamma = \Psi^4 \mathbf{f}, \quad (9.113)$$

with the conformal factor⁶

$$\Psi = 1 + \frac{m}{2\bar{r}}. \quad (9.114)$$

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 (9.108a). But much more than a mere change of coordinate $r \leftrightarrow \bar{r}$ is involved: the two coordinate systems do not cover the same part of the extended Schwarzschild spacetime: Schwarzschild-Droste coordinates cover the region $\mathcal{M}_I \cup \mathcal{M}_{II}$, while isotropic coordinates cover the region $\mathcal{M}_I \cup \mathcal{M}_{III}$. In particular, isotropic coordinates cannot be used to describe the black hole interior (i.e. \mathcal{M}_{II}). This last feature can be inferred directly from the square in the g_{tt} component read on (9.111), which implies an everywhere timelike Killing vector ∂_t , while ∂_t is spacelike in \mathcal{M}_{II} (cf. Sec. 6.2.5).

9.5.4 Recapitulation: coordinates on Schwarzschild spacetime

At this point, we have introduced many coordinate systems on Schwarzschild spacetime:

- Schwarzschild-Droste (t, r, θ, φ) (Sec. 6.2.3);
- ingoing Eddington-Finkelstein $(\tilde{t}, r, \theta, \varphi)$ (Sec. 6.3.2);
- null ingoing Eddington-Finkelstein (v, r, θ, φ) (Sec. 6.3.2);
- null outgoing Eddington-Finkelstein (u, r, θ, φ) (Sec. 6.4);
- Kruskal-Szekeres (T, X, θ, φ) (Sec. 9.2.1);
- global null (U, V, θ, φ) (Sec. 9.3.2);
- Penrose-Frolov-Novikov $(\tilde{U}, \tilde{V}, \theta, \varphi)$ (Sec. 9.4.3);

⁶Note 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 : $\Omega = \Psi^2$.

- isotropic $(t, \bar{r}, \theta, \varphi)$ (Sec. 9.5.3).

Two other coordinate systems will be introduced in Chap. 14:

- Lemaître synchronous $(\tau, \chi, \theta, \varphi)$ (Sec. 14.2.6);
- Painlevé-Gullstrand $(\tau, r, \theta, \varphi)$ (Sec. 14.2.7).

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 \mathcal{M}_I and \mathcal{M}_{III} . Indeed, it is clear from the Carter-Penrose diagram of Fig. 9.11 that the past null cone of any event in \mathcal{M}_I or \mathcal{M}_{III} encounters the curvature singularity $r' = 0$ at the past boundary of \mathcal{M}_{IV} . In other words, any observer in \mathcal{M}_I or \mathcal{M}_{III} can receive signals from this singularity. Contrary to the singularity $r = 0$ at the future boundary of \mathcal{M}_{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 [407] (see also Sec. 10.1 of Wald’s textbook [498]).

9.6.2 Astrophysical relevance

We shall see in Chap. 14 that, in spherical symmetry, the formation of a black hole by the astrophysical process of gravitational collapse of a star or a cloud of matter yields a spacetime that contains some parts of regions \mathcal{M}_I and \mathcal{M}_{II} of Schwarzschild spacetime, but that do not contain any part of \mathcal{M}_{III} nor \mathcal{M}_{IV} (cf. the Carter-Penrose diagram of Fig. 14.4). 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 instance, we shall see in Chap. 14 that there exist solutions of the Einstein equation describing a spherically symmetric ball of matter that is expanding from an initial singularity, reaches some maximal extension and then collapses and whose exterior contains parts of regions \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{IV} , and even \mathcal{M}_{III} (in which case, it is called a *semiclosed world*), cf. Figs. 14.5 and 14.6, as well as Remark 3 on p. 557. For the above reasons, the maximal extension of the Schwarzschild spacetime is sometimes called the *eternal Schwarzschild black hole*.

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 [353], 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. [340]).

Part III

Kerr black hole

Chapter 10

Kerr black hole

Contents

10.1 Introduction	321
10.2 The Kerr solution	322
10.3 Extension of the spacetime manifold through $\Delta = 0$	332
10.4 Principal null geodesics	338
10.5 Event horizon	345
10.6 Global quantities	349
10.7 Families of observers in Kerr spacetime	352
10.8 Maximal analytic extension	362
10.9 Further reading	369

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. 5.6).

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. 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 discusses 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 specific angular momentum $a = J/m$, J being the spacetime total angular momentum.

In this chapter, we focus on Kerr solutions for which

$$0 < a < m, \quad (10.1)$$

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 Boyer-Lindquist coordinates cover the manifold $\mathbb{R}^2 \times \mathbb{S}^2$, with t spanning \mathbb{R} , r spanning¹ \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 \mathcal{M}_{BL} of $\mathbb{R}^2 \times \mathbb{S}^2$ formed by the disjoint union of the following three connected components (cf. Fig. 10.1):

$$\mathcal{M}_{\text{BL}} := \mathcal{M}_I \cup \mathcal{M}_{\text{II}} \cup \mathcal{M}_{\text{III}}, \quad (10.2a)$$

$$\mathcal{M}_I := \mathbb{R} \times (r_+, +\infty) \times \mathbb{S}^2 \quad (10.2b)$$

$$\mathcal{M}_{\text{II}} := \mathbb{R} \times (r_-, r_+) \times \mathbb{S}^2 \quad (10.2c)$$

$$\mathcal{M}_{\text{III}} := \mathbb{R} \times (-\infty, r_-) \times \mathbb{S}^2 \setminus \mathcal{R}, \quad (10.2d)$$

where

$$\boxed{r_+ := m + \sqrt{m^2 - a^2}} \quad \text{and} \quad \boxed{r_- := m - \sqrt{m^2 - a^2}} \quad (10.3)$$

and \mathcal{R} is the subset of $\mathbb{R}^2 \times \mathbb{S}^2$ defined in terms of Boyer-Lindquist coordinates (t, r, θ, φ) by

$$\mathcal{R} := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = 0 \text{ and } \theta(p) = \frac{\pi}{2} \right\}. \quad (10.4)$$

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 \mathcal{M}_{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 $\mathcal{M} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathcal{R}$. Even latter on, after having

¹This contrasts with r spanning only $(0, +\infty)$ for the standard spherical coordinates (r, θ, φ) on \mathbb{R}^3 .

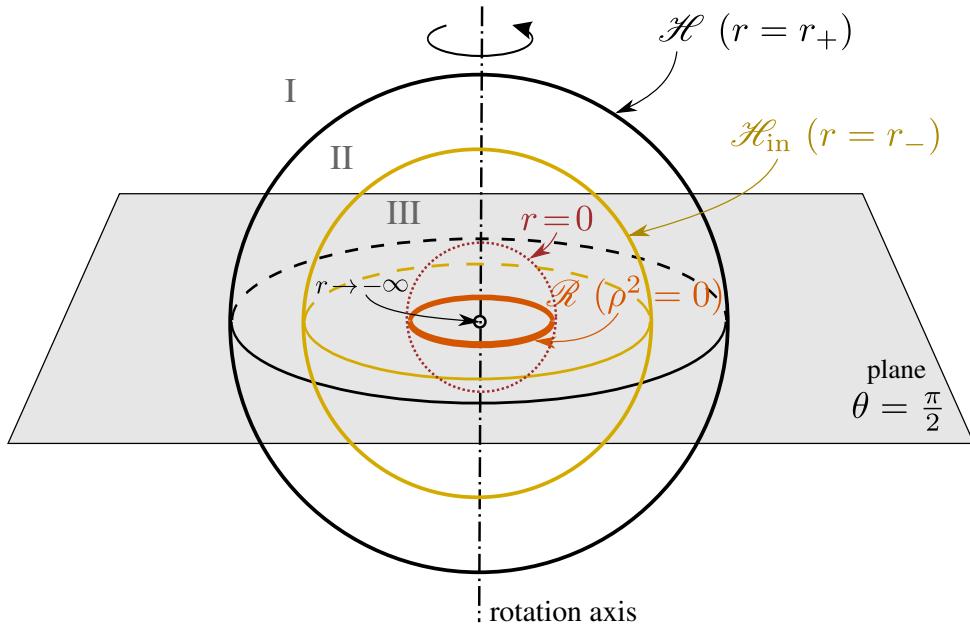


Figure 10.1: View of a section $t = \text{const}$ of $\mathbb{R}^2 \times \mathbb{S}^2$ in terms of the coordinates (R, θ, φ) , with $R := e^r$, so that the region $r \rightarrow -\infty$ is reduced to a single point at the centre of all the pictured spheres. Such coordinates have been introduced for pictorial purposes by O'Neill [390].

noticed that (\mathcal{M}, 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 \mathcal{S}_0 diffeomorphic to \mathbb{S}^2 . On the spacetime manifold \mathcal{M}_{BL} , the points $r = 0$ are part of \mathcal{M}_{III} and because of the exclusion of \mathcal{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

$$0 < r_- < m < r_+ < 2m. \quad (10.5)$$

Note also that \mathcal{R} is spanned by the coordinates (t, φ) and is diffeomorphic to the 2-dimensional cylinder $\mathbb{R} \times \mathbb{S}^1$:

$$\mathcal{R} \simeq \mathbb{R} \times \mathbb{S}^1. \quad (10.6)$$

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 \mathcal{M}_{BL} are defined in terms of the Boyer-Lindquist coordinates (t, r, θ, φ) by

$$\forall p \in \mathcal{M}_{\text{BL}}, \quad p \in \mathcal{M}_I \iff r(p) > r_+ \quad (10.7a)$$

$$p \in \mathcal{M}_{\text{II}} \iff r_- < r(p) < r_+ \quad (10.7b)$$

$$p \in \mathcal{M}_{\text{III}} \iff r(p) < r_- \text{ and } \left(r(p) \neq 0 \text{ or } \theta(p) \neq \frac{\pi}{2} \right). \quad (10.7c)$$

The **Kerr metric** is defined in terms of the Boyer-Lindquist coordinates (t, r, θ, φ) by

$$\begin{aligned} g = & -\left(1 - \frac{2mr}{\rho^2}\right) dt^2 - \frac{4amr \sin^2 \theta}{\rho^2} dt d\varphi + \frac{\rho^2}{\Delta} dr^2 \\ & + \rho^2 d\theta^2 + \left(r^2 + a^2 + \frac{2a^2mr \sin^2 \theta}{\rho^2}\right) \sin^2 \theta d\varphi^2, \end{aligned} \quad (10.8)$$

with

$$\rho^2 := r^2 + a^2 \cos^2 \theta \quad (10.9)$$

and

$$\Delta := r^2 - 2mr + a^2 = (r - r_-)(r - r_+). \quad (10.10)$$

Remark 3: Since $dt d\varphi := 1/2 (dt \otimes d\varphi + d\varphi \otimes dt)$ [cf. Eq. (A.39)], the metric component $g_{t\varphi}$ is *one half* of the coefficient of the $dt d\varphi$ term in Eq. (10.8), i.e. $g_{t\varphi} = -2amr \sin^2 \theta / \rho^2$.

Remark 4: On \mathcal{M}_{BL} , $\rho \neq 0$ and $\Delta \neq 0$ (by construction of \mathcal{M}_{BL} !), so that the metric components (10.8) are regular in \mathcal{M}_{BL} , except for the standard singularities of the spherical coordinates (θ, φ) .

By means of a computer algebra system (cf. the SageMath notebook D.5.1), one easily checks:

Property 10.1: Kerr metric as a solution of the vacuum Einstein equation

The Kerr metric g [Eq. (10.8)] is a solution of the vacuum Einstein equation (1.44) on \mathcal{M}_{BL} .

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 [311]. 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 [439]. 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, known today as *Kerr coordinates* and to be discussed in Sec. 10.3.1 (cf. the historical note page 334). For more details about this fantastic discovery, see the account by Kerr himself in Refs. [312, 323]. Boyer-Lindquist coordinates have been introduced in 1966 by Robert H. Boyer (see the historical note on p. 87) and Richard W. Lindquist [71] (compare Eq. (2.13) of Ref. [71] 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 [311], cf. the historical note on p. 334).

10.2.2 Basic properties

Various properties of the Kerr metric are immediate:

Property 10.2: asymptotic flatness of Kerr spacetime

The spacetime $(\mathcal{M}_{BL}, \mathbf{g})$ has two asymptotically flat ends: one in \mathcal{M}_I for $r \rightarrow +\infty$, which is equivalent to the asymptotics of a Schwarzschild spacetime of mass m and the other one in \mathcal{M}_{III} for $r \rightarrow -\infty$, which is equivalent to the asymptotics of a Schwarzschild spacetime of (negative!) mass $-m$.

Proof. For $r \rightarrow +\infty$ or $r \rightarrow -\infty$, one has $\rho^2 \sim r^2$ and $\rho^2/\Delta \sim (1 - 2m/r)^{-1}$, and $4amr/\rho^2 dt d\varphi \sim 4am/r^2 dt r d\varphi$, so that the metric (10.8) becomes

$$\mathbf{g} \simeq - \left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) + O\left(\frac{1}{r^2}\right) \quad (10.11)$$

For $r > 0$, we recognize the Schwarzschild metric expressed in Schwarzschild-Droste coordinates [Eq. (6.14)]. For $r < 0$, the change of coordinate $r' = -r$ leads to the Schwarzschild metric as well, but with the negative mass parameter $m' = -m$. \square

Property 10.3: stationarity and axisymmetry of Kerr spacetime

The spacetime $(\mathcal{M}_{BL}, \mathbf{g})$ admits two isometries: stationarity and axisymmetry, generated respectively by the Killing vectors

$$\xi := \partial_t \quad \text{and} \quad \eta := \partial_\varphi. \quad (10.12)$$

Proof. All the metric components $g_{\alpha\beta}$ in Eq. (10.8) are independent from t and φ . This implies that the vector fields (10.12) are Killing vectors (cf. Sec. 3.3.1). Since t spans \mathbb{R} , the isometry group generated by ξ is clearly the translation group $(\mathbb{R}, +)$. Moreover, in view of (10.11), we have $\xi \cdot \xi = g_{tt} < 0$ as $r \rightarrow +\infty$, which means that the Killing vector ξ 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 η is the rotation group $SO(2) = U(1)$. Moreover, ∂_φ is spacelike for $r \rightarrow +\infty$ (even for $r > 0$). Hence, the Kerr spacetime is axisymmetric (cf. the definition given in Sec. 5.3.6). \square

Property 10.4: non-staticity for $a \neq 0$

For $a \neq 0$, as we have assumed in (10.1), the Kerr spacetime is not static (cf. the definition in Sec. 5.2.1): the stationary Killing vector ξ is not hypersurface-orthogonal.

Proof. From Eq. (10.8), we have $\xi \cdot \eta = g_{t\varphi} \neq 0$ for $a \neq 0$. Since $\eta = \partial_\varphi$ is tangent to the hypersurfaces $t = \text{const}$, this implies that ξ is not normal to these hypersurfaces. That there does not exist any hypersurface family orthogonal to ξ can be seen from the non-vanishing of

the twist 3-form defined by Eq. (3.63): $\omega := \underline{\xi} \wedge d\underline{\xi}$. We have indeed (see the notebook D.5.1 for the computation)

$$\omega = \frac{2am\sin\theta}{\rho^4} \left[\sin\theta(a^2\cos^2\theta - r^2) dt \wedge dr \wedge d\varphi + 2\Delta r \cos\theta dt \wedge d\theta \wedge d\varphi \right].$$

Clearly $\omega \neq 0$ as soon as $a \neq 0$. According to the Frobenius theorem (see e.g. Eq. (B.3.6) in Wald's textbook [498]), it follows that $\underline{\xi}$ is not hypersurface-orthogonal. \square

Property 10.5: Schwarzschild limit

For $a \rightarrow 0$, the Kerr metric reduces to the Schwarzschild metric.

Proof. When $a \rightarrow 0$, we have $r_+ \rightarrow 2m$, $r_- \rightarrow 0$, $\rho^2 \sim r^2$, and $\rho^2/\Delta \sim (1 - 2m/r)^{-1}$, and we see on (10.8) that the Kerr metric reduces to the Schwarzschild metric expressed in Schwarzschild-Droste coordinates [Eq. (6.14)]. \square

Property 10.6: the double flat disk at $r = 0$

The metric induced by the Kerr metric on the hypersurface $r = 0$ of \mathcal{M}_{BL} is

$$h = -dt^2 + a^2(\cos^2\theta d\theta^2 + \sin^2\theta d\varphi^2). \quad (10.13)$$

For $a \neq 0$, h is a flat Lorentzian metric, which can be brought to a manifestly Minkowskian form via the change coordinates $x := a \sin\theta \cos\varphi$, $y := a \sin\theta \sin\varphi$: $h = -dt^2 + dx^2 + dy^2$. In particular, for a fixed value of t , the $r = 0$ subset

$$\mathcal{S}_{0,t} := \{p \in \mathcal{M}_{\text{III}}, r(p) = 0, t(p) = t\} \quad (10.14)$$

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 \mathcal{M}_{III} (cf. Remark 2 above).

Proof. On the hypersurface $r = 0$, we have $\rho^2 = a^2 \cos^2\theta$, $\Delta = a^2$ and $dr = 0$, so that the Kerr metric (10.8) induces the metric (10.13). That $\mathcal{S}_{0,t}$ is made of two disks is obvious from the range of variation of the coordinates (x, y) , which by construction obey $x^2 + y^2 < a^2$ (cf. Fig. C.3 in Appendix C, where the Kerr-Schild coordinates (x, y) coincide with the current coordinates (x, y) , up to some rotation). \square

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 g with respect to Boyer-Lindquist coordinates is deduced from (10.8); it takes a relatively simple form (see the notebook D.5.1 for the computation):

$$\det(g_{\alpha\beta}) = -\rho^4 \sin^2 \theta. \quad (10.15)$$

The inverse metric is (see the notebook D.5.1 for the computation)

$$g^{\alpha\beta} = \begin{pmatrix} -\frac{1}{\Delta} \left(r^2 + a^2 + \frac{2a^2 mr \sin^2 \theta}{\rho^2} \right) & 0 & 0 & -\frac{2amr}{\rho^2 \Delta} \\ 0 & \frac{\Delta}{\rho^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\rho^2} & 0 \\ -\frac{2amr}{\rho^2 \Delta} & 0 & 0 & \frac{1}{\Delta \sin^2 \theta} \left(1 - \frac{2mr}{\rho^2} \right) \end{pmatrix}. \quad (10.16)$$

10.2.4 Ergoregion

Let us investigate the causal character of the stationary Killing vector ξ . We have, according to (10.8) and (10.9),

$$\xi \cdot \xi = g_{tt} = -1 + \frac{2mr}{r^2 + a^2 \cos^2 \theta}.$$

Thus

$$\xi \text{ timelike} \iff r^2 - 2mr + a^2 \cos^2 \theta > 0 \iff r < r_{\mathcal{E}^-}(\theta) \quad \text{or} \quad r > r_{\mathcal{E}^+}(\theta),$$

with

$$r_{\mathcal{E}^\pm}(\theta) := m \pm \sqrt{m^2 - a^2 \cos^2 \theta}. \quad (10.17)$$

Comparing with (10.3), we note that

$$0 \leq r_{\mathcal{E}^-}(\theta) \leq r_- \leq m \leq r_+ \leq r_{\mathcal{E}^+}(\theta) \leq 2m, \quad (10.18)$$

with

$$r_{\mathcal{E}^-}(0) = r_{\mathcal{E}^-}(\pi) = r_- \quad \text{and} \quad r_{\mathcal{E}^-}(\pi/2) = 0 \quad (10.19a)$$

$$r_{\mathcal{E}^+}(0) = r_{\mathcal{E}^+}(\pi) = r_+ \quad \text{and} \quad r_{\mathcal{E}^+}(\pi/2) = 2m. \quad (10.19b)$$

Given the definition of \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{III} , we conclude that:

Property 10.7: ergoregion of Kerr spacetime

The stationary Killing vector ξ of Kerr spacetime obeys:

- ξ is timelike in the region of \mathcal{M}_I defined by $r > r_{\mathcal{E}^+}(\theta)$ and in the region of \mathcal{M}_{III} defined by $r < r_{\mathcal{E}^-}(\theta)$;
- ξ is null on the hypersurface \mathcal{E}^+ of \mathcal{M}_I defined by $r = r_{\mathcal{E}^+}(\theta)$ and on the hypersurface \mathcal{E}^- of \mathcal{M}_{III} defined by $r = r_{\mathcal{E}^-}(\theta)$;

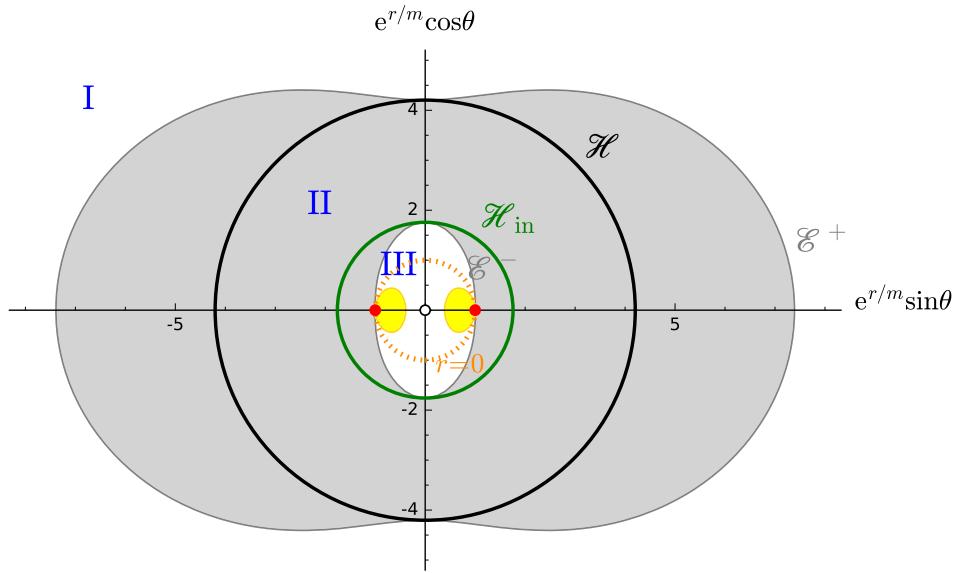


Figure 10.2: Meridional view of a section $t = \text{const}$ of Kerr spacetime with $a/m = 0.90$ in O'Neill exponential coordinates $x = e^{r/m} \sin \theta$ and $z = e^{r/m} \cos \theta$ (cf. Fig. 10.1). The right (resp. left) half of the figure corresponds to $\varphi = 0$ (resp. $\varphi = \pi$). The Roman numbers I, II, III denote the components \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{III} of the manifold \mathcal{M}_{BL} . The dotted orange circle marks the location of $r = 0$, while the small black circle at the center of the figure corresponds to $r \rightarrow -\infty$. The two red dots marks the curvature singularity \mathcal{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).

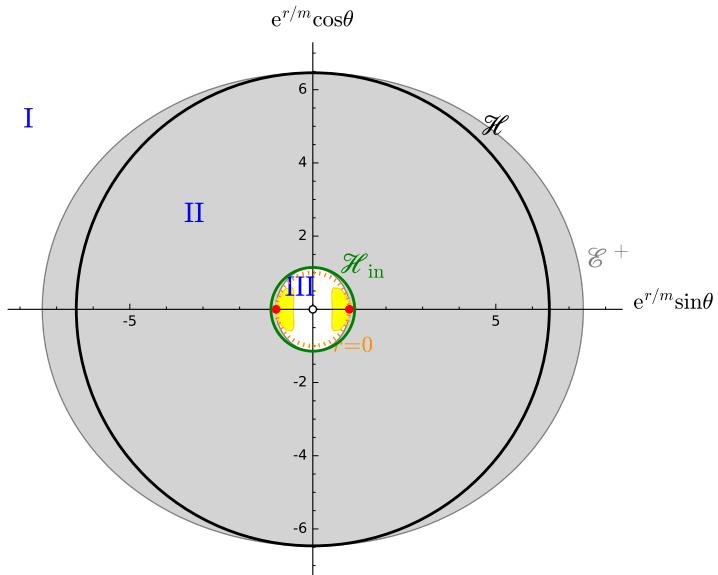


Figure 10.3: Same as Fig. 10.2 but for $a/m = 0.50$.

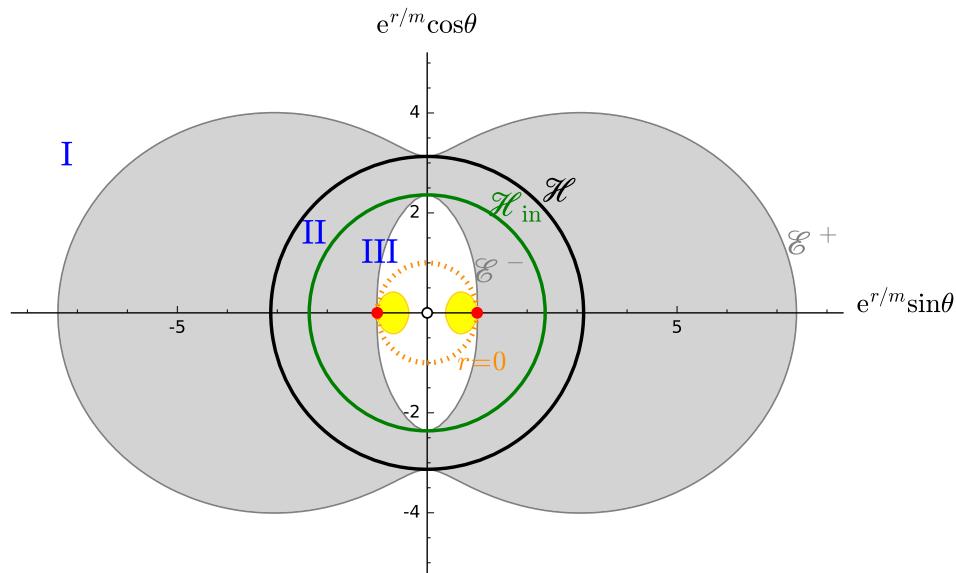


Figure 10.4: Same as Fig. 10.2 but for $a/m = 0.99$.

- ξ is spacelike in all \mathcal{M}_{II} and in the region \mathcal{G}^+ of \mathcal{M}_{I} defined by $r < r_{\mathcal{E}^+}(\theta)$, as well as in the region \mathcal{G}^- of \mathcal{M}_{III} defined by $r > r_{\mathcal{E}^-}(\theta)$.

According to the nomenclature introduced in Sec. 5.4.2, one calls \mathcal{E}^+ (resp. \mathcal{E}^-) the **outer ergosphere** (resp. **inner ergosphere**) and \mathcal{G}^+ (resp. \mathcal{G}^-) the **outer ergoregion** (resp. **inner ergoregion**). The part of \mathcal{M}_{BL} where ξ is spacelike, i.e. $\mathcal{G} = \mathcal{G}^+ \cup \mathcal{M}_{\text{II}} \cup \mathcal{G}^-$, is called the **ergoregion**.

Following the standard notation in topology, we shall denote by $\overline{\mathcal{G}}$ the closure of \mathcal{G} , i.e. the union of the ergoregion and the two ergospheres:

$$\overline{\mathcal{G}} := \mathcal{G} \cup \mathcal{E}^- \cup \mathcal{E}^+. \quad (10.20)$$

The locus where the Killing vector ξ is timelike is then $\mathcal{M}_{\text{BL}} \setminus \overline{\mathcal{G}}$.

The ergoregion is depicted in Figs. 10.2–10.4. It is worth noticing that the locations of the inner and outer ergospheres in the equatorial plane ($\theta = \pi/2$) do not depend on the spin parameter a : $r_{\mathcal{E}^-}(\pi/2) = 0$ [Eq. (10.19a)] and $r_{\mathcal{E}^+}(\pi/2) = 2m$ [Eq. (10.19b)], a feature that appears clearly in Figs. 10.2–10.4, given that $e^2 \simeq 7.39$.

Remark 5: By taking the limit $a \rightarrow 0$ in Eq. (10.17), we get $r_{\mathcal{E}^+}(\theta) \rightarrow 2m$. Since $r_+ \rightarrow 2m$ in the same limit, we conclude that, for a Schwarzschild black hole, the outer ergosphere coincides with the event horizon. Consequently, there is no outer ergoregion for such a black hole: $\mathcal{G}^+ = \emptyset$.

Remark 6: Sometimes the word **ergosurface** is used instead of **ergosphere**. One may encounter as well the names *infinite redshift surface* (e.g. [494]) and *static limit* (cf. Sec. 10.7.2 below) for **ergosphere**, especially in old texts.

Remark 7: By construction, the Killing vector ξ is null on the ergospheres \mathcal{E}^+ and \mathcal{E}^- . It is moreover tangent to these hypersurfaces (see below). However, for $a \neq 0$, \mathcal{E}^+ and \mathcal{E}^- are *not* Killing horizons,

as defined in Sec. 3.3.2, since ξ fails to be normal to them, or equivalently, \mathcal{E}^+ and \mathcal{E}^- are not null hypersurfaces. They are actually timelike hypersurfaces, except on the rotation axis. This can be seen by considering \mathcal{E}^\pm as the level set $f = 0$ of the scalar field $f := r - r_{\mathcal{E}^\pm}(\theta) = r - m \mp (m^2 - a^2 \cos^2 \theta)^{1/2}$ [cf. Eq. (10.17)]. The differential of f is $df = dr \pm a^2 \sin \theta \cos \theta (m^2 - a^2 \cos^2 \theta)^{-1/2} d\theta$, so that $\langle df, \xi \rangle = 0$, which shows that ξ is tangent to \mathcal{E}^\pm . The scalar square of the gradient vector field $\vec{\nabla} f$ associated to df by metric duality is easily evaluated via the inverse metric (10.16):

$$\vec{\nabla} f \cdot \vec{\nabla} f = g^{\mu\nu} \partial_\mu f \partial_\nu f = \frac{1}{\rho^2} \left(\Delta + \frac{a^4 \sin^2 \theta \cos^2 \theta}{m^2 - a^2 \cos^2 \theta} \right) \stackrel{\mathcal{E}^\pm}{=} \frac{a^2 m^2 \sin^2 \theta}{\rho^2 (m^2 - a^2 \cos^2 \theta)}.$$

Away from the rotation axis, i.e. for $\sin \theta \neq 0$, we have clearly $\vec{\nabla} f \cdot \vec{\nabla} f > 0$ for $a \neq 0$, which shows that the normal $\vec{\nabla} f$ to \mathcal{E}^\pm is spacelike, making \mathcal{E}^\pm a timelike hypersurface. Further details about the outer ergosphere \mathcal{E}^+ , especially the 2-dimensional geometry of its time slices, are discussed in Ref. [294].

We shall see in Sec. 11.3.2 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, η . From (10.8) and (10.9), we have

$$\eta \cdot \eta = g_{\varphi\varphi} = \left(r^2 + a^2 + \frac{2a^2mr \sin^2 \theta}{r^2 + a^2 \cos^2 \theta} \right) \sin^2 \theta. \quad (10.21)$$

Hence

$$\eta \text{ spacelike} \iff (r^2 + a^2)(r^2 + a^2 \cos^2 \theta) + 2a^2mr \sin^2 \theta > 0.$$

For $\theta \rightarrow 0$ or $\theta \rightarrow \pi$, 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 \mathcal{T} defined by $r_{\mathcal{T}}(\theta) < r < 0$ for some negative function $r_{\mathcal{T}}(\theta)$, such that $g_{\varphi\varphi} < 0$. Since \mathcal{T} corresponds to negative values of r , we have $\mathcal{T} \subset \mathcal{M}_{\text{III}}$. Hence we conclude:

Property 10.8: Carter time machine

The axisymmetric Killing vector η of Kerr spacetime obeys:

- η is spacelike in all \mathcal{M}_{I} and \mathcal{M}_{II} , as well as outside the region \mathcal{T} in \mathcal{M}_{III} ;
- η is timelike in the subset \mathcal{T} of \mathcal{M}_{III} ;
- η is null at the boundary of \mathcal{T} .

The region \mathcal{T} is called **Carter time machine**. This name stems from the fact that thanks to \mathcal{T} , there is a future-directed timelike curve connecting any two points of \mathcal{M}_{III} .

See e.g. Proposition 2.4.7 of O'Neill's textbook [390] for a demonstration, or Carter's original article [89]) for the proof of the last sentence. The Carter time machine \mathcal{T} is depicted in yellow in the meridional diagrams of Figs. 10.2-10.4.

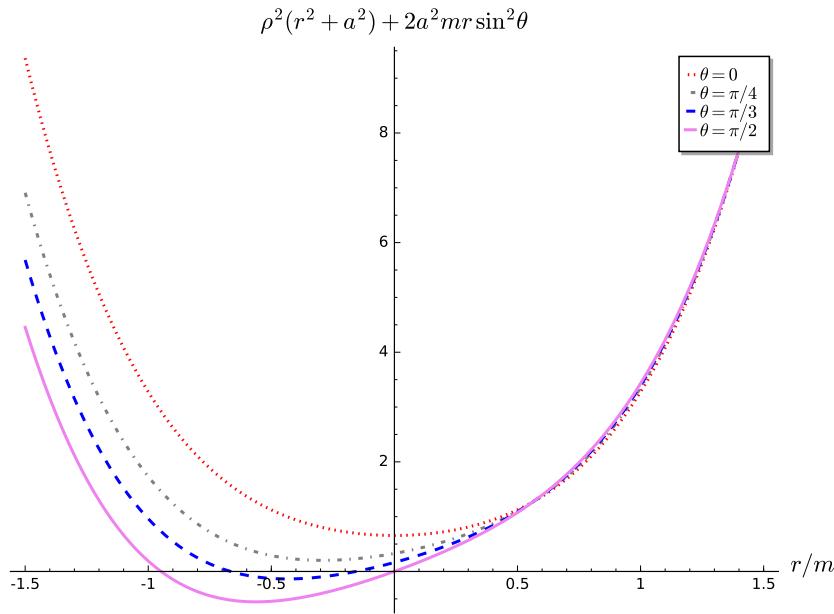


Figure 10.5: Graph of the function giving the sign of $g_{\varphi\varphi}$ for $a = 0.9m$ and various values of θ .

10.2.6 Singularities

The components $g_{\alpha\beta}$ of the Kerr metric as given by (10.8) are diverging at various locations:

- when $\rho^2 \rightarrow 0$, which, given (10.9) and assuming $a \neq 0$, is equivalent to approaching the cylinder \mathcal{R} defined by (10.4);
- when $\Delta \rightarrow 0$, which, given (10.10), is equivalent to either $r \rightarrow r_-$ or $r \rightarrow r_+$; the first case corresponds to the boundary (within $\mathbb{R}^2 \times \mathbb{S}^2$) between \mathcal{M}_{II} and \mathcal{M}_{III} and the second case to the boundary between \mathcal{M}_{I} and \mathcal{M}_{II} .

The divergence when $\rho^2 \rightarrow 0$ corresponds to a *curvature singularity*:

Property 10.9: ring singularity

The curvature of Kerr spacetime, as measured by the Kretschmann scalar $K := R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma}$ [cf Eq. (6.44)], diverges for $\rho^2 \rightarrow 0$, i.e. for $r \rightarrow 0$ and $\theta \rightarrow \pi/2$. Accordingly the subset \mathcal{R} of $\mathbb{R}^2 \times \mathbb{S}^2$ defined by $\rho^2 = 0$ [Eq. (10.4)] is called the **ring singularity** of Kerr spacetime; the word *ring* reflects that $t = \text{const}$ sections of \mathcal{R} are circles [cf. Eq. (10.6)].

Proof. The Kretschmann scalar expressed in terms of Boyer-Lindquist coordinates is (cf. the notebook D.5.1 for the computation)

$$K = 48 \frac{m^2}{\rho^{12}} (r^6 - 15r^4 a^2 \cos^2 \theta + 15r^2 a^4 \cos^4 \theta - a^6 \cos^6 \theta). \quad (10.22)$$

The value for $\theta = \pi/2$ is thus $K = 48m^2/r^6$, which clearly diverges for $r \rightarrow 0$ (i.e. $\rho^2 \rightarrow 0$). \square

See Ref. [123] for an extended discussion of the ring singularity.

On the contrary, the divergence of the metric components (10.8) when $\Delta \rightarrow 0$ corresponds to a mere *coordinate singularity*, i.e. to a pathology of Boyer-Lindquist coordinates. The latter can be cured by switching to other coordinates, as we shall see in the next section.

10.3 Extension of the spacetime manifold through $\Delta = 0$

10.3.1 Advanced Kerr coordinates

The **advanced 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

$$\boxed{\mathbf{d}v = \mathbf{d}t + \frac{r^2 + a^2}{\Delta} \mathbf{d}r} \quad (10.23a)$$

$$\boxed{\mathbf{d}\tilde{\varphi} = \mathbf{d}\varphi + \frac{a}{\Delta} \mathbf{d}r}. \quad (10.23b)$$

If $a = 0$, we note that the advanced Kerr coordinates are nothing but the *null ingoing Eddington-Finkelstein coordinates* on Schwarzschild spacetime (cf. Sec. 6.3.2 and compare (10.23a) with (6.28)).

Remark 1: In the literature, the advanced Kerr coordinates are often simply called *Kerr coordinates* (e.g. [370, 463]). However, we prefer to keep this short name for closely related coordinates to be introduced in Sec. 10.3.3. Both systems originate in Kerr's seminal article [311] (cf. the historical notes on p. 334 and p. 338). The advanced Kerr coordinates are also called *Kerr-star coordinates* by O'Neill [390].

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

$$\boxed{v = t + r + \frac{m}{\sqrt{m^2 - a^2}} \left(r_+ \ln \left| \frac{r - r_+}{2m} \right| - r_- \ln \left| \frac{r - r_-}{2m} \right| \right)} \quad (10.24a)$$

$$\boxed{\tilde{\varphi} = \varphi + \frac{a}{2\sqrt{m^2 - a^2}} \ln \left| \frac{r - r_+}{r - r_-} \right|}, \quad (10.24b)$$

up to some additive constants. When $a \rightarrow 0$, we have $r_+ \rightarrow 2m$ and $r_- \rightarrow 0$ and by comparing Eq. (10.24a) with Eq. (6.27), we recover that the advanced Kerr coordinates reduces to the null ingoing Eddington-Finkelstein ones in this limit.

The expression of the metric tensor \mathbf{g} in terms of the advanced Kerr coordinates $(v, r, \theta, \tilde{\varphi})$ is computed from that in terms the Boyer-Lindquist ones, as given by Eq. (10.8), via (10.23).

One gets (cf. Appendix D or Eq. (5.31) of Ref. [265], or Lemma 2.5.2 of [390]):

$$\begin{aligned} \mathbf{g} = & -\left(1 - \frac{2mr}{\rho^2}\right) \mathbf{d}v^2 + 2\mathbf{d}v \mathbf{d}r - \frac{4amr \sin^2 \theta}{\rho^2} \mathbf{d}v \mathbf{d}\tilde{\varphi} \\ & - 2a \sin^2 \theta \mathbf{d}r \mathbf{d}\tilde{\varphi} + \rho^2 \mathbf{d}\theta^2 + \left(r^2 + a^2 + \frac{2a^2 mr \sin^2 \theta}{\rho^2}\right) \sin^2 \theta \mathbf{d}\tilde{\varphi}^2. \end{aligned} \quad (10.25)$$

The inverse metric is pretty simple (cf. the notebook D.5.2-1):

$$g^{\hat{\alpha}\hat{\beta}} = \frac{1}{\rho^2} \begin{pmatrix} a^2 \sin^2 \theta & r^2 + a^2 & 0 & a \\ r^2 + a^2 & \Delta & 0 & a \\ 0 & 0 & 1 & 0 \\ a & a & 0 & \frac{1}{\sin^2 \theta} \end{pmatrix}. \quad (10.26)$$

We note that the metric components (10.25) do not show any divergence when $\Delta \rightarrow 0$, 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)

$$\mathcal{H} := \{p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = r_+\} \quad (10.27)$$

and

$$\mathcal{H}_{\text{in}} := \{p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = r_-\}. \quad (10.28)$$

The hypersurface \mathcal{H} is actually the boundary between the regions \mathcal{M}_I and \mathcal{M}_{II} , while \mathcal{H}_{in} is the boundary between \mathcal{M}_{II} and \mathcal{M}_{III} (cf. Eq. (10.7) and Fig. 10.2). We thus consider

$$\boxed{\mathcal{M} := \mathcal{M}_{\text{BL}} \cup \mathcal{H} \cup \mathcal{H}_{\text{in}} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathcal{R}} \quad (10.29)$$

as the spacetime manifold. In order for \mathbf{g} defined by (10.25) to be a well-defined metric on \mathcal{M} , it does not suffice that the components $g_{\hat{\alpha}\hat{\beta}}$ do not diverge at \mathcal{H} and \mathcal{H}_{in} : one shall check as well that the bilinear form \mathbf{g} is non-degenerate there. This is easily proven by considering the determinant of the metric components (10.25), which turns out to have a simple form:

$$\det(g_{\hat{\alpha}\hat{\beta}}) = -\rho^4 \sin^2 \theta. \quad (10.30)$$

Except at $\theta = 0$ and $\theta = \pi$ (the usual singularities of spherical coordinates), we have $\det(g_{\hat{\alpha}\hat{\beta}}) \neq 0$ everywhere on \mathcal{M} , since ρ vanishes only on \mathcal{R} , which is excluded from \mathcal{M} . Hence we conclude

Property 10.10: Kerr spacetime

The symmetric bilinear form \mathbf{g} given by Eq. (10.25) is regular and non-degenerate on the manifold \mathcal{M} defined by (10.29) and thus $(\mathcal{M}, \mathbf{g})$ is a well-behaved spacetime – our **Kerr spacetime** from now on. We note that, contrary to \mathcal{M}_{BL} , \mathcal{M} is a connected manifold.

We deduce from (10.23) that

$$\frac{\partial v}{\partial t} \Big|_{r,\theta,\varphi} = 1, \quad \frac{\partial v}{\partial r} \Big|_{t,\theta,\varphi} = \frac{r^2 + a^2}{\Delta}, \quad \frac{\partial \tilde{\varphi}}{\partial r} \Big|_{t,\theta,\varphi} = \frac{a}{\Delta}, \quad \frac{\partial \tilde{\varphi}}{\partial \varphi} \Big|_{t,r,\theta} = 1.$$

It follows from the chain rule that the advanced Kerr coordinate frame is related to the Boyer-Lindquist coordinate frame by

$$\partial_v = \partial_t \tag{10.31a}$$

$$\partial_{\hat{r}} = \partial_r - \frac{a^2 + r^2}{\Delta} \partial_t - \frac{a}{\Delta} \partial_\varphi \tag{10.31b}$$

$$\partial_\theta = \partial_\theta \tag{10.31c}$$

$$\partial_{\tilde{\varphi}} = \partial_\varphi. \tag{10.31d}$$

Note that we are using the notation $\partial_{\hat{r}}$ for the $\partial/\partial r$ vector of the advanced Kerr coordinates $(x^\alpha) = (v, r, \theta, \tilde{\varphi})$, to distinguish it from the $\partial/\partial r$ vector of Boyer-Lindquist coordinates.

Remark 2: Contrary to the Schwarzschild case (cf. Sec. 6.3.2), v is not a null coordinate for $a \neq 0$, but rather a spacelike one, i.e. the hypersurfaces of constant v are timelike (cf. the definitions given in Sec. A.3.2). Indeed, we read on (10.26) that $g^{vv} = a^2 \sin^2 \theta / \rho^2$. For $a \neq 0$, this implies $g^{vv} > 0$ outside the rotation axis. The criterion (A.56c) allows us to conclude that v is a spacelike coordinate.

Historical note : The advanced Kerr coordinates are those in which Roy P. Kerr originally presented his solution in 1963 [311]. As noted by Kerr himself later on [312], 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. [312]). Actually, it was shown in 1964 by Robert H. Boyer and T.G. Price [72] that the angular momentum about the rotation axis of the Kerr solution is $J = -am$, where a is Kerr’s one. Taking this into account, the correspondence between our notations and those of Kerr’s article [311] is $v \leftrightarrow u$ and $a \leftrightarrow -a$. Then one can check that the metric (10.25) coincides with that given by an unnumbered (!) equation in Kerr’s article [311].

10.3.2 Time orientation of Kerr spacetime

We read on (10.25) that $\mathbf{g}(\partial_{\hat{r}}, \partial_{\hat{r}}) = g_{rr} = 0$, which implies that $\partial_{\hat{r}}$ is a global null vector field on \mathcal{M} . We may then use it to set the time orientation of $(\mathcal{M}, \mathbf{g})$ (cf. Sec. 1.2.2):

Property 10.11: time orientation of Kerr spacetime

The Kerr spacetime $(\mathcal{M}, \mathbf{g})$ is time-orientable, and we choose its time orientation such that

$$\mathbf{k} := -\partial_{\hat{r}} \tag{10.32}$$

is a future-directed null vector field in all \mathcal{M} .

Remark 3: The minus sign in the above definition, along with Eq. (10.31b), ensures that

$$\mathbf{k} \sim \partial_t - \partial_r \quad \text{when } r \rightarrow +\infty,$$

which shows that the time orientation set by \mathbf{k} agrees asymptotically with that of $\xi = \partial_t$. The latter vector field could not have been chosen to set the time orientation of (\mathcal{M}, g) since it is not causal everywhere, being spacelike in the ergoregion.

The field lines of \mathbf{k} are future-directed null curves, which may be qualified of *ingoing* since, by definition, $-\partial_{\hat{r}}$ points towards decreasing values of r . Note that, by the very definition of the coordinate vector $\partial_{\hat{r}}$, the values of the coordinates $(v, \theta, \tilde{\varphi})$ are fixed along each of these null curves. We may therefore denote them by $\mathcal{L}_{(v, \theta, \tilde{\varphi})}^{\text{in}}$. We shall see in Sec. 10.4 that each $\mathcal{L}_{(v, \theta, \tilde{\varphi})}^{\text{in}}$ is actually a null geodesic.

Decaying of r towards the future in \mathcal{M}_{II}

The scalar square of the vector ∂_r of Boyer-Lindquist coordinates is read from the metric components (10.8): $\partial_r \cdot \partial_r = g(\partial_r, \partial_r) = g_{rr} = \rho^2/\Delta$. Since Δ is positive in \mathcal{M}_{I} and \mathcal{M}_{III} and negative in \mathcal{M}_{II} , we conclude that ∂_r is spacelike in $\mathcal{M}_{\text{I}} \cup \mathcal{M}_{\text{III}}$ and timelike in \mathcal{M}_{II} . Moreover, the above choice of time orientation leads to

Property 10.12

In region \mathcal{M}_{II} , the vector ∂_r of Boyer-Lindquist coordinates is a past-directed timelike vector.

Proof. Applying Lemma 1.2 (Sec. 1.2.2) with $\mathbf{u} = \mathbf{k}$ and $\mathbf{v} = \partial_r$, we get that ∂_r is past-directed iff $g(\mathbf{k}, \partial_r) > 0$. Now, in terms of the Boyer-Lindquist components (10.8), $g(\mathbf{k}, \partial_r) = g_{\mu r} k^\mu = g_{rr} k^r = (\rho^2/\Delta) k^r$. The Boyer-Lindquist component k^r is given by Eq. (10.31b) where $\partial_{\hat{r}} = -\mathbf{k}$; we get $k^r = -1$. Hence $g(\mathbf{k}, \partial_r) = -\rho^2/\Delta > 0$ in \mathcal{M}_{II} , for $\Delta < 0$ there. \square

An important consequence of the above property is

Property 10.13: decreasing of r in \mathcal{M}_{II}

In region \mathcal{M}_{II} , the coordinate r must decrease towards the future along any causal (i.e. timelike or null) worldline.

Proof. Let \mathcal{L} be a causal curve in region \mathcal{M}_{II} and λ a parameter along \mathcal{L} increasing towards the future. The associated tangent vector $\mathbf{v} = dx/d\lambda$ is then future-directed. According to the above result, $-\partial_r$ is a future-directed timelike vector in \mathcal{M}_{II} , so that we can apply Lemma 1.1 (Sec. 1.2.2) with $\mathbf{u} = -\partial_r$ and get $g(-\partial_r, \mathbf{v}) < 0$. Now, using Boyer-Lindquist components, we have

$$g(-\partial_r, \mathbf{v}) = -g_{r\mu} v^\mu = -g_{rr} v^r = -g_{rr} \frac{dr}{d\lambda} = -\frac{\rho^2}{\Delta} \frac{dr}{d\lambda}.$$

Since $-\rho^2/\Delta > 0$ in \mathcal{M}_{II} , $g(-\partial_r, \mathbf{v}) < 0$ is thus equivalent to $dr/d\lambda < 0$, which proves that r is decreasing along \mathcal{L} as λ increases. \square

10.3.3 Kerr coordinates

As in Sec. 6.3.2, we shall move from the coordinate v to a (asymptotically) timelike one by setting

$$\boxed{\tilde{t} = v - r} \iff \boxed{v = \tilde{t} + r} \quad (10.33)$$

so that v appears as the advanced time $\tilde{t} + r$ (compare with Eq. (6.30)). We thus consider the coordinates $(x^{\tilde{\alpha}}) = (\tilde{t}, r, \theta, \tilde{\varphi})$, 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 $d\tilde{t} = dv - dr$:

$$\boxed{d\tilde{t} = dt + \frac{2mr}{\Delta} dr} \quad (10.34a)$$

$$\boxed{d\tilde{\varphi} = d\varphi + \frac{a}{\Delta} dr}. \quad (10.34b)$$

The integrated version is obtained by substituting (10.33) in Eq. (10.24):

$$\boxed{\tilde{t} = t + \frac{m}{\sqrt{m^2 - a^2}} \left(r_+ \ln \left| \frac{r - r_+}{2m} \right| - r_- \ln \left| \frac{r - r_-}{2m} \right| \right)} \quad (10.35a)$$

$$\boxed{\tilde{\varphi} = \varphi + \frac{a}{2\sqrt{m^2 - a^2}} \ln \left| \frac{r - r_+}{r - r_-} \right|}. \quad (10.35b)$$

Since the transformation (10.33) leads to $dv = d\tilde{t} + dr$, the expression of the metric with respect to the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ is easily deduced from (10.25):

$$\boxed{\begin{aligned} g = & - \left(1 - \frac{2mr}{\rho^2} \right) d\tilde{t}^2 + \frac{4mr}{\rho^2} d\tilde{t} dr - \frac{4amr \sin^2 \theta}{\rho^2} d\tilde{t} d\tilde{\varphi} \\ & + \left(1 + \frac{2mr}{\rho^2} \right) dr^2 - 2a \left(1 + \frac{2mr}{\rho^2} \right) \sin^2 \theta dr d\tilde{\varphi} \\ & + \rho^2 d\theta^2 + \left(r^2 + a^2 + \frac{2a^2 mr \sin^2 \theta}{\rho^2} \right) \sin^2 \theta d\tilde{\varphi}^2. \end{aligned}} \quad (10.36)$$

Since we kept r, θ and $\tilde{\varphi}$ and simply changed v to \tilde{t} via (10.33) when moving from the advanced Kerr coordinates to the Kerr ones, we easily get the link between the two coordinate frames:

$$\partial_{\tilde{t}} = \partial_v \quad (10.37a)$$

$$\partial_{\tilde{r}} = \partial_v + \partial_{\hat{r}} \quad (10.37b)$$

$$\partial_{\theta} = \partial_{\theta} \quad (10.37c)$$

$$\partial_{\tilde{\varphi}} = \partial_{\tilde{\varphi}}. \quad (10.37d)$$

Note that we have denoted by $\partial_{\tilde{r}}$ 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 ∂_r of

the advanced Kerr coordinates $(x^{\hat{\alpha}}) = (v, r, \theta, \tilde{\varphi})$, as well as from the coordinate vector ∂_r of the Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$.

By combining (10.31) and (10.37), we get the relation between the Kerr coordinate frame and the Boyer-Lindquist coordinate frame:

$$\partial_{\tilde{t}} = \partial_t \quad (10.38a)$$

$$\partial_{\tilde{r}} = \partial_r - \frac{2mr}{\Delta} \partial_t - \frac{a}{\Delta} \partial_\varphi \quad (10.38b)$$

$$\partial_\theta = \partial_\theta \quad (10.38c)$$

$$\partial_{\tilde{\varphi}} = \partial_\varphi. \quad (10.38d)$$

We notice on (10.38a) and (10.38d) that the coordinate frame vectors $\partial_{\tilde{t}}$ and $\partial_{\tilde{\varphi}}$ coincide with the Killing vectors ξ and η :

$$\boxed{\partial_{\tilde{t}} = \xi} \quad \text{and} \quad \boxed{\partial_{\tilde{\varphi}} = \eta}. \quad (10.39)$$

That $\partial_{\tilde{t}}$ and $\partial_{\tilde{\varphi}}$ 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:

$$\det(g_{\tilde{\alpha}\tilde{\beta}}) = -\rho^4 \sin^2 \theta. \quad (10.40)$$

The inverse metric also takes a rather simple form in terms of Kerr coordinates (see the notebook D.5.2-2 for the computation):

$$g^{\tilde{\alpha}\tilde{\beta}} = \begin{pmatrix} -1 - \frac{2mr}{\rho^2} & \frac{2mr}{\rho^2} & 0 & 0 \\ \frac{2mr}{\rho^2} & \frac{\Delta}{\rho^2} & 0 & \frac{a}{\rho^2} \\ 0 & 0 & \frac{1}{\rho^2} & 0 \\ 0 & \frac{a}{\rho^2} & 0 & \frac{1}{\rho^2 \sin^2 \theta} \end{pmatrix}. \quad (10.41)$$

Comparing (10.36) 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: $g_{\tilde{t}\tilde{r}}$ and $g_{\tilde{r}\tilde{\varphi}}$. However the determinant (10.40) and the inverse metric (10.41) are pretty simple. Moreover 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 \mathcal{H} and \mathcal{H}_{in} , contrary to Boyer-Lindquist coordinates. The last feature is all the more important that \mathcal{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 of the black hole feature, by means of Kerr coordinates.

Remark 4: The coordinate \tilde{t} is not everywhere timelike, i.e. the hypersurfaces of constant \tilde{t} are not everywhere spacelike. Indeed, according to the criterion (A.56), \tilde{t} is timelike iff $g^{\tilde{t}\tilde{t}} < 0$ and we read on Eq. (10.41) that $g^{\tilde{t}\tilde{t}} = -1 - 2mr/\rho^2$. Hence \tilde{t} is timelike iff $\rho^2 + 2mr > 0$, or equivalently $r^2 + 2mr + a^2 \cos^2 \theta > 0$. This quadratic polynomial in r is positive everywhere except in the region of \mathcal{M}_{III} defined by

$$-m - \sqrt{m^2 - a^2 \cos^2 \theta} \leq r \leq -m + \sqrt{m^2 - a^2 \cos^2 \theta}. \quad (10.42)$$

Note that this region is contained in the negative- r part of \mathcal{M}_{III} . We conclude that the coordinate \tilde{t} is timelike in \mathcal{M}_{I} , \mathcal{M}_{II} and in the part of \mathcal{M}_{III} outside the region defined by (10.42). Consequently, the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ are related to a 3+1 slicing of spacetime only outside the region (10.42). By **3+1 slicing**, it is meant a foliation of \mathcal{M} by spacelike hypersurfaces (see e.g. [226]).

Historical note : The coordinate \tilde{t} has been introduced by Roy Kerr in the 1963 discovery article [311], by exactly the same transformation as (10.33) (\tilde{t} is denoted by t and v by u in Ref. [311]). 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 [311] and not in the subsequent article by Kerr and Schild [313] (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. 334), called here *advanced Kerr coordinates*, and the Kerr-Schild coordinates (\tilde{t}, x, y, z) . The Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ have been first explicitly considered by Robert Boyer and Richard Lindquist in 1966 [71], 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 metric (10.36) coincides with Eq. (2.7) of Boyer and Lindquist's article [71], 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. 334).

10.4 Principal null geodesics

10.4.1 Ingoing principal null geodesics

We have seen that the advanced Kerr coordinates $(v, r, \theta, \tilde{\varphi})$ introduced in Sec. 10.3.1 are such that the curves $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ defined by $(v, \theta, \tilde{\varphi}) = \text{const}$ are null curves (cf. Sec. 10.3.2). Their future-directed tangent vector field is $\mathbf{k} = -\partial_{\tilde{r}}$ [Eq. (10.32)], which can be expressed in terms of the Kerr basis via (10.37):

$$\boxed{\mathbf{k} = \partial_{\tilde{t}} - \partial_{\tilde{r}}}. \quad (10.43)$$

The 1-form $\underline{\mathbf{k}}$ associated to \mathbf{k} by \mathbf{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.

$$\boxed{\underline{\mathbf{k}} = -d\tilde{t} - dr + a \sin^2 \theta d\tilde{\varphi}}. \quad (10.44)$$

A direct computation (cf. the notebook D.5.2-2) shows that

$$\nabla_{\underline{\mathbf{k}}} \underline{\mathbf{k}} = 0. \quad (10.45)$$

It follows that each curve $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ is a geodesic and that the parameter λ associated with \mathbf{k} is an affine parameter of this geodesic (cf. Sec. B.2.1 in Appendix B). Since Eq. (10.43) implies $k^r = dr/d\lambda = -1$, we have, up to some additive constant,

$$\lambda = -r. \quad (10.46)$$

The geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ are called the **ingoing principal null geodesics**. The qualifier *ingoing* stems from the fact that r is decreasing towards the future along $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$, which is an immediate

consequence of $\lambda = -r$ being a future-directed parameter along $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$. The qualifier *principal* arises from a specific connection between \mathbf{k} and the Weyl conformal curvature tensor \mathbf{C} of Kerr spacetime (cf. Sec. A.5.4), namely:

$$C_{\mu\nu[\beta}^{\alpha} k_{\gamma]} k^{\mu} k^{\nu} = 0 \quad \text{and} \quad {}^*C_{\mu\nu[\beta}^{\alpha} k_{\gamma]} k^{\mu} k^{\nu} = 0, \quad (10.47)$$

where ${}^*\mathbf{C}$ stands for the ***dual of the Weyl tensor***:

$${}^*C_{\beta\gamma\delta}^{\alpha} := \frac{1}{2} C_{\beta\mu\nu}^{\alpha} \epsilon^{\mu\nu}_{\gamma\delta}, \quad (10.48)$$

ϵ being the Levi-Civita tensor (cf. Sec. A.3.4). In view of (10.47), one says that the vector field \mathbf{k} constitutes a ***doubly degenerate, principal null direction*** of \mathbf{C} (see e.g. Chap. 5 of O'Neill textbook [390] for details). We note that the ingoing principal null geodesics form a *congruence*: through each point of \mathcal{M} , there is one, and only one, curve $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$.

Remark 1: For the Kerr spacetime, as for any solution of the vacuum Einstein equation ($\mathbf{R} = 0$), the Weyl conformal curvature tensor \mathbf{C} is equal to the Riemann curvature tensor **Riem**: setting $\mathbf{R} = 0$ in Eq. (A.114) yields **Riem** = \mathbf{C} .

Remark 2: At the Schwarzschild limit, $a = 0$, $\tilde{\varphi} = \varphi$ and the ingoing principal null geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ reduce to the ingoing radial null geodesics $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$ discussed in Secs. 6.3.1 and 6.3.5 [compare Eqs. (6.48) and (10.43)].

The components of \mathbf{k} with respect to the Boyer-Lindquist coordinate frame are immediately deduced from Eq. (10.31b) and $\mathbf{k} = -\partial_{\hat{r}}$:

$$\mathbf{k} = \frac{r^2 + a^2}{\Delta} \partial_t - \partial_r + \frac{a}{\Delta} \partial_{\varphi}. \quad (10.49)$$

Substituting Eqs. (10.34) for $d\tilde{t}$ and $d\tilde{\varphi}$ in Eq. (10.44), we get the expression of the associated 1-form $\underline{\mathbf{k}}$ with respect to the Boyer-Lindquist coordinate coframe:

$$\underline{\mathbf{k}} = -dt - \frac{\rho^2}{\Delta} dr + a \sin^2 \theta d\varphi. \quad (10.50)$$

Remark 3: Expressions (10.49) and (10.50) are singular on the two horizons \mathcal{H} and \mathcal{H}_{in} , where $\Delta = 0$. This reflects the singularity of Boyer-Lindquist coordinates on \mathcal{H} and \mathcal{H}_{in} , not any pathology of the vector field \mathbf{k} . Indeed, the expressions of \mathbf{k} and $\underline{\mathbf{k}}$ in terms of Kerr coordinates [Eqs. (10.43) and (10.44)] are perfectly regular in all \mathcal{M} .

10.4.2 Outgoing principal null geodesics

One can construct a second congruence of principal null geodesics by considering the *retarded* Kerr coordinates instead of the advanced ones discussed in Sec. 10.3.1. The ***retarded Kerr***

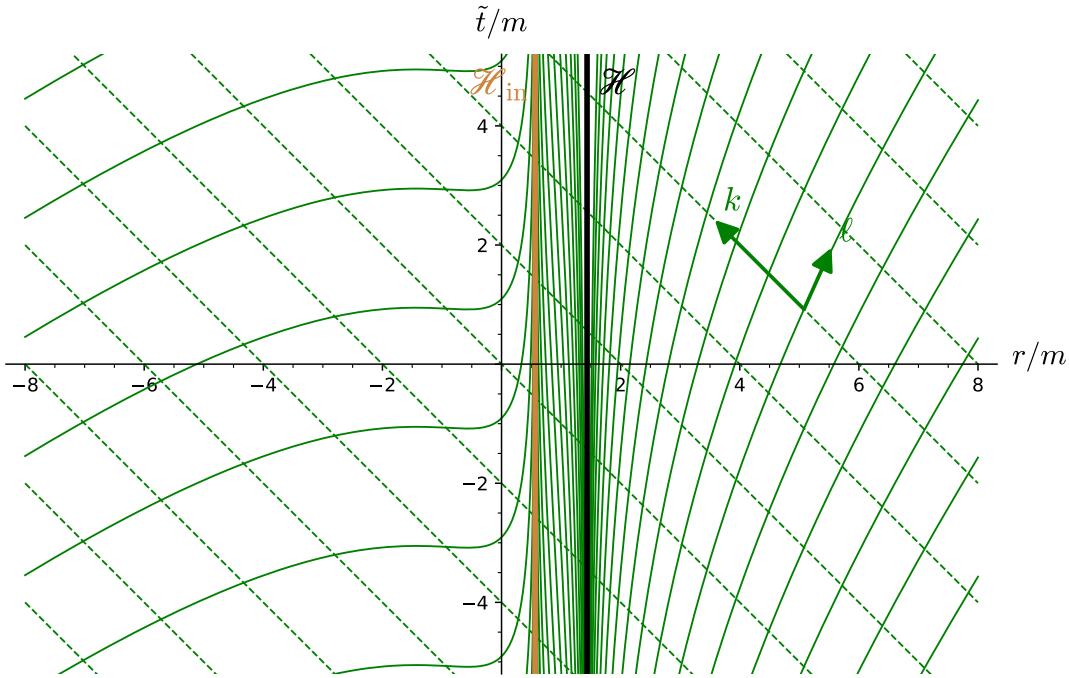


Figure 10.6: Principal null geodesics of Kerr spacetime viewed in terms of the Kerr coordinates (\tilde{t}, \tilde{r}) for $a/m = 0.9$. The solid (resp. dashed) curves correspond to outgoing geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ (resp. ingoing geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$), as given by Eq. (10.53) with $u = \text{const}$ (resp. Eq. (10.33) 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 \mathcal{H} (resp. \mathcal{H}_{in}) when $r \rightarrow r_+$ (resp. $r \rightarrow r_-$). 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.5.3]

coordinates $(u, r, \theta, \tilde{\varphi})$ are defined by relations to Boyer-Lindquist coordinates that are similar to (10.23), up to a change of sign:

$$du = dt - \frac{r^2 + a^2}{\Delta} dr \quad (10.51a)$$

$$d\tilde{\varphi} = d\varphi - \frac{a}{\Delta} dr. \quad (10.51b)$$

The coordinates $(u, r, \theta, \tilde{\varphi})$ 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 $(u, \tilde{\varphi})$ into $(-v, -\tilde{\varphi})$, it is clear that the curves $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ defined by $(u, \theta, \tilde{\varphi}) = \text{const}$ constitute a second congruence of principal null geodesics, called the **outgoing principal null geodesics**. A priori, this congruence is defined only in \mathcal{M}_{BL} , i.e. where $\Delta \neq 0$, but we shall see below that we can extend it to all \mathcal{M} . As $-r$ was an affine parameter along the ingoing principal null geodesics, r is an affine parameter along the outgoing principal null geodesics in \mathcal{M}_{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 \mathcal{M}_I and \mathcal{M}_{III} ; in region \mathcal{M}_{II} , r is decreasing towards the future, in agreement with the general Property 10.13 (Sec. 10.3.2).

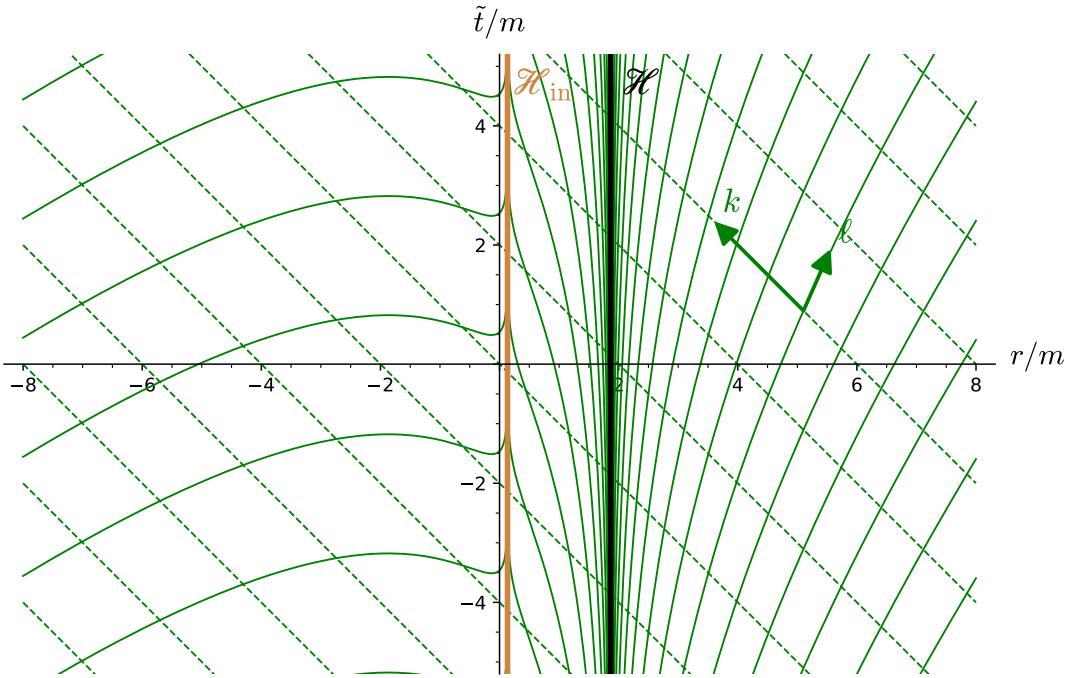


Figure 10.7: Same as Fig. 10.6, but for $a/m = 0.5$.

The fact that the Weyl tensor C admits two, and exactly two, congruences of principal null geodesics means that the Kerr metric is an *algebraically special* solution of the Einstein equation: it belongs to the so-called *Petrov type D* [390].

Let us find the expression of the outgoing principal null geodesics in terms of Kerr coordinates (which have been constructed on the *ingoing* principal null congruence). Combining (10.51) with (10.34), we get

$$du = d\tilde{t} - \frac{r^2 + 2mr + a^2}{\Delta} dr \quad (10.52a)$$

$$d\tilde{\varphi} = d\tilde{\varphi} - \frac{2a}{\Delta} dr. \quad (10.52b)$$

These equations can be integrated (cf. the computation leading to Eq. (10.24)), yielding

$$u = \tilde{t} - r - \frac{2m}{\sqrt{m^2 - a^2}} \left(r_+ \ln \left| \frac{r - r_+}{2m} \right| - r_- \ln \left| \frac{r - r_-}{2m} \right| \right) \quad (10.53a)$$

$$\tilde{\varphi} = \tilde{\varphi} - \frac{a}{\sqrt{m^2 - a^2}} \ln \left| \frac{r - r_+}{r - r_-} \right|, \quad (10.53b)$$

The quantities $u - \tilde{t}$ and $\tilde{\varphi} - \tilde{\varphi}$ 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 \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{III} . Accordingly, the 3-tuple $(u, \theta, \tilde{\varphi})$ defines uniquely an outgoing principal null geodesic only in each of these three regions. In other words, the congruence of outgoing principal null geodesics can be split in three families $\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}, I}$, $\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}, II}$ and $\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}, III}$, each family being labelled by $(u, \theta, \tilde{\varphi})$. In what follows, we shall denote by $\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}}$ any member of one of these three families.

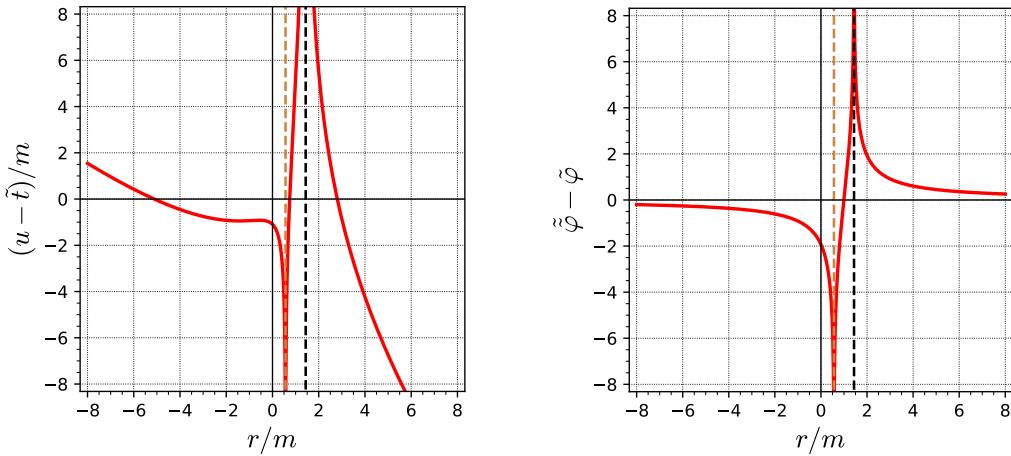


Figure 10.8: $u - \tilde{t}$ (left panel) and $\tilde{\varphi} - \varphi$ (right panel) as functions of r given by Eqs. (10.53) for $a/m = 0.9$. The dashed vertical lines correspond to $r = r_-$ (\mathcal{H}_{in}) and $r = r_+$ (\mathcal{H}) and delimitate the regions \mathcal{M}_{III} , \mathcal{M}_{II} and \mathcal{M}_1 (from the left to the right). [Figure generated by the notebook D.5.3]

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 $\tilde{\varphi}$ held fixed, is not shown in these figures, which represent only traces in the (\tilde{t}, r) plane. Actually the geodesics $\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}}$ are winding with respect the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ at the coordinate speed

$$\frac{d\tilde{\varphi}}{d\tilde{t}} \Big|_{\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}}} = \frac{2a}{r^2 + 2mr + a^2}. \quad (10.54)$$

Proof. Along a null geodesic $\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}}$, we have $du = 0$ and $d\tilde{\varphi} = 0$, so that (10.52) yields

$$\frac{d\tilde{t}}{dr} \Big|_{\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}}} = \frac{r^2 + 2mr + a^2}{\Delta} \quad \text{and} \quad \frac{d\tilde{\varphi}}{dr} \Big|_{\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}}} = \frac{2a}{\Delta}. \quad (10.55)$$

Dividing the second expression by the first one yields (10.54). \square

Note that in both asymptotically flat regions, when $r \rightarrow \pm\infty$, the winding speed (10.54) goes to zero and the two congruences of geodesics tend to $\pm 45^\circ$ 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 \mathcal{M}_{II} (between \mathcal{H}_{in} and \mathcal{H}), i.e. have r decreasing towards the future, in agreement with Property 10.13 (Sec. 10.3.2).

Remark 4: In Fig. 10.7, the outgoing geodesics seem to go “backward in time” in the region $-2m \lesssim r \lesssim 0$. This is an artefact due to the hypersurfaces $\tilde{t} = \text{const}$ being not spacelike there, as discussed in Remark 4 p. 337. 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 $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ when $\Delta = 0$, i.e. on \mathcal{H} and \mathcal{H}_{in} . Actually, the family $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ is not even defined there, since we read on (10.53) and see on Fig. 10.8 that $u \rightarrow \pm\infty$ and $\tilde{\varphi} \rightarrow \pm\infty$ when $r \rightarrow r_{\pm}$ (see also Fig. 6.5 for a pictorial view of $u \rightarrow +\infty$ when $r \rightarrow r_+$ in the Schwarzschild limit). In order to extend the outgoing principal null family to \mathcal{H} and \mathcal{H}_{in} , let us introduce along any geodesic $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ in \mathcal{M}_{BL} the parameter λ that is related to the affine parameter r by

$$\frac{dr}{d\lambda} = \frac{\Delta}{2(r^2 + a^2)}. \quad (10.56)$$

The tangent vector to $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ associated to λ , ℓ say, has the following components w.r.t. Kerr coordinates, obtained by combining Eqs. (10.55) and (10.56):

$$\begin{aligned}\ell^{\tilde{t}} &= \frac{d\tilde{t}}{d\lambda} = \frac{d\tilde{t}}{dr} \times \frac{dr}{d\lambda} = \frac{r^2 + 2mr + a^2}{2(r^2 + a^2)} = \frac{1}{2} + \frac{mr}{r^2 + a^2} \\ \ell^r &= \frac{dr}{d\lambda} = \frac{\Delta}{2(r^2 + a^2)} = \frac{1}{2} - \frac{mr}{r^2 + a^2} \\ \ell^\theta &= \frac{d\theta}{d\lambda} = 0 \\ \ell^{\tilde{\varphi}} &= \frac{d\tilde{\varphi}}{d\lambda} = \frac{d\tilde{\varphi}}{dr} \times \frac{dr}{d\lambda} = \frac{a}{r^2 + a^2}.\end{aligned}$$

In other words, we have

$$\boxed{\ell = \left(\frac{1}{2} + \frac{mr}{r^2 + a^2} \right) \partial_{\tilde{t}} + \left(\frac{1}{2} - \frac{mr}{r^2 + a^2} \right) \partial_r + \frac{a}{r^2 + a^2} \partial_{\tilde{\varphi}}}. \quad (10.57)$$

It is clear that this vector field is regular everywhere in \mathcal{M} . Given the metric components (10.36), it is easy to check that $g(\ell, \ell) = 0$, i.e. that ℓ is a null vector. Moreover, an explicit computation (cf. the notebook D.5.2-2) reveals that

$$\nabla_\ell \ell = \kappa_\ell \ell, \quad \text{with} \quad \kappa_\ell := \frac{m(r^2 - a^2)}{(r^2 + a^2)^2}. \quad (10.58)$$

Equation (10.58) shows that the integral curves of the vector field ℓ are geodesics (cf. Secs. 2.3.3 and B.2.2). In \mathcal{M}_{BL} , they are nothing but the null geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$. On \mathcal{H} or \mathcal{H}_{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

$$\psi := \tilde{\varphi} - \frac{a}{2mr_{\pm}} \tilde{t}. \quad (10.59)$$

Indeed, in view of the components (10.57), we have along these geodesics

$$\frac{d\tilde{\varphi}}{d\tilde{t}} = \frac{d\tilde{\varphi}}{d\lambda} \frac{d\lambda}{d\tilde{t}} = \frac{\ell^{\tilde{\varphi}}}{\ell^{\tilde{t}}} = \frac{2a}{r_{\pm}^2 + 2mr_{\pm} + a^2} = \frac{a}{2mr_{\pm}}.$$

Since $a/(2mr_{\pm})$ is constant, we get $\tilde{\varphi} = a/(2mr_{\pm}) \tilde{t} + \psi$, where ψ is some integration constant.

Property 10.14: outgoing principal null geodesics

Given that ℓ is a smooth vector field, the congruence of **outgoing principal null geodesics** can be smoothly extended to include the integral curves of ℓ on \mathcal{H} and \mathcal{H}_{in} , which we shall denote by $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ and $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}_{\text{in}}}$, given that θ and ψ [Eq. (10.59)] are constant along them. Consequently, the congruence of outgoing principal null geodesics on \mathcal{M} is made of 5 families: $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},\text{I}}$, $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},\text{II}}$, $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},\text{III}}$, $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ and $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}_{\text{in}}}$, each family covering a different subset of \mathcal{M} .

Remark 5: At the Schwarzschild limit, $a = 0$, $\tilde{\varphi} = \varphi$, $\psi = \varphi$ and the outgoing principal null geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ and $\mathcal{L}_{(\theta,\psi)}^{\text{out}}$ reduce respectively to the outgoing radial null geodesics $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ and $\mathcal{L}_{(\theta,\varphi)}^{\text{out}}$ discussed in Secs. 6.3.1 and 6.3.5. This can be seen from the fact that, for $a = 0$, the tangent vector ℓ given by Eq. (10.57) reduces to the tangent vector ℓ given by Eq. (6.53).

The 1-form $\underline{\ell}$ associated to ℓ by 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):

$$\underline{\ell} = -\frac{\Delta}{2(r^2 + a^2)} dt + \frac{2\rho^2 - \Delta}{2(r^2 + a^2)} dr + \frac{a\Delta \sin^2 \theta}{2(r^2 + a^2)} d\tilde{\varphi}. \quad (10.60)$$

The components of ℓ and $\underline{\ell}$ with respect to the Boyer-Lindquist coordinate frame/coframe are obtained by combining Eq. (10.57) with Eq. (10.38) and Eq. (10.60) with Eq. (10.34):

$$\ell = \frac{1}{2}\partial_t + \frac{\Delta}{2(r^2 + a^2)} \partial_r + \frac{a}{2(r^2 + a^2)} \partial_\varphi. \quad (10.61)$$

$$\underline{\ell} = -\frac{\Delta}{2(r^2 + a^2)} dt + \frac{\rho^2}{2(r^2 + a^2)} dr + \frac{a\Delta \sin^2 \theta}{2(r^2 + a^2)} d\varphi. \quad (10.62)$$

Remark 6: The reader may have noticed a certain dissymmetry between the chosen tangent vector k of ingoing principal null geodesics, which obeys $\nabla_k k = 0$ [Eq. (10.45)] and the tangent vector ℓ of the outgoing ones, which obeys $\nabla_\ell \ell \neq 0$ [Eq. (10.58)]. The last property implies that the parameter λ associated to ℓ is not affine, while the parameter $-r$ associated to k is (cf. Sec. B.2.2). The non-affine choice is the price to pay to have a parametrization of the outgoing congruence well-defined everywhere in \mathcal{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 (10.61), one can easily evaluate the scalar product of k and ℓ :

$$g(k, \ell) = -\frac{\rho^2}{r^2 + a^2}. \quad (10.63)$$

This implies that $g(k, \ell) < 0$ everywhere (note that $\rho^2 = 0$ would correspond to the ring singularity \mathcal{R} , which has been excluded from the spacetime manifold \mathcal{M}). Using Lemma 1.2 (Sec. 1.2.2) with $u = k$ and $v = \ell$, we conclude

Property 10.15: ℓ future-directed

The tangent vector ℓ to the outgoing principal null geodesics is future-directed in all the Kerr spacetime \mathcal{M} .

Remark 7: Instead of ℓ , another natural choice for the tangent vector to the outgoing principal null geodesics would have been the tangent vector $\hat{\ell}$ such that $\mathbf{g}(\mathbf{k}, \hat{\ell}) = -1$. In view of (10.63), the two tangent vectors are related by $\hat{\ell} = (r^2 + a^2)/\rho^2 \ell$. As for ℓ , the tangent vector $\hat{\ell}$ does not correspond to an affine parameter of the outgoing principal null geodesics, i.e. its non-affinity coefficient $\kappa_{\hat{\ell}}$ is nonzero. While $\hat{\ell}$ is used in many studies, we prefer ℓ here because it coincides with the Killing vector $\xi + \Omega_{\mathcal{H}}\eta$ on the event horizon, as we shall see in Sec. 10.5 [Eq. (10.72)]. In particular, its non-affinity coefficient κ_{ℓ} yields the so-called *black hole's surface gravity* (Sec. 10.5.4 below).

10.5 Event horizon

10.5.1 The two Killing horizons of Kerr spacetime

Let us consider the hypersurfaces of \mathcal{M} defined by a fixed value of the coordinate r . \mathcal{H} and \mathcal{H}_{in} are two particular cases, corresponding to $r = r_+$ and $r = r_-$ respectively. The normal 1-form to a hypersurface $r = \text{const}$ is dr ; the corresponding gradient vector field is $\vec{\nabla}r$, the components of which with respect to Kerr coordinates are $\nabla^{\tilde{\alpha}}r = g^{\tilde{\alpha}\tilde{\mu}}\partial_{\tilde{\mu}}r = g^{\tilde{\alpha}r}$, with $g^{\tilde{\alpha}r}$ read on the second raw of the matrix (10.41). Hence

$$\vec{\nabla}r = \frac{2mr}{\rho^2}\partial_{\tilde{t}} + \frac{\Delta}{\rho^2}\partial_{\tilde{r}} + \frac{a}{\rho^2}\partial_{\tilde{\varphi}}.$$

Instead of $\vec{\nabla}r$, it is quite natural to consider the rescaled vector field $\mathbf{n} := \rho^2\vec{\nabla}r$ as the normal to the hypersurfaces $r = \text{const}$, for it has simpler components in the Kerr frame:

$$\mathbf{n} = 2mr\partial_{\tilde{t}} + \Delta\partial_{\tilde{r}} + a\partial_{\tilde{\varphi}}. \quad (10.64)$$

The scalar square of \mathbf{n} is $\mathbf{n} \cdot \mathbf{n} = n_{\mu}n^{\mu} = \rho^2(\nabla_{\mu}r)n^{\mu} = \rho^2n^r$. Hence, in view of (10.64),

$$\mathbf{n} \cdot \mathbf{n} = \rho^2\Delta. \quad (10.65)$$

Since $\rho^2 > 0$ everywhere on \mathcal{M} and $\Delta = (r - r_+)(r - r_-)$ [Eq. (10.10)], we conclude:

Property 10.16: causal type of the hypersurfaces of constant r

- The hypersurfaces $r = \text{const}$ are timelike in regions \mathcal{M}_I and \mathcal{M}_{III} ;
- The hypersurfaces $r = \text{const}$ are spacelike in region \mathcal{M}_{II} ;
- \mathcal{H} (where $r = r_+$) and \mathcal{H}_{in} (where $r = r_-$) are null hypersurfaces.

On \mathcal{H} and \mathcal{H}_{in} , $\Delta = 0$, so that Eq. (10.64) yields

$$\mathbf{n} = 2mr_{\pm}\partial_{\tilde{t}} + a\partial_{\tilde{\varphi}} = 2mr_{\pm}\boldsymbol{\xi} + a\boldsymbol{\eta}, \quad (10.66)$$

where we have used (10.39) and r_{\pm} stands for r_+ on \mathcal{H} and r_- on \mathcal{H}_{in} . On \mathcal{H} , we may rewrite this expression as

$$\mathbf{n} = 2mr_+\boldsymbol{\chi}, \quad (10.67)$$

with

$$\boxed{\boldsymbol{\chi} := \boldsymbol{\xi} + \Omega_{\mathcal{H}}\boldsymbol{\eta}} \quad (10.68)$$

and

$$\boxed{\Omega_{\mathcal{H}} := \frac{a}{2mr_+} = \frac{a}{r_+^2 + a^2} = \frac{a}{2m(m + \sqrt{m^2 - a^2})}.} \quad (10.69)$$

$\Omega_{\mathcal{H}}$ being a constant, the vector field $\boldsymbol{\chi}$ defined by (10.68) is a Killing vector field of (\mathcal{M}, g) . Moreover, Eq. (10.67) shows that this Killing vector is normal to the null hypersurface \mathcal{H} . In view of the definition given in Sec. 3.3.2, we conclude:

Property 10.17: \mathcal{H} as a Killing horizon

The hypersurface \mathcal{H} defined by $r = r_+$ is a Killing horizon with respect to the Killing vector field $\boldsymbol{\chi}$ [Eq. (10.68)].

Similarly, on \mathcal{H}_{in} , we may rewrite (10.66) as $\mathbf{n} = 2mr_-\boldsymbol{\chi}_{\text{in}}$, with

$$\boldsymbol{\chi}_{\text{in}} := \boldsymbol{\xi} + \Omega_{\text{in}}\boldsymbol{\eta} \quad (10.70)$$

and

$$\Omega_{\text{in}} := \frac{a}{2mr_-} = \frac{a}{r_-^2 + a^2} = \frac{a}{2m(m - \sqrt{m^2 - a^2})}, \quad (10.71)$$

thereby arriving at:

Property 10.18: \mathcal{H}_{in} as a Killing horizon

The hypersurface \mathcal{H}_{in} defined by $r = r_-$ is a Killing horizon with respect to the Killing vector field $\boldsymbol{\chi}_{\text{in}}$ [Eq. (10.70)].

We shall call \mathcal{H}_{in} the **inner horizon**. We shall see in Sec. 10.8.3 that \mathcal{H}_{in} is actually (part of) a so-called *Cauchy horizon*.

Historical note : The identification of the hypersurfaces \mathcal{H} and \mathcal{H}_{in} as the only two null hypersurfaces of Kerr spacetime that are stationary (i.e. that are Killing horizons in the modern language) has been first performed in 1964 by Robert H. Boyer and T.G. Price [72], who claimed: “These are ‘horizons’ in the sense that there can be a flow of matter or radiation across them in only one direction. They are the analogues of the Schwarzschild null sphere or ‘singularity’.”

10.5.2 Black hole character

As a null hypersurface, \mathcal{H} is a one-way membrane (cf. Sec. 2.2.2), therefore any (massive or null) particle that crossed it from \mathcal{M}_I to \mathcal{M}_{II} can never be back in \mathcal{M}_I . Let us show that \mathcal{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 \mathcal{M}_I is the same as that of Schwarzschild spacetime. Hence one can perform a conformal completion of (\mathcal{M}_I, g) endowed with a future null infinity \mathcal{I}^+ and a past null infinity \mathcal{I}^- (an explicit construction of \mathcal{I}^+ and \mathcal{I}^- for Schwarzschild spacetime has been performed in Sec. 9.4.3). A conformal diagram representing \mathcal{M}_I along with \mathcal{I}^+ and \mathcal{I}^- is given in Fig. 10.9 below.

Let us show that the causal past of the future null infinity coincide with \mathcal{M}_I : $J^-(\mathcal{I}^+) = \mathcal{M}_I$. Since, as stressed above, no future-directed causal curve can move from \mathcal{M}_{II} to \mathcal{M}_I and \mathcal{I}^+ is a boundary of \mathcal{M}_I , we have $\mathcal{M}_{II} \cap J^-(\mathcal{I}^+) = \emptyset$. A fortiori $\mathcal{M}_{III} \cap J^-(\mathcal{I}^+) = \emptyset$. We have thus $J^-(\mathcal{I}^+) \subset \mathcal{M}_I$. To show the equality between the two sets there remains to show that any point $p \in \mathcal{M}_I$ can emit a signal reaching \mathcal{I}^+ . Let \mathcal{L} be the null geodesic through p of the outgoing principal null congruence $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ introduced in Sec. 10.4, i.e. \mathcal{L} is the unique geodesic departing from p with the tangent vector ℓ given by (10.57). Along \mathcal{L} , one has

$$\frac{dr}{d\lambda} = \ell^r = \frac{1}{2} - \frac{mr}{r^2 + a^2},$$

where λ is the parameter associated with ℓ . In particular, at p , if we denote by r_0 the r -coordinate of p in the Kerr system,

$$\left. \frac{dr}{d\lambda} \right|_p = \frac{1}{2} - \frac{mr_0}{r_0^2 + a^2} > 0.$$

The above inequality simply translates the fact that $r_0 > r_+$ wherever p lies in \mathcal{M}_I . Hence, initially r is increasing along \mathcal{L} and we get, since $-mr/(r^2 + a^2)$ is an increasing function of r ,

$$\frac{dr}{d\lambda} \geq \frac{1}{2} - \frac{mr_0}{r_0^2 + a^2} =: \alpha > 0.$$

Since α is a constant, we deduce that

$$r \geq r_0 + \alpha(\lambda - \lambda_0),$$

where λ_0 is the value of \mathcal{L} 's parameter at p . When $\lambda \rightarrow +\infty$, we get $r \rightarrow +\infty$, which proves that the null curve \mathcal{L} reaches \mathcal{I}^+ . Hence we conclude:

Property 10.19: black hole region in Kerr spacetime

$\mathcal{B} = \mathcal{M} \setminus \mathcal{M}_I$ is the black hole region, the event horizon of which is \mathcal{H} .

Incidentally, since we have already shown that \mathcal{H} is a Killing horizon (Property 10.17), this illustrates the strong rigidity theorem (Property 5.25 in Sec. 5.4.2): the black hole event horizon of Kerr spacetime is a Killing horizon.

According to the discussion in Sec. 5.4.2, we may then call the quantity $\Omega_{\mathcal{H}}$ introduced in Eqs. (10.68)-(10.69) the **black hole rotation velocity**.

10.5.3 Null generators of the event horizon

The null vector field ℓ defined by Eq. (10.57) coincides with the Killing vector χ on \mathcal{H} , since $mr/(r^2 + a^2) \stackrel{\mathcal{H}}{=} 1/2$ and $a/(r^2 + a^2) \stackrel{\mathcal{H}}{=} \Omega_{\mathcal{H}}$:

$$\ell \stackrel{\mathcal{H}}{=} \chi. \quad (10.72)$$

Since (i) χ is tangent to the null geodesic generators of \mathcal{H} , being a null normal to it (cf. Sec. 2.3.3) and (ii) ℓ is the tangent vector to the outgoing principal null geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ (Sec. 10.4), we get

Property 10.20: null generators of the event horizon

The null generators of the event horizon \mathcal{H} are the geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ of the outgoing principal null congruence.

Similarly, we have

Property 10.21: null generators of the inner horizon

The null generators of the inner horizon \mathcal{H}_{in} are the geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}_{\text{in}}}$ of the outgoing principal null congruence.

The reader is referred to Fig. 5.1 for a pictorial view of the event horizon \mathcal{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 \mathcal{H} (resp. \mathcal{H}_{in}).

Remark 1: Since $\Delta = 0$ on \mathcal{H} , we read immediately from Eq. (10.60) that

$$\underline{\ell} \stackrel{\mathcal{H}}{=} \frac{\rho^2}{r^2 + a^2} dr,$$

which shows that, at any point of \mathcal{H} , the vector ℓ is normal to the hypersurface $r = \text{const}$ through this point. This hypersurface being nothing but \mathcal{H} itself, we recover the fact that ℓ is normal to \mathcal{H} .

In view of expression (10.69) for $\Omega_{\mathcal{H}}$, Eq. (10.59) for the parameter ψ labelling the outgoing principal null geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ can be rewritten as

$$\psi = \tilde{\varphi} - \Omega_{\mathcal{H}} \tilde{t}. \quad (10.73)$$

Hence ψ appears as a corotating azimuthal coordinate on \mathcal{H} . Moreover, we verify that the winding speed of the outgoing principal null geodesics given by Eq. (10.54) tends toward $\Omega_{\mathcal{H}}$ when approaching the event horizon:

$$\lim_{r \rightarrow r_+} \left. \frac{d\tilde{\varphi}}{d\tilde{t}} \right|_{\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}} = \Omega_{\mathcal{H}}. \quad (10.74)$$

This follows immediately from $\ell \stackrel{\mathcal{H}}{=} \chi = \xi + \Omega_{\mathcal{H}} \eta = \partial_{\tilde{t}} + \Omega_{\mathcal{H}} \partial_{\tilde{\varphi}}$ and the identity $r_+^2 + a^2 = 2mr_+$ (compare Eq. (10.54) with $r \rightarrow r_+$ and Eq. (10.69)).

10.5.4 Surface gravity

In view of the pre-geodesic equation (10.58) satisfied by ℓ and the identity (10.72), we deduce that

$$\nabla_\chi \chi^{\mathcal{H}} = \kappa \chi, \quad (10.75)$$

with the non-affinity coefficient given by Eq. (10.58): $\kappa = \kappa_\ell|_{r=r_+} = m(r_+^2 - a^2)/(r_+^2 + a^2)^2$. Since r_+ is a zero of Δ , we have $r_+^2 + a^2 = 2mr_+$, so that κ can be rewritten as

$$\kappa = \frac{r_+^2 - a^2}{2r_+(r_+^2 + a^2)} = \frac{r_+^2 - a^2}{4mr_+^2}. \quad (10.76)$$

Substituting (10.3) for r_+ , we get an expression involving the two basic Kerr parameters:

$$\kappa = \frac{\sqrt{m^2 - a^2}}{2m(m + \sqrt{m^2 - a^2})}. \quad (10.77)$$

Given the strict inequality $a < m$ assumed in this chapter [Eq. (10.1)], we have $\kappa \neq 0$. According to the classification introduced in Sec. 3.3.6, we may state:

Property 10.22: non-degeneracy of the event horizon

As long as $a < m$, the event horizon \mathcal{H} is a non-degenerate Killing horizon.

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 dynamics* established in Sec. 3.3.5 (cf. in particular Example 14 in that section).

Remark 2: As a check, if we let $a \rightarrow 0$ in Eq. (10.77), we get $\kappa = 1/(4m)$, i.e. we recover the Schwarzschild horizon value computed in Example 10 of Chap. 2 [cf. Eq. (2.29)].

10.6 Global quantities

10.6.1 Mass

We have seen in Sec. 10.2.2 that when $r \rightarrow +\infty$, 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 a stationary Killing vector ξ , as the Kerr spacetime, there is a well-defined concept of mass, namely the *Komar mass* introduced in Sec. 5.3.3. Let us show explicitly that for the Kerr spacetime, the Komar mass coincides with m . To this aim, we shall use expression (5.46) with $n = 4$ for the Komar mass. We shall consider the Boyer-Lindquist coordinates $(x^\alpha) = (t, r, \theta, \varphi)$ and define \mathcal{S} to be a sphere $(t, r) = \text{const.}$ Coordinates on \mathcal{S} are then $(x^2, x^3) = (\theta, \varphi)$. Moreover, since the value of $M_{\mathcal{S}}$ does not depend on the choice of \mathcal{S} (cf. Property 5.15), we may set $r \rightarrow +\infty$

and use the asymptotic flatness of Kerr metric to get simple expressions. The unit normals to \mathcal{S} are then $\mathbf{n} = \partial_t$ and $\mathbf{s} = \partial_r$. Moreover, when $r \rightarrow +\infty$, $\sqrt{q} = r^2 \sin \theta$. Hence Eq. (5.46) with $n = 4$ yields

$$M = -\frac{1}{8\pi} \lim_{r \rightarrow +\infty} \int_{\mathcal{S}} (\partial_r \xi_t - \underbrace{\partial_t \xi_r}_0) r^2 \sin \theta d\theta d\varphi,$$

with

$$\xi_t = g_{t\mu} \xi^\mu = g_{t\mu} \delta^\mu_t = g_{tt} \simeq -1 + \frac{2m}{r},$$

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

$$\boxed{M = m}. \quad (10.78)$$

10.6.2 Angular momentum

Since the Kerr spacetime is axisymmetric, a well-defined notion of angular momentum is provided by the *Komar angular momentum* J introduced in Sec. 5.3.6. Let us compute J via Eq. (5.66) by means of Boyer-Lindquist coordinates, choosing for \mathcal{S} a 2-sphere $(t, r) = \text{const.}$ In evaluating the terms $s^\mu \partial_\mu \eta_\nu n^\nu$ and $n^\mu \partial_\mu \eta_\nu s^\nu$ as $r \rightarrow +\infty$, we have to be a little more cautious than in Sec. 10.6.1, since one of the components η_α is diverging when $r \rightarrow +\infty$:

$$\eta_\alpha = g_{\alpha\mu} \eta^\mu = g_{\alpha\mu} \delta^\mu_\varphi = g_{\alpha\varphi} = (g_{t\varphi}, 0, 0, g_{\varphi\varphi})$$

with, according to (10.11), $g_{\varphi\varphi} \sim r^2 \sin^2 \theta$ as $r \rightarrow +\infty$. Moreover, given the value of $g_{t\varphi}$ read on (10.8), we may write

$$\eta_\alpha \sim \left(-\frac{2am \sin^2 \theta}{r}, 0, 0, r^2 \sin^2 \theta \right) \quad \text{when } r \rightarrow +\infty.$$

Let us choose for the timelike normal \mathbf{n} to \mathcal{S} the future-directed unit normal to the hypersurfaces $t = \text{const.}$

$$\underline{\mathbf{n}} = -N dt, \quad (10.79)$$

where N is a normalization factor ensuring $\mathbf{n} \cdot \mathbf{n} = -1$. We do not need the precise value² of N , but simply the property $N \rightarrow 1$ as $r \rightarrow +\infty$. We have then $n_\alpha = (-N, 0, 0, 0)$, so that

$$n^\alpha = g^{\alpha\mu} n_\mu = g^{\alpha t} (-N) = (-Ng^{tt}, 0, 0, -Ng^{t\varphi}),$$

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 $r \rightarrow +\infty$; hence

$$n^\alpha \sim \left(1, 0, 0, \frac{2am}{r^3} \right) \quad \text{when } r \rightarrow +\infty.$$

The choice of \mathbf{n} completely determines that of \mathbf{s} :

$$s^\alpha = \left(0, \frac{\sqrt{\Delta}}{\rho}, 0, 0 \right) \sim (0, 1, 0, 0) \quad \text{when } r \rightarrow +\infty.$$

²It is given by Eq. (10.93) below.

Indeed, given the metric components (10.8), we immediately check that $\mathbf{n} \cdot \mathbf{s} = 0$, $\mathbf{s} \cdot \mathbf{s} = 1$ and $\underline{\mathbf{s}} = (\rho/\sqrt{\Delta}) \mathbf{d}r$, which does imply that \mathbf{s} is normal to \mathcal{S} .

Given the above expressions for η_α , n^α and s^α , we get, for $r \rightarrow +\infty$,

$$s^\mu \partial_\mu \eta_\nu n^\nu \sim \partial_r \left(-\frac{2am \sin^2 \theta}{r} \right) \times 1 + \partial_r (r^2 \sin^2 \theta) \times \frac{2am}{r^3} \sim \frac{6am \sin^2 \theta}{r^2}$$

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. (5.66) leads to

$$J = \frac{1}{16\pi} \lim_{r \rightarrow +\infty} \int_{\mathcal{S}} \frac{6am \sin^2 \theta}{r^2} \times r^2 \sin \theta \mathbf{d}\theta \mathbf{d}\varphi = \frac{3am}{8\pi} \int_{\mathcal{S}} \sin^3 \theta \mathbf{d}\theta \mathbf{d}\varphi = \frac{3am}{4} \underbrace{\int_0^\pi \sin^3 \theta \mathbf{d}\theta}_{4/3},$$

i.e.

$$\boxed{J = am}. \quad (10.80)$$

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 \mathcal{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 Boyer-Lindquist coordinates as for M and J , because they are singular on \mathcal{H} ; we shall use rather the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$, which are regular on \mathcal{H} . \mathcal{H} is defined by $r = r_+$ and it is natural to consider a cross-section of it, \mathcal{S} say, defined by $\{\tilde{t} = \text{const}, r = r_+\}$. Then \mathcal{S} is spanned by the coordinates $(x^a) = (\theta, \tilde{\varphi})$ and the metric \mathbf{q} induced on it by the spacetime metric is obtained by setting $r = r_+$, $\mathbf{d}\tilde{t} = 0$ and $\mathbf{d}r = 0$ in (10.36):

$$\mathbf{q} = (r_+^2 + a^2 \cos^2 \theta) \mathbf{d}\theta^2 + \left(r_+^2 + a^2 + \frac{2a^2 m r_+ \sin^2 \theta}{r_+^2 + a^2 \cos^2 \theta} \right) \sin^2 \theta \mathbf{d}\tilde{\varphi}^2.$$

Now, since r_+ is a zero of Δ [cf. Eq. (10.10)], we have $2mr_+ = r_+^2 + a^2$. This brings us to

$$\mathbf{q} = (r_+^2 + a^2 \cos^2 \theta) \mathbf{d}\theta^2 + \frac{(r_+^2 + a^2)^2}{r_+^2 + a^2 \cos^2 \theta} \sin^2 \theta \mathbf{d}\tilde{\varphi}^2. \quad (10.81)$$

The area of the cross-section \mathcal{S} is

$$A = \int_{\mathcal{S}} \sqrt{q} \mathbf{d}\theta \mathbf{d}\tilde{\varphi} = (r_+^2 + a^2) \underbrace{\int_{\mathcal{S}} \sin \theta \mathbf{d}\theta \mathbf{d}\tilde{\varphi}}_{4\pi}$$

where we have used (10.81) to write $q = \det(q_{ab}) = (r_+^2 + a^2)^2 \sin^2 \theta$. We have thus

$$A = 4\pi(r_+^2 + a^2) = 8\pi m r_+. \quad (10.82)$$

Via (10.3), one may recast this result to let appear only m and a :

$$A = 8\pi m(m + \sqrt{m^2 - a^2}). \quad (10.83)$$

10.6.4 Smarr formula

By combining the relations $\Omega_{\mathcal{H}} = a/(2mr_+)$ [Eq. (10.69)], $\kappa = (r_+^2 - a^2)/(4mr_+^2)$ [Eq. (10.76)] and $A = 8\pi m r_+$ [Eq. (10.82)], we get an interesting identity:

$$\frac{\kappa A}{4\pi} + 2\Omega_{\mathcal{H}} am = m.$$

If we let appear the total angular momentum $J = am$ [Eq. (10.80)] and the Komar mass $M = m$ [Eq. (10.78)], we can turn this identity into

$$M = \frac{\kappa A}{4\pi} + 2\Omega_{\mathcal{H}} J. \quad (10.84)$$

This is actually the particular case $T = 0$ (vacuum) of the generic *Smarr formula* (5.91) established in Chap. 5, or the particular case $Q_{\mathcal{H}} = 0$ (vanishing electric charge) of the electrovacuum Smarr formula (5.104).

10.7 Families of 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 \mathcal{O} in Kerr spacetime whose 4-velocity \mathbf{u} is a linear combination of the Killing vectors ξ and η with constant coefficients³:

$$\mathbf{u} = \alpha \xi + \beta \eta, \quad \alpha = \text{const}, \beta = \text{const}. \quad (10.85)$$

It follows that the worldline \mathcal{L} of \mathcal{O} is an orbit of the isometry group $\mathbb{R} \times \text{SO}(2)$ of Kerr spacetime. In physical terms, this means that the spacetime geometry as perceived by observer \mathcal{O} does not evolve, hence the name *stationary observer*.

³In the definition (10.85), 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.

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 $\xi = \partial_t$ and $\eta = \partial_\varphi$ and we deduce from (10.85) and the definition (1.16) of the 4-velocity that along the worldline \mathcal{L} :

$$\frac{dt}{d\tau} = \alpha, \quad \frac{dr}{d\tau} = 0, \quad \frac{d\theta}{d\tau} = 0, \quad \frac{d\varphi}{d\tau} = \beta,$$

where τ is the observer's proper time. It follows that the stationary observer evolves at fixed values of the coordinates r and θ .

Outside the Carter time machine \mathcal{T} (cf. Sec. 10.2.5), we have necessarily $\alpha \neq 0$, because η is spacelike in $\mathcal{M} \setminus \mathcal{T}$ and the 4-velocity u is necessarily timelike. Hence, by introducing $\omega := \beta/\alpha$, we may rewrite (10.85) as

$$\boxed{u = \alpha(\xi + \omega\eta)}, \quad \alpha = \text{const}, \quad \omega = \text{const}. \quad (10.86)$$

The coefficient ω gives the rate of variation of the azimuthal coordinate along the observer's worldline \mathcal{L} in any coordinate system adapted to the spacetime symmetries, like the Boyer-Lindquist ones (t, r, θ, φ) , the advanced Kerr ones $(v, r, \theta, \tilde{\varphi})$ or the Kerr ones $(\tilde{t}, r, \theta, \tilde{\varphi})$, according to

$$\boxed{\omega = \left. \frac{d\varphi}{dt} \right|_{\mathcal{L}} = \left. \frac{d\tilde{\varphi}}{dv} \right|_{\mathcal{L}} = \left. \frac{d\tilde{\varphi}}{d\tilde{t}} \right|_{\mathcal{L}}}. \quad (10.87)$$

Proof. Let us consider Boyer-Lindquist coordinates (t, r, θ, φ) . Denoting by τ the proper time of \mathcal{O} , we have $d\varphi/dt|_{\mathcal{L}} = d\varphi/d\tau \times d\tau/dt$. Now by the definition (1.16) of the 4-velocity, $d\varphi/d\tau = u^\varphi$ and $d\tau/dt = (u^t)^{-1}$. We have thus $d\varphi/dt|_{\mathcal{L}} = u^\varphi/u^t$. From Eq. (10.86) along with $\xi = \partial_t$ and $\eta = \partial_\varphi$, we read $u^t = \alpha$ and $u^\varphi = \alpha\omega$. Hence $d\varphi/dt|_{\mathcal{L}} = \omega$. The same demonstration applies to Kerr and advanced Kerr coordinates because $\xi = \partial_v = \partial_{\tilde{t}}$ and $\eta = \partial_{\tilde{\varphi}}$. \square

Remark 1: That the ratios $d\varphi/dt$, $d\tilde{\varphi}/dv$ and $d\tilde{\varphi}/d\tilde{t}$ along \mathcal{L} are all equal, as expressed in (10.87), is not surprising if one considers the links between the various coordinates expressed by Eqs. (10.23a), (10.23b) and (10.34). Setting $dr = 0$ in these equations, since r is constant along \mathcal{L} , we get

$$dt|_{\mathcal{L}} = dv|_{\mathcal{L}} = d\tilde{t}|_{\mathcal{L}} \quad \text{and} \quad d\varphi|_{\mathcal{L}} = d\tilde{\varphi}|_{\mathcal{L}}.$$

Equation (10.87) allows for a nice physical interpretation of ω . Indeed, we have seen in Sec. 7.3.3 [cf. Eq. (7.58)] that $d\varphi/dt|_{\mathcal{L}}$ 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 $d\varphi/dt$ was constant along \mathcal{L} , so it applies to the present case as well.

10.7.2 Static observers

A *static observer* is a stationary observer \mathcal{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 \mathcal{O} 's worldline, since $\omega = 0$ in Eq. (10.87) implies $\varphi = \text{const}$ and $\tilde{\varphi} = \text{const}$. Moreover, such an observer appears not moving to an asymptotic inertial observer.

According to (10.86) with $\omega = 0$, the 4-velocity of \mathcal{O} is collinear to the Killing vector ξ : $u = \alpha \xi$. We conclude that a static observer can exist only where ξ 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 6 on p. 329).

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 \mathcal{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 dt . The associated normal vector, N say, is obtained by metric duality [cf. Eq. (A.46)-(A.47)]:

$$N = -\vec{dt} = -\vec{\nabla}t. \quad (10.88)$$

In components:

$$N^\alpha = -g^{\alpha\mu}(dt)_\mu = -g^{\alpha\mu}\delta^t_\mu = -g^{t\alpha}.$$

The minus sign has been chosen to have N future-directed, as we shall see below. In view of the Boyer-Lindquist components (10.16) of the inverse metric, we get

$$N = \frac{1}{\Delta} \left(r^2 + a^2 + \frac{2a^2mr \sin^2 \theta}{\rho^2} \right) \partial_t + \frac{2amr}{\rho^2 \Delta} \partial_\varphi. \quad (10.89)$$

The scalar square of N is

$$N \cdot N = g(N, N) = N_\mu N^\mu = \delta^t_\mu N^\mu = N^t = -\frac{1}{\Delta} \left(r^2 + a^2 + \frac{2a^2mr \sin^2 \theta}{\rho^2} \right). \quad (10.90)$$

Now, the quantity inside the parentheses is positive everywhere except in the Carter time machine $\mathcal{T} \subset \mathcal{M}_{\text{III}}$ discussed in Sec. 10.2.5. Indeed, up to the factor $\sin^2 \theta \geq 0$, it coincides with expression (10.21) of $\eta \cdot \eta$. Since $\Delta = (r - r_+)(r - r_-)$ is positive on \mathcal{M}_{I} and \mathcal{M}_{III} , and negative on \mathcal{M}_{II} , we conclude that the locus where N is timelike is

$$\mathcal{M}_{\text{ZAMO}} := \mathcal{M}_{\text{I}} \cup (\mathcal{M}_{\text{III}} \setminus \mathcal{T}). \quad (10.91)$$

This is thus the region where the observer \mathcal{O} is defined. Note that it does not contain the horizons \mathcal{H} and \mathcal{H}_{in} , which is not surprising since 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 $\mathcal{M}_{\text{ZAMO}}$, the timelike vector \mathbf{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 \mathbf{k} and we have, using the Boyer-Lindquist components (10.49) of \mathbf{k} :

$$\mathbf{k} \cdot \mathbf{N} = N_\mu k^\mu = -k^t = -\frac{r^2 + a^2}{\Delta} < 0 \text{ on } \mathcal{M}_{\text{ZAMO}},$$

so that Lemma 1.2 [cf. Eq. (1.6a)] let us conclude that \mathbf{N} is future-directed.

Choosing the observer \mathcal{O} as having his worldline orthogonal to Σ_t means that \mathcal{O} 's 4-velocity \mathbf{n} is the unit timelike vector introduced by Eq (10.79). Equivalently, \mathbf{n} is \mathbf{N} rescaled to form a unit vector:

$$\mathbf{n} = N \mathbf{N} = -N \vec{\nabla} t, \quad (10.92)$$

with⁴ $N := (-\mathbf{N} \cdot \mathbf{N})^{-1/2}$. In view of Eq. (10.90), we get

$$N = \sqrt{\Delta} \left(r^2 + a^2 + \frac{2a^2mr \sin^2 \theta}{\rho^2} \right)^{-1/2}. \quad (10.93)$$

N is called the ***lapse function***, for it relates the increment $d\tau$ in the proper time of \mathcal{O} to the change dt of the coordinate t when moving from Σ_t to Σ_{t+dt} via

$$d\tau = N dt. \quad (10.94)$$

Proof. The infinitesimal vector that connects the point of proper time τ to that of proper time $\tau + d\tau$ along \mathcal{O} 's worldline is $d\mathbf{x} = d\tau \mathbf{n}$ (by the very definition of the 4-velocity \mathbf{n} , compare Eq. (1.16)). The corresponding increment in the coordinate t is given by formula (A.21): $dt = \langle dt, d\mathbf{x} \rangle$; we have then, using Eq. (10.92) to express $\vec{\nabla} t$,

$$dt = \langle dt, d\mathbf{x} \rangle = \vec{\nabla} t \cdot d\mathbf{x} = d\tau \vec{\nabla} t \cdot \mathbf{n} = -\frac{d\tau}{N} \underbrace{\mathbf{n} \cdot \mathbf{n}}_{=1} = \frac{d\tau}{N},$$

hence formula (10.94). \square

A ***zero-angular-momentum observer (ZAMO)*** is an observer \mathcal{O} of the above type, i.e. whose worldline \mathcal{L} is normal to the hypersurfaces Σ_t of constant Boyer-Lindquist time t , and whose orthonormal frame $(\mathbf{e}_{(\alpha)})$ is related to the Boyer-Lindquist coordinate frame (∂_α) by

$$\mathbf{e}_{(0)} = \mathbf{n} = \frac{\sqrt{\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta}}{\rho \sqrt{\Delta}} \partial_t + \frac{2amr}{\rho \sqrt{\Delta[\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta]}} \partial_\varphi \quad (10.95a)$$

$$\mathbf{e}_{(r)} = \frac{\sqrt{\Delta}}{\rho} \partial_r \quad (10.95b)$$

$$\mathbf{e}_{(\theta)} = \frac{1}{\rho} \partial_\theta \quad (10.95c)$$

$$\mathbf{e}_{(\varphi)} = \frac{\rho}{\sin \theta \sqrt{\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta}} \partial_\varphi, \quad (10.95d)$$

⁴Note that N is not defined as the (pseudo)norm of \mathbf{N} , as the notation might suggest, but rather as the inverse of it.

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 (10.95a) for \mathbf{n} has been obtained by combining Eqs. (10.92), (10.93) and (10.89). Given the Boyer-Lindquist components (10.8) of \mathbf{g} , one readily check that $(\mathbf{e}_{(r)}, \mathbf{e}_{(\theta)}, \mathbf{e}_{(\varphi)})$ is an orthonormal basis of spacelike vectors (see the notebook D.5.5 for a SageMath computation). Moreover, these vectors are all tangent to Σ_t , since $(\partial_r, \partial_\theta, \partial_\varphi)$ are. They are thus orthogonal to $\mathbf{n} = \mathbf{e}_{(0)}$, which is a unit timelike vector. This completes the proof that $(\mathbf{e}_{(\alpha)})$ is an orthonormal frame.

The **ZAMO coframe** is the 4-tuple of 1-forms $(\mathbf{e}^{(\alpha)})_{0 \leq \alpha \leq 3}$ that constitutes, at each point $p \in \mathcal{L}$, a dual basis of $(\mathbf{e}_{(\alpha)}|_p)$, namely $(\mathbf{e}^{(\alpha)})$ obeys $\langle \mathbf{e}^{(\alpha)}, \mathbf{e}_{(\beta)} \rangle = \delta^\alpha_\beta$ (cf. Eq. (A.24) in Appendix A). Its expression in terms of the Boyer-Lindquist coordinate coframe (dx^α) is (cf. the notebook D.5.5)

$$\mathbf{e}^{(0)} = \frac{\rho \sqrt{\Delta}}{\sqrt{\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta}} dt \quad (10.96a)$$

$$\mathbf{e}^{(r)} = \frac{\rho}{\sqrt{\Delta}} dr \quad (10.96b)$$

$$\mathbf{e}^{(\theta)} = \rho d\theta \quad (10.96c)$$

$$\mathbf{e}^{(\varphi)} = -\frac{2amr \sin \theta}{\rho \sqrt{\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta}} dt + \frac{\sin \theta}{\rho} \sqrt{\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta} d\varphi. \quad (10.96d)$$

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 $\mathcal{M}_{\text{ZAMO}}$. 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 $\partial_t = \xi$, are not orthogonal to Σ_t , except for $a = 0$. The orthogonal decomposition of ξ into a part along the normal \mathbf{n} and a part tangent to Σ_t defines the **shift vector** β :

$$\boxed{\xi = N \mathbf{n} + \boldsymbol{\beta}}, \quad \mathbf{n} \cdot \boldsymbol{\beta} = 0. \quad (10.97)$$

From Eq. (10.95a), we get

$$\boldsymbol{\beta} = -\frac{2amr}{\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta} \partial_\varphi. \quad (10.98)$$

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. [10, 43, 226, 455]). 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. (10.86) and Eq. (10.97) rewritten as $\mathbf{n} = N^{-1}(\xi - \beta^\varphi \boldsymbol{\eta})$. We thus get immediately $\omega = -\beta^\varphi$. Hence

$$\boxed{\omega := \left. \frac{d\varphi}{dt} \right|_{\mathcal{L}} = \frac{2amr}{\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta}}. \quad (10.99)$$

Note that $\omega = 0$ for $a = 0$ (Schwarzschild black hole) and that ω decays quite rapidly with r :

$$\omega \underset{r \rightarrow \pm\infty}{\sim} \frac{2am}{r^3} = \frac{2J}{r^3}, \quad (10.100)$$

where Eq. (10.80) 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 $\mathbf{a} := \nabla_{\mathbf{n}} \mathbf{n}$ is nonzero. Indeed, for any family of observers orthogonal to a spacelike foliation $(\Sigma_t)_{t \in \mathbb{R}}$, it can be shown that \mathbf{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. [226]): $\mathbf{a} = \vec{\nabla} \ln N + (\nabla_{\mathbf{n}} \ln N) \mathbf{n}$. In the present case, this expression simplifies since $\nabla_{\mathbf{n}} \ln N = n^\mu \partial_\mu \ln N = 0$, as a result of $\mathbf{n} = n^t \partial_t + n^\varphi \partial_\varphi$. We then get $\mathbf{a} = \vec{\nabla} \ln N$, i.e. $a^\alpha = g^{\alpha\mu} \partial_\mu \ln N$, so that

$$\mathbf{a} = \frac{\Delta}{\rho^2 N} \frac{\partial N}{\partial r} \partial_r + \frac{1}{\rho^2 N} \frac{\partial N}{\partial \theta} \partial_\theta.$$

Using expression (10.93) for N , we get⁵

$$\boxed{\mathbf{a} = \frac{m[\rho^2(r^4 - a^4) + 2\Delta(ra \sin \theta)^2]}{\rho^3 \sqrt{\Delta} [\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta]} \mathbf{e}_{(r)} - \frac{2a^2mr(r^2 + a^2) \sin \theta \cos \theta}{\rho^3 [\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta]} \mathbf{e}_{(\theta)}} \quad (10.101)$$

We have thus $\mathbf{a} \neq 0$ 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

$$\mathbf{a} \underset{r \rightarrow \pm\infty}{\sim} \frac{m}{r^2} \mathbf{e}_{(r)}. \quad (10.102)$$

In the asymptotic regions, the non-relativistic gravitational field felt by the observer is $\mathbf{g} = -\mathbf{a}$, so that we recover the standard Newtonian expression for $r \rightarrow +\infty$ (\mathcal{M}_1 side). In the asymptotic region of \mathcal{M}_{III} , i.e. for $r \rightarrow -\infty$, $\mathbf{e}_{(r)} = (\sqrt{\Delta}/\rho) \partial_r$ is oriented towards the black hole, so that \mathbf{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 \mathcal{M}_{III} discussed in Sec. 10.2.2.

Despite ω as given by Eq. (10.99) 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 $\ell := \boldsymbol{\eta} \cdot \mathbf{n}$ [cf. Eq. (7.26)], is identically zero, the observer's 4-velocity \mathbf{n} being orthogonal to $\boldsymbol{\eta} = \partial_\varphi$. By *loosely 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. 11.3.4 [cf. Eq. (11.72)].

The name *locally non-rotating observer*, initially given to a ZAMO (cf. historical note below), conveys other specificities of such a observer \mathcal{O} :

⁵See the notebook D.5.5; see also Eqs. (70)-(71) of Ref. [449] or Eqs. (A.9) and (A.10) of Ref. [60].

- \mathcal{O} does not measure any component along $e_{(\varphi)}$ for the 3-momentum \mathbf{P} of any null or timelike geodesic that has a zero conserved angular momentum ($L = 0$):

$$P^{(\varphi)} = e_{(\varphi)} \cdot \mathbf{P} = e_{(\varphi)} \cdot [\mathbf{p} + (\mathbf{n} \cdot \mathbf{p}) \mathbf{n}] = e_{(\varphi)} \cdot \mathbf{p} = e_{(\varphi)}^\varphi \boldsymbol{\eta} \cdot \mathbf{p} = e_{(\varphi)}^\varphi L = 0,$$

where \mathbf{p} stands for the particle's 4-momentum and we have used expression (1.24) for the 3-momentum, along with $e_{(\varphi)} \cdot \mathbf{n} = 0$, formula (10.95d) for $e_{(\varphi)}$ and the definition $L := \boldsymbol{\eta} \cdot \mathbf{p}$ of the conserved angular momentum [cf. Eq. (7.1b), as well as Eq. (11.2b) below].

- for \mathcal{O} , the directions $e_{(\varphi)}$ and $-e_{(\varphi)}$ are equivalent, insofar as the proper time measured by \mathcal{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. [36] 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 \mathcal{O} of 4-velocity \mathbf{n} and 4-acceleration \mathbf{a} is defined as the spacelike vector $\boldsymbol{\omega}_{\text{rot}}$ orthogonal to \mathbf{n} such that the evolution of \mathcal{O} 's orthonormal frame $(e_{(\alpha)})$ along \mathcal{O} 's worldline \mathcal{L} takes the form (see e.g. Sec. 13.6 of Ref. [370] or Sec. 4.5 of Ref. [227])

$$\nabla_{\mathbf{n}} e_{(\alpha)} = \underbrace{(\mathbf{a} \cdot e_{(\alpha)}) \mathbf{n} - (\mathbf{n} \cdot e_{(\alpha)}) \mathbf{a}}_{\Omega_{\text{FW}}(e_{(\alpha)})} + \boldsymbol{\omega}_{\text{rot}} \times_{\mathbf{n}} e_{(\alpha)}, \quad (10.103)$$

where the cross product in the hyperplane orthogonal to \mathbf{n} , $\boldsymbol{\omega}_{\text{rot}} \times_{\mathbf{n}} e_{(\alpha)}$, is the unique vector orthogonal to both \mathbf{n} and $\boldsymbol{\omega}_{\text{rot}}$ such that $(\boldsymbol{\omega}_{\text{rot}} \times_{\mathbf{n}} e_{(\alpha)}) \cdot \mathbf{v} = \epsilon(\mathbf{n}, \boldsymbol{\omega}_{\text{rot}}, e_{(\alpha)}, \mathbf{v})$ for any vector \mathbf{v} , ϵ being the Levi-Civita tensor (cf. Sec. A.3.4). The Ω_{FW} operator that appears in the right-hand side of Eq. (10.103) is called the **Fermi-Walker operator**. It appears as soon as the observer is accelerating, even if he is non-rotating. It corrects the parallel transport of $(e_{(\alpha)})$ along \mathcal{L} , which would be realized by $\nabla_{\mathbf{n}} e_{(\alpha)} = 0$, to ensure that $(e_{(\alpha)})$ remains an orthonormal frame. A vector field \mathbf{v} that obeys $\nabla_{\mathbf{n}} \mathbf{v} = \Omega_{\text{FW}}(\mathbf{v})$ is said to be **Fermi-Walker transported** along \mathcal{L} . Physically, Fermi-Walker transport is realized by a free gyroscope: its spin vector with respect to \mathcal{O} is Fermi-Walker transported along \mathcal{L} . Hence the 4-rotation $\boldsymbol{\omega}_{\text{rot}}$ of an observer, which is an *absolute* quantity (like the 4-acceleration \mathbf{a} , it does not depend on any other observer), measures the motion of his spacelike triad $(e_{(i)})$ with respect to an orthonormal triad whose vectors are aligned along gyroscopes axes. For the ZAMO, the 4-rotation turns out to be⁶

$$\boldsymbol{\omega}_{\text{rot}} = -\frac{\omega}{\rho^3} \left\{ a^2 \sqrt{\Delta} \cos \theta \sin^2 \theta \mathbf{e}_{(r)} + \left[r(r^2 + a^2) + \frac{\rho^2}{2r}(r^2 - a^2) \right] \sin \theta \mathbf{e}_{(\theta)} \right\}, \quad (10.104)$$

where ω is given by Eq. (10.99). Hence, as soon as $a \neq 0$, $\boldsymbol{\omega}_{\text{rot}} \neq 0$. Far from the black hole, $\rho \sim |r|$ and we get

$$\boldsymbol{\omega}_{\text{rot}} \underset{r \rightarrow \pm\infty}{\sim} -\frac{3\omega}{2} \frac{r}{|r|} \sin \theta \mathbf{e}_{(\theta)} = -\frac{3J}{|r|^3} \sin \theta \mathbf{e}_{(\theta)}. \quad (10.105)$$

⁶See the notebook D.5.5 for the computation; see also Eqs. (73)-(74) of Ref. [449].

In particular, in the equatorial plane ($\theta = \pi/2$), $\omega_{\text{rot}} \sim 3J/|r|^3 e_z$, where $e_z = -e_{(\theta)}$ 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 u is defined as the “magnetic” part Ω in the “electric/magnetic” decomposition⁷ of the 2-form $d\underline{u}$ with respect to u (cf. Sec. A.4.3):

$$d\underline{u} = \underline{a} \wedge \underline{u} + \Omega \quad \text{with} \quad \Omega(u, \cdot) = 0, \quad (10.106)$$

where \underline{a} is the 1-form associated by metric duality to the 4-acceleration $a := \nabla_u u$ of the worldlines. That $\Omega = 0$ for the ZAMO congruence follows from its orthogonality to a family of hypersurfaces, namely $(\Sigma_t)_{t \in \mathbb{R}}$. Indeed, we deduce from (10.92) that $d\underline{n} = -d(Ndt) = -dN \wedge dt$, since $ddt = 0$ [cf. Eq. (A.92)]. Using (10.92) again, we get $d\underline{n} = d \ln N \wedge \underline{n}$. Comparing with (10.106) with $u = n$, we conclude that $\Omega = 0$ for the ZAMO congruence⁸. Physically, this means that if each ZAMO sets up a orthonormal spatial frame (e'_i) with axes aligned along the spin of free gyroscopes, instead of the frame $(e_{(i)})$ defined by (10.95), 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 [36] under the name of **locally non-rotating 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 [41]. The name **ZAMO** has been coined by Kip S. Thorne and Douglas MacDonald in 1982 [479].

10.7.4 Carter observers

A **Carter observer** is a stationary observer \mathcal{O} defined in the region $\mathcal{M}_I \cup \mathcal{M}_{III}$ of Kerr spacetime, whose worldline \mathcal{L} has the following equation in Boyer-Lindquist coordinates:

$$t(\tau) = \frac{r^2 + a^2}{\rho\sqrt{\Delta}}\tau + \text{const}, \quad r(\tau) = \text{const}, \quad \theta(\tau) = \text{const}, \quad \varphi(\tau) = \frac{a}{\rho\sqrt{\Delta}}\tau + \text{const}, \quad (10.107)$$

where τ is \mathcal{O} 's proper time, and which is equipped with the following orthonormal frame:

$$\varepsilon_{(0)} = \frac{r^2 + a^2}{\rho\sqrt{\Delta}} \partial_t + \frac{a}{\rho\sqrt{\Delta}} \partial_\varphi \quad (10.108a)$$

$$\varepsilon_{(r)} = \frac{\sqrt{\Delta}}{\rho} \partial_r \quad (10.108b)$$

$$\varepsilon_{(\theta)} = \frac{1}{\rho} \partial_\theta \quad (10.108c)$$

$$\varepsilon_{(\varphi)} = \frac{a}{\rho} \sin \theta \partial_t + \frac{1}{\rho \sin \theta} \partial_\varphi. \quad (10.108d)$$

⁷This name stems from the decomposition of the electromagnetic field 2-form F (cf. Sec. 1.5.2) with respect to an observer of 4-velocity u into the electric field E and the magnetic field B , both measured by that observer, according to $F = \underline{u} \wedge \underline{E} + \star(\underline{u} \wedge \underline{B})$, where \star stands for the Hodge dual, cf. Eq. (5.38).

⁸Incidentally, we also get $\underline{a} = d \ln N + \alpha \underline{n}$ and Eq. (10.101) shows that α is actually zero.

We note that the frame $(\varepsilon_{(\alpha)})$ is well-defined because $\Delta > 0$ in $\mathcal{M}_I \cup \mathcal{M}_{III}$. It is an easy exercise to check that $(\varepsilon_{(\alpha)})$ is an orthonormal frame (see the notebook D.5.6). In particular, we have $\varepsilon_{(0)} \cdot \varepsilon_{(0)} = -1$ and, using the Boyer-Lindquist components (10.50) of \mathbf{k} :

$$\mathbf{k} \cdot \varepsilon_{(0)} = k_\mu \varepsilon_{(0)}^\mu = -\frac{r^2 + a^2}{\rho \sqrt{\Delta}} + a \sin^2 \theta \frac{a}{\rho \sqrt{\Delta}} = -\frac{\rho}{\sqrt{\Delta}} < 0,$$

which, by virtue of Lemma 1.2 (Sec. 1.2.2), shows that $\varepsilon_{(0)}$ is future-directed. Moreover, we notice from (10.107) and (10.108a) that $\varepsilon_{(0)}^\alpha = dx^\alpha/d\tau$, which proves that the worldline \mathcal{L} is timelike and $\varepsilon_{(0)}$ is the 4-velocity of observer \mathcal{O} . Since r and θ are constant along \mathcal{L} , we note also that expression (10.108a) for $\varepsilon_{(0)}$ agrees with the general form (10.85) 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 $(\varepsilon^{(\alpha)})_{0 \leq \alpha \leq 3}$ such that $\langle \varepsilon^{(\alpha)}, \varepsilon_{(\beta)} \rangle = \delta_\beta^\alpha$ (cf. Sec. A.2.4 in Appendix A). It is related to the Boyer-Lindquist coordinate coframe (dx^α) by⁹

$$\varepsilon^{(0)} = \frac{\sqrt{\Delta}}{\rho} (dt - a \sin^2 \theta d\varphi) \quad (10.109a)$$

$$\varepsilon^{(r)} = \frac{\rho}{\sqrt{\Delta}} dr \quad (10.109b)$$

$$\varepsilon^{(\theta)} = \rho d\theta \quad (10.109c)$$

$$\varepsilon^{(\varphi)} = \frac{\sin \theta}{\rho} (-adt + (r^2 + a^2)d\varphi). \quad (10.109d)$$

Remark 4: By definition of an orthonormal frame, the spacetime metric can be written

$$g = -\varepsilon^{(0)} \otimes \varepsilon^{(0)} + \varepsilon^{(r)} \otimes \varepsilon^{(r)} + \varepsilon^{(\theta)} \otimes \varepsilon^{(\theta)} + \varepsilon^{(\varphi)} \otimes \varepsilon^{(\varphi)}. \quad (10.110)$$

Substituting Eqs. (10.109) for the 1-forms $\varepsilon^{(\alpha)}$ and using the notation $\omega^2 := \omega \otimes \omega$ for the tensor-product square of a 1-form [cf. Eq. (A.39)], we get the following expression of the metric with respect to Boyer-Lindquist coordinates:

$$g = -\frac{\Delta}{\rho^2} (dt - a \sin^2 \theta d\varphi)^2 + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2 + \frac{\sin^2 \theta}{\rho^2} ((r^2 + a^2)d\varphi - adt)^2. \quad (10.111)$$

This expression, which is equivalent to (10.8), is often found in textbooks (cf. e.g. Eq. (33.2) in Ref. [370]). It is called the *canonical form* by Carter [94].

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. (10.108a) and (10.108b) and comparing the result with Eqs. (10.49) and (10.61), we can express the tangent vectors \mathbf{k} and ℓ of respectively the ingoing and outgoing principal null geodesics as

$$\boxed{\mathbf{k} = \frac{\rho}{\sqrt{\Delta}} (\varepsilon_{(0)} - \varepsilon_{(r)})} \quad \text{and} \quad \boxed{\ell = \frac{\rho \sqrt{\Delta}}{2(r^2 + a^2)} (\varepsilon_{(0)} + \varepsilon_{(r)})}. \quad (10.112)$$

⁹cf. the notebook D.5.6.

In other words, the principal null vectors \mathbf{k} and ℓ lie in $\text{Span}(\boldsymbol{\varepsilon}_{(0)}, \boldsymbol{\varepsilon}_{(r)})$. Physically, this means that the Carter observer see the principal null geodesics having a pure radial motion (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. [94, 390]).

By comparing expression (10.108a) for a Carter observer's 4-velocity $\boldsymbol{\varepsilon}_{(0)}$ with Eq. (10.86), we get immediately the angular velocity around the symmetry axis of the Carter observer as measured by an asymptotically distant static observer:

$$\boxed{\omega_C := \left. \frac{d\varphi}{dt} \right|_{\mathcal{L}} = \frac{a}{r^2 + a^2}.} \quad (10.113)$$

Remark 5: Expression (10.113) is much simpler than that of the angular velocity ω of a ZAMO, as given by Eq. (10.99). In particular, ω_C does not depend on θ , contrary to ω .

The 4-acceleration $\mathbf{a}_C := \nabla_{\boldsymbol{\varepsilon}_{(0)}} \boldsymbol{\varepsilon}_{(0)}$ of the Carter observer is¹⁰

$$\boxed{\mathbf{a}_C = \frac{m(r^2 - a^2) + a^2(m - r) \sin^2 \theta}{\rho^3 \sqrt{\Delta}} \boldsymbol{\varepsilon}_{(r)} - \frac{a^2}{\rho^3} \sin \theta \cos \theta \boldsymbol{\varepsilon}_{(\theta)}}. \quad (10.114)$$

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

$$\boxed{\mathbf{a}_C \underset{r \rightarrow \pm\infty}{\sim} \frac{m}{r^2} \boldsymbol{\varepsilon}_{(r)}}, \quad (10.115)$$

i.e. the same Newtonian behavior as for the ZAMO 4-acceleration [Eq. (10.102)], with the attractive character in \mathcal{M}_I and the repelling one in \mathcal{M}_{III} .

The 4-rotation vector $\boldsymbol{\omega}_{\text{rot}}^C$ of the Carter observer is computed in a way similar to that of the ZAMO [Eq. (10.104)], i.e. by subtracting the Fermi-Walker part from $\nabla_{\boldsymbol{\varepsilon}_{(0)}} \boldsymbol{\varepsilon}_{(i)}$ for $i \in \{1, 2, 3\}$ and writing it as $\boldsymbol{\omega}_{\text{rot}}^C \times_{\boldsymbol{\varepsilon}_{(0)}} \boldsymbol{\varepsilon}_{(i)}$. We get¹¹

$$\boxed{\boldsymbol{\omega}_{\text{rot}}^C = \frac{a}{\rho^3} \left[\sqrt{\Delta} \cos \theta \boldsymbol{\varepsilon}_{(r)} - r \sin \theta \boldsymbol{\varepsilon}_{(\theta)} \right]}. \quad (10.116)$$

Due to $\sqrt{\Delta} \sim |r|$ and $\rho \sim |r|$ when $r \rightarrow \pm\infty$, the asymptotic behavior is

$$\boxed{\boldsymbol{\omega}_{\text{rot}}^C \underset{r \rightarrow \pm\infty}{\sim} \frac{a}{r^2} \left[\cos \theta \boldsymbol{\varepsilon}_{(r)} - (\text{sgn } r) \sin \theta \boldsymbol{\varepsilon}_{(\theta)} \right]}, \quad (10.117)$$

with $\text{sgn } r = +1$ in \mathcal{M}_I and $\text{sgn } r = -1$ in the asymptotic region of \mathcal{M}_{III} . Introducing in both the region \mathcal{M}_I and in the asymptotic region of \mathcal{M}_{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

$$\boxed{\boldsymbol{\omega}_{\text{rot}}^C \underset{r \rightarrow \pm\infty}{\sim} (\text{sgn } r) \frac{a}{r^2} \partial_z}. \quad (10.118)$$

¹⁰See the notebook D.5.6 for the computation; see as well Eqs. (90)-(91) of Ref. [449].

¹¹See the notebook D.5.6 for the computation; see as well Eq. (93) of Ref. [449].

We have thus ω_{rot}^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 ω_{rot} : Eq. (10.105) shows that ω_{rot} is collinear to ∂_z only at $\theta = \pi/2$. Note also that ω_{rot} decays faster than $\omega_{\text{rot}}^C: |r|^{-3}$ versus r^{-2} . Besides note that the full expression of ω_{rot}^C , Eq. (10.116), is much simpler than that of ω_{rot} , Eq. (10.104). The same remark holds for the 4-acceleration a_C [Eq. (10.114)] versus the ZAMO 4-acceleration a [Eq. (10.101)]. 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. (10.112).

Historical note : The Carter frame has been introduced by Brandon Carter in 1968 [90], 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. [90]). The explicit form for Kerr spacetime, i.e. the system (10.109), can be found in Carter's lecture at the 1972 Les Houches Summer School [94] (Eqs. (5.19)-(5.24) and (5.31)-(5.34); cf. footnote 41 in the reprinted version [94]). The Carter frame has been popularized by Roman Znajek in 1977 [532] for studying electromagnetic processes around a black hole and has been heavily used in Barrett O'Neill's monograph about Kerr black holes [390], where it is called the *Boyer-Lindquist frame field*. In particular O'Neill's *canonical Kerr vector fields* V and W are $V = \rho\sqrt{\Delta}\varepsilon_{(0)}$ and $W = \rho\sin\theta\varepsilon_{(\varphi)}$.

10.7.5 Asymptotic inertial observers

For $r \rightarrow \pm\infty$, we see on Eqs. (10.102), (10.105), (10.115) and (10.117) 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 $r \rightarrow \pm\infty$ in Eqs. (10.100) and (10.113)). 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 ξ .

10.8 Maximal analytic extension

10.8.1 Carter-Penrose diagrams

It is useful to have a “compactified view” of the whole Kerr spacetime (\mathcal{M}, 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**¹² of Kerr spacetime (\mathcal{M}, g) the image $\Pi(\mathcal{M})$ of a differentiable map $\Pi : \mathcal{M} \rightarrow \mathbb{R}^2$, $(\tilde{t}, r, \theta, \tilde{\varphi}) \mapsto (T, X)$ such that

1. $T = T(\tilde{t}, r), X = X(\tilde{t}, r);$
2. $\Pi(\mathcal{M})$ is a bounded region of \mathbb{R}^2 ;
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 (\mathcal{M}, g) in both cases.

¹²See the historical notes on p. 298 and p. 366.

Property 1 implies that Π acts as a *projection*, leaving out any information in the non-ignorable coordinate θ . Property 2 provides the “compactified” view¹³, and Property 3 makes $\Pi(\mathcal{M})$ look “conformal” to the 2-dimensional Minkowski spacetime generated by the metric $f = -dT \otimes dT + dX \otimes dX$ on \mathbb{R}^2 . However, let us stress that Π is not a priori connected to a proper conformal completion of (\mathcal{M}, 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 \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{III} ,

$$T(\tilde{t}, r) = U(u) + V(v) \quad \text{and} \quad X(\tilde{t}, r) = V(v) - U(u), \quad (10.119)$$

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 \mathcal{M}_I and \mathcal{M}_{III} and decreasing in \mathcal{M}_{II} (cf. Fig. 10.8) and (ii) $V(v)$ is a monotonic increasing function of the advanced Kerr time $v = \tilde{t} + r$ [Eq. (10.33)].

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 \mathcal{H} and \mathcal{H}_{in} , which are generated by outgoing principal null geodesics, appear as lines inclined by $+45^\circ$ 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. [125] (see also Chap. 7 of Ref. [118]); it differs from that considered here by demanding, instead of Properties 1 to 3, that Π maps any timelike curve of¹⁴ $(\mathcal{M} \setminus \mathcal{T}, g)$ to a timelike curve of the Minkowski spacetime (\mathbb{R}^2, f) and that any timelike curve in $\Pi(\mathcal{M} \setminus \mathcal{T})$ is the image of a timelike curve of $(\mathcal{M} \setminus \mathcal{T}, g)$.

A schematic Carter-Penrose diagram of the Kerr spacetime (\mathcal{M}, g) is shown in Fig. 10.9. 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. (10.119), with $k = 0, 1$ and 2 on respectively \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{III} , would lead to such a diagram. By virtue of Property 1, one may think of each point in the diagram of Fig. 10.9 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 \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{III} of the Kerr spacetime \mathcal{M} is mapped to the interior of a square tilted by 45° (a “diamond”) in the diagram of Fig. 10.9. 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 \mathcal{M}_I and \mathcal{M}_{III} and in the opposite direction in \mathcal{M}_{II} . These ranges and directions follow directly from $v = \tilde{t} + r$ and the graph of $u - \tilde{t}$ shown in Fig. 10.8.

The dotted curves in Fig. 10.9 represent some hypersurfaces $r = \text{const}$. According to the results of Sec. 10.5.1, such hypersurfaces are timelike in \mathcal{M}_I and \mathcal{M}_{III} , spacelike in \mathcal{M}_{II} and null for $r = r_-$ or $r = r_+$. Since the Killing vector field ξ is tangent to the hypersurfaces $r = \text{const}$, the dotted curves in Fig. 10.9 can also be seen as the projections of the field lines of ξ , i.e. of the orbits of the stationary group action.

¹³Property 2 refers to a “bounded region” instead of “compact region”, because $\Pi(\mathcal{M})$ is in general an open subset of \mathbb{R}^2 and therefore not a compact subset.

¹⁴ \mathcal{T} is the Carter time machine discussed in Sec. 10.2.5.



Figure 10.9: Carter-Penrose diagram of the Kerr spacetime (\mathcal{M}, g) , with $\mathcal{M} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathcal{R}$ (cf. Eq. (10.29) 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 $r = \text{const}$.

As pointed out in Sec. 10.5.2, the asymptotic structure of region \mathcal{M}_I is identical to that of \mathcal{M}_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 \mathcal{I}^+ (resp. \mathcal{I}^-) to the diagram of Fig. 10.9. Note that \mathcal{I}^+ corresponds to the limit $v \rightarrow +\infty$ in \mathcal{M}_I , while \mathcal{I}^- corresponds to the limit $u \rightarrow -\infty$ in \mathcal{M}_I .

10.8.2 Constructing the maximal extension

The ingoing principal null geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ (dashed green lines in Fig. 10.9) 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 $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ (solid green lines in Fig. 10.9) are not complete: in \mathcal{M}_I , their affine parameter $\lambda = r$ is bounded from below by r_+ ; in \mathcal{M}_{II} , their affine parameter $\lambda = -r$ ranges in $(-r_+, -r_-)$ only and in \mathcal{M}_{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 \mathcal{M}_{III} , this indicates that the spacetime (\mathcal{M}, g) can be extended. Moreover, for $0 \leq a < m$, the event horizon \mathcal{H} is a *non-degenerate* Killing horizon (cf. Sec. 10.5.4) and we have seen in Sec. 3.4.2 that, generically, such horizons are part of a *bifurcate Killing horizon* in an extended

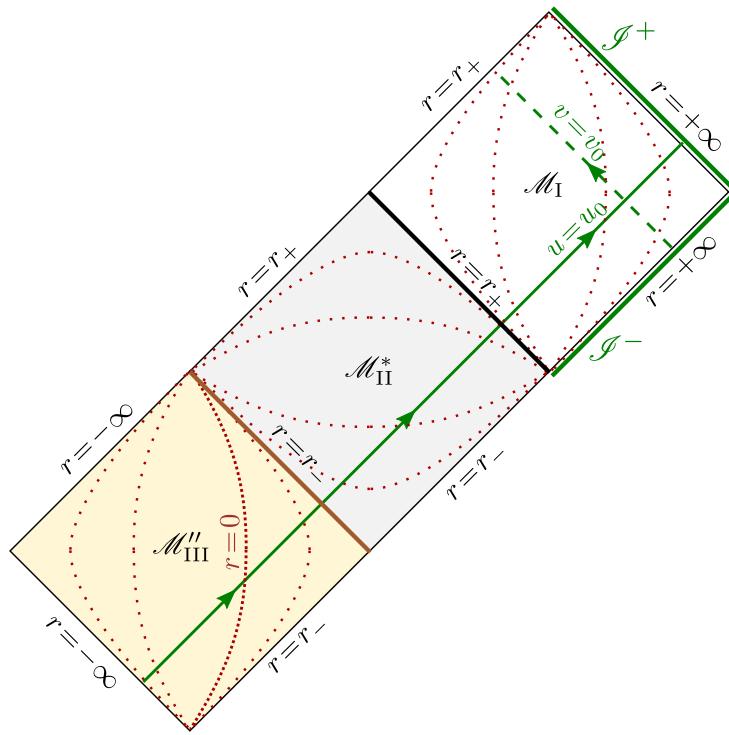


Figure 10.10: Carter-Penrose diagram of the minimal extension of \mathcal{M}_I to ensure complete outgoing principal null geodesics (one of them is drawn as a solid green line).

spacetime.

A spacetime extending (\mathcal{M}_I, g) “to the past”, so that the outgoing principal null geodesics are complete, is shown in Fig. 10.10. It is made by attaching to (\mathcal{M}_I, g) a time-reversed copy¹⁵ of (\mathcal{M}_{II}, g) , (\mathcal{M}_{II}^*, g) say, and then attaching to the latter a copy of (\mathcal{M}_{III}, g) , (\mathcal{M}_{III}'', g) say¹⁶. 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. 10.10 is bounded from above by $-r_+$. Actually the situation is completely symmetric to that of the original Kerr spacetime (\mathcal{M}, g) (Fig. 10.9). In particular, the region $\mathcal{M}_{II}^* \cup \mathcal{H}_{in} \cup \mathcal{M}_{III}''$, where \mathcal{H}_{in} is the (null) hypersurface $r = r_-$, is a *white hole*, since it is the complement of $J^+(\mathcal{J}^-)$ (cf. the definition (4.41) of a white hole). The hypersurface $r = r_+$ (black thick line in Fig. 10.10) is the corresponding *past event horizon*.

By combining the diagrams of Figs. 10.9 and 10.10, one obtains a spacetime which still contains incomplete null geodesics: the outgoing ones in regions \mathcal{M}_{II} and \mathcal{M}_{III} , and the ingoing ones in regions \mathcal{M}_{II}^* and \mathcal{M}_{III}'' . To go further, one should add new regions isometric to one of the three blocks (\mathcal{M}_I, g) , (\mathcal{M}_{II}, g) and (\mathcal{M}_{III}, g) and iterate indefinitely, leading to the Carter-Penrose diagram of Fig. 10.11. 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

¹⁵By *copy*, it is meant that (\mathcal{M}_{II}^*, g) is a spacetime isometric to (\mathcal{M}_{II}, g) and by *time-reversed* that r is increasing towards the future in \mathcal{M}_{II}^* , while it is decreasing towards the future in \mathcal{M}_{II} .

¹⁶The notation \mathcal{M}_{III}'' instead of \mathcal{M}_{III}' is for later convenience.

here and refer the reader to the seminal articles by Boyer & Lindquist [71] and Carter [89], the famous Les Houches lectures by Carter [94] or the textbook by O'Neill [390].

In the diagram of Fig. 10.11, it is clear that the event horizon \mathcal{H} and the inner horizon \mathcal{H}_{in} have been extended to bifurcate Killing horizons (cf. Sec. 3.4), the bifurcation surface of which being a 2-sphere depicted by a circular dot.

Historical note : In 1966, Brandon Carter [87] obtained the maximal analytic extension of the 2-dimensional manifold constituted by the rotation axis¹⁷ \mathcal{A} of the Kerr spacetime and drew a diagram similar to that of Fig. 10.11 (Fig. 1a of Ref. [87]). More precisely, Carter introduced on \mathcal{A} a coordinate system (T, X) (denoted (ψ, ξ) by him) in which the metric induced by \mathbf{g} on \mathcal{A} is explicitly conformal the 2-dimensional Minkowski metric $f = -dT \otimes dT + dX \otimes dX$. Thus Fig. 1a of Carter's article [87], which is reproduced as Fig. 28 of Hawking & Ellis' textbook [265], is a true conformal representation of \mathcal{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 \mathcal{A} is a 2-dimensional manifold. For the whole Kerr spacetime, the “projection” aspect is inevitable. In the same article [87], Carter suggested that the maximal analytic extension of the whole 4-dimensional manifold would be similar to that of \mathcal{A} . This was proven rigorously a year later by Robert H. Boyer and Richard W. Lindquist [71] and generalized to Kerr-Newmann spacetimes by Carter himself in 1968 [89].

10.8.3 Cauchy horizon

Let us consider a spacelike hypersurface Σ running from the asymptotically flat end of \mathcal{M}_I to the asymptotically flat end of \mathcal{M}'_I , possibly through¹⁸ \mathcal{M}_{II} or \mathcal{M}_{II}^* (cf. Fig. 10.12) 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 [265]. 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. 10.12. The spacetime metric at every point in $D^+(\Sigma)$ is entirely determined by initial data on Σ through the Einstein equation, in its 3+1 form (see e.g. Chap. 5 of Ref. [226]); 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 intersects¹⁹ Σ . It is clear on Fig. 10.12 that Σ is a Cauchy surface for $(\mathcal{M}_I, \mathbf{g})$, with $\mathcal{M}_I := \mathcal{M}_I \cup \mathcal{M}'_I \cup \mathcal{M}_{II} \cup \mathcal{M}_{II}^*$, 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 $\mathcal{M}_I \cup \mathcal{M}'_I$. We shall therefore denote it by \mathcal{H}_C . It is depicted as the blue thick line in Fig. 10.12. \mathcal{H}_C is the union of what we called the *inner horizon* in Sec. 10.5.1, i.e. the Killing horizon \mathcal{H}_{in} , and the null

¹⁷Let us recall that the *rotation axis* is the 2-dimensional Lorentzian submanifold \mathcal{A} where the Killing vector η vanishes (cf. the definition given in Sec. 5.3.6); in terms of the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$, \mathcal{A} is defined by $\theta = 0$ or π and is naturally spanned by the coordinates (\tilde{t}, r) .

¹⁸In Fig. 10.12, this is not the case: Σ goes from \mathcal{M}_I to \mathcal{M}'_I via the bifurcation sphere.

¹⁹There is a single intersection since Σ is assumed acausal; the *Cauchy surface* definition given here agrees thus with that given on p. 125.

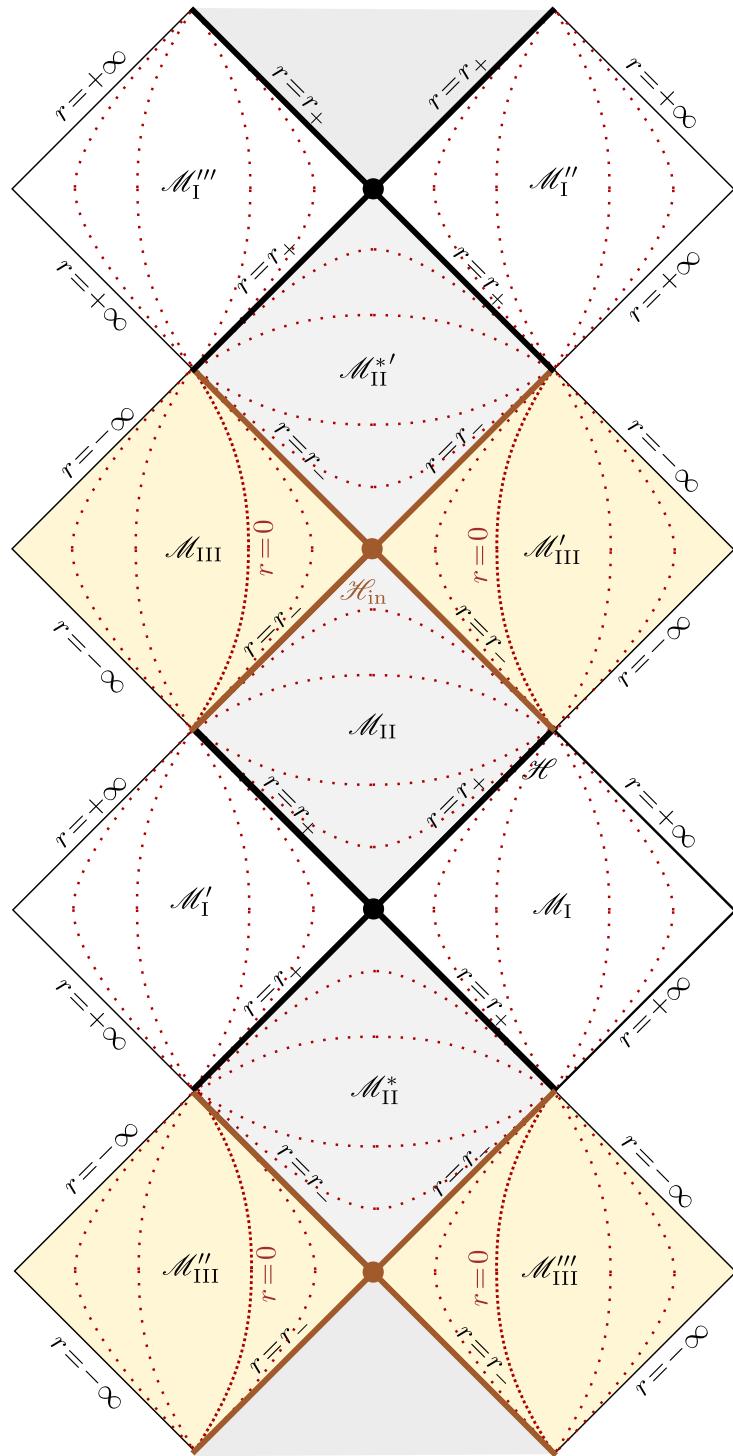


Figure 10.11: Carter-Penrose diagram of the maximal analytic extension of the Kerr spacetime. As in Figs. 10.9 and 10.10, the dotted curves mark some hypersurfaces $r = \text{const}$. The central black or light brown dots mark the bifurcation spheres of bifurcate Killing horizons.

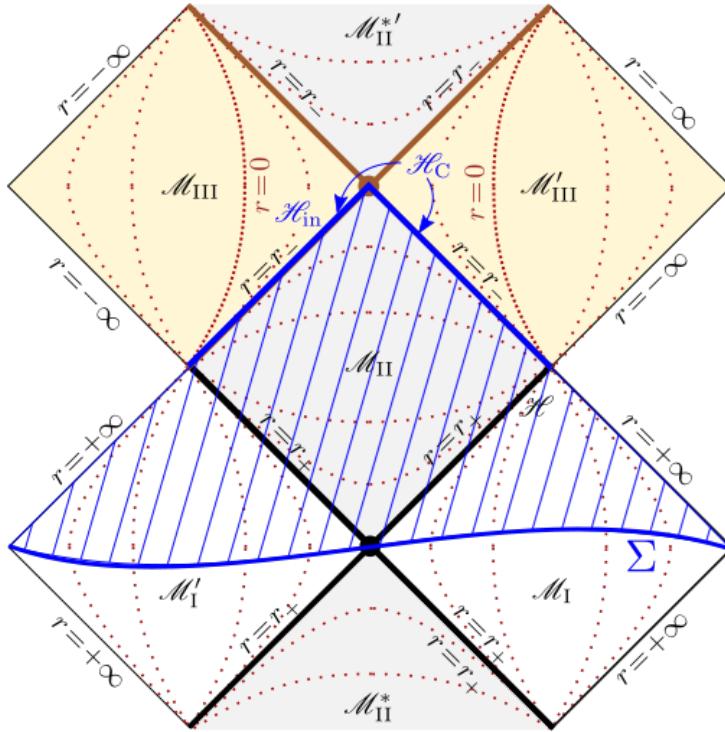


Figure 10.12: The partial Cauchy surface Σ , its future Cauchy development $D^+(\Sigma)$ (hatched) and the Cauchy horizon \mathcal{H}_C . As in Figs. 10.9–10.11, the dotted curves marks some hypersurfaces $r = \text{const}$.

boundary between \mathcal{M}_{II} and $\mathcal{M}'_{\text{III}}$, \mathcal{H}'_{in} say, which is a Killing horizon as well:

$$\mathcal{H}_C = \mathcal{H}_{\text{in}} \cup \mathcal{H}'_{\text{in}}. \quad (10.120)$$

Note that \mathcal{H}_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 \mathcal{H}_{in} given in Eq. (10.28).

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 κ_{in} of the Killing generator χ_{in} of the part \mathcal{H}_{in} of the Cauchy horizon [cf. Eq. (10.70)]. Since $mr_-/(r_-^2 + a^2) = 1/2$, we notice, from Eqs. (10.57), (10.70) and (10.71), that the Killing vector χ_{in} coincides with the vector ℓ tangent to the outgoing principal null geodesics on \mathcal{H}_{in} :

$$\chi_{\text{in}} \stackrel{\mathcal{H}_{\text{in}}}{=} \ell. \quad (10.121)$$

Hence, as the event horizon \mathcal{H} , the inner horizon \mathcal{H}_{in} is generated by outgoing principal null geodesics [compare Eq. (10.72) and see Figs. 10.6 and 10.7]. Equation (10.58) implies then

$$\nabla_{\chi_{\text{in}}} \chi_{\text{in}} \stackrel{\mathcal{H}_{\text{in}}}{=} \kappa_{\text{in}} \chi_{\text{in}}, \quad (10.122)$$

with the non-affinity coefficient κ_{in} obtained by specializing Eq. (10.58) to $r = r_-$: $\kappa_{\text{in}} = \kappa_\ell|_{r=r_-} = m(r_-^2 - a^2)/(r_-^2 + a^2)^2$. Using the expression of r_- in terms of m and a [Eq. (10.3)], we get

$$\kappa_{\text{in}} = -\frac{\sqrt{m^2 - a^2}}{2m(m - \sqrt{m^2 - a^2})}. \quad (10.123)$$

Given the assumption $0 < a < m$, we have $\kappa_{\text{in}} \neq 0$, which implies (cf. Sec. 3.3.6)

Property 10.23: non-degeneracy of the inner horizon

As long as $a < m$, the part \mathcal{H}_{in} of the Cauchy horizon \mathcal{H}_{C} is a non-degenerate Killing horizon.

We note that $\kappa_{\text{in}} < 0$. According to the results of Sec. 3.4.2, this implies

Property 10.24: Cauchy horizon as part of a bifurcate Killing horizon

\mathcal{H}_{in} is contained in a *bifurcate Killing horizon*, the bifurcation surface of which being the future boundary of \mathcal{H}_{in} (light brown small disk in Fig. 10.12). Actually, the whole Cauchy horizon \mathcal{H}_{C} is the past part of the bifurcate Killing horizon, the future part being formed by the future boundaries $r = r_-$ of \mathcal{M}_{III} and $\mathcal{M}'_{\text{III}}$ (light brown thick lines in Fig. 10.12).

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 the Jebsen-Birkhoff theorem, cf. Sec. 14.2.5), 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: it has been shown to be unstable, suffering from the so-called *mass inflation instability* discovered by Éric Poisson and Werner Israel [416] (see Ref. [77] 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 [390], as well to the review articles by Heinicke & Hehl [269], Teukolsky [475] and Visser [496].

Chapter 11

Geodesics in Kerr spacetime: generic and timelike cases

Contents

11.1 Introduction	371
11.2 Equations of geodesic motion	372
11.3 Main properties of geodesics	386
11.4 Timelike geodesics	408
11.5 Circular timelike orbits in the equatorial plane	416
11.6 Going further	439

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 can be inferred 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; they 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 \mathcal{P} in the Kerr spacetime $(\mathcal{M}, \mathbf{g})$, under the hypothesis that \mathcal{P} feels only gravity, as described by \mathbf{g} (freely falling particle). The worldline \mathcal{L} of \mathcal{P} is then necessarily a geodesic¹ of $(\mathcal{M}, \mathbf{g})$. It is a timelike geodesic if \mathcal{P} is a massive particle and a null geodesic if \mathcal{P} is massless (e.g. a photon).

In a given coordinate system (x^α) , the geodesic \mathcal{L} is described by a system of the form² $x^\alpha = x^\alpha(\lambda)$, where λ is an affine parameter³ along \mathcal{L} . We choose λ to be the affine parameter associated with \mathcal{P} 's 4-momentum \mathbf{p} , i.e. such that $dx^\alpha/d\lambda = p^\alpha$ [Eq. (1.21)]. In particular, λ is dimensionless and is increasing towards the future⁴ along \mathcal{L} . The curve \mathcal{L} is a geodesic iff the functions $x^\alpha(\lambda)$ obey the geodesic equation [Eq. (B.10) in Appendix B]:

$$\frac{d^2 x^\alpha}{d\lambda^2} + \Gamma_{\mu\nu}^\alpha \frac{dx^\mu}{d\lambda} \frac{dx^\nu}{d\lambda} = 0, \quad 0 \leq \alpha \leq 3, \quad (11.1)$$

where the $\Gamma_{\mu\nu}^\alpha$'s are the Christoffel symbols of the metric \mathbf{g} with respect to the coordinates (x^α) , as given by Eq. (A.71). Equation (11.1) is the component expression of $\nabla_p \mathbf{p} = 0$ [Eq. (1.11)]. 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 $dx^\alpha/d\lambda = F^\alpha(x^0, x^1, x^2, x^3)$.

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).

11.2.2 Integrals of motion from spacetime symmetries

As for Schwarzschild spacetime (cf. Sec. 7.2.1), we have the following property:

¹The definition and basic properties of geodesics are recalled in Appendix B; see also Sec. 1.3.2.

²In Appendices A and B, a different symbol is used for the function $X^\alpha(\lambda)$ defining \mathcal{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.

³See Sec. B.2.1 for the definition of an affine parameter along a geodesic.

⁴Let us recall that the Kerr spacetime is time-oriented, cf. Sec. 10.3.2.

Property 11.1: conserved quantities along causal geodesics

The Killing vectors ξ and η of Kerr spacetime, associated respectively with stationarity and axisymmetry [cf. Eq. (10.12)], give birth to two conserved quantities along any (causal) geodesic \mathcal{L} :

$$E := -\xi \cdot p = -g(\xi, p) \quad (11.2a)$$

$$L := \eta \cdot p = g(\eta, p), \quad (11.2b)$$

where p is the 4-momentum of the particle \mathcal{P} having \mathcal{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 \mathcal{P} , while L is called the **conserved angular momentum** or **angular momentum at infinity** of \mathcal{P} .

In particular, if \mathcal{L} reaches the asymptotic region $|r| \rightarrow +\infty$, E is nothing but \mathcal{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 ξ [compare Eq. (1.23) with Eq. (11.2a)]. Similarly, L is the component along the rotation axis of \mathcal{P} 's (total) angular momentum vector \mathbf{L}_{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 $\xi = \partial_t$ and $\eta = \partial_\varphi$, for instance Boyer-Lindquist coordinates (Sec. 10.2.1), advanced 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{p} associated to p by metric duality:

$$E = -p_t \quad (11.3a)$$

$$L = p_\varphi \quad (11.3b)$$

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 \mathcal{L} a null geodesic generator of the event horizon \mathcal{H} (cf. Sec. 10.5.3). \mathcal{L} is then an outgoing principal null geodesic $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ and the 4-momentum vector p is proportional to the null vector $\ell \stackrel{\mathcal{H}}{=} \chi = \xi + \Omega_{\mathcal{H}} \eta$ [Eq. (10.72) and (10.68)]. By definition of a generator of a null hypersurface (cf. Sec. 2.3.3), p is normal to \mathcal{H} . Since the Killing vector fields ξ and η are tangent to \mathcal{H} (for \mathcal{H} is globally preserved by the spacetime symmetries), we get immediately from the definitions (11.2) $E = 0$ and $L = 0$. Similarly, if \mathcal{L} is a null geodesic generator of the inner horizon \mathcal{H}_{in} , \mathcal{L} is an outgoing principal null geodesic $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}_{\text{in}}}$, with p proportional to the null normal $\ell \stackrel{\mathcal{H}_{\text{in}}}{=} \chi_{\text{in}} = \xi + \Omega_{\text{in}} \eta$ [Eq. (10.121) and (10.70)], so that we have $E = 0$ and $L = 0$ as well. Hence we conclude:

$$\mathcal{L} \text{ null geodesic generator of } \mathcal{H} \text{ or } \mathcal{H}_{\text{in}} \implies E = 0 \quad \text{and} \quad L = 0. \quad (11.4)$$

Example 2 (Ingoing principal null geodesics): For the ingoing principal null geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ introduced in Sec. 10.4, the 4-momentum vector is

$$\mathbf{p} = \alpha \mathbf{k}, \quad (11.5)$$

where α is a (positive) constant, since \mathbf{k} is a geodesic vector [Eq. (10.45)], as \mathbf{p} [Eq. (1.11)]. Equations (11.3) with the components relative to Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ lead then to

$$E = -\alpha k_{\tilde{t}} = -\alpha(-1) = \alpha \quad \text{and} \quad L = \alpha k_{\tilde{\varphi}} = \alpha a \sin^2 \theta,$$

where the components $k_{\tilde{t}}$ and $k_{\tilde{\varphi}}$ have been read on Eq. (10.44). Hence for any ingoing principal null geodesic $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$,

$$E > 0 \quad \text{and} \quad L = aE \sin^2 \theta. \quad (11.6)$$

Recall that θ is constant along $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$, 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 $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$, $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ and $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}_{\text{in}}}$ introduced in Sec. 10.4, the 4-momentum vector is

$$\mathbf{p} = \beta(\lambda) \ell. \quad (11.7)$$

where $\beta(\lambda)$ is a function of the affine parameter λ associated to \mathbf{p} that obeys $\beta(\lambda) > 0$, since both \mathbf{p} and ℓ are future-directed (cf. Sec. 10.4). Contrary to the coefficient α in Eq. (11.5), $\beta(\lambda)$ is not constant because ℓ is not a geodesic vector, but only a pregeodesic one: it fulfills $\nabla_{\ell} \ell = \kappa_{\ell} \ell$ with $\kappa_{\ell} \neq 0$ [Eq. (10.58)] (cf. Remark 6 on p. 344). The geodesic equation $\nabla_{\mathbf{p}} \mathbf{p} = 0$ implies that β obeys the differential equation $\nabla_{\mathbf{p}} \beta + \kappa_{\ell} \beta^2 = 0$, or equivalently

$$\frac{d\beta}{d\lambda} + \kappa_{\ell} \beta^2 = 0. \quad (11.8)$$

The outgoing principal null geodesics either (i) are confined to one of the horizons \mathcal{H} and \mathcal{H}_{in} , generating them (case of $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ and $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}_{\text{in}}}$), or (ii) never intersect them (case of $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$, cf. the solid curves in Figs. 10.6 – 10.7). In the first case, κ_{ℓ} is a constant on \mathcal{H} and \mathcal{H}_{in} , given by Eq. (10.58) with $r = r_+$ for \mathcal{H} and $r = r_-$ for \mathcal{H}_{in} . The solution of Eq. (11.8) is then

$$\beta(\lambda) = \frac{1}{\kappa_{\ell}(\lambda - \lambda_0)},$$

where λ_0 is a constant. Outside the horizons, κ_{ℓ} is the function of r given by Eq. (10.58). 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 $dr/d\lambda = p^r = \beta \ell^r$ with ℓ^r read on Eq. (10.57), we get the following linear differential equation for B :

$$(r^2 - 2mr + a^2) \frac{dB}{dr} + 2m \frac{r^2 - a^2}{r^2 + a^2} B = 0.$$

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 \mathcal{M}_1 and \mathcal{M}_{III} and a negative constant in \mathcal{M}_{II} , to ensure that $\beta(\lambda) > 0$.

Plugging (11.7) into Eqs. (11.3) with the components relative to Kerr coordinates yields

$$E = -\beta(\lambda)\ell_{\tilde{t}} = \beta(\lambda)\frac{\Delta}{2(r^2 + a^2)} \quad \text{and} \quad L = \beta(\lambda)\ell_{\tilde{\varphi}} = \beta(\lambda)\frac{a\Delta \sin^2 \theta}{2(r^2 + a^2)},$$

where the components $\ell_{\tilde{t}}$ and $\ell_{\tilde{\varphi}}$ have been read on Eq. (10.60). Given the above results for $\beta(\lambda)$, we get $E = 0$ and $L = 0$ on \mathcal{H} and \mathcal{H}_{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,

$$\text{in } \mathcal{H} \cup \mathcal{H}_{\text{in}}, \quad E = 0 \quad \text{and} \quad L = 0 \quad (11.9a)$$

$$\text{in } \mathcal{M}_I \cup \mathcal{M}_{\text{III}}, \quad E > 0 \quad \text{and} \quad L = aE \sin^2 \theta \quad (11.9b)$$

$$\text{in } \mathcal{M}_{\text{II}}, \quad E < 0 \quad \text{and} \quad L = aE \sin^2 \theta. \quad (11.9c)$$

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 \mathbf{g} in these coordinates, evaluating E and L via $E = -g_{t\mu} p^\mu$ and $L = g_{\varphi\mu} p^\mu$ yields

$$E = \left(1 - \frac{2mr}{\rho^2}\right) p^t + \frac{2amr \sin^2 \theta}{\rho^2} p^\varphi. \quad (11.10)$$

$$L = -\frac{2amr \sin^2 \theta}{\rho^2} p^t + \left(r^2 + a^2 + \frac{2a^2mr \sin^2 \theta}{\rho^2}\right) \sin^2 \theta p^\varphi, \quad (11.11)$$

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 \mathcal{L} in terms of the affine parameter λ by $p^\alpha = dx^\alpha/d\lambda$ [cf. Eq. (1.21)], i.e.

$$p^t = \frac{dt}{d\lambda}, \quad p^r = \frac{dr}{d\lambda}, \quad p^\theta = \frac{d\theta}{d\lambda}, \quad p^\varphi = \frac{d\varphi}{d\lambda}. \quad (11.12)$$

11.2.3 Mass as an integral of motion

The mass μ of particle \mathcal{P} is related to the scalar square of the 4-momentum vector \mathbf{p} via Eq. (1.10):

$$\mu^2 = -\mathbf{g}(\mathbf{p}, \mathbf{p}). \quad (11.13)$$

The fact that \mathcal{L} is a geodesic implies that μ is constant along \mathcal{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

$$\begin{aligned} \mu^2 &= \frac{1}{\Delta} \left(r^2 + a^2 + \frac{2a^2mr \sin^2 \theta}{\rho^2} \right) E^2 - \frac{4amr}{\rho^2 \Delta} EL - \frac{1}{\Delta} \left(1 - \frac{2mr}{\rho^2} \right) \frac{L^2}{\sin^2 \theta} \\ &\quad - \frac{\rho^2}{\Delta} (p^r)^2 - \rho^2 (p^\theta)^2, \end{aligned} \quad (11.14)$$

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 \mathbf{g} :

$$p_r = \frac{\rho^2}{\Delta} p^r \quad \text{and} \quad p_\theta = \rho^2 p^\theta. \quad (11.15)$$

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 K** , which is the symmetric tensor of type $(0, 2)$ defined by

$$\boxed{K := (r^2 + a^2)(\underline{k} \otimes \underline{\ell} + \underline{\ell} \otimes \underline{k}) + r^2 \mathbf{g}}, \quad (11.16)$$

where $\underline{\ell}$ and \underline{k} are the 1-forms associated by metric duality to the null vector fields \mathbf{k} and $\mathbf{\ell}$ tangent to the principal null geodesics introduced in Sec. 10.4. In index notation, Eq. (11.16) writes

$$K_{\alpha\beta} = (r^2 + a^2)(k_\alpha \ell_\beta + \ell_\alpha k_\beta) + r^2 g_{\alpha\beta}. \quad (11.17)$$

K is called a **Killing tensor** because its symmetrized covariant derivative vanishes identically:

$$\boxed{\nabla_{(\alpha} K_{\beta\gamma)} = 0}. \quad (11.18)$$

This property can be seen as a generalization of the Killing equation (3.19) to tensors of valence 2. That the tensor K defined by (11.16) obeys the Killing identity (11.18) is established in the SageMath notebook D.5.7.

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 \mathcal{L} [cf. Eq. (B.46)]:

$$\boxed{\mathcal{K} := K(\mathbf{p}, \mathbf{p}) = K_{\mu\nu}p^\mu p^\nu}. \quad (11.19)$$

\mathcal{K} is named **Carter constant** (cf. the historical note on p. 385). From the definition (11.16) of K , we have

$$\mathcal{K} = 2(r^2 + a^2)\langle \underline{k}, \mathbf{p} \rangle \langle \underline{\ell}, \mathbf{p} \rangle + r^2 \mathbf{g}(\mathbf{p}, \mathbf{p}). \quad (11.20)$$

Now by Eq. (11.13), $\mathbf{g}(\mathbf{p}, \mathbf{p}) = -\mu^2$. Besides, we have $\langle \underline{k}, \mathbf{p} \rangle = k_\mu p^\mu = k^\mu p_\mu$. 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 \mathbf{k} as given by Eq. (10.49), we get

$$\langle \underline{k}, \mathbf{p} \rangle = -\frac{r^2 + a^2}{\Delta} E - p_r + \frac{a}{\Delta} L.$$

Similarly, from the components (10.61) of $\mathbf{\ell}$, we obtain

$$\langle \underline{\ell}, \mathbf{p} \rangle = -\frac{1}{2}E + \frac{\Delta}{2(r^2 + a^2)} p_r + \frac{a}{2(r^2 + a^2)} L.$$

Accordingly, Eq. (11.20) becomes

$$\mathcal{K} = \frac{1}{\Delta} [(r^2 + a^2)E + \Delta p_r - aL] [(r^2 + a^2)E - \Delta p_r - aL] - r^2 \mu^2,$$

which can be rewritten as

$$\mathcal{K} = \frac{1}{\Delta} \left[((r^2 + a^2)E - aL)^2 - \rho^4 (p^r)^2 \right] - r^2 \mu^2. \quad (11.21)$$

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 \mathcal{P} following the geodesic \mathcal{L} visits the asymptotic region $r \rightarrow +\infty$. 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 $r \rightarrow +\infty$, we deduce from Eq. (11.21) that

$$\mathcal{K} \underset{r \rightarrow +\infty}{\sim} r^2 [E^2 - \mu^2 - (p^r)^2]. \quad (11.22)$$

As discussed above, E is \mathcal{P} 's energy as measured by the asymptotic inertial observer \mathcal{O} . Then, according to Einstein's formula (1.26), $E^2 - \mu^2 = \mathbf{P} \cdot \mathbf{P}$, where \mathbf{P} is \mathcal{P} 's linear momentum as measured by \mathcal{O} . Given that asymptotically, $p^r \sim P^r$ [cf. Eq. (1.22)], Eq. (11.22) becomes

$$\mathcal{K} \underset{r \rightarrow +\infty}{\sim} r^2 [\mathbf{P} \cdot \mathbf{P} - (P^r)^2] = r^2 [(P^{(\theta)})^2 + (P^{(\varphi)})^2],$$

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 \mathbf{P} in the orthonormal basis $(\mathbf{e}_{(r)}, \mathbf{e}_{(\theta)}, \mathbf{e}_{(\varphi)}) := (\partial_r, r^{-1}\partial_\theta, (r \sin \theta)^{-1}\partial_\varphi)$. Now the total angular momentum of \mathcal{P} measured by \mathcal{O} is

$$\mathbf{L}_{\text{tot}} := \mathbf{r} \times \mathbf{P} = -rP^{(\varphi)}\mathbf{e}_{(\theta)} + rP^{(\theta)}\mathbf{e}_{(\varphi)}.$$

Hence we may conclude that asymptotically, the Carter constant coincides with the square of \mathcal{P} 's angular momentum as measured by the inertial observer \mathcal{O} :

$$\mathcal{K} \underset{r \rightarrow +\infty}{\sim} \mathbf{L}_{\text{tot}} \cdot \mathbf{L}_{\text{tot}}. \quad (11.23)$$

We shall see later that one has always $\mathcal{K} \geq 0$. For now, let us establish the following characterization of null geodesics with $\mathcal{K} = 0$:

Property 11.2: null geodesics with vanishing Carter constant

A null geodesic \mathcal{L} has a vanishing Carter constant \mathcal{K} if, and only if, \mathcal{L} is a principal null geodesic:

$$\mathcal{K} = 0 \iff \mathcal{L} = \mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}} \quad \text{or} \quad \mathcal{L} = \mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}, \quad (11.24)$$

where $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ (resp. $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$) is one of the ingoing (resp. outgoing) principal null geodesic introduced in Sec. 10.4, with $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ standing for $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$, $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{K}}$ or $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{K}_{\text{in}}}$.

Proof. Consider expression (11.20) for \mathcal{K} . If \mathcal{L} is a null geodesic, the term $\mathbf{g}(\mathbf{p}, \mathbf{p})$ vanishes identically, so that one is left with

$$\mathcal{K} = 2(r^2 + a^2)(\mathbf{k} \cdot \mathbf{p})(\ell \cdot \mathbf{p}).$$

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

$$\mathcal{K} = 0 \iff \mathbf{k} \cdot \mathbf{p} = 0 \quad \text{or} \quad \mathbf{\ell} \cdot \mathbf{p} = 0.$$

The vectors \mathbf{k} , $\mathbf{\ell}$ and \mathbf{p} are all null. Now, according to Property 1.4 in Sec. 1.2.2, two null vectors are orthogonal iff they are collinear. Hence $\mathcal{K} = 0$ is equivalent to $\mathbf{p} = \alpha \mathbf{k}$ or $\mathbf{p} = \alpha \mathbf{\ell}$ with $\alpha > 0$. Since \mathbf{k} (resp. $\mathbf{\ell}$) is tangent to $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ (resp. $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{out}}$), this completes the proof. \square

Remark 2: Property 11.2 is consistent with the interpretation (11.23) of \mathcal{K} , since asymptotically the principal null geodesics are purely radial⁵, and hence have $\mathbf{L}_{\text{tot}} = 0$.

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 \mathcal{H} and the inner horizon \mathcal{H}_{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 $\mathcal{K} = 0$. As $\mu = 0$ by definition of a null geodesic, we see that the four integrals of motion μ , E , L and \mathcal{K} all vanish for these geodesics.

The converse is true:

Property 11.3: characterization of the null generators of the two horizons

The null geodesic generators of the event horizon \mathcal{H} and of the inner horizon \mathcal{H}_{in} are the only geodesics of Kerr spacetime having their four integrals of motion vanishing:

$$\mathcal{L} \text{ null geodesic generator of } \mathcal{H} \text{ or } \mathcal{H}_{\text{in}} \iff (\mu, E, L, \mathcal{K}) = (0, 0, 0, 0). \quad (11.25)$$

Proof. If a geodesic \mathcal{L} has $(\mu, \mathcal{K}) = (0, 0)$, \mathcal{L} is necessary null and (11.24) shows that \mathcal{L} is a principal null geodesic. Moreover Eq. (11.21) with $(\mu, E, L, \mathcal{K}) = (0, 0, 0, 0)$ implies $p^r = 0$, i.e. \mathcal{L} lies at a constant value of r . If \mathcal{L} is ingoing, then $\mathbf{p} \propto \mathbf{k}$, with the Boyer-Lindquist components of \mathbf{k} given by Eq. (10.49). Since $k^r = -1$, this precludes $p^r = 0$. Hence \mathcal{L} is an outgoing principal null geodesic and one has $\mathbf{p} \propto \mathbf{\ell}$, with the Boyer-Lindquist components of $\mathbf{\ell}$ given by Eq. (10.61). We read $\ell^r = \Delta/(2(r^2 + a^2))$ with $\Delta \neq 0$, except precisely on \mathcal{H} and \mathcal{H}_{in} ⁶. Hences $p^r = 0$ is possible only on \mathcal{H} and \mathcal{H}_{in} . \square

Historical note : That $\mathcal{K} = 0$ for principal null geodesics has been pointed out by Jiří Bičák and Zdeněk Stuchlík in 1976 [59].

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 \mathcal{K} [Eq. (11.21)]. In the expressions of each of these integrals,

⁵Indeed, we deduce from Eqs. (10.49) and (10.61) that, for $r \rightarrow +\infty$, $\mathbf{k} \sim \partial_t - \partial_r$ and $\mathbf{\ell} \sim (\partial_t + \partial_r)/2$.

⁶This is graphically confirmed by Figs. 10.6 and 10.7, which show that \mathcal{H} and \mathcal{H}_{in} are the only locations where a principal null geodesic can have $r = \text{const.}$

p^α has to be thought of as the first order derivative $dx^\alpha/d\lambda$ [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 \mathcal{K} , are quadratic. Furthermore, Eqs. (11.10) and (11.11) constitute a decoupled subsystem for (p^t, p^φ) , which can easily be solved⁷, yielding

$$\rho^2 p^t = \frac{1}{\Delta} [(r^2 + a^2)^2 E - 2amrL] - a^2 E \sin^2 \theta \quad (11.26)$$

$$\rho^2 p^\varphi = \frac{L}{\sin^2 \theta} + \frac{a}{\Delta} (2mrE - aL). \quad (11.27)$$

Besides, Eq. (11.21) involves only p^r and can be recast as

$$\rho^4 (p^r)^2 = R(r), \quad (11.28)$$

with

$$R(r) := [(r^2 + a^2)E - aL]^2 - \Delta(r^2 \mu^2 + \mathcal{K}). \quad (11.29)$$

Recall that Δ is the function of r given by Eq. (10.10): $\Delta := r^2 - 2mr + a^2$. All other quantities appearing in Eq. (11.29) are constant. Accordingly, $R(r)$ is a 4th order polynomial in r . Equation (11.28) implies that this polynomial must be non-negative along the geodesic \mathcal{L} :

$$R(r) \geq 0. \quad (11.30)$$

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,

$$\rho^4 (p^\theta)^2 = \Theta(\theta), \quad (11.31)$$

with

$$\Theta(\theta) := \mathcal{K} - \left(\frac{L}{\sin \theta} - aE \sin \theta \right)^2 - \mu^2 a^2 \cos^2 \theta. \quad (11.32)$$

Equation. (11.31) imposes that $\Theta(\theta)$ is non-negative along the geodesic \mathcal{L} :

$$\Theta(\theta) \geq 0. \quad (11.33)$$

The following constant is often used instead of \mathcal{K} :

$$Q := \mathcal{K} - (L - aE)^2 \quad (11.34)$$

Thanks to it, we may rewrite (11.32) as

$$\Theta(\theta) = Q + \cos^2 \theta \left[a^2(E^2 - \mu^2) - \frac{L^2}{\sin^2 \theta} \right]. \quad (11.35)$$

⁷An 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 $(r^2 + a^2)E - aL = \Delta(p^t - a \sin^2 \theta p^\varphi)$.

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* \mathcal{K} . As we shall see in Sec. 11.3.7, Q is well adapted to the description of the θ -motion of geodesics. On the other hand, a nice property of \mathcal{K} , which is not shared by Q , is to be always non-negative, as Eqs. (11.32) and (11.33) show:

$$\boxed{\mathcal{K} \geq 0}. \quad (11.36)$$

If the particle \mathcal{P} reaches the asymptotic region $r \gg m$, we deduce from Eqs. (11.34) and (11.23) the following behavior of Q :

$$Q \underset{r \rightarrow +\infty}{\sim} \mathbf{L}_{\text{tot}} \cdot \mathbf{L}_{\text{tot}} - L^2 + aE(2L - aE). \quad (11.37)$$

Hence, if $a = 0$, Q can be interpreted as the square of the part of \mathcal{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 \mathcal{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

$$Q = -a^2 E^2 \cos^4 \theta, \quad (11.38)$$

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 \mathcal{H} and \mathcal{H}_{in} , where it reduces to $Q = 0$, for $E = 0$ there [Eq. (11.9a)]. Equation (11.38) implies

$$Q \leq 0, \quad (11.39)$$

with $Q = 0$ only for principal null geodesics lying in the equatorial plane or for the outgoing principal null geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ and $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}_{\text{in}}}$ generating the horizons \mathcal{H} and \mathcal{H}_{in} .

Remark 3: As for \mathcal{K} , one may derive the Carter constant Q from a Killing tensor. Indeed, from the Walker-Penrose Killing tensor \mathbf{K} [Eq. (11.16)], let us form the tensor field

$$\tilde{\mathbf{K}} := \mathbf{K} - \tilde{\boldsymbol{\eta}} \otimes \tilde{\boldsymbol{\eta}}, \quad \text{where} \quad \tilde{\boldsymbol{\eta}} := \boldsymbol{\eta} + a\boldsymbol{\xi}. \quad (11.40)$$

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 $\tilde{\boldsymbol{\eta}} \otimes \tilde{\boldsymbol{\eta}}$ is a Killing tensor (cf. Example 3 in Sec. B.5.2), so that $\tilde{\mathbf{K}}$ a Killing tensor, the Killing equation (11.18) being linear. Applying $\tilde{\mathbf{K}}$ to the 4-momentum \mathbf{p} , we get

$$\tilde{\mathbf{K}}(\mathbf{p}, \mathbf{p}) = \underbrace{\mathbf{K}(\mathbf{p}, \mathbf{p})}_{\mathcal{K}} - (\underbrace{\langle \boldsymbol{\eta}, \mathbf{p} \rangle}_L + a \underbrace{\langle \boldsymbol{\xi}, \mathbf{p} \rangle}_{-E})^2.$$

Comparing with the definition (11.34), we conclude that

$$Q = \tilde{\mathbf{K}}(\mathbf{p}, \mathbf{p}). \quad (11.41)$$

The Boyer-Lindquist components of the contravariant tensor associated to $\tilde{\mathbf{K}}$ by metric duality have an expression particularly simple in terms of those of the inverse metric (cf. the notebook D.5.7):

$$\tilde{K}^{\alpha\beta} = -a^2 \cos^2 \theta g^{\alpha\beta} + \text{diag}(-a^2 \cos^2 \theta, 0, 1, \tan^{-2} \theta)^{\alpha\beta}. \quad (11.42)$$

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

$$\rho^2 \frac{dt}{d\lambda} = \frac{1}{\Delta} [(r^2 + a^2)^2 E - 2amrL] - a^2 E \sin^2 \theta \quad (11.43a)$$

$$\rho^2 \frac{dr}{d\lambda} = \epsilon_r \sqrt{R(r)} \quad (11.43b)$$

$$\rho^2 \frac{d\theta}{d\lambda} = \epsilon_\theta \sqrt{\Theta(\theta)} \quad (11.43c)$$

$$\rho^2 \frac{d\varphi}{d\lambda} = \frac{L}{\sin^2 \theta} + \frac{a}{\Delta} (2mrE - aL), \quad (11.43d)$$

where $\epsilon_r := \operatorname{sgn} p^r = \pm 1$, $\epsilon_\theta := \operatorname{sgn} p^\theta = \pm 1$ and the functions $R(r)$ and $\Theta(\theta)$ are defined by Eq. (11.29) and Eq. (11.32) or (11.35). Since $p^r = dr/d\lambda$, ϵ_r is +1 (resp. -1) in the parts of the geodesic \mathcal{L} where r increases (resp. decreases) with λ . Similarly, ϵ_θ is +1 (resp. -1) in the parts of \mathcal{L} where θ increases (resp. decreases) with λ .

We may rewrite Eq. (11.29) for $R(r)$ in terms of Q instead of \mathcal{K} , via Eq. (11.34):

$$R(r) := [(r^2 + a^2)E - aL]^2 - \Delta [r^2 \mu^2 + Q + (L - aE)^2]. \quad (11.44)$$

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 \mathcal{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 \mathcal{L} be a geodesic that is not stuck at some constant value of the coordinate r . We define a *r-turning point* of \mathcal{L} as a point $p_0 \in \mathcal{L}$, the r -coordinate r_0 of which obeys

$$R(r_0) = 0 \quad \text{and} \quad R'(r_0) \neq 0, \quad (11.45)$$

i.e. r_0 is a simple root of the polynomial R .

We have then

$$\left. \frac{dr}{d\lambda} \right|_{\lambda_0} = 0 \quad \text{and} \quad \left. \frac{d^2r}{d\lambda^2} \right|_{\lambda_0} = \frac{R'(r_0)}{2\rho_0^4} \neq 0, \quad (11.46)$$

where λ_0 is the value of the affine parameter λ at p_0 and $\rho_0 := \rho(p_0)$.

Proof. The vanishing of $dr/d\lambda$ at λ_0 follows immediately from Eq. (11.43b) with $R(r_0) = 0$, since ρ^2 never vanishes on \mathcal{M} . Besides, by taking the derivative of Eq. (11.43b) with respect to λ , we get

$$2 \left(r \frac{dr}{d\lambda} + a^2 \cos \theta \sin \theta \frac{d\theta}{d\lambda} \right) \frac{dr}{d\lambda} + \rho^2 \frac{d^2r}{d\lambda^2} = \epsilon_r \frac{R'(r)}{2\sqrt{R(r)}} \frac{dr}{d\lambda} = \frac{R'(r)}{2\rho^2}.$$

At $\lambda = \lambda_0$, the first term in left-hand side vanishes identically, due to the $dr/d\lambda$ factor, and we get the second part of (11.46). \square

We deduce from the result (11.46) that at $\lambda = \lambda_0$, $dr/d\lambda$ 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 $dr/d\lambda = 0$ at λ_0 , the condition $d^2r/d\lambda^2 \neq 0$ at λ_0 is sufficient but not necessary for r to change its direction of variation there. Indeed, the same property holds with $d^2r/d\lambda^2 = 0$ and higher order derivatives vanishing up to some even order k for which $d^k r/d\lambda^k \neq 0$. However, in the present case, $d^2r/d\lambda^2 = 0$ would imply $R'(r_0) = 0$ and we shall see Sec. 11.3.6 that a geodesic can reach such a point only asymptotically, i.e. for $\lambda \rightarrow +\infty$. Hence it cannot be a turning point.

Similarly, if \mathcal{L} is a geodesic that is not stuck at some constant value of the coordinate θ , we define a θ -**turning point** of \mathcal{L} as a point $p_0 \in \mathcal{L}$, the θ -coordinate θ_0 of which obeys

$$\Theta(\theta_0) = 0 \quad \text{and} \quad \Theta'(\theta_0) \neq 0. \quad (11.47)$$

We deduce then from the equation of motion (11.43c):

$$\left. \frac{d\theta}{d\lambda} \right|_{\lambda_0} = 0 \quad \text{and} \quad \left. \frac{d^2\theta}{d\lambda^2} \right|_{\lambda_0} = \frac{\Theta'(\theta_0)}{2\rho_0^4} \neq 0. \quad (11.48)$$

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 $d\theta/d\lambda = 0$ for some finite value of λ cannot have $d^2\theta/d\lambda^2 = 0$ at the same point, since we shall see in Sec. 11.3.6 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 \mathcal{L} such that

$$d\lambda' = \frac{d\lambda}{\rho^2} = \frac{d\lambda}{r(\lambda)^2 + a^2 \cos^2 \theta(\lambda)}. \quad (11.49)$$

Since ρ^2 never vanishes on the spacetime manifold \mathcal{M} (by construction of the latter, cf. Eq. (10.29)), the above relation leads to a well-defined parameter⁸ along \mathcal{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 \mathcal{L} in general⁹, contrary to λ . The parameter λ' is called **Mino parameter** [367].

⁸Note however that λ' may blow up if \mathcal{L} comes arbitrarily close to the ring singularity, i.e. if $\rho \rightarrow 0$.

⁹The 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' .

In terms of Mino parameter, the system (11.43) becomes

$$\frac{dt}{d\lambda'} = \frac{1}{\Delta} [(r^2 + a^2)^2 E - 2amrL] - a^2 E \sin^2 \theta \quad (11.50a)$$

$$\frac{dr}{d\lambda'} = \epsilon_r \sqrt{R(r)} \quad (11.50b)$$

$$\frac{d\theta}{d\lambda'} = \epsilon_\theta \sqrt{\Theta(\theta)} \quad (11.50c)$$

$$\frac{d\varphi}{d\lambda'} = \frac{L}{\sin^2 \theta} + \frac{a}{\Delta} (2mrE - aL), \quad (11.50d)$$

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 \mathcal{L} where $R(r) \neq 0$, this yields

$$\lambda' - \lambda'_0 = \int_{r_0}^r \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}}, \quad (11.51)$$

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 \mathcal{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 to a r -turning point of \mathcal{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 as¹⁰

$$\frac{1}{\sqrt{-R'(r_1)}} \int_{r_0}^{r_1} \frac{d\bar{r}}{\sqrt{r_1 - \bar{r}}},$$

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, \dots, 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

$$\lambda' - \lambda'_0 = \int_{r_0}^{r_1} \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}} + (M-1) \int_{r_1}^{r_2} \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}} + \int_{r_{1,2}}^r \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}}, \quad (11.52)$$

¹⁰We assume here $r_0 < r_1$, so that the constraint $R(\bar{r}) \geq 0$ on the interval $[r_0, r_1]$ implies $R'(r_1) < 0$.

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 abridged by means of a slash notation:

$$\lambda' - \lambda'_0 = \int_{r_0}^r \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}}. \quad (11.53)$$

Similarly, if $\Theta(\theta) \neq 0$ along \mathcal{L} , except possibly at some θ -turning points, Eq. (11.50c) can be integrated as

$$\begin{aligned} \lambda' - \lambda'_0 &= \int_{\theta_0}^{\theta} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} \\ &= \begin{cases} \int_{\theta_0}^{\theta} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} & \text{if } N = 0 \\ \int_{\theta_0}^{\theta_1} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} + (N-1) \int_{\theta_1}^{\theta_2} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} + \int_{\theta_{1,2}}^{\theta} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} & \text{if } N \geq 1, \end{cases} \end{aligned} \quad (11.54)$$

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:

Property 11.4: solution for geodesic motion

Let \mathcal{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. (11.95) below), and the associated function $\Theta(\theta)$, as defined by Eq. (11.35), do not vanish along \mathcal{L} except possibly at some r -turning points or θ -turning points. Let λ be the affine parameter of \mathcal{L} associated with the 4-momentum p and λ' the Mino parameter. If for $\lambda = \lambda_0$ \mathcal{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 \mathcal{L} obey

$$\lambda' - \lambda'_0 = \int_{r_0}^r \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}} = \int_{\theta_0}^{\theta} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} \quad (11.55a)$$

$$\lambda - \lambda_0 = \int_{r_0}^r \frac{\epsilon_r \bar{r}^2 d\bar{r}}{\sqrt{R(\bar{r})}} + a^2 \int_{\theta_0}^{\theta} \frac{\epsilon_\theta \cos^2 \bar{\theta} d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} \quad (11.55b)$$

$$t - t_0 = \int_{r_0}^r \frac{(\bar{r}^2 + a^2)^2 E - 2am\bar{r}L}{\bar{r}^2 - 2m\bar{r} + a^2} \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}} - a^2 E \int_{\theta_0}^{\theta} \sin^2 \bar{\theta} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} \quad (11.55c)$$

$$\varphi - \varphi_0 = a \int_{r_0}^r \frac{2m\bar{r}E - aL}{\bar{r}^2 - 2m\bar{r} + a^2} \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}} + L \int_{\theta_0}^{\theta} \frac{1}{\sin^2 \bar{\theta}} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}}. \quad (11.55d)$$

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

$$d\lambda = r^2 d\lambda' + a^2 \cos^2 \theta d\lambda'$$

and to substitute $d\lambda'$ by $\epsilon_r dr / \sqrt{R(r)}$ in the first term [cf. Eq. (11.50b)] and by $\epsilon_\theta d\theta / \sqrt{\Theta(\theta)}$ in the second term [cf. Eq. (11.50c)]. Similarly, by rewriting Eq. (11.50a) as

$$dt = \frac{(r^2 + a^2)^2 E - 2amrL}{r^2 - 2mr + a^2} d\lambda' - a^2 E \sin^2 \theta d\lambda'$$

and performing the same substitutions for $d\lambda'$ as above, we get Eq. (11.55c). Finally, Eq. (11.55d) is deduced in the same fashion from Eq. (11.50d). \square

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. 250) 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

$$\int_{\theta_0}^{\theta} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}} = - \int_{\zeta_0}^{\zeta} \frac{\epsilon_\theta d\bar{\zeta}}{\sqrt{Q(1 - \bar{\zeta}^2) + \bar{\zeta}^2[a^2(E^2 - \mu^2)(1 - \bar{\zeta}^2) - L^2]}},$$

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. [205] for bound timelike geodesics and to Ref. [235] for null geodesics.

Historical note : The constant of motion \mathcal{K} has been discovered by Brandon Carter in 1968 [89], in a study about Kerr-Newman spacetimes, which generalize the Kerr ones to nonzero global electric charge. Carter actually did not get \mathcal{K} from the Killing tensor \mathbf{K} , which was discovered only two years later

by Martin Walker and Roger Penrose [507]; 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 case, i.e. Kerr) and discovered that the Hamilton-Jacobi equation is separable, i.e. can be solved by separation of variables. \mathcal{K} appeared then as a separation constant. Note that Carter used the Kerr coordinates¹¹ $(v, r, \theta, \tilde{\varphi})$ 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 $dr/d\lambda$ and $d\theta/d\lambda$, 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 [521], 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 [89]: Eq. (11.55a) (without the $\lambda' - \lambda'_0$ part) is Eq. (58) of Ref. [89] and Eq. (11.55b) is Eq. (59) of Ref. [89]. 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 Main properties of geodesics

11.3.1 Sign of E

We have:

Property 11.5: positivity of the conserved energy outside the ergoregion

If a null or timelike geodesic \mathcal{L} has some part lying in the exterior of the ergoregion \mathcal{G} (cf. Sec. 10.2.4), then the conserved energy E defined by Eq. (11.2a) is necessarily positive:

$$\mathcal{L} \not\subset \overline{\mathcal{G}} \implies E > 0. \quad (11.56)$$

Proof. By the very definition of the ergoregion \mathcal{G} (cf. Sec. 10.2.4), the Killing vector ξ is timelike in the exterior of \mathcal{G} , i.e. in $\mathcal{M} \setminus \overline{\mathcal{G}}$. Moreover, it is future-directed there, given the time orientation defined in Sec. 10.3.2. The 4-momentum p is either timelike or null and always future-directed. By Eq. (1.5a) in Lemma 1.1, one has then necessarily $\xi \cdot p < 0$; hence $E := -\xi \cdot p > 0$ in $\mathcal{M} \setminus \overline{\mathcal{G}}$. Since E is constant along \mathcal{L} , it follows that $E > 0$ everywhere. \square

In particular, any timelike or null geodesic that reaches one of the asymptotic regions $r \rightarrow \pm\infty$ has $E > 0$.

Remark 1: Property 11.5 generalizes Property 7.2 obtained for Schwarzschild spacetime to the case $a \neq 0$. Indeed, for Schwarzschild spacetime, the exterior of the ergoregion is nothing but the exterior of the black hole region.

Inside the ergoregion, the Killing vector ξ 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

¹¹Carter's u is our v .

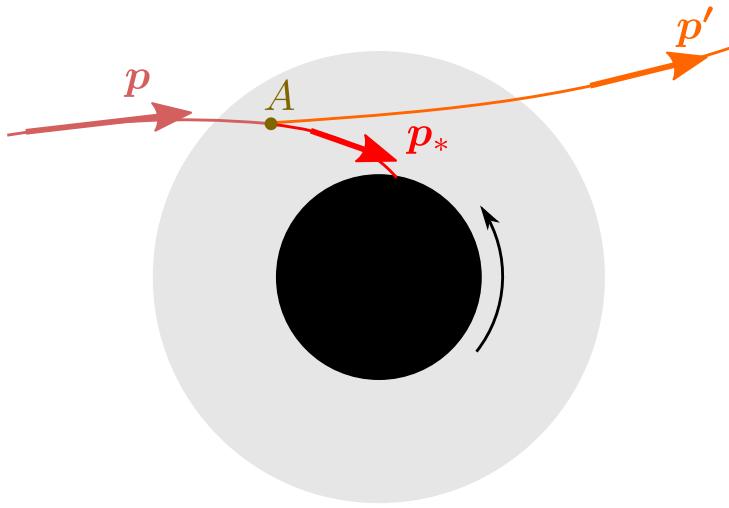


Figure 11.1: Projection in the equatorial plane $t = \text{const}$ and $\theta = \pi/2$ of the worldline of a particle and its 4-momentum p , which decay at event A in two particles: one with 4-momentum p' , which leaves to infinity, and one with 4-momentum p_* , which falls into the black hole (black region). The grey zone is the outer ergoregion.

zero-energy particle). Note that according to Property 11.5, neither a negative-energy particle nor a zero-energy one can exist outside the ergoregion.

Example 6 (Generators of the event and inner horizons): 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 \mathcal{H} and \mathcal{H}_{in} [cf. Eq. (11.4)].

Example 7 (Outgoing principal null geodesics in \mathcal{M}_{II}): We have seen in Example 3 (p. 374) that in the region \mathcal{M}_{II} , the outgoing principal null geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$ have $E < 0$ [Eq. (11.9c)]. This is consistent with Property 11.5 because \mathcal{M}_{II} is entirely contained in the ergoregion \mathcal{G} (cf. Fig. 10.2).

Remark 2: It should be stressed that, for a particle lying in the ergoregion, the quantity E is *not* the particle's energy measured by some physical observer, as given by formula (1.23). Indeed, since ξ is a spacelike vector in the ergoregion, it cannot be identified with any observer 4-velocity. In other words, 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 such particles.

11.3.2 The Penrose process

The possibility of having negative-energy particles in the outer ergoregion opens the path to a mechanism of energy extraction from a rotating black hole. To see this, consider a particle \mathcal{P} in free fall from infinity into the outer ergoregion \mathcal{G}^+ . At some point $A \in \mathcal{G}^+$, suppose that \mathcal{P} splits (or decays) into two particles: \mathcal{P}' and \mathcal{P}_* , such that \mathcal{P}' subsequently exits the ergoregion and leaves to infinity, while \mathcal{P}_* falls into the black hole (cf. Fig. 11.1). The particles \mathcal{P} , \mathcal{P}' , and \mathcal{P}_* can be either massive or massless and we assume that, once created, they are subject only to gravitation, so that their worldlines are timelike or null geodesics. Let p , p' and

\mathbf{p}_* be the 4-momentum vectors of respectively \mathcal{P} , \mathcal{P}' and \mathcal{P}_* . We define the *energy gain* in the above process by

$$\Delta E := E_{\text{out}} - E_{\text{in}}, \quad (11.57)$$

where E_{in} (resp. E_{out}) is the energy of \mathcal{P} (resp. \mathcal{P}') as measured by an asymptotic inertial observer at rest with respect to the black hole (cf. Sec. 10.7.5). E_{in} and E_{out} are actually nothing but the conserved energies of particles \mathcal{P} and \mathcal{P}' as defined in Sec. 11.2.2. They are thus given by Eq. (11.2a):

$$E_{\text{in}} = -\xi \cdot \mathbf{p}|_{\infty} \quad \text{and} \quad E_{\text{out}} = -\xi \cdot \mathbf{p}'|_{\infty}, \quad (11.58)$$

where the subscript ∞ refers to the location of the asymptotic inertial observer. Since E_{in} and E_{out} are constant along the geodesic worldlines of \mathcal{P} and \mathcal{P}' (Property 11.1), we have $E_{\text{in}} = -\xi \cdot \mathbf{p}|_A$ and $E_{\text{out}} = -\xi \cdot \mathbf{p}'|_A$. Accordingly, Eqs. (11.57) and (11.58) lead to

$$\Delta E = \xi \cdot \mathbf{p}|_A - \xi \cdot \mathbf{p}'|_A = \xi|_A \cdot (\mathbf{p}|_A - \mathbf{p}'|_A).$$

Now, the conservation of energy-momentum at event A implies $\mathbf{p}|_A = \mathbf{p}'|_A + \mathbf{p}_*|_A$, hence

$$\Delta E = \xi \cdot \mathbf{p}_*|_A = -E_*, \quad (11.59)$$

where E_* is the conserved energy of particle \mathcal{P}_* . If \mathcal{P}_* happens to be a negative-energy particle, which is allowed since \mathcal{P}_* is confined to the ergoregion, then $E_* < 0$ and Eq. (11.59) implies $\Delta E > 0$, i.e. the outgoing particle has more energy than the ingoing one. The above process is then called a **Penrose process**. It is particularly relevant to astrophysics, via some electromagnetic variants of it, notably the so-called *Blandford-Znajek mechanism* (see e.g. Ref. [338] for an extended discussion and Box. 15.1 of Ref. [256] for an elementary presentation of the Blandford-Znajek mechanism).

Remark 3: No peculiar property of the Kerr spacetime, except for the existence of an outer ergoregion, has been invoked in the above reasoning. Consequently, the Penrose process can occur for any stationary rotating body (not even necessarily a black hole) that is endowed with an ergoregion in which a round trip from infinity is possible.

Since the particle \mathcal{P}_* is assumed to fall into the black hole, it must cross the event horizon \mathcal{H} at some point, B say. At B , the Killing vector $\chi := \xi + \Omega_{\mathcal{H}} \eta$ of Kerr spacetime [Eq. (10.68)] is null (being a null normal to \mathcal{H} , cf. Eq. (10.72)) and future-directed (Property 10.15). On the other side, the 4-momentum \mathbf{p}_* is necessarily a future-directed timelike or null vector. Moreover, \mathbf{p}_* is transverse to \mathcal{H} at B (for \mathcal{P}_* is crossing \mathcal{H}) and therefore cannot be collinear to χ , which is tangent to \mathcal{H} . We may then invoke Lemma 1.2 (case (1.6a)) to assert that $\chi \cdot \mathbf{p}_*|_B < 0$, or equivalently [thanks to Eq. (10.68)],

$$\underbrace{\xi \cdot \mathbf{p}_*|_B}_{-E_*} + \Omega_{\mathcal{H}} \underbrace{\eta \cdot \mathbf{p}_*|_B}_{L_*} < 0,$$

where L_* is the conserved angular momentum of \mathcal{P}_* [Eq. (11.2b)]. We thus get

$$E_* > \Omega_{\mathcal{H}} L_*. \quad (11.60)$$

Given that $\Omega_{\mathcal{H}} > 0$ [cf. Eqs. (10.69) and (10.1)], we deduce that if a Penrose process occurs, i.e. if $E_* < 0$, then necessarily $L_* < 0$. Similarly to the energy gain (11.57), we may define the angular momentum gain of the process $\mathcal{P} \rightarrow \mathcal{P}' + \mathcal{P}_*$ by

$$\Delta L := L_{\text{out}} - L_{\text{in}}, \quad (11.61)$$

where L_{in} (resp. L_{out}) is the angular momentum of \mathcal{P} (resp. \mathcal{P}') as measured by an asymptotic inertial observer, i.e. the conserved angular momenta of particles \mathcal{P} and \mathcal{P}' as defined in Sec. 11.2.2: $L_{\text{in}} = \boldsymbol{\eta} \cdot \mathbf{p}|_\infty$ and $L_{\text{out}} = \boldsymbol{\eta} \cdot \mathbf{p}'|_\infty$ [Eq. (11.2b)]. The same computation as above for ΔE , except for ξ replaced by $\boldsymbol{\eta}$, shows that the energy-momentum conservation law $\mathbf{p}|_A = \mathbf{p}'|_A + \mathbf{p}_*|_A$ leads to a result similar to (11.59):

$$\Delta L = -\boldsymbol{\eta} \cdot \mathbf{p}_*|_A = -L_*. \quad (11.62)$$

Along with Eq. (11.59), this relation turns the inequality (11.60) into $\Omega_{\mathcal{H}}\Delta L > \Delta E$. This shows that any Penrose process yields $\Delta L > 0$ in addition to $\Delta E > 0$. Let us summarize:

Property 11.6: Penrose process

The Penrose process as described above and illustrated in Fig. 11.1 extracts both some energy $\Delta E > 0$ and some angular momentum $\Delta L > 0$ from a rotating black hole, in such a way that

$$\Omega_{\mathcal{H}}\Delta L > \Delta E > 0, \quad (11.63)$$

where $\Omega_{\mathcal{H}}$ is the black hole rotation velocity [cf. Eq. (10.69)]. The secondary particle \mathcal{P}_* falling into the black hole during the process has both a negative conserved energy E_* and a negative conserved angular momentum L_* , which obey Eq. (11.60): $\Omega_{\mathcal{H}}L_* < E_* < 0$.

Remark 4: Setting $\Omega_{\mathcal{H}} = 0$ in Eq. (11.63) shows that the Penrose process is not possible for a non-rotating (Schwarzschild) black hole. This is not surprising since there is no outer ergoregion for such a black hole, the ergosphere coinciding with the event horizon (cf. Remark 5 on p. 329).

Ultimately, the energy and angular momentum extracted via the Penrose process are taken from the mass m and the angular momentum $J = am$ of the black hole, so that these two quantities must decrease slightly (i.e. in a infinitely small manner, given that \mathcal{P} , \mathcal{P}' and \mathcal{P}_* are assumed to be “test particles” (geodesic worldlines), which implies $\Delta E \ll m$ and $\Delta L \ll J$). This will be discussed in details in Sec. 16.3.1.

The Penrose process can be extended from particles to fields; it is then known as the phenomenon of **superradiance**, also called **superradiant scattering**. This occurs when the scattering of a (scalar, vector or tensor) field by a rotating black hole results in a outgoing field energy larger than the ingoing one. The energy of the field is defined from its energy-momentum tensor \mathbf{T} by constructing the “energy density” vector \mathbf{P} as $P^\alpha := -T^\alpha_\mu \xi^\mu$ (compare with Eq. (11.2a)). The energy of the field through a hypersurface Σ is then $E_\Sigma := \int_\Sigma \star \underline{\mathbf{P}}$, where $\star \underline{\mathbf{P}}$ is the the Hodge dual of the 1-form $\underline{\mathbf{P}}$, i.e. the 3-form defined by $(\star \underline{\mathbf{P}})_{\alpha\beta\gamma} := P^\mu \varepsilon_{\mu\alpha\beta\gamma}$ [cf. Eq. (5.38) with $n = 4$ and $p = 1$]. Now, by combining the energy-momentum conservation law $\nabla_\mu T^\mu_\alpha = 0$ [Eq. (1.45)] and the Killing equation for ξ , it is easy to show that \mathbf{P} is a

conserved current: $\nabla_\mu P^\mu = 0$. This is equivalent to the 3-form $\star \underline{P}$ being closed: $d\star \underline{P} = 0$, since $d\star \underline{P} = (\nabla_\mu P^\mu) \varepsilon$. The Stokes theorem (A.94) leads then to $E_\Sigma = 0$ as soon as Σ is a closed oriented hypersurface. Let us choose $\Sigma = \Sigma_1 \cup \Delta \mathcal{H} \cup \Sigma_2 \cup \Sigma_{\text{ext}}$, where (i) Σ_1 and Σ_2 are two spacelike hypersurfaces intersecting the event horizon \mathcal{H} and such that Σ_2 lies in the future of Σ_1 ¹², (ii) $\Delta \mathcal{H}$ is the portion of \mathcal{H} between $\mathcal{H} \cap \Sigma_1$ and $\mathcal{H} \cap \Sigma_2$ and (iii) Σ_{ext} is a timelike hypersurface connecting Σ_1 and Σ_2 far from \mathcal{H} . We have then $E_\Sigma = E_{\Sigma_1} + E_{\Delta \mathcal{H}} + E_{\Sigma_2} + E_{\Sigma_{\text{ext}}} = 0$. Assuming that T decays sufficiently fast so that $E_{\Sigma_{\text{ext}}} = 0$, we arrive at

$$E_2 = E_1 - E_*, \quad (11.64)$$

where $E_2 := E_{\Sigma_2}$, $E_1 := -E_{\Sigma_1}$ (the change of sign occurs because the orientation of Σ_1 inherited from that of Σ is the opposite of that induced by the future-directed unit normal n to Σ_1) and $E_* := E_{\Delta \mathcal{H}} = \int_{\Delta \mathcal{H}} \star \underline{P}$. Viewing the spacelike hypersurfaces Σ_1 and Σ_2 as “time slices”, we may say that E_1 (resp. E_2) is the field energy at “time Σ_1 ” (resp. “time Σ_2 ”) ¹³, while E_* is the field energy captured by the black hole. Superradiance occurs iff $E_2 > E_1$, i.e. iff $E_* < 0$, as in the Penrose process. For a scalar field of the form $\Phi = \Phi_0(r, \theta) e^{i(\omega t - m\varphi)}$ obeying the Klein-Gordon equation, it can be shown (by a direct computation of $E_* = \int_{\Delta \mathcal{H}} \star \underline{P}$) that $E_* < 0$ is equivalent to [98, 498, 338]

$$0 < \omega < m\Omega_{\mathcal{H}}. \quad (11.65)$$

Actually the same superradiance condition holds for the Fourier modes of electromagnetic waves [476, 98], as well as gravitational waves [476].

Historical note : The Penrose process has first been suggested by Roger Penrose in 1969, in the review article [407] (cf. the footnote 7 in this article); the detailed calculation has been presented subsequently in an article with Roger Floyd published in 1971 [408]. In particular, Penrose and Floyd have derived the inequality (11.60) [Eq. (4) in Ref. [408], taking into account that $a/(2mr_+) = \Omega_{\mathcal{H}}$, cf. Eq. (10.69)]. The extension to fields (superradiance) was introduced by Yakov Zeldovich in 1971 [526, 527] and explicit computations for a scalar field in the Kerr metric have been performed by his doctoral student Alexei Starobinsky in 1972 (published: 1973) [459], as well as by William Press and Saul Teukolsky at the same time [419]. It seems that the superradiance of a scalar field scattered by a Kerr black hole has also been discovered around 1971 by Charles Misner: mention of unpublished results by him can be found in his article [368], as well as in a 1973 article by Jacob Bekenstein [50].

11.3.3 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 $p = dx/d\lambda$ is oriented everywhere towards the future. The latter condition is equivalent to have the affine parameter λ increase towards the future along the geodesic \mathcal{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 \mathcal{L} is timelike.

¹²For instance, Σ_1 is a hypersurface of constant Kerr coordinate $\tilde{t} = \tilde{t}_1$ and Σ_2 is a hypersurface $\tilde{t} = \tilde{t}_2$, with $\tilde{t}_2 > \tilde{t}_1$.

¹³One can easily express E_1 and E_2 in terms of the normal volume element vector dV introduced by Eq. (5.48): $E_{1,2} = \int_{\Sigma_{1,2}} P_\mu dV^\mu$.

As discussed in Sec. 10.3.2, the time orientation of the Kerr spacetime $(\mathcal{M}, \mathbf{g})$ is set by the global null vector field \mathbf{k} generating the ingoing principal null geodesics, cf. Eq. (10.32). By virtue of Lemma 1.2 [cf. Eq. (1.6a)], we have then, for \mathbf{p} not collinear with \mathbf{k} (in particular for \mathcal{L} timelike),

$$\mathbf{p} \text{ future-directed} \iff \mathbf{g}(\mathbf{k}, \mathbf{p}) < 0 \iff k^\mu p_\mu < 0 \iff \frac{r^2 + a^2}{\Delta} p_t - p_r + \frac{a}{\Delta} p_\varphi < 0,$$

where the last inequality follows from the Boyer-Lindquist components (10.49) of \mathbf{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:

Property 11.7

$$\begin{aligned} \mathcal{L} \text{ timelike or} \\ \mathbf{k} \text{ not tangent to } \mathcal{L} \end{aligned} \implies \boxed{\frac{1}{\Delta} [(r^2 + a^2)E - aL + \epsilon_r \sqrt{R(r)}] > 0}. \quad (11.66)$$

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 $(\mathcal{M}, \mathbf{g})$ is the null tangent ℓ to the outgoing principal null geodesics, introduced in Sec. 10.4 [Eq. (10.57)]. If we apply Lemma 1.2 with ℓ instead of \mathbf{k} , we get, using the Boyer-Lindquist components (10.61) of ℓ ,

Property 11.8

$$\begin{aligned} \mathcal{L} \text{ timelike or} \\ \ell \text{ not tangent to } \mathcal{L} \end{aligned} \implies \boxed{(r^2 + a^2)E - aL - \epsilon_r \sqrt{R(r)} > 0}. \quad (11.67)$$

In $\mathcal{M}_I \cup \mathcal{M}_{III}$, $\Delta > 0$ and we can combine (11.66) and (11.67) 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 :

Property 11.9

$$\mathcal{L} \text{ not principal null geod.} \implies \boxed{(r^2 + a^2)E - aL > \sqrt{R(r)}} \quad \text{in } \mathcal{M}_I \cup \mathcal{M}_{III}. \quad (11.68)$$

A related constraint in $\mathcal{M}_I \cup \mathcal{M}_{III}$ 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 \mathbf{k} and ℓ [cf. Eq. (10.112)] and is a future-directed timelike vector in all $\mathcal{M}_I \cup \mathcal{M}_{III}$. We can then apply Lemma 1.1, using expression (10.108a) of $\boldsymbol{\varepsilon}_{(0)}$ in terms of $\boldsymbol{\xi} = \partial_t$ and $\boldsymbol{\eta} = \partial_\varphi$:

$$\mathbf{p} \text{ future-directed in } \mathcal{M}_I \cup \mathcal{M}_{III} \iff \mathbf{g}(\boldsymbol{\varepsilon}_{(0)}, \mathbf{p}) < 0 \iff (r^2 + a^2) \underbrace{\mathbf{g}(\boldsymbol{\xi}, \mathbf{p})}_{-E} + a \underbrace{\mathbf{g}(\boldsymbol{\eta}, \mathbf{p})}_L < 0.$$

Hence the constraint:

Property 11.10

$$(r^2 + a^2)E - aL > 0 \quad \text{in } \mathcal{M}_I \cup \mathcal{M}_{III}. \quad (11.69)$$

Note that if \mathcal{L} is not a principal null geodesic, (11.69) is a mere consequence of (11.68), given that $\sqrt{R(r)} \geq 0$.

An immediate corollary of Property 11.10 is:

Property 11.11

In the outer ergoregion $\mathcal{G}^+ = \mathcal{G} \cap \mathcal{M}_I$, a geodesic that has $E \leq 0$ must have $L < 0$.

Another constraint is obtained by considering the vector field N normal to the hypersurfaces of constant Boyer-Lindquist coordinate t , which has been introduced in Sec. 10.7.3: $N = -\vec{dt}$ [Eq. (10.88)]. As shown in Sec. 10.7.3, N is future-directed timelike in the subpart $\mathcal{M}_{ZAMO} := \mathcal{M}_I \cup (\mathcal{M}_{III} \setminus \mathcal{T})$ of Kerr spacetime. Lemma 1.1 [cf. Eq. (1.5a)] then yields

$$\mathbf{p} \text{ future-directed in } \mathcal{M}_{ZAMO} \iff g(N, \mathbf{p}) < 0 \iff -\underbrace{\langle dt, \mathbf{p} \rangle}_{p^t} < 0 \iff p^t > 0.$$

Hence the Boyer-Lindquist component p^t of the 4-momentum \mathbf{p} must be positive in \mathcal{M}_{ZAMO} :

Property 11.12

$$p^t > 0 \quad \text{in } \mathcal{M}_I \cup (\mathcal{M}_{III} \setminus \mathcal{T}). \quad (11.70)$$

Let us recall that \mathcal{T} is the Carter time machine (cf. Sec. 10.2.5), which occupies a very limited portion of \mathcal{M}_{III} . Since $p^t = dt/d\lambda$ and λ increases towards the future (cf. Sec. 11.2.1), Property 11.12 implies

Property 11.13

In the region $\mathcal{M}_I \cup (\mathcal{M}_{III} \setminus \mathcal{T})$ of Kerr spacetime, the Boyer-Lindquist coordinate t increases towards the future along any geodesic \mathcal{L} .

Using Eq. (11.43a) with $dt/d\lambda = p^t$ and the fact that $\Delta > 0$ in $\mathcal{M}_I \cup \mathcal{M}_{III}$, we can rewrite Property 11.12 as

Property 11.14

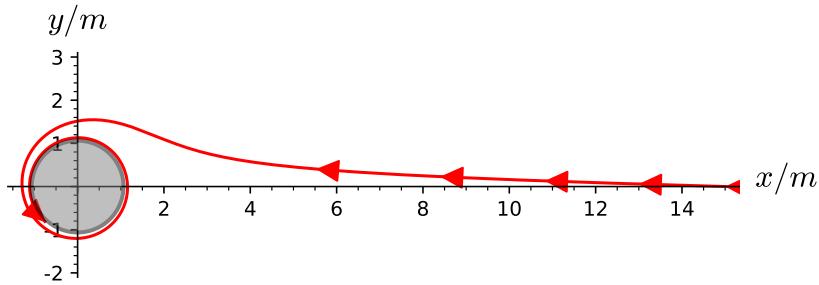


Figure 11.2: Trajectory in the equatorial plane of an incoming timelike geodesic with $E = \mu$, $L = 0$ and $Q = 0$, plunging into a Kerr black hole with $a = 0.998\text{ m}$. The figure is drawn in terms of the Cartesian Boyer-Lindquist coordinates (x, y) defined by Eq. (11.73) with $\theta = \pi/2$ (equatorial plane) and the grey disk marks the black hole region. [Figure generated by the notebook D.5.8]

$$\boxed{\rho^2(r^2 + a^2)E - 2amr(L - aE \sin^2 \theta) > 0} \text{ in } \mathcal{M}_I \cup (\mathcal{M}_{III} \setminus \mathcal{T}). \quad (11.71)$$

Remark 5: In the outer ergoregion $\mathcal{G}^+ = \mathcal{G} \cap \mathcal{M}_I$ (i.e. the part of the ergoregion outside of the black hole), $E \leq 0$ is allowed, but Property 11.12 shows that $p^t \leq 0$ is not allowed.

Historical note : That $E \leq 0$ in the outer ergoregion implies $L < 0$ (Property 11.11) has been shown for equatorial geodesics by George Contopoulos in 1984 [140].

11.3.4 Lense-Thirring effect

Let us consider a null or timelike geodesic \mathcal{L} with a vanishing angular momentum: $L = 0$. The equations of motion (11.43a) and (11.43d) reduce to

$$\begin{aligned} \rho^2 \frac{dt}{d\lambda} &= E \left(\frac{(r^2 + a^2)^2}{\Delta} - a^2 \sin^2 \theta \right) = \frac{E}{\Delta} [\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta] \\ \rho^2 \frac{d\varphi}{d\lambda} &= \frac{2amrE}{\Delta}, \end{aligned}$$

so that we get

$$\boxed{\left. \frac{d\varphi}{dt} \right|_{\mathcal{L}} = \frac{2amr}{\rho^2(r^2 + a^2) + 2a^2mr \sin^2 \theta}}_{L=0}. \quad (11.72)$$

Hence, for $a \neq 0$, a geodesic with zero angular momentum about the symmetry axis has nevertheless a rotational motion around that axis. Moreover, this motion is independent of the geodesic's characteristics, i.e. (μ, E, Q) . 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. 11.2 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. 11.3.8). The trajectory is initially radial, but as r decreases, $d\varphi/dt$ increases according to formula (11.72). In Fig. 11.2 and in all figures of this chapter, we are using the

Cartesian Boyer-Lindquist coordinates (t, x, y, z) , which are defined 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:

$$x := r \sin \theta \cos \varphi, \quad y := r \sin \theta \sin \varphi, \quad z := r \cos \theta. \quad (11.73)$$

We note that, at a given point (r, θ) , the angular velocity (11.72) coincides with that of the zero-angular momentum observer (ZAMO) at that point [compare Eq. (10.99)].

11.3.5 Winding near the event horizon and the inner horizon

Let us consider a null or timelike geodesic \mathcal{L} in the vicinity of the black hole event horizon \mathcal{H} . On \mathcal{H} , $r = r_+$ and $\Delta = 0$. Then, the term Δ^{-1} in Eq. (11.43a) makes $dt/d\lambda$ diverge as $r \rightarrow r_+$, except in the very special where $(r_+^2 + a^2)E - aL = 0$, which is equivalent to $E = \Omega_{\mathcal{H}}L$ according to Eq. (10.69). Similarly, $d\varphi/d\lambda$, as given by Eq. (11.43d), diverges as $r \rightarrow r_+$, except for $E = \Omega_{\mathcal{H}}L$. These two divergences are not a pathology of \mathcal{L} per se; they reflect merely the singularity of Boyer-Lindquist coordinates (t, r, θ, φ) at \mathcal{H} (cf. Sec. 10.2.6). However, we read on Eqs. (11.43a) and (11.43d) that the ratio $d\varphi/dt|_{\mathcal{L}} := d\varphi/d\lambda \times (dt/d\lambda)^{-1}$ converges to a finite value:

$$\lim_{r \rightarrow r_+} \frac{d\varphi}{dt} \Big|_{\mathcal{L}} = \frac{a}{r_+^2 + a^2} = \Omega_{\mathcal{H}}, \quad (11.74)$$

where the second equality follows from Eq. (10.69), making the black hole rotation velocity $\Omega_{\mathcal{H}}$ appear. Hence we conclude:

Property 11.15: behavior of φ near the event horizon

Any null or timelike geodesic that approaches the event horizon \mathcal{H} is winding around it in terms of the Boyer-Lindquist coordinates at exactly the black hole rotation velocity $\Omega_{\mathcal{H}}$.

Remark 6: For a timelike geodesic on a circular orbit, we have argued in Sec. 7.3.3 that $d\varphi/dt|_{\mathcal{L}}$ is the angular velocity of as seen by an asymptotic inertial observer. The reasoning was given in the Schwarzschild spacetime context but it actually involved only the spacetime symmetry by translation in t , so it is applicable here.

Remark 7: When $a \neq 0$, a geodesic that starts far from the black hole with $d\varphi/dt|_{\mathcal{L}} < 0$ [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 (11.74), where $\Omega_{\mathcal{H}}$ is positive. This is illustrated in Fig. 11.3 and is in sharp contrast with the Schwarzschild case, where φ is always a monotonic function of λ , as shown in Sec. 7.2.2.

Remark 8: The winding property does not hold for the Kerr or advanced Kerr coordinates. Indeed, we have seen in Sec. 10.4 that the ingoing principal null geodesics $\mathcal{L}_{(v, \theta, \tilde{\varphi})}^{\text{in}}$ are geodesics along which $\tilde{\varphi}$ is constant. They are therefore not winding in terms of either the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ nor the advanced Kerr ones $(v, r, \theta, \tilde{\varphi})$. This difference of (coordinate) behavior is understandable if one

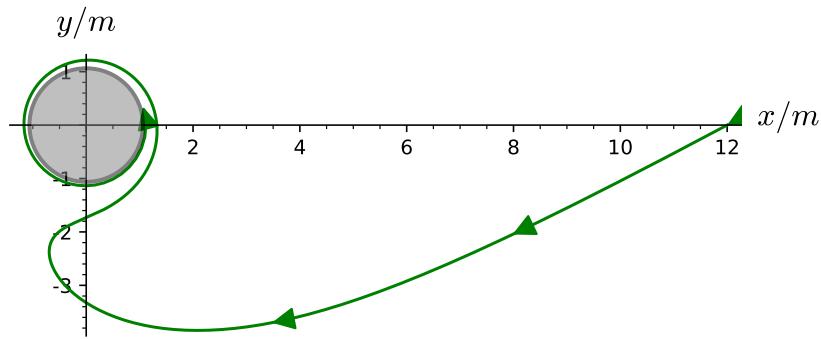


Figure 11.3: Trajectory in the equatorial plane of an incoming null geodesic with $L = -6E < 0$ and $Q = 0$, plunging into a Kerr black hole with $a = 0.998\text{ m}$. Note the turning point in φ and the final winding in the direction of the black hole rotation (counterclockwise in the figure). The figure is drawn in terms of the Cartesian Boyer-Lindquist coordinates (x, y) defined by Eq. (11.73) with $\theta = \pi/2$ (equatorial plane) and the grey disk marks the black hole region. [Figure generated by the notebook D.5.8]

considers the diverging behavior in Δ^{-1} of the relation $d\tilde{\varphi} = d\varphi + a/\Delta dr$ [Eq. (10.23b)] between the angular coordinates φ and $\tilde{\varphi}$.

Regarding the Boyer-Lindquist coordinate behavior in the vicinity of the inner horizon \mathcal{H}_{in} (corresponding to the second root r_- of Δ), we deduce from Eqs. (11.43a) and (11.43d) that

$$\lim_{r \rightarrow r_-} \frac{d\varphi}{dt} \Big|_{\mathcal{L}} = \frac{a}{r_-^2 + a^2} = \Omega_{\text{in}}, \quad (11.75)$$

where the second equality follows from Eq. (10.71); it involves the rotation velocity Ω_{in} of the inner horizon \mathcal{H}_{in} . Hence

Property 11.16: behavior of φ near the inner horizon

Any null or timelike geodesic that approaches the inner horizon \mathcal{H}_{in} is winding around it in terms of the Boyer-Lindquist coordinates at exactly the rotation velocity Ω_{in} of \mathcal{H}_{in} .

11.3.6 Asymptotic r -values and θ -values

Let \mathcal{L} be a geodesic and p_0 a point of \mathcal{L} at which r varies, i.e. such that $R(r_0) \neq 0$, where $r_0 = r(p_0)$. Let us assume that \mathcal{L} comes close to $r = r_1$, where r_1 is a double root of R , i.e. fulfills $R(r_1) = 0$, $R'(r_1) = 0$ 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

$$\frac{\sqrt{2}}{\sqrt{R''(r_1)}} \int_{r_0}^{r_1} \frac{d\bar{r}}{r_1 - \bar{r}},$$

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: $\lambda' \rightarrow +\infty$. Since $\rho^2 > 0$, provided that ρ^2 does not tend to 0 (the ring singularity) as $\lambda' \rightarrow +\infty$, this implies that the affine parameter λ tends to $+\infty$ as well, given the relation (11.49) between the two parameters. Hence we conclude:

Property 11.17: asymptotic r -value

Let \mathcal{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 \mathcal{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 \mathcal{L} , i.e. in the limit $\lambda \rightarrow \pm\infty$ of the affine parameter λ . We call r_1 an **asymptotic r -value** of the geodesic \mathcal{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. 382).

By the same reasoning, we have

Property 11.18: asymptotic θ -value

Let \mathcal{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 \mathcal{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 \mathcal{L} , i.e. in the limit $\lambda \rightarrow \pm\infty$ of the affine parameter λ . We call θ_1 an **asymptotic θ -value** of the geodesic \mathcal{L} .

11.3.7 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

$$\left(\frac{d\theta}{d\lambda'} \right)^2 + V(\theta) = Q, \quad (11.76)$$

with

$$V(\theta) := \cos^2 \theta \left[a^2(\mu^2 - E^2) + \frac{L^2}{\sin^2 \theta} \right]. \quad (11.77)$$

Note that $V(\theta)$ is related to $\Theta(\theta)$ by

$$\Theta(\theta) = Q - V(\theta). \quad (11.78)$$

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. 11.3.6) correspond to $V(\theta) = Q$ and $V'(\theta) = 0$.

We recognize in Eq. (11.76) 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. (11.76) playing the role of the constant “total energy”. Since $(d\theta/d\lambda')^2 \geq 0$, Eq. (11.76) implies

$$V(\theta) \leq Q. \quad (11.79)$$

Accordingly, given a plot of $V(\theta)$, as in Figs. 11.4 and 11.5, 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. 11.4 (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 plane¹⁴: $\theta_m := \arccos \sqrt{Q/(a^2(\mu^2 - E^2))}$ and $\pi - \theta_m$ (cf. the trajectory $Q = Q_1$ in Fig. 11.4, left).
 - $Q = a^2(\mu^2 - E^2)$: $\theta = 0$ and $\theta = \pi$ are either unstable positions or asymptotic θ -values (cf. Sec. 11.3.6), i.e. the geodesic reaches the rotation axis for $\lambda \rightarrow \pm\infty$.
 - $Q > a^2(\mu^2 - E^2)$: the range of θ is not limited (cf. the trajectory $Q = Q_2$ in Fig. 11.4, left) and each time the geodesic reaches the rotation axis ($\theta = 0$ or $\theta = \pi$), it crosses it, since the velocity $d\theta/d\lambda'$ 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 .
- **Case $a^2(E^2 - \mu^2) = 0 \iff a = 0$ or $|E| = \mu$:** $V(\theta) = 0$ and Eq.(11.76) reduces to $(d\theta/d\lambda')^2 = Q$. 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. 11.4 (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.

¹⁴The index m in θ_m stands for *minimal*.

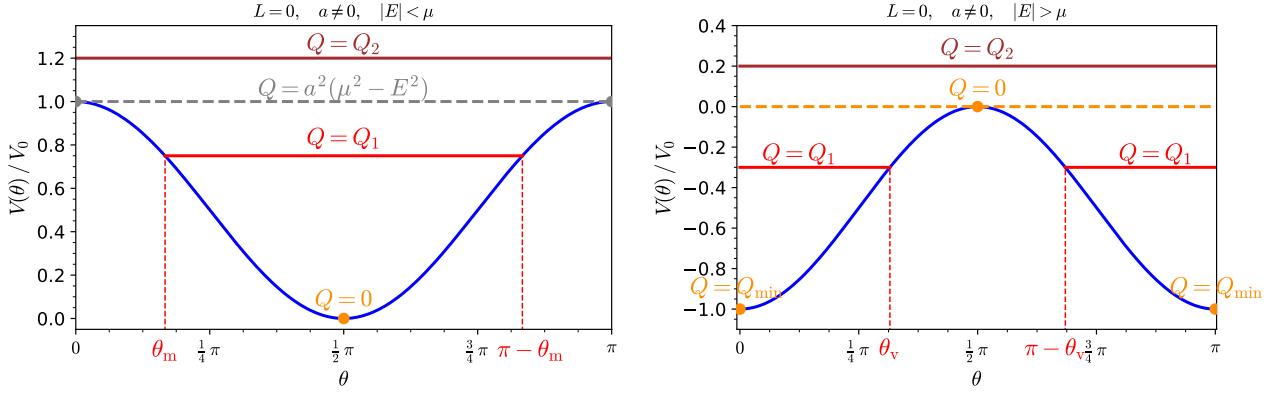


Figure 11.4: Effective θ -potential $V(\theta)$ in the case $L = 0$ and $a \neq 0$. $V(\theta)$ is plotted in units of $V_0 := a^2|\mu^2 - E^2|$. The left figure is for $|E| < \mu$, with colored dots or horizontal lines corresponding to geodesic trajectories for four values of the Carter constant Q : 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 : $Q_{\min} = -a^2(E^2 - \mu^2)$, $Q_1 = -0.3V_0$, 0 and $Q_2 = 0.2V_0$. Dashed lines indicate trajectories with asymptotic θ -values.

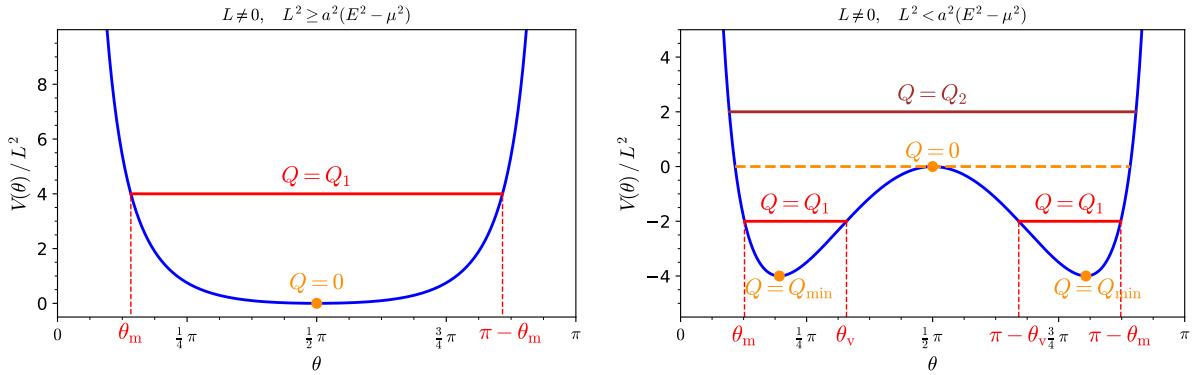


Figure 11.5: Effective θ -potential $V(\theta)$ in the case $L \neq 0$. The left figure is for $L^2 \geq a^2(E^2 - \mu^2)$, with colored dots or horizontal lines corresponding to geodesic trajectories for two values of the Carter constant Q : 0 and $Q_1 = 4L^2$. The right figure is for $L^2 < a^2(E^2 - \mu^2)$, with trajectories corresponding to four values of Q : Q_{\min} , $Q_1 = -2L^2$, 0 and $Q_2 = 2L^2$. The dashed line ($Q = 0$) indicates the trajectory with an asymptotic θ -value, which is $\pi/2$.

- $-a^2(E^2 - \mu^2) < Q < 0$: the geodesic oscillates about the rotation axis, without reaching the equator; one has either $\theta \in [0, \theta_v]$ or $\theta \in [\pi - \theta_v, \pi]$, with the turning point value¹⁵ $\theta_v := \arccos \sqrt{|Q|/(a^2(E^2 - \mu^2))} < \pi/2$ (cf. the trajectories $Q = Q_1$ in Fig. 11.4, 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 $\lambda \rightarrow \pm\infty$.
- $Q > 0$: 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. 11.4, right).

¹⁵The index v in θ_v stands for *vortical*, the definition of which is given at the end of this section.

Geodesics with $L \neq 0$

When $L \neq 0$, we see from expression (11.77) that $V(\theta) \rightarrow +\infty$ when $\theta \rightarrow 0$ or π . We conclude immediately that the geodesic cannot reach the rotation axis for $L \neq 0$. Moreover, we have

$$\frac{dV}{d\theta} = -2 \cos \theta \sin \theta \left[a^2(\mu^2 - E^2) + \frac{L^2}{\sin^4 \theta} \right].$$

The extrema of V in $(0, \pi)$ are then given by

$$\frac{dV}{d\theta} = 0 \iff \theta = \frac{\pi}{2} \quad \text{or} \quad \sin^4 \theta = \frac{L^2}{a^2(E^2 - \mu^2)}. \quad (11.80)$$

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 $V(\theta) \rightarrow +\infty$ at the boundaries of the interval $[0, \pi]$ (cf. left part of Fig. 11.5). 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 $\theta = \theta_m$ and $\pi - \theta_m$, θ_m being the solution of $V(\theta_m) = Q$ in $(0, \pi/2)$ (cf. the trajectory $Q = Q_1$ in Fig. 11.5, 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. 11.5), which are located at $\theta = \theta_*$, $\pi/2$ and $\pi - \theta_*$ with

$$\theta_* := \arcsin \sqrt{\frac{|L|}{a \sqrt{E^2 - \mu^2}}} \quad \text{and} \quad 0 < \theta_* < \frac{\pi}{2}. \quad (11.81)$$

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. (11.79)]:

$$Q_{\min} = - \left(a\sqrt{E^2 - \mu^2} - |L| \right)^2. \quad (11.82)$$

We have then (cf. right panel of Fig. 11.5)

- $Q = Q_{\min}$: the geodesic stays stably at a fixed value of θ , either θ_* or $\pi - \theta_*$.
- $Q_{\min} < Q < 0$: 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. 11.5, right).

- $Q = 0$: the geodesic lies unstably in the equatorial plane or moves asymptotically towards it (possibly after a turning point at $\theta_m \neq \pi/2$), $\pi/2$ being an asymptotic θ -value, since $\Theta'(\pi/2) = -V'(\pi/2) = 0$ (cf. Sec. 11.3.6).
- $Q > 0$: the geodesic oscillates about the equatorial plane, between two turning points at $\theta = \theta_m$ and $\pi - \theta_m$, θ_m being the solution of $V(\theta_m) = Q$ in $(0, \pi/2)$ (cf. the trajectory $Q = Q_2$ in Fig. 11.5, right).

Expression of the θ -turning points

The θ -turning points θ_m and θ_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

$$x := \cos^2 \theta, \quad (11.83)$$

the equation $\Theta(\theta) = 0$ is equivalent to

$$a^2(E^2 - \mu^2)x^2 + [L^2 + Q - a^2(E^2 - \mu^2)]x - Q = 0. \quad (11.84)$$

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

$$\theta_m = \arccos \sqrt{\frac{Q}{L^2 + Q}}. \quad (11.85)$$

In the rest of this section, we assume that $a^2(E^2 - \mu^2) \neq 0$. Equation (11.84) is then a quadratic equation in x , the discriminant of which is

$$\begin{aligned} \Delta &= [L^2 + Q - a^2(E^2 - \mu^2)]^2 + 4Qa^2(E^2 - \mu^2) \\ &= [L^2 + Q + a^2(E^2 - \mu^2)]^2 - 4L^2a^2(E^2 - \mu^2). \end{aligned}$$

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. (11.82), we realize that $Q_2 = Q_{\min}$, so that $Q \geq Q_2$ always holds and $\Delta \geq 0$. Accordingly the two roots of Eq. (11.84) are real and are given by

$$x_{\pm} = \frac{1}{2} \left(1 - \frac{L^2 + Q}{a^2(E^2 - \mu^2)} \pm \sqrt{\left[1 - \frac{L^2 + Q}{a^2(E^2 - \mu^2)} \right]^2 + \frac{4Q}{a^2(E^2 - \mu^2)}} \right). \quad (11.86)$$

Since $x := \cos^2 \theta$ and $\theta = 0$ and $\theta = \pi/2$ cannot correspond to a θ -turning point, acceptable solutions must obey

$$0 < x_{\pm} < 1. \quad (11.87)$$

For $L = 0$, Eq. (11.86) simplifies to¹⁶

$$x_+ = 1 \quad \text{and} \quad x_- = -\frac{Q}{a^2(E^2 - \mu^2)}$$

Now, $x_+ = 1$ is excluded by (11.87). 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:

$$\boxed{\theta_m = \theta_v = \arccos \sqrt{\frac{Q}{a^2(\mu^2 - E^2)}}}_{L=0} \quad (11.88)$$

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. 11.4), otherwise there is no θ -turning point (as for the trajectories $Q = Q_2$ in both panels of Fig. 11.4). More precisely, for $Q > 0$ and $\mu^2 - E^2 > 0$, it leads to θ_m , while for $Q < 0$ and $\mu^2 - E^2 < 0$, it leads to θ_v .

For $L \neq 0$, one shall distinguish between two cases, since the case $|E| = \mu$ has been covered above [Eq. (11.85)]:

- if $|E| < \mu$, then only x_- fulfills the criterion (11.87), leading to the turning point value:

$$\boxed{\theta_m = \arccos \sqrt{\frac{1}{2} \left[1 + \frac{L^2 + Q}{a^2(\mu^2 - E^2)} - \sqrt{\left[1 + \frac{L^2 + Q}{a^2(\mu^2 - E^2)} \right]^2 - \frac{4Q}{a^2(\mu^2 - E^2)}} \right]}} \quad (11.89)$$

- if $|E| > \mu$, then x_+ always fulfills the criterion (11.87), leading to the turning point value:

$$\boxed{\theta_m = \arccos \sqrt{\frac{1}{2} \left[1 - \frac{L^2 + Q}{a^2(E^2 - \mu^2)} + \sqrt{\left[1 - \frac{L^2 + Q}{a^2(E^2 - \mu^2)} \right]^2 + \frac{4Q}{a^2(E^2 - \mu^2)}} \right]}} \quad (11.90)$$

If, in addition $Q < 0$, then x_- also fulfills (11.87) and we get a second turning point value (cf. the case $Q = Q_1$ in the right panel of Fig. 11.5):

$$\boxed{\theta_v = \arccos \sqrt{\frac{1}{2} \left[1 - \frac{L^2 + Q}{a^2(E^2 - \mu^2)} - \sqrt{\left[1 - \frac{L^2 + Q}{a^2(E^2 - \mu^2)} \right]^2 + \frac{4Q}{a^2(E^2 - \mu^2)}} \right]}} \quad (11.91)$$

Remark 9: At the limit $L \rightarrow 0$, formulas (11.89) and (11.91) reduce both to (11.88).

¹⁶A switch between x_+ and x_- is performed if $1 + Q/(a^2(E^2 - \mu^2)) < 0$.

Summary

We can summarize the above results by

Property 11.19: latitudinal motion of Kerr geodesics

- A geodesic \mathcal{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:

$$Q \geq 0. \quad (11.92)$$

- 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:

$$Q \geq Q_{\min} = - \left(a \sqrt{E^2 - \mu^2} - |L| \right)^2. \quad (11.93)$$

A geodesic with $Q < 0$ is called **vortical**; it never encounters the equatorial plane.

- If $Q > 0$, \mathcal{L} oscillates symmetrically about the equatorial plane, between two θ -turning points, at $\theta = \theta_m$ and $\theta = \pi - \theta_m$, where $\theta_m \in (0, \pi/2)$ is given by Eq. (11.85) for $a^2(E^2 - \mu^2) = 0$, Eq. (11.88) for $L = 0$ and $Q < a^2(\mu^2 - E^2)$, Eq. (11.89) for $|E| < \mu$ and Eq. (11.90) for $|E| > \mu$, except for two subcases with $L = 0$: (i) $Q = a^2(\mu^2 - E^2)$: \mathcal{L} lies unstably along the rotation axis or approaches it asymptotically and (ii) $Q > a^2(\mu^2 - E^2)$: \mathcal{L} crosses repeatedly the rotation axis, with θ taking all values in the range $[0, \pi]$.
- If $Q = 0$, \mathcal{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 $L^2 < a^2(E^2 - \mu^2)$, \mathcal{L} either lies unstably in the equatorial plane or approaches it asymptotically from one side, while for $L^2 = a^2(E^2 - \mu^2) = 0$, \mathcal{L} lies at a constant value $\theta = \theta_0 \in [0, \pi]$.
- If $Q_{\min} < Q < 0$, \mathcal{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 $L \neq 0$, \mathcal{L} oscillates between two θ -turning points, at $\theta = \theta_m$ and $\theta = \theta_v$ (Northern hemisphere) or at $\theta = \pi - \theta_v$ and $\theta = \pi - \theta_m$ (Southern hemisphere), where θ_m and θ_v are given by Eqs. (11.90) and (11.91) respectively; if $L = 0$, \mathcal{L} oscillates about the rotation axis, with a θ -turning point at $\theta = \theta_v$ or $\theta = \pi - \theta_v$, where θ_v is given by Eq. (11.88).

- If $Q = Q_{\min}$, \mathcal{L} lies stably at a constant value $\theta = \theta_*$ or $\theta = \pi - \theta_*$, with $\theta_* \in [0, \pi/2)$ given by^a

$$\theta_* := \arcsin \sqrt{\frac{|L|}{a\sqrt{E^2 - \mu^2}}}. \quad (11.94)$$

^aThis is Eq. (11.81) generalized to encompass the case $L = 0$.

Remark 10: 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, $L = \boldsymbol{\eta} \cdot \mathbf{p}$ [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 $\text{SO}(2)$. Hence $L = 0$ on the rotation axis. Since L is constant along the geodesic, the result follows immediately.

Remark 11: 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 8 (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 \mathcal{L} under consideration. So θ may vary along \mathcal{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. (11.92)].

If $Q = 0$, then for $L = 0$, \mathcal{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 $\mathbf{L}_{\text{tot}} = 0$ in that case [set $Q = 0$, $L = 0$ and $a = 0$ in Eq. (11.37)]. If $L \neq 0$, still with $Q = 0$, \mathcal{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: \mathcal{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 $L \neq 0$, \mathcal{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 \mathcal{L} is then $\theta_m = \pi/2 - \iota$.

11.3.8 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 :

$$R(r) = (E^2 - \mu^2)r^4 + 2m\mu^2r^3 + [a^2(E^2 - \mu^2) - Q - L^2]r^2 + 2m[Q + (L - aE)^2]r - a^2Q. \quad (11.95)$$

Geodesics with $|E| < \mu$

For $r \rightarrow \pm\infty$ and $|E| \neq \mu$, we have $R(r) \sim (E^2 - \mu^2)r^4$ and in particular $\lim_{r \rightarrow \pm\infty} R(r) = +\infty$ for $|E| > \mu$ and $\lim_{r \rightarrow \pm\infty} R(r) = -\infty$ 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. 11.3.7 lead to $Q \geq 0$ [cf. Eq. (11.92)]. Therefore, $-a^2Q \leq 0$. It follows then that r cannot be negative. Indeed, we see on expression (11.95) that for $|E| < \mu$ (which implies $\mu > 0$ and $Q \geq 0$) and $r < 0$,

$$R(r) = \underbrace{(E^2 - \mu^2)r^4}_{<0} + \underbrace{2m\mu^2r^3}_{<0} + \underbrace{[a^2(E^2 - \mu^2) - Q - L^2]r^2}_{\leq 0} + \underbrace{2m[Q + (L - aE)^2]r}_{\leq 0} - a^2Q, \leq 0$$

which contradicts $R(r) \geq 0$. Hence we conclude:

Property 11.20: bound orbits

Any geodesic \mathcal{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 arbitrarily large values of r . \mathcal{L} is called a **geodesic with a bound orbit**, or in short, a **bound geodesic**.

Remark 12: That \mathcal{L} cannot reach the asymptotic region $r \rightarrow +\infty$ could have been found directly from the definition of E as the “energy at infinity” of the particle \mathcal{P} having \mathcal{L} as worldline [Eq. (11.2a)]. Indeed, since \mathcal{L} is timelike, the 4-momentum of \mathcal{P} is $\mathbf{p} = \mu \mathbf{u}$ [Eq. (1.18)], where \mathbf{u} is the 4-velocity of \mathcal{P} , so that Eq. (11.2a) yields $E = -\mu \xi \cdot \mathbf{u}$. Now, when $r \rightarrow +\infty$, ξ tends to the 4-velocity of an inertial observer (cf. Sec. 10.7.5). If \mathcal{L} could reach the asymptotic region, the scalar product of the two 4-velocities ξ and \mathbf{u} would be necessarily lower or equal to -1 , or equivalently $\Gamma := -\xi \cdot \mathbf{u} \geq 1$ (the proof lies in expression (1.35) of Γ with $0 \leq \mathbf{V} \cdot \mathbf{V} < 1$), 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. 373). We have

Property 11.21

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. (11.92)], which is evidently fulfilled for $|E| = \mu$. Furthermore, for $|E| = \mu$, expression (11.95) for $R(r)$ simplifies to

$$R(r) = 2m\mu^2r^3 - (Q + L^2)r^2 + 2m[Q + (L - aE)^2]r - a^2Q.$$

For $r < 0$, all the four terms in the above sum are ≤ 0 . If \mathcal{L} is timelike, then $\mu \neq 0$ and the first term is < 0 , so that $r < 0 \implies R(r) < 0$, which violates the constraint $R(r) \geq 0$. Let us now assume that \mathcal{L} is null. We have then $\mu = 0$ and $E = 0$, so that

$$R(r) = -(Q + L^2)r^2 + 2m(Q + L^2)r - a^2Q = (Q + L^2)r(2m - r) - a^2Q.$$

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 \mathcal{K} is then zero as well, since $\mathcal{K} = Q + (L - aE)^2$ [Eq. (11.34)]. Then by the result (11.25), \mathcal{L} is a null geodesic generator of either the event horizon \mathcal{H} , which is located at $r = r_+$, or the inner horizon \mathcal{H}_{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. \square

A timelike geodesic \mathcal{L} with $E = \mu$ can reach the asymptotic region $r \rightarrow +\infty$. It has then a Lorentz factor with respect to the asymptotic static observer of 4-velocity ξ equal to one (cf. Remark 12 above), which implies that p is collinear to ξ . In that sense, \mathcal{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 $r \rightarrow +\infty$. Actually, it cannot even exist outside the ergoregion, by virtue of Property 11.5.

Example 9: The null geodesics generating the horizons \mathcal{H} and \mathcal{H}_{in} considered in Example 1 on p. 373 have $E = 0$ [Eq. (11.4)] and are indeed fully located in the ergoregion \mathcal{G} , since $\mathcal{H} \subset \mathcal{G}$ and $\mathcal{H}_{\text{in}} \subset \mathcal{G}$ (cf. Fig. 10.2).

Geodesics with $|E| > \mu$

An immediate corollary of the properties obtained for $|E| < \mu$ and $|E| = \mu$ is

Property 11.22: necessary condition to visit the negative- r region

Only geodesics with $|E| > \mu$ may have some part in the region $r < 0$ of Kerr spacetime, \mathcal{M}_- :

$$\mathcal{L} \cap \mathcal{M}_- \neq \emptyset \implies |E| > \mu. \quad (11.96)$$

Decay of r towards the future in \mathcal{M}_{II}

As a particular case of the result established in Sec. 10.3.2 for any causal worldline (not necessarily a geodesic), we have

Property 11.23: decreasing of r in \mathcal{M}_{II}

In region \mathcal{M}_{II} , the coordinate r must decrease towards the future along any timelike or null geodesic:

$$\boxed{\frac{dr}{d\lambda} < 0}_{\mathcal{M}_{\text{II}}} \quad (11.97)$$

Example 10 (Principal null geodesics): For the ingoing principal null geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$, a future-directed tangent vector is k and we have $k^r = dr/d\lambda = -1$, since the affine parameter associated with k is $\lambda = -r$ [Eq. (10.46)]. Hence for these geodesics, r is decreasing towards the future everywhere in \mathcal{M} , and in particular in \mathcal{M}_{II} . For the outgoing principal null geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out}}$, a future-directed tangent vector is ℓ and the associated (non-affine) parameter obeys Eq. (10.56):

$$\frac{dr}{d\lambda} = \frac{\Delta}{2(r^2 + a^2)}.$$

Thus $dr/d\lambda < 0$ in \mathcal{M}_{II} , since $\Delta < 0$ there.

11.3.9 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 \mathcal{R} discussed in Sec. 10.2.6.

Formally, \mathcal{R} is not part of the Kerr spacetime \mathcal{M} but of the larger manifold $\mathbb{R}^2 \times \mathbb{S}^2$ [cf. the construction (10.29)]. It is located at $\rho^2 = 0$, i.e. at $r = 0$ and $\theta = \pi/2$. We shall say that a geodesic \mathcal{L} of affine parameter λ (oriented towards the future) **approaches the ring singularity** if \mathcal{L} has both $r(\lambda) \rightarrow 0$ and $\theta(\lambda) \rightarrow \pi/2$ as $\lambda \rightarrow \lambda_*$, with λ_* being finite or infinite. If λ_* is finite, we shall say that \mathcal{L} **hits the ring singularity** for $\lambda \rightarrow \lambda_*^-$ and **emanates from the ring singularity** for $\lambda \rightarrow \lambda_*^+$. If $\lambda_* = \pm\infty$, we shall say that \mathcal{L} **asymptotically approaches the ring singularity**.

A first key result is

Property 11.24: vanishing of Q required to approach the ring singularity

A null or timelike geodesic \mathcal{L} that approaches the ring singularity has a vanishing Carter constant Q .

Proof. From Eqs. (11.95) and (11.35), we have

$$\lim_{r \rightarrow 0} R(r) = -a^2 Q \quad \text{and} \quad \lim_{\theta \rightarrow \pi/2} \Theta(\theta) = Q.$$

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 $Q = 0$. \square

The above property can be refined¹⁷:

¹⁷This 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 \mathcal{H} have $Q = 0$ [Eq. (11.25)] but those with $\theta \neq \pi/2$ lie outside the equatorial plane.

Property 11.25: ring singularity reachable only from the equatorial plane

A null or timelike geodesic cannot approach the ring singularity unless it lies entirely in the equatorial plane.

Proof. Let \mathcal{L} be a causal geodesic that approaches \mathcal{R} . From the previous result, \mathcal{L} has $Q = 0$. Reviewing the subcases $Q = 0$ among all the cases considered in Sec. 11.3.7, we see that \mathcal{L} can approach $\theta = \pi/2$ iff (i) \mathcal{L} has $|E| \leq \sqrt{\mu^2 + L^2/a^2}$ and lies stably in the equatorial plane or (ii) \mathcal{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 $r \rightarrow 0$. For $Q = 0$, expression (11.95) for $R(r)$ reduces to

$$R(r) = r \left[(E^2 - \mu^2)r^3 + 2m\mu^2r^2 + (a^2(E^2 - \mu^2) - L^2)r + 2m(L - aE)^2 \right],$$

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 \mathcal{L} would reach both $r = 0$ and $\theta = \pi/2$ when λ tends to some value λ_* , then Eq. (11.55a) would yield

$$\lambda'_* - \lambda'_0 = \int_{r_0}^0 \frac{\epsilon_r dr}{\sqrt{R(r)}} = \int_{\theta_0}^{\frac{\pi}{2}} \frac{\epsilon_\theta d\theta}{\sqrt{\Theta(\theta)}}, \quad (11.98)$$

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 $\rho^2 \rightarrow 0$ when $\lambda \rightarrow \lambda_*$. 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. (11.98) 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. \square

Property 11.25 has been obtained for a generic approach to the ring singularity, i.e. for a geodesic \mathcal{L} that hits \mathcal{R} , emanates from \mathcal{R} or asymptotically approach \mathcal{L} in the future or the past. If we apply it to null geodesics (light rays) emanating from \mathcal{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.

11.3.10 Moving to the negative- r side

If it maintains $\theta \neq \pi/2$ in the vicinity of $r = 0$, a geodesic \mathcal{L} can a priori move from the region $r > 0$ of \mathcal{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. 11.3.6). In both cases, \mathcal{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 (11.95) of R , we have $R(0) = -a^2 Q$, so that $R(0) > 0 \iff Q < 0$. We thus conclude

Property 11.26: crossing $r = 0$ only possible for vortical geodesics

Only a vortical geodesic ($Q < 0$) can cross the hypersurface $r = 0$ and thus move from the positive- r region of spacetime (\mathcal{M}_+) to the negative- r one (\mathcal{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. 11.3.7. We note that it implies $|E| > \mu$, which is consistent with the property (11.96) required to travel in \mathcal{M}_- .

Example 11: 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 10.9, 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 fulfill $|E| > \sqrt{\mu^2 + L^2/a^2}$.

11.4 Timelike geodesics

11.4.1 Parametrization

Whenever the geodesic \mathcal{L} is timelike, it is relevant to rescale everything by the mass μ of the particle \mathcal{P} whose worldline is \mathcal{L} . In particular, as in the Schwarzschild case treated in Sec. 7.3, let us parameterize \mathcal{L} by the proper time τ , which is the affine parameter τ related to the affine parameter λ associated to the 4-momentum p by

$$\tau = \mu\lambda \quad (11.99)$$

and let us introduce the *specific conserved energy* ε , *specific conserved angular momentum* ℓ , and *reduced Carter constant* q :

$$\boxed{\varepsilon := \frac{E}{\mu}}, \quad \boxed{\ell := \frac{L}{\mu}} \quad \text{and} \quad \boxed{q := \frac{Q}{\mu^2}}. \quad (11.100)$$

We shall refer to ε , ℓ and q as the *reduced integrals of motion* of the geodesic \mathcal{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:

$$\frac{dt}{d\tau'} = T_1(r) + T_2(\theta) \quad (11.101a)$$

$$\left(\frac{dr}{d\tau'} \right)^2 - \mathcal{R}(r) = 0 \quad (11.101b)$$

$$\left(\frac{d\theta}{d\tau'} \right)^2 - \tilde{\Theta}(\theta) = 0 \quad (11.101c)$$

$$\frac{d\varphi}{d\tau'} = \Phi_1(r) + \Phi_2(\theta) \quad (11.101d)$$

where τ' is related to the Mino parameter λ' by

$$\tau' = \mu\lambda' \quad (11.102)$$

and

$$T_1(r) := \frac{\varepsilon(r^2 + a^2)^2 - 2am\ell r}{r^2 - 2mr + a^2}, \quad T_2(\theta) := -a^2\varepsilon \sin^2 \theta, \quad (11.103)$$

$$\Phi_1(r) := \frac{a(2m\varepsilon r - a\ell)}{r^2 - 2mr + a^2}, \quad \Phi_2(\theta) := \frac{\ell}{\sin^2 \theta}, \quad (11.104)$$

$$\mathcal{R}(r) := (\varepsilon^2 - 1)r^4 + 2mr^3 + [a^2(\varepsilon^2 - 1) - q - \ell^2]r^2 + 2m[q + (\ell - a\varepsilon)^2]r - a^2q, \quad (11.105)$$

$$\tilde{\Theta}(\theta) := q + \cos^2 \theta \left[a^2(\varepsilon^2 - 1) - \frac{\ell^2}{\sin^2 \theta} \right]. \quad (11.106)$$

Note that $\mathcal{R}(r) = R(r)/\mu^2$, $\tilde{\Theta}(\theta) := \Theta(\theta)/\mu^2$ and that expressions (11.105) and (11.106) follow respectively from Eqs. (11.95) and (11.35).

Using Eqs. (11.99) and (11.49), we can relate τ' to the proper time τ :

$$d\tau' = \frac{d\tau}{\rho^2} = \frac{d\tau}{r(\tau)^2 + a^2 \cos^2 \theta(\tau)}. \quad (11.107)$$

We shall call τ' **Mino time** along the geodesic \mathcal{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 \mathcal{L} with $|E| < \mu$, or equivalently

$$|\varepsilon| < 1. \quad (11.108)$$

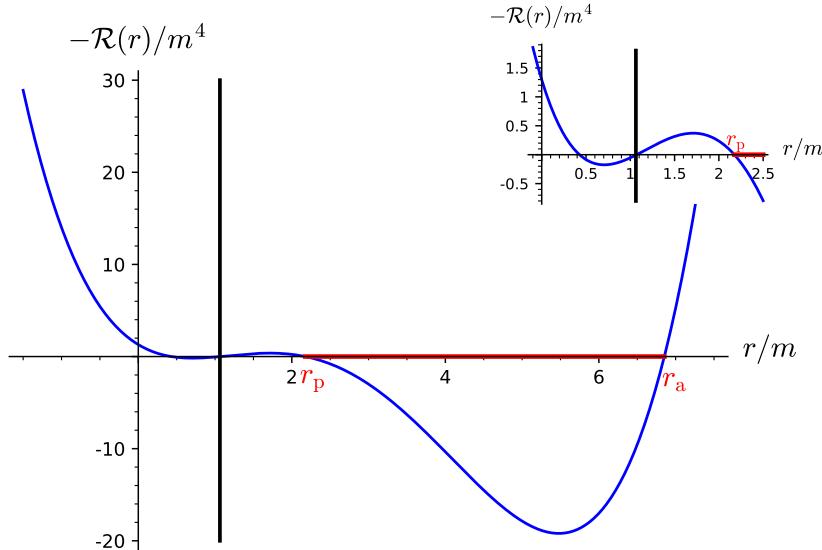


Figure 11.6: Effective potential $-\mathcal{R}(r)$ corresponding to $a = 0.998 m$, $\varepsilon = 0.9$, $\ell = 2m$ and $q = 1.3 m^2$. The black vertical line marks the black hole horizon at $r = r_+ \simeq 1.063 m$. Note that the polynomial $-\mathcal{R}(r)$ has four real roots, one of them being very 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.5.8]

As shown in Sec. 11.3.8, such a geodesic has necessarily a non-negative Carter constant:

$$q \geq 0 \quad (11.109)$$

and is located in a radially bounded part of the $r \geq 0$ region of Kerr spacetime.

Equation (11.101b) 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. (11.101b). 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 (11.108) implies that the coefficient of r^4 in formula (11.105) for $\mathcal{R}(r)$ is negative. It follows then that $\lim_{r \rightarrow \pm\infty} U(r) = +\infty$ and the segments with $U(r) \leq 0$ are located between two roots of $\mathcal{R}(r)$ (cf. Fig. 11.6). Let us denote the lower of these two roots by r_p , for **periastron**, and the larger one by r_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. (11.101b) (cf. Fig. 11.6) is periodic:

$$\forall n \in \mathbb{Z}, \quad r(\tau' + n\Lambda_r) = r(\tau'), \quad (11.110)$$

the period Λ_r being the Mino time $\Delta\tau'$ spent to perform a round-trip between r_p and r_a . By rewriting Eq. (11.101b) as $d\tau' = \pm dr / \sqrt{\mathcal{R}(r)}$, we get

$$\Lambda_r = 2 \int_{r_p}^{r_a} \frac{dr}{\sqrt{\mathcal{R}(r)}}.$$

(11.111)

Since $\mathcal{R}(r)$ is a polynomial of degree 4 [cf. Eq. (11.105)], the integral in the right-hand side of Eq. (11.111) can be evaluated by means of elliptic integrals. We shall not give the detail here, referring the interested reader to the article [205].

Regarding the θ -motion, we are in the case $L^2 > a^2(E^2 - \mu^2)$ considered in Sec. 11.3.7, 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. 11.4 for $\ell = 0$ and in the left panel of Fig. 11.5 for $\ell \neq 0$. We can then distinguish three types of bound orbits:

- **polar orbit:** \mathcal{L} crosses the rotational axis an infinite number of times; this occurs iff $\ell = 0$ and $q > a^2(1 - \varepsilon^2)$;
- **equatorial orbit:** \mathcal{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:** \mathcal{L} never crosses the rotational axis and oscillates symmetrically about the equatorial plane between two θ -turning points at $\theta = \theta_m \in (0, \pi/2)$ and $\theta = \pi - \theta_m$; 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. 11.3.6 and the grey dashed curve in the left panel of Fig. 11.4).

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 $\theta_m \leq \theta \leq \pi - \theta_m$, with θ_m given by Eq. (11.89):

$$\theta_m = \arccos \sqrt{\frac{1}{2} \left[1 + \frac{\ell^2 + q}{a^2(1 - \varepsilon^2)} - \sqrt{\left[1 + \frac{\ell^2 + q}{a^2(1 - \varepsilon^2)} \right]^2 - \frac{4q}{a^2(1 - \varepsilon^2)}} \right]}, \quad (11.112)$$

which reduces to Eq. (11.88) for $\ell = 0$:

$$\theta_m = \arccos \sqrt{\frac{q}{a^2(1 - \varepsilon^2)}} \Big|_{\ell=0}. \quad (11.113)$$

From the viewpoint of Eq. (11.101c) above, non-polar orbits have a periodic θ -motion in the potential well $-\tilde{\Theta}(\theta)$:

$$\forall n \in \mathbb{Z}, \quad \theta(\tau' + n\Lambda_\theta) = \theta(\tau'). \quad (11.114)$$

The period Λ_θ is the Mino time $\Delta\tau'$ spent in a round-trip between θ_m and $\pi - \theta_m$. By rewriting Eq. (11.101c) as $d\tau' = \pm d\theta / \sqrt{\tilde{\Theta}(\theta)}$, we get

$$\Lambda_\theta = 2 \int_{\theta_m}^{\pi - \theta_m} \frac{d\theta}{\sqrt{\tilde{\Theta}(\theta)}} = 4 \int_{\theta_m}^{\pi/2} \frac{d\theta}{\sqrt{\tilde{\Theta}(\theta)}}, \quad (11.115)$$

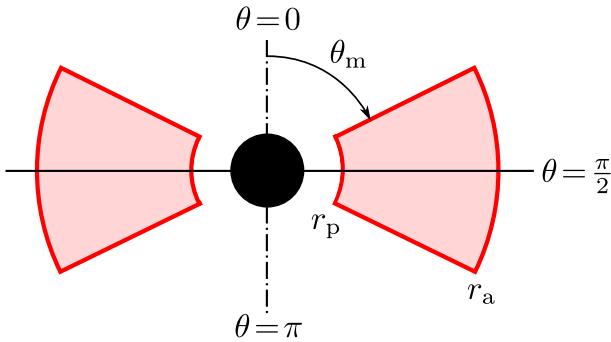


Figure 11.7: Meridional section of the torus occupied by a bound timelike geodesic. The dot-dashed line is the rotation axis and the solid line marks the equatorial plane.

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**:

$$\boxed{\Upsilon_r := \frac{2\pi}{\Lambda_r}} \quad \text{and} \quad \boxed{\Upsilon_\theta := \frac{2\pi}{\Lambda_\theta}}. \quad (11.116)$$

For non-polar orbits, the motion is restricted by

$$\boxed{r_p \leq r \leq r_a} \quad \text{and} \quad \boxed{\theta_m \leq \theta \leq \pi - \theta_m}. \quad (11.117)$$

This means that in terms of the Cartesian Boyer-Lindquist coordinates (11.73), the geodesic is confined inside a torus (cf. Fig. 11.7), 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. [243].

As an illustration, Figs. 11.8–11.10 show a bound timelike geodesic \mathcal{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 $\theta_m \simeq 1.060 \text{ rad} \simeq 60.75^\circ$. The effective radial potential $-\mathcal{R}(r)$ of this geodesic is the one depicted in Fig. 11.6. Figures 11.8–11.10 show actually the segment $0 \leq \tau \leq 600 m$ of \mathcal{L} , starting at the point of Boyer-Lindquist coordinates $(0, (r_p + r_a)/2, \pi/2, 0)$. That $\theta(\tau)$ lies in the interval $[\theta_m, \pi - \theta_m] \sim [60^\circ, 120^\circ]$ appears clearly in the right panel of Fig. 11.9. On Figs. 11.9 (left panel) and 11.10, we notice a generic feature of eccentric orbits in the Kerr metric, known as **zoom-whirl**¹⁸: 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. (11.101a) and (11.101d). Since r and θ are periodic functions of τ' , so are the terms

¹⁸According to Ref. [222], the name “zoom-whirl” has been forged by Curt Cutler and Eric Poisson and may have been suggested by Kip Thorne.

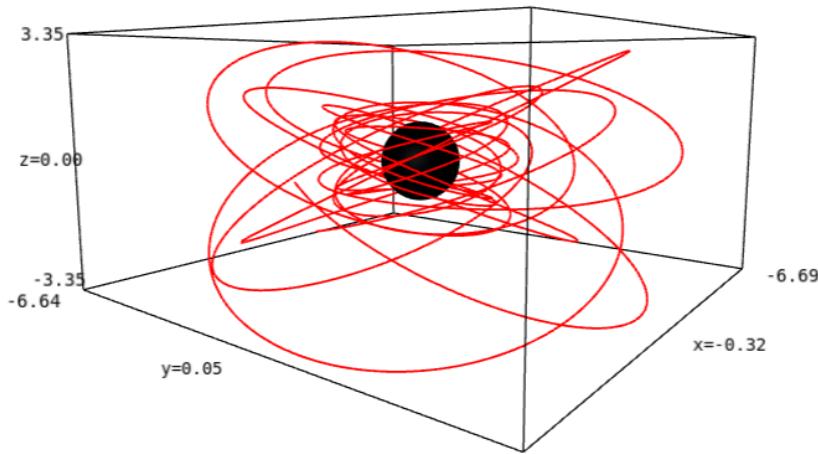


Figure 11.8: Trajectory with respect to the Cartesian Boyer-Lindquist coordinates (x, y, z) [Eq. (11.73)] of a bound timelike geodesic of reduced integrals of motion $\varepsilon = 0.9$, $\ell = 2m$ and $q = 1.3 m^2$, orbiting around a Kerr black hole with $a = 0.998 m$. These parameters are the same as for the $-\mathcal{R}(r)$ plot in Fig. 11.6. [Figure generated by the notebook D.5.8]

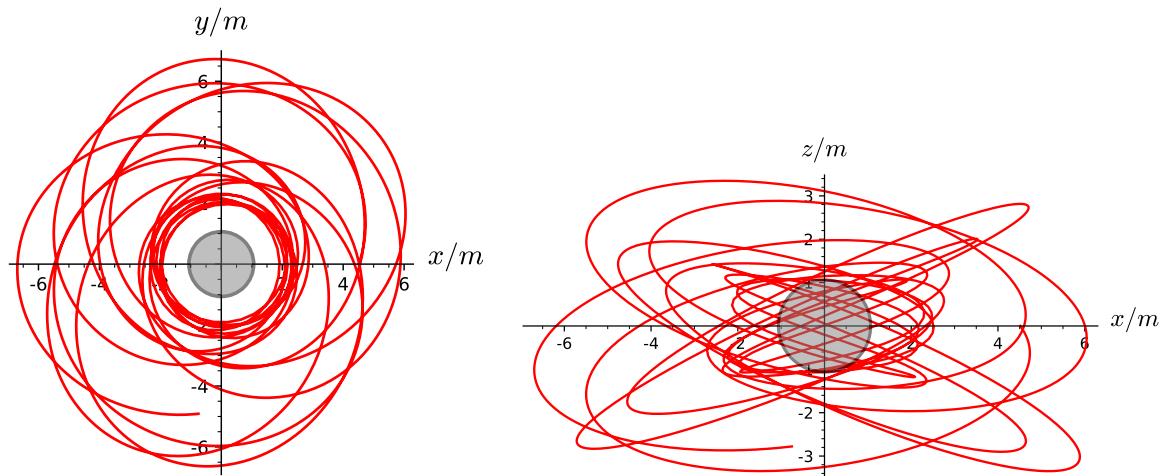


Figure 11.9: Projection in the (x, y) plane (equatorial plane; left panel) and the (x, z) plane (meridional plane; right panel) of the bound timelike geodesic considered in Fig. 11.8. The grey disk depicts the black hole region. On the left panel, there appears clearly that the minimal value of r is $r_p \simeq 2.2 m$, while its maximal value is $r_a \simeq 6.9 m$. On the right panel, we check that $\theta_m \leq \theta \leq \pi - \theta_m$ with $\theta_m \sim 60^\circ$, in agreement with the generic behavior illustrated in Fig. 11.7. [Figure generated by the notebook D.5.8]

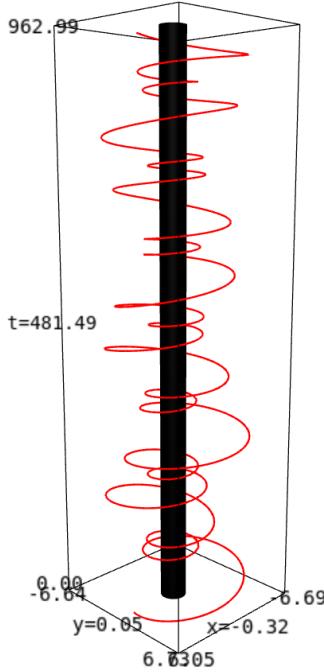


Figure 11.10: Spacetime diagram based on the Cartesian Boyer-Lindquist coordinates (t, x, y) [Eq. (11.73)] depicting the bound timelike geodesic considered in Figs. 11.8 and 11.9. The black cylinder is the black hole event horizon. [Figure generated by the notebook D.5.8]

$T_1(r)$, $T_2(\theta)$, $\Phi_1(r)$ and $\Phi_2(\theta)$ that appear in the right-hand side of Eqs. (11.101a) and (11.101d). We can thus expand them in Fourier series:

$$T_1(r) = \sum_{k=-\infty}^{+\infty} \hat{T}_{1k} e^{ik\Upsilon_r \tau'} = \langle T_1(r) \rangle + \sum_{k=1}^{+\infty} (\hat{T}_{1k} e^{ik\Upsilon_r \tau'} + \text{c.c.}) \quad (11.118a)$$

$$T_2(\theta) = \sum_{k=-\infty}^{+\infty} \hat{T}_{2k} e^{ik\Upsilon_\theta \tau'} = \langle T_2(\theta) \rangle + \sum_{k=1}^{+\infty} (\hat{T}_{2k} e^{ik\Upsilon_\theta \tau'} + \text{c.c.}) \quad (11.118b)$$

$$\Phi_1(r) = \sum_{k=-\infty}^{+\infty} \hat{\Phi}_{1k} e^{ik\Upsilon_r \tau'} = \langle \Phi_1(r) \rangle + \sum_{k=1}^{+\infty} (\hat{\Phi}_{1k} e^{ik\Upsilon_r \tau'} + \text{c.c.}) \quad (11.118c)$$

$$\Phi_2(\theta) = \sum_{k=-\infty}^{+\infty} \hat{\Phi}_{2k} e^{ik\Upsilon_\theta \tau'} = \langle \Phi_2(\theta) \rangle + \sum_{k=1}^{+\infty} (\hat{\Phi}_{2k} e^{ik\Upsilon_\theta \tau'} + \text{c.c.}), \quad (11.118d)$$

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:

$$\langle T_1(r) \rangle = \frac{1}{\Lambda_r} \int_0^{\Lambda_r} T_1(r(\tau')) d\tau' = \frac{2}{\Lambda_r} \int_{r_p}^{r_a} \frac{T_1(r)}{\sqrt{R(r)}} dr \quad (11.119a)$$

$$\langle T_2(\theta) \rangle = \frac{1}{\Lambda_\theta} \int_0^{\Lambda_\theta} T_2(\theta(\tau')) d\tau' = \frac{4}{\Lambda_\theta} \int_{\theta_m}^{\pi/2} \frac{T_2(\theta)}{\sqrt{\tilde{\Theta}(\theta)}} d\theta, \quad (11.119b)$$

with similar formulas for $\langle \Phi_1(r) \rangle$ and $\langle \Phi_2(\theta) \rangle$.

We may then rewrite Eqs. (11.101a) and (11.101d) as

$$\frac{dt}{d\tau'} = \Gamma + \sum_{k=1}^{+\infty} \left(\hat{T}_{1k} e^{ik\Upsilon_r \tau'} + \text{c.c.} \right) + \sum_{k=1}^{+\infty} \left(\hat{T}_{2k} e^{ik\Upsilon_\theta \tau'} + \text{c.c.} \right) \quad (11.120a)$$

$$\frac{d\varphi}{d\tau'} = \Upsilon_\varphi + \sum_{k=1}^{+\infty} \left(\hat{\Phi}_{1k} e^{ik\Upsilon_r \tau'} + \text{c.c.} \right) + \sum_{k=1}^{+\infty} \left(\hat{\Phi}_{2k} e^{ik\Upsilon_\theta \tau'} + \text{c.c.} \right), \quad (11.120b)$$

with

$$\Gamma := \langle T_1(r) \rangle + \langle T_2(\theta) \rangle, \quad (11.121)$$

and

$$\Upsilon_\varphi := \langle \Phi_1(r) \rangle + \langle \Phi_2(\theta) \rangle. \quad (11.122)$$

Γ and Υ_φ represent the constant (or average) parts of respectively $dt/d\tau'$ and $d\varphi/d\tau'$, while the oscillatory parts are the terms involved in the sums over $k \geq 1$. Since Γ is the average value of $dt/d\tau'$, we can define the **average angular frequencies with respect to Boyer-Lindquist time** by [172, 205]

$$\Omega_r := \frac{\Upsilon_r}{\Gamma}, \quad \Omega_\theta := \frac{\Upsilon_\theta}{\Gamma}, \quad \Omega_\varphi := \frac{\Upsilon_\varphi}{\Gamma}. \quad (11.123)$$

We may summarize the above results as follows:

Property 11.27: periodicities of bound timelike orbits

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 \mathcal{L} of Kerr spacetime, one may define its **eccentricity** e and (dimensionless) **semilatus rectum** p by means of Keplerian-like formulas [446, 173, 460]:

$$r_p =: \frac{pm}{1+e} \quad \text{and} \quad r_a =: \frac{pm}{1-e}, \quad (11.124)$$

or equivalently

$$p = \frac{2r_a r_p}{m(r_a + r_p)} \quad \text{and} \quad e = \frac{r_a - r_p}{r_a + r_p}. \quad (11.125)$$

One may also introduce the *inclination angle* θ_{inc} by

$$\theta_{\text{inc}} := \frac{\pi}{2} - (\text{sgn } \ell) \theta_m. \quad (11.126)$$

The sign of ℓ appears in this formula to enforce $\theta_{\text{inc}} \in [0, \pi/2]$ for prograde orbits ($\ell > 0$) and $\theta_{\text{inc}} \in [\pi/2, \pi]$ for retrograde orbits ($\ell < 0$).

Instead of (ε, ℓ, q) , a bound timelike geodesic can be parameterized¹⁹ by $(p, e, \theta_{\text{inc}})$. Actually, there is a one-to-one correspondence between (ε, ℓ, q) and $(p, e, \theta_{\text{inc}})$: it is clear from formulas (11.125) and (11.126) that $(p, e, \theta_{\text{inc}})$ 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. (11.105)] and (ii) θ_m in the right-hand side of expression (11.126) for θ_{inc} is the function of (ε, ℓ, q) given by Eq. (11.112). Conversely, to express (ε, ℓ, q) in terms of $(p, e, \theta_{\text{inc}})$, one shall write

$$\begin{cases} \mathcal{R}(r_p) = 0 \\ \mathcal{R}(r_a) = 0 \\ \tilde{\Theta}(\theta_m) = 0, \end{cases}$$

where r_p and r_a are the functions of (e, p) given by Eq. (11.124) and θ_m is the function of θ_{inc} given by Eq. (11.126). The above system is then a system of 3 equations for the 3 unknowns (ε, ℓ, q) ; see Appendix B of Ref. [446] or Appendix A of Ref. [173] for details.

Example 12: The geodesic \mathcal{L} depicted in Figs. 11.8–11.10 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 $\theta_m = 1.060 \text{ rad} = 60.75^\circ$. Then Eqs. (11.125) and (11.126) yield $p = 3.302$, $e = 0.518$ and $\theta_{\text{inc}} = 29.26^\circ$.

The parametrization $(p, e, \theta_{\text{inc}})$ allows one to introduce along the geodesic \mathcal{L} the radial phase angle $\psi(\tau)$ and the meridional phase angle $\chi(\tau)$ as monotonically increasing functions of τ such that

$$r(\tau) =: \frac{pm}{1 + e \cos \psi(\tau)} \quad \text{and} \quad \theta(\tau) =: \arccos(\cos \theta_m \cos \chi(\tau)). \quad (11.127)$$

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 $dr/d\tau$ and $d\theta/d\tau$ vanish and change sign [172].

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 $d\theta/d\tau' = 0$ and Eq. (11.101c) leads to $\tilde{\Theta}(\theta) = 0$. Given expression (11.106) for $\tilde{\Theta}(\theta)$, this implies

$$q = 0. \quad (11.128)$$

¹⁹Some authors use $x := \cos \theta_{\text{inc}}$ instead of θ_{inc} [460].

Remark 1: The above result follows as well from the analysis of the θ -motion performed in Sec. 11.3.7 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. 11.3.7).

When $\theta = \pi/2$, we have $\sin^2 \theta = 1$, $\rho^2 = r^2$ and $d\tau' = r^{-2}d\tau$ [cf. Eq. (11.107)], so that the system (11.101) reduces to

$$\frac{dt}{d\tau} = \frac{1}{r^2 - 2mr + a^2} \left[\varepsilon(r^2 + a^2) + \frac{2am}{r}(a\varepsilon - \ell) \right] \quad (11.129a)$$

$$\left(\frac{dr}{d\tau} \right)^2 + \mathcal{V}(r) = 0 \quad (11.129b)$$

$$\frac{d\varphi}{d\tau} = \frac{1}{r^2 - 2mr + a^2} \left[\ell \left(1 - \frac{2m}{r} \right) + \frac{2am\varepsilon}{r} \right], \quad (11.129c)$$

with

$$\mathcal{V}(r) := 1 - \varepsilon^2 - \frac{2m}{r} + \frac{\ell^2 + a^2(1 - \varepsilon^2)}{r^2} - \frac{2m(\ell - a\varepsilon)^2}{r^3}. \quad (11.130)$$

Note that $\mathcal{V}(r) = -\mathcal{R}(r)/r^4 = -R(r)/(\mu^2 r^4)$ [cf. Eqs. (11.105) and (11.95)].

Remark 2: At the Schwarzschild limit ($a = 0$) we recover the differential equations of Chap. 7, namely Eq. (11.129a) reduces to Eq. (7.34), Eq. (11.129b) to Eq. (7.24) and Eq. (11.129c) 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. (11.129b), as we did for Eq. (7.24) to make it match the first integral of a 1-dimensional motion in a potential well.

Historical note : The potential (11.130), which is ruling the radial motion of timelike geodesics confined to the equatorial plane, has been exhibited first by Robert Boyer and Richard Lindquist in 1966 (published 1967) [71], in the form $r^3\mathcal{V}(r)$, cf. their Eq. (5.6), taking into account that they were using $-a$ instead of a (cf. the historical note on p. 338).

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

$$\forall \tau \in \mathbb{R}, \quad \theta(\tau) = \pi/2 \quad \text{and} \quad r(\tau) = r_0 = \text{const.} \quad (11.131)$$

The constant r_0 is called the *radius* of the circular orbit. One has then $dr/d\tau = 0$, so that Eq. (11.129b) implies

$$\mathcal{V}(r_0) = 0. \quad (11.132)$$

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 (11.129) (which includes (11.132)) but that are not solution of the geodesic equation (11.1) (see Ref. [488] 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:

$$\frac{dr}{d\tau} \left[2 \frac{d^2r}{d\tau^2} + \mathcal{V}'(r) \right] = 0.$$

It is obtained by differentiating Eq. (11.129b) 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 $dr/d\tau \neq 0$ and the above relation implies

$$\mathcal{V}'(r) = -2 \frac{d^2r}{d\tau^2}.$$

Since in the vicinity of any circular geodesic there exist non-circular ones (those with small eccentricity e , as defined by Eq. (11.125)), by continuity, we shall ask that this relation holds for circular geodesics as well. For the latter ones, $r = r_0$ and $d^2r/d\tau^2 = 0$, so that it simplifies to

$$\boxed{\mathcal{V}'(r_0) = 0}. \quad (11.133)$$

One can verify²⁰ that this relation, in conjunction with Eq. (11.132), is sufficient to eliminate the non-geodesic circular worldlines.

To summarize, circular timelike geodesics in the equatorial plane are obtained by solving the system (11.132)-(11.133). Given expression (11.130) for \mathcal{V} , this system takes the form

$$\left\{ \begin{array}{l} (1 - \varepsilon^2)r_0^3 - 2mr_0^2 + [\ell^2 + a^2(1 - \varepsilon^2)]r_0 - 2m(\ell - a\varepsilon)^2 = 0 \\ mr_0^2 - [\ell^2 + a^2(1 - \varepsilon^2)]r_0 + 3m(\ell - a\varepsilon)^2 = 0. \end{array} \right. \quad (11.134a)$$

$$\left\{ \begin{array}{l} (1 - \varepsilon^2)r_0^3 - 2mr_0^2 + [\ell^2 + a^2(1 - \varepsilon^2)]r_0 - 2m(\ell - a\varepsilon)^2 = 0 \\ mr_0^2 - [\ell^2 + a^2(1 - \varepsilon^2)]r_0 + 3m(\ell - a\varepsilon)^2 = 0. \end{array} \right. \quad (11.134b)$$

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 (11.134) for (ε, ℓ) . The change of variable

$$\tilde{\ell} := \ell - a\varepsilon \quad (11.135)$$

turns (11.134) into the system

$$\left\{ \begin{array}{l} r_0^3 \varepsilon^2 = m\tilde{\ell}^2 + r_0^2(r_0 - m) \\ 2ar_0 \varepsilon \tilde{\ell} = (3m - r_0)\tilde{\ell}^2 + r_0(mr_0 - a^2). \end{array} \right. \quad (11.136a)$$

$$\left\{ \begin{array}{l} r_0^3 \varepsilon^2 = m\tilde{\ell}^2 + r_0^2(r_0 - m) \\ 2ar_0 \varepsilon \tilde{\ell} = (3m - r_0)\tilde{\ell}^2 + r_0(mr_0 - a^2). \end{array} \right. \quad (11.136b)$$

Let us consider the square of Eq. (11.136b):

$$4a^2r_0^2 \varepsilon^2 \tilde{\ell}^2 = \left[(3m - r_0)\tilde{\ell}^2 + r_0(mr_0 - a^2) \right]^2, \quad (11.137)$$

and substitute (11.136a) for ε^2 in it; we get

$$[r_0(r_0 - 3m)^2 - 4a^2m] \tilde{\ell}^4 + 2r_0^2 [(3m^2 - a^2)r_0 - m(r_0^2 + a^2)] \tilde{\ell}^2 + r_0^3(mr_0 - a^2)^2 = 0.$$

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 $r_0 < 0$:

²⁰See Ref. [488] for details.

Property 11.28

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. 11.3.8 show that if a circular orbit would exist there, it would necessarily have $|E| > \mu$ [cf. Eq. (11.96)], 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

$$r_0 > 0. \quad (11.138)$$

We have then $\sqrt{\Delta} = 4ar_0\sqrt{mr_0}|r_0^2 - 2mr_0 + a^2|$ and the solutions of the quadratic equation are²¹

$$\tilde{\ell}_\pm^2 = r_0 \frac{(a^2 - 3m^2)r_0^2 + mr_0(r_0^2 + a^2) \mp 2a\sqrt{mr_0}(r_0^2 - 2mr_0 + a^2)}{r_0(r_0 - 3m)^2 - 4a^2m}. \quad (11.139)$$

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 (11.139), we may use the identity $\mp 2a\sqrt{mr_0} = A_\mp - r_0(r_0 - 3m)$ and get, after simplification,

$$\tilde{\ell}_\pm^2 = \frac{r_0^2}{A_+ A_-} [A_\mp(r_0^2 - 2mr_0 + a^2) - A_+ A_-] = \frac{r_0^2}{A_\pm} (r_0^2 - 2mr_0 + a^2 - A_\pm).$$

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

$$\tilde{\ell}_\pm^2 = \frac{r_0^2(a \mp \sqrt{mr_0})^2}{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}. \quad (11.140)$$

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

$$r_0^{3/2} - 3mr_0^{1/2} \pm 2a\sqrt{m} > 0. \quad (11.141)$$

²¹For future convenience, we have chosen the sign \mp in the numerator to be the opposite of the label \pm in $\tilde{\ell}_\pm^2$.

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 $x_0 x_1 x_2 = -q = \mp 2a\sqrt{m}$, from which we deduce that for $\pm = +$ in Eq. (11.141), 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:

$$\text{for } \tilde{\ell}_+^2 : \quad 0 < r_0 < r_{\text{ph}}^* \quad \text{or} \quad r_0 > r_{\text{ph}}^+ \quad (11.142\text{a})$$

$$\text{for } \tilde{\ell}_-^2 : \quad r_0 > r_{\text{ph}}^-, \quad (11.142\text{b})$$

with

$$r_{\text{ph}}^\pm := 4m \cos^2 \left[\frac{1}{3} \arccos \left(\mp \frac{a}{m} \right) \right] \quad (11.143)$$

and

$$r_{\text{ph}}^* := 4m \cos^2 \left[\frac{1}{3} \arccos \left(-\frac{a}{m} \right) + \frac{4\pi}{3} \right]. \quad (11.144)$$

The index “ph” stands for *photon*, because we shall see in Sec. 11.5.5 and in Sec. 12.3 that these radii actually correspond to photon orbits (circular null geodesics). r_{ph}^\pm and r_{ph}^* are plotted as functions of a in Fig. 11.11 (green curves). Note that $m \leq r_{\text{ph}}^+ \leq 3m$ and $3m \leq r_{\text{ph}}^- \leq 4m$, with

$$\lim_{a \rightarrow 0} r_{\text{ph}}^\pm = 3m, \quad \lim_{a \rightarrow m} r_{\text{ph}}^+ = m, \quad \lim_{a \rightarrow m} r_{\text{ph}}^- = 4m. \quad (11.145)$$

Besides, $0 \leq r_{\text{ph}}^* \leq m$, with

$$\lim_{a \rightarrow 0} r_{\text{ph}}^* = 0, \quad \lim_{a \rightarrow m} r_{\text{ph}}^* = m. \quad (11.146)$$

Furthermore, one can check that

$$r_{\text{ph}}^\pm \geq r_+ \quad (11.147)$$

and

$$0 \leq r_{\text{ph}}^* \leq r_-, \quad (11.148)$$

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 \mathcal{H} and the inner horizon \mathcal{H}_{in} . Regarding the upper bound in Eq. (11.148), one can see from Fig. 11.11 that r_{ph}^* is always very close to r_- . The maximum discrepancy is $r_- - r_{\text{ph}}^* \simeq 0.032m$ and is achieved for $a \simeq 0.9m$ (cf. the notebook D.5.9). We conclude that the first permitted range for circular orbits in (11.142a) lies entirely in the region \mathcal{M}_{III} of Kerr spacetime, while the second range in (11.142a) lies in entirely in \mathcal{M}_{I} . Regarding the orbits for $\tilde{\ell}_-^2$, the range (11.142b) lies in \mathcal{M}_{I} as well.

Substituting expression (11.140) for $\tilde{\ell}^2$ in Eq. (11.136a), we get, after simplification,

$$\varepsilon_\pm^2 = \frac{(r_0^2 - 2mr_0 \pm a\sqrt{mr_0})^2}{r_0^2(r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0})}. \quad (11.149)$$

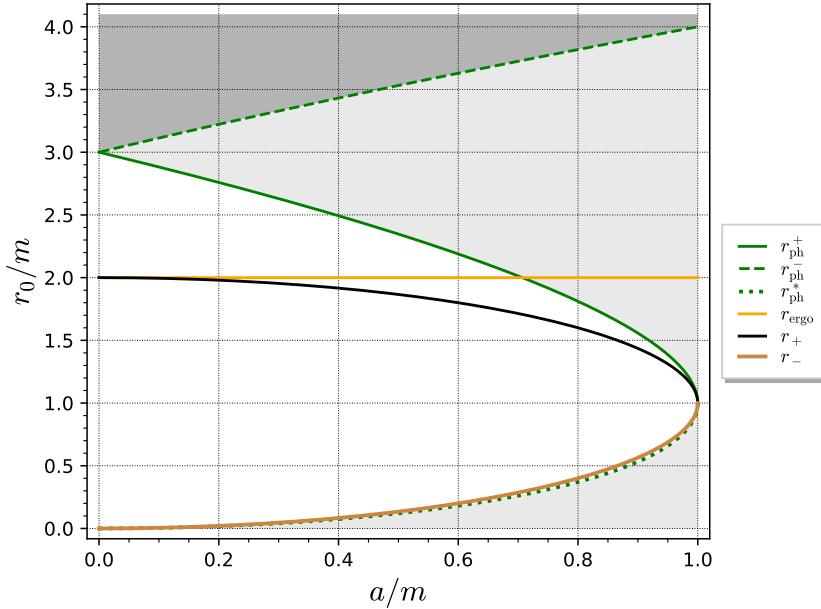


Figure 11.11: Domain of existence of circular equatorial timelike orbits in the (a, r_0) plane: orbits of the (ε_+, ℓ_+) family exist in both the light and dark grey regions, while orbits of the (ε_-, ℓ_-) family exist only in the dark grey region. The orange line marks the location of the outer ergosphere in the equatorial plane, which is $r_{\text{ergo}} = r_{\mathcal{E}^+}(\pi/2) = 2m$ [cf. Eq. (10.19b)], the black curve corresponds to the black hole event horizon \mathcal{H} ($r = r_+$) and the maroon one to the inner horizon \mathcal{H}_{in} ($r = r_-$). [Figure generated by the notebook D.5.9]

The \pm 's in Eqs. (11.140) and (11.149) 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 (11.137) but not all of them satisfy the original equation (11.136b). To see this, let us consider the following square roots of Eqs. (11.149) and (11.140):

$$\varepsilon_{\pm} = \frac{r_0^2 - 2mr_0 \pm a\sqrt{mr_0}}{r_0\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}}, \quad (11.150)$$

$$\tilde{\ell}_{\pm} = -\frac{r_0(a \mp \sqrt{mr_0})}{\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}} = \frac{r_0(\pm\sqrt{mr_0} - a)}{\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}} \quad (11.151)$$

and write the eight solutions of (11.137) as

$$(\varepsilon, \tilde{\ell}) = (\epsilon_1 \varepsilon_{\pm}, \epsilon_2 \tilde{\ell}_{\pm}), \quad \text{with } \epsilon_1 = \pm 1 \quad \text{and} \quad \epsilon_2 = \pm 1.$$

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. (11.136b) 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:

$$(\varepsilon, \tilde{\ell}) = (\varepsilon_+, \tilde{\ell}_+) \quad \text{or} \quad (-\varepsilon_+, -\tilde{\ell}_+) \quad \text{or} \quad (\varepsilon_-, \tilde{\ell}_-) \quad \text{or} \quad (-\varepsilon_-, -\tilde{\ell}_-). \quad (11.152)$$

A further reduction of the number of solutions is provided by the future-directed condition (11.66). Since $R(r_0) = -\mu^2 r_0^4 \mathcal{V}(r_0) = 0$ by virtue of Eq. (11.132), for circular equatorial orbits, Eq. (11.66) 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 \mathcal{M}_I or \mathcal{M}_{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

$$r_0^2 \varepsilon - a\tilde{\ell} > 0. \quad (11.153)$$

Let us check each of the four solutions (11.152):

$$\begin{aligned} r_0^2 \varepsilon_+ - a\tilde{\ell}_+ &= \frac{r_0 \Delta_0}{\sqrt{r_0^2 - 3mr_0 + 2a\sqrt{mr_0}}} > 0 \\ r_0^2 (-\varepsilon_+) - a(-\tilde{\ell}_+) &= -(r_0^2 \varepsilon_+ - a\tilde{\ell}_+) < 0 \\ r_0^2 \varepsilon_- - a\tilde{\ell}_- &= \frac{r_0 \Delta_0}{\sqrt{r_0^2 - 3mr_0 - 2a\sqrt{mr_0}}} > 0 \\ r_0^2 (-\varepsilon_-) - a(-\tilde{\ell}_-) &= -(r_0^2 \varepsilon_- - a\tilde{\ell}_-) < 0. \end{aligned}$$

We conclude that only two solutions remain:

$$(\varepsilon, \tilde{\ell}) = (\varepsilon_+, \tilde{\ell}_+) \quad \text{or} \quad (\varepsilon_-, \tilde{\ell}_-). \quad (11.154)$$

Given that $\ell = \tilde{\ell} + a\varepsilon$ [Eq. (11.135)] and expressions (11.150) and (11.151) for respectively ε_\pm and $\tilde{\ell}_\pm$, these two solutions can be reexpressed in terms of (ε, ℓ) as

$$(\varepsilon, \ell) = \boxed{(\varepsilon_+, \ell_+)} \quad \text{or} \quad \boxed{(\varepsilon_-, \ell_-)}, \quad (11.155)$$

where

$$\boxed{\ell_\pm = \pm \sqrt{\frac{m}{r_0} \frac{r_0^2 + a^2 \mp 2a\sqrt{mr_0}}{\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}}}}. \quad (11.156)$$

The quantities ε_\pm and ℓ_\pm are plotted in terms of r_0 in Figs. 11.12 and 11.13. 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 $r_0^2 + a^2 - 2a\sqrt{mr_0}$. Outside the event horizon, i.e. in \mathcal{M}_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

$$r_0 > r_{\text{ph}}^+ \implies \ell_+ > 0. \quad (11.157)$$

But for $r_0 < r_{\text{ph}}^*$, i.e. in region \mathcal{M}_{III} , we may have $\ell_+ < 0$ (cf. Fig. 11.13). In view of this and the result (11.142), we introduce the following nomenclature:

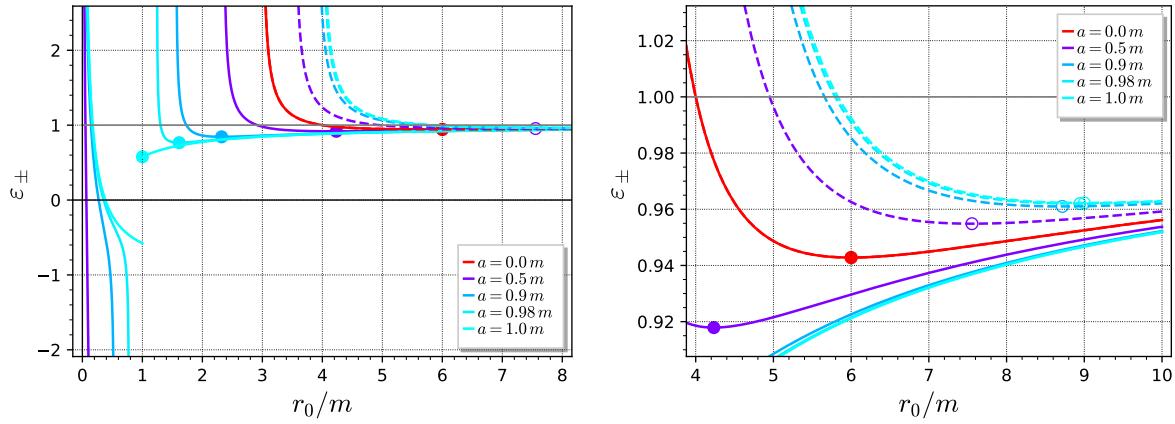


Figure 11.12: Specific conserved energy $\varepsilon = \varepsilon_+$ (solid curves) or $\varepsilon = \varepsilon_-$ (dashed curves) along circular timelike orbits in the equatorial plane, as a function of the orbital radius r_0 [Eq. (11.150)], for selected values of the Kerr spin parameter a . Curves in the inner region ($r_0 < m$) terminate by a vertical asymptote at $r = r_{\text{ph}}^*(a)$ given by Eq. (11.144), while the other curves start along a vertical asymptote at $r = r_{\text{ph}}^\pm(a)$ given by Eq. (11.143). Dots on ε_+ curves and open circles on ε_- ones mark the ISCO: all configurations on the left of these points are unstable. The right panel is a zoom on the region $4m \leq r_0 \leq 10m$. Note that the red curve ($a = 0$) is the same as in Fig. 7.8. [Figure generated by the notebook D.5.9]

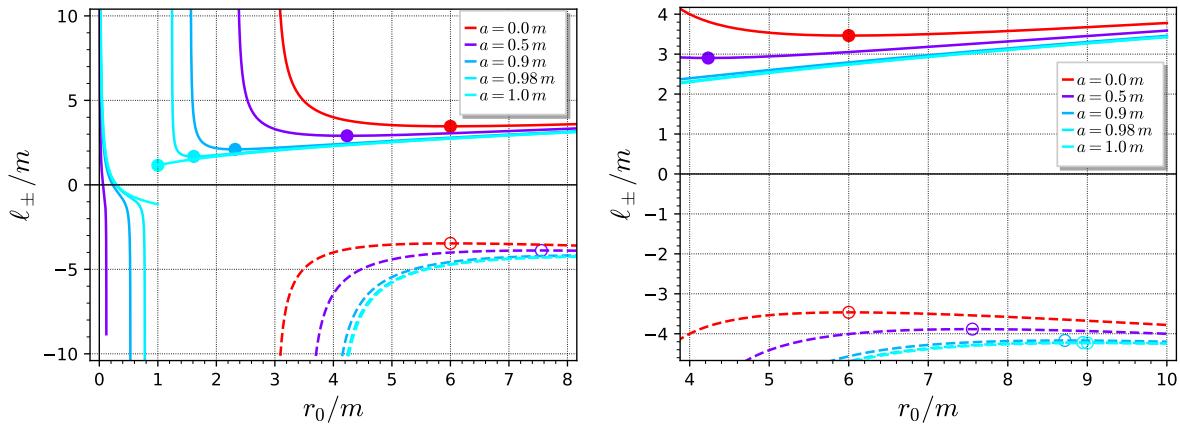


Figure 11.13: Specific conserved angular momentum $\ell = \ell_+$ (solid curves) or $\ell = \ell_-$ (dashed curves) along circular timelike orbits in the equatorial plane, as a function of the orbital radius r_0 [Eq. (11.156)], for selected values of the Kerr spin parameter a . Curves in the inner region ($r_0 < m$) terminate by a vertical asymptote at $r = r_{\text{ph}}^*(a)$ given by Eq. (11.144), while the other curves start along a vertical asymptote at $r = r_{\text{ph}}^\pm(a)$ given by Eq. (11.143). Dots on ℓ_+ curves and open circles on ℓ_- ones mark the ISCO: all configurations on the left of these points are unstable. The right panel is a zoom on the region $4m \leq r_0 \leq 10m$. [Figure generated by the notebook D.5.9]

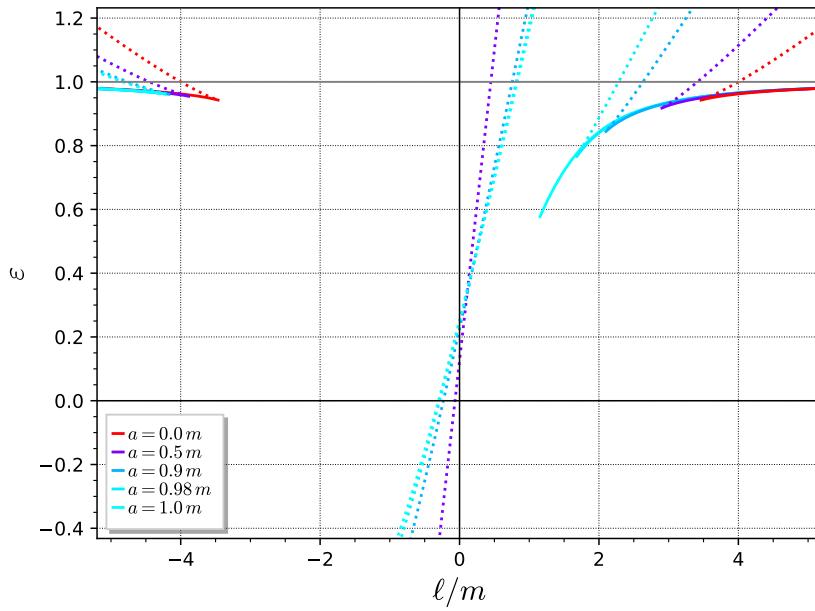


Figure 11.14: Circular timelike orbits in the (ℓ, ε) plane, ℓ being the specific conserved angular momentum and ε the specific conserved energy. Solid (resp. dotted) curves correspond to stable (unstable) orbits. Retrograde outer circular orbits are on the left side, inner circular orbits in the middle 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.5.9]

- orbits of the (ε_+, ℓ_+) family with $r_0 > r_{\text{ph}}^+$ are called ***prograde outer circular orbits***;
- orbits of the (ε_+, ℓ_+) family with $0 < r_0 < r_{\text{ph}}^*$ are called ***inner circular orbits***;
- all orbits of the (ε_-, ℓ_-) family are called ***retrograde outer circular orbits***.

We note from Figs. 11.12 and 11.13, or from formulas (11.150) and (11.156), 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. 11.5.5.

Figure 11.14 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:

Property 11.29: latitudinal stability of equatorial circular timelike orbits

All equatorial circular timelike orbits are stable with respect to any perturbation away from the equatorial plane.

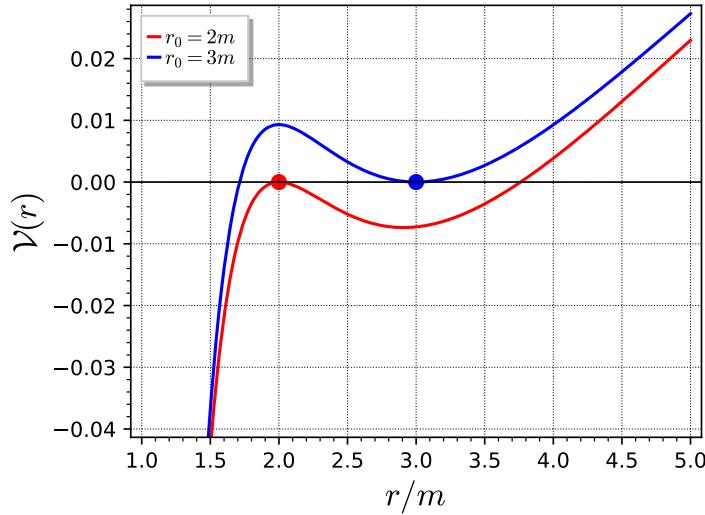


Figure 11.15: Function $\mathcal{V}(r)$, as defined by Eq. (11.130), with $a = 0.9\text{ m}$, $\varepsilon = \varepsilon_+(r_0)$ [Eq. (11.150)], $\ell = \ell_+(r_0)$ [Eq. (11.156)] for $r_0 = 2\text{ m}$ (red curve) and $r_0 = 3\text{ m}$ (blue curve). By definition of ε_+ and ℓ_+ , one has both $\mathcal{V}(r_0) = 0$ and $\mathcal{V}'(r_0) = 0$. [Figure generated by the notebook D.5.9]

Proof. All equatorial orbits have a vanishing Carter constant Q [cf. Eq. (11.128)] and we have seen in Sec. 11.3.7 that a $Q = 0$ geodesic stays stably at $\theta = \pi/2$ iff $a^2(E^2 - \mu^2) \leq L^2$, i.e. iff

$$a^2(\varepsilon^2 - 1) \leq \ell^2 \quad (11.158)$$

for a timelike geodesic. Now, for a circular orbit of radius r_0 , Eq. (11.134b) 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. (11.138)]. Hence the stability condition (11.158) is always fulfilled. \square

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. (11.132) and (11.133)]. 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. 11.15) or a local minimum (blue curve in Fig. 11.15). Now, in view of Eq. (11.129b) rewritten as $\mathcal{V}(r) = -(dr/d\tau)^2$, 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. 11.15). 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. 11.15). 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

Property 11.30: criterion for radial stability of circular orbits

$$\text{The circular orbit of radius } r_0 \text{ is stable} \iff \mathcal{V}''(r_0) > 0. \quad (11.159)$$

From the definition (11.130) of $\mathcal{V}(r)$ and the relation $\ell = \tilde{\ell} + a\varepsilon$ [Eq. (11.135)], 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. (11.136b) 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 (11.151) 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. (11.138) and (11.141), the sign of $\mathcal{V}''(r_0)$ is entirely determined by the numerator, so that we may rewrite (11.159) as

$$\begin{aligned} \text{stable orbit (radius } r_0) &\iff r_0^2 - 6mr_0 \pm 8a\sqrt{mr_0} - 3a^2 > 0 \\ &\iff P(x) := x^4 - 6x^2 \pm 8\bar{a}x - 3\bar{a}^2 > 0, \end{aligned} \quad (11.160)$$

where \pm is $+$ (resp. $-$) for prograde (resp. retrograde) orbits and we have introduced the dimensionless variables

$$x := \sqrt{\frac{r_0}{m}} \quad \text{and} \quad \bar{a} := \frac{a}{m}. \quad (11.161)$$

The problem amounts to finding the range of x where the quartic polynomial $P(x)$ 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

$$P(x) = (x^2 - Z_1)^2 - T(x), \quad \text{with} \quad T(x) := 2(3 - Z_1)x^2 \mp 8\bar{a}x + Z_1^2 + 3\bar{a}^2. \quad (11.162)$$

The above expression is an identity, which holds for any value of Z_1 ; the core of Ferrari's method is 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

$$P(x) = 0 \iff x^2 - Z_1 - S(x) = 0 \quad \text{or} \quad x^2 - Z_1 + S(x) = 0. \quad (11.163)$$

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

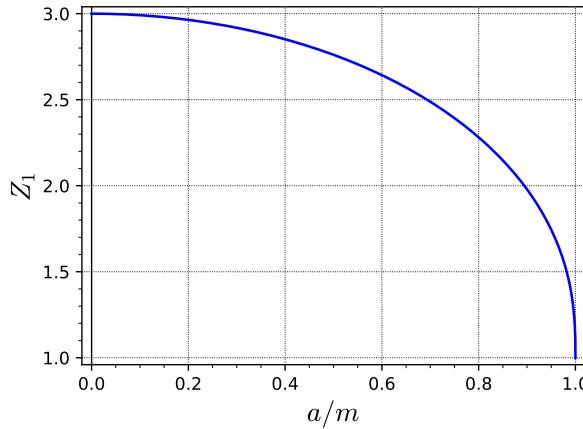


Figure 11.16: Function Z_1 , as defined by Eq. (11.165). [Figure generated by the notebook D.5.9]

$x^2 - Z_1 + S(x) = 0$. A necessary and sufficient condition for $T(x)$ to have a double root is that its discriminant vanishes, which is equivalent to

$$Z_1^3 - 3Z_1^2 + 3\bar{a}^2Z_1 - \bar{a}^2 = 0. \quad (11.164)$$

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

$$Z_1 = 1 + \sqrt[3]{1 - \bar{a}^2} \left(\sqrt[3]{1 + \bar{a}} + \sqrt[3]{1 - \bar{a}} \right). \quad (11.165)$$

Z_1 is plotted in terms of $\bar{a} = a/m$ in Fig. 11.16. In particular, we notice that $1 \leq Z_1 \leq 3$. When Z_1 takes the value (11.165), 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 (11.162) 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 $\bar{a} \rightarrow 0$, 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. (11.164). 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. (11.163)]

$$x^2 - \sqrt{2(3 - Z_1)} x - Z_1 \pm Z_2 = 0 \quad (11.166a)$$

$$x^2 + \sqrt{2(3 - Z_1)} x - Z_1 \mp Z_2 = 0. \quad (11.166b)$$

Moreover, physically acceptable solutions must obey $x > 0$ (recall that $x := \sqrt{r_0/m}$). The discriminant of (11.166a) 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 (11.166a) 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 (11.166b) 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 (11.166b) 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 $(0, +\infty)$:

$$\boxed{r_{\text{ISCO}}^\pm = m \left[3 + Z_2 \mp \sqrt{(3 - Z_1)(3 + Z_1 + 2Z_2)} \right]}, \quad (11.167)$$

$$Z_1 := 1 + \sqrt[3]{1 - \bar{a}^2} \left(\sqrt[3]{1 + \bar{a}} + \sqrt[3]{1 - \bar{a}} \right); \quad Z_2 := \sqrt{Z_1^2 + 3\bar{a}^2}; \quad \bar{a} := a/m.$$

where \pm indicates which family among (ϵ_+, ℓ_+) and (ϵ_-, ℓ_-) is considered and *ISCO* stands for **innermost stable circular orbit**. Indeed, r_{ISCO}^\pm being the unique zero of $r_0^2 - 6mr_0 \pm 8a\sqrt{mr_0} - 3a^2$, (11.160) is equivalent to

Property 11.31: stability of circular orbits

$$\text{A circular orbit of radius } r_0 \text{ is radially stable} \iff r_0 > r_{\text{ISCO}}^\pm. \quad (11.168)$$

Note that r_{ISCO}^\pm/m , as given by Eq. (11.167), is a function of $\bar{a} := a/m$ only; this function is depicted in Fig. 11.17. We notice immediately from it that

$$r_{\text{ph}}^* < r_{\text{ISCO}}^+ \quad (11.169)$$

and conclude that all the inner circular orbits are unstable. We also see on Fig. 11.17 that in the limit $a = 0$, $r_{\text{ISCO}}^+ = r_{\text{ISCO}}^- = 6m$, i.e. we recover the Schwarzschild ISCO discussed in Sec. 7.3.3.

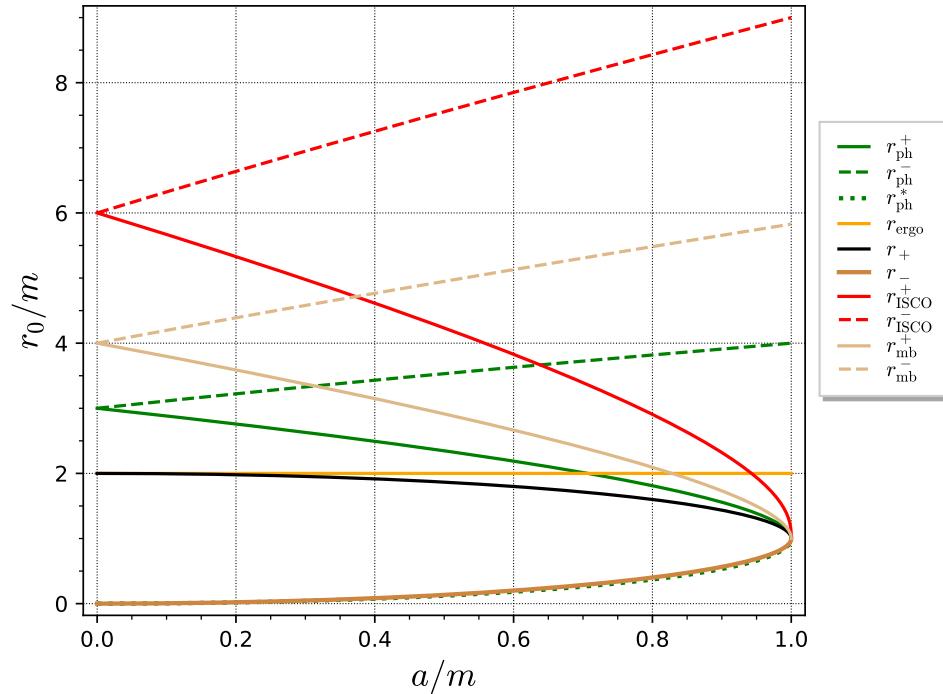


Figure 11.17: Various critical radii for circular orbits as functions of the Kerr spin parameter a . The red curves correspond to the innermost stable circular orbit r_{ISCO}^{\pm} [Eq. (11.167)], the light brown ones to the marginally bound circular orbit r_{mb}^{\pm} [Eq. (11.186)], while the other curves are the same as in Fig. 11.11. [Figure generated by the notebook D.5.9]

The function $r_{\text{ISCO}}^{\pm}(\bar{a})$ is easily inverted by solving the quadratic equation $(r_{\text{ISCO}}^{\pm})^2 - 6mr_{\text{ISCO}}^{\pm} \pm 8a\sqrt{mr_{\text{ISCO}}^{\pm}} - 3a^2 = 0$ for a [cf. Eq. (11.160)]; one gets

$$a = \pm \frac{m}{3} \sqrt{\frac{r_{\text{ISCO}}^{\pm}}{m}} \left(4 - \sqrt{3 \frac{r_{\text{ISCO}}^{\pm}}{m} - 2} \right), \quad (11.170)$$

where the overall factor \pm must be the same as in²² r_{ISCO}^{\pm} .

Remark 4: The ISCO is also called ***marginally stable circular orbit*** by some authors (e.g. [41]), r_{ISCO} is then denoted by r_{ms} .

Example 13: For $a = 0.9 m$, Eq. (11.167) yields $r_{\text{ISCO}}^+ = 2.32088 m$. Accordingly, the prograde circular orbit at $r_0 = 3 m$ is stable, while that at $r_0 = 2 m$ is unstable, in agreement with the plot of $\mathcal{V}(r)$ in Fig. 11.15.

An interesting property of the ISCO is that it corresponds to extrema of $\varepsilon_{\pm}(r_0)$ and $\ell_{\pm}(r_0)$:

²²Note that for $\pm = -$, formula (11.170) still yields $a \geq 0$, since $r_{\text{ISCO}}^- \geq 6m$.

Property 11.32: characterization of the ISCO

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.

Proof. In what precedes, we have considered \mathcal{V} , defined by Eq. (11.130), 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. (11.132) and (11.133) can be rewritten as

$$\mathcal{V}(r_0, \varepsilon_\pm(r_0), \ell_\pm(r_0)) = 0 \quad (11.171a)$$

$$\frac{\partial\mathcal{V}}{\partial r}(r_0, \varepsilon_\pm(r_0), \ell_\pm(r_0)) = 0. \quad (11.171b)$$

These equations are valid for any circular orbit, i.e. any value of r_0 . Let us take the derivative of Eq. (11.171a) with respect to r_0 . Using the chain rule, we get

$$\underbrace{\frac{\partial\mathcal{V}}{\partial r}\Big|_0}_{=0} + \frac{\partial\mathcal{V}}{\partial \varepsilon}\Big|_0 \varepsilon'_\pm(r_0) + \frac{\partial\mathcal{V}}{\partial \ell}\Big|_0 \ell'_\pm(r_0) = 0,$$

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. (11.171b). Similarly, deriving Eq. (11.171b) with respect to r_0 yields

$$\frac{\partial^2\mathcal{V}}{\partial r^2}\Big|_0 + \frac{\partial^2\mathcal{V}}{\partial \varepsilon \partial r}\Big|_0 \varepsilon'_\pm(r_0) + \frac{\partial^2\mathcal{V}}{\partial \ell \partial r}\Big|_0 \ell'_\pm(r_0) = 0.$$

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:

$$\begin{cases} \frac{\partial\mathcal{V}}{\partial \varepsilon}\Big|_0 \varepsilon'_\pm(r_0) + \frac{\partial\mathcal{V}}{\partial \ell}\Big|_0 \ell'_\pm(r_0) = 0 \\ \frac{\partial^2\mathcal{V}}{\partial \varepsilon \partial r}\Big|_0 \varepsilon'_\pm(r_0) + \frac{\partial^2\mathcal{V}}{\partial \ell \partial r}\Big|_0 \ell'_\pm(r_0) = 0 \end{cases} \quad (11.172)$$

This constitutes a linear homogeneous system for the two unknowns $(\varepsilon'_\pm(r_0), \ell'_\pm(r_0))$. Since there are no obvious reason for its determinant to vanish, we deduce that the only solution is $(\varepsilon'_\pm(r_0), \ell'_\pm(r_0)) = (0, 0)$. Hence the ISCO realizes an extremum of both $\varepsilon_\pm(r_0)$ and $\ell_\pm(r_0)$. A quick look at Figs. 11.12 and 11.13, 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 $\ell_-(r_0)$. \square

Property 11.32 explains why the ISCO is located at the cusp in the (ℓ, ε) curves shown in Fig. 11.14: this corresponds to an extremum of both $\varepsilon(r_0)$ and $\ell(r_0)$. Property 11.32 has also a particular astrophysical significance, to be discussed in Sec. 11.5.4.

Summary

Property 11.33: existence and stability of equatorial circular timelike orbits

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 \mathcal{M}_1), having a Boyer-Lindquist radius r_0 ranging from $r_{\text{ph}}^+ = 4m \cos^2[\arccos(-a/m)/3]$ [Eq. (11.143)] to $+\infty$; the minimal radius r_{ph}^+ decreases with a monotonically from $3m$ ($a = 0$) to m ($a = m$). These orbits are radially unstable for $r_{\text{ph}}^+ < r_0 \leq r_{\text{ISCO}}^+$ and stable for $r_0 > r_{\text{ISCO}}^+$, where r_{ISCO}^+ is given by Eq. (11.167), 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. (11.150) and $\ell_+(r_0)$ by Eq. (11.156), both being plotted as the solid curves in the region $r_0 \geq m$ of Figs. 11.12 and 11.13. 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 \mathcal{M}_1), having a Boyer-Lindquist radius r_0 ranging from $r_{\text{ph}}^- = 4m \cos^2[\arccos(a/m)/3]$ [Eq. (11.143)] to $+\infty$; the minimal radius r_{ph}^- increases with a monotonically from $3m$ ($a = 0$) to $4m$ ($a = m$). These orbits are radially unstable for $r_{\text{ph}}^- < r_0 \leq r_{\text{ISCO}}^-$ and stable for $r_0 > r_{\text{ISCO}}^-$, where r_{ISCO}^- is given by Eq. (11.167), 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. (11.150) and $\ell_-(r_0)$ by Eq. (11.156), both being plotted as the dashed curves in Figs. 11.12 and 11.13. 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 \mathcal{M}_{III}), having a Boyer-Lindquist radius r_0 ranging from 0 (the ring singularity) to $r_{\text{ph}}^* = 4m \cos^2[\arccos(-a/m)/3 + 4\pi/3]$ [Eq. (11.144)]; the maximal radius r_{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 Eq. (11.150) and $\ell_+(r_0)$ by Eq. (11.156), both being plotted as the solid curves in the region $r_0 \leq m$ of Figs. 11.12 and 11.13.

In particular, there are no equatorial circular orbits in region \mathcal{M}_{II} of Kerr spacetime.

Historical note : The solutions for timelike circular orbits in the equatorial plane (outer cases only, i.e. excluding the case 3 above) have been first published by James M. Bardeen, William H. Press and Saul A. Teukolsky in 1972 [41]. In particular, they have obtained Eq. (11.150) for $\varepsilon_{\pm}(r_0)$, Eq. (11.156) for $\ell_{\pm}(r_0)$ and Eq. (11.167) for r_{ISCO}^{\pm} . In his lecture notes at the 1972 Les Houches School [37], Bardeen says that these solutions have been derived by Teukolsky. In the article [41], it is mentioned that they have been obtained by means of computer algebra techniques.

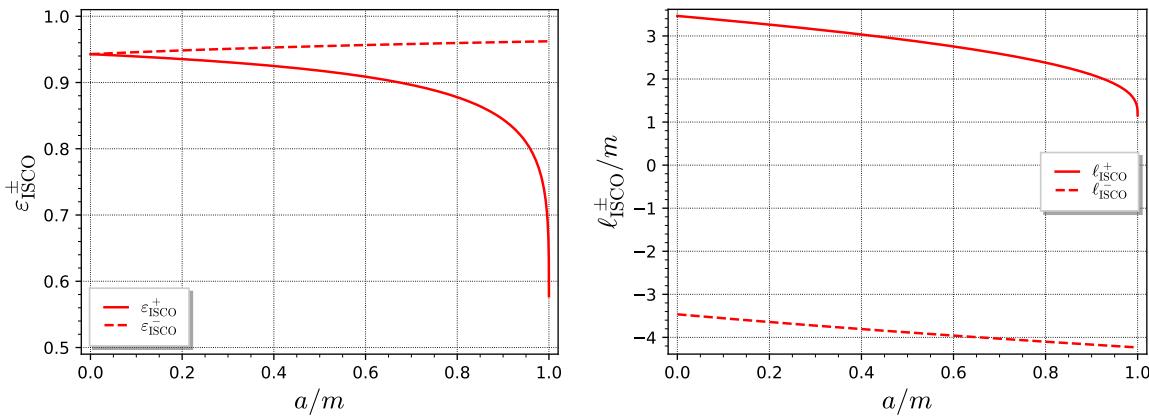


Figure 11.18: Specific conserved energy ε (left) and specific conserved angular momentum ℓ (right) at the prograde ISCO (solid curve) and retrograde ISCO (dashed curve), as functions of the Kerr spin parameter a . [Figure generated by the notebook D.5.9]

11.5.4 Radiated energy from an accretion disk around a black hole

The prograde ISCO around a Kerr black hole plays an important role in astrophysics: infall of matter onto black holes proceeds via an accretion disk (cf. Sec. 8.5.5) and the inner edge of the disk is located at the ISCO, either exactly (low luminosity thin disk) or very close to it [2, 315, 430]. Matter slowly inspirals in the accretion disk because it loses energy and angular momentum by the friction resulting from the differential rotation in the disk. This friction heats the disk at high temperatures and the energy is taken away by electromagnetic radiation. At a given orbit in the disk, the conserved energy of a particle of mass μ is $E = \mu\varepsilon$ and the difference $E_{\text{rad}} = |E - \mu| = \mu(1 - \varepsilon)$ can be considered as the energy radiated away to reach this orbit starting from rest far from the black hole, where its energy was $E_\infty = \mu$ ($\mu - E$ is generally called the *binding energy*, cf. Property 11.20). According to Property 11.32, E_{rad} is maximal when the particle has reached the prograde ISCO. Let us then evaluate ε there.

The values of ε and ℓ at the ISCO depend on a (cf. Fig. 11.18); they take a simple form when expressed in terms of r_{ISCO}^\pm , instead of directly a :

Property 11.34: specific conserved energy and angular momentum at the ISCO

The values $\varepsilon_{\text{ISCO}}^+$ and ℓ_{ISCO}^+ (resp. $\varepsilon_{\text{ISCO}}^-$ and ℓ_{ISCO}^-) of the specific conserved energy and angular momentum of a particle at the prograde ISCO (resp. retrograde ISCO) are

$$\varepsilon_{\text{ISCO}}^\pm = \sqrt{1 - \frac{2m}{3r_{\text{ISCO}}^\pm}} \quad \text{and} \quad \ell_{\text{ISCO}}^\pm = \pm \frac{2m}{3\sqrt{3}} \left(2\sqrt{3\frac{r_{\text{ISCO}}^\pm}{m} - 2} + 1 \right). \quad (11.173)$$

Proof. Let us set $u := r_{\text{ISCO}}^\pm/m$, so that Eq. (11.170) becomes $a = \pm m\sqrt{u}(4 - \sqrt{3u - 2})/3$.

Substituting this value for a in Eq. (11.150) yields

$$\varepsilon_{\text{ISCO}}^{\pm} = \frac{3u - 2 - \sqrt{3u - 2}}{\sqrt{3u}\sqrt{3u - 1 - 2\sqrt{3u - 2}}}.$$

Noticing that $3u - 1 - 2\sqrt{3u - 2} = (\sqrt{3u - 2} - 1)^2$ leads to the first formula in Eq. (11.173). Similarly, Eq. (11.156) becomes

$$\ell_{\text{ISCO}}^{\pm} = \pm \frac{2m}{3\sqrt{3}} \frac{6u - 5 - \sqrt{3u - 2}}{\sqrt{3u - 1 - 2\sqrt{3u - 2}}} = \pm \frac{2m}{3\sqrt{3}} \frac{6u - 5 - \sqrt{3u - 2}}{\sqrt{3u - 2} - 1}.$$

Multiplying both the numerator and denominator by $\sqrt{3u - 2} + 1$ and simplifying yields the second formula in Eq. (11.173). \square

For $r_{\text{ISCO}}^{\pm} = 6m$ (i.e. $a = 0$), formula (11.173) yields

$$\varepsilon_{\text{ISCO}}^{\pm} = \frac{2\sqrt{2}}{3} \simeq 0.943 \quad \text{and} \quad \ell_{\text{ISCO}}^{\pm} = \pm 2\sqrt{3}m \simeq 3.46m \quad (a = 0), \quad (11.174)$$

while for $r_{\text{ISCO}}^{+} = m$ and $r_{\text{ISCO}}^{-} = 9m$ (i.e. $a = m$), formula (11.173) yields

$$\varepsilon_{\text{ISCO}}^{+} = \frac{1}{\sqrt{3}} \simeq 0.577 \quad \text{and} \quad \ell_{\text{ISCO}}^{+} = \frac{2m}{\sqrt{3}} \simeq 1.15m \quad (a = m) \quad (11.175a)$$

$$\varepsilon_{\text{ISCO}}^{-} = \frac{5}{3\sqrt{3}} \simeq 0.962 \quad \text{and} \quad \ell_{\text{ISCO}}^{-} = -\frac{22m}{3\sqrt{3}} \simeq -4.23m \quad (a = m). \quad (11.175b)$$

For $a = 0$, we recover the value of $\varepsilon_{\text{ISCO}}$ obtained in Chap. 7 for the Schwarzschild metric [Eq. (7.56)], while the value of $|\ell_{\text{ISCO}}^{\pm}|$ coincides with the value ℓ_{crit} given by Eq. (7.32), as it should. For $a = m$, the values of $\varepsilon_{\text{ISCO}}^{\pm}$ and ℓ_{ISCO}^{\pm} are the extremal ones among all values of a , as it appears clearly on Fig. 11.18.

From the numerical values (11.174)-(11.175a), the energy per unit mass radiated away to reach the prograde ISCO, i.e. $E_{\text{rad}}/\mu = 1 - \varepsilon_{\text{ISCO}}^{+}$, is 0.057 for $a = 0$ and 0.423 for $a = m$. This leads us to the important conclusion:

Property 11.35: energy release from an accretion disk around a black hole

Accretion onto a black hole through an accretion disk releases a significant fraction of the rest mass energy of matter: from 6% for a non-rotating black hole (Schwarzschild) up to 42% for a maximally rotating black hole (extremal Kerr):

$$E_{\text{rad}} \simeq 0.057\mu \quad (a = 0) \quad (11.176a)$$

$$E_{\text{rad}} \simeq 0.423\mu \quad (a = m). \quad (11.176b)$$

It must be stressed that the above numbers are much larger than the energy release achieved through thermonuclear reactions: the most efficient of such reactions is hydrogen fusion — the main energy source for stars — and it releases only 0.8% of the rest mass! This explains

why accretion onto a black hole is invoked to account for the most energetic sources in the Universe, such as quasars and active galactic nuclei (AGNs) (cf. e.g. [134]).

Remark 5: The energy-release mechanism described above is distinct from the Penrose process discussed in Sec. 11.3.2. In the former, the energy is extracted from the thermal energy of the disk, while in the latter, it is extracted from the “rotational energy” of the black hole, as we shall discuss in Sec. 16.3.1, and this requires the existence of an ergoregion, contrary to the disk accretion process.

Historical note : In 1969, Donald Lynden-Bell [350] suggested that the large luminosity of quasars is due to an accretion disk surrounding a massive black hole and terminating at the ISCO. He focused on Schwarzschild black holes, for which the radiated energy per unit mass is $E_{\text{rad}}/\mu = 0.057$ [Eq. (11.176a)]. Soon after, in 1970, James M. Bardeen [35] considered the rotating case, i.e. Kerr black holes. He obtained expression (11.170) for a , as well as formulas (11.173) for $\varepsilon_{\text{ISCO}}^{\pm}$ and ℓ_{ISCO}^{\pm} , and he pointed out that E_{rad}/μ can reach much higher values than in the non-rotating case, up to 0.423 for $a = m$ [Eq. (11.176b)] (the number 0.432 given in Ref. [35] is a typo and should be 0.423). In 1974, Kip Thorne [477] declared that Bardeen’s article [35] is “*a landmark paper — a paper that persuaded astrophysicists to stop using the Schwarzschild metric to describe black holes and start using the Kerr metric*”.

11.5.5 4-velocity and angular velocities

The orbiting angular velocity as seen by an asymptotic inertial observer is $\Omega := d\varphi/dt|_{\mathcal{L}}$ (cf. Sec. 7.3.3 for the Schwarzschild case and Sec. 10.7.1 for the extension to Kerr spacetime). We evaluate it by combining Eqs. (11.129a) and (11.129c):

$$\Omega = \frac{d\varphi}{d\tau} \times \frac{dt}{d\tau} = \frac{\left(1 - \frac{2m}{r_0}\right)\ell + \frac{2am\varepsilon}{r_0}}{(r_0^2 + a^2)\varepsilon + \frac{2am}{r_0}(a\varepsilon - \ell)}.$$

Using $\tilde{\ell} = \ell - a\varepsilon$ [Eq. (11.135)] 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 ε_{\pm} and $\tilde{\ell}_{\pm}$ taken on a circular orbit [Eqs. (11.150) and (11.151)], we obtain, after simplification,

$$\Omega_{\pm} = \pm \frac{\sqrt{m}}{r_0^{3/2} \pm a\sqrt{m}},$$

(11.177)

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 terms of r_0 in Fig. 11.19. Note that for $r_0 \gtrsim 7m$, the effect of spin parameter a on Ω_{\pm} is hardly perceptible.

Remark 6: For $a \rightarrow 0$, Eq. (11.177) reduces to Eq. (7.60), as it should.

From the very definition (11.131) of a circular orbit in the equatorial plane, the 4-velocity \mathbf{u} along such an orbit obeys $u^r = dr/d\tau = 0$ and $u^\theta = d\theta/d\tau = 0$. Since moreover $\Omega = d\varphi/dt = u^\varphi/u^t$, $\partial_t = \xi$ and $\partial_\varphi = \eta$, we can write the 4-velocity as

$$\mathbf{u} = u^t (\xi + \Omega_{\pm} \eta).$$

(11.178)

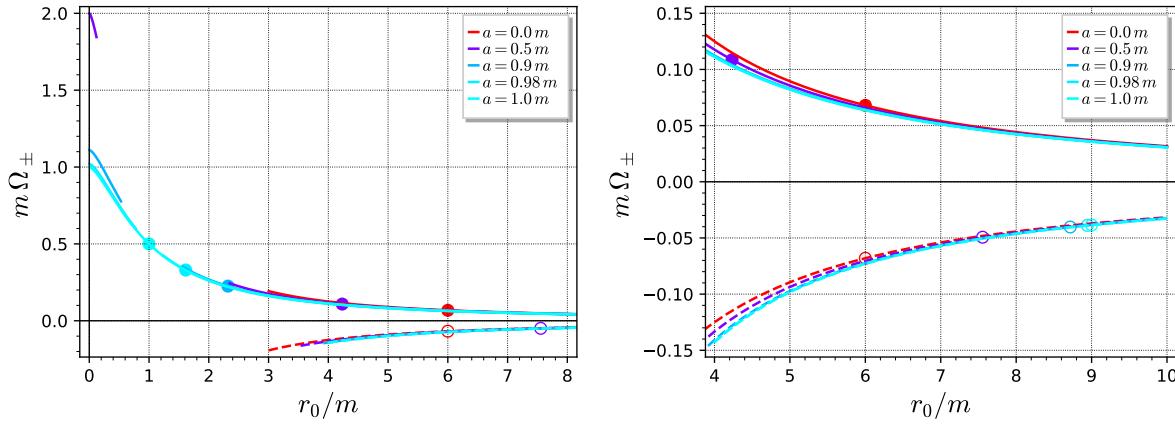


Figure 11.19: Angular velocity $\Omega := d\varphi/dt = \Omega_+$ (solid curves) or $\Omega = \Omega_-$ (dashed curves) along circular timelike orbits in the equatorial plane, as a function of the orbital radius r_0 [Eq. (11.177)], for selected values of the Kerr spin parameter a . Dots on Ω_+ curves and open circles on Ω_- ones mark the ISCO: all configurations on the left of these points are unstable. The right panel is a zoom on the region $4m \leq r_0 \leq 10m$. [Figure generated by the notebook D.5.9]

The component $u^t = dt/d\tau$ is given by Eq. (11.129a). Substituting in it Eq. (11.150) for ε and Eq. (11.151) for $\tilde{\ell} := \ell - a\varepsilon$, we get, after simplification:

$$u^t = \frac{r_0 \pm a\sqrt{m/r_0}}{\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}}. \quad (11.179)$$

Equation (11.178) shows that the 4-velocity is a linear combination of the two Killing vectors ξ and η with coefficients that are constant along the worldline \mathcal{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. (10.86)]. 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 $\Omega_{\mathcal{P}} = d\varphi/d\tau = d\varphi/dt \times dt/d\tau = u^t \Omega$. Using the values (11.177) and (11.179), we get

$$\Omega_{\mathcal{P}}^\pm = \pm \frac{\sqrt{m} \pm amr_0^{-3/2}}{(\sqrt{r_0} \pm a\sqrt{m}/r_0)\sqrt{r_0^2 - 3mr_0 \pm 2a\sqrt{mr_0}}}. \quad (11.180)$$

The orbital period measured by the circular geodesic observer is $T_{\mathcal{P}}^\pm = 2\pi/|\Omega_{\mathcal{P}}^\pm|$.

It is instructive to evaluate the velocity \mathbf{V} of a particle \mathcal{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, $\mathcal{M}_{\text{ZAMO}} = \mathcal{M}_I \cup (\mathcal{M}_{\text{III}} \setminus \mathcal{T})$ [Eq. (10.91)]. Indeed, the outer circular orbits are in \mathcal{M}_I and the inner ones are in the part $r > 0$ of \mathcal{M}_{III} , while the Carter time machine \mathcal{T} is located in the part $r < 0$ (cf. Sec. 10.2.5). Let us then rewrite formula (11.178) for \mathcal{P} 's 4-velocity \mathbf{u} by expressing ξ in terms of the ZAMO's 4-velocity \mathbf{n} , the lapse

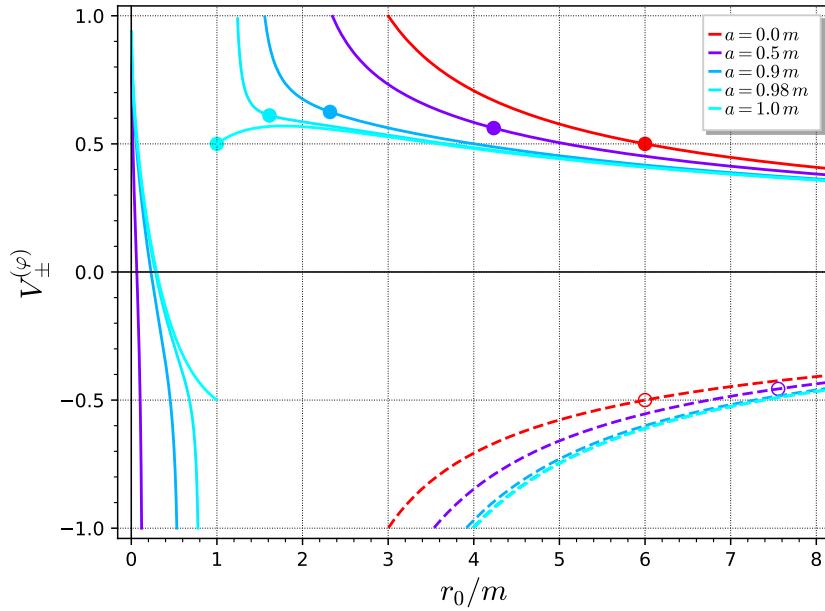


Figure 11.20: Component $V_+^{(\varphi)}$ (solid curves) or $V_-^{(\varphi)}$ (dashed curves) of the velocity $\mathbf{V} = V_\pm^{(\varphi)} e_{(\varphi)}$ of a particle moving along a circular orbit as measured by the ZAMO [Eq. (11.183)] for selected values of the Kerr spin parameter a . Dots on $V_+^{(\varphi)}$ 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.5.9]

N and the shift vector β via Eq. (10.97):

$$\mathbf{u} = u^t (N \mathbf{n} + \boldsymbol{\beta} + \Omega_\pm \boldsymbol{\eta}) = u^t [N \mathbf{n} + (\Omega_\pm - \omega) \boldsymbol{\eta}] = N u^t \left[\mathbf{n} + \frac{1}{N} (\Omega_\pm - \omega) \boldsymbol{\eta} \right],$$

where the second equality results from $\boldsymbol{\beta} = \beta^\varphi \partial_\varphi = \beta^\varphi \boldsymbol{\eta} = -\omega \boldsymbol{\eta}$, ω being the rotation angular velocity of the ZAMO seen from infinity [Eq. (10.99)]. Since $\mathbf{n} \cdot \boldsymbol{\eta} = 0$, the above formula constitute the orthogonal decomposition of \mathbf{u} with respect to the ZAMO's 4-velocity \mathbf{n} , which we can compare to the generic formula (1.36) and thereby conclude that the velocity of \mathcal{P} with respect to the ZAMO is

$$\mathbf{V} = \frac{1}{N} (\Omega_\pm - \omega) \boldsymbol{\eta} \quad (11.181)$$

and the Lorentz factor of \mathcal{P} with respect to the ZAMO is

$$\Gamma = (1 - \mathbf{V} \cdot \mathbf{V})^{-1/2} = N u^t. \quad (11.182)$$

Let us expand \mathbf{V} on the ZAMO's orthonormal fame ($e_{(\alpha)}$) [Eq. (10.95)]. Given relation (10.95d) between $\boldsymbol{\eta} = \partial_\varphi$ and $e_{(\varphi)}$ we get, after substituting Eq. (10.93) (with $\theta = \pi/2$) for N :

$$\mathbf{V} = \frac{r_0^2 + a^2 + 2a^2 m/r_0}{\sqrt{r_0^2 - 2mr_0 + a^2}} (\Omega_\pm - \omega) e_{(\varphi)}.$$

Finally, let us substitute Eq. (11.177) for Ω_\pm and Eq. (10.99) (with $\theta = \pi/2$) for ω . We obtain,

after simplification:

$$\boxed{\mathbf{V} = V_{\pm}^{(\varphi)} \mathbf{e}_{(\varphi)}} \quad \text{with} \quad \boxed{V_{\pm}^{(\varphi)} = \pm \frac{\sqrt{m} (r_0^2 \mp 2a\sqrt{mr_0} + a^2)}{(r_0^{3/2} \pm a\sqrt{m}) \sqrt{r_0^2 - 2mr_0 + a^2}}}. \quad (11.183)$$

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. 11.20. We note that, except for $a = m$,

$$\lim_{r_0 \rightarrow 0} V_{+}^{(\varphi)} = 1, \quad \lim_{r_0 \rightarrow r_{\text{ph}}^*} V_{+}^{(\varphi)} = -1, \quad \lim_{r_0 \rightarrow r_{\text{ph}}^+} V_{+}^{(\varphi)} = 1, \quad \lim_{r_0 \rightarrow r_{\text{ph}}^-} V_{-}^{(\varphi)} = -1. \quad (11.184)$$

Since $\mathbf{e}_{(\varphi)}$ 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 ($r_0 \rightarrow 0$) and near the outer boundary of their domain of existence ($r_0 \rightarrow r_{\text{ph}}^*$), 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 ($r_0 \rightarrow r_{\text{ph}}^{\pm}$).

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. 11.12 and 11.13: $\varepsilon := E/\mu$ 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. ??.

11.5.6 Marginally bound circular orbit

We see on Fig. 11.12 that for $a \neq m$ and r_0 lower than a critical value, r_{mb}^+ say, the prograde outer circular orbits have $\varepsilon > 1$ (cf. the grey horizontal line in Fig. 11.12), which is equivalent to $E > \mu$; they are thus unbound (cf. Property 11.20). Conversely, all orbits with $r_0 > r_{\text{mb}}^+$ 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_{mb}^- say. The orbit at $r_0 = r_{\text{mb}}^{\pm}$ is called the **marginally bound circular orbit**.

We note from Fig. 11.12 that

$$r_{\text{ph}}^{\pm} < r_{\text{mb}}^{\pm} < r_{\text{ISCO}}^{\pm}. \quad (11.185)$$

Accordingly, all unbound orbits are unstable.

To evaluate r_{mb}^{\pm} , let us solve the equation $\varepsilon_{\pm}(r_0) = 1$ for r_0 . Using expression (11.150) 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

$$r_{\text{mb}}^\pm = 2m \mp a + 2\sqrt{m(m \mp a)}. \quad (11.186)$$

r_{mb}^\pm is plotted as a function of a in Fig. 11.17.

11.5.7 Circular orbits in the ergoregion

We see in Fig. 11.17 that for a sufficiently large, prograde outer circular orbits can exist in the outer ergoregion \mathcal{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 \mathcal{G}^+ (the outer ergosphere) is simply $r_{\mathcal{E}^+}(\pi/2) = 2m$ [Eq. (10.19b)] – the orange horizontal line in Fig. 11.17. Thus, prograde outer circular orbits exist in the ergoregion if $r_{\text{ph}}^+ < 2m$ and they can be stable if $r_{\text{ISCO}}^+ < 2m$. 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. (11.141) with the + sign: $(2m)^{3/2} - 3m\sqrt{2m} + 2a\sqrt{m} > 0$, from which we get immediately

Property 11.36: timelike circular orbits in the ergoregion

$$\left(\begin{array}{l} \text{timelike circular orbits} \\ \text{exist in the outer ergoregion} \end{array} \right) \iff a > \frac{m}{\sqrt{2}} \simeq 0.707 m. \quad (11.187)$$

Similarly, the limiting value of a for the stability of circular orbits in the ergoregion is obtained by setting $r_0 = 2m$ in Eq. (11.160) 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 $a \leq m$,

Property 11.37: stable timelike circular orbits in the ergoregion

$$\left(\begin{array}{l} \text{timelike circular orbits} \\ \text{exist stably in the ergoregion} \end{array} \right) \iff a > \frac{2\sqrt{2}}{3}m \simeq 0.943 m. \quad (11.188)$$

Both in (11.187) and (11.188), the orbits referred to belong to the prograde outer family.

Remark 7: Regarding the limit (11.187), 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 $0 < r_0 < r_{\text{ph}}^*$, 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 $r_- > r_{\text{ph}}^*$. On the contrary, in the limit (11.188), 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 8: As discussed in Sec. 11.3.1, negative-energy or zero-energy particles can exist in the outer ergoregion. However, Fig. 11.12 shows that none of them can follow a circular orbit.

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 [390]. For more details about bound orbits, in particular their description in terms of action angles, see Sec. 6 of the recent review by Pound and Wardell [418] and references therein.

Chapter 12

Null geodesics and images in Kerr spacetime

Contents

12.1 Introduction	441
12.2 Main properties of null geodesics	441
12.3 Spherical photon orbits	458
12.4 Black hole shadow and critical curve	477
12.5 Images	493

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 [6].

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 \mathcal{L} with $E = 0$ cannot exist outside the ergoregion \mathcal{G} , by virtue of the result (11.56). In particular, it cannot exist far from the black hole.

Another property of \mathcal{L} is to have a non-negative Carter constant:

$$\boxed{Q \geq 0}_{E=0}. \quad (12.1)$$

This follows immediately Property 11.19 in Sec. 11.3.7, which, among other things, states 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 \mathcal{L} has some part in \mathcal{M}_I (necessarily in the outer ergoregion) or in \mathcal{M}_{III} (necessarily in the inner ergoregion), the constraint (11.69) reduces to $L < 0$:

$$\mathcal{L} \cap (\mathcal{M}_I \cup \mathcal{M}_{III}) \neq \emptyset \implies L < 0. \quad (12.2)$$

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 \mathcal{L} with $\mu = 0$ and $E = 0$:

$$\frac{dt}{d\lambda'} = -\frac{2amLr}{\Delta} \quad (12.3a)$$

$$\frac{dr}{d\lambda'} = \epsilon_r \sqrt{R(r)} \quad (12.3b)$$

$$\frac{d\theta}{d\lambda'} = \epsilon_\theta \sqrt{\Theta(\theta)} \quad (12.3c)$$

$$\frac{d\varphi}{d\lambda'} = \frac{L}{\Delta \sin^2 \theta} (r^2 - 2mr + a^2 \cos^2 \theta), \quad (12.3d)$$

with [cf. Eqs. (11.95) and (11.35)]:

$$R(r) = -(Q + L^2)r^2 + 2m(Q + L^2)r - a^2Q \quad (12.4)$$

$$\Theta(\theta) = Q - \frac{L^2}{\tan^2 \theta}. \quad (12.5)$$

By combining (12.3a) and (12.3d), we get

$$\boxed{\left. \frac{d\varphi}{dt} \right|_{\mathcal{L}} = \frac{2mr - r^2 - a^2 \cos^2 \theta}{2amr \sin^2 \theta}}_{E=0}. \quad (12.6)$$

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 $\overline{\mathcal{G}}$ 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 \mathcal{L} is necessarily confined. Since moreover $r > 0$ in $\overline{\mathcal{G}}$, we conclude that

$$\frac{d\varphi}{dt} \Big|_{\mathcal{L}} \geq 0. \quad (12.7)$$

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

$$\bar{L} := \frac{L}{\sqrt{Q}} \quad (12.8)$$

and rewrite expression (12.4) for $R(r)$ as

$$R(r)/Q = -(1 + \bar{L}^2)r^2 + 2m(1 + \bar{L}^2)r - a^2. \quad (12.9)$$

Since $1 + \bar{L}^2 \neq 0$, this is a second-order polynomial in r , the two roots of which are

$$r_{\min} = m - \sqrt{m^2 - \frac{a^2}{1 + \bar{L}^2}} \quad \text{and} \quad r_{\max} = m + \sqrt{m^2 - \frac{a^2}{1 + \bar{L}^2}}. \quad (12.10)$$

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

$$r_{\min} \leq r \leq r_{\max}, \quad (12.11)$$

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

$$0 \leq r_{\min} \leq r_- \leq m \leq r_+ \leq r_{\max} \leq 2m, \quad (12.12)$$

with $r_{\min} = 0$ for $a = 0$ or $\bar{L}^2 \rightarrow +\infty$, $r_{\min} = r_-$ for $L = 0$, $r_{\max} = 2m$ for $a = 0$ or $\bar{L}^2 \rightarrow +\infty$ and $r_{\max} = r_+$ for $L = 0$. If $L \neq 0$ and $a \neq 0$, then $r_{\max} > r_+$, so that \mathcal{L} has a part in the outer ergoregion and (12.2) implies that $L < 0$. Hence

$$a \neq 0 \implies L \leq 0. \quad (12.13)$$

Let us consider a zero-energy null geodesic \mathcal{L} emitted outward (i.e. with $\epsilon_r = +1$) from a point A in the outer ergoregion \mathcal{G}^+ . The coordinate r increases along \mathcal{L} from r_A to r_{\max} , which corresponds to a r -turning point. Then r decreases to r_+ , which means that \mathcal{L} crosses the black hole event horizon \mathcal{H} and enters the region \mathcal{M}_{II} . In all \mathcal{M}_{II} , r keeps decreasing and reaches r_- . There \mathcal{L} crosses the inner horizon \mathcal{H}_{in} and enters the region \mathcal{M}_{III} , where r continues to decrease until it reaches r_{\min} . The latter corresponding to a r -turning point, r starts to increase and reaches r_- again. There one might think that \mathcal{L} crosses the inner horizon \mathcal{H}_{in} and enters into \mathcal{M}_{II} . But this is impossible since \mathcal{H}_{in} is a 1-way membrane: it can be crossed by a causal curve from \mathcal{M}_{II} to \mathcal{M}_{III} but not in the reverse way. Moreover, r could

not continue to increase into \mathcal{M}_{II} since r must be decreasing towards the future in all this region (this follows from the hypersurfaces $r = \text{const}$ being spacelike in \mathcal{M}_{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 \mathcal{M}_{III} is not entirely constituted by \mathcal{H}_{in} : it also comprises a null hypersurface that separates \mathcal{M}_{III} from a region distinct from \mathcal{M}_{II} in the maximally extended Kerr spacetime, cf. Fig. 10.11. This region is a “time-reversed” copy of \mathcal{M}_{II} and is denoted by $\mathcal{M}_{\text{II}}^{* \prime}$ in Fig. 10.11 (cf. Sec. 10.8 for details). So actually, when it reaches $r = r_-$, the null geodesic \mathcal{L} enters $\mathcal{M}_{\text{II}}^{* \prime}$. There r necessarily increases towards the future, at the opposite of \mathcal{M}_{II} . It reaches then $r = r_+$, where \mathcal{L} crosses a white hole horizon and emerges into the asymptotically flat region \mathcal{M}_{I}'' , as illustrated in Fig. 12.1. The region \mathcal{M}_{I}'' is similar to \mathcal{M}_{I} . In particular, \mathcal{L} is confined into the outer ergoregion of \mathcal{M}_{I}'' , having a r -turning point at $r = r_{\max}$ (same value (12.10) as in \mathcal{M}_{I}). Then a new cycle begins, with \mathcal{L} entering the future event horizon of \mathcal{M}_{I}'' .

The θ -motion of \mathcal{L} is constrained by $\Theta(\theta) \geq 0$ [Eq. (11.33)], which, given expression (12.5) for Θ , is equivalent to

$$\theta_m \leq \theta \leq \pi - \theta_m \quad \text{with} \quad \theta_m := \arctan(-\bar{L}). \quad (12.14)$$

Remark 1: The general formula for θ_m in the case $a^2(E^2 - \mu^2) = 0$, Eq. (11.85), which holds here since $\mu = 0$ and $E = 0$, yields $\theta_m = \arccos \sqrt{1/(1 + \bar{L}^2)} = \arctan |\bar{L}|$. Hence we recover the above formula.

For $L = 0$, one has $\theta_m = 0$, so that θ takes all values in the range $[0, \pi]$, which means that \mathcal{L} crosses repeatedly the rotation axis. For $L < 0$, one has $0 < \theta_m < \pi/2$ and \mathcal{L} oscillates symmetrically about the equatorial plane, having two θ -turning points, at θ_m and $\pi - \theta_m$. Of course, we recover the general results for $Q > 0$ of Sec. 11.3.7.

We can obtain r as a function of θ along \mathcal{L} by evaluating the integrals in the identity (11.55a):

$$\int_{r_0}^r \frac{\epsilon_r d\bar{r}}{\sqrt{R(\bar{r})}} = \int_{\theta_0}^{\theta} \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\Theta(\bar{\theta})}}$$

Using (12.9) and (12.5), we get on any portion of \mathcal{L} where ϵ_r and ϵ_θ are constant,

$$\epsilon_r \int_{r_0}^r \frac{d\bar{r}}{\sqrt{-(1 + \bar{L}^2)\bar{r}^2 + 2m(1 + \bar{L}^2)\bar{r} - a^2}} = \epsilon_\theta \int_{\theta_0}^{\theta} \frac{d\bar{\theta}}{\sqrt{1 - \bar{L}^2/\tan^2 \bar{\theta}}}.$$

The changes of variables

$$x = \frac{r/m - 1}{\sqrt{1 - \frac{a^2}{m^2(1 + \bar{L}^2)}}} \quad \text{and} \quad \mu = \cos \theta$$

lead to

$$\frac{\epsilon_r}{\sqrt{1 + \bar{L}^2}} \int_{x_0}^x \frac{d\bar{x}}{\sqrt{1 - \bar{x}^2}} = -\epsilon_\theta \int_{\cos \theta_0}^{\cos \theta} \frac{d\mu}{\sqrt{1 - (1 + \bar{L}^2)\mu^2}}.$$

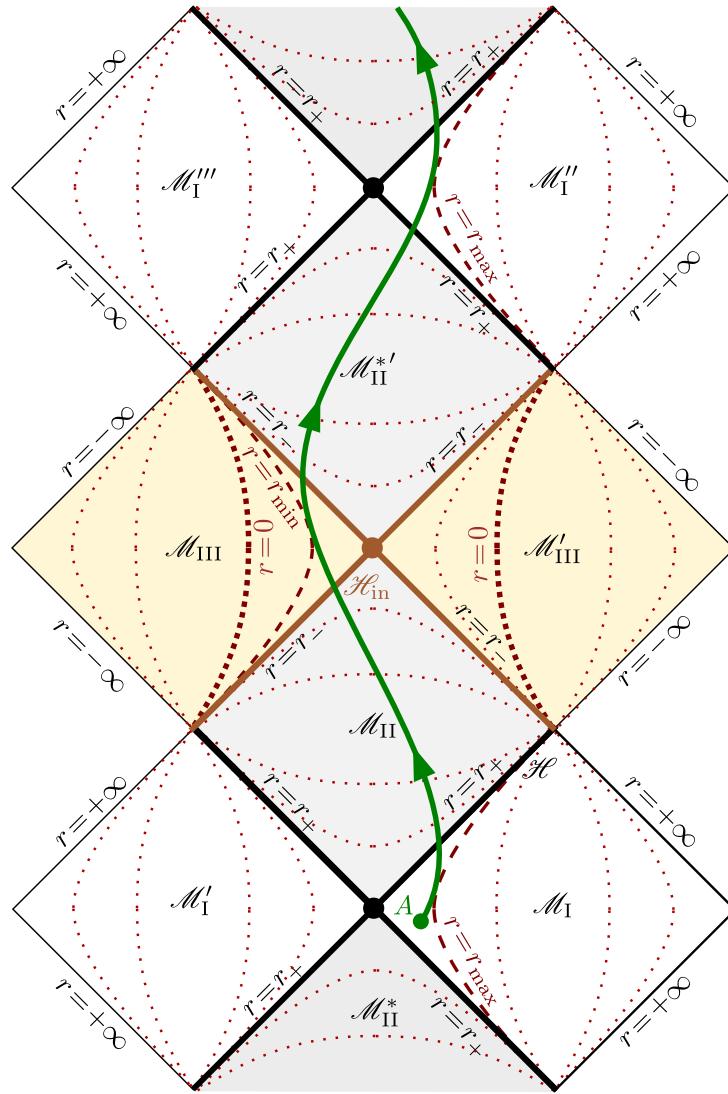


Figure 12.1: Trajectory in the extended Kerr spacetime of a null geodesic with $E = 0$, $Q > 0$ and $L < 0$, emitted from a point A in the outer ergoregion.

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

$$r = m + m \sqrt{1 - \frac{a^2}{m^2(1 + \bar{L}^2)}} \sin \left[K - \epsilon_r \epsilon_\theta \arcsin \left(\sqrt{1 + \bar{L}^2} \cos \theta \right) \right]. \quad (12.15)$$

Since $\sqrt{1 + \bar{L}^2} \cos \theta_m = 1$, we see that the constant K is related to the value of r at $\theta = \theta_m$ by

$$K = \arcsin \left(\frac{r(\theta_m)/m - 1}{\sqrt{1 - \frac{a^2}{m^2(1 + \bar{L}^2)}}} \right) + \epsilon_r \epsilon_\theta \frac{\pi}{2}. \quad (12.16)$$

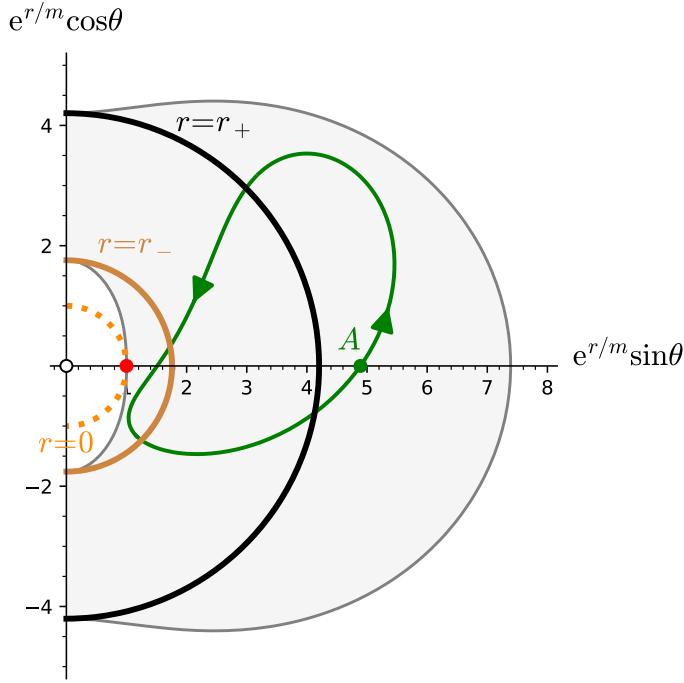


Figure 12.2: Trajectory in a meridional plane, as given by Eq. (12.15), of a null geodesic (green curve) with $E = 0$, $Q > 0$, $L = -\sqrt{Q}$ and $r(\theta_m) = 1.5 m$ in the Kerr spacetime with $a/m = 0.9$. The meridional plane is described in terms of O'Neill exponential coordinates $x = e^{r/m} \sin \theta$ 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 \mathcal{M}_I and \mathcal{M}_I'' in Fig. 12.1, the area $r_- < r < r_+$ corresponds to the regions \mathcal{M}_{II} and \mathcal{M}_{II}' in Fig. 12.1 and the area $r < r_-$ corresponds to the region \mathcal{M}_{III} in Fig. 12.1. [Figure generated by the notebook D.5.10]

Note that K is not a constant all along \mathcal{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 \mathcal{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 θ_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 $r(\theta_m) = 1.5 m$. It has $\theta_m = \pi/4$, $r_{\min} \simeq 0.229 m$ and $r_{\max} \simeq 1.771 m$, 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 \mathcal{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), \mathcal{L} is nothing but a null geodesic generator of the event horizon \mathcal{H} or of the inner

horizon \mathcal{H}_{in} .

In the second case ($\theta = \pi/2$), \mathcal{L} is confined to the equatorial plane. If $L = 0$, we are back to the first case: \mathcal{L} is null geodesic generator of \mathcal{H} or \mathcal{H}_{in} lying in the equatorial plane. If $L \neq 0$, the radial motion of \mathcal{L} is governed by Eq. (12.3b) with the expression (12.4) of $R(r)$ reduced to

$$R(r) = L^2 r(2m - r). \quad (12.17)$$

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. (10.19b). Hence we have necessarily $\mathcal{L} \cap \mathcal{M}_I \neq \emptyset$ 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, \mathcal{L} starts at the ring singularity in a \mathcal{M}_{III} -type region (cf. Fig. 12.3), has r increasing, enters a $\mathcal{M}_{\text{II}}^*$ -type region (time reversed copy of \mathcal{M}_{II}), emerges in \mathcal{M}_I via the white hole horizon at $r = r_+$ and reaches a r -turning point at $r = 2m$, then r decreases continuously until \mathcal{L} terminates at the ring singularity of \mathcal{M}_{III} , after having crossed the black hole horizon \mathcal{H} and the inner horizon \mathcal{H}_{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 \mathcal{M}_I -type regions.

Remark 2: For $Q \neq 0$ and $L \neq 0$, the limit $Q \rightarrow 0$ corresponds to $\bar{L}^2 \rightarrow +\infty$ [cf. Eq. (12.8)], so that Eq. (12.10) yields $r_{\min} \rightarrow 0$ and $r_{\max} \rightarrow 2m$. We recover then the range $[0, 2m]$ for r obtained here for $Q = 0$.

We conclude:

Property 12.1: null geodesic with $E = 0$ and $Q = 0$

Any null geodesic with $E = 0$ and $Q = 0$ is either a null generator of one of the two Killing horizons \mathcal{H} or \mathcal{H}_{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 get:

Property 12.2: negative angular momentum for zero-energy null geodesics

For $a \neq 0$, any null geodesic with $E = 0$ has necessarily

$$\boxed{L \leq 0}_{\substack{a \neq 0 \\ E=0}}. \quad (12.18)$$

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 [465]; some corrections and refinement of his results have been performed by George Contopoulos in 1984 [140], who studied zero-energy timelike geodesics as well.

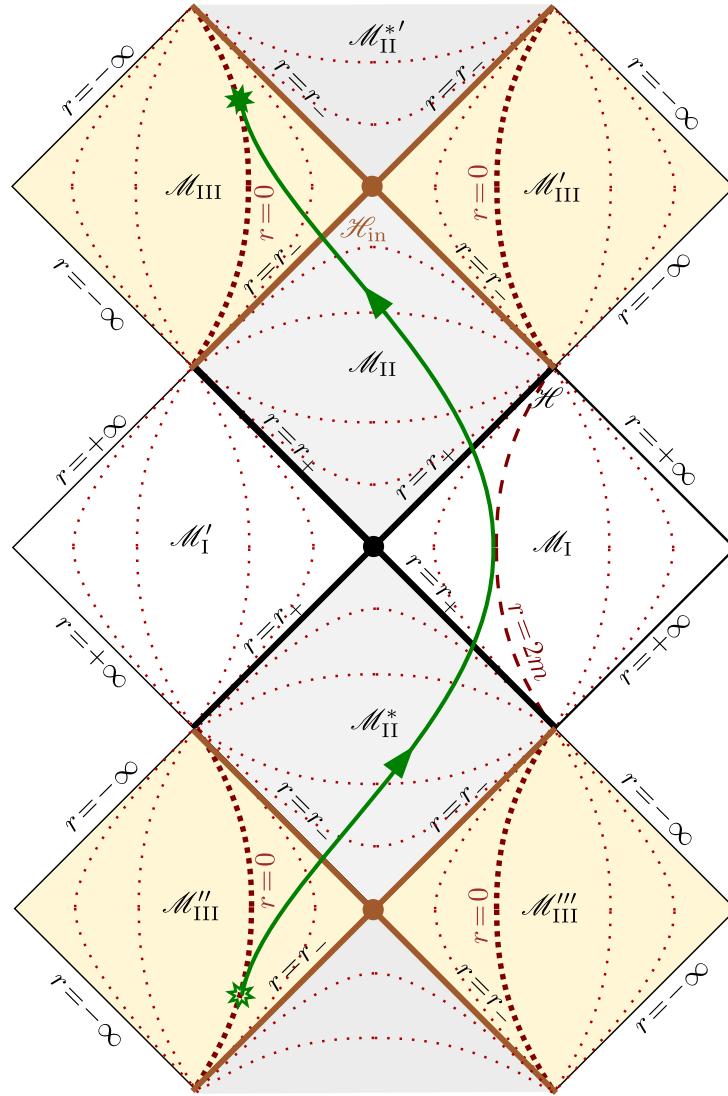


Figure 12.3: Trajectory in the extended Kerr spacetime of a null geodesic with $E = 0, Q = 0$ and $L < 0$.

12.2.2 Equations of geodesic motion for $E \neq 0$

For any null geodesic \mathcal{L} with $E \neq 0$, we introduce the reduced constants of motion

$$\ell := \frac{L}{E} \quad \text{and} \quad q := \frac{Q}{E^2}. \quad (12.19)$$

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. [37, 235, 168]) or by ξ (e.g. Ref. [106]), while q is sometimes denoted by η (e.g. Refs. [37, 106, 235]). 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. [164, 237, 168]).

Remark 4: Contrary to L and Q , the quantities ℓ and q are independent from the affine parametrization of the geodesic \mathcal{L} . Indeed, if instead of the affine parameter λ associated with the particle's 4-momentum p , one considers the affine parameter $\lambda' = \alpha\lambda$, where α is a positive constant, the tangent vector field becomes $p' = \alpha^{-1}p$, so that the associated conserved quantities are $E' = -\xi \cdot p' = \alpha^{-1}E$ [cf. Eq. (11.2a)], $L' = \eta \cdot p' = \alpha^{-1}L$ [cf. Eq. (11.2b)] and $Q' = \tilde{\mathbf{K}}(p', p') = \alpha^{-2}Q$ [cf. Eq. (11.41)], so that $\ell' := L'/E' = \ell$ and $q' := Q'/E'^2 = q$.

The non-negative property of Carter constant \mathcal{K} [Eq. (11.36)], along with the identity (11.34) leads to the following constraint on the parameters ℓ and q :

$$\boxed{q + (\ell - a)^2 \geq 0}. \quad (12.20)$$

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 \mathcal{H} or \mathcal{H}_{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 $Q = -a^2 E^2 \cos^2 \theta_0$ [Eq. (11.38)], where θ_0 is the constant value of θ along the geodesic. We have thus

$$\ell = a \sin^2 \theta_0 \quad \text{and} \quad q = -a^2 \cos^4 \theta_0. \quad (12.21)$$

This yields

$$\boxed{q + (\ell - a)^2 = 0}, \quad (12.22)$$

so that the inequality (12.20) is saturated for principal null geodesics.

The equations of motion in terms of the Mino parameter λ' , Eqs. (11.50) specialized to $\mu = 0$, can be rewritten as

$$\boxed{\frac{1}{E} \frac{dt}{d\lambda'} = \frac{1}{\Delta} [(r^2 + a^2)^2 - 2am\ell r] - a^2 \sin^2 \theta} \quad (12.23a)$$

$$\boxed{\frac{1}{|E|} \frac{dr}{d\lambda'} = \epsilon_r \sqrt{\mathcal{R}(r)}} \quad (12.23b)$$

$$\boxed{\frac{1}{|E|} \frac{d\theta}{d\lambda'} = \epsilon_\theta \sqrt{\tilde{\Theta}(\theta)}} \quad (12.23c)$$

$$\boxed{\frac{1}{E} \frac{d\varphi}{d\lambda'} = \frac{\ell}{\sin^2 \theta} + \frac{a}{\Delta} (2mr - a\ell)}, \quad (12.23d)$$

with

$$\mathcal{R}(r) := \frac{R(r)}{E^2} \quad \text{and} \quad \tilde{\Theta}(\theta) := \frac{\Theta(\theta)}{E^2}. \quad (12.24)$$

From the general expressions (11.44), (11.95) and (11.35) specialized to $\mu = 0$, we get

$$\boxed{\mathcal{R}(r) = (r^2 + a^2 - a\ell)^2 - \Delta [q + (\ell - a)^2]} \quad (12.25a)$$

$$\boxed{\mathcal{R}(r) = r^4 + (a^2 - \ell^2 - q)r^2 + 2m [q + (\ell - a)^2] r - a^2 q} \quad (12.25b)$$

and

$$\tilde{\Theta}(\theta) = q + \cos^2 \theta \left(a^2 - \frac{\ell^2}{\sin^2 \theta} \right). \quad (12.26)$$

It suffices to use the parameter $\lambda'':=|E|\lambda'$ to make E disappear from the system (12.23). We therefore conclude:

Property 12.3

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} :

$$\mathcal{R}(r) = (r^2 + a^2 \cos^2 \theta_0)^2. \quad (12.27)$$

We note that $\mathcal{R}(r) = r^4$, 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:

$$\mathcal{R}(r) = r^4 \quad \left(\theta_0 = \frac{\pi}{2} \right). \quad (12.28)$$

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 \mathcal{L} reaches the asymptotic region $r \rightarrow +\infty$. Indeed, let us consider an asymptotic inertial observer \mathcal{O} located at (fixed) Boyer-Lindquist coordinates $(r_\mathcal{O}, \theta_\mathcal{O}, \varphi_\mathcal{O})$, with $r_\mathcal{O} \gg m$ (cf. Fig. 12.4). In order to reach \mathcal{O} , the geodesic \mathcal{L} must be such that the constraints constraints $\mathcal{R}(r_\mathcal{O}) \geq 0$ [Eq. (12.46)] and $\tilde{\Theta}(\theta_\mathcal{O}) \geq 0$ [Eq. (11.33)] are fulfilled. The first one is always satisfied due to the assumption that \mathcal{O} is an asymptotic observer, since $\mathcal{R}(r) \sim r^4$ for $r \rightarrow +\infty$ [cf. Eq. (12.25b)]. In view of expression (12.26) for $\tilde{\Theta}$, the second constraint is equivalent to

$$(q + a^2 \cos^2 \theta_\mathcal{O}) \sin^2 \theta_\mathcal{O} - \ell^2 \cos^2 \theta_\mathcal{O} \geq 0. \quad (12.29)$$

If $\sin \theta_\mathcal{O}$ is small, this constraint limits significantly the amplitude of ℓ .

Observer not located on the rotation axis

Here we treat the generic case of an observer \mathcal{O} that is not located on the rotation axis, i.e. we assume $\theta_\mathcal{O} \notin \{0, \pi\}$. The orthonormal frame of \mathcal{O} is $e_{(0)} = \xi$, $e_{(r)} = \partial_r$, $e_{(\theta)} = r_\mathcal{O}^{-1} \partial_\theta$ and $e_{(\varphi)} = (r_\mathcal{O} \sin \theta_\mathcal{O})^{-1} \partial_\varphi$ [Eq. (10.95) with $r = r_\mathcal{O} \rightarrow +\infty$]. 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 $e_{(r)}$ is the

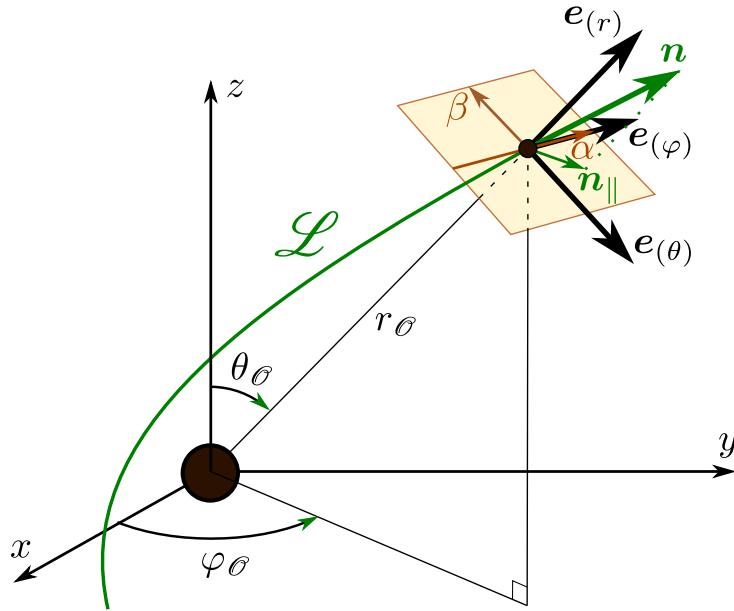


Figure 12.4: Impact of a null geodesic \mathcal{L} onto the screen of a remote observer. (x, y, z) are the Cartesian Boyer-Lindquist coordinates defined by Eq. (11.73).

normal to the screen. The 4-momentum of the photon having \mathcal{L} as worldline at the location of \mathcal{O} writes

$$\mathbf{p} = p^t \boldsymbol{\xi} + p^r \mathbf{e}_{(r)} + p^\theta r_\mathcal{O} \mathbf{e}_{(\theta)} + p^\varphi r_\mathcal{O} \sin \theta_\mathcal{O} \mathbf{e}_{(\varphi)} = E(\boldsymbol{\xi} + \mathbf{n}),$$

where the second equality follows from the orthogonal decomposition (1.31) with respect to \mathcal{O} , $\boldsymbol{\xi}$ being the 4-velocity of \mathcal{O} and the unit spacelike vector \mathbf{n} being the photon's velocity¹ relative to \mathcal{O} ; it obeys $\boldsymbol{\xi} \cdot \mathbf{n} = 0$ and $\mathbf{n} \cdot \mathbf{n} = 1$ [Eq. (1.37)]. The appearance of the conserved energy E in the above equation is a direct consequence of the definition (11.2a): $E := -\boldsymbol{\xi} \cdot \mathbf{p}$, with $\boldsymbol{\xi} \cdot \boldsymbol{\xi} = -1$ for $r_\mathcal{O} \rightarrow +\infty$. The above identity implies $p^t = E$ and

$$\mathbf{n} = \frac{p^r}{E} \mathbf{e}_{(r)} + \frac{p^\theta}{E} r_\mathcal{O} \mathbf{e}_{(\theta)} + \frac{p^\varphi}{E} r_\mathcal{O} \sin \theta_\mathcal{O} \mathbf{e}_{(\varphi)}.$$

With respect to \mathcal{O} , the incoming direction of the photon is given by the vector $-\mathbf{n}$ and the trace on the screen is indicated by the component of $-\mathbf{n}$ that is tangent to the screen, namely

$$\mathbf{m} = -\mathbf{n}_{\parallel} = -\frac{p^\theta}{E} r_\mathcal{O} \mathbf{e}_{(\theta)} - \frac{p^\varphi}{E} r_\mathcal{O} \sin \theta_\mathcal{O} \mathbf{e}_{(\varphi)} \quad (12.30)$$

since the screen plane is spanned by $\mathbf{e}_{(\theta)}$ and $\mathbf{e}_{(\varphi)}$ (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

$$\mathbf{m} = \alpha \mathbf{e}_{(\alpha)} + \beta \mathbf{e}_{(\beta)}, \quad \text{with} \quad \mathbf{e}_{(\alpha)} = \mathbf{e}_{(\varphi)} \quad \text{and} \quad \mathbf{e}_{(\beta)} = -\mathbf{e}_{(\theta)}. \quad (12.31)$$

¹ \mathbf{n} is denoted by \mathbf{V} in Eq. (1.31).

By comparing with (12.30), we get

$$\alpha = -\frac{p^\varphi}{E} r_\mathcal{O} \sin \theta_\mathcal{O} \quad \text{and} \quad \beta = \frac{p^\theta}{E} r_\mathcal{O}.$$

Now, for $r_\mathcal{O} \rightarrow +\infty$,

$$\begin{aligned} \frac{p^\theta}{E} &= \frac{1}{E} \frac{d\theta}{d\lambda} = \frac{1}{Er_\mathcal{O}^2} \frac{d\theta}{d\lambda'} = \frac{\epsilon_\theta}{r_\mathcal{O}^2} \sqrt{\tilde{\Theta}(\theta_\mathcal{O})} \\ \frac{p^\varphi}{E} &= \frac{1}{E} \frac{d\varphi}{d\lambda} = \frac{1}{Er_\mathcal{O}^2} \frac{d\varphi}{d\lambda'} = \frac{\ell}{r_\mathcal{O}^2 \sin^2 \theta_\mathcal{O}}, \end{aligned}$$

where we have used Eqs. (12.23c) and (12.23d), with the term involving a/Δ neglected since $\Delta \sim r^2$ for $r \rightarrow +\infty$. By inserting these formula into the above expressions of α and β , and using Eq. (12.26) for $\tilde{\Theta}(\theta_\mathcal{O})$, we get the sought relation between the constants of motion (ℓ, q) and the screen coordinates:

$$\boxed{\alpha = -\frac{\ell}{r_\mathcal{O} \sin \theta_\mathcal{O}}} \quad (12.32a)$$

$$\boxed{\beta = \frac{\epsilon_\theta}{r_\mathcal{O}} \sqrt{q + \cos^2 \theta_\mathcal{O} \left(a^2 - \frac{\ell^2}{\sin^2 \theta_\mathcal{O}} \right)}} \quad (12.32b)$$

We have defined (α, β) as dimensionless Cartesian coordinates on the screen, cf. Eq. (12.31), where m is dimensionless and $(e_{(\alpha)}, e_{(\beta)})$ is the screen's orthonormal basis. In practice, their values are tiny, being exactly zero at the limit $r_\mathcal{O} \rightarrow +\infty$. (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 \mathcal{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

$$b = |\ell|. \quad (12.33)$$

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 $\theta_\mathcal{O} = \pi/2$ and $q = 0$. Equations (12.32) yield then

$$|\alpha| = \frac{b}{r_\mathcal{O}} = \hat{b} \quad \text{and} \quad \beta = 0, \quad (12.34)$$

where \hat{b} is the angle introduced by formula (8.118).

Remark 6: The angular impact parameters (α, β) depend on the geodesic \mathcal{L} as a curve in spacetime and not on any affine parametrization of \mathcal{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 \mathcal{L} , as noticed in Remark 4 p. 449.

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:

$$\alpha^2 + \beta^2 = \frac{1}{r_\mathcal{O}^2} (\ell^2 + q + a^2 \cos^2 \theta_\mathcal{O}). \quad (12.35)$$

Observer on the rotation axis

If the asymptotic inertial observer \mathcal{O} is located on the rotation axis, i.e. if $\theta_{\mathcal{O}} = 0$ or $\theta_{\mathcal{O}} = \pi$, the only value of ℓ compatible with the constraint (12.29) is

$$\ell = 0 \quad (\theta_{\mathcal{O}} = 0 \text{ or } \theta_{\mathcal{O}} = \pi). \quad (12.36)$$

Given that $\ell = L/E$, we recover one of the properties listed in Sec. 11.3.7: a geodesic cannot encounter the rotation axis unless it has $L = 0$.

Moreover, on the rotation axis, the vectors $e_{(\theta)}$ and $e_{(\varphi)}$ 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 $(e_{(\alpha)}, e_{(\beta)})$ 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 $(e_{(\alpha)}, e_{(\beta)})$. Another way to see that (12.35) is still valid is to notice that it admits a well-defined limit for $\theta_{\mathcal{O}} \rightarrow 0$ or π . Taking into account $\ell = 0$, we obtain

$$\alpha^2 + \beta^2 = \frac{1}{r_{\mathcal{O}}^2} (q + a^2) \quad (\theta_{\mathcal{O}} = 0 \text{ or } \theta_{\mathcal{O}} = \pi). \quad (12.37)$$

12.2.4 Latitudinal motion

Specializing the general results expressed in Property 11.19 to $\mu = 0$ and $E \neq 0$, we get

Property 12.4: latitudinal motion of null geodesics

- A null geodesic \mathcal{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:

$$q \geq 0. \quad (12.38)$$

- 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:

$$q \geq q_{\min} = -(a - |\ell|)^2. \quad (12.39)$$

If $q < 0$, \mathcal{L} is called a **vortical null geodesic**; it never encounters the equatorial plane.

- If $q > 0$ and $\ell \neq 0$, \mathcal{L} oscillates symmetrically about the equatorial plane, between two θ -turning points, at $\theta = \theta_m$ and $\theta = \pi - \theta_m$, where $\theta_m \in (0, \pi/2)$ is given by

Eqs. (11.85) and (11.90):

$$\theta_m = \arccos \sqrt{\frac{q}{\ell^2 + q}} \quad \text{for } a = 0 \quad (12.40)$$

$$\theta_m = \arccos \sqrt{\frac{1}{2} \left[1 - \frac{\ell^2 + q}{a^2} + \sqrt{\left(1 - \frac{\ell^2 + q}{a^2}\right)^2 + \frac{4q}{a^2}} \right]} \quad \text{for } a \neq 0. \quad (12.41)$$

If $q > 0$ and $\ell = 0$, \mathcal{L} crosses repeatedly the rotation axis, with θ taking all values in the range $[0, \pi]$.

- If $q = 0$, \mathcal{L} is stably confined to the equatorial plane for $|\ell| > a$ or $|\ell| = a \neq 0$; for $|\ell| < a$, \mathcal{L} either lies unstably in the equatorial plane or approaches it asymptotically from one side, while for $\ell = 0$ and $a = 0$, \mathcal{L} lies at a constant value $\theta = \theta_0 \in [0, \pi]$.
- If $q_{\min} < q < 0$, \mathcal{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 $\ell \neq 0$, \mathcal{L} oscillates between two θ -turning points, at $\theta = \theta_m$ and $\theta = \theta_v$ (Northern hemisphere) or at $\theta = \pi - \theta_v$ and $\theta = \pi - \theta_m$ (Southern hemisphere), where θ_m is given by Eq. (12.41) above and θ_v is given by Eq. (11.91):

$$\theta_v = \arccos \sqrt{\frac{1}{2} \left[1 - \frac{\ell^2 + q}{a^2} - \sqrt{\left(1 - \frac{\ell^2 + q}{a^2}\right)^2 + \frac{4q}{a^2}} \right]}, \quad (12.42)$$

if $\ell = 0$, \mathcal{L} oscillates about the rotation axis, with a θ -turning point at $\theta = \theta_v$ or $\theta = \pi - \theta_v$, where θ_v is given by Eq. (11.88), or equivalently by the $\ell \rightarrow 0$ limit of Eq. (12.42):

$$\theta_v = \arccos \left(\frac{\sqrt{-q}}{a} \right) \quad (\ell = 0). \quad (12.43)$$

- If $q = q_{\min}$, \mathcal{L} lies stably at a constant value $\theta = \theta_*$ or $\theta = \pi - \theta_*$, with $\theta_* \in [0, \pi/2)$ given by

$$\theta_* := \arcsin \sqrt{\frac{|\ell|}{a}}. \quad (12.44)$$

Remark 7: For $\ell < 0$, the constraint (12.39) is tighter than (12.20).

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 $q_{\min} = -a^2 \cos^4 \theta_0$. Comparing with the value of q given by Eq. (12.21), we note that

$$q = q_{\min}, \quad (12.45)$$

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 $\sqrt{|\ell|/a} = \sin \theta_0$.

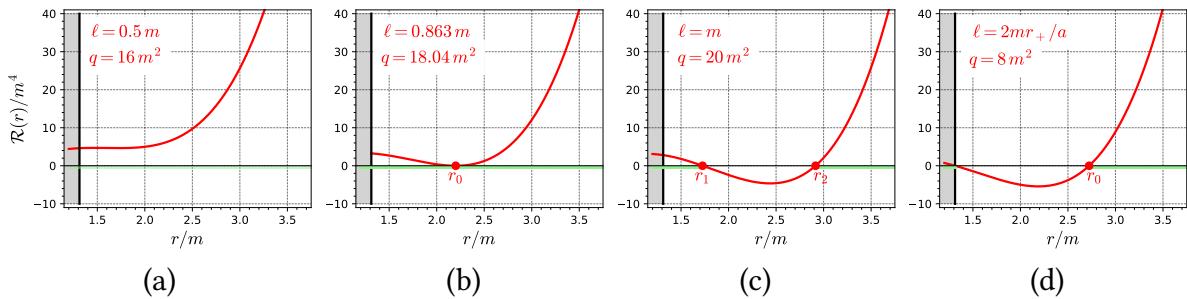


Figure 12.5: Quartic polynomial $\mathcal{R}(r)$ in the region \mathcal{M}_1 for four values of (ℓ, q) and for $a = 0.95 \text{ m}$. The grey area marks the black hole region, with the vertical black line at the event horizon, at $r = r_+ \simeq 1.312 \text{ m}$. The green part of the r axis corresponds to $\mathcal{R}(r) \geq 0$, i.e. to regions where the geodesic motion is allowed. [Figure generated by the notebook [D.5.13](#)]

12.2.5 Radial motion

As for any geodesic, the radial motion of a null geodesic \mathcal{L} is ruled by $R(r) \geq 0$ [Eq. (11.30)], which, in terms of $\mathcal{R}(r)$ [Eq. (12.25)], writes

$$\boxed{\mathcal{R}(r) \geq 0}. \quad (12.46)$$

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. 10.9) and outgoing principal null geodesics travel from $r = -\infty$ to $r = +\infty$ (cf. the solid green curve in Fig. 10.10).

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 12.5: behavior of $\mathcal{R}(r)$ in \mathcal{M}_I

In region \mathcal{M}_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 \mathcal{M}_I . This is clear for the asymptotic flat end since, according to expression (12.25b), $\mathcal{R}(r) \sim r^4 > 0$ when $r \rightarrow +\infty$. At the inner end, namely for $r \rightarrow r_+$, we have $\Delta \rightarrow 0$ 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. \square

Remark 8: Case 4 of Lemma 12.5 can be seen as the limit $r_1 \rightarrow r_+$ of case 3.

Remark 9: In case 4, $\mathcal{R}(r_+) = 0$ occurs for $\Omega_{\mathcal{H}}\ell = 1$, where $\Omega_{\mathcal{H}}$ is the black hole rotation velocity, as given by Eq. (10.69). Similarly, $\mathcal{R}(r_-) = 0$ for $\Omega_{\text{in}}\ell = 1$, where Ω_{in} is the rotation velocity of the inner horizon, cf. Eq. (10.71).

Remark 10: The region \mathcal{M}_{III} shares with \mathcal{M}_I that $\mathcal{R}(r)$ is non-negative at each of its ends: $\mathcal{R}(r) \sim r^4 > 0$ when $r \rightarrow -\infty$ and $\mathcal{R}(r_-) = (r_-^2 + a^2 - a\ell)^2 \geq 0$. However, the argument used to limit the number of roots in \mathcal{M}_I cannot be applied to \mathcal{M}_{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 \mathcal{M}_{III} .

According to the definition given in Sec. 11.2.6, a r -turning of a null geodesic \mathcal{L} is a point $p_0 \in \mathcal{L}$ such that $r_0 := r(p_0)$ is a simple root of $\mathcal{R}(r)$. Lemma 12.5 leads then to:

Property 12.6: r -turning points of null geodesics

A null geodesic in Kerr spacetime has

- at most one r -turning point in region \mathcal{M}_I ;
- no r -turning point in region \mathcal{M}_{II} .

Proof. If a geodesic would have two turning points in \mathcal{M}_I , this would mean that there exist two simple roots in \mathcal{M}_I and that $\mathcal{R}(r)$ is positive between them (so that the motion is possible there). But this is excluded by Lemma 12.5. In region \mathcal{M}_{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.

$$r^2 = a(\ell - a) \quad \text{and} \quad q = -(\ell - a)^2.$$

The second equation implies $q \leq 0$. The case $q = 0$ would lead to $\ell = a$ and $r^2 = 0$, which is impossible in \mathcal{M}_{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 \mathcal{M}_{II} . \square

Remark 11: That no r -turning point can exist in \mathcal{M}_{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 (11.97), namely that r must be a strictly decreasing function of λ at any point in \mathcal{M}_{II} .

Let us consider case 2 of Lemma 12.5 (double root of $\mathcal{R}(r)$ in \mathcal{M}_1); it corresponds to two distinct situations regarding the null geodesic \mathcal{L} having $\mathcal{R}(r)$ as radial polynomial. First, \mathcal{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); \mathcal{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 \mathcal{L} , then according to the definition in Sec. 11.3.6, \mathcal{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 $dr/d\lambda \neq 0$ all along \mathcal{L} , so that $r(\lambda) \rightarrow r_0$ for $\lambda \rightarrow +\infty$ (future asymptotic value) or $\lambda \rightarrow -\infty$ (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:

Property 12.7: radial behavior of null geodesics

In the region \mathcal{M}_1 of Kerr spacetime, any null geodesic \mathcal{L} has one of the following behaviors. For generic values of the constant of motions (ℓ, q) , the possibilities are:

1. \mathcal{L} arises from the past null infinity of \mathcal{M}_1 , \mathcal{I}^- , has a r -turning point, which we may call the **periastron**, and terminates at the future null infinity of \mathcal{M}_1 , \mathcal{I}^+ ;
2. \mathcal{L} arises from the past null infinity \mathcal{I}^- , has r decreasing monotonically and crosses the black hole event horizon \mathcal{H} ;
3. \mathcal{L} arises from the past event horizon \mathcal{H}^- separating \mathcal{M}_1 from the white hole region $\mathcal{M}_{\text{II}}^*$, has r increasing monotonically and terminates at the future null infinity \mathcal{I}^+ ;
4. \mathcal{L} arises from the past event horizon \mathcal{H}^- , has a r -turning point, which we may call the **apoastron**, and crosses the black hole event horizon \mathcal{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. \mathcal{L} evolves at a fixed value of r ;

6. \mathcal{L} arises from the past null infinity \mathcal{I}^- and has r decreasing monotonically to a future asymptotic r -value at $r_0 > r_+$;
7. \mathcal{L} arises from a past asymptotic r -value at $r_0 > r_+$, has r increasing monotonically and terminates at the future null infinity \mathcal{I}^+ ;
8. \mathcal{L} arises from a past asymptotic r -value at $r_0 > r_+$, has r decreasing monotonically and crosses the black hole event horizon \mathcal{H} ;
9. \mathcal{L} arises from the past event horizon \mathcal{H}^- , has r increasing monotonically to a future asymptotic r -value at $r_0 > r_+$.

^aThis 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 Lemma 12.5. 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 case 1 of Lemma 12.5 (no root of $\mathcal{R}(r)$ in \mathcal{M}_1). Cases 5–9 correspond to case 2 of Lemma 12.5 (double root of $\mathcal{R}(r)$ in \mathcal{M}_1). 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 \mathcal{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 \mathcal{M}_{III} . There can thus exist null geodesics that are trapped between two distinct values of r in \mathcal{M}_{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 \mathcal{L} is *spherical* or is a *spherical photon orbit* iff \mathcal{L} lies at a constant value of the coordinate r , r_0 say. We have already encountered such geodesics, namely the null generators $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ and $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}_{\text{in}}}$ of the two Killing horizons \mathcal{H} and \mathcal{H}_{in} (cf. Secs. 10.5.3 and 10.8.3). Indeed, they lie at a constant value of r : $r_0 = r_+$ for \mathcal{H} and $r_0 = r_-$

for \mathcal{H}_{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:

Property 12.8: zero-energy spherical photon orbits

In Kerr spacetime, the only spherical photon orbits with $E = 0$ are the null generators of the two Killing horizons \mathcal{H} and \mathcal{H}_{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 [143, 144], namely a null geodesic \mathcal{L} of affine parameter λ such that (i) \mathcal{L} is restricted a spatially compact region and (ii) there exists $\Lambda \in \mathbb{R}^+$ such that

$$\forall \lambda \in \mathbb{R}, \quad \exists \Phi \in \mathbb{R} \times \text{SO}(2), \quad p(\lambda + \Lambda) = \Phi(p(\lambda)),$$

where $p(\lambda)$ stands for the point of affine parameter λ along \mathcal{L} and $\mathbb{R} \times \text{SO}(2)$ is the symmetry group implementing stationarity and axisymmetry. Basically (i) says that \mathcal{L} is a bound orbit and (ii) that \mathcal{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 \mathcal{L} , the quartic polynomial $\mathcal{R}(r)$ associated to (ℓ, q) by Eq. (12.25) must obey

$$\mathcal{R}(r_0) = 0 \quad \text{and} \quad \mathcal{R}'(r_0) = 0. \quad (12.47)$$

Proof. Since $r = r_0$ with r_0 constant, we have $dr/d\lambda' = 0$ where λ' is the Mino parameter along \mathcal{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 \mathcal{L} (cf. Eq. (11.46)), which would contradict r being constant. \square

In other words, r_0 is a double root the polynomial $\mathcal{R}(r)$. A spherical photon orbit in \mathcal{M}_I correspond then to case 2 of Lemma 12.5 and an example of $\mathcal{R}(r)$ is shown for $a = 0.95 \text{ m}$ and $r_0 = 2.2 \text{ m}$ in Fig. 12.5b.

In view of expression (12.25a) for $\mathcal{R}(r)$, the system (12.47) is equivalent to

$$(r_0^2 + a^2 - a\ell)^2 - \tilde{q}\Delta_0 = 0 \quad (12.48a)$$

$$2r_0(r_0^2 + a^2 - a\ell) - \tilde{q}(r_0 - m) = 0, \quad (12.48b)$$

where $\Delta_0 := r_0^2 - 2mr_0 + a^2$ and \tilde{q} is the reduced Carter constant \mathcal{K} :

$\tilde{q} := \frac{\mathcal{K}}{E^2} = q + (\ell - a)^2,$

(12.49)

where the second equality follows from relation (11.34) between \mathcal{K} and Q . The system (12.48) has 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

$$\begin{cases} (m^2 + a^2 - a\ell)^2 - \tilde{q}(a^2 - m^2) = 0 \\ m^2 + a^2 - a\ell = 0 \end{cases} \iff \begin{cases} \ell = a + \frac{m^2}{a} \\ \tilde{q} = 0 \quad \text{or} \quad a = m. \end{cases}$$

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

$$r_0 = m \iff \begin{cases} a = m \\ \ell = 2m. \end{cases}$$

(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

$$\tilde{q} = \frac{2r_0(r_0^2 + a^2 - a\ell)}{r_0 - m}. \quad (12.51)$$

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:

$$\ell = a + \frac{r_0^2}{a} \quad (12.52)$$

or

$$\ell = a + \frac{r_0}{a(r_0 - m)} [r_0(r_0 - m) - 2\Delta_0]. \quad (12.53)$$

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

$$\tilde{q} = \frac{4r_0^2\Delta_0}{(r_0 - m)^2}, \quad (12.54)$$

so that Eqs. (12.49) and (12.53) yield, after simplifications,

$$q = \frac{r_0^3}{a^2(r_0 - m)^2} [-r_0^3 + 6mr_0^2 - 9m^2r_0 + 4a^2m]. \quad (12.55)$$

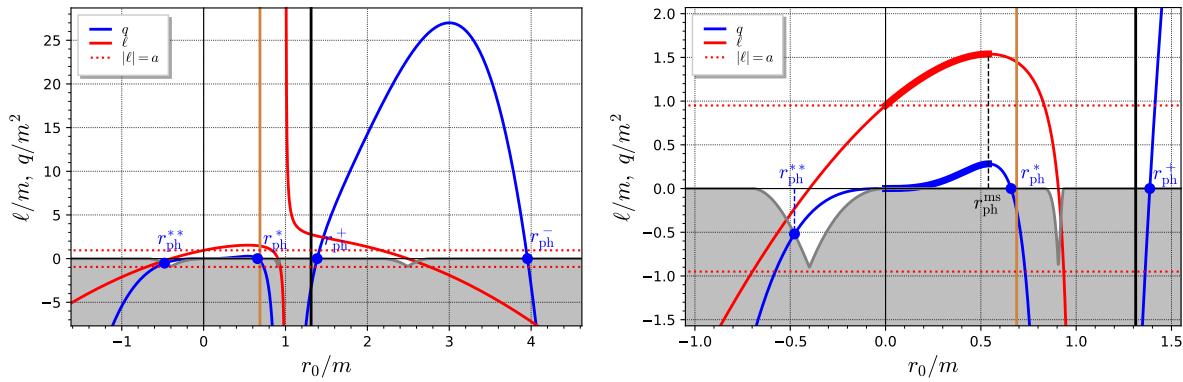


Figure 12.6: Functions $\ell_c(r_0)$ (in red) and $q_c(r_0)$ (in blue) giving the reduced angular momentum and reduced Carter constant of a spherical photon orbit of radius r_0 , according to Eqs. (12.56) and (12.57) and for $a = 0.95 \text{ m}$. The right figure is a zoom on the part $-m \leq r_0 \leq 3m/2$. The thick vertical lines mark the two horizons: \mathcal{H} (black) and \mathcal{H}_{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 unphysical, 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 $r_{\text{ph}}^{**} \leq r_0 \leq r_{\text{ph}}^*$ and $r_{\text{ph}}^+ \leq r_0 \leq r_{\text{ph}}^-$. The thick segments of the ℓ and q curves correspond to stable orbits, for which $0 \leq r_0 \leq r_{\text{ph}}^{\text{ms}}$. [Figure generated by the notebook D.5.11]

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

$$\ell = \ell_c(r_0) := \frac{r_0^2(3m - r_0) - a^2(r_0 + m)}{a(r_0 - m)}, \quad (12.56)$$

$$q = q_c(r_0) := \frac{r_0^3}{a^2(r_0 - m)^2} [4a^2m - r_0(r_0 - 3m)^2]. \quad (12.57)$$

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 \mathcal{M}_{II} of Kerr spacetime. This is in agreement with the general result (11.97).

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

$$q \geq 0 \quad \text{if } |\ell| \geq a \quad (12.58a)$$

$$q \geq -(a - |\ell|)^2 \quad \text{if } |\ell| < a. \quad (12.58b)$$

The solutions ℓ and q given by Eqs. (12.56)-(12.57) are plotted as functions of r_0 for $a = 0.95 \text{ m}$ 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:

$$r_0 \in [r_{\text{ph}}^{**}, 0], \quad r_0 \in (0, r_{\text{ph}}^*], \quad \text{and} \quad r_0 \in [r_{\text{ph}}^+, r_{\text{ph}}^-], \quad (12.59)$$

where r_{ph}^* , r_{ph}^+ and r_{ph}^- are the three (ordered) roots distinct from 0 of the equation $q_c(r_0) = 0$ (boundary for condition (12.58a)) and r_{ph}^{**} is the unique root of the equation $q_c(r_0) = -(a - |\ell_c(r_0)|)^2$ when $|\ell_c(r_0)| < a$ (boundary for condition (12.58b)). The above reasoning is based on Fig. 12.6, which has been drawn for $a = 0.95 \text{ m}$; however the conclusions (12.59) hold for any value of a (see the notebook D.5.11 for figures with $a = 0.5 \text{ m}$ or $a = 0.998 \text{ m}$). 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_{ph}^* , r_{ph}^+ and r_{ph}^- are the three roots of the cubic equation

$$r_0(r_0 - 3m)^2 - 4a^2m = 0. \quad (12.60)$$

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:

$$\begin{aligned} r_0(r_0 - 3m)^2 - 4a^2m = 0 &\iff r_0 \geq 0 \text{ and } \sqrt{r_0}|r_0 - 3m| = 2a\sqrt{m} \\ &\iff r_0 \geq 0 \text{ and } \sqrt{r_0}(r_0 - 3m) \pm 2a\sqrt{m} = 0 \\ &\iff r_0 \geq 0 \text{ and } r_0^{3/2} - 3mr_0^{1/2} \pm 2a\sqrt{m} = 0, \end{aligned}$$

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. (11.141). As shown in Sec. 11.5.2, there are three real roots, r_{ph}^* , r_{ph}^+ and r_{ph}^- , with $r_{\text{ph}}^*, r_{\text{ph}}^+ \leq 3m$ ($\pm = +$) and $r_{\text{ph}}^- \geq 3m$ ($\pm = -$). They are given by Eqs. (11.144) and (11.143):

$$r_{\text{ph}}^* := 4m \cos^2 \left[\frac{1}{3} \arccos \left(-\frac{a}{m} \right) + \frac{4\pi}{3} \right]. \quad (12.61)$$

$$r_{\text{ph}}^\pm := 4m \cos^2 \left[\frac{1}{3} \arccos \left(\mp \frac{a}{m} \right) \right] \quad (12.62)$$

As for r_{ph}^{**} , since $q_c(r_0) = -(a - |\ell_c(r_0)|)^2$ with $|\ell_c(r_0)| < a$ occurs in a region where $\ell_c(r_0) < 0$ (cf. Fig. 12.6), we get that r_{ph}^{**} is a solution of the equation $q_c(r_0) = -(a + \ell_c(r_0))^2$. Given expressions (12.56) and (12.57) for respectively $\ell_c(r_0)$ and $q_c(r_0)$, we get that r_{ph}^{**} is a root of the cubic equation

$$2r_0^3 - 3mr_0^2 + a^2m = 0. \quad (12.63)$$

By the change of variable $r_0 =: x + m/2$, we turn this equation into a depressed one:

$$x^3 - \frac{3}{4}m^2x + \frac{m}{2} \left(a^2 - \frac{m^2}{2} \right) = 0,$$

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

$$x_1 = m \cos \left[\frac{1}{3} \arccos \left(1 - 2 \frac{a^2}{m^2} \right) + \frac{2\pi}{3} \right] = m \cos \left[\frac{2}{3} \arcsin \left(\frac{a}{m} \right) + \frac{2\pi}{3} \right],$$

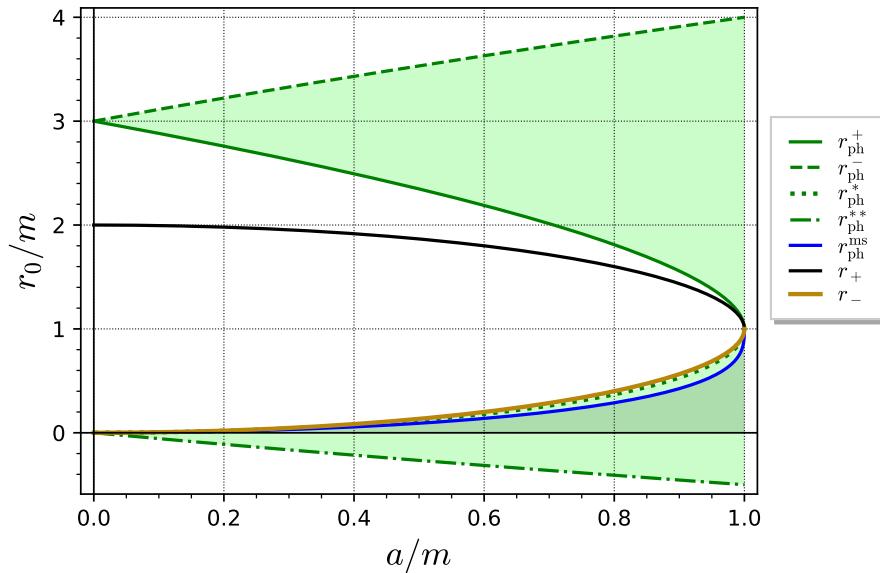


Figure 12.7: Domain of existence of spherical photon orbits in the (a, r_0) plane (in green). The boundaries of the domain are the radii r_{ph}^{**} , r_{ph}^* , r_{ph}^+ and r_{ph}^- given by Eqs. (12.64), (12.61) and (12.62). The shaded area correspond to stable spherical orbits; its upper boundary (blue curve) is the radius $r_{\text{ph}}^{\text{ms}}$ given by Eq. (12.78). The black curve indicates the black hole horizon \mathcal{H} and the light brown one the inner horizon \mathcal{H}_{in} . [Figure generated by the notebook D.5.11]

from which we obtain

$$r_{\text{ph}}^{**} = \frac{m}{2} + m \cos \left[\frac{2}{3} \arcsin \left(\frac{a}{m} \right) + \frac{2\pi}{3} \right]. \quad (12.64)$$

The four critical radii r_{ph}^{**} , r_{ph}^* , r_{ph}^+ and r_{ph}^- are plotted in terms of a in Fig. 12.7. We notice that

$$-\frac{m}{2} \leq r_{\text{ph}}^{**} \leq 0 \leq r_{\text{ph}}^* \leq r_- \leq m \leq r_+ \leq r_{\text{ph}}^+ \leq 3m \leq r_{\text{ph}}^- \leq 4m, \quad (12.65)$$

with the limits (11.145) and (11.146). In addition,

$$\lim_{a \rightarrow 0} r_{\text{ph}}^{**} = 0 \quad \text{and} \quad \lim_{a \rightarrow m} r_{\text{ph}}^{**} = -\frac{m}{2}. \quad (12.66)$$

As already stressed in Sec. 11.5.2, r_{ph}^* is lower than, but very close to, the inner horizon radius r_- , with $\max(r_- - r_{\text{ph}}^*) \simeq 0.032 m$, achieved for $a \simeq 0.9 m$. In view of the above inequalities and the ranges (12.59), we conclude:

Property 12.9: spherical photon orbits with $E \neq 0$

Spherical photon orbits with $E \neq 0$ exist in two regions of Kerr spacetime:

- orbits with $r_0 \in [r_{\text{ph}}^{**}, 0) \cup (0, r_{\text{ph}}^*]$ are located in \mathcal{M}_{III} ; we shall call them the **inner spherical photon orbits**;

- orbits with $r_0 \in [r_{\text{ph}}^+, r_{\text{ph}}^-]$ are located in \mathcal{M}_I ; we shall call them the *outer spherical photon orbits*.

Sign of E

All $E \neq 0$ spherical photon orbits lie in $\mathcal{M}_I \cup \mathcal{M}_{\text{III}}$. We may then apply the general result (11.69) 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 $\mathcal{M}_I \cup \mathcal{M}_{\text{III}}$. In view of (12.65), we conclude:

Property 12.10: energy sign of spherical photon orbits

All outer spherical photon orbits have $E > 0$, as well as inner spherical photon orbits with $r_0 \in [r_{\text{ph}}^{**}, 0)$, while inner spherical photon orbits with $r_0 \in (0, r_{\text{ph}}^*]$ 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: $r_0 = r_{\text{ph}}^{\text{pol,in}}$ and $r_0 = r_{\text{ph}}^{\text{pol}}$ with $r_{\text{ph}}^{**} < r_{\text{ph}}^{\text{pol,in}} < 0$ and $r_{\text{ph}}^+ < r_{\text{ph}}^{\text{pol}} < r_{\text{ph}}^-$. The value $r_{\text{ph}}^{\text{pol,in}}$ (resp. $r_{\text{ph}}^{\text{pol}}$) 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 $r_{\text{ph}}^{\text{pol,in}}$ and $r_{\text{ph}}^{\text{pol}}$ are obtained by solving $\ell_c(r_0) = 0$, with $\ell_c(r_0)$ given by Eq. (12.56). We get the cubic equation $r_0^3 - 3mr_0^2 + a^2r_0 + a^2m = 0$. 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 $\ell_c(r_0)$ seen on

Fig. 12.6, which lies in the unphysical range of r_0 — blue curve in the grey region):

$$r_{\text{ph}}^{\text{pol}} = m + 2\sqrt{m^2 - \frac{a^2}{3}} \cos \left[\frac{1}{3} \arccos \left(\frac{m(m^2 - a^2)}{(m^2 - a^2/3)^{3/2}} \right) \right] \quad (12.67\text{a})$$

$$r_{\text{ph}}^{\text{pol,in}} = m + 2\sqrt{m^2 - \frac{a^2}{3}} \cos \left[\frac{1}{3} \arccos \left(\frac{m(m^2 - a^2)}{(m^2 - a^2/3)^{3/2}} \right) + \frac{2\pi}{3} \right]. \quad (12.67\text{b})$$

As a function of a , $r_{\text{ph}}^{\text{pol}}$ decays monotonically from $\lim_{a \rightarrow 0} r_{\text{ph}}^{\text{pol}} = 3m$ to $\lim_{a \rightarrow m} r_{\text{ph}}^{\text{pol}} = (1 + \sqrt{2})m \simeq 2.414214 m$, while $r_{\text{ph}}^{\text{pol,in}}$ decays monotonically from $\lim_{a \rightarrow 0} r_{\text{ph}}^{\text{pol,in}} = 0$ to $\lim_{a \rightarrow m} r_{\text{ph}}^{\text{pol,in}} = (1 - \sqrt{2})m \simeq -0.414214 m$.

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 $r_{\text{ph}}^{\text{pol,in}}$ and $r_{\text{ph}}^{\text{pol}}$ determined above and whose denominator is positive iff $r_0 > m$, which occurs only for outer spherical orbits. We get then

Property 12.11: sign of ℓ for spherical photon orbits

- Outer spherical orbits with $r_0 \in [r_{\text{ph}}^+, r_{\text{ph}}^{\text{pol}})$ have $\ell > 0$;
- outer spherical orbits with $r_0 \in (r_{\text{ph}}^{\text{pol}}, r_{\text{ph}}^-]$ have $\ell < 0$;
- inner spherical orbits with $r_0 \in (r_{\text{ph}}^{\text{pol,in}}, r_{\text{ph}}^*)$ have $\ell > 0$;
- inner spherical orbits with $r_0 \in [r_{\text{ph}}^{**}, r_{\text{ph}}^{\text{pol,in}})$ 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

Property 12.12: latitudinal behavior of spherical photon orbits

- All outer spherical photon orbits with $r_0 \notin \{r_{\text{ph}}^+, r_{\text{ph}}^-\}$ and the inner ones with $r_0 \in (0, r_{\text{ph}}^*)$ have $q > 0$ (cf. Fig. 12.6); they therefore cross the equatorial plane. Moreover,
 - those with $r_0 = r_{\text{ph}}^{\text{pol}}$ (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 $r_0 \neq r_{\text{ph}}^{\text{pol}}$ oscillate about the equatorial plane, having two θ -turning points symmetrical about it, at $\theta = \theta_m$ and $\theta = \pi - \theta_m$, with θ_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 $r_0 \in (r_{\text{ph}}^{**}, 0)$ have $q < 0$ (cf. right panel of Fig. 12.6); they are thus vortical and never encounter the equatorial plane. Moreover,
 - those with $r_0 = r_{\text{ph}}^{\text{pol,in}}$ (inner polar spherical photon orbits) have $\ell = 0$ and oscillate about the rotation axis, with a θ -turning point at $\theta = \theta_v$ (Northern hemisphere) or $\theta = \pi - \theta_v$ (Southern hemisphere), where θ_v is given by Eq. (12.43) in which q is to be considered as the function (12.57) of $r_{\text{ph}}^{\text{pol,in}}$.
 - those with $r_0 \neq r_{\text{ph}}^{\text{pol,in}}$ neither encounter the rotation axis nor the equatorial plane, having either $\theta \in [\theta_m, \theta_v]$ (Northern hemisphere) or $\theta \in [\pi - \theta_v, \pi - \theta_m]$ (Southern hemisphere), with θ_m and θ_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 $r_0 = r_{\text{ph}}^{**}$, r_{ph}^* , r_{ph}^+ or r_{ph}^- , which have been excluded from the above list, will be discussed in details in Sec. 12.3.3.

In addition, we have

Property 12.13: periodicity of the θ -motion

For a spherical photon orbit, θ is either a constant or a periodic function of the affine parameter λ , the period being

$$\Lambda_\theta = \frac{2}{|E|} \int_{\theta_{\min}}^{\theta_{\max}} \frac{r_0^2 + a^2 \cos^2 \theta}{\sqrt{\tilde{\Theta}(\theta)}} d\theta, \quad (12.68)$$

with $(\theta_{\min}, \theta_{\max}) = (0, \pi)$ (outer polar orbit), $(0, \theta_v)$ (Northern inner polar orbit), $(\pi - \theta_v, \pi)$ (Southern inner polar orbit), $(\theta_m, \pi - \theta_m)$ (non-polar orbit with $r_0 \in (0, r_{\text{ph}}^*) \cup (r_{\text{ph}}^+, r_{\text{ph}}^-)$), (θ_m, θ_v) (Northern non-polar vortical inner orbit) or $(\pi - \theta_v, \pi - \theta_m)$ (Southern non-polar vortical inner orbit).

Proof. Switching from the Mino parameter λ' to the affine parameter λ via Eq. (11.49) with $r(\lambda) = r_0$, we may rewrite the equation of motion (12.23c) as

$$\left(\frac{d\theta}{d\lambda} \right)^2 + \mathcal{V}(\theta) = 0, \quad \text{with } \mathcal{V}(\theta) := -\frac{E^2 \tilde{\Theta}(\theta)}{(r_0^2 + a^2 \cos^2 \theta)^2}.$$

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 $d\lambda = \epsilon_\theta (r_0^2 + a^2 \cos^2 \theta)^2 / (|E| \sqrt{\tilde{\Theta}(\theta)})$ over a “round-trip” to the same values of θ and $d\theta/d\lambda$. \square

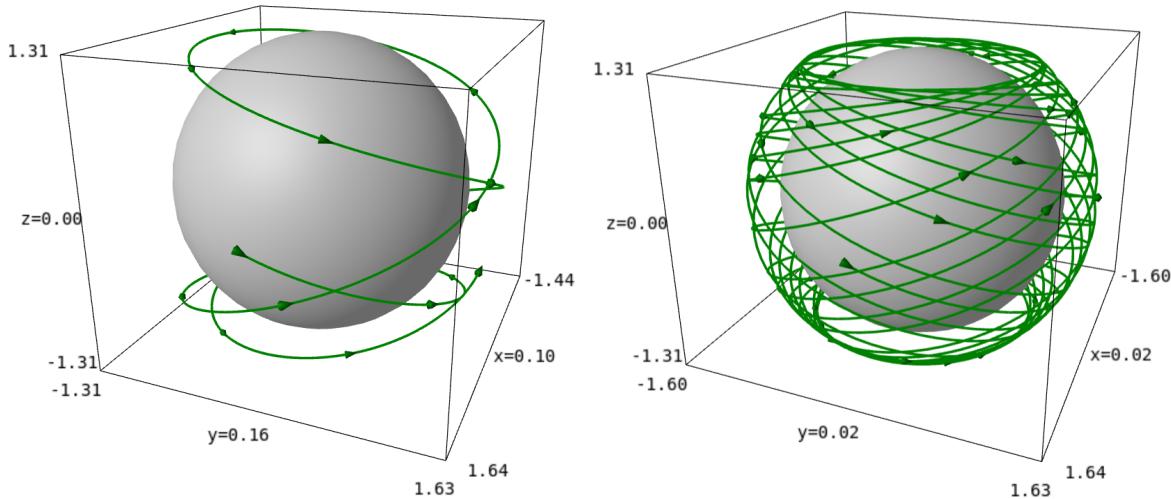


Figure 12.8: Spherical photon orbit at $r_0 = 1.6 \text{ m}$ around a Kerr black hole with $a = 0.95 \text{ m}$, depicted in terms of the Cartesian Boyer-Lindquist coordinates (x, y, z) defined by Eq. (11.73) and scaled in units of m . The geodesic starts at $\lambda = 0$ and $\varphi = 0$ in the equatorial plane, in the direction of the Southern hemisphere ($d\theta/d\lambda > 0$). The left panel corresponds to $0 \leq \lambda \leq 7.7 \text{ m}/E$, while the right one extends the range to $0 \leq \lambda \leq 70 \text{ m}/E$. The grey sphere is the black hole event horizon. [Figure generated by the notebook D.5.12]

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 $d\varphi/d\lambda' = \text{const}$, 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 $a = 0.95 \text{ m}$ (same value as in Fig. 12.6). The event and inner horizon radii are $r_+ = 1.312 \text{ m}$ and $r_- = 0.688 \text{ m}$ and one has $r_{\text{ph}}^{**} = -0.478 \text{ m}$, $r_{\text{ph}}^* = 0.658 \text{ m}$, $r_{\text{ph}}^+ = 1.386 \text{ m}$ and $r_{\text{ph}}^- = 3.955 \text{ m}$. Some outer spherical photon orbits are plotted in Figs. 12.8–12.11 for various values of r_0 :

- $r_0 = 1.6 \text{ m}$ (Fig. 12.8): this orbit has $\ell = 2.171 \text{ m}$, $q = 5.976 \text{ m}^2$, $\theta_m = 0.706 \text{ rad}$ and $d\varphi/d\lambda > 0$;
- $r_0 = 2.8 \text{ m}$ (Fig. 12.9): this orbit has $\ell = -1.089 \text{ m}$, $q = 26.260 \text{ m}^2$ and $\theta_m = 0.206 \text{ rad}$; $\varphi(\lambda)$ is not a monotonic function (cf. the projection onto the xy -plane): $d\varphi/d\lambda < 0$ almost everywhere, except near the equator, where the Lense-Thirring effect (cf. Sec. 11.3.4) is the strongest² and enforces $d\varphi/d\lambda > 0$;
- $r_0 = 3 \text{ m}$ (Fig. 12.10): this orbit has $\ell = -1.9 \text{ m}$, $q = 27 \text{ m}^2$, $\theta_m = 0.346 \text{ rad}$ and $d\varphi/d\lambda \leq 0$, with $d\varphi/d\lambda = 0$ in the equatorial plane, as it can be easily checked by plugging $r = 3 \text{ m}$, $\theta = \pi/2$ 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);
- $r_0 = 3.9 \text{ m}$ (Fig. 12.11): this orbit has $\ell = -6.574 \text{ m}$, $q = 3.525 \text{ m}^2$, $\theta_m = 1.290 \text{ rad}$ and $d\varphi/d\lambda < 0$.

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

²This can be seen by considering the limit $\theta \rightarrow \pi/2$ in Eq. (12.23d).

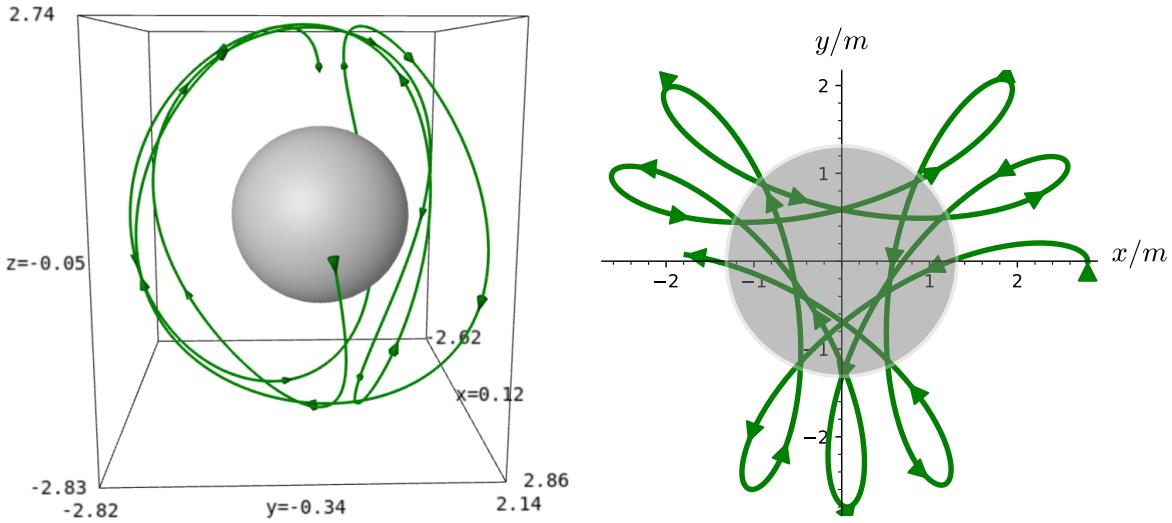


Figure 12.9: Spherical photon orbit at $r_0 = 2.8 \text{ m}$ around a Kerr black hole with $a = 0.95 \text{ m}$, depicted in terms of the Cartesian Boyer-Lindquist coordinates (x, y, z) defined by Eq. (11.73) and scaled in units of m , with the grey sphere representing the black hole event horizon. The 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 $0 \leq \lambda \leq 38 \text{ m}/E$. [Figure generated by the notebook D.5.12]

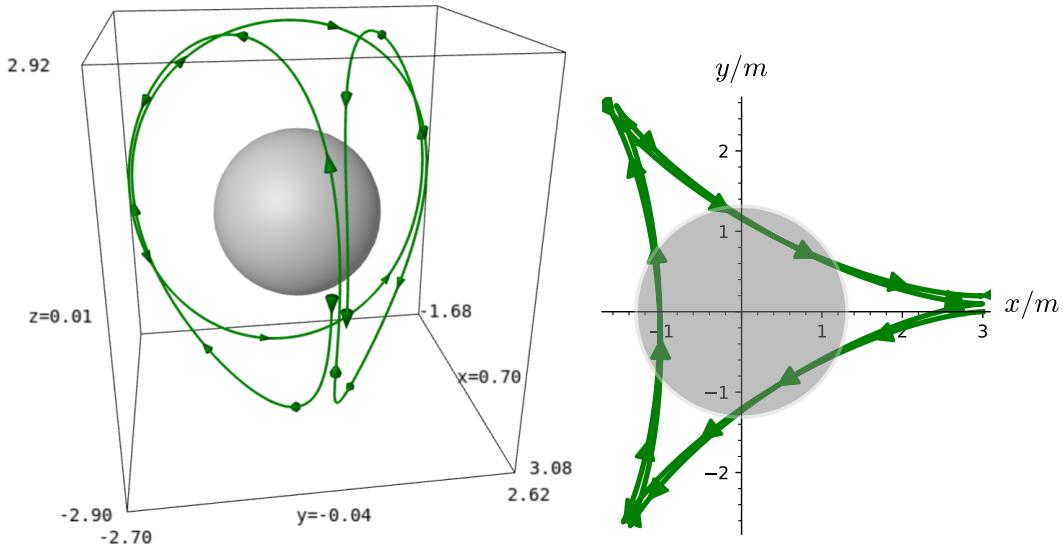


Figure 12.10: Same as in Fig. 12.9 but for a spherical photon orbit at $r_0 = 3 \text{ m}$, with $0 \leq \lambda \leq 32 \text{ m}/E$. [Figure generated by the notebook D.5.12]

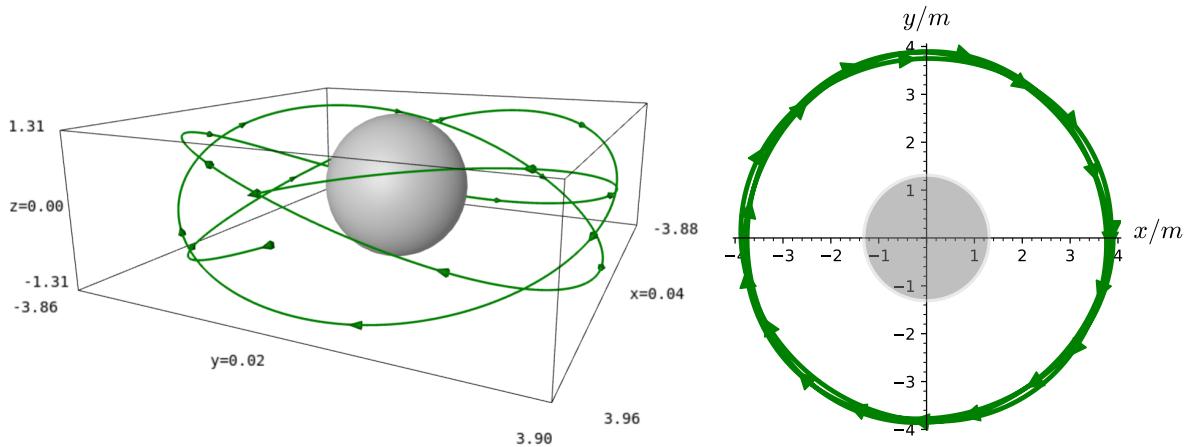


Figure 12.11: Same as in Fig. 12.9 but for a spherical photon orbit at $r_0 = 3.9 \text{ m}$, with $0 \leq \lambda \leq 55 \text{ m}/E$. [Figure generated by the notebook D.5.12]

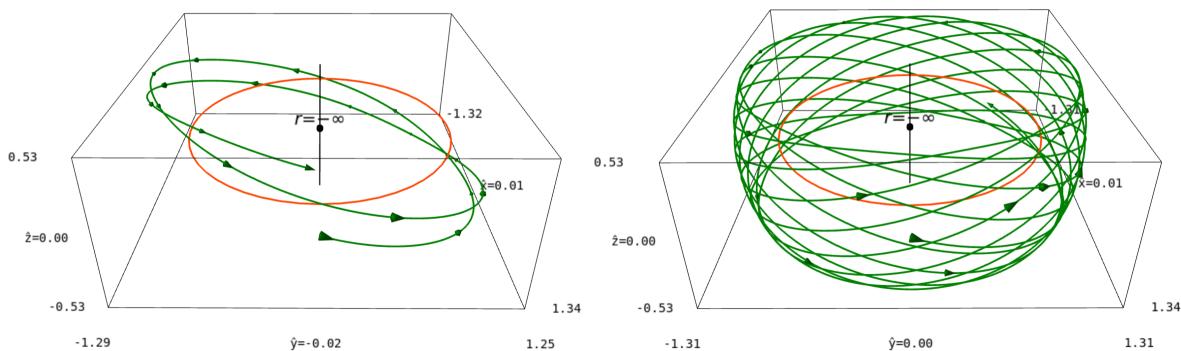


Figure 12.12: Marginally stable spherical photon orbit inside a Kerr black hole with $a = 0.95 \text{ m}$, depicted in terms of the coordinates $(\hat{x}, \hat{y}, \hat{z})$ defined by Eq. (12.69) and scaled in units of m . The geodesic is located at $r_0 = 0.540 \text{ m}$ and starts at $\lambda = 0$ and $\varphi = 0$ in the equatorial plane, in the direction of the Southern hemisphere ($d\theta/d\lambda > 0$). The left panel corresponds to $0 \leq \lambda \leq 3 \text{ m}/|E|$, while the right one extends the range to $0 \leq \lambda \leq 20 \text{ m}/|E|$. 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 $r \rightarrow -\infty$. [Figure generated by the notebook D.5.12]

$r < 0$. 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

$$\hat{r} := \frac{1}{2} \left(r + \sqrt{r^2 + 4m^2} \right) \quad (12.69a)$$

$$\hat{x} := \hat{r} \sin \theta \cos \varphi, \quad \hat{y} := \hat{r} \sin \theta \sin \varphi, \quad \hat{z} := \hat{r} \cos \theta. \quad (12.69b)$$

As O'Neill coordinates, this brings the whole range $(-\infty, +\infty)$ for r to $(0, +\infty)$ for \hat{r} , with $\hat{r} \rightarrow 0$ corresponding to $r \rightarrow -\infty$. Contrary to O'Neill coordinate $e^{r/m}$, 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 \mathcal{M}_I , \hat{r} is asymptotically equivalent to r : $\hat{r} \sim r$ as $r \rightarrow +\infty$.

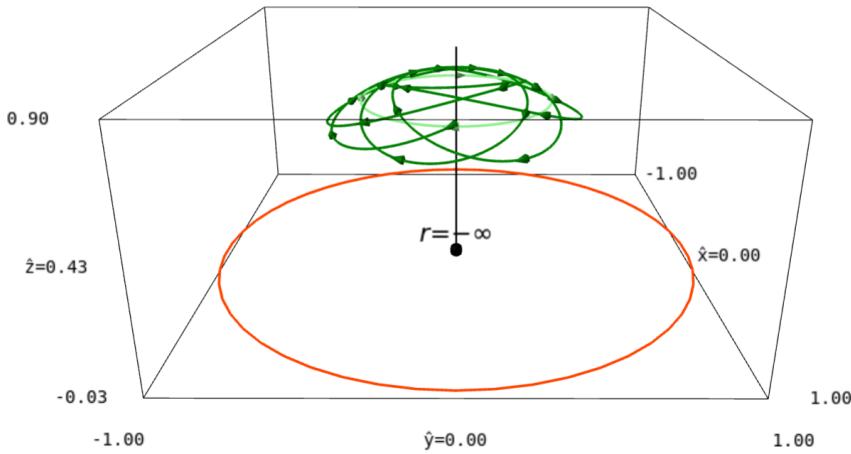


Figure 12.13: Same as in Fig. 12.12 but for a vortical spherical photon orbit at $r_0 = -0.46 \text{ m}$, with $0 \leq \lambda \leq 20 \text{ m}/E$. Also shown is the Northern vortical circular photon orbit at $r = r_{\text{ph}}^{**}$ (in light green). [Figure generated by the notebook D.5.12]

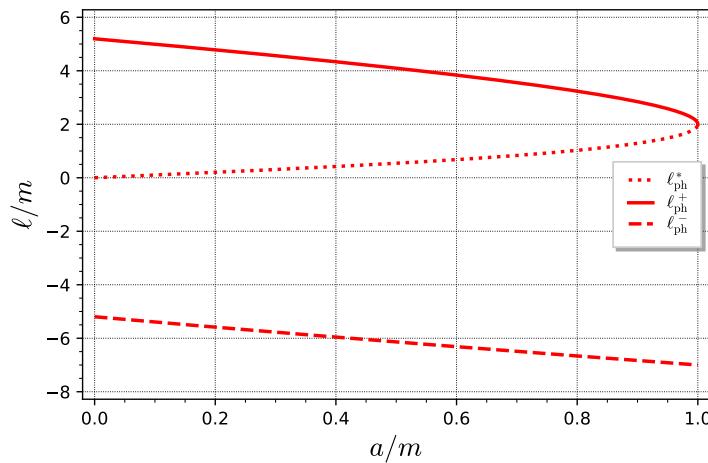


Figure 12.14: Reduced angular momentum ℓ of the three circular photon orbits in the equatorial plane, as a function of the Kerr spin parameter a . [Figure generated by the notebook D.5.11]

Example 7 (Inner spherical photon orbits): We consider the same Kerr spacetime as in Example 6, i.e. $a = 0.95 \text{ m}$, but this time, we focus on inner spherical photon orbits:

- $r_0 = 0.540 \text{ m}$ (Fig. 12.12): this orbit has $\ell = 1.539 \text{ m}$, $q = 0.282 \text{ m}^2$, $\theta_m = 1.173 \text{ rad}$, $d\varphi/d\lambda > 0$, $E < 0$ and $L < 0$; it is actually a marginally stable photon orbit ($r_0 = r_{\text{ph}}^{\text{ms}}$), as we shall see in Sec. 12.3.4;
- $r_0 = -0.46 \text{ m}$ (Fig. 12.13): this orbit has $\ell = -0.176 \text{ m}$, $q = -0.461 \text{ m}^2$ (hence it is vortical), $\theta_m = 0.282 \text{ rad}$, $\theta_v = 0.731 \text{ rad}$, $d\varphi/d\lambda < 0$, $E > 0$ and $L < 0$;
- $r_0 = -0.478 \text{ m} = r_{\text{ph}}^{**}$ (Fig. 12.13, light green curve): this orbit has $\ell = -0.229 \text{ m}$, $q = -0.519 \text{ m}^2$ (hence it is vortical), $\theta_m = \theta_v = \theta_{\text{ph}}^{**} = 0.514 \text{ rad}$, $d\varphi/d\lambda < 0$, $E > 0$ 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 \mathcal{L} whose tangent vector field $\mathbf{p} = dx/d\lambda$ is a linear combination of the two Killing vectors ξ and η generating respectively the stationarity and the axisymmetry, with a non-vanishing component along η :

$$\mathbf{p} = \alpha \xi + \beta \eta, \quad \text{with } \beta \neq 0. \quad (12.70)$$

Note that the above definition is independent from any coordinate system. It is worth to compare Eq. (12.70) with expression (11.178) for the 4-velocity $\mathbf{u} = \mu^{-1} \mathbf{p}$ of a circular timelike orbit in the equatorial plane.

Remark 3: Circular photon orbits are sometimes called *light rings* [143].

In Kerr spacetime, when using coordinates adapted to the spacetime symmetries, such as Boyer-Lindquist coordinates (t, r, θ, φ) , one has $\xi = \partial_t$, $\eta = \partial_\varphi$, $\alpha = p^t$, $\beta = p^\varphi$ and the definition (12.70) is equivalent to $p^r = dr/d\lambda = 0$ and $p^\theta = d\theta/d\lambda = 0$. 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 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 $\theta' = \text{const}$ 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 \mathcal{H} and \mathcal{H}_{in} are circular photon orbits, since they are the outgoing principal null geodesics $\mathcal{L}_{(\theta, \psi)}^{\text{out}, \mathcal{H}}$ and $\mathcal{L}_{(\theta, \psi)}^{\text{out}, \mathcal{H}_{\text{in}}}$, which have $\theta = \text{const}$.

The spherical photon orbits with $r_0 = r_{\text{ph}}^*$, r_{ph}^+ and r_{ph}^- have $q = 0$ and $|\ell| > a$. According to the results of Sec. 12.2.4, they necessarily 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 $q_c(r_0) = 0$ is $r_0 = 0$ (see also Fig. 12.6), which would correspond to the ring singularity. Let us denote by ℓ_{ph}^* , ℓ_{ph}^+ and ℓ_{ph}^- the reduced angular momentum ℓ of respectively the circular photon orbit at r_{ph}^* , r_{ph}^+ and r_{ph}^- . These values of ℓ are deduced from Eq. (12.56) and are plotted in terms of a in Fig. 12.14. We have $\ell_{\text{ph}}^* > 0$, $\ell_{\text{ph}}^+ > 0$ and $\ell_{\text{ph}}^- < 0$. Moreover,

$$\lim_{a \rightarrow 0} \ell_{\text{ph}}^+ = \lim_{a \rightarrow 0} |\ell_{\text{ph}}^-| = 3\sqrt{3}m \simeq 5.196 m, \quad (12.71)$$

in agreement with the Schwarzschild result (8.21), while

$$\lim_{a \rightarrow m} \ell_{\text{ph}}^+ = \lim_{a \rightarrow m} \ell_{\text{ph}}^* = 2m, \quad (12.72)$$

in agreement with Eq. (12.50), since $\lim_{a \rightarrow m} r_{\text{ph}}^+ = \lim_{a \rightarrow m} r_{\text{ph}}^* = m$.

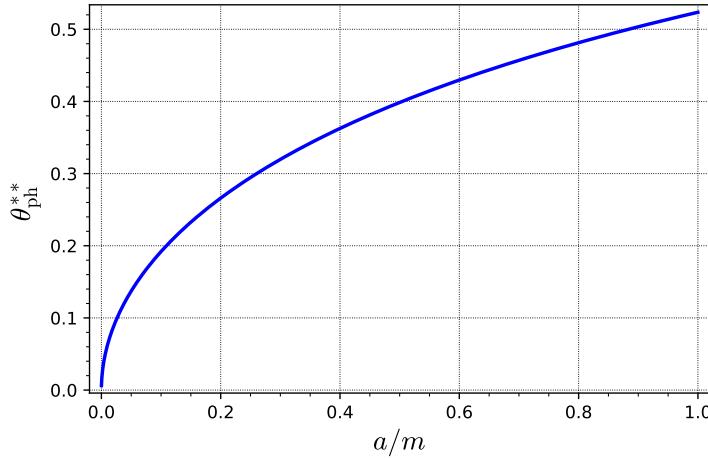


Figure 12.15: Angle θ of the Northern vortical circular photon orbit at $r = r_{\text{ph}}^{**}$ as a function of Kerr spin parameter a . [Figure generated by the notebook D.5.11]

The spherical orbits with $r_0 = r_{\text{ph}}^{**}$ have $q < 0$, i.e. are vortical geodesics. Moreover, they fulfill $q = q_{\min} = -(a - |\ell|)^2$. 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 $\theta = \theta_{\text{ph}}^{**}$ (Northern hemisphere) and $\theta = \pi - \theta_{\text{ph}}^{**}$ (Southern hemisphere), where θ_{ph}^{**} is given by Eq. (12.44), using for ℓ the value (12.56) with $r_0 = r_{\text{ph}}^{**}$. Since $3m(r_{\text{ph}}^{**})^2 = 2(r_{\text{ph}}^{**})^3 + a^2m$ by virtue of Eq. (12.63), some simplification occurs and we get

$$\theta_{\text{ph}}^{**} = \arcsin \sqrt{\frac{|r_{\text{ph}}^{**}|}{m - r_{\text{ph}}^{**}} \left(1 - \frac{(r_{\text{ph}}^{**})^2}{a^2}\right)}. \quad (12.73)$$

θ_{ph}^{**} is an increasing function of a/m , plotted in Fig. 12.15. We have $\lim_{a \rightarrow 0} \theta_{\text{ph}}^{**} = 0$ (the rotation axis) and $\lim_{a \rightarrow m} \theta_{\text{ph}}^{**} = \pi/6$. 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 ℓ_{ph}^{**} and q_{ph}^{**} respectively, are depicted in terms of a in Fig. 12.16. They tend to zero as $a \rightarrow 0$ and obey

$$\lim_{a \rightarrow m} \ell_{\text{ph}}^{**} = -\frac{m}{4} \quad \text{and} \quad \lim_{a \rightarrow m} q_{\text{ph}}^{**} = -\frac{9}{16}m^2. \quad (12.74)$$

We may summarize the above results by

Property 12.14: circular photon orbits in Kerr spacetime

The geodesics at the boundaries of the domain of existence of spherical photon orbits [Eq. (12.59)] are circular orbits:

- the orbit at $(r, \theta) = (r_{\text{ph}}^{**}, \theta_{\text{ph}}^{**})$ is called the **Northern vortical circular photon orbit**;
- the orbit at $(r, \theta) = (r_{\text{ph}}^{**}, \pi - \theta_{\text{ph}}^{**})$ is called the **Southern vortical circular photon orbit**;

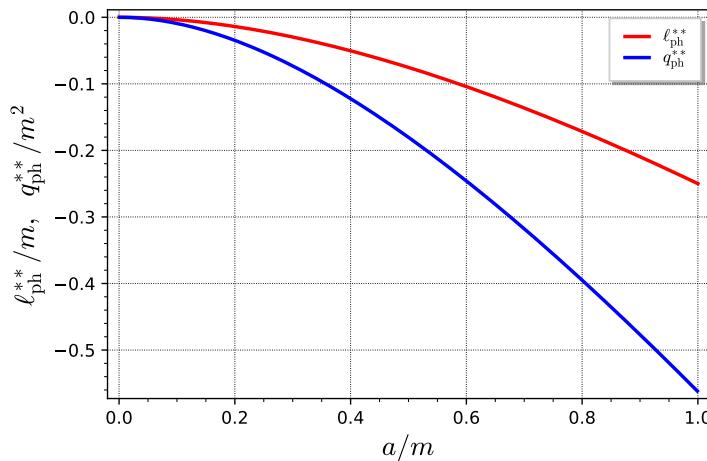


Figure 12.16: Reduced angular momentum ℓ_{ph}^{**} and reduced Carter constant q_{ph}^{**} of the two vortical circular photon orbits at $r = r_{\text{ph}}^{**}$, as functions of Kerr spin parameter a . [Figure generated by the notebook D.5.11]

- the orbit at $(r, \theta) = (r_{\text{ph}}^*, \pi/2)$ is called the **equatorial inner circular photon orbit**;
- the orbit at $(r, \theta) = (r_{\text{ph}}^+, \pi/2)$ is called the **prograde outer circular photon orbit**;
- the orbit at $(r, \theta) = (r_{\text{ph}}^-, \pi/2)$ 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 \mathcal{H} and \mathcal{H}_{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 $q_{\min} < q < 0$ (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:

Property 12.15: stability criterion for spherical photon orbits

$$\text{The spherical photon orbit of radius } r_0 \text{ is stable} \iff \mathcal{R}''(r_0) < 0. \quad (12.75)$$

Remark 5: The criterion (11.159) 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 Lemma 12.5 is

Property 12.16: instability of spherical photon orbits in the black hole exterior

All spherical photon orbits in the black hole exterior are unstable.

Proof. The quartic polynomial $\mathcal{R}(r)$ of spherical photon orbits in \mathcal{M}_I belong to case 2 of Lemma 12.5, 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). \square

We are however going to see that there exist stable spherical photon orbits in region \mathcal{M}_{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

$$\mathcal{R}''(r_0) = \frac{8r_0}{(r_0 - m)^2} \left[(r_0 - m)^3 + m^3 \left(1 - \frac{a^2}{m^2} \right) \right]. \quad (12.76)$$

Hence the stability criterion (12.75) becomes

$$\mathcal{R}''(r_0) < 0 \iff \begin{cases} r_0 < 0 & \text{and} \quad (r_0 - m)^3 + m^3 \left(1 - \frac{a^2}{m^2} \right) > 0 \\ \text{or} \\ r_0 > 0 & \text{and} \quad (r_0 - m)^3 + m^3 \left(1 - \frac{a^2}{m^2} \right) < 0. \end{cases} \quad (12.77)$$

Now

$$(r_0 - m)^3 + m^3 \left(1 - \frac{a^2}{m^2} \right) > 0 \iff r_0 - m > -m \left(1 - \frac{a^2}{m^2} \right)^{1/3} \iff r_0 > r_{ph}^{ms},$$

where

$$r_{ph}^{ms} := m \left[1 - \left(1 - \frac{a^2}{m^2} \right)^{1/3} \right]. \quad (12.78)$$

Since obviously $r_{ph}^{ms} \geq 0$, we conclude that the first case in (12.77) is excluded and that the second case holds for $0 < r_0 < r_{ph}^{ms}$:

Property 12.17: stability region for spherical photon orbits

$$\text{A spherical photon orbit of radius } r_0 \text{ is stable} \iff 0 < r_0 < r_{ph}^{ms}. \quad (12.79)$$

The index ‘ms’ in r_{ph}^{ms} stands for **marginally stable**. r_{ph}^{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

$$r_{ph}^{ms} \leq r_{ph}^*, \quad (12.80)$$

with equality iff $a = m$. We conclude:

Property 12.18: stability of spherical photon orbits

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 $r_0 \in (0, r_{\text{ph}}^{\text{ms}})$, where $r_{\text{ph}}^{\text{ms}}$ is given by Eq. (12.78). In particular, all spherical photon orbits in \mathcal{M}_1 (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 Property 12.10, given the inequality (12.80).

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, θ_m and $\pi - \theta_m$, such that $0 < \theta_m < \pi/2$. In particular, $r_0 = r_{\text{ms}}$ 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 $\theta = \theta_m$ (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:

Property 12.19: characterization of marginally stable spherical photon orbits

Among all inner spherical photon orbits, the marginally stable orbits at $r_0 = r_{\text{ph}}^{\text{ms}}$ 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

$$\frac{dq_c}{dr_0} = -\frac{4r_0^2(r_0 - 3m)}{a^2(r_0 - m)^3} \left[(r_0 - m)^3 + m^3 \left(1 - \frac{a^2}{m^2} \right) \right] \quad (12.81)$$

$$\frac{d\ell_c}{dr_0} = -2 \frac{(r_0 - m)^3 + m^3 \left(1 - \frac{a^2}{m^2} \right)}{a(r_0 - m)^2} \quad (12.82)$$

Since $r_{\text{ph}}^{\text{ms}}$ 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 $\ell_c(r_0)$ has a unique extremum, which is achieved by the marginally stable orbits. From the graph of $\ell_c(r_0)$ shown in Fig. 12.6, we see that this extremum is a maximum. Regarding the function $q_c(r_0)$, the above expression of dq_c/dr_0 leads to two extrema: $r_0 = r_{\text{ph}}^{\text{ms}}$ 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 $r_0 = r_{\text{ph}}^{\text{ms}}$ realizes a maximum of q_c among all inner spherical orbits. \square

Remark 6: Another proof can be given by using the same general argument as that employed in Sec. 11.5.3 for showing that the ISCO realizes an extremum of the specific energy and specific angular momentum of timelike circular equatorial orbits (Property 11.32). 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

$$\mathcal{R}(r_0, \ell_c(r_0), q_c(r_0)) = 0, \quad \frac{\partial \mathcal{R}}{\partial r}(r_0, \ell_c(r_0), q_c(r_0)) = 0, \quad \frac{\partial^2 \mathcal{R}}{\partial r^2}(r_0, \ell_c(r_0), q_c(r_0)) = 0.$$

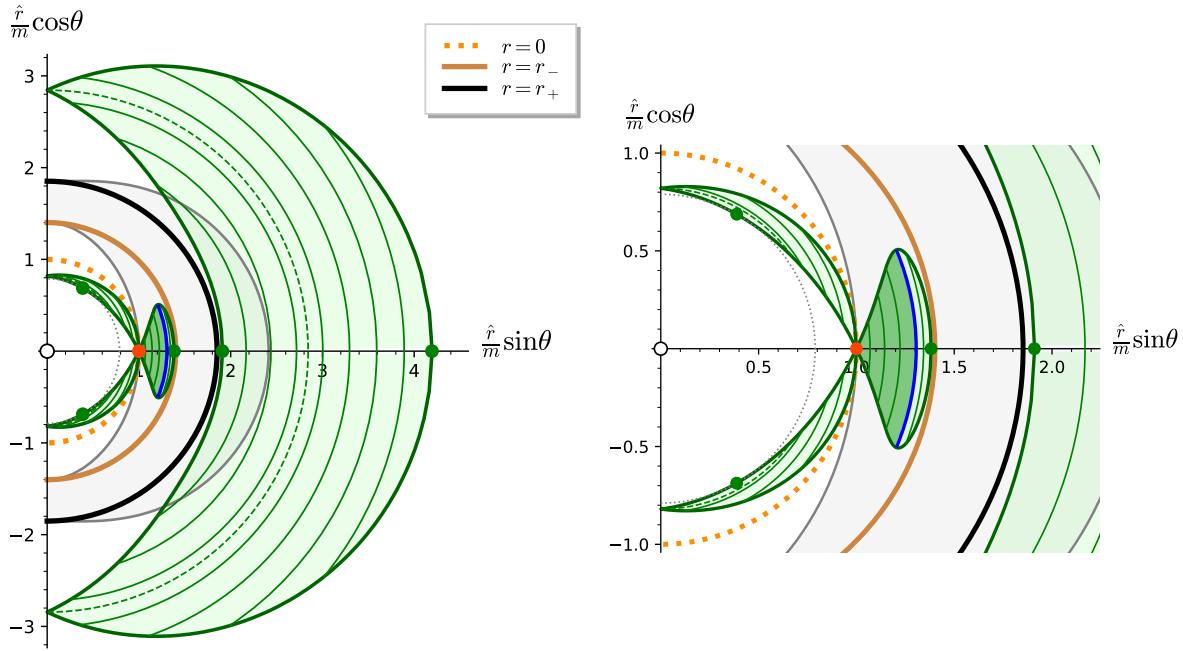


Figure 12.17: Trace of the photon region (pale and dark green areas) in a meridional plane $(t, \varphi) = \text{const}$ of a Kerr spacetime with $a = 0.95 m$ (as in Fig. 12.6). The right panel is a zoom on the inner part of the left one. The meridional plane is described in terms of the coordinates (\hat{r}, θ) 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 $r = r_{\text{ph}}^{\text{ms}}$ drawn in blue. Green dots mark photon circular orbits: from the left to the right, they are the two vortical circular orbits at $r = r_{\text{ph}}^{**}$, the equatorial inner circular orbit at $r = r_{\text{ph}}^*$, the prograde outer circular orbit at $r = r_{\text{ph}}^+$ and the retrograde outer circular orbit at $r = r_{\text{ph}}^-$. The thin dotted grey half-circle marks $r = r_{\text{ph}}^{**}$. [Figure generated by the notebook D.5.11]

Deriving the first and second equations with respect to r_0 and using the third equation, we get a homogeneous linear system for $(d\ell_c/dr_0, dq_c/dr_0)$, similar to the system (11.172). The unique solution is then $(d\ell_c/dr_0, dq_c/dr_0) = (0, 0)$, yielding the extremum in the functions $\ell_c(r_0)$ and $q_c(r_0)$ at $r_0 = r_{\text{ph}}^{\text{ms}}$.

Equation (12.81) shows that, in addition to that at $r_0 = r_{\text{ph}}^{\text{ms}}$, the function $q_c(r_0)$ admits a second extremum at $r_0 = 3m$, i.e. for outer spherical orbits. This extremum is actually a maximum and its value, obtained by setting $r_0 = 3m$ in Eq. (12.57), turns out to be independent from a :

$$\max q_c(r_0) = 27m^2. \quad (12.83)$$

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 \mathcal{M}_I , a photon cannot be trapped (i.e. have 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** [303]. 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 θ_m given by Eq. (12.41) ($\theta_m = \pi/2$ for the three circular orbits at $r_0 = r_{\text{ph}}^*$, $r_0 = r_{\text{ph}}^+$ and $r_0 = r_{\text{ph}}^-$, which lie in the equatorial plane). On the other side, the inner orbits with $r_0 < 0$ are all vertical, with the limiting angles θ_m and θ_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 $r_0 = r_{\text{ph}}^{\text{pol}}$ and $r_0 = r_{\text{ph}}^{\text{pol,in}}$ (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 \mathcal{H}_{in} and \mathcal{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 [41]. The existence of stable spherical photon orbits under the inner horizon of a Kerr black hole has been shown by Zdeněk Stuchlík in 1981 [465]. The systematic study of spherical photon orbits in the black hole exterior has been performed by Edward Teo in 2003 [474].

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 $|\ell| = b_c = 3\sqrt{3} m$. In the Kerr case with $0 < a < m$, we have instead a 1-parameter family of critical values: the family $(\ell_c(r_0), q_c(r_0))$ given by Eqs. (12.56)–(12.57), the parameter being r_0 . By construction, the polynomial $\mathcal{R}(r)$ of a critical null geodesic \mathcal{L} of parameter³ r_0 has a double root at $r = r_0$ [cf. Eq. (12.47)]. More precisely, inserting formula

³The 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 \mathcal{L} .

(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

$$\mathcal{R}(r) = (r - r_0)^2 \left(r^2 + 2r_0r - \frac{a^2 q_c(r_0)}{r_0^2} \right), \quad (12.84)$$

where $q_c(r_0)$ is the function (12.57). This writing of $\mathcal{R}(r)$ clearly exhibits the double root at $r = r_0$. As discussed in Sec. 11.3.6 and 12.2.5, a consequence of the double-root behavior is that any critical null geodesic \mathcal{L} of parameter r_0 has r_0 as an asymptotic r -value, i.e. $r(\lambda) \rightarrow r_0$ when $\lambda \rightarrow +\infty$ or $\lambda \rightarrow -\infty$, where λ is the affine parameter along \mathcal{L} .

Let us consider a critical null geodesic \mathcal{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 \mathcal{L} ,

$$\oint_{r_A}^r \frac{\epsilon_r d\bar{r}}{|\bar{r} - r_0| \sqrt{\bar{r}^2 + 2r_0\bar{r} - a^2 q_c(r_0)/r_0^2}} = \oint_{\theta_A}^\theta \frac{\epsilon_\theta d\bar{\theta}}{\sqrt{\tilde{\Theta}(\bar{\theta})}}. \quad (12.85)$$

When $r \rightarrow r_0$, 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 \mathcal{L} has an endless oscillatory θ -motion as $r \rightarrow r_0$. Similarly the integrated equation of motion (11.55d) with expression (12.84) for $\mathcal{R}(r)$ becomes

$$\varphi - \varphi_{\text{em}} = a \oint_{r_A}^r \frac{(2m\bar{r} - a\ell) \epsilon_r d\bar{r}}{|\bar{r} - r_0| (\bar{r}^2 - 2m\bar{r} + a^2) \sqrt{\bar{r}^2 + 2r_0\bar{r} - a^2 q_c(r_0)/r_0^2}} + \ell \oint_{\theta_A}^\theta \frac{\epsilon_\theta d\bar{\theta}}{\sin^2 \bar{\theta} \sqrt{\tilde{\Theta}(\bar{\theta})}}. \quad (12.86)$$

Again, the first integral diverges when $r \rightarrow r_0$. 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 $\epsilon_r d\bar{r} > 0$ and $\epsilon_\theta d\bar{\theta} > 0$, we conclude that the integral on r and the path integral on θ both tend to $+\infty$ when $r \rightarrow r_0$. If $\ell \geq 0$, we get then immediately $\varphi \rightarrow +\infty$ when $r \rightarrow r_0$. 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 ,

$$\varphi \rightarrow \pm\infty \quad \text{when} \quad r \rightarrow r_0. \quad (12.87)$$

We may summarize the above results as follows:

Property 12.20: radial motion of critical null geodesics

A critical null geodesic \mathcal{L} of parameter r_0 has r_0 as an asymptotic r -value, either in the future ($r(\lambda) \rightarrow r_0$ when $\lambda \rightarrow +\infty$) or in the past ($r(\lambda) \rightarrow r_0$ when $\lambda \rightarrow -\infty$), λ being the (future-directed) affine parameter of \mathcal{L} . In particular, \mathcal{L} never crosses the sphere $r = r_0$. Moreover \mathcal{L} is winding endlessly on the sphere $r = r_0$ when either $\lambda \rightarrow +\infty$ or $\lambda \rightarrow -\infty$,

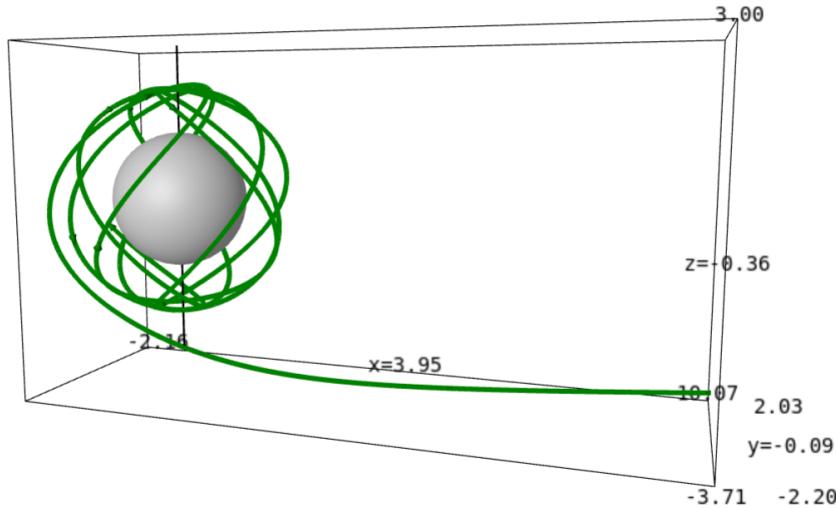


Figure 12.18: Critical null geodesic of parameter $r_0 = 2.2 \text{ m}$ in a Kerr spacetime with $a = 0.95 \text{ m}$, depicted in terms of the Cartesian Boyer-Lindquist coordinates (x, y, z) [Eq. (11.73)]. This geodesic has $\ell = 0.863 \text{ m}$, $q = 18.042 \text{ m}^2$ and is emitted at $\lambda = 0$ from the point of Boyer-Lindquist coordinates $(r, \theta, \varphi) = (40 \text{ m}, \pi/2, 0) \iff (x, y, z) = (40 \text{ m}, 0, 0)$, towards the black hole ($\epsilon_r = -1$). The drawing is interrupted at $\lambda = 80 \text{ m}/E$. The grey sphere is the black hole event horizon at $r = r_+ = 1.312 \text{ m}$. [Figure generated by the notebook D.5.13]

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 $r \sim 10 \text{ m}$). We note the winding around the sphere $r = r_0 = 2.2 \text{ m}$, in the same fashion as a spherical photon orbit (compare with Figs. 12.8 and 12.9).

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 \mathcal{M}_1 , 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 Lemma 12.5, this family separates null geodesics whose quartic polynomial \mathcal{R} has no root in \mathcal{M}_1 from those for which \mathcal{R} has a double root in \mathcal{M}_1 . It follows that the family of critical null geodesics separates null geodesics that have a r -turning point in \mathcal{M}_1 from those that do not have any, given that a null geodesic has at most one r -turning point in \mathcal{M}_1 (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, $\ell_c(r_0)$, up to a relative difference of 10^{-4} . The geodesic \mathcal{L}_1 (green curve) has $\ell = 1.0001 \ell_c(r_0)$; it performs a few turns onto (actually very close to) the sphere \mathcal{S}_0 of coordinate radius $r_0 = 2.2 \text{ m}$ and eventually depart to infinity; this means that \mathcal{L}_1 has a r -turning point, somewhere

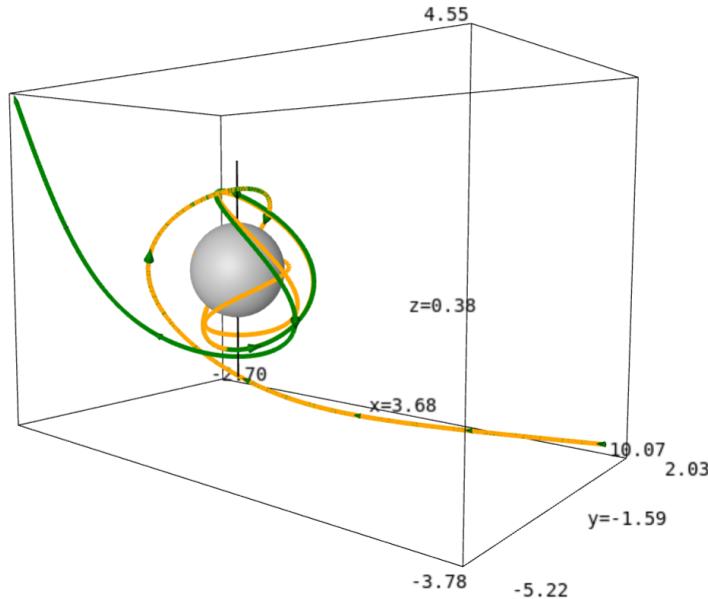


Figure 12.19: Two null geodesics with constants of motion (ℓ, q) close to those of a critical null geodesic, $(\ell_c(r_0), q_c(r_0))$ [Eqs. (12.56)–(12.57)]. Here, as in Fig. 12.18, $r_0 = 2.2 \text{ m}$ and $a = 0.95 \text{ m}$. The two geodesics depart from the same point at $(x, y, z) = (40 \text{ m}, 0, 0)$ (outside the scope of the figure) as the critical null geodesic shown in Fig. 12.18. Both geodesics have $q = q_c(r_0)$, but the green one has $\ell = 1.0001 \ell_c(r_0)$ while the orange one has $\ell = 0.9999 \ell_c(r_0)$. [Figure generated by the notebook D.5.13]

close to \mathcal{S}_0 . The geodesic \mathcal{L}_2 (orange curve) has $\ell = 0.9999 \ell_c(r_0)$; it is graphically indistinguishable from \mathcal{L}_1 until a full turn onto \mathcal{S}_0 (this is best seen in the interactive 3D view in the online notebook D.5.13), it then clearly departs from it and eventually terminates into the black hole. Hence \mathcal{L}_2 has no r -turning point.

Let us consider an asymptotic inertial observer \mathcal{O} (cf. Sec. 10.7.5), i.e. a static observer located at Boyer-Lindquist coordinates $(t, r_\mathcal{O}, \theta_\mathcal{O}, \varphi_\mathcal{O})$, where $(r_\mathcal{O}, \theta_\mathcal{O}, \varphi_\mathcal{O})$ are constant and $r_\mathcal{O} \gg m$. As in the Schwarzschild case (Sec. 8.5.4), the concept of black hole shadow is defined by considering an emitting large sphere \mathcal{S} of constant Boyer-Lindquist coordinate $r_\mathcal{S}$ that encompasses both the black hole and observer \mathcal{O} ; this means that $r_\mathcal{S} > r_\mathcal{O}$. In a more astrophysical setting, one could image \mathcal{S} as being made of many far-away light sources. We assume that \mathcal{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 \mathcal{O} , the black hole and the emitting sphere \mathcal{S} , any photon emitted from \mathcal{S} that reaches \mathcal{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 \mathcal{S} and \mathcal{O} . Reciprocally, any null geodesic that impacts \mathcal{O} 's screen and has no r -turning point in its past cannot have emerged from \mathcal{S} and therefore results in a black dot on the screen. According to the above discussion, the boundary of the black area on \mathcal{O} 's screen is made by the impact points of critical null geodesics and is called the **critical curve** [233, 234]; we shall denote it by \mathcal{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 \mathcal{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 \mathcal{O} is not located on the black hole rotation axis, i.e. $\theta_{\mathcal{O}} \notin \{0, \pi\}$, leaving the special case $\theta_{\mathcal{O}} = 0$ or π to Sec. 12.4.3. Then $\sin \theta_{\mathcal{O}} \neq 0$ and we have seen in Sec. 12.2.3 that the impact point of a null geodesic \mathcal{L} on \mathcal{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 \mathcal{L} by formulas (12.32). Given the above definition, the critical curve \mathcal{C} is obtained by using for (ℓ, q) in (12.32) the critical values $(\ell_c(r_0), q_c(r_0))$ given by Eqs. (12.56)–(12.57):

$$\alpha = -\frac{\ell_c(r_0)}{r_{\mathcal{O}} \sin \theta_{\mathcal{O}}} \quad (12.88a)$$

$$\beta = \frac{\epsilon_{\theta}}{r_{\mathcal{O}}} \sqrt{q_c(r_0) + \cos^2 \theta_{\mathcal{O}} \left(a^2 - \frac{\ell_c(r_0)^2}{\sin^2 \theta_{\mathcal{O}}} \right)}. \quad (12.88b)$$

This provides the equation of \mathcal{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 \mathcal{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 $r_{\text{ph}}^+ \leq r_0 \leq r_{\text{ph}}^-$ [Eq. (12.59)], with r_{ph}^+ and r_{ph}^- given by Eq. (12.62). The corresponding range of $\ell = \ell_c(r_0)$ is then $\ell_c(r_{\text{ph}}^-) \leq \ell \leq \ell_c(r_{\text{ph}}^+)$ (cf. the red curve in Fig. 12.6) and the corresponding range of $q = q_c(r_0)$ 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 \mathcal{O} , they have to obey the constraint (12.29):

$$(q_c(r_0) + a^2 \cos^2 \theta_{\mathcal{O}}) \sin^2 \theta_{\mathcal{O}} - \ell_c(r_0)^2 \cos^2 \theta_{\mathcal{O}} \geq 0. \quad (12.89)$$

Except for $\theta_{\mathcal{O}} = \pi/2$ (observer in the equatorial plane), this constraint limits the range of r_0 to a subinterval $[r_0^{\min}, r_0^{\max}]$ of $[r_{\text{ph}}^+, r_{\text{ph}}^-]$. We note that the radius $r_{\text{ph}}^{\text{pol}}$ of polar spherical photon orbits, as given by Eq. (12.67a), lies necessarily in that subinterval; indeed by definition $\ell_c(r_{\text{ph}}^{\text{pol}}) = 0$, which ensures that the constraint (12.89) is fulfilled for $r_0 = r_{\text{ph}}^{\text{pol}}$, whatever the value of $\theta_{\mathcal{O}}$.

The critical curve \mathcal{C} in the observer's screen is obtained by first solving Eq. (12.89) with equality in the sign \geq to determine r_0^{\min} and r_0^{\max} . Then, the upper half of \mathcal{C} is computed by means of formulas (12.88) with $\epsilon_{\theta} = +1$ and r_0 ranging in $[r_0^{\min}, r_0^{\max}]$. 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 \mathcal{O} . For $\theta_{\mathcal{O}} = \pi/2$ the parameter r_0 of the critical null geodesics forming \mathcal{C} spans the full interval $[r_{\text{ph}}^+, r_{\text{ph}}^-] \simeq [1.386 m, 3.955 m]$. For $\theta_{\mathcal{O}} = \pi/6$, the range is restricted to $[r_0^{\min}, r_0^{\max}] \simeq [1.768 m, 3.237 m]$, while for $\theta_{\mathcal{O}} = 0$ (to be discussed

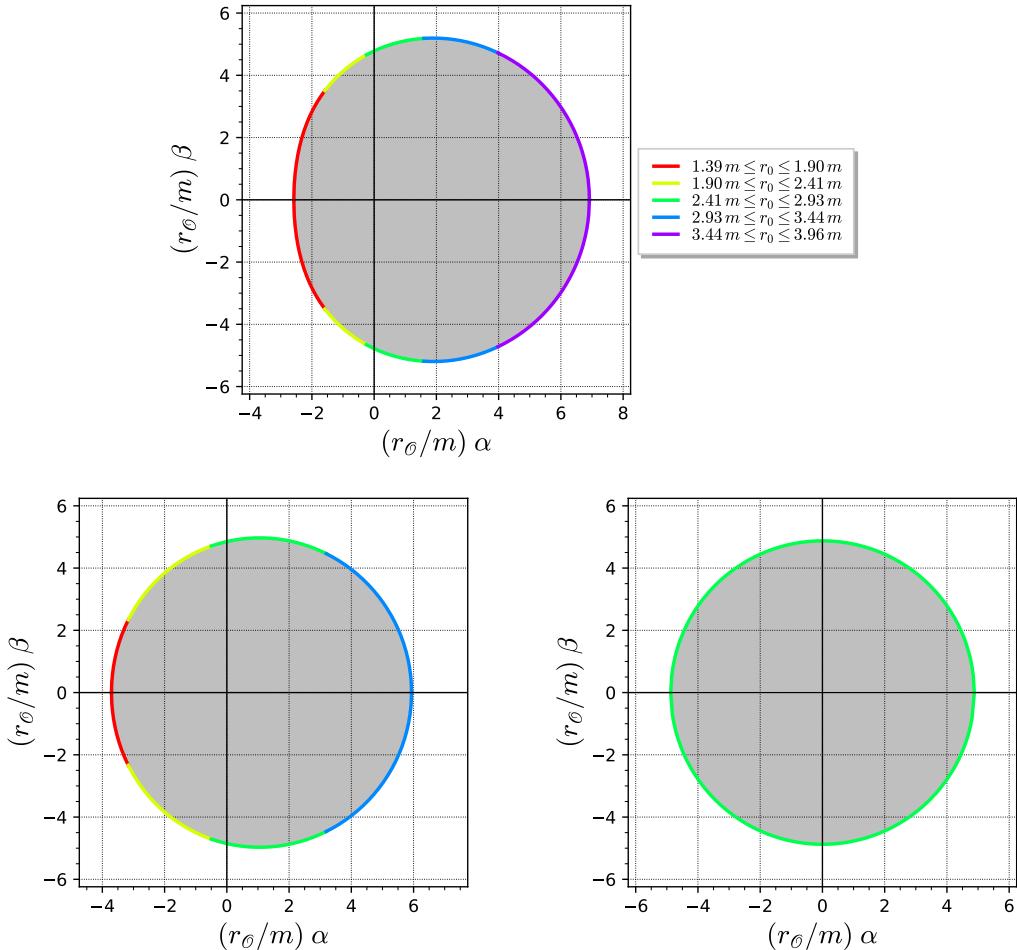


Figure 12.20: Shadow of a Kerr black hole of mass m and spin parameter $a = 0.95 m$ on the screen of an asymptotic inertial observer \mathcal{O} located at $r = r_\mathcal{O}$ and $\theta = \theta_\mathcal{O}$, for three values of $\theta_\mathcal{O}$: $\theta_\mathcal{O} = \pi/2$ (\mathcal{O} in the equatorial plane; upper panel), $\theta_\mathcal{O} = \pi/6$ (lower left) and $\theta_\mathcal{O} = 0$ (\mathcal{O} on the rotation axis; lower right). The screen is spanned by the angular coordinates (α, β) , rescaled by the inverse of the factor $m/r_\mathcal{O}$, whose values for some astrophysical black holes can be found in Table 12.1. The shadow is bounded by the critical curve \mathcal{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.5.14]

	Sgr A*	M87*	M31*	Cyg X-1
$m [M_\odot]$	$4.1 \cdot 10^6$	$6.2 \cdot 10^9$	$1.5 \cdot 10^8$	15
$r_\theta [\text{kpc}]$	8.12	$1.67 \cdot 10^4$	$7.6 \cdot 10^2$	1.86
m/r_θ	$2.4 \cdot 10^{-11}$	$1.8 \cdot 10^{-11}$	$9.4 \cdot 10^{-12}$	$3.9 \cdot 10^{-16}$
$m/r_\theta [\mu\text{as}]$	5.0	3.7	1.9	$8.0 \cdot 10^{-5}$

Table 12.1: Scale factor m/r_θ 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. [4]), the supermassive black hole M87* in the nucleus of the galaxy Messier 87 (data from Refs. [213, 6]; see also Appendix I and Table 9 of Ref. [8]), the supermassive black hole M31* in the nucleus of the Andromeda Galaxy (Messier 31) (data from Ref. [54]) and the stellar black hole Cygnus X-1 (data from Refs. [392, 427, 224]). The last line gives m/r_θ in microarcseconds ($1 \mu\text{as} = 4.848 \cdot 10^{-12} \text{ rad}$).

in Sec. 12.4.3), r_0 can take only one value: $r_0 = r_{\text{ph}}^{\text{pol}} \simeq 2.493 m$. At a fixed value of a/m , the black hole shadow depends on m and r_θ only through the dimensionless ratio m/r_θ , which sets the global scale of the shadow. Accordingly, in Fig. 12.20, the screen angular coordinates (α, β) have been rescaled by $(m/r_\theta)^{-1}$. Values of m/r_θ for some astrophysical black holes are provided in Table 12.1; the observer's radial coordinate r_θ is then nothing but the distance of the black hole to the Earth. These values of m/r_θ are tiny, being at most $2.4 \cdot 10^{-11} = 5.0 \mu\text{as}$ (microarcseconds; $1 \mu\text{as} = 4.848 \cdot 10^{-12} \text{ rad}$) 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_θ used in Fig. 12.20, this means that the angular size of Sgr A* shadow as seen from Earth is only $\sim 50 \mu\text{as}$. Such a small value⁴ has entered recently in the realm of observational astronomy, with the advent of the Event Horizon Telescope [6, 520], whose angular resolution is of order $20 \mu\text{as}$. 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 $\theta_\theta = \pi/6$ and even more for $\theta_\theta = \pi/2$, the shadow is shifted to the right and its left edge is flattened, as compared to the shadow for $\theta_\theta = 0$ (or π). This can be understood by noticing that the critical null geodesics forming the left edge arise from 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 $r_0 = r_{\text{ph}}^+$ 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 $r_0 = r_{\text{ph}}^-$. 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.

⁴For comparison, the angular resolution of the Hubble Space Telescope is $\sim 0.1'' = 10^5 \mu\text{as}$!

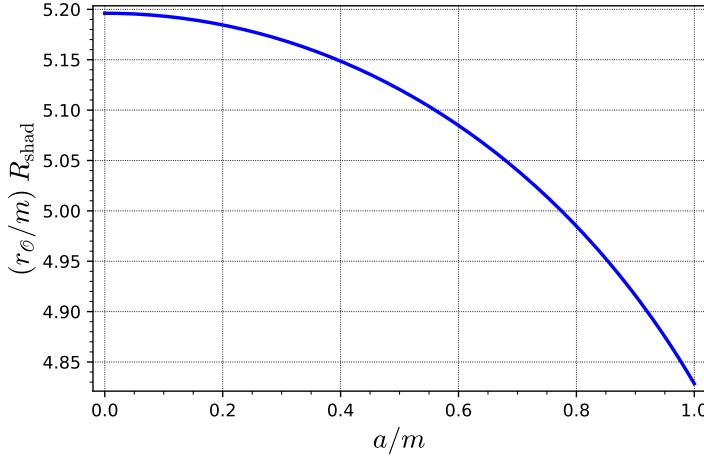


Figure 12.21: Angular radius R_{shad} of the (circular) critical curve of a Kerr black hole as seen by an asymptotic inertial observer located at $r = r_{\mathcal{O}}$ on the black hole's rotation axis, as a function of the Kerr spin parameter a [Eq. (12.90)]. R_{shad} is given in units of $m/r_{\mathcal{O}}$ (cf. Table 12.1). [Figure generated by the notebook D.5.14]

12.4.3 Shadow for an observer on the rotation axis

As stressed in Sec. 12.2.3, if $\theta_{\mathcal{O}} \in \{0, \pi\}$, the screen angular coordinates (α, β) can no longer be defined from the orthonormal frame $(e_{(\theta)}, e_{(\varphi)})$ associated with the Boyer-Lindquist coordinates (θ, φ) . In this case, we pick an arbitrary orthonormal frame $(e_{(\alpha)}, e_{(\beta)})$ in the screen plane to define (α, β) . Due to the axisymmetry of spacetime, when \mathcal{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 \mathcal{C} , is then a circle. A critical null geodesic impacting the screen on \mathcal{C} has necessarily $\ell = 0$ [Eq. (12.36)]. It follows that the parameter r_0 can take only a single value: $r_0 = r_{\text{ph}}^{\text{pol}}$, which is given by Eq. (12.67a). The radius of \mathcal{C} is then given by Eq. (12.37), with $q = q_c(r_{\text{ph}}^{\text{pol}})$:

$$\alpha^2 + \beta^2 = \frac{1}{r_{\mathcal{O}}^2} \left(q_c(r_{\text{ph}}^{\text{pol}}) + a^2 \right).$$

Given expression (12.57) of the function q_c and the fact that $r_{\text{ph}}^{\text{pol}}$ obeys $(r_{\text{ph}}^{\text{pol}})^4(r_{\text{ph}}^{\text{pol}} - 3m)^2 = a^4(r_{\text{ph}}^{\text{pol}} + m)^2$, as a solution of $\ell_c(r_0) = 0$, we get

$$R_{\text{shad}} := \sqrt{\alpha^2 + \beta^2} = 2 \frac{m}{r_{\mathcal{O}}} \sqrt{\frac{r_{\text{ph}}^{\text{pol}}}{m}} \sqrt{\frac{(r_{\text{ph}}^{\text{pol}})^2 - a^2}{r_{\text{ph}}^{\text{pol}} - m}} \quad (\theta_{\mathcal{O}} = 0 \text{ or } \theta_{\mathcal{O}} = \pi) \quad (12.90)$$

where $r_{\text{ph}}^{\text{pol}}$ is the function (12.67a) of (m, a) and the scale factor $m/r_{\mathcal{O}}$ 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 $r_{\text{ph}}^{\text{pol}} \simeq 2.493 m$, so that Eq. (12.90) results in $R_{\text{shad}} \simeq 4.875 m/r_{\mathcal{O}}$. We note that the critical curve \mathcal{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 $r_0 = r_{\text{ph}}^{\text{pol}}$ as all the critical null geodesics forming \mathcal{C} for $\theta_{\mathcal{O}} = 0$ or π .

The shadow radius R_{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 $\lim_{a \rightarrow 0} r_{\text{ph}}^{\text{pol}} = 3m$ and $\lim_{a \rightarrow m} r_{\text{ph}}^{\text{pol}} = (\sqrt{2} + 1)m$ (cf. Sec. 12.3.2) yield respectively

$$\lim_{a \rightarrow 0} R_{\text{shad}} = 3\sqrt{3}\frac{m}{r_{\mathcal{O}}} \simeq 5.196 \frac{m}{r_{\mathcal{O}}} \quad \text{and} \quad \lim_{a \rightarrow m} R_{\text{shad}} = 2(\sqrt{2} + 1)\frac{m}{r_{\mathcal{O}}} \simeq 4.828 \frac{m}{r_{\mathcal{O}}}. \quad (12.91)$$

The value for $a = m$ is thus only 7% lower than the value for $a = 0$.

Remark 2: The limit $a \rightarrow 0$ 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 $\ell_c(r_0)$ and $q_c(r_0)$ simplify significantly:

$$\ell_c(r_0) = -\frac{r_0^2}{m} + 2r_0 + m \quad (a = m) \quad (12.92)$$

$$q_c(r_0) = \frac{r_0^3}{m^2}(4m - r_0) \quad (a = m). \quad (12.93)$$

Substituting these values into the system (12.88) leads to the screen angular coordinates determining the critical curve \mathcal{C} of an extremal Kerr black hole:

$$\alpha = \frac{(r_0 - m)^2 - 2m^2}{mr_{\mathcal{O}} \sin \theta_{\mathcal{O}}} \quad (12.94a)$$

$$\beta = \frac{\epsilon_{\theta}}{mr_{\mathcal{O}} \sin \theta_{\mathcal{O}}} \sqrt{r_0^3(4m - r_0) - m^2 \cos^2 \theta_{\mathcal{O}} [2r_0(r_0 + 2m) + m^2 \cos^2 \theta_{\mathcal{O}}]}. \quad (12.94b)$$

Besides, Eq. (12.62) yields

$$r_{\text{ph}}^+ = m \quad \text{and} \quad r_{\text{ph}}^- = 4m \quad (a = m). \quad (12.95)$$

The range of r_0 is determined by $r_{\text{ph}}^+ \leq r_0 \leq r_{\text{ph}}^-$ and $\tilde{\Theta}(\theta_{\mathcal{O}}) \geq 0$. The last condition is equivalent to demanding that the quantity under the square root in expression (12.94b) for β is non-negative. This leads⁵ to some interval $[r_{\min}, r_{\max}] \subset [r_{\text{ph}}^+, r_{\text{ph}}^-]$. As in the case $a < m$ treated above, we have $r_{\max} \leq r_{\text{ph}}^-$ with $r_{\max} = r_{\text{ph}}^- \iff \theta_{\mathcal{O}} = \pi/2$. However, r_{\min} behaves

⁵The values of r_{\min} and r_{\max} can be computed exactly as $r_{\min} = \max(r_1, m)$ and $r_{\max} = r_2$, 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.5.14, is sufficient in practice.

differently. As we are going to see, $r_{\min} = r_{\text{ph}}^+ (= m)$ for a finite-width interval of values of $\theta_{\mathcal{O}}$ around $\pi/2$ and not only for $\pi/2$. To see this, let us start by noticing that expression (12.93) results in $q_c(r_{\text{ph}}^-) = 0$, while $q_c(r_{\text{ph}}^+) = 3m^2 \neq 0$. This last result seems to contradict the fact that r_{ph}^+ has been obtained in Sec. 12.3.1 by searching for the zeros of q_c . However, the generic expression (12.57) for q_c has a factor $(r_0 - m)^2$ in its denominator, so that taking the limits $a \rightarrow m$ and $r_0 \rightarrow r_{\text{ph}}^+ = m$ yields the indeterminate form “0/0”. As a consequence of $q_c(r_{\text{ph}}^+) \neq 0$, one has $\beta \neq 0$ for $r_0 = r_{\text{ph}}^+ = m$ and $|\theta_{\mathcal{O}} - \pi/2|$ sufficiently small. This implies that $r_{\min} = m$ for a finite range of values of $\theta_{\mathcal{O}}$ around $\pi/2$. More precisely, according to (12.94), the screen coordinates for $r_0 = r_{\text{ph}}^+ = m$ are

$$\alpha|_{r_0=m} = -2 \frac{m}{r_{\mathcal{O}} \sin \theta_{\mathcal{O}}} \quad (12.96a)$$

$$\beta|_{r_0=m} = \frac{\epsilon_{\theta} m}{r_{\mathcal{O}} \sin \theta_{\mathcal{O}}} \sqrt{3 - 6 \cos^2 \theta_{\mathcal{O}} - \cos^4 \theta_{\mathcal{O}}} \quad (12.96b)$$

In particular, for $\theta_{\mathcal{O}} = \pi/2$ (observer \mathcal{O} in the equatorial plane), we get

$$\alpha|_{r_0=m} = -2 \frac{m}{r_{\mathcal{O}}} \quad \text{and} \quad \beta|_{r_0=m} = \epsilon_{\theta} \sqrt{3} \frac{m}{r_{\mathcal{O}}} \quad \left(\theta_{\mathcal{O}} = \frac{\pi}{2} \right). \quad (12.97)$$

Equation (12.96b) shows that, for $\epsilon_{\theta} = +1$, $\beta|_{r_0=m} > 0 \iff \theta_{\text{crit}} < \theta_{\mathcal{O}} < \pi - \theta_{\text{crit}}$, where θ_{crit} is such that $\cos^2 \theta_{\text{crit}}$ is the positive root of the quadratic polynomial $-x^2 - 6x + 3 = 0$. We get $\cos^2 \theta_{\text{crit}} = 2\sqrt{3} - 3$, from which $\sin^2 \theta_{\text{crit}} = 4 - 2\sqrt{3} = (\sqrt{3} - 1)^2$, hence

$$\theta_{\text{crit}} = \arcsin(\sqrt{3} - 1) \simeq 0.8213 \text{ rad} \simeq 47.06^\circ. \quad (12.98)$$

For $\theta_{\mathcal{O}} = \theta_{\text{crit}}$ or $\pi - \theta_{\text{crit}}$, we have exactly $\beta|_{r_0=m} = 0$, while for $\theta_{\mathcal{O}} < \theta_{\text{crit}}$ or $\theta_{\mathcal{O}} > \pi - \theta_{\text{crit}}$, $\beta|_{r_0=m}$ is imaginary. This means that for $\theta_{\text{crit}} \leq \theta_{\mathcal{O}} \leq \pi - \theta_{\text{crit}}$, the range of the parameter r_0 is $[m, r_{\max}]$, while for $\theta_{\mathcal{O}} < \theta_{\text{crit}}$ or $\theta_{\mathcal{O}} > \pi - \theta_{\text{crit}}$, the range is $[r_{\min}, r_{\max}]$ with $r_{\min} > m$, as for the case $\theta_{\mathcal{O}} \neq \pi/2$ of the shadows with $a < m$ discussed in Sec. 12.4.2. In this last case, the critical curve is a smooth curve with $\beta|_{r_0=r_{\min}} = 0$ (in addition to $\beta|_{r_0=r_{\max}} = 0$, which always holds).

Let us focus on the first case, i.e. $\theta_{\text{crit}} \leq \theta_{\mathcal{O}} \leq \pi - \theta_{\text{crit}}$. 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 $\theta_{\mathcal{O}} = \pi/2$ (one point for $\epsilon_{\theta} = +1$, and the other one for $\epsilon_{\theta} = -1$). For $\theta_{\text{crit}} < \theta_{\mathcal{O}} < \pi - \theta_{\text{crit}}$, 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 $\theta_{\mathcal{O}} = \pi/2$). 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 value of q corresponds to a spherical photon orbit associated with a critical null geodesic that reaches the asymptotic inertial observer \mathcal{O} . Indeed, q must give birth to a

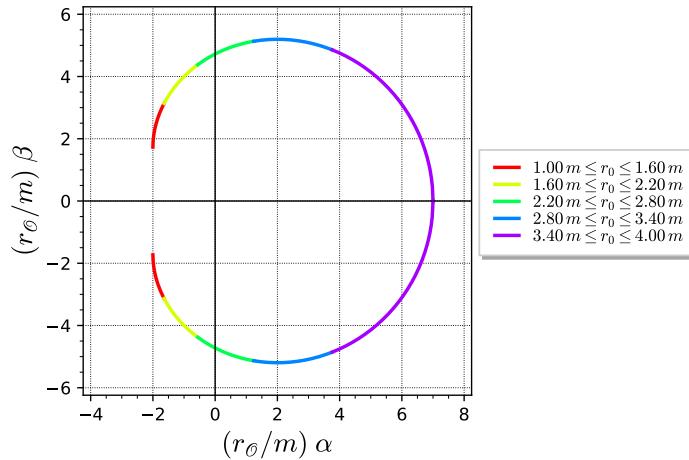


Figure 12.22: Part of the critical curve of an extremal Kerr black hole ($a = m$) in the screen of an asymptotic inertial observer \mathcal{O} in the equatorial plane ($\theta_{\mathcal{O}} = \pi/2$), as defined by the parametric equations (12.94), with $r_0 \in [m, 4m]$. The two endpoints are achieved 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 $\bar{\alpha} := \alpha r_{\mathcal{O}}/m$ and $\bar{\beta} := \beta r_{\mathcal{O}}/m$. [Figure generated by the notebook D.5.14]

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

$$\mathcal{R}(r) = (r - m)^2(r^2 + 2mr - q). \quad (12.99)$$

$r = m$ 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 :

$$r_q^{\pm} = \pm\sqrt{m^2 + q} - m. \quad (12.100)$$

Since $q \geq 0$, we have $r_q^+ \geq 0$ and $r_q^- \leq -2m$. It is then clear (cf. Fig. 12.23) that

$$(\forall r \in (m, +\infty), \quad \mathcal{R}(r) > 0) \iff r_q^+ \leq m. \quad (12.101)$$

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:

Property 12.21

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

$$\ell = 2m \quad \text{and} \quad 0 \leq q \leq 3m^2. \quad (12.102)$$

Furthermore, such a critical null geodesic can reach the asymptotic inertial observer \mathcal{O} , who is located at $\theta = \theta_{\mathcal{O}}$, only if the $\tilde{\Theta}$ function associated with (ℓ, q) obeys $\tilde{\Theta}(\theta_{\mathcal{O}}) \geq 0$. Setting

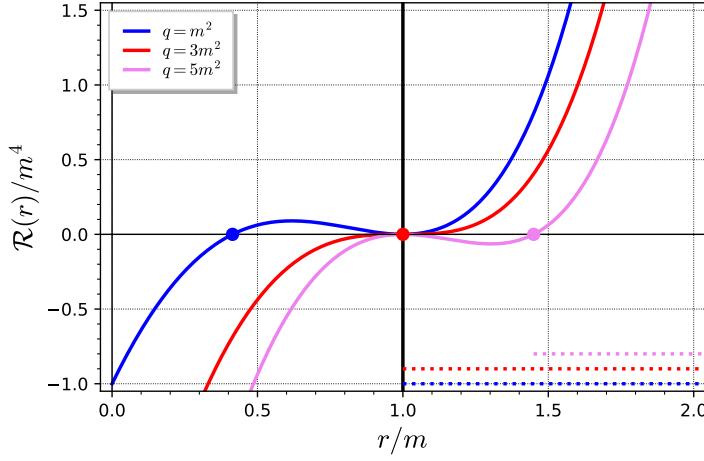


Figure 12.23: Radial polynomial $\mathcal{R}(r)$ for $a = m$, $\ell = 2m$ and three values of q . All the polynomials have a double root at $r = m$ and the dots mark the root r_q^+ [Eq. (12.100)], the fourth root $r_q^- \leq -2m$ being out of the figure's scope. The dotted horizontal lines indicate the range of r where a null geodesic motion to the asymptotic inertial observer is possible. [Figure generated by the notebook D.5.14]

$a = m$ and $\ell = 2m$ into expression (12.26) for $\tilde{\Theta}$, we get the criterion

$$q \geq \frac{3 + \cos^2 \theta_\varnothing}{\tan^2 \theta_\varnothing} m^2. \quad (12.103)$$

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):

$$\alpha = -\frac{2m}{r_\varnothing \sin \theta_\varnothing} \quad (12.104a)$$

$$\beta = \frac{\epsilon_\varnothing m}{r_\varnothing \sin \theta_\varnothing} \sqrt{\frac{q}{m^2} \sin^2 \theta_\varnothing - \cos^2 \theta_\varnothing (3 + \cos^2 \theta_\varnothing)}, \quad \frac{3 + \cos^2 \theta_\varnothing}{\tan^2 \theta_\varnothing} \leq \frac{q}{m^2} \leq 3. \quad (12.104b)$$

The system (12.104) defines a curve parameterized by q , with the range of q obtained by combining (12.102) and (12.103). For \varnothing in the equatorial plane ($\theta_\varnothing = \pi/2$), this parametric equation simplifies to

$$\alpha = -2 \frac{m}{r_\varnothing} \quad \text{and} \quad \beta = \epsilon_\varnothing \frac{m}{r_\varnothing} \sqrt{\frac{q}{m^2}}, \quad 0 \leq \frac{q}{m^2} \leq 3 \quad \left(\theta_\varnothing = \frac{\pi}{2} \right). \quad (12.105)$$

In all cases, the curve (12.104) is actually a segment of a vertical straight line on \varnothing 's screen, since it has $\alpha = \text{const}$. Following Ref. [237], we shall call this segment the **NHEK line**, NHEK standing for *Near-Horizon Extremal Kerr*, given that spherical photon orbits with r_0 close to m are located in the near-horizon region, which we shall study in details in Sec. 13.4. The key feature 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 $\theta_\varnothing = \pi/2$ and $\theta_\varnothing = \pi/3 > \theta_{\text{crit}}$.

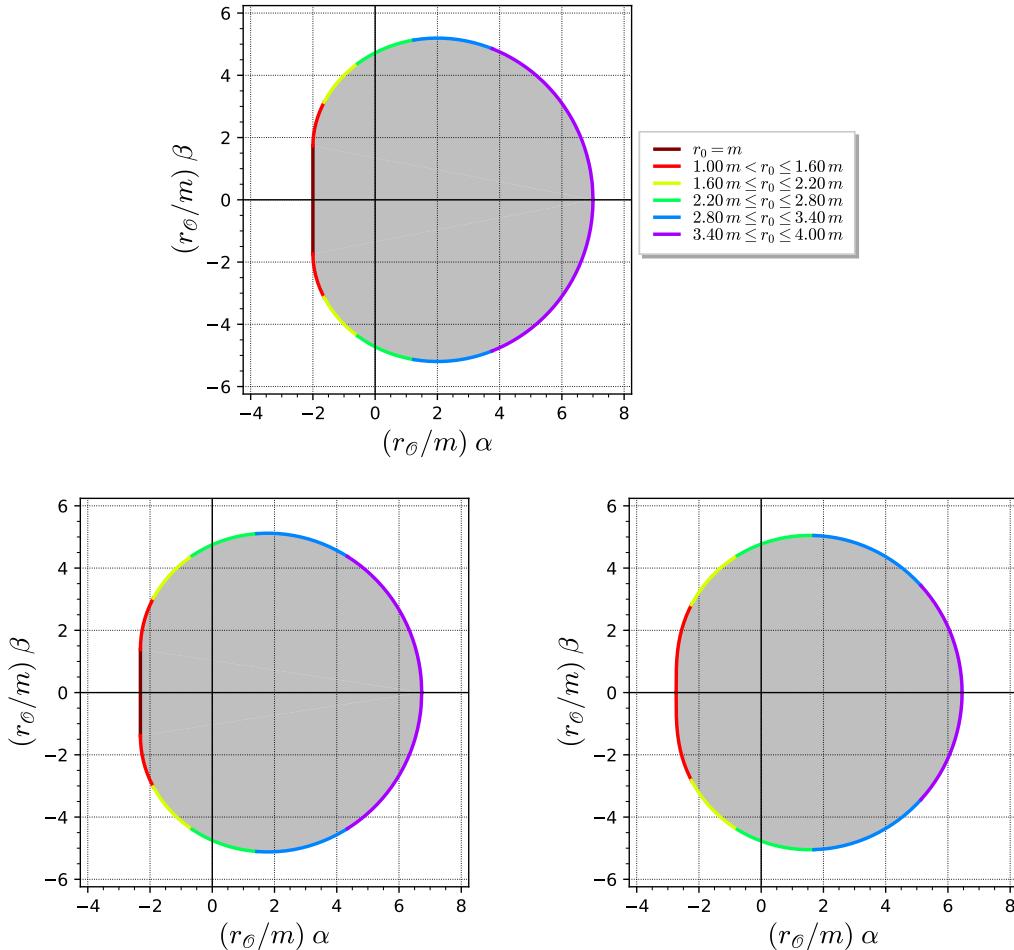


Figure 12.24: Critical curve \mathcal{C} (colored curve) and shadow (grey area) of an extremal Kerr black hole on the screen of an asymptotic inertial observer \mathcal{O} located at $r = r_{\mathcal{O}}$ and $\theta = \theta_{\mathcal{O}}$, for three values of $\theta_{\mathcal{O}}$: $\theta_{\mathcal{O}} = \pi/2$ (\mathcal{O} in the equatorial plane; upper panel), $\theta_{\mathcal{O}} = \pi/3$ (lower left) and $\theta_{\mathcal{O}} = \theta_{\text{crit}} = \arcsin(\sqrt{3} - 1) \simeq 0.821$ [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 \mathcal{C} for $\theta < \theta_{\text{crit}}$ or $\theta > \pi - \theta_{\text{crit}}$. As in Fig. 12.20, the color code corresponds to some selected ranges for the parameter r_0 of the critical null geodesics forming \mathcal{C} . [Figure generated by the notebook D.5.14]

Cardioid shape

As seen from Fig. 12.24, the departure of the critical curve \mathcal{C} from a perfect circle is maximal when the observer lies in the equatorial plane, i.e. when $\theta_\phi = \pi/2$. 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 [192]. Indeed, for $\theta_\phi = \pi/2$, equations (12.94), which govern the part of \mathcal{C} parameterized by r_0 , simplify to

$$\alpha = \frac{(r_0 - m)^2 - 2m^2}{mr_\phi} \quad (12.106a)$$

$$\beta = \frac{\epsilon_\theta}{mr_\phi} r_0 \sqrt{r_0(4m - r_0)}. \quad (12.106b)$$

Let us then extract r_0 from Eq. (12.106a):

$$r_0 = m \left(1 + \sqrt{2 + \frac{r_\phi}{m} \alpha} \right)$$

and substitute it into the square of Eq. (12.106b); we get (using $\epsilon_\theta^2 = 1$)

$$\beta^2 = \frac{m^2}{r_\phi^2} \left(1 + \sqrt{2 + \frac{r_\phi}{m} \alpha} \right)^3 \left(3 - \sqrt{2 + \frac{r_\phi}{m} \alpha} \right). \quad (12.107)$$

At this stage, it is worth to introduce explicitly the rescaled angular coordinates used in Figs. 12.22 and 12.24:

$$\bar{\alpha} := \frac{r_\phi}{m} \alpha \quad \text{and} \quad \bar{\beta} := \frac{r_\phi}{m} \beta. \quad (12.108)$$

Then, by expanding its right-hand side, we can rewrite Eq. (12.107) as

$$\bar{\beta}^2 = -\bar{\alpha}^2 + 2\bar{\alpha} + 8\sqrt{\bar{\alpha} + 2} + 11. \quad (12.109)$$

Let us introduce new screen coordinates $(\hat{\alpha}, \hat{\beta})$ by shifting the origin to $(\bar{\alpha}, \bar{\beta}) = (-1, 0)$:

$$\hat{\alpha} := \bar{\alpha} + 1 \quad \text{and} \quad \hat{\beta} := \bar{\beta}. \quad (12.110)$$

We can then recast Eq. (12.109) as

$$\hat{\alpha}^2 + \hat{\beta}^2 - 4\hat{\alpha} = 8 \left(\sqrt{\hat{\alpha} + 1} + 1 \right). \quad (12.111)$$

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

$$(\hat{\alpha}^2 + \hat{\beta}^2 - 4\hat{\alpha})^2 = 16 (\hat{\alpha}^2 + \hat{\beta}^2). \quad (12.112)$$

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

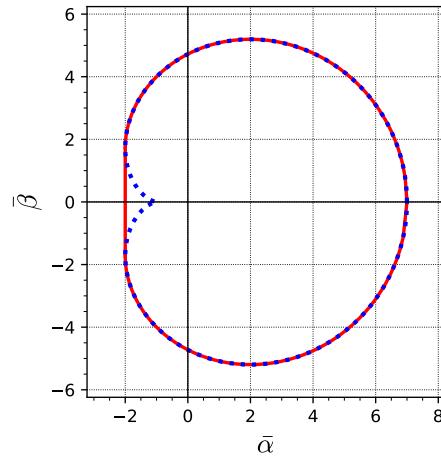


Figure 12.25: Cardioid defined by the polar equation $\rho = 4(1 + \cos \phi)$ (dotted blue curve) and critical curve of an extremal Kerr black hole seen from the equatorial plane (red curve). The screen coordinates $(\bar{\alpha}, \bar{\beta})$ are defined by Eq. (12.108), while (ρ, ϕ) are polar coordinates 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.5.14]

$(\rho - 4 \cos \phi)^2 = 16$, 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

$$\rho = 4(1 + \cos \phi). \quad (12.113)$$

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 $\theta_\phi \neq \pi/2$, the critical curve \mathcal{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, \mathcal{C} is a classical elementary curve, namely a *Cartesian oval*, also known as *oval of Descartes*, as shown in Ref. [236]. More precisely, \mathcal{C} is exactly such an oval for $\theta_\phi \leq \theta_{\text{crit}}$ or $\theta_\phi \geq \pi - \theta_{\text{crit}}$. For $\theta_{\text{crit}} < \theta_\phi < \pi - \theta_{\text{crit}}$, i.e. when the NHEK line is present, \mathcal{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 θ_ϕ is known, as we shall see for M87* in Sec. 12.5.3. Figure 12.26 compares then the critical curves at a fixed

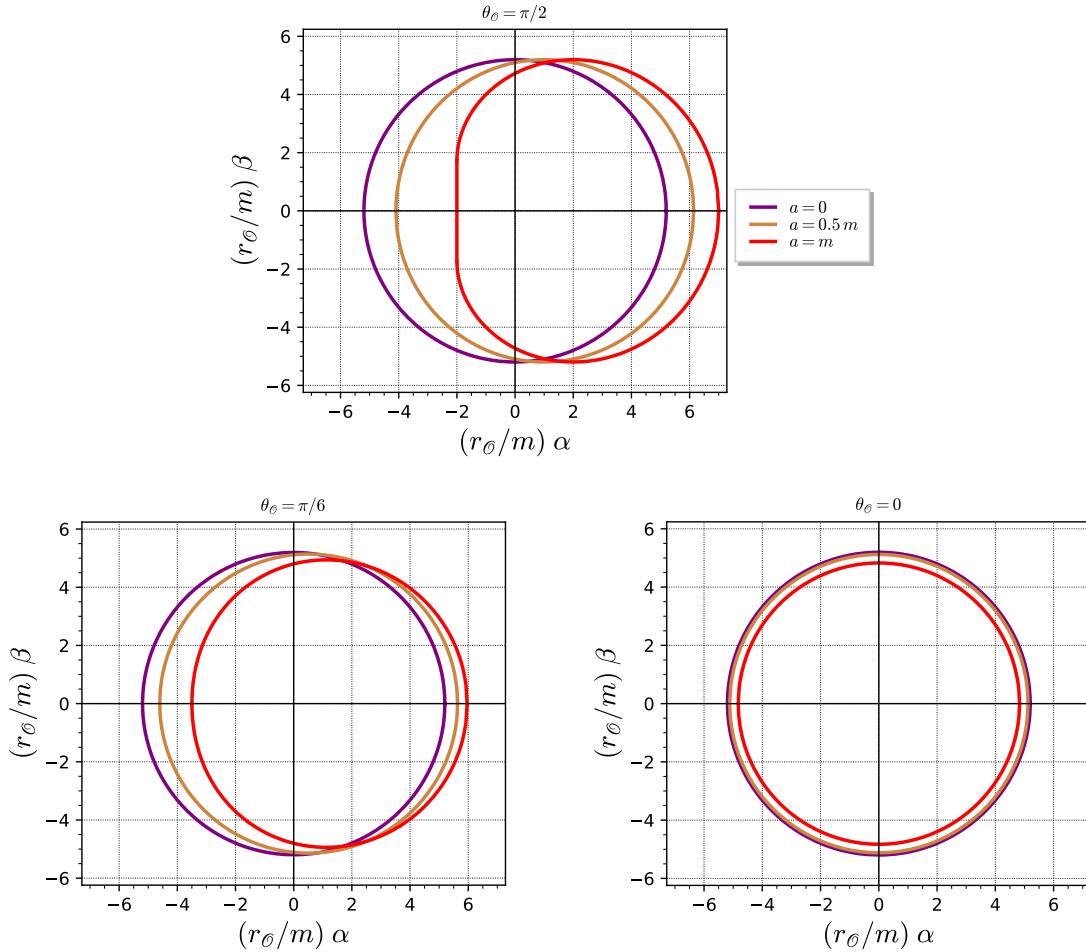


Figure 12.26: Critical curves for different values of a at fixed observer's inclination angle: $\theta_\mathcal{O} = \pi/2$ (top panel), $\theta_\mathcal{O} = \pi/6$ (lower left panel) and $\theta_\mathcal{O} = 0$ (lower right panel). [Figure generated by the notebook D.5.14]

value of $\theta_\mathcal{O}$ for various values of the black hole spin parameter a . For $\theta_\mathcal{O} = 0$, we recover the feature found in Sec. 12.4.3, namely that the critical curve depends very weakly on the spin.

Historical note : The critical curve of a Kerr black hole has been computed first by James M. Bardeen in 1972 [37] (cf. the historical note on p. 273), in the form of the parametric system (12.88). He wrote: *The rim of the “black hole”⁶ 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 [6, 520], this has become astrophysically important! For $a = m$ and $\theta_\mathcal{O} = \pi/2$, Bardeen derived the system (12.106) governing the part of the critical curve parameterized by r_0 . For the NEHK line, he simply noted⁷: *The non-uniform nature of the limit $a \rightarrow m$ 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. [37]) similar to that shown in the top panel of Fig. 12.24.

⁶For Bardeen, “black hole” (with quotes) stands for the black spot on the observer’s screen.

⁷ 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 [191], cf. the historical note on p. 499.

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 \mathcal{M}_I), gives birth to two infinite sequences of images on the screen of the asymptotic inertial observer, both sequences converging to the critical curve \mathcal{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* [4, 5], or in the form of an accretion disk [2]. 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 [6, 520]. 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. [493]. 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): $r_{\text{in}} = r_{\text{ISCO}}^+$, with $r_{\text{ISCO}}^+ \simeq 4.233 m$ for $a = 0.5 m$ (Fig. 12.27) and $r_{\text{ISCO}}^+ \simeq 1.937 m$ 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. [493] for details). The images have been generated by the open-source ray-tracing code Gyoto [492] (cf. Appendix E).

Figure 12.27 provides images for moderate black hole spin: $a = 0.5 m$. For a given inclination θ_ϕ , 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 ($\theta_\phi = 0$), one could use a figure similar to Fig. 8.26, with the orange null geodesics generating the primary image, the brown ones the secondary image and the red ones

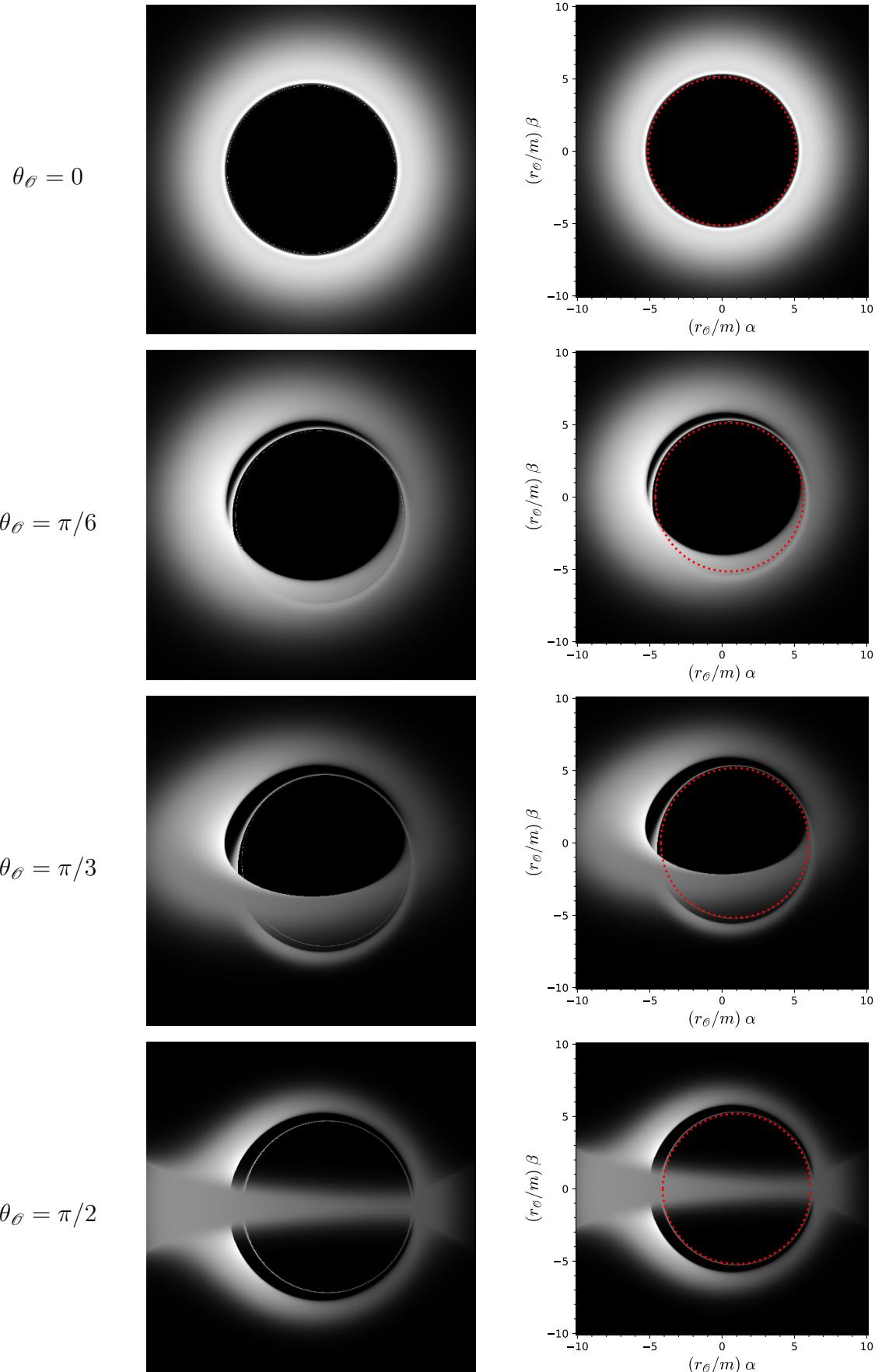


Figure 12.27: Images of a thick accretion disk around a Kerr black hole with $a = 0.5 m$ for three observer's inclination angles θ_θ . The right column shows the critical curve (red dotted line) superposed on the image. [Images produced by Gyoto with the input files given in Sec. E.2.2; the addition of the critical curve in the right column has been performed by means of the SageMath notebook D.5.15]

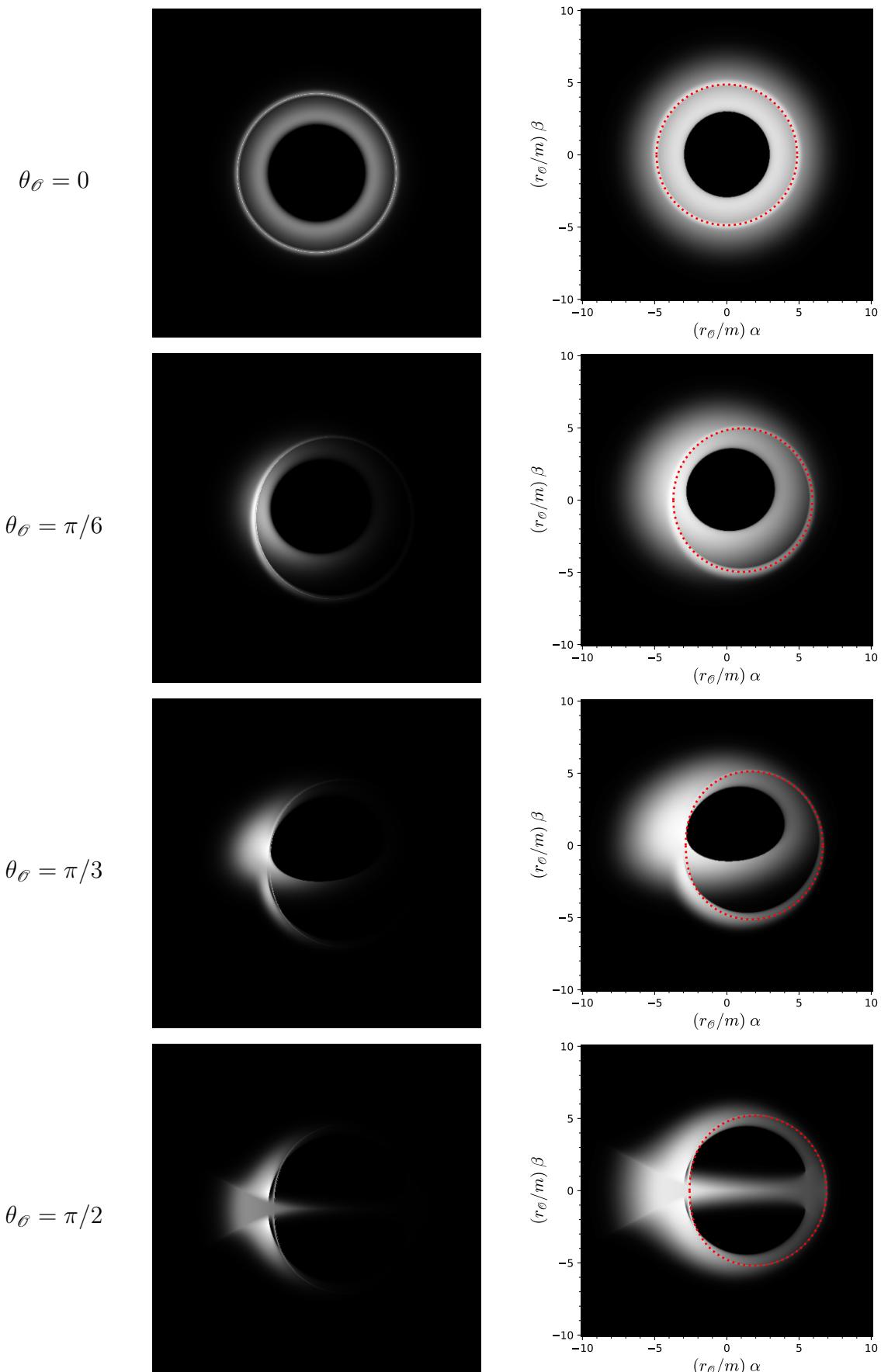


Figure 12.28: Same as Fig. 12.27 but for $a = 0.95\text{ m}$ and the images in the right column using a log scale to reveal the faint parts. The critical curves are the same as in Fig. 12.20.

the tertiary image. The green circle would then mark the location of polar spherical photon orbits, at $r_{\text{ph}}^{\text{pol}} \simeq 2.883 \text{ m}$ for $a = 0.5 \text{ m}$, since only the $\ell = 0$ critical null geodesics matter for $\theta_\phi = 0$. 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 $r = r_{\text{ph}}^{\text{pol}}$ 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 \mathcal{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 \mathcal{C} , in agreement with the exponential convergence of high order images to \mathcal{C} . This exponential convergence has been established for $a = 0$ by Eq. (8.131), but it holds for $a \neq 0$ as well [234].

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 $\theta_\phi \neq 0$. 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 $\theta_\phi = 0$ and $\theta_\phi = \pi/6$ 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 θ_ϕ 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 $\theta_\phi = 0$, 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 $r_{\text{ISCO}}^+ \simeq 4.233 \text{ m}$ for $a = 0.5 \text{ m}$ and $r_{\text{ISCO}}^+ = 6 \text{ m}$ 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 $a = 0.95 \text{ m}$. 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 $a = 0.95 \text{ m}$, we have $r_{\text{ISCO}}^+ \simeq 1.937 \text{ m}$, while the radius of the retrograde circular photon orbit is $r_{\text{ph}}^- \simeq 3.995 \text{ m}$ and that of polar spherical orbits is $r_{\text{ph}}^{\text{pol}} \simeq 2.493 \text{ m}$. Given that $r_{\text{ph}}^+ \simeq 1.386 \text{ m}$, we have thus $r_{\text{ph}}^+ < r_{\text{ISCO}}^+ < r_{\text{ph}}^{\text{pol}} < r_{\text{ph}}^-$. More generally, using formulas (11.167), (12.62) and (12.67a), we have

$$r_{\text{ISCO}}^+ < r_{\text{ph}}^- \iff a > 0.638 \text{ m} \quad \text{and} \quad r_{\text{ISCO}}^+ < r_{\text{ph}}^{\text{pol}} \iff a > 0.853 \text{ m}. \quad (12.114)$$

These properties are illustrated in Fig. 12.29. Hence, for large spins, if the inner boundary of

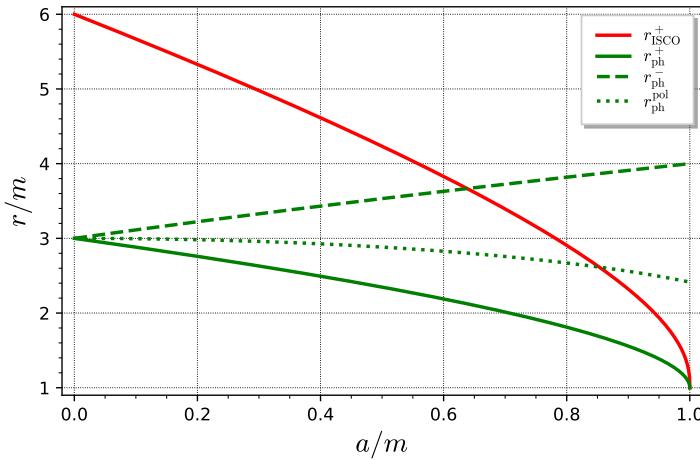


Figure 12.29: Radius of the timelike prograde ISCO, r_{ISCO}^+ [Eq. (11.167)], compared with the radii of the prograde and retrograde outer circular photon orbits, r_{ph}^+ and r_{ph}^- [Eq. (12.62)] and the radius of polar spherical photon orbits, $r_{\text{ph}}^{\text{pol}}$ [Eq. (12.67a)]. [Figure generated by the notebook D.5.15]

the accretion disk is set by the prograde ISCO, a large part of the primary image lies strictly inside the critical curve \mathcal{C} . Let us for instance consider the image for $\theta_\phi = 0$ in Fig. 12.28, which is the easiest to understand since for $\theta_\phi = 0$, \mathcal{C} involves only the $\ell = 0$ critical null geodesics (cf. Sec. 12.4.3), i.e. those that are rolling indefinitely around the sphere $r_0 = r_{\text{ph}}^{\text{pol}}$ in their asymptotic past. Since $r_{\text{ISCO}}^+ < r_{\text{ph}}^{\text{pol}}$ in the present case, this implies that the part of the disk located at $r \in [r_{\text{ISCO}}^+, r_{\text{ph}}^{\text{pol}})$ is an emitting region below the polar spherical photon orbits at $r_0 = r_{\text{ph}}^{\text{pol}}$. The emitted photons encounter then the screen strictly inside \mathcal{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 $A(r_{\text{em}}) < 0$ for $r_{\text{em}} < 3m = r_{\text{ph}}^{\text{pol}}(a = 0)$ (cf. Fig. 8.22) yields impact parameters b_n^\pm that are lower than b_c , the latter being nothing but the radius of \mathcal{C} for $a = 0$. Hence we conclude:

Property 12.22: black spot in the image and black hole shadow

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 \mathcal{C} . The black spot is actually smaller than the shadow, lying strictly inside \mathcal{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. [233] or Fig. 2 in Ref. [493]). 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. [191, 376].

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 θ_ϕ , 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:

Property 12.23: photon ring

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 \mathcal{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 [234]. We note on Figs. 12.27 and 12.28 that the $n = 2$ image (tertiary image) is already very close to \mathcal{C} . We conclude:

Property 12.24: photon ring and critical curve

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 [303]. 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 $r = r_{\text{ph}}^{\pm}$ in Kerr

spacetime (see e.g. Ref. [83] 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. [46, 301, 303, 234, 236]. Recently the term *secondary ring*, has been introduced [493] 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* [233].

Historical note : In 1973, Christopher T. Cunningham and James M. Bardeen [145] computed the multiple images of a star orbiting an extremal Kerr black hole ($a = m$). In 1979, Jean-Pierre Luminet [347] predicted that a Schwarzschild black hole illuminated by a parallel light beam would appear to a remote observer as a black disk of radius $R_{\text{shad}} = 3\sqrt{3}m/r_\phi$ (cf. Eq. (12.91)) surrounded by a sequence of rings, called *ghost rings* by Luminet, converging exponentially 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 [347], Luminet presented the first computed image of an accretion disk around a Schwarzschild black hole (cf. historical note on p. 278). 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 [490], for $a = 0.998 m$. In 1997, Michał Jaroszyński and Andrzej Kuprowski [298] 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 [191] (see also Ref. [190]) 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 Kuprowski'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 [191], 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 [6].

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 [6, 520]. 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 (points in the image Fourier plane) obtained by very-long-baseline interferometry (VLBI) in April 2017. A new image has been published in 2024 from data obtained in April 2018 [9].

In order to interpret the EHT image, it is natural to superpose the (theoretical) critical curve \mathcal{C} of a Kerr black hole onto it. This is all the more meaningful that \mathcal{C} is 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_ϕ , (ii) the angle θ_ϕ 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*

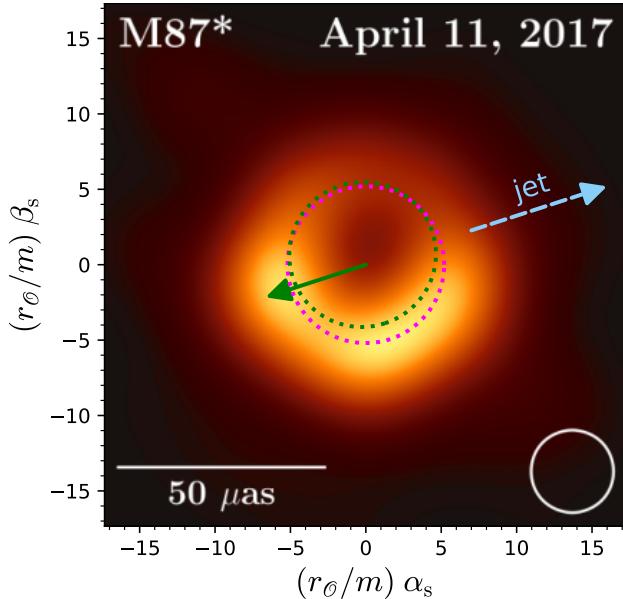


Figure 12.30: Image of the immediate vicinity of the supermassive black hole M87*, as released by the EHT collaboration in 2019 [6], with two critical curves superposed: that of a Schwarzschild black hole (magenta dotted circle) and that of an extremal Kerr black hole seen under the inclination $\theta_\theta = 163^\circ$ (green dotted curve), with the projection of the spin axis onto the screen indicated by the green arrow (position angle $\Theta = 108^\circ$ with respect to the β_s -axis). The critical curves are scaled by assuming $m/r_\theta = 3.7 \mu\text{as}$ (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. $20 \mu\text{as}$). [Figure generated by the notebook D.5.16; source of the EHT image: Fig. 3 of Ref. [6]]

these three quantities can be estimated. Indeed, the distance from the Earth to M87* has been measured to $r_\theta \simeq 16.7 \text{ Mpc}$ and the mass of M87* has been deduced from stellar dynamics in the inner part of the galactic nucleus, resulting in⁸ $m \simeq 6.2 \cdot 10^9 M_\odot$ (cf. Table 12.1 and references therein), hence the scale factor $m/r_\theta \simeq 3.7 \mu\text{as}$. Regarding θ_θ and Θ , both angles can be determined via reasonable assumptions, thanks to the large relativistic jet emanating from the close vicinity of the black hole⁹, 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 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 $\Theta_{\text{jet}} \sim -72^\circ$ (see e.g. Fig. 1 in Ref. [508]); 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 $\iota \sim 17^\circ$ [364, 508]. In particular, the detection of apparent superluminal motions, up to $v_{\text{app}} \sim 6c$, implies both relativistic velocities and ι being a small

⁸Another method, based on the dynamics of the gas in the nucleus disk at $r < 40 \text{ pc}$, yields a mass of only $m = 3.5 \cdot 10^9 M_\odot$ [513]; the discrepancy with the value arising from stellar dynamics has not been explained yet.

⁹This jet has been discovered in 1918 [147], 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. 11.3.2, see e.g. Ref. [2].

angle (see e.g. Sec. 5.7.4 of Ref. [227]). Assuming that the jet axis is aligned with the black hole's rotation axis, we have either $\theta_\phi = \iota$ or $\theta_\phi = \pi - \iota$.

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) [7]. 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. 11.3.4), this implies that the spin of black hole is pointing away from us, i.e. that $\theta_\phi = \pi - \iota \sim 163^\circ$ and $\Theta = \pi + \Theta_{\text{jet}} \sim 108^\circ$. 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 $m/r_\phi = 3.7 \mu\text{as}$ 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 circle) 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, θ_ϕ being close to π . To have a stronger discrepancy among various possible values of a , it would have been better to have θ_ϕ 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 $6 \cdot 10^9 M_\odot$. Unfortunately, the resolution of the EHT image ($\sim 20 \mu\text{as} \sim 1/6$ 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. [232, 223] 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 [304, 349], 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. [166, 167, 168, 234]. In particular, Ref. [234] 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 [433].

Chapter 13

Extremal Kerr black hole

Contents

13.1 Introduction	503
13.2 Definition and basic properties	504
13.3 Maximal analytic extension	511
13.4 Near-horizon extremal Kerr (NHEK) geometry	521
13.5 Going further	535

13.1 Introduction

The Kerr solution of the 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 $0 < a < m$ and the case $a = 0$ (Schwarzschild solution) has been treated in Chaps. 6–9. Here we focus on the case $a = m$, which is called the *extremal Kerr spacetime*. This corresponds to 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 (1.44), but it describes a *naked singularity* (cf. Sec. 9.6.1): the ring singularity is not hidden by any horizon to asymptotic observers.

The Kerr spacetime with $a = m$ is not just the maximally spinning black hole of the Kerr family. As we going to see, it has important properties that are not shared by Kerr spacetimes 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 . This corresponds to an *extremal black hole*, according to the terminology introduced in Sec. 5.4.3. Besides, there is no internal horizon and the maximal analytic extension, to be constructed in Sec. 13.3 below, is simpler than that of the Kerr spacetime with $a < m$. Another specific property of the extremal Kerr spacetime (actually related to the degeneracy of the horizon) regards the geometry near the horizon: it admits an enlarged symmetry group, which is generated by four independent Killing vectors, instead of two for the geometry of the global solution. This is called the *NHEK* geometry (for

near-horizon extremal Kerr) and has received considerable attention in the recent literature. We shall discuss it in Sec. 13.4.

13.2 Definition and basic properties

13.2.1 The extremal Kerr solution

Let us consider the manifold $\mathbb{R}^2 \times \mathbb{S}^2$ described 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 (\mathcal{M}, g) where the manifold \mathcal{M} is the following open subset of $\mathbb{R}^2 \times \mathbb{S}^2$:

$$\mathcal{M} := \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathcal{R} \quad (13.1)$$

with

$$\mathcal{R} := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = 0 \text{ and } \theta(p) = \frac{\pi}{2} \right\}, \quad (13.2)$$

and the metric g has the following expression in terms of the coordinates $(x^{\tilde{\alpha}}) = (\tilde{t}, r, \theta, \tilde{\varphi})$:

$$\begin{aligned} g = & - \left(1 - \frac{2mr}{\rho^2} \right) d\tilde{t}^2 + \frac{4mr}{\rho^2} d\tilde{t} dr - \frac{4m^2 r \sin^2 \theta}{\rho^2} d\tilde{t} d\tilde{\varphi} \\ & + \left(1 + \frac{2mr}{\rho^2} \right) dr^2 - 2m \left(1 + \frac{2mr}{\rho^2} \right) \sin^2 \theta dr d\tilde{\varphi} \\ & + \rho^2 d\theta^2 + \left(r^2 + m^2 + \frac{2m^3 r \sin^2 \theta}{\rho^2} \right) \sin^2 \theta d\tilde{\varphi}^2, \end{aligned} \quad (13.3)$$

with

$$\rho^2 := r^2 + m^2 \cos^2 \theta. \quad (13.4)$$

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 $a \rightarrow m$ of expression (10.36) for the Kerr metric with $a < m$.

The metric (13.3) is regular in all \mathcal{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 \mathcal{R} that has precisely been excluded from \mathcal{M} in 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 $\rho \rightarrow 0$. Therefore, as for the Kerr spacetime with $a < m$ (cf. Sec. 10.2.6), we shall call \mathcal{R} the **ring singularity** of the extremal Kerr spacetime. Note that, formally, it is not part of the spacetime manifold \mathcal{M} [cf. Eq. (13.1)].

Moreover, the Ricci tensor of the metric (13.3) is identically zero in all \mathcal{M} (see the notebook D.5.17 for the computation). Hence, we have:

Property 13.1: extremal Kerr metric as a solution of the vacuum Einstein equation

The metric g of the extremal Kerr spacetime is a solution of the vacuum Einstein equation $R = 0$ [Eq. (1.44)].

The inverse metric is

$$g^{\tilde{\alpha}\tilde{\beta}} = \begin{pmatrix} -1 - \frac{2mr}{\rho^2} & \frac{2mr}{\rho^2} & 0 & 0 \\ \frac{2mr}{\rho^2} & \frac{(r-m)^2}{\rho^2} & 0 & \frac{m}{\rho^2} \\ 0 & 0 & \frac{1}{\rho^2} & 0 \\ 0 & \frac{m}{\rho^2} & 0 & \frac{1}{\rho^2 \sin^2 \theta} \end{pmatrix}. \quad (13.5)$$

13.2.2 Boyer-Lindquist coordinates

For $a < m$, the Kerr manifold \mathcal{M} has been split in three open regions, \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{III} , separated by the two Killing horizons \mathcal{H} and \mathcal{H}_{in} [cf. Eqs. (10.29) and (10.2a)]. Since \mathcal{H} was defined by $r = r_+ := m + \sqrt{m^2 - a^2}$ [Eq. (10.27)] and \mathcal{H}_{in} by $r = r_- := m - \sqrt{m^2 - a^2}$ [Eq. 10.28], we notice that $r_+ = r_- = m$ in the limit $a \rightarrow m$. This implies that \mathcal{H} and \mathcal{H}_{in} coincide when $a \rightarrow m$ and the region \mathcal{M}_{II} , which is bounded by \mathcal{H} and \mathcal{H}_{in} , disappears. Accordingly, we shall split the extremal Kerr manifold \mathcal{M} in two open regions only, \mathcal{M}_I and \mathcal{M}_{III} , separated by a single hypersurface \mathcal{H} :

$$\boxed{\mathcal{M} = \mathcal{M}_I \cup \mathcal{H} \cup \mathcal{M}_{III}}, \quad (13.6)$$

with

$$\mathcal{M}_I := \{p \in \mathcal{M}, \quad r(p) > m\} \quad (13.7a)$$

$$\mathcal{H} := \{p \in \mathcal{M}, \quad r(p) = m\} \quad (13.7b)$$

$$\mathcal{M}_{III} := \{p \in \mathcal{M}, \quad r(p) < m\}. \quad (13.7c)$$

Remark 1: We are using the notation \mathcal{M}_{III} for the “second” region, and not \mathcal{M}_{II} , to be consistent with Chaps. 10–12, i.e. with the limit $a \rightarrow m$ 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 $a \rightarrow m$. Its double root, $r = m$, defines the hypersurface \mathcal{H} .

In the region $\mathcal{M}_{BL} := \mathcal{M} \setminus \mathcal{H} = \mathcal{M}_I \cup \mathcal{M}_{III}$, 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

$$t = \tilde{t} + \frac{2m^2}{r - m} - 2m \ln \left| \frac{r - m}{m} \right| \quad (13.8a)$$

$$\varphi = \tilde{\varphi} + \frac{m}{r - m}. \quad (13.8b)$$

Differentiating these relations leads to

$$d\tilde{t} = dt + \frac{2mr}{(r - m)^2} dr \quad \text{and} \quad d\tilde{\varphi} = d\varphi + \frac{m}{(r - m)^2} dr. \quad (13.9)$$

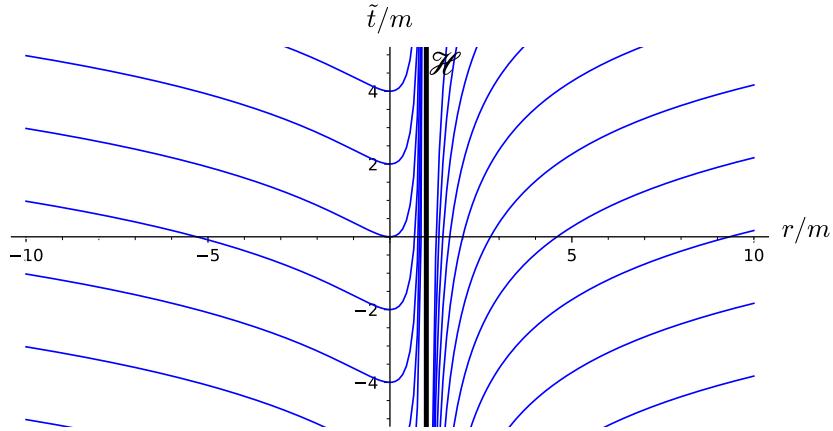


Figure 13.1: Trace of the hypersurfaces of constant Boyer-Lindquist time t in the plane (\tilde{t}, r) . [Figure generated by the notebook D.5.17]

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 $a < m$ counterpart (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.

It follows from the transformations (13.8) that the Boyer-Lindquist coordinate frame (∂_α) and the Kerr coordinate frame $(\partial_{\tilde{\alpha}})$ are related by¹

$$\partial_t = \partial_{\tilde{t}} \quad (13.10a)$$

$$\partial_r = \partial_{\tilde{r}} + \frac{2mr}{(r-m)^2} \partial_{\tilde{t}} + \frac{m}{(r-m)^2} \partial_{\tilde{\varphi}} \quad (13.10b)$$

$$\partial_\theta = \partial_\theta \quad (13.10c)$$

$$\partial_\varphi = \partial_{\tilde{\varphi}}. \quad (13.10d)$$

Remark 3: As in Chap. 10, we have denoted by $\partial_{\tilde{r}}$ 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 ∂_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

$$\begin{aligned} g = & - \left(1 - \frac{2mr}{\rho^2} \right) dt^2 - \frac{4m^2 r \sin^2 \theta}{\rho^2} dt d\varphi + \frac{\rho^2}{(r-m)^2} dr^2 \\ & + \rho^2 d\theta^2 + \left(r^2 + m^2 + \frac{2m^3 r \sin^2 \theta}{\rho^2} \right) \sin^2 \theta d\varphi^2. \end{aligned} \quad (13.11)$$

This expression can be obtained either by taking the limit $a \rightarrow m$ of Eq. (10.8) or by using (13.9) to substitute $d\tilde{t}$ and $d\tilde{\varphi}$ in Eq. (13.3). We note that $g_{rr} \rightarrow +\infty$ when $r \rightarrow m$, which reflects the

¹See also the limit $a \rightarrow m$ of Eq. (10.38).

singularity of Boyer-Lindquist coordinates on \mathcal{H} and explains why the latter was excluded in the definition of \mathcal{M}_{BL} . 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 \mathcal{H} , without crossing it, so that the points on \mathcal{H} do not belong to any hypersurface $t = \text{const}$. That the $t = \text{const}$ hypersurfaces do not provide a regular slicing of the extremal Kerr spacetime (\mathcal{M}, g) is also manifest on the Carter-Penrose diagram shown in Fig. 13.4.

The Boyer-Lindquist components of the inverse metric g^{-1} are

$$g^{\alpha\beta} = \begin{pmatrix} -\frac{1}{(r-m)^2} \left(r^2 + m^2 + \frac{2m^3 r \sin^2 \theta}{\rho^2} \right) & 0 & 0 & -\frac{2m^2 r}{\rho^2(r-m)^2} \\ 0 & \frac{(r-m)^2}{\rho^2} & 0 & 0 \\ 0 & 0 & \frac{1}{\rho^2} & 0 \\ -\frac{2m^2 r}{\rho^2(r-m)^2} & 0 & 0 & \frac{1}{(r-m)^2 \sin^2 \theta} \left(1 - \frac{2mr}{\rho^2} \right) \end{pmatrix}. \quad (13.12)$$

They can also be obtained by taking the limit $a \rightarrow m$ of Eq. (10.16).

13.2.3 Symmetries

The extremal Kerr metric (13.3) is stationary² and axisymmetric. The corresponding isometry group is $\mathbb{R} \times \text{SO}(2) \simeq \mathbb{R} \times \text{U}(1)$ and is generated by two commuting Killing vectors ξ and η . 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 from the metric components (13.3) and (13.11). Accordingly, one can normalize the Killing vectors so that

$$\boxed{\xi = \partial_{\tilde{t}} = \partial_t} \quad \text{and} \quad \boxed{\eta = \partial_{\tilde{\varphi}} = \partial_\varphi}. \quad (13.13)$$

13.2.4 Principal null geodesics

As discussed in Sec. 10.4, the Kerr spacetime is endowed with two congruences of null geodesics 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 $a \rightarrow m$. We can summarize them as follows:

Property 13.2: principal null geodesics of the extremal Kerr spacetime

- The *ingoing principal null geodesics* are the curves

$$\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}} : (v, \theta, \tilde{\varphi}) = \text{const} \in \mathbb{R} \times [0, \pi] \times [0, 2\pi], \quad (13.14)$$

²Cf. Sec. 5.2.1 for the definition of *stationary* and some discussion about the terminology.

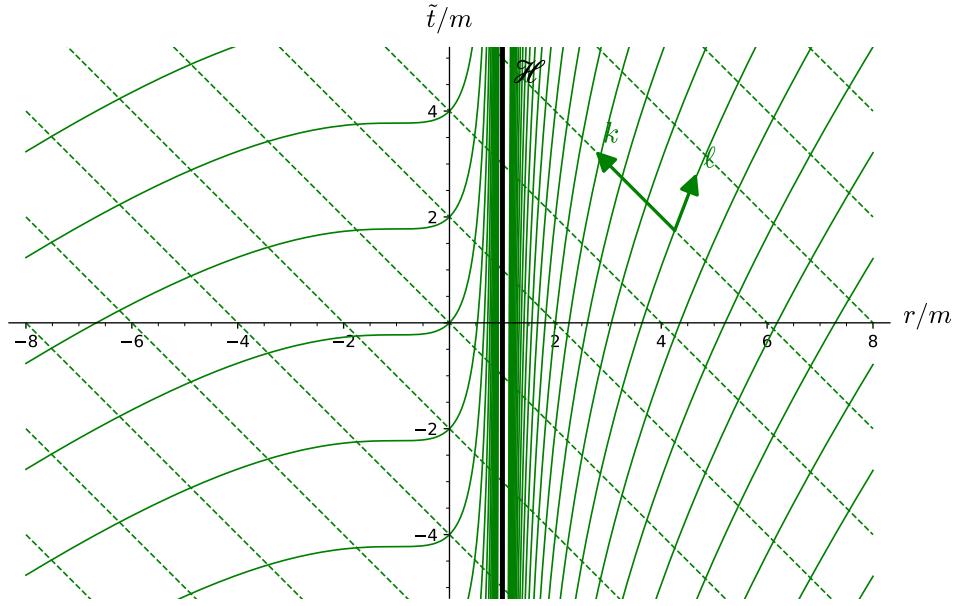


Figure 13.2: Trace of the principal null geodesics in the plane $(\tilde{t}/m, r)$. The dashed lines correspond to the ingoing principal null geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ and the solid curves to the outgoing principal null geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$ for $r > m$ and $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},III}$ for $r < m$. [Figure generated by the notebook D.5.17]

where v is the Kerr advanced time:

$$v := \tilde{t} + r = t + r - \frac{2m^2}{r-m} + 2m \ln \left| \frac{r-m}{m} \right|. \quad (13.15)$$

Along any geodesic $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$, $-r$ is an affine parameter increasing towards the future; the corresponding tangent vector is

$$\mathbf{k} = \partial_{\tilde{t}} - \partial_{\tilde{r}} \quad (13.16a)$$

$$\mathbf{k} = \frac{r^2 + m^2}{(r-m)^2} \partial_t - \partial_r + \frac{m}{(r-m)^2} \partial_\varphi \quad \text{in } \mathcal{M} \setminus \mathcal{H}. \quad (13.16b)$$

- The **outgoing principal null geodesics** are the curves

$$\text{in } \mathcal{M}_I : \quad \mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I} : \quad (u, \theta, \tilde{\varphi}) = \text{const} \in \mathbb{R} \times [0, \pi] \times [0, 2\pi], \quad (13.17a)$$

$$\text{in } \mathcal{M}_{III} : \quad \mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},III} : \quad (u, \theta, \tilde{\varphi}) = \text{const} \in \mathbb{R} \times [0, \pi] \times [0, 2\pi], \quad (13.17b)$$

$$\text{on } \mathcal{H} : \quad \mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}} : \quad (\theta, \psi) = \text{const} \in [0, \pi] \times [0, 2\pi], \quad (13.17c)$$

where u is the Kerr retarded time:

$$u := \tilde{t} - r + \frac{4m^2}{r-m} - 4m \ln \left| \frac{r-m}{m} \right| = t - r + \frac{2m^2}{r-m} - 2m \ln \left| \frac{r-m}{m} \right| \quad (13.18)$$

and $\tilde{\varphi}$ and ψ are defined by

$$\tilde{\varphi} := \tilde{\varphi} + \frac{2m}{r-m} = \varphi + \frac{m}{r-m} \quad (13.19)$$

and

$$\psi := \tilde{\varphi} - \frac{\tilde{t}}{2m}. \quad (13.20)$$

Along the geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},\text{I}}$ and $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},\text{III}}$, r is an affine parameter increasing towards the future, while along $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$, such an affine parameter is \tilde{t} . The tangent vector ℓ to the outgoing principal null geodesics that coincides with the Killing vector $\xi + (2m)^{-1}\eta$ on \mathcal{H} is

$$\ell = \frac{(r+m)^2}{2(r^2+m^2)} \partial_{\tilde{t}} + \frac{(r-m)^2}{2(r^2+m^2)} \partial_{\tilde{r}} + \frac{m}{r^2+m^2} \partial_{\tilde{\varphi}} \quad (13.21\text{a})$$

$$\ell = \frac{1}{2} \partial_t + \frac{(r-m)^2}{2(r^2+m^2)} \partial_r + \frac{m}{2(r^2+m^2)} \partial_\varphi \quad \text{in } \mathcal{M} \setminus \mathcal{H}. \quad (13.21\text{b})$$

Proof. The second equality in Eq. (13.15) follows from Eq. (13.8a). Equation (13.16b) follows from Eq. (13.16a) via Eq. (13.10). Equations (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. (10.59), (10.57) and (10.61). All the other statements follow from the limit $a \rightarrow m$ of results of Sec. 10.4, except for \tilde{t} being an affine parameter along $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$, which is peculiar to the extremal Kerr horizon and will be proven in Sec. 13.2.5. \square

The principal null geodesic congruences are depicted in terms of the (\tilde{t}, r) coordinates in Fig. 13.2. Note that the outgoing geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},\text{I}}$ and $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},\text{III}}$ tend to become tangent to \mathcal{H} for $r \rightarrow m$; this agrees with \mathcal{H} being generated by some members of the outgoing principal null congruence, namely the geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$, 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 (\mathcal{M}, g) shown in Fig. 13.3, in which both families of geodesics appear as straight lines.

13.2.5 The degenerate horizon

\mathcal{H} is the hypersurface of \mathcal{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 [Eq. (13.5)] vanishes at $r = m$, we have $g^{\tilde{u}\tilde{v}} \partial_{\tilde{u}} r \partial_{\tilde{v}} r = 0$ on \mathcal{H} , which implies that the gradient $\vec{\nabla} r$ is a null vector there and that \mathcal{H} is a null hypersurface. Moreover, since the components of $\vec{\nabla} r$ are $\nabla^{\tilde{\alpha}} r = g^{\tilde{\alpha}r}$,

we read on Eq. (13.5) that

$$\vec{\nabla}_r \mathcal{H} = \frac{2m^2}{\rho^2} \partial_{\tilde{t}} + \frac{m}{\rho^2} \partial_{\tilde{\varphi}} \stackrel{\mathcal{H}}{=} \frac{2m^2}{\rho^2} \chi, \quad (13.22)$$

where χ is the Killing vector field

$$\boxed{\chi := \xi + \Omega_{\mathcal{H}} \eta}, \quad \text{with} \quad \boxed{\Omega_{\mathcal{H}} := \frac{1}{2m}}. \quad (13.23)$$

It follows immediately that \mathcal{H} is a *Killing horizon*, i.e. a null hypersurface that admits a Killing vector as null normal (cf. Sec. 3.3.2).

From expression (13.21a) for ℓ , we have immediately

$$\ell \stackrel{\mathcal{H}}{=} \chi. \quad (13.24)$$

This means that the null generators of \mathcal{H} are the outgoing principal null geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$.

The *surface gravity* κ of the Killing horizon \mathcal{H} has been defined in Sec. 3.3.5 as the non-affinity coefficient of the Killing-vector normal χ to \mathcal{H} : $\nabla_{\chi} \chi \stackrel{\mathcal{H}}{=} \kappa \chi$ [Eq. (3.29)]. Given the identity (13.24), κ coincides with the value on \mathcal{H} of the non-affinity coefficient κ_{ℓ} of the tangent ℓ to the outgoing principal null geodesics: $\nabla_{\ell} \ell = \kappa_{\ell} \ell$. A direct computation (cf. the notebook D.5.17) reveals that

$$\kappa_{\ell} = m \frac{r^2 - m^2}{(r^2 + m^2)^2}. \quad (13.25)$$

In particular, κ_{ℓ} vanishes for $r = m$, i.e. on \mathcal{H} . Hence $\kappa = \kappa_{\ell}|_{\mathcal{H}}$ leads to³

$$\boxed{\kappa = 0}. \quad (13.26)$$

According to the classification introduced in Sec. 3.3.6, it follows that \mathcal{H} is a *degenerate Killing horizon*. The vanishing of the non-affinity coefficient κ means that ℓ is a geodesic vector on \mathcal{H} , and not only a pregeodesic one (cf. Remark 1 in Sec. 1.3.2). Equivalently, at any given point $p \in \mathcal{H}$, ℓ is the tangent vector associated to an affine parameter λ of the null geodesic $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ through p . Moreover, the affine parameter λ coincides with \tilde{t} , up to some additive constant. Indeed, Eqs. (13.24) and (13.23) imply

$$\ell^{\tilde{t}} = \frac{d\tilde{t}}{d\lambda} = \chi^{\tilde{t}} = 1,$$

from which $\lambda = \tilde{t} + \text{const.}$ Since the range of \tilde{t} is $(-\infty, +\infty)$, we conclude that $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ is a *complete* geodesic. This contrasts with the null generators of a non-degenerate Killing horizon, which are incomplete, as shown in Sec. 3.4.2.

Let us summarize the results obtained above:

³The vanishing of κ can also be obtained by taking the limit $a \rightarrow m$ of expression (10.77), which has been derived for $a < m$.

Property 13.3: \mathcal{H} as a degenerate Killing horizon

In the extremal Kerr spacetime, the hypersurface \mathcal{H} defined by $r = m$ is a degenerate Killing horizon. Its generators are the outgoing principal null geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$, which are complete geodesics and which admit the Kerr coordinate \tilde{t} as an affine parameter. The tangent vector associated to this affine parameter is the Killing vector $\chi = \xi + \Omega_{\mathcal{H}}\eta$ [Eq. (13.23)], which coincides on \mathcal{H} with the tangent vector ℓ to the outgoing principal null congruence.

13.2.6 Black hole character

As a Killing horizon, \mathcal{H} is a null hypersurface and thus a one-way membrane (cf. Sec. 2.2.2). Since the ingoing principal null geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ cross it from \mathcal{M}_I to \mathcal{M}_{III} (cf. Fig. 13.2), we conclude that no (massive or null) particle can cross \mathcal{H} from \mathcal{M}_{III} to \mathcal{M}_I . In order to show that \mathcal{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:

Property 13.4: black hole region in the extremal Kerr spacetime

The extremal Kerr spacetime (\mathcal{M}, g) can be endowed with a conformal completion at null infinity such that the future and past null infinities \mathcal{J}^+ and \mathcal{J}^- are located at the boundary of \mathcal{M}_I . The region \mathcal{M}_{III} is then the interior of a black hole, the event horizon of which is the Killing horizon \mathcal{H} . Since \mathcal{H} is degenerate, the black hole is *extremal*, according to the terminology introduced in Sec. 5.4.3.

The future null infinity \mathcal{J}^+ and the past null infinity \mathcal{J}^- relative to \mathcal{M}_I are depicted in the Carter-Penrose diagram of Figs. 13.3 - 13.4. In this diagram, it is clear that \mathcal{M}_{III} is a black hole region for \mathcal{M}_I and that \mathcal{H} is the corresponding event horizon.

13.3 Maximal analytic extension

13.3.1 Extension of \mathcal{M}_I for complete outgoing principal null geodesics

Figure 13.3 depicts a Carter-Penrose diagram of the extremal Kerr spacetime (\mathcal{M}, g) built by means of the projection map⁴ $\Pi : \mathcal{M} \rightarrow \mathbb{R}^2, (\tilde{t}, r, \theta, \tilde{\varphi}) \mapsto (T, X)$ defined by

$$T = T_0(u, v) \quad \text{and} \quad X = X_0(u, v) \quad \text{in } \mathcal{M}_I \tag{13.27a}$$

$$T = \arctan(v/2) + \pi/2 \quad \text{and} \quad X = \arctan(v/2) - \pi/2 \quad \text{on } \mathcal{H} \tag{13.27b}$$

$$T = T_0(u, v) + \pi \quad \text{and} \quad X = X_0(u, v) - \pi \quad \text{in } \mathcal{M}_{III} \tag{13.27c}$$

⁴cf. the definition of a Carter-Penrose diagram given in Sec. 10.8.1.

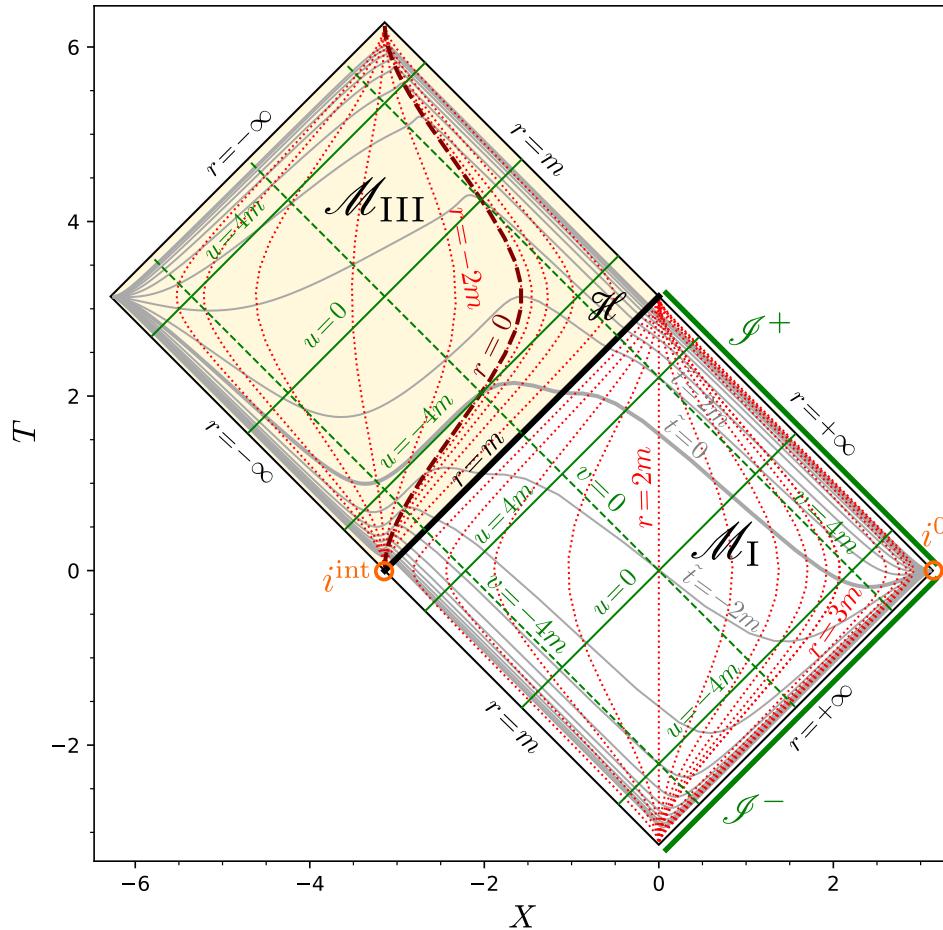


Figure 13.3: Carter-Penrose diagram of the extremal Kerr spacetime (\mathcal{M}, g) constructed via the projection map $\Pi : \mathcal{M} \rightarrow \mathbb{R}^2, (\tilde{t}, r, \theta, \tilde{\varphi}) \mapsto (T, X)$ defined by Eqs. (13.27)-(13.28). The grey curves represent hypersurfaces $\tilde{t} = \text{const.}$, 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 $\delta r = 0.2m$ 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). i^{int} is the internal infinity discussed in Sec. 13.4.1. [Figure generated by the notebook D.5.17]

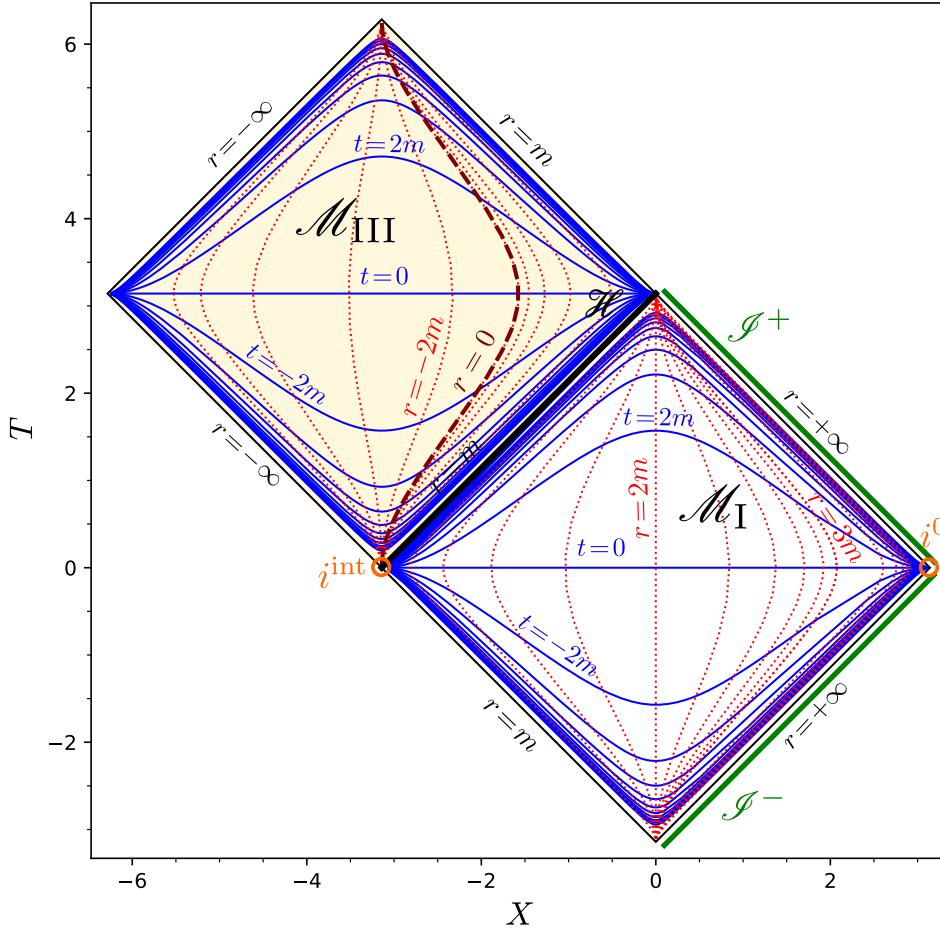


Figure 13.4: Same as Fig. 13.3 but for the time slicing associated to Boyer-Lindquist coordinates (t, r, θ, φ) . The blue curves represent hypersurfaces $t = \text{const}$, with $t \in [-20m, 20m]$ and the increment $\delta t = 2m$ between two successive hypersurfaces. Note that the spacetime slicing by the hypersurfaces $t = \text{const}$ is singular at \mathcal{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.5.17]

with

$$T_0(u, v) := \arctan\left(\frac{u}{2}\right) + \arctan\left(\frac{v}{2}\right), \quad X_0(u, v) := \arctan\left(\frac{v}{2}\right) - \arctan\left(\frac{u}{2}\right), \quad (13.28)$$

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 $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$ and $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},\text{III}}$ are plotted in Fig. 13.3 for selected values of u (solid green lines): $u = -4m$, $u = 0$ and $u = 4m$. Since r is an affine parameter along the null geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$, 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 \mathcal{M}_{I} in Fig. 13.3), without any possible extension into \mathcal{M}_{I} from there. To extend \mathcal{M}_{I} so that all geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$ are complete, let us introduce a coordinate system on \mathcal{M}_{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** $(x^{\tilde{\alpha}}) = (\tilde{t}, r, \theta, \tilde{\varphi})$ by

$$u = \tilde{t} - r \iff \tilde{t} = u + r, \quad (13.29)$$

where u is the retarded Kerr time (13.18) and $\tilde{\varphi}$ 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 $\tilde{t} = u + r$ and using Eq. (13.19) linking $\tilde{\varphi}$ to $\tilde{\varphi}$, we get the transition map between the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ and the outgoing Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$:

$$\text{on } \mathcal{M}_1, \quad \begin{cases} \tilde{t} = \tilde{t} + \frac{4m^2}{r-m} - 4m \ln \left(\frac{r-m}{m} \right) \\ \tilde{\varphi} = \tilde{\varphi} + \frac{2m}{r-m}. \end{cases} \quad (13.30)$$

By construction, the tangent vector to $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$ associated with the affine parameter r is then

$$\ell' = \partial_{\tilde{t}} + \partial_{\tilde{r}}, \quad (13.31)$$

where $\partial_{\tilde{r}}$ stands for the vector $\partial/\partial r$ of the coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$. Indeed $\ell'^{\tilde{\alpha}} = dx^{\tilde{\alpha}}/dr = (1, 1, 0, 0)$ since along $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$, $\tilde{t} = r + u$, with u constant, and both θ and $\tilde{\varphi}$ are constant. An explicit computation (cf. the notebook D.5.18) shows that ℓ' is a geodesic vector:

$$\nabla_{\ell'} \ell' = 0, \quad (13.32)$$

which confirms that r is an affine parameter along $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$. ℓ' is thus similar to k , which is the tangent vector to the *ingoing* principal null geodesics $\mathcal{L}_{(v,\theta,\tilde{\varphi})}^{\text{in}}$ associated with the affine parameter $-r$ along them. In this respect, note the symmetry between the relations $\ell' = \partial_{\tilde{t}} + \partial_{\tilde{r}}$ and $k = \partial_{\tilde{t}} - \partial_{\tilde{r}}$ [Eq. (13.16a)]. The link between ℓ' and the tangent vector ℓ to $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$ introduced in Sec. 13.2.4 is easily obtained from the definition of a tangent vector to a curve:

$$\ell' = \frac{dx}{dr} = \frac{dx}{d\lambda} \frac{d\lambda}{dr} = \left(\frac{dr}{d\lambda} \right)^{-1} \ell,$$

where λ is the (non-affine) parameter of $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$ associated with ℓ . We read on the Kerr components (13.21a) of ℓ , as well as on the Boyer-Lindquist ones (13.21b), that $dr/d\lambda = \ell^r = (r-m)^2/(2(r^2+m^2))$. Hence

$$\ell' = 2 \frac{r^2 + m^2}{(r-m)^2} \ell. \quad (13.33)$$

Differentiating Eq. (13.30) leads to

$$d\tilde{t} = d\tilde{t} - \frac{4mr}{(r-m)^2} dr \quad \text{and} \quad d\tilde{\varphi} = d\tilde{\varphi} - \frac{2m}{(r-m)^2} dr. \quad (13.34)$$

From these relations and the chain rule, we get immediately the link between the outgoing Kerr coordinate frame and the Kerr coordinate frame:

$$\partial_{\tilde{t}} = \partial_t, \quad \partial_{\tilde{r}} = \partial_r + \frac{4mr}{(r-m)^2} \partial_t + \frac{2m}{(r-m)^2} \partial_{\tilde{\varphi}}, \quad \partial_\theta = \partial_\theta, \quad \partial_{\tilde{\varphi}} = \partial_{\tilde{\varphi}}. \quad (13.35)$$

In view of Eq. (13.13), we conclude that $\partial_{\tilde{t}}$ and $\partial_{\tilde{\varphi}}$ coincide with the Killing vectors ξ and η of the Kerr metric:

$$\partial_{\tilde{t}} = \xi \quad \text{and} \quad \partial_{\tilde{\varphi}} = \eta. \quad (13.36)$$

The link between the outgoing Kerr coordinates and the Boyer-Lindquist ones is obtained by substituting Eq. (13.18) for u into $\tilde{t} = u + r$ [Eq. (13.29)]:

$$\tilde{t} = t + \frac{2m^2}{r-m} - 2m \ln \left| \frac{r-m}{m} \right|. \quad (13.37)$$

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 $(\tilde{t}, r, \theta, \tilde{\varphi})$. Differentiating Eqs. (13.37) and (13.19) leads to

$$d\tilde{t} = dt - \frac{2mr}{(r-m)^2} dr \quad \text{and} \quad d\tilde{\varphi} = d\varphi - \frac{m}{(r-m)^2} dr. \quad (13.38)$$

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 $(\tilde{t}, r, \theta, \tilde{\varphi})$ 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 expression of the metric tensor with respect to the outgoing Kerr coordinates is easily obtained by substituting dt and $d\varphi$ from Eq. (13.38) into the Boyer-Lindquist expression (13.11) (see also the notebook D.5.18); we get

$$\begin{aligned} g = & - \left(1 - \frac{2mr}{\rho^2} \right) d\tilde{t}^2 - \frac{4mr}{\rho^2} d\tilde{t} dr - \frac{4m^2 r \sin^2 \theta}{\rho^2} d\tilde{t} d\tilde{\varphi} + \left(1 + \frac{2mr}{\rho^2} \right) dr^2 \\ & + 2m \left(1 + \frac{2mr}{\rho^2} \right) \sin^2 \theta dr d\tilde{\varphi} + \rho^2 d\theta^2 + \left(r^2 + m^2 + \frac{2m^3 r \sin^2 \theta}{\rho^2} \right) \sin^2 \theta d\tilde{\varphi}^2. \end{aligned} \quad (13.39)$$

These metric components are very similar to those in Kerr coordinates, as given by Eq. (13.3): the only differences are $g_{\tilde{t}\tilde{t}}$ and $g_{r\tilde{\varphi}}$, which have a sign opposite to that of respectively $g_{\tilde{t}\tilde{t}}$ and $g_{r\tilde{\varphi}}$. Apart from the standard singularities of the spherical coordinates $(\theta, \tilde{\varphi})$ 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 \mathcal{M}_I . In particular there is no divergence for $r \rightarrow m$. This can be used to extend smoothly the spacetime (\mathcal{M}_I, g) to $r \in (-\infty, m]$, so that the outgoing principal null geodesics $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$ with $\theta \neq \pi/2$ become complete. But the extension to $r < m$ cannot be \mathcal{M}_{III} as it appears clearly on Fig. 13.3 that the end point of $\mathcal{L}_{(u,\theta,\tilde{\varphi})}^{\text{out},I}$ for $r \rightarrow m^+$ is not located at the boundary between \mathcal{M}_I and \mathcal{M}_{III} . We thus introduce a spacetime (\mathcal{M}', g) from a new copy of $\mathbb{R}^2 \times \mathbb{S}^2$ with \mathbb{R}^2 spanned by the coordinates (\tilde{t}, r) and \mathbb{S}^2 spanned by the coordinates $(\theta, \tilde{\varphi})$, such that (i) the manifold \mathcal{M}' is

$$\mathcal{M}' := \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathcal{R}' \quad \text{where} \quad \mathcal{R}' := \left\{ p \in \mathbb{R}^2 \times \mathbb{S}^2, \quad r(p) = 0 \text{ and } \theta(p) = \frac{\pi}{2} \right\}, \quad (13.40)$$

(ii) \mathcal{M}_I is identified with the part $r > m$ of \mathcal{M}' and (iii) in all \mathcal{M}' , \mathbf{g} has the components given by expression (13.39). Furthermore, we define

$$\mathcal{H}' := \{p \in \mathcal{M}', \quad r(p) = m\} \quad \text{and} \quad \mathcal{M}'_{III} := \{p \in \mathcal{M}', \quad r(p) < m\}. \quad (13.41)$$

We have then $\mathcal{M}_I = \mathcal{M} \cap \mathcal{M}'$. In \mathcal{M}'_{III} , one can define Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ from $(\tilde{t}, r, \varphi, \tilde{\varphi})$ via formulas (13.18) (with $u = \tilde{t} - r$) and (13.19). It appears then immediately that $(\mathcal{M}'_{III}, \mathbf{g})$ is isometric to $(\mathcal{M}_{III}, \mathbf{g})$. It follows that \mathbf{g} obeys the vacuum Einstein equation in all \mathcal{M}' . The vector field $\ell' := \partial_{\tilde{t}} + \partial_{\tilde{r}}$ [cf. Eq. (13.31)] is a smooth non-vanishing null vector field on \mathcal{M}' . Since it is future directed in $(\mathcal{M}_I, \mathbf{g})$ considered as a part of $(\mathcal{M}, \mathbf{g})$, we use it to set the time orientation in all \mathcal{M}' .

The tangent vector \mathbf{k} to ingoing principal null geodesics has the following components with respect to the outgoing Kerr coordinates:

$$\mathbf{k} = \frac{(r+m)^2}{(r-m)^2} \partial_{\tilde{t}} - \partial_{\tilde{r}} + \frac{2m}{(r-m)^2} \partial_{\tilde{\varphi}}. \quad (13.42)$$

This follows immediately from $\mathbf{k} = \partial_{\tilde{t}} - \partial_{\tilde{r}}$ [Eq. (13.16a)] and using Eq. (13.35) to substitute $\partial_{\tilde{t}}$ and $\partial_{\tilde{r}}$. Extending Equation (13.42) to all \mathcal{M}' leads to a vector field that is singular on \mathcal{H}' . To get a vector field everywhere regular on \mathcal{M}' , we rescale it by the inverse of the factor connecting ℓ to ℓ' in Eq. (13.33), i.e. we define

$$\mathbf{k}' := \frac{(r-m)^2}{2(r^2+m^2)} \mathbf{k}. \quad (13.43)$$

Hence

$$\mathbf{k}' = \frac{(r+m)^2}{2(r^2+m^2)} \partial_{\tilde{t}} - \frac{(r-m)^2}{2(r^2+m^2)} \partial_{\tilde{r}} + \frac{m}{r^2+m^2} \partial_{\tilde{\varphi}}. \quad (13.44)$$

This vector field is clearly regular in all \mathcal{M}' . Accordingly, it can be used to extend smoothly the family of ingoing principal null geodesics to \mathcal{H}' . The price to pay is that \mathbf{k}' is only a pregeodesic vector field, while \mathbf{k} was geodesic, being associated with the affine parameter $-r$. The pair $(\mathbf{k}, \mathbf{k}')$ plays actually the same role as the pair (ℓ', ℓ) (note the order!) regarding the outgoing principal null geodesics.

As \mathcal{H} in $(\mathcal{M}, \mathbf{g})$, \mathcal{H}' is a degenerate Killing horizon of $(\mathcal{M}', \mathbf{g})$. Indeed, by the same reasoning as in Sec. 13.2.5, we get that the null normal to \mathcal{H}' is the Killing vector $\chi = \xi + 1/(2m) \boldsymbol{\eta}$ [Eq. (13.23) extended to \mathcal{M}']. This normal coincides with \mathbf{k}' on \mathcal{H}' , as we can see by setting $r = m$ in Eq. (13.44). This implies that the null geodesic generators of \mathcal{H}' belong to the ingoing principal null congruence. The non-affinity coefficient of \mathbf{k}' is (cf. the notebook D.5.18):

$$\kappa_{\mathbf{k}'} = m \frac{m^2 - r^2}{(r^2 + m^2)^2}. \quad (13.45)$$

We have thus $\kappa_{\mathbf{k}'} = 0$ on \mathcal{H}' , so that \mathcal{H}' is a degenerate Killing horizon.

With \mathcal{M}' , our extended extremal Kerr spacetime is thus $(\mathcal{M}_0, \mathbf{g})$ with

$$\boxed{\mathcal{M}_0 := \mathcal{M} \cup \mathcal{M}' = \mathcal{M}_I \cup \mathcal{M}_{III} \cup \mathcal{M}'_{III} \cup \mathcal{H} \cup \mathcal{H}'}. \quad (13.46)$$

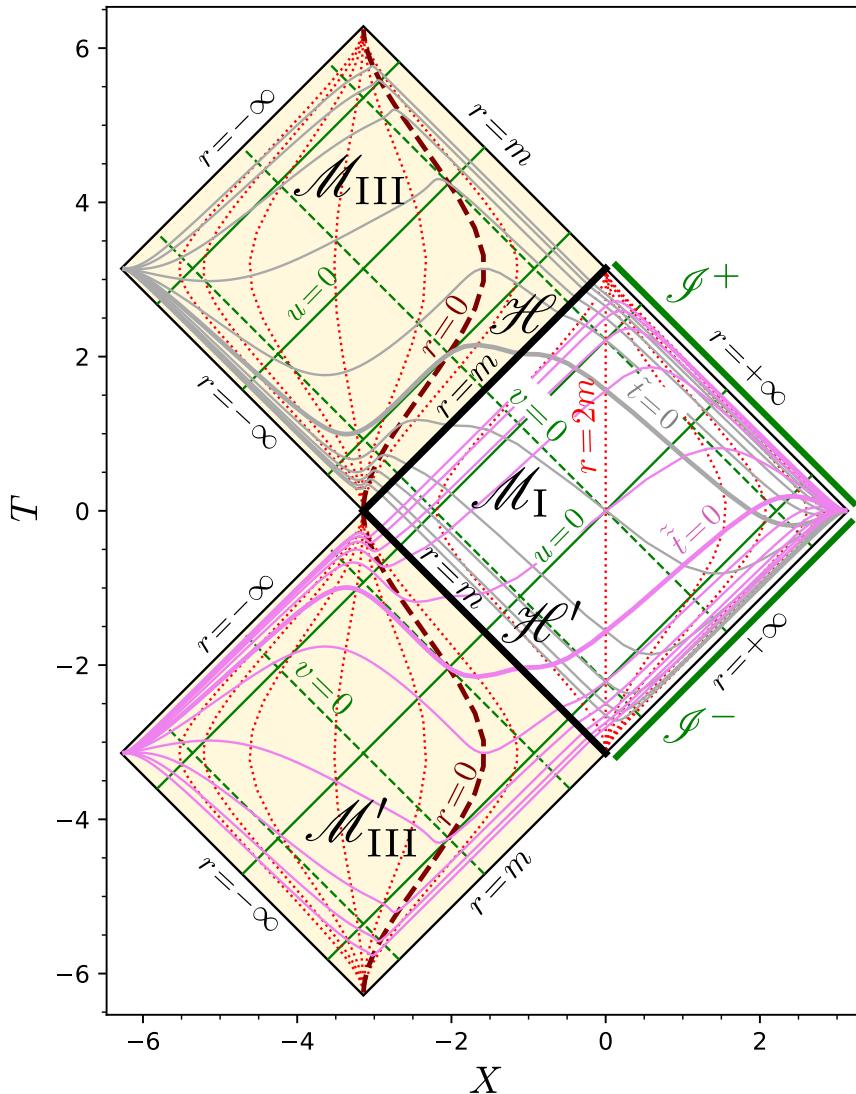


Figure 13.5: Carter-Penrose diagram of the (partially) extended extremal Kerr spacetime $(\mathcal{M}_0, \mathbf{g})$. The grey curves represent hypersurfaces $\tilde{t} = \text{const}$ in \mathcal{M} , with $\tilde{t} \in [-10m, 10m]$ and the increment $\delta\tilde{t} = 2m$ between two successive hypersurfaces. The hypersurface $\tilde{t} = 0$ is singled out by a larger thickness. The purple curves represent hypersurfaces $\tilde{t} = \text{const}$ in \mathcal{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 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.5.18](#)]

A Carter-Penrose diagram of \mathcal{M}_0 is shown in Fig. 13.5. The projection operator used to build this diagram (cf. Sec. 10.8.1) is $\Pi : \mathcal{M}_0 \rightarrow \mathbb{R}^2$, with \mathbb{R}^2 spanned by coordinates (T, X) , such that Π is defined by Eqs. (13.27)-(13.28) on \mathcal{M} , while on \mathcal{M}' , Π is defined by

$$T = T_0(u, v) \quad \text{and} \quad X = X_0(u, v) \quad \text{in } \mathcal{M}_I \quad (13.47a)$$

$$T = \arctan(u/2) - \pi/2 \quad \text{and} \quad X = -\arctan(u/2) - \pi/2 \quad \text{on } \mathcal{H}' \quad (13.47b)$$

$$T = T_0(u, v) - \pi \quad \text{and} \quad X = X_0(u, v) - \pi \quad \text{in } \mathcal{M}'_{III} \quad (13.47c)$$

where $T_0(u, v)$ and $X_0(u, v)$ are given by Eq. (13.28) and u and v are functions of (\tilde{t}, r) defined by respectively $u = \tilde{t} - r$ [Eq. (13.29)] and

$$v = \tilde{t} + r - \frac{4m^2}{r-m} + 4m \ln \left| \frac{r-m}{m} \right|. \quad (13.48)$$

The last equation is obtained by combining Eqs. (13.15) and (13.18).

Remark 2: As we have constructed it, the manifold \mathcal{M}_0 is covered by two coordinate charts: $(\tilde{t}, r, \theta, \tilde{\varphi})$ on $\mathcal{M}_I \cup \mathcal{H} \cup \mathcal{M}_{III}$ and $(\tilde{t}, r, \theta, \tilde{\varphi})$ on $\mathcal{M}_I \cup \mathcal{H}' \cup \mathcal{M}'_{III}$. These two charts overlap in \mathcal{M}_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 \mathcal{M}_0 by a single chart, in a fashion similar to the covering of the entire maximal extension of Schwarzschild spacetime 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 \mathcal{H}' , $\tilde{\varphi}$ diverges on \mathcal{H} and the Boyer-Lindquist coordinate φ diverges on both \mathcal{H} and \mathcal{H}' . Accordingly, none of $(T, X, \theta, \tilde{\varphi})$, $(T, X, \theta, \tilde{\varphi})$ and (T, X, θ, φ) would provide a regular chart of \mathcal{M}_0 .

Remark 3: The spacetime (\mathcal{M}_0, g) is analytic: (i) \mathcal{M}_0 is an analytic manifold, given that the change of coordinates (13.30) between the two charts covering \mathcal{M}_0 is analytic (cf. Remark 4 in Sec. A.2.1) and (ii) the components (13.3) and (13.39) of g are analytic functions of the coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ and $(\tilde{t}, r, \theta, \tilde{\varphi})$ respectively.

The Killing horizon \mathcal{H}' is actually a white hole horizon from the point of view of \mathcal{M}_I . More precisely, we can endow (\mathcal{M}_0, g) with the same conformal completion at null infinity as that used for (\mathcal{M}, g) in Sec. 13.2.6, i.e. such that the conformal boundary \mathcal{J} is constituted by the same future and past null infinities \mathcal{J}^+ and \mathcal{J}^- located at the boundary of \mathcal{M}_I (cf. Fig. 13.5). It appears then that

$$\mathcal{M}'_{III} \cup \mathcal{H}' = \mathcal{M}_0 \setminus (J^+(\mathcal{J}^-) \cap \mathcal{M}_0) \quad (13.49)$$

In view of the definition (4.41), we conclude:

Property 13.5: white hole region in the extended extremal Kerr spacetime

\mathcal{M}'_{III} is the interior of a white hole region, the boundary of which is \mathcal{H}' .

Since \mathcal{M}_{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:

Property 13.6: domain of outer communications

\mathcal{M}_I is the domain of outer communications of the spacetime $(\mathcal{M}_0, \mathbf{g})$.

13.3.2 Construction of the maximal analytic extension

With $(\mathcal{M}_0, \mathbf{g})$, we have achieved our first goal: all the outgoing principal null geodesics crossing \mathcal{M}_I and not lying in the equatorial plane (i.e. the geodesics extending the $\mathcal{L}_{(u, \theta, \tilde{\varphi})}^{\text{out}, I}$ family with $\theta \neq \pi/2$ to the past) are complete. However, there remains incomplete geodesics in \mathcal{M}_0 : the outgoing principal null geodesics crossing \mathcal{M}_{III} all stop at the value $r = m$ of their affine parameter, while the ingoing principal null geodesics crossing \mathcal{M}'_{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 encounter the curvature singularity at $r = 0$ and $\theta = \pi/2$, one introduces an infinite number of copies of \mathcal{M}_0 , $(\mathcal{M}_n)_{n \in \mathbb{Z}}$ say, and identify the region \mathcal{M}'_{III} of \mathcal{M}_n with the region \mathcal{M}_{III} of \mathcal{M}_{n-1} for all $n \in \mathbb{Z}$. The manifold hence obtained,

$$\mathcal{M}_* := \bigcup_{n \in \mathbb{Z}} \mathcal{M}_n, \quad (13.50)$$

is depicted via a Carter-Penrose diagram in Fig. 13.6. The region \mathcal{M}_I (resp. \mathcal{M}_{III}) of \mathcal{M}_n is denoted by $\mathcal{M}_I^{(n)}$ (resp. $\mathcal{M}_{III}^{(n)}$). Similarly, the Killing horizon \mathcal{H} (resp. \mathcal{H}') of \mathcal{M}_n is denoted by $\mathcal{H}_{(n)}^+$ (resp. $\mathcal{H}_{(n)}^-$). Note that $\mathcal{H}_{(n)}^+$ is a future event horizon (black hole horizon), while $\mathcal{H}_{(n)}^-$ is a past event horizon (white hole horizon), for the conformal completion at null infinity with the future and past null infinities $\mathcal{I}_{(n)}^+$ and $\mathcal{I}_{(n)}^-$ as copies of \mathcal{I}^+ and \mathcal{I}^- introduced for \mathcal{M}_0 .

It is clear that $(\mathcal{M}_*, \mathbf{g})$ is an analytic spacetime, since $(\mathcal{M}_0, \mathbf{g})$ is (cf. Remark 3 on p. 518).

By construction, all principal null geodesics are complete in \mathcal{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 $\mathcal{H}_{(n)}^+$ and for the ingoing ones generating $\mathcal{H}_{(n)}^-$. 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 representing a complete principal null geodesic.

It can be shown that actually *all* timelike or null geodesics of $(\mathcal{M}_*, \mathbf{g})$, and not only the principal null ones, are complete (see Carter's article [89] for the proof), so that we can conclude:

Property 13.7: maximal analytic extension of the extremal Kerr spacetime

$(\mathcal{M}_*, \mathbf{g})$ 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 $a < m$ 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 $(\tilde{\tilde{t}}, r, \theta, \tilde{\varphi})$, one ends up with a manifold

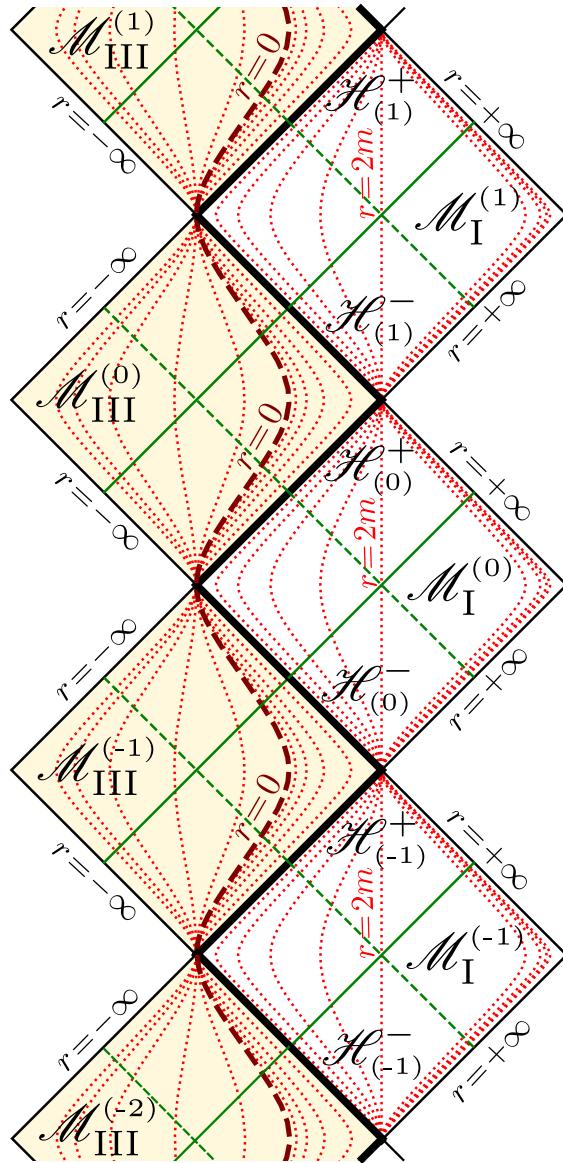


Figure 13.6: Carter-Penrose diagram of the maximal analytic extension (\mathcal{M}_*, g) of the extremal Kerr spacetime. The red dotted curves represent hypersurfaces $r = \text{const}$, with the increment δr between two successive hypersurfaces being $\delta r = 2m$ for $r < 0$ and $r > 2m$ and $\delta r = 0.2m$ 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.5.18](#)]

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. 10.11), 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. 3.4), 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 $\mathcal{H}_{(n)}^+$ or $\mathcal{H}_{(n)}^-$ is degenerate and therefore made of complete null geodesics. In particular, $\mathcal{H}_{(n)}^+$ and $\mathcal{H}_{(n)}^-$ are not part of a bifurcate Killing horizon. In other words, there is no bifurcation sphere 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 [87] obtained the maximal analytic extension of the rotation axis \mathcal{A} of the extremal Kerr spacetime (cf. historical note on p. 366). In particular, he drew a diagram (Fig. 1b of Ref. [87]) similar to that of Fig. 13.6, the difference being that Carter's one is a true conformal representation, owing to the fact that \mathcal{A} is 2-dimensional, while the diagram of Fig. 13.6 is a mere projection of the 4-dimensional manifold \mathcal{M}_* . In their 1967 classical study entitled *Maximal Analytic Extension of the Kerr metric* [71], Robert H. Boyer and Richard W. Lindquist focused on the case $a < m$. For $a = m$, they referred to Carter's study [87] 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 [89], 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 [94].

13.4 Near-horizon extremal Kerr (NHEK) geometry

13.4.1 The extremal Kerr throat

Let us consider a hypersurface Σ_t of constant Boyer-Lindquist time t in the external region \mathcal{M}_I of the extremal Kerr spacetime, i.e. one of the hypersurfaces shown in blue in Figs. 13.1 and 13.4. Σ_t is a spacelike hypersurface, since it is clear from the line element (13.11) that the metric induced by \mathbf{g} on Σ_t is Riemannian (i.e. positive definite) in⁵ \mathcal{M}_I . We may thus visualize the geometry of Σ_t by some isometric embedding of 2-dimensional slices of Σ_t into the Euclidean space \mathbb{R}^3 , as we did in Sec. 9.5.2 for the hypersurfaces of constant Kruskal-Szekeres time in Schwarzschild spacetime. It is then natural to consider the slices $\Sigma_{t,\theta}$ where θ is held constant, in addition to t , with $\Sigma_{t,\frac{\pi}{2}}$ representing the equatorial “plane”. $\Sigma_{t,\theta}$ is spanned by the coordinates $(x^a) = (r, \varphi)$ and the Riemannian metric \mathbf{q} induced on it by the spacetime metric \mathbf{g} is obtained by setting $dt = 0$ and $d\theta = 0$ in (13.11):

$$\mathbf{q} = \frac{r^2 + m^2 \cos^2 \theta}{(r - m)^2} dr^2 + \left(r^2 + m^2 + \frac{2m^3 r \sin^2 \theta}{r^2 + m^2 \cos^2 \theta} \right) \sin^2 \theta d\varphi^2. \quad (13.51)$$

⁵The induced metric would not be Riemannian in the Carter time machine (cf. Sec. 10.2.5), where $g_{\varphi\varphi} < 0$; but this requires $r < 0$, while $r > m$ in \mathcal{M}_I .

On the other side, the Euclidean space \mathbb{R}^3 can be described by cylindrical coordinates $(X^i) = (\varpi, z, \varphi)$, such that the Euclidean metric \mathbf{f} takes the form

$$\mathbf{f} = d\varpi^2 + dz^2 + \varpi^2 d\varphi^2. \quad (13.52)$$

Let us define the embedding of the 2-surface $\Sigma_{t,\theta}$ into \mathbb{R}^3 by

$$\begin{aligned} \Phi : \Sigma_{t,\theta} &\longrightarrow \mathbb{R}^3 \\ (r, \varphi) &\longmapsto (\varpi(r), z(r), \varphi). \end{aligned} \quad (13.53)$$

Along $\Phi(\Sigma_{t,\theta})$, we have then $d\varpi = \varpi'(r) dr$ and $dz = z'(r) dr$, so that the metric \mathbf{h} induced by \mathbf{f} on $\Phi(\Sigma_{t,\theta})$ is

$$\mathbf{h} = (\varpi'(r)^2 + z'(r)^2) dr^2 + \varpi(r)^2 d\varphi^2. \quad (13.54)$$

Φ performs an isometric embedding iff the two metrics (13.51) and (13.54) coincide⁶, i.e. iff

$$\varpi(r) = \sin \theta \sqrt{r^2 + m^2 + \frac{2m^3 r \sin^2 \theta}{r^2 + m^2 \cos^2 \theta}} \quad (13.55a)$$

$$\varpi'(r)^2 + z'(r)^2 = \frac{r^2 + m^2 \cos^2 \theta}{(r - m)^2}. \quad (13.55b)$$

We deduced from Eq. (13.55b) that

$$z(r) = \int_{2m}^r \frac{1}{\bar{r} - m} \sqrt{\bar{r}^2 + m^2 \cos^2 \theta - \varpi'(\bar{r})^2(\bar{r} - m)^2} d\bar{r}, \quad (13.56)$$

up to some additive constant, which we set to zero by choosing arbitrarily $z(2m) = 0$. In the integrand, $\varpi'(\bar{r})$ should be substituted by the value obtained by differentiating Eq. (13.55a).

The functions $\varpi(r)$ and $z(r)$, provided respectively by Eqs. (13.55a) and (13.56), define fully the isometric embedding (13.53) of the Riemannian 2-manifold $(\Sigma_{t,\theta}, \mathbf{q})$ into the Euclidean space $(\mathbb{R}^3, \mathbf{f})$. The outcome is depicted in Fig. 13.7 for two values of θ : $\pi/2$ (left plot) and $\pi/6$ (right plot). In both cases, it appears that the embedded surface is infinite in two limits: for $r \rightarrow +\infty$ and for $r \rightarrow m$.

The first limit, which implies $\varpi \rightarrow +\infty$ and $z \rightarrow +\infty$ according to Eqs. (13.55a) and (13.56), is towards the asymptotic flat end of \mathcal{M}_1 . More precisely, since t is held fixed on $\Sigma_{t,\theta}$, the asymptotic direction appears as the spacelike infinity point i^0 at the right-most corner of the Carter-Penrose diagram of Fig. 13.4.

The second limit, $r \rightarrow m$, corresponds to $\varpi \rightarrow 2m \frac{\sin \theta}{\sqrt{1 + \cos^2 \theta}}$ and $z \rightarrow -\infty$, according to Eqs. (13.55a) and (13.56). This describes an infinite cylinder, since ϖ tends to a constant value, while z decays to $-\infty$. The divergence of z for $r \rightarrow m$ is of course due to the factor $(\bar{r} - m)^{-1}$ in the integrand of (13.56), which leads to the following behavior:

$$z(r) \sim m \sqrt{1 + \cos^2 \theta} \ln \left(\frac{r - m}{m} \right) \quad \text{when } r \rightarrow m. \quad (13.57)$$

⁶Formally, one says that the metric \mathbf{q} is the pullback of \mathbf{h} by Φ : $\mathbf{q} = \Phi^* \mathbf{h}$.

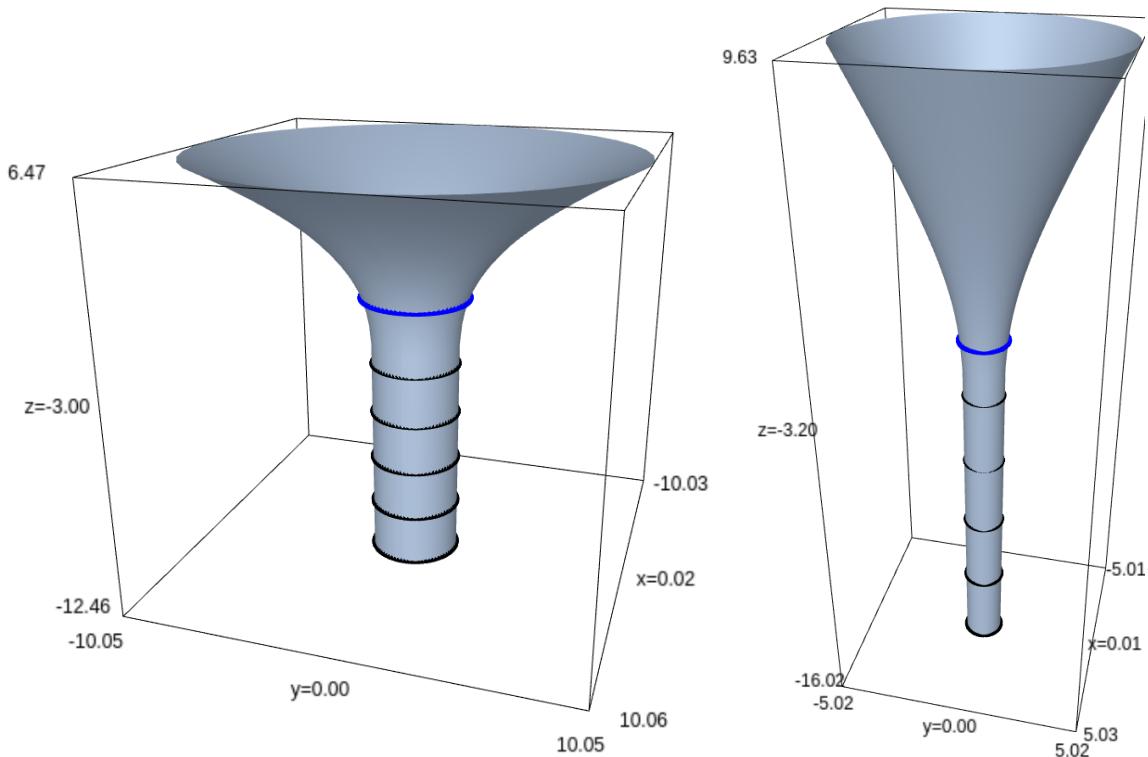


Figure 13.7: Isometric embeddings of 2-surfaces $\Sigma_{t,\theta}$ of constant Boyer-Lindquist time t and constant θ of the extremal Kerr spacetime in the Euclidean 3-space, for two values of θ : $\theta = \pi/2$ (equatorial plane) (left figure) and $\theta = \pi/6$ (right figure). The drawings are truncated at $r = (1 + 10^{-5})m$ (bottom boundary) and at $r = 10m$ (top boundary). The blue circle marks the ergosphere, which is located at $r = 2m$ for $\theta = \pi/2$ and $r = 3m/2$ for $\theta = \pi/6$. The five black circles correspond to $r = 1.1m, 1.01m, 1.001m, 1.0001m$ and $1.00001m$, from top to bottom. [Figure generated by the notebook [D.5.19](#)]

This logarithmic divergence is clearly apparent in Fig. 13.7: the circles at $r - m = 10^{-1}m, 10^{-2}m, \dots, 10^{-5}m$ look equally spaced, all the more than r is close to m . In the Carter-Penrose diagram of Fig. 13.4, the limit $r \rightarrow m$ at fixed t appears as the point marked i^{int} . This point is actually the projection of a 2-sphere (spanned by (θ, φ)) called the ***internal infinity of the extremal Kerr spacetime***. Let us stress that, similarly to the spacelike infinity i^0 and the null infinities \mathcal{I}^+ and \mathcal{I}^- , the sphere i^{int} is not part of the physical spacetime (\mathcal{M}, g) .

The property of i^{int} being located infinitely far from any point at $r = r_0 > m$ on a hypersurface Σ_t appears clearly on the metric (13.11). Indeed, according to the latter, the distance along a curve $(t, \theta, \varphi) = \text{const}$ between the points at $r = r_0$ and $r = m$ is

$$\ell = \int_m^{r_0} \sqrt{g_{rr}} dr = \int_m^{r_0} \frac{\sqrt{r^2 + m^2 \cos^2 \theta}}{r - m} dr.$$

Again, the factor $(r - m)^{-1}$ makes the integral diverge, so that $\ell = +\infty$.

Another perspective of i^{int} being “infinitely far” is obtained by noticing that on the Carter-Penrose diagram of Fig. 13.4, i^{int} is located at the past end of the event horizon \mathcal{H} . Since \mathcal{H} is a degenerate Killing horizon (cf. Sec. 13.2.5), it is generated by *complete* null geodesics: the principal null geodesics $\mathcal{L}_{(\theta, \psi)}^{\text{out}, \mathcal{H}}$ introduced in Sec. 13.2.4. Now, since $\mathcal{L}_{(\theta, \psi)}^{\text{out}, \mathcal{H}}$ is complete, its

past end at i^{int} is obtained at the limit $\tilde{t} \rightarrow -\infty$ of its affine parameter \tilde{t} (see also the grey curves in Fig. 13.3, which reveal clearly that the limit $\tilde{t} \rightarrow -\infty$ at $r = m$ is i^{int}). Hence, in terms of affine length, i^{int} is infinitely far in the past along any of the null geodesics $\mathcal{L}_{(\theta,\psi)}^{\text{out},\mathcal{H}}$ generating \mathcal{H} .

The spacetime region in the vicinity of $r = m$ is called the ***extremal Kerr throat***. We are going to explore it in more details in the next sections.

13.4.2 The NHEK metric

In order to zoom in on the near-horizon region, let us introduce a small parameter $\varepsilon > 0$ and the ***Bardeen-Horowitz coordinates***⁷ (T, R, θ, Φ) , which are defined on $\mathcal{M}_{\text{BL}} = \mathcal{M}_I \cup \mathcal{M}_{\text{III}}$ and related to the Boyer-Lindquist coordinates (t, r, θ, φ) by

$$T = \varepsilon \frac{t}{2m}, \quad R = \frac{r - m}{\varepsilon m}, \quad \Phi = \varphi - \frac{t}{2m}, \quad (13.58)$$

the reverse transformation being

$$t = 2m \frac{T}{\varepsilon}, \quad r = m(1 + \varepsilon R), \quad \varphi = \Phi + \frac{T}{\varepsilon}. \quad (13.59)$$

Note that on \mathcal{M}_{BL} , the range of T is \mathbb{R} , while that of R is $(-\infty, 0) \cup (0, +\infty)$, with $R > 0$ on \mathcal{M}_I and $R < 0$ on \mathcal{M}_{III} . The coordinate choice (13.59) ensures, for a fixed value of (T, R, Φ) with $T > 0$,

$$\lim_{\varepsilon \rightarrow 0} t = +\infty, \quad \lim_{\varepsilon \rightarrow 0} r = m, \quad \lim_{\varepsilon \rightarrow 0} \varphi = +\infty, \quad \frac{d\varphi}{dt} = \frac{1}{2m} \quad (13.60)$$

and is motivated as follows. For $\varepsilon \rightarrow 0$, the coordinate R clearly implements the zoom on the region $r \simeq m$, i.e. the region close to the event horizon \mathcal{H} . Regarding the change $t \rightarrow T$, we may notice that the Boyer-Lindquist coordinate t diverges near \mathcal{H} — more precisely, t diverges along the worldline of any particle that reaches \mathcal{H} (this is clear on Figs. 13.1 and 13.4). So, we may keep $T = \varepsilon t / (2m)$ finite in the vicinity of \mathcal{H} , while having t diverging. The division by $2m$ makes T dimensionless, as R , with the factor 2 being chosen for later convenience. Similarly, the Boyer-Lindquist coordinate φ diverges along the worldline of any particle that reaches \mathcal{H} . For geodesics, this divergence occurs at the rate $d\varphi/dt = \Omega_{\mathcal{H}} = 1/(2m)$ [Eqs. (11.74) and (13.23)], which is ensured at fixed Φ by the choice (13.59).

Let us denote by $G_{\alpha\beta}(T, R, \theta, \Phi, \varepsilon)$ the components of the metric tensor \mathbf{g} with respect to Bardeen-Horowitz coordinates $(X^\alpha) := (T, R, \theta, \Phi)$. They depend on ε since the change of coordinates (13.59)-(13.58) does. The explicit expressions of the $G_{\alpha\beta}(T, R, \theta, \Phi, \varepsilon)$'s are rather complicated and can be found in the notebook D.5.20. We are concerned here only by the limit $\varepsilon \rightarrow 0$. In other words, we consider the tensor field \mathbf{h} on \mathcal{M}_{BL} whose components with respect to the Bardeen-Horowitz coordinates are the limits

$$h_{\alpha\beta}(T, R, \theta, \Phi) := \lim_{\varepsilon \rightarrow 0} G_{\alpha\beta}(T, R, \theta, \Phi, \varepsilon). \quad (13.61)$$

⁷Cf. the historical note on page 534.

Evaluating the limits leads to the explicit expression of \mathbf{h} (cf. the notebook D.5.20):

$$\boxed{\mathbf{h} = m^2(1 + \cos^2 \theta) \left[-R^2 dT^2 + \frac{dR^2}{R^2} + d\theta^2 + \frac{4 \sin^2 \theta}{(1 + \cos^2 \theta)^2} (d\Phi + R dT)^2 \right].} \quad (13.62)$$

A priori the limits (13.61) lead only to a symmetric bilinear form. To be a proper metric tensor, \mathbf{h} has to be non-degenerate as well (cf. Sec. A.3.1 in Appendix A). This can be checked by computing the determinant of the components (13.62); one gets (cf. the notebook D.5.20):

$$\det(h_{\alpha\beta}) = -4m^2(1 + \cos^2 \theta)^2 \sin^2 \theta. \quad (13.63)$$

Hence $\det(h_{\alpha\beta})$ vanishes only for $\sin \theta = 0$, i.e. on the rotation axis \mathcal{A} . This merely signals the standard coordinate singularity of spherical-type coordinates on \mathcal{A} . Away from \mathcal{A} , one has $\det(h_{\alpha\beta}) < 0$, which shows that \mathbf{h} is non-degenerate and has the signature $(-, +, +, +)$ or $(+, -, -, -)$. Since the (diagonal) (R, θ) -block has clearly the signature $(+, +)$, we get that only $(-, +, +, +)$ is possible. Hence we may conclude:

Property 13.8: NHEK metric

Equation (13.62) defines a Lorentzian metric tensor \mathbf{h} on $\mathcal{M}_{\text{BL}} = \mathcal{M}_I \cup \mathcal{M}_{\text{III}}$. We shall call it the ***near-horizon extremal Kerr (NHEK) metric***.

The NHEK metric \mathbf{h} has been obtained in Bardeen-Horowitz coordinates (T, R, θ, Φ) . Let us express it in terms of Boyer-Lindquist coordinates (t, r, θ, φ) . From the transformation law (13.58), we have

$$R dT = \frac{r - m}{2m^2} dt, \quad \frac{dR}{R} = \frac{dr}{r - m} \quad \text{and} \quad d\Phi + R dT = d\varphi + \frac{r - 2m}{2m^2} dt. \quad (13.64)$$

The components of \mathbf{h} with respect to Boyer-Lindquist coordinates are then immediately deduced from Eq. (13.62):

$$\boxed{\mathbf{h} = m^2(1 + \cos^2 \theta) \left[-\frac{(r - m)^2}{4m^4} dt^2 + \frac{dr^2}{(r - m)^2} + d\theta^2 + \frac{4 \sin^2 \theta}{(1 + \cos^2 \theta)^2} \left(d\varphi + \frac{r - 2m}{2m^2} dt \right)^2 \right]} \quad (13.65)$$

Remark 1: It is remarkable that, while the change of coordinates $(t, r, \theta, \varphi) \leftrightarrow (T, R, \theta, \Phi)$ depends on the parameter ε [Eqs. (13.59)-(13.58)], the Boyer-Lindquist expression (13.65) of \mathbf{h} is independent of ε . This occurs because ε disappears in the combinations (13.64). In particular, the NHEK metric is unique, i.e. it does not depend on the value of ε , contrary to the Bardeen-Horowitz coordinates on \mathcal{M}_{BL} .

Remark 2: As we have introduced it, the NHEK metric is defined on the part \mathcal{M}_{BL} of the Kerr manifold \mathcal{M} , where $R \neq 0$. In Sec. 13.4.4, we shall consider instead that \mathbf{h} is defined on a manifold distinct from \mathcal{M} , where it can be extended across $R = 0$.

13.4.3 Anti-de Sitter features

It appears from (13.62) that the NEHK metric \mathbf{h} has striking similarities with the metric of anti-de Sitter spacetime (AdS): compare the dT^2 and dR^2 terms with the dt^2 and du^2 terms of the components (3.54) of the AdS₄ metric in Poincaré coordinates, presented in Example 18 of Chap. 3, noticing that both the Poincaré horizon in AdS₄ ($u = 0$) and the Kerr horizon ($R = 0$) are degenerate Killing horizons. More precisely, the metric induced by \mathbf{h} on the rotation axis \mathcal{A} ($\sin \theta = 0$), which is a 2-dimensional submanifold of \mathcal{M} , is exactly the metric of AdS₂ (the 2-dimensional anti-de Sitter spacetime), expressed in the Poincaré coordinates $(X^a) = (T, R)$:

$$\mathbf{h}^{\mathcal{A}} = 2m^2 \left(-R^2 dT^2 + \frac{dR^2}{R^2} \right). \quad (13.66)$$

Moreover, the metric induced by \mathbf{h} on the hypersurface $\theta = \theta_*$, with $\theta_* := \arcsin(\sqrt{3} - 1) \simeq 47.06^\circ$ (so that $4\sin^2 \theta_*/(1 + \cos^2 \theta_*) = 1$), is that of AdS₃:

$$\mathbf{h}^{\theta=\theta_*} = 2(\sqrt{3} - 1)m^2 \left[-R^2 dT^2 + \frac{dR^2}{R^2} + (d\Phi + R dT)^2 \right]. \quad (13.67)$$

Note however that the above expression is not that of AdS₃ metric in Poincaré coordinates. Indeed, by expanding the last term, we get

$$\mathbf{h}^{\theta=\theta_*} = 2(\sqrt{3} - 1)m^2 \left(2R dT d\Phi + \frac{dR^2}{R^2} + d\Phi^2 \right). \quad (13.68)$$

In particular, $h_{TT}^{\theta=\theta_*} = 0$, which implies that ∂_T is a null vector field, while it should be a timelike one for Poincaré coordinates. To prove that (13.67) is indeed the metric of AdS₃, let us introduce the so-called **global NHEK coordinates** (τ, y, θ, ψ) , which, for their (τ, y) part, are linked to (T, R) by the same relationship as that between global and Poincaré coordinates in AdS₂, namely

$$\begin{cases} T = \frac{\sqrt{1+y^2} \sin \tau}{y + \sqrt{1+y^2} \cos \tau} \\ R = y + \sqrt{1+y^2} \cos \tau \\ \theta = \theta \\ \Phi = \psi + \ln \left| \frac{\cos \tau + y \sin \tau}{1 + \sqrt{1+y^2} \sin \tau} \right| \end{cases} \iff \begin{cases} \tau = \arctan2(2TR^2, R^2 - T^2R^2 + 1) \\ + \pi H(-R) \\ y = \frac{(1+T^2)R^2 - 1}{2R} \\ \theta = \theta \\ \psi = \Phi - \ln \left(\frac{(1-TR)^2 + R^2}{\sqrt{[(1+T^2)R^2 - 1]^2 + 4R^2}} \right), \end{cases} \quad (13.69)$$

where H stands for the Heaviside function: $H(-R) = 1$ for $R < 0$ and 0 for $R > 0$. Beside the AdS₂-type transformation for $(\tau, y) \leftrightarrow (T, R)$, the new azimuthal coordinate ψ is introduced to insure $d\Phi + R dT = d\psi + y d\tau$. In terms of the coordinates (τ, y, θ, ψ) , the NHEK metric \mathbf{h} reads (cf. the notebook D.5.20):

$$\boxed{\mathbf{h} = m^2(1 + \cos^2 \theta) \left[-(1 + y^2) d\tau^2 + \frac{dy^2}{1 + y^2} + d\theta^2 + \frac{4 \sin^2 \theta}{(1 + \cos^2 \theta)^2} (d\psi + y d\tau)^2 \right]}. \quad (13.70)$$

On the rotation axis ($\sin \theta = 0$), we recover the AdS_2 metric expressed in global AdS coordinates (τ, y) . On the hypersurface $\theta = \theta_*$ considered above, we get

$$\mathbf{h}^{\theta=\theta_*} = 2(\sqrt{3} - 1)m^2 \left[-(1 + y^2)d\tau^2 + \frac{dy^2}{1 + y^2} + (d\psi + y d\tau)^2 \right]. \quad (13.71)$$

This is the metric of AdS_3 in some standard coordinates⁸ (compare e.g. with Eq. (17) of Ref. [56] with $y = \sinh \omega$). On constant θ hypersurfaces, with $\theta \neq \theta_*$ or $\pi - \theta_*$, the metric induced by \mathbf{h} is not that of AdS_3 but that of a so-called **warped AdS**₃, the θ -term in front of $(d\psi + y d\tau)^2$ in Eq. (13.70) being the *warp factor* (see e.g. Refs. [136, 56, 246]).

Remark 3: The value $\theta_* = \arcsin(\sqrt{3} - 1)$ for which the hypersurface $\theta = \theta_*$ inherits AdS_3 metric from \mathbf{h} is exactly the value θ_{crit} considered in Sec. 12.4.4 [cf. Eq. (12.98)], namely the minimal value of the colatitude θ of a distant observer for the critical curve (boundary of the black hole shadow) to contain a vertical straight line segment (the so-called *NHEK line*; maroon line in Fig. 12.24). A priori, this is a mere coincidence.

13.4.4 NHEK spacetime

The Bardeen-Horowitz components (13.62) of the NHEK metric \mathbf{h} are singular at $R = 0$. But this is a mere coordinate singularity, since, in terms of the global NHEK coordinates (τ, y, θ, ψ) , $R = 0$ is the locus $y + \sqrt{1 + y^2} \cos \tau = 0$ [cf. Eq. (13.69)] and the components (13.70) of \mathbf{h} are regular there. They are actually regular in all the range $\tau \in \mathbb{R}$, $y \in \mathbb{R}$, $\theta \in (0, \pi)$ and $\psi \in (0, 2\pi)$. As it is introduced in Eq. (13.69), the coordinate τ is actually restricted to $(-\pi, 2\pi)$ on \mathcal{M} . We may therefore define an “extended” spacetime as follows:

The **NHEK spacetime** is the spacetime $(\mathcal{N}, \mathbf{h})$, where the manifold \mathcal{N} is diffeomorphic to $\mathbb{R}^2 \times \mathbb{S}^2$ and \mathbf{h} is the Lorentzian metric defined by Eq. (13.70) in terms of the coordinates (τ, y, θ, ψ) , with $(\tau, y) \in \mathbb{R}^2$ and $(\theta, \psi) \in (0, \pi) \times (0, 2\pi)$ being standard spherical coordinates on \mathbb{S}^2 .

By comparing the metrics (13.70) and (3.49), with the changes of notation $y \leftrightarrow r$ and $\psi \leftrightarrow \varphi$, we note that $(\mathcal{N}, \mathbf{h})$ bears some similarities with AdS_4 , but is distinct from it. First of all, r ranges in $(0, +\infty)$ only on AdS_4 , while y ranges in⁹ $(-\infty, +\infty)$ on \mathcal{N} . Furthermore, there is a global $(1 + \cos^2 \theta)$ term in (13.70), the $d\theta^2$ terms differ drastically in the two expressions, the one in (13.70) being not multiplied by y^2 . The $d\varphi^2$ and $d\psi^2$ are also truly distinct. Moreover, the NHEK metric contains the off-diagonal term $h_{\tau\psi}$, while the AdS_4 one is purely diagonal.

⁸These coordinates are linked to the property of AdS_3 being a fiber bundle over AdS_2 with fiber \mathbb{R} (in the same way as \mathbb{S}^3 is a fiber bundle over \mathbb{S}^2 with fiber \mathbb{S}^1 (Hopf fibration)); ψ is then the fiber coordinate and spans \mathbb{R} . Since ψ is 2π -periodic on the hypersurface $\theta = \theta_*$, we see that the latter is actually not the whole of AdS_3 , but a quotient of it, obtained by identifying the points ψ and $\psi + 2\pi$.

⁹Note that there is no coordinate singularity at $y = 0$ in (13.70), while there is one at $r = 0$ in (3.49) and the latter is such that AdS_4 cannot be extended analytically to $r < 0$.

The inverse NHEK metric \mathbf{h}^{-1} takes a rather simple form (cf. the notebook D.5.21 for the computation):

$$h^{\alpha\beta} = \frac{1}{1 + \cos^2 \theta} \begin{pmatrix} -\frac{1}{1+y^2} & 0 & 0 & \frac{y}{1+y^2} \\ 0 & 1+y^2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \frac{y}{1+y^2} & 0 & 0 & \frac{(1+\cos^2 \theta)^2 + y^2(\cos^4 \theta + 6\cos^2 \theta - 3)}{4(1+y^2)\sin^2 \theta} \end{pmatrix}. \quad (13.72)$$

As AdS_4 , the NHEK spacetime is not asymptotically flat. The metrics of both spacetimes are solutions of the Einstein equation with $T = 0$ [Eq. (1.43)], but with $\Lambda < 0$ for AdS_4 metric and with $\Lambda = 0$ for \mathbf{h} :

Property 13.9: NHEK metric as a solution of the vacuum Einstein equation

The NHEK metric \mathbf{h} is a solution of the vacuum Einstein equation (1.44):

$$\text{Ric}(\mathbf{h}) = 0, \quad (13.73)$$

where $\text{Ric}(\mathbf{h})$ stands for the Ricci tensor of \mathbf{h} .

Proof. \mathbf{h} has been defined on \mathcal{M}_{BL} via the limit expression (13.61). Now, for $\varepsilon > 0$, the $G_{\alpha\beta}$'s in the right-hand side of Eq. (13.61) are the components of a metric that has a vanishing Ricci tensor, since it is nothing but the extremal Kerr metric, expressed in the Bardeen-Horowitz coordinates (T, R, θ, Φ) , which form a perfectly regular coordinate system on \mathcal{M}_{BL} for $\varepsilon > 0$ (cf. the transformations (13.59)-(13.58)). By continuity at the limit $\varepsilon \rightarrow 0$, it follows that \mathbf{h} has a vanishing Ricci tensor as well. A priori this holds only on \mathcal{M}_{BL} , but since the components of \mathbf{h} in the global NHEK coordinates (τ, y, θ, ψ) are identical on \mathcal{M}_{BL} and \mathcal{N} , the result is immediately extended by analyticity to all the NHEK spacetime. \square

Remark 4: For the reader skeptical about the limit process and the extension to the whole of \mathcal{N} , a direct computation of the Ricci tensor of \mathbf{h} from the global-coordinate components (13.70) is performed in the notebook D.5.21 and the outcome is indeed identically vanishing.

Total angular momentum

The quantity m^2 appears as a global scale factor of the metric \mathbf{h} in each of the three expressions (13.62), (13.70) and (13.65). This is of course the square of the mass of the extremal Kerr spacetime from which the NHEK spacetime $(\mathcal{N}, \mathbf{h})$ arises. However, one cannot associate intrinsically a mass to the latter. Indeed, one cannot use the Komar integral (5.36), as in the Kerr case, since the the NHEK spacetime lacks any global asymptotically timelike Killing vector¹⁰.

¹⁰In particular, the Killing vector ∂_τ is spacelike for θ close to $\pi/2$ and large enough values of y , since we read from (13.70) that $\mathbf{h}(\partial_\tau, \partial_\tau) = h_{\tau\tau} = m^2(1+\cos^2 \theta) \left[\frac{4y^2 \sin^2 \theta}{(1+\cos^2 \theta)^2} - 1 - y^2 \right] \sim \frac{m^2 y^2}{1+\cos^2 \theta} [4 \sin^2 \theta - (1 + \cos^2 \theta)^2]$ for $|y| \rightarrow +\infty$, so that $\mathbf{h}(\partial_\tau, \partial_\tau) > 0$ for $\sin \theta > (1 + \cos^2 \theta)/2$.

An alternative concept of mass, which does not require any Killing vector, is that of *ADM mass* (cf. Sec. 5.3.5). It is however not applicable either since $(\mathcal{N}, \mathbf{h})$ is not asymptotically flat. Incidentally, this last fact would anyway preclude any unambiguous definition of the Komar mass, since asymptotic flatness is required to normalize the timelike Killing field at spatial infinity (scalar square -1).

On the contrary, the **angular momentum** J of the NHEK spacetime can properly be defined from the axisymmetry Killing vector $\eta = \partial_\varphi = \partial_\Phi = \partial_\psi$, via the Komar formula (5.63). The Hodge dual $\star(d\eta)$ [cf. Eq. (5.38)], which appears in this formula, has of course to be taken with respect to the metric \mathbf{h} . Since \mathbf{h} fulfills the vacuum Einstein equation [Eq. (13.73)], the value of J does not depend upon the choice of the integration 2-surface \mathcal{S} in formula (5.63) (Property 5.21). One finds (cf. the notebook D.5.21 for the computation):

$$\boxed{J = m^2}. \quad (13.74)$$

This is exactly the value $J = am = m^2$ of the angular momentum of the extremal Kerr spacetime [Eq. (10.80)]. Hence we conclude:

Property 13.10: angular momentum of NHEK spacetime

The NHEK spacetime $(\mathcal{N}, \mathbf{h})$ has no well defined mass, but it has a well defined Komar angular momentum, whose value is nothing but the angular momentum $J = m^2$ of the extremal Kerr spacetime $(\mathcal{M}, \mathbf{g})$ from which $(\mathcal{N}, \mathbf{h})$ arises.

Remark 5: For this reason, many authors replace the overall factor m^2 of the NHEK metric in Eq. (13.62) or (13.70) by J .

“Conformal” coordinates

The range of the coordinate y is the whole of \mathbb{R} . For pictural purposes (e.g. drawing Carter-Penrose-like diagrams), it would be convenient to have instead a coordinate spanning some finite range. As in the AdS_4 case presented in Example 18 of Chap. 3, it is quite natural to introduce

$$\chi := \arctan y \iff y =: \tan \chi \quad (13.75)$$

so that¹¹ $\chi \in (-\pi/2, \pi/2)$. The relation (13.69) to the Bardeen-Horowitz coordinates (T, R, θ, Φ) can be then rewritten as

$$\left\{ \begin{array}{l} T = \frac{\sin \tau}{\cos \tau + \sin \chi} \\ R = \frac{\cos \tau + \sin \chi}{\cos \chi} \\ \theta = \theta \\ \Phi = \psi + \ln \left| \frac{\cos(\tau - \chi)}{\sin \tau + \cos \chi} \right| \end{array} \right. \iff \left\{ \begin{array}{l} \tau = \arctan 2(2TR^2, R^2 - T^2R^2 + 1) + \pi H(-R) \\ \chi = \arctan \left(\frac{(1+T^2)R^2 - 1}{2R} \right) \\ \theta = \theta \\ \psi = \Phi - \ln \left(\frac{(1-TR)^2 + R^2}{\sqrt{[(1+T^2)R^2 - 1]^2 + 4R^2}} \right), \end{array} \right. \quad (13.76)$$

¹¹In Example 18, the range of χ was only $(0, \pi/2)$ because $r = \tan \chi \geq 0$ in AdS_4 , while here $y = \tan \chi$ spans the whole of \mathbb{R} .

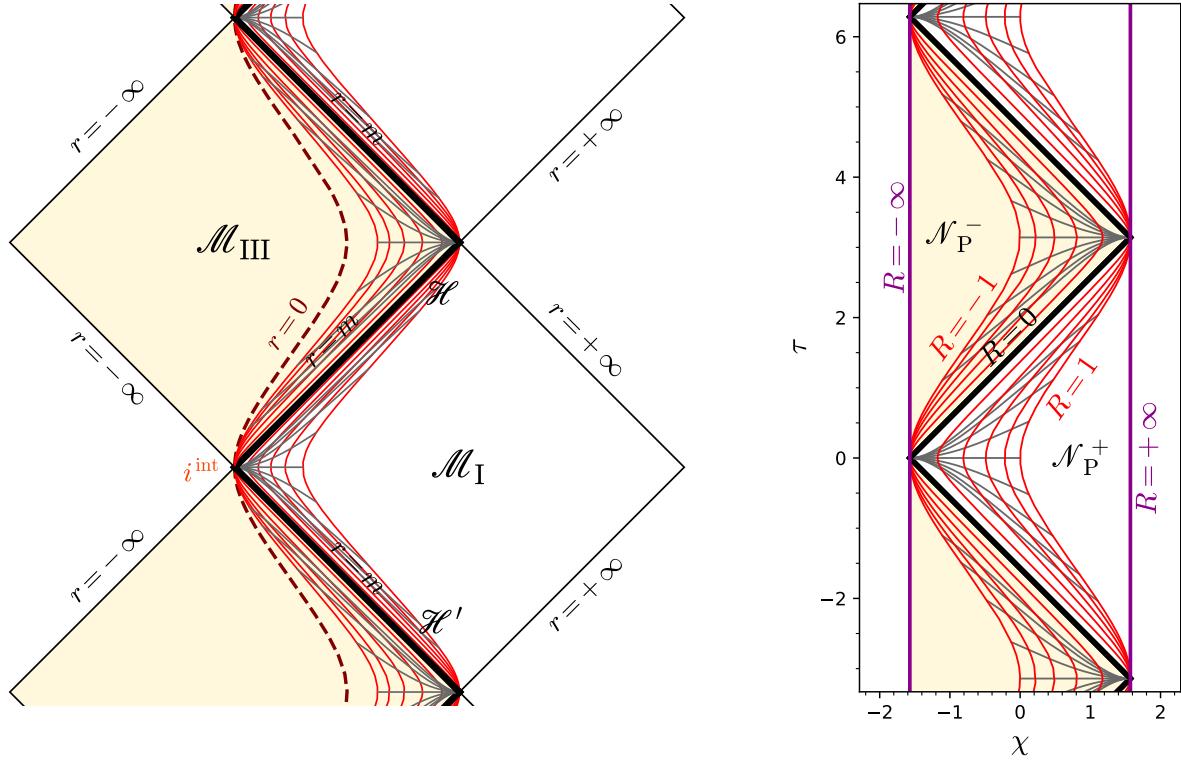


Figure 13.8: 2-dimensional views of the extremal Kerr spacetime (\mathcal{M}, g) (left) and of the NHEK spacetime (\mathcal{N}, h) (right). The left figure is a zoom on the $(\mathcal{M}_I^{(0)}, \mathcal{M}_{III}^{(0)})$ part of the Carter-Penrose diagram shown in Fig. 13.6. The right figure is a “compactified” view of \mathcal{N} based on the coordinates (τ, χ) . The near-horizon region in the Kerr spacetime is visualized by means of red curves, which are curves (actually hypersurfaces) of constant r , ranging from $r = m/2$ to $r = 3m/2$, i.e. within $\Delta r = m/2$ of the horizon value $r = m$. The r -increment between two nearby curves is $\delta r = m/10$. Grey curves represent hypersurfaces of constant Boyer-Lindquist coordinate t in the near horizon region, with t varying by steps of $\delta t = 2m$ in the range $[-8m, 8m]$. Each of these constant- r or constant- t curves is mapped to the NHEK spacetime (right figure) by means of the transformation (13.58) with $\varepsilon = 1/2$, giving birth to constant- R curves ranging from $R = -1$ to $R = 1$ (red curves) and to constant- T curves ranging from $T = -2$ to $T = 2$ (grey curves). [Figures generated by the notebooks D.5.18 (left) and D.5.21 (right)]

These transformations can be used to map the Boyer-Lindquist domain \mathcal{M}_{BL} of Kerr spacetime to the region $-\pi/2 < \tau + \chi < 3\pi/2$ of the NHEK spacetime \mathcal{N} , as illustrated in Fig. 13.8. This assumes a fixed value of $\varepsilon > 0$ in Eqs. (13.59)-(13.58), so that (T, R, θ, Φ) constitute a regular coordinate system on $\mathcal{M}_{BL} = \mathcal{M}_I \cup \mathcal{M}_{III}$. More precisely, \mathcal{M}_I , where $R > 0$, is mapped to the subregion \mathcal{N}_P^+ defined by $-\chi - \pi/2 < \tau < \chi + \pi/2$, which implies $\cos \tau + \sin \chi > 0$ and hence $R > 0$ via (13.76), since $\cos \chi > 0$ for $\chi \in (-\pi/2, \pi/2)$. On the other hand, \mathcal{M}_{III} , where $R < 0$, is mapped to the subregion \mathcal{N}_P^- defined by $\chi + \pi/2 < \tau < -\chi + 3\pi/2$, which implies $\cos \tau + \sin \chi < 0$ and hence $R < 0$ via (13.76). In both \mathcal{N}_P^+ and \mathcal{N}_P^- , the coordinate T spans the whole of \mathbb{R} . We shall call each of the regions \mathcal{N}_P^+ and \mathcal{N}_P^- a **Poincaré patch of NHEK spacetime**. They appear as the interiors of triangles having one vertical edge at the NHEK

boundary in Fig. 13.8. Note the similarity¹² with the Poincaré patch in AdS_4 spacetime plotted in Fig. 3.6.

The NHEK metric components with respect to the coordinates $(\tau, \chi, \theta, \psi)$ are easily deduced from (13.70):

$$\boxed{\mathbf{h} = m^2(1 + \cos^2 \theta) \left[\frac{1}{\cos^2 \chi} (-d\tau^2 + d\chi^2) + d\theta^2 + \frac{4 \sin^2 \theta}{(1 + \cos^2 \theta)^2} (d\psi + \tan \chi d\tau)^2 \right].} \quad (13.77)$$

Note the factor $1/\cos^2 \chi$ in front of the $-d\tau^2 + d\chi^2$ term, which is analogous to the conformal factor $1/\cos^2 \chi$ of the AdS_4 metric in Eq. (3.50), except that in the present case, $1/\cos^2 \chi$ does not factor all the metric, so that it cannot be considered as a proper conformal factor. Since $\cos \chi$ never vanishes for $\chi \in (-\pi/2, \pi/2)$, the factor $1/\cos^2 \chi$ is regular in all \mathcal{N} . It simply blows up at the “boundaries” $\chi \rightarrow \pm\pi/2$ of \mathcal{N} , which correspond to $y \rightarrow \pm\infty$ and to $R \rightarrow \pm\infty$ in the Poincaré patches \mathcal{N}_P^+ and \mathcal{N}_P^- .

One can provide a pictorial view of \mathbf{h} being the “near-horizon limit” of \mathbf{g} as follows:

The two red strips $r \in [m/2, 3m/2]$ and $R \in [-1, 1]$ in Fig. 13.8 have been drawn by setting $\varepsilon = 1/2$ in the transformation (13.58) linking the Boyer-Lindquist coordinates (t, r, θ, φ) to the Bardeen-Horowitz ones (T, R, θ, Φ) . Now, if one maintains the red strip $R \in [-1, 1]$ in the NHEK spacetime unchanged and let ε decays to 0, the red strip in the extremal Kerr spacetime shrinks to a more and more narrow strip around the horizon. As $\varepsilon \rightarrow 0$, this narrow strip, endowed with the Kerr metric \mathbf{g} , tends to be isometric to the fixed NHEK strip $R \in [-1, 1]$ endowed with the metric \mathbf{h} .

Remark 6: While the extremal Kerr spacetime and the NHEK one are similar in the near-horizon region, it is clear on Fig. 13.8 that they have very different asymptotics for $r \rightarrow \pm\infty$ and $R \rightarrow \pm\infty$.

13.4.5 NHEK symmetries

Given the similarities with the AdS spacetime, which is a maximally symmetric spacetime¹³, it should come as no surprise that the NHEK spacetime $(\mathcal{N}, \mathbf{h})$ has more symmetries than the extremal Kerr spacetime $(\mathcal{M}, \mathbf{g})$. First of all, it is obvious on the expression (13.62) of \mathbf{h} in Bardeen-Horowitz coordinates that T and Φ are ignorable coordinates, so that the two vector fields

$$\xi_1 := \partial_T = \frac{2m}{\varepsilon} \partial_t + \frac{1}{\varepsilon} \partial_\varphi \quad \text{and} \quad \eta = \partial_\Phi = \partial_\varphi = \partial_\psi \quad (13.78)$$

¹²In Fig. 3.6, the Poincaré patch (where $u > 0$) has its vertical edge on the left, while in Fig. 13.8, \mathcal{N}_P^+ (where $R > 0$) has its vertical edge on the right, but this results simply from a distinct convention in choosing χ .

¹³A spacetime of dimension n is said to be **maximally symmetric** if, and only if, its group of isometries has the maximum possible dimension, which is $n(n+1)/2$. It is equivalent to say that there are $n(n+1)/2$ linearly independent Killing vector fields when considering only linear combinations with constant coefficients. For $n = 4$, one has $n(n+1)/2 = 10$, which, among others, is the dimension of the Poincaré group (isometries of Minkowski spacetime). The group of isometries of AdS_n is the pseudo-orthogonal group $O(2, n-1)$, the dimension of which is exactly $n(n+1)/2$ (cf. e.g. Ref. [389]).

are two Killing vectors of \mathbf{h} . In the above equation, the second expression of ξ_1 is that in the Boyer-Lindquist coordinate frame and follows directly from the transformation law (13.59). In particular, note that ξ_1 is distinct from the Killing vector $\xi = \partial_t$ of the Kerr metric \mathbf{g} . It is rather connected to the horizon-generating Killing vector χ defined by Eq. (13.23):

$$\xi_1 = \frac{2m}{\varepsilon} \chi. \quad (13.79)$$

Besides, in Eq. (13.78), the vector field ∂_Φ is identical to the Killing vector field $\eta = \partial_\varphi$ of the Kerr metric \mathbf{g} as a consequence of the transformation law (13.59). Similarly, the transformation law (13.69) implies that $\eta = \partial_\psi$.

A third symmetry apparent on the metric components (13.62) is the invariance under the transformation $(T, R) \mapsto (\alpha T, R/\alpha)$, for any $\alpha > 0$. Such transformations, which are called **squeeze mappings** (or **hyperbolic rotations**) of the (T, R) plane, form a 1-parameter group action, whose generator ξ_2 is given by formula (3.14) with $t = \lambda := \alpha - 1$ (so that the identity element is for $\lambda = 0$). By considering an infinitesimal transformation of parameter $d\lambda$, formula (3.14) leads to the components $\xi_2^T = [(1 + d\lambda)T - T]/d\lambda = T$ and $\xi_2^R = [(1 - d\lambda)R - R]/d\lambda = -R$. Hence the Killing vector field

$$\xi_2 = T \partial_T - R \partial_R = t \partial_t + (m - r) \partial_r + \frac{t}{2m} \partial_\varphi. \quad (13.80)$$

The second expression stems from the transformation law (13.59) and expresses ξ_2 in terms of the Boyer-Lindquist coordinate frame. We note that this expression is independent of ε , contrary to the Boyer-Lindquist expression of ξ_1 .

Finally, a fourth Killing vector of \mathbf{h} shows up clearly on the components (13.70) with respect to the global NHEK coordinates (τ, y, θ, ψ) , since these components are independent of τ . Hence ∂_τ is a Killing vector of \mathbf{h} . In what follows, we will use instead the linear combination¹⁴

$$\xi_3 := \partial_\tau - \frac{1}{2} \xi_1, \quad (13.81)$$

since it has slightly simpler components with respect to Bardeen-Horowitz coordinates. Given the transformation law (13.69), we get (cf. the notebook D.5.20)

$$\xi_3 = \frac{1}{2} \left(T^2 + \frac{1}{R^2} \right) \partial_T - RT \partial_R - \frac{1}{R} \partial_\Phi. \quad (13.82)$$

The Boyer-Lindquist expression of ξ_3 is found via the transformation law (13.59)-(13.58):

$$\xi_3 = \varepsilon \left\{ \left[\frac{t^2}{4m} + \frac{m^3}{(r-m)^2} \right] \partial_t + \frac{t(m-r)}{2m} \partial_r + \frac{1}{2} \left[\frac{t^2}{4m^2} + \frac{m(3m-2r)}{(r-m)^2} \right] \partial_\varphi \right\}. \quad (13.83)$$

We have thus four Killing vectors of the NEHK metric \mathbf{h} : η , ξ_1 , ξ_2 and ξ_3 . Since none of them is a linear combination with constant coefficients of the three others, the isometry group G generated by these Killing vectors is 4-dimensional. This is 2 dimensions more than the

¹⁴Let us recall that a linear combination with constant coefficients of Killing vectors is a Killing vector.

isometry group of Kerr spacetime, which is $\mathbb{R} \times \text{U}(1)$. In order to fully characterize G , let us determine its Lie algebra by evaluating the commutator¹⁵ of each pair of the four generators η , ξ_1 , ξ_2 and ξ_3 . We get (cf. the notebook D.5.20)

$$[\eta, \xi_1] = 0, \quad [\eta, \xi_2] = 0, \quad [\eta, \xi_3] = 0 \quad (13.84)$$

and

$$[\xi_2, \xi_1] = -\xi_1, \quad [\xi_2, \xi_3] = \xi_3, \quad [\xi_1, \xi_3] = \xi_2. \quad (13.85)$$

Equation (13.84) shows that η commutes with all the other generators of G . Since η is the generator of the axisymmetry group $\text{U}(1)$, we may write $G = G_3 \times \text{U}(1)$, where G_3 is a 3-dimensional Lie group, generated by ξ_1 , ξ_2 and ξ_3 . The commutation relations (13.85) show¹⁶ that the Lie algebra of G_3 is the special linear algebra $\mathfrak{sl}(2, \mathbb{R})$. At this stage, G_3 can be one of the following three Lie groups: $\text{SL}(2, \mathbb{R})$, $\text{PSL}(2, \mathbb{R}) = \text{SL}(2, \mathbb{R})/\mathbb{Z}_2$ or $\overline{\text{SL}}(2, \mathbb{R})$ (the universal covering group of $\text{SL}(2, \mathbb{R})$). It cannot be $\text{PSL}(2, \mathbb{R})$ because, as it appears clearly on (13.62), the transformation $(T, R) \mapsto (-T, -R)$ is an element of G_3 and, in $\text{PSL}(2, \mathbb{R})$, this element would be identified with the identity due to the quotient by $\mathbb{Z}_2 = \{\text{Id}, -\text{Id}\}$, where Id is the identity element. G_3 is actually the special linear group $\text{SL}(2, \mathbb{R})$. We conclude:

Property 13.11: isometry group of NEHK spacetime

The isometry group of the NEHK spacetime is

$$G = \boxed{\text{SL}(2, \mathbb{R}) \times \text{U}(1)}, \quad (13.86)$$

with $\text{U}(1)$ generated by the axisymmetry Killing vector η and $\text{SL}(2, \mathbb{R})$ generated by the three Killing vectors ξ_1 , ξ_2 and ξ_3 given by Eqs. (13.78), (13.80) and (13.82).

This property is often phrased as follows: the near-horizon region of the extremal Kerr spacetime (the extremal Kerr throat) has some *emergent symmetries*, i.e. it has more symmetries than the extremal Kerr spacetime itself. Let us stress that no such thing occurs for non-extremal Kerr black holes.

Remark 7: $\text{SL}(2, \mathbb{R})$ is isomorphic to the spin group $\text{Spin}(2, 1)$, which is the isometry group of the 2-dimensional anti-de Sitter spacetime (AdS_2). The spin group $\text{Spin}(2, 1)$ is the double cover of the pseudo-orthogonal group $\text{SO}(2, 1)$, the latter being the isometry group of the “time-cyclic” 2-dimensional anti-de Sitter spacetime, i.e. the quadric surface $X^2 - U^2 - V^2 = -1$ in \mathbb{R}^3 endowed with the flat metric $-\text{d}U^2 - \text{d}V^2 + \text{d}X^2$.

Remark 8: Having an enhanced symmetry group near the horizon is not specific to the extremal Kerr black hole; this is actually a generic feature, which occurs near *degenerate* Killing horizons [326, 136].

¹⁵The definition of the commutator of two vector fields is recalled in Appendix A [cf. Eq. (A.27)].

¹⁶More precisely, in the representation of $\mathfrak{sl}(2, \mathbb{R})$ by the algebra of 2×2 real matrices with zero trace, a standard basis is $E = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, $F = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $H = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. This basis has the following commutation relations: $[H, E] = 2E$, $[H, F] = -2F$ and $[E, F] = H$. One recovers the last ones from (13.85) by identifying ξ_1 with $F/\sqrt{2}$, ξ_2 with $H/2$ and ξ_3 with $-E/\sqrt{2}$.

Pure geometric arguments suffice to show that a degenerate Killing horizon associated with some Killing vector ξ_1 has necessarily a near-horizon geometry endowed with a second Killing vector, which is similar to ξ_2 [326]. Axisymmetry provides the third Killing vector η and the Einstein equation leads to the existence of the fourth Killing vector ξ_3 [328].

Historical note : The NHEK metric h has been first exhibited by Brandon Carter in 1973 [94]: compare Eq. (5.63) of Ref. [94] (with $Q = P = 0$) with Eq. (13.62), taking into account the change of notations $(\tau, \lambda, \varphi) \leftrightarrow (T, -R, \Phi)$. Actually the NHEK metric was found by Carter when searching for the most general stationary and axisymmetric solutions of the source-free Einstein-Maxwell equations (cf. Sec. 1.5.2), in order to get the Kerr-Newman solution. The latter appeared then as the generic case and NHEK-like metrics as special cases. The limit $Q = P = 0$ of Carter's metric (5.63) corresponds to the case of vanishing electric charge Q and magnetic monopole P , so that it is a solution of the vacuum Einstein equation [Eq. (13.73) above]. Carter stated that the metric h has a 4-dimensional isometry group and that the hypersurfaces of constant θ are homogeneous and partially isotropic. Even if he did not get h by some near-horizon limit process, Carter stressed that in the neighborhood of the extremal Kerr horizon, the Kerr metric g can be approximated arbitrarily closely by h (cf. Fig. 6.3 in Ref. [94], which is similar to our Fig. 13.8). In the related case of the *static* extremal black hole, constituted by the (spherically symmetric) Reissner-Nordström solution with electric charge equal to the mass ($Q = \sqrt{4\pi/\mu_0} m$, cf. Eq. (5.107)), Carter exhibited a change of coordinates similar¹⁷ to that between Boyer-Lindquist coordinates and Bardeen-Horowitz coordinates [Eq. (13.59)] and he showed that the near-horizon limit brings the Reissner-Nordström metric to the *Bertotti-Robinson metric*. The latter is the product metric of $\text{AdS}_2 \times \mathbb{S}^2$ and constitutes a static solution of the source-free Einstein-Maxwell equations that is highly symmetric, having the 6-dimensional group $\text{SL}(2, \mathbb{R}) \times \text{SO}(3)$ as isometry group (cf. e.g. Ref. [212]). The NHEK metric has been rediscovered by James Bardeen and Gary Horowitz in 1999 [40]. They derived it from the Boyer-Lindquist expression of the extremal Kerr metric by performing the change of coordinates¹⁸ (13.58)-(13.59) and taking the limit $\varepsilon \rightarrow 0$, as presented in Sec. 13.4.2. In the same article [40], Bardeen and Horowitz introduced the global NHEK coordinates¹⁹ (τ, y, θ, ψ) via the transformation (13.69) and obtained expression (13.70) for h [their Eq. (2.9)]. They have analyzed the global properties of the NHEK spacetime, in particular its geodesic completeness. Bardeen and Horowitz have also shown that the NHEK isometry group is $\text{SL}(2, \mathbb{R}) \times \text{U}(1)$ [Eq. (13.86)] and have given formulas (13.78), (13.80) and (13.82) for the four Killing vectors ξ_1, ξ_2, ξ_3 and η [their Eq. (2.14)].

13.4.6 Applications

The NHEK geometry is at the core of the so-called *Kerr/CFT correspondence*, initiated by Monica Guica, Thomas Hartman, Wei Song and Andrew Strominger in 2009 [246], which conjectures that quantum gravity in the vicinity of the extremal Kerr horizon is dual to a two-dimensional conformal field theory (CFT) (see Ref. [136] for a review of this topic).

¹⁷Carter defined the near-horizon coordinates (τ, λ) by $t = m^2\tau$ and $r = \lambda + m$ and took the near-horizon limit as $m \rightarrow +\infty$ (cf. the paragraph surrounding Eq. (4.31) in Ref. [94]); this is equivalent to Eq. (13.59) with $\varepsilon = 1/m$.

¹⁸The coordinates actually used by Bardeen and Horowitz were $\tilde{T} := 2mT$ and $\tilde{R} = mR$, denoted respectively t and r by them.

¹⁹ ψ is denoted by ϕ in Ref. [40].

Another domain of application of the NHEK geometry regards the computation of signals emanating from the neighborhood of an extremal Kerr black hole, either electromagnetic ones (e.g. [237, 138, 307]) or gravitational-wave ones (e.g. [139, 238]).

13.5 Going further

Apart from the principal null geodesics, we have not discussed the geodesics of the extremal Kerr black hole in this chapter. Many of their properties can be obtained by taking the limit $a \rightarrow m$ of the Kerr geodesics discussed in Chaps. 11 and 12. The specific case of critical null geodesics, delineating the black hole shadow, has been treated in Sec. 12.4.4 for $a = m$. For an extensive study of geodesics close to the horizon, making use of the NHEK geometry, see Refs. [307, 138].

We have not discussed the stability of extremal Kerr black holes. Contrary to what happens for subextremal Kerr black holes ($a < m$), Stefanos Aretakis has shown in 2012²⁰ that axisymmetric scalar perturbations (solutions of the wave equation $\square_g \Psi = 0$) have transverse second derivatives²¹ that blow up when $\tilde{t} \rightarrow +\infty$ along the event horizon \mathcal{H} of an extremal Kerr black hole, while Ψ itself decays at least like $\tilde{t}^{-1/2}$ and its transverse first derivative remains bounded [18] (see also the monograph [19] and Theorem A in Ref. [206]). In 2016, Marc Casals, Samuel Gralla and Peter Zimmerman [105] have shown that *non-axisymmetric* scalar perturbations with initial data supported away from \mathcal{H} are unstable along \mathcal{H} , with a transverse *first* derivative growing like $\tilde{t}^{1/2}$, which is a higher rate than the axisymmetric perturbations considered by Aretakis. This has been recently confirmed by a rigorous mathematical analysis by Dejan Gajic [206]. Gralla and Zimmerman have shown in 2018 [239] that this instability extends to electromagnetic and gravitational perturbations and can be considered as a critical phenomenon associated with the emergent symmetry generated by the NHEK Killing vector ξ_2 (squeeze mappings). Finally let us point out that Aretakis [17] has also shown that solutions of the non-linear wave equation $\square_g \Psi = N(\Psi, \nabla \Psi)$, where N is a non-linear function, blow up in a finite time along \mathcal{H} .

²⁰The work [18] appeared on arXiv in 2012, but was published in 2015.

²¹By *transverse derivative* it is meant $\partial \Psi / \partial r$, $\partial^2 \Psi / \partial r^2$, etc. (recall that \mathcal{H} is located at $r = \text{const}$ and is spanned by the coordinates $(\tilde{t}, \theta, \tilde{\varphi})$). If the initial data has support away from \mathcal{H} , the divergence occurs only for the third derivative.

Part IV

Dynamical black holes

Chapter 14

Black hole formation 1: dust collapse

Contents

14.1 Introduction	539
14.2 Lemaître-Tolman equations and their solutions	540
14.3 Oppenheimer-Snyder collapse	548
14.4 Observing the black hole formation	566
14.5 Going further	580

14.1 Introduction

After having investigated black holes in equilibrium, through the Schwarzschild and Kerr solutions, we move to dynamical black holes, 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 modeling the gravitational collapse of a stellar core or a gas cloud. A drawback of this approach is that it forbids the study of gravitational waves, since by virtue of the Jebsen-Birkhoff theorem (to be proven in Sec. 14.2.5), the exterior of any spherically symmetric collapsing object is a piece of Schwarzschild spacetime, which does not contain any gravitational radiation. The second major simplifying approximation is to consider *pressureless matter*, commonly referred to as *dust*.

We start by deriving from the Einstein equation the system of partial differential equations yielding the metric for spherically symmetric dust (Lemaître-Tolman system) and write down the most general solutions (Sec. 14.2). Then we specialize these solutions to the case of a homogeneous ball of dust collapsing from rest (Oppenheimer-Snyder collapse, Sec. 14.3), which we study in details, in particular regarding the birth of the event horizon and the existence of trapped surfaces. Finally, we consider the observational appearance of the gravitational collapse and the black hole formation to a remote observer (Sec. 14.4).

14.2 Lemaître-Tolman equations and their solutions

14.2.1 Hypotheses

As mentioned in the Introduction, we shall restrict ourselves to spherically symmetric¹ space-times, and for concreteness, to 4-dimensional ones. The most general spherically symmetric 4-dimensional spacetime $(\mathcal{M}, \mathbf{g})$ can be described in terms of coordinates $(x^\alpha) = (\tau, \chi, \theta, \varphi)$ such that the metric tensor writes

$$\mathbf{g} = -d\tau^2 + a(\tau, \chi)^2 d\chi^2 + r(\tau, \chi)^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (14.1)$$

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 simplicity, we consider only pressureless matter, in the form of a perfect fluid with zero pressure. The matter energy-momentum tensor is then

$$\mathbf{T} = \rho \underline{\mathbf{u}} \otimes \underline{\mathbf{u}}, \quad (14.2)$$

where $\underline{\mathbf{u}}$ is the 1-form metric-dual to the fluid 4-velocity \mathbf{u} , i.e. the 1-form of components $u_\alpha = g_{\alpha\mu} u^\mu$ [cf. Eq. (A.45)], and ρ is a scalar field that can be interpreted as the fluid energy density measured in the fluid frame, by virtue of the identity $\rho = \mathbf{T}(\underline{\mathbf{u}}, \underline{\mathbf{u}})$, which follows from $\langle \underline{\mathbf{u}}, \underline{\mathbf{u}} \rangle = \mathbf{g}(\mathbf{u}, \mathbf{u}) = -1$. Let us recall that the energy-momentum tensor of a generic **perfect fluid** is $\mathbf{T} = (\rho + p)\underline{\mathbf{u}} \otimes \underline{\mathbf{u}} + p\mathbf{g}$, where p is the fluid pressure. 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 any fluid particle stays at fixed values of (χ, θ, φ) . Because the 4-velocity obeys $u^\alpha = dx^\alpha/d\tau_f$, where τ_f is the fluid proper time [cf. Eq. (1.16)], this amounts to $u^\chi = u^\theta = u^\varphi = 0$, hence

$$\mathbf{u} = \partial_\tau. \quad (14.3)$$

A priori, one should only have $\mathbf{u} = u^\tau \partial_\tau$, but the synchronous coordinate condition $g_{\tau\tau} = -1$ along with the normalization $\mathbf{g}(\mathbf{u}, \mathbf{u}) = -1$ implies $u^\tau = 1$. Since $u^\tau = d\tau/d\tau_f$, we get $\tau = \tau_f$ (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 $\nabla \cdot \vec{\mathbf{T}} = 0$ [Eq. (1.45)], which follows from the Einstein equation (1.40) and the contracted Bianchi identity (A.110) (cf. Sec. 1.5), implies:

¹See Sec. 6.2.2 for the precise definition of *spherically symmetric*.

Property 14.1: geodesic flow

The worldlines of the fluid particles of the pressureless matter model (14.2) are timelike geodesics of $(\mathcal{M}, \mathbf{g})$.

Proof. If one plugs the energy-momentum tensor (14.2) in the energy-momentum conservation law (14.5), one obtains $\nabla_\mu(\rho u^\mu u^\alpha) = 0$, i.e.

$$\nabla_\mu(\rho u^\mu)u^\alpha + \rho u^\mu \nabla_\mu u^\alpha = 0. \quad (14.4)$$

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 \mathbf{u} [Eq. (1.17)]: $\mathbf{u} \cdot \nabla_{\mathbf{u}} \mathbf{u} = 0$. In particular, \mathbf{u} is a timelike vector, while the 4-acceleration $\nabla_{\mathbf{u}} \mathbf{u}$ is a spacelike one. Thus the only way for Eq. (14.4) to hold is that each term in the left-hand side vanishes separately:

$$\nabla_\mu(\rho u^\mu) = 0 \quad \text{and} \quad u^\mu \nabla_\mu u^\alpha = 0.$$

The second equation is nothing but the geodesic equation [Eq. (B.1)] for the field lines of \mathbf{u} , i.e. the fluid worldlines. \square

Each fluid particle is thus in free-fall and moves independently of its neighbors, 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 (1.40) 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 the notebook D.6.1, if one disregards the peculiar case² $\partial r / \partial \chi = 0$, the $\tau\chi$ component yields

$$a(\tau, \chi) = \frac{1}{f(\chi)} \frac{\partial r}{\partial \chi}, \quad (14.5)$$

where $f(\chi)$ is an arbitrary function of χ . Accordingly, we may rewrite the metric (14.1) as

$$\mathbf{g} = -d\tau^2 + \frac{1}{f(\chi)^2} \left(\frac{\partial r}{\partial \chi} \right)^2 d\chi^2 + r(\tau, \chi)^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (14.6)$$

Property 14.2: Lemaître-Tolman system

The metric (14.6) is a solution of the Einstein equation (1.40) with the energy-momentum

²For $\Lambda = 0$, this case leads to Datt solution [156].

tensor (14.2)-(14.3) if, and only if,

$$\left(\frac{\partial r}{\partial \tau}\right)^2 = f(\chi)^2 - 1 + \frac{2M(\chi)}{r(\tau, \chi)} + \frac{\Lambda}{3}r(\tau, \chi)^2 \quad (14.7a)$$

$$\frac{dM}{d\chi} = 4\pi r(\tau, \chi)^2 \rho(\tau, \chi) \frac{\partial r}{\partial \chi}, \quad (14.7b)$$

where $M(\chi)$ is an arbitrary function of χ .

Proof. Taking into account (14.5), the $\chi\chi$ and $\tau\tau$ components of the Einstein equation yield respectively to (14.7a) and (14.7b), cf. the notebook D.6.1. There is no other independent component of the Einstein equation. \square

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 [369], although it has been introduced by Lemaître [345] more than 30 years earlier. This quantity is invariantly defined for any spherically symmetric spacetime from the areal radius r :

$$M := \frac{r}{2} \left(1 - \nabla_\mu r \nabla^\mu r - \frac{\Lambda}{3} r^2 \right). \quad (14.8)$$

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 for $M = M(\chi)$.

Historical note : The Lemaître-Tolman system (14.7) has been first derived in 1932 by Georges Lemaître [345]: Eqs. (14.6), (14.7a) and (14.7b) are respectively Eqs. (8.1), (8.2) and (8.3) of Ref. [345], up to some slight change of notations. The system (14.7) however became known as *Tolman model* or *Tolman-Bondi model*, in reference to posterior works by Richard Tolman (1934) [481] and by Hermann Bondi (1947) [65]. This happened despite Tolman fully acknowledging Lemaître's work [345] in his article [481] (Tolman actually met Lemaître in 1932-33 during the latter's trip to United States [181]) and Bondi [65] 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 [181] for a detailed historical study of Lemaître's paper [345] (see also Krasiński's note [322]). We follow the suggestion of Plebański & Krasiński [413] to call the system (14.7) *Lemaître-Tolman*, and not merely *Lemaître*, in order to distinguish 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 then be rewritten as

$$\frac{1}{2}\dot{r}^2 - \frac{M(\chi)}{r} = E(\chi), \quad (14.9)$$

where $\dot{r} := \partial r / \partial \tau$ and

$$E(\chi) := \frac{f(\chi)^2 - 1}{2}. \quad (14.10)$$

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

$$\begin{cases} \tau = \frac{M(\chi)}{(2E(\chi))^{3/2}} (\sinh \eta - \eta) + \tau_0(\chi) \\ r(\tau, \chi) = \frac{M(\chi)}{2E(\chi)} (\cosh \eta - 1) \end{cases} \quad (14.11)$$

- if $E(\chi) = 0$, the solution is

$$r(\tau, \chi) = \left(\frac{9M(\chi)}{2} (\tau - \tau_0(\chi))^2 \right)^{1/3} \quad (14.12)$$

- if $E(\chi) < 0$, the solution is given in parameterized form (parameter η) by

$$\begin{cases} \tau = \frac{M(\chi)}{|2E(\chi)|^{3/2}} (\eta + \sin \eta) + \tau_0(\chi) \\ r(\tau, \chi) = \frac{M(\chi)}{|2E(\chi)|} (1 + \cos \eta) \end{cases} \quad (14.13)$$

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).

We may summarize the above results as follows:

Property 14.3: solution for spherical dust collapse

The procedure to get a full solution of spherical dust collapse with $\Lambda = 0$ is

1. choose arbitrary functions $f(\chi)$, $M(\chi)$ and $\tau_0(\chi)$;
2. evaluate $E(\chi)$ via (14.10);

3. depending on the value of $E(\chi)$, use (14.11), (14.12) or (14.13) to get the solution for $r(\tau, \chi)$;
4. 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 Vacuum solution and the Jebsen-Birkhoff theorem

Let us search for a solution in the vacuum case, i.e. $\rho = 0$. Equation (14.7b) with $\rho = 0$ implies that $M(\chi)$ is a constant, which we shall denote by m . Equations (14.9)-(14.10) then yield

$$\frac{\partial r}{\partial \tau} = \varepsilon \sqrt{f(\chi)^2 - 1 + \frac{2m}{r}}, \quad \text{with } \varepsilon = \pm 1. \quad (14.14)$$

Hence

$$dr = \varepsilon \sqrt{f(\chi)^2 - 1 + \frac{2m}{r}} d\tau + \frac{\partial r}{\partial \chi} d\chi.$$

Using this identity to substitute $(\partial r / \partial \chi) d\chi$ in the metric expression (14.6), we get

$$\begin{aligned} g &= \frac{1}{f(\chi)^2} \left[-\left(1 - \frac{2m}{r}\right) d\tau^2 - 2\varepsilon \sqrt{f(\chi)^2 - 1 + \frac{2m}{r}} d\tau dr + dr^2 \right] \\ &\quad + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \end{aligned} \quad (14.15)$$

which can be rearranged as

$$\begin{aligned} g &= -\left(1 - \frac{2m}{r}\right) \left(\frac{1}{f(\chi)} d\tau + \varepsilon \frac{\sqrt{f(\chi)^2 - 1 + 2m/r}}{f(\chi)(1 - 2m/r)} dr \right)^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 \\ &\quad + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \end{aligned} \quad (14.16)$$

If there exists a scalar function $t = t(\tau, r)$ such that

$$dt = \frac{1}{f(\chi)} d\tau + \varepsilon \frac{\sqrt{f(\chi)^2 - 1 + 2m/r}}{f(\chi)(1 - 2m/r)} dr, \quad (14.17)$$

then Eq. (14.16) would be nothing but Schwarzschild metric expressed in Schwarzschild-Droste coordinates (t, r, θ, φ) [cf. Eq. (6.14)]. The necessary and sufficient condition for (14.17) to hold is

$$\frac{\partial}{\partial r} \left(\frac{1}{f(\chi)} \right) \Big|_{\tau} = \frac{\partial}{\partial \tau} \left(\varepsilon \frac{\sqrt{f(\chi)^2 - 1 + 2m/r}}{f(\chi)(1 - 2m/r)} \right) \Big|_r,$$

this relation expressing the identity $\partial^2 t / \partial r \partial \tau = \partial^2 t / \partial \tau \partial r$. In it, χ has to be considered as a function of (τ, r) , which is given implicitly by one of the three expressions (14.11)-(14.13) with $M(\chi) = m$. Accordingly, the above condition becomes

$$-\frac{f'(\chi)}{f(\chi)^2} \frac{\partial \chi}{\partial r} \Big|_{\tau} = \frac{\varepsilon}{\sqrt{f(\chi)^2 - 1 + 2m/r}} \frac{f'(\chi)}{f(\chi)^2} \frac{\partial \chi}{\partial \tau} \Big|_r.$$

Simplifying and invoking (14.14), we arrive at a condition independent of $f(\chi)$:

$$\frac{\partial \chi}{\partial \tau} \Big|_r + \frac{\partial \chi}{\partial r} \Big|_\tau \frac{\partial r}{\partial \tau} \Big|_\chi = 0.$$

Now this condition is always fulfilled: by the chain rule, the left-hand side is nothing but the partial derivative $\frac{\partial \chi}{\partial \tau} \Big|_\chi$, which is trivially zero if one considers χ as a function of (τ, χ) . We conclude that there exists a function t obeying (14.17), which proves that g is Schwarzschild metric, whatever the choice of the function $f(\chi)$. Since our starting point was the most general metric for a (possibly non-stationary) spherically symmetric spacetime [Eq. (14.1)], we have proven a famous theorem of general relativity:

Property 14.4: Jebsen-Birkhoff theorem

In vacuum, the unique spherically symmetric solution of the 4-dimensional Einstein equation with $\Lambda = 0$ [Eq. (1.44)] is the 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 it may oscillate radially, keeping its spherical symmetry). In particular, there are no gravitational waves in spherical symmetry.

Remark 1: The Jebsen-Birkhoff theorem can be viewed as a generalization of the shell theorem version of Gauss's law in Newtonian gravity: the gravitational field outside any spherical source is entirely determined by the total mass m of the source, being identical to that generated by a point of mass m located at the symmetry center.

Historical note : The Jebsen-Birkhoff theorem has been first proved by the Norwegian physicist Jørg Tofte Jebsen in 1920 [299]. It was independently established by the American mathematician George D. Birkhoff in 1923 [61] and became famous under the name *Birkhoff's theorem*, the work of Jebsen remaining largely unknown until recently (see Ref. [302] for details).

14.2.6 Schwarzschild solution in Lemaître coordinates

The above derivation of the Jebsen-Birkhoff theorem offers a way to express the Schwarzschild metric in new sets of coordinates, not considered in Part II, namely Lemaître and Painlevé-Gullstrand coordinates. Since we have all freedom on the function $f(\chi)$, we will focuss on a subclass of Lemaître coordinates for which $f(\chi) = 1$. Then $E(\chi) = 0$ and $r(\tau, \chi)$ is given by Eq. (14.12). Since $M(\chi) = m$ is constant, we cannot choose $\tau_0(\chi)$ in (14.12) to be a constant, otherwise we would have $\partial r / \partial \chi = 0$ and the metric (14.6) would be degenerate. The simplest non-constant choice is $\tau_0(\chi) = \chi$. Equation (14.12) yields then

$$r(\tau, \chi) = \left(\frac{9m}{2} \right)^{1/3} (\chi - \tau)^{2/3}. \quad (14.18)$$

In what follows, we assume $\chi \geq \tau$. Then

$$\chi - \tau = \frac{1}{3} \sqrt{\frac{2}{m}} r^{3/2} \quad (14.19)$$

and

$$\frac{\partial r}{\partial \chi} = \left(\frac{4m}{3} \right)^{1/3} (\chi - \tau)^{-1/3} = \sqrt{\frac{2m}{r}}. \quad (14.20)$$

Accordingly, Eq. (14.6) yields

$$\mathbf{g} = -\mathbf{d}\tau^2 + \frac{2m}{r(\tau, \chi)} \mathbf{d}\chi^2 + r(\tau, \chi)^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2), \quad (14.21)$$

where $r(\tau, \chi)$ is given by Eq. (14.18). This is the expression of the Schwarzschild metric in terms of the Lemaître coordinates $(\tau, \chi, \theta, \varphi)$.

The relation between Lemaître coordinates and Schwarzschild-Droste ones is obtained by setting $f(\chi) = 1$ and³ $\varepsilon = -1$ in Eq. (14.17):

$$\mathbf{d}\tau = \mathbf{d}t + \frac{\sqrt{\frac{2m}{r}}}{1 - \frac{2m}{r}} \mathbf{d}r. \quad (14.22)$$

By integration, we get

$$\tau = t + 4m \sqrt{\frac{r}{2m}} + 2m \ln \left(\frac{\sqrt{r/2m} - 1}{\sqrt{r/2m} + 1} \right) + \text{const.} \quad (14.23)$$

The expression of χ in terms of (t, r) is then deduced from Eq. (14.19):

$$\chi = t + 4m \sqrt{\frac{r}{2m}} \left(1 + \frac{r}{6m} \right) + 2m \ln \left(\frac{\sqrt{r/2m} - 1}{\sqrt{r/2m} + 1} \right) + \text{const.} \quad (14.24)$$

We deduce easily from these formulas the expression of the stationarity Killing vector ξ of Schwarzschild spacetime in terms of the Lemaître coordinates. Since $\xi = \partial_t$ [Eq. (6.6)], and the above formulas imply $\partial\tau/\partial t = 1$ and $\partial\chi/\partial t = 1$, we get, applying the chain rule $\partial/\partial t = \partial/\partial\tau \times \partial\tau/\partial t + \partial/\partial\chi \times \partial\chi/\partial t$,

$$\xi = \partial_\tau + \partial_\chi. \quad (14.25)$$

Remark 2: Although very simple, the above relation shows that Lemaître coordinates are *not* adapted to the spacetime symmetry generated by the Killing vector ξ : the latter does not coincide with any coordinate vector ∂_α of Lemaître coordinates. This reflects the fact that the metric components (14.21) depend on τ (via the function $r(\tau, \chi)$), in addition to χ .

³ $\varepsilon = -1$ follows from Eqs. (14.14) and (14.18).

The vacuum hypothesis implies that we can no longer interpret Lemaître coordinates as comoving with some free-falling dust, as in Sec. 14.2.1. However, the geodesic character of these coordinates remains. Indeed, the vector $\mathbf{u} := \partial_\tau$ is geodesic with respect to the metric (14.1): $\nabla_u u = 0$, as a direct calculation shows (cf. the notebook D.6.1). This implies that the curves $(\chi, \theta, \varphi) = \text{const}$ are timelike geodesics. Moreover, the conserved energy per unit mass along these geodesics is (cf. Sec. B.5)

$$\varepsilon = -\xi \cdot \mathbf{u} = -(\partial_\tau + \partial_\chi) \cdot \partial_\tau = -\underbrace{g_{\tau\tau}}_{-1} - \underbrace{g_{\chi\tau}}_0 = 1,$$

where use has been made of Eq. (14.25). $\varepsilon = 1$ means that the geodesics are marginally bound: they describe the free fall from rest at infinity.

As it is clear on the metric components (14.21), a key feature of Lemaître coordinates is to be regular at $r = 2m$, i.e. across the event horizon of Schwarzschild spacetime, contrary to Schwarzschild-Droste coordinates.

Historical note : The Schwarzschild metric in the form (14.21) has been obtained in 1932 by Georges Lemaître [345], as a vacuum solution of the Lemaître-Tolman system: cf. Eq. (11.12) of Ref. [345]. Remarkably, Lemaître pointed out that the metric components (14.21) 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. 192, eight years before, Arthur Eddington [176] exhibited a coordinate system that is regular at $r = 2m$ but he did not mention this feature.

14.2.7 Schwarzschild solution in Painlevé-Gullstrand coordinates

The **Painlevé-Gullstrand coordinates** on Schwarzschild spacetime are defined from the Lemaître coordinates $(\tau, \chi, \theta, \varphi)$ by using r , instead of χ , as the radial coordinate. They are thus the coordinates $(\tau, r, \theta, \varphi)$. The expression of the metric tensor in the Painlevé-Gullstrand coordinates is provided by Eq. (14.15) with $f(\chi) = 1$ and $\varepsilon = -1$:

$$\mathbf{g} = -\left(1 - \frac{2m}{r}\right) d\tau^2 + 2\sqrt{\frac{2m}{r}} d\tau dr + dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (14.26)$$

The Painlevé-Gullstrand coordinates $(\tau, r, \theta, \varphi)$ have a noticeable feature: the hypersurfaces $\tau = \text{const}$ are flat 3-manifolds, i.e. the metric induced of them by \mathbf{g} is the flat Euclidean metric. This is immediate: by setting $d\tau = 0$ in Eq. (14.26), one gets

$$\mathbf{g}^{\tau=\text{const}} = dr^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2),$$

which is nothing but the 3-dimensional Euclidean metric expressed in spherical coordinates (r, θ, φ) . This proves that Schwarzschild spacetime can be sliced by a family of flat hypersurfaces. The associated 3+1 decomposition of the metric tensor is revealed by rewriting (14.26) as

$$\mathbf{g} = -d\tau^2 + \left(dr + \sqrt{\frac{2m}{r}} d\tau\right)^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2). \quad (14.27)$$

One reads on this expression that the *lapse function* (see e.g. Ref. [226]) is $N = 1$ and that the *shift vector* is $\beta = \sqrt{2m/r} \partial_r$. Finding a lapse function equal to one simply reflects that the coordinate time τ is the proper time of some observer, namely the observer following the marginally bound radial geodesics discussed above.

Another interesting property of Painlevé-Gullstrand coordinates, which they share with Lemaître ones, is to be regular at $r = 2m$: despite the vanishing of $g_{\tau\tau}$ there, as read on (14.26), the determinant of the metric components (14.26) is not vanishing, thanks to the off-diagonal term $g_{\tau r}$. Indeed, $\det(g_{\alpha\beta}) = -r^4 \sin^2 \theta$, which is clearly nonzero, except on the axis $\theta = 0$ or π .

Remark 3: Painlevé-Gullstrand coordinates $(\tau, r, \theta, \varphi)$ can be seen as a timelike analog of the ingoing Eddington-Finkelstein (IEF) coordinates $(\tilde{t}, r, \theta, \varphi)$ introduced in Sec. 6.3.2. Both coordinate systems are based on ingoing radial geodesics of Schwarzschild spacetime, these geodesics being null for IEF and timelike for Painlevé-Gullstrand.

The reader is referred to Ref. [356] for a more detailed discussion of Painlevé-Gullstrand coordinates.

Historical note : Painlevé-Gullstrand coordinates have been introduced in 1921 by the French mathematician Paul Painlevé [396], as well as by the Swedish physicist and ophthalmologist Allvar Gullstrand (1911 laureate of the Nobel Prize in Medicine) in 1922 [247].

14.3 Oppenheimer-Snyder collapse

14.3.1 Pressureless collapse of a star from rest

An important example of dust collapse is that of a spherical star in hydrostatic equilibrium that suddenly loses pressure support to counterbalance gravity. This roughly models the astrophysical phenomenon of gravitational collapse of the iron core of a massive star at the end of thermonuclear evolution — a phenomenon that can ultimately give birth to a supernova if some bounce occurs. In this astrophysical event, the pressure never vanishes, but after the collapse has reached a certain stage, it plays a negligible role on the dynamics, so that the pressureless (dust) approximation developed in this chapter is quite good.

Given the assumed spherical symmetry, the Jebsen-Birkhoff theorem (Property 14.4) implies that the spacetime metric outside the star is Schwarzschild metric. Let us then characterize the star by its areal radius r_0 (i.e. the Schwarzschild-Droste coordinate r of the star's surface, cf. Sec. 6.2.2) at the start of the collapse and its gravitational mass m . The latter is nothing but the mass parameter of the Schwarzschild metric. It stays constant during all the collapse. We shall describe the spacetime exterior to the star by means the ingoing Eddington-Finkelstein coordinates (IEF) $(\tilde{t}, r, \theta, \varphi)$ introduced in Sec. 6.3.2, because they are regular on the black hole event horizon, contrary to Schwarzschild-Droste coordinates. The exterior metric is then given by Eq. (6.33):

$$\mathbf{g} = - \left(1 - \frac{2m}{r} \right) d\tilde{t}^2 + \frac{4m}{r} d\tilde{t} dr + \left(1 + \frac{2m}{r} \right) dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (14.28)$$

Let us denote by $r_s(\tau)$ the areal radius of the star, i.e. the coordinate r of its surface, as a function of the proper time τ of a matter particle located at the surface. If the origin of τ is set to the start of the collapse, we have

$$r_s(0) = r_0. \quad (14.29)$$

Due to the pressureless hypothesis, the surface particles follow timelike radial geodesics of Schwarzschild spacetime, as studied in Sec. 7.3.2. Since the collapse starts from rest, the function $r_s(\tau)$ is given by the geodesic solution (7.41):

$$\boxed{\begin{cases} \tau = \sqrt{\frac{r_0^3}{8m}} (\eta + \sin \eta) \\ r_s(\tau) = \frac{r_0}{2} (1 + \cos \eta) \end{cases} \quad 0 \leq \eta \leq \pi,} \quad (14.30)$$

where the parameter η is zero at the start of the collapse and π at its end. The function $r_s(\tau)$ is plotted in Fig. 14.1 (right). Its graph is an arch of cycloid, the parameter η corresponding to the angle of rotation of the rolling circle defining the cycloid.

Remark 1: The “cycloidal” evolution law $r_s(\tau)$, as given by Eq. (14.30), is identical to that of dust ball in Newtonian gravity, with τ being then the (universal) Newtonian time. This coincidence follows from the Lemaître-Tolman equation (14.9) being identical to the first integral of a purely radial motion in Newtonian gravity.

The IEF coordinate time \tilde{t} corresponding to the surface proper time τ is given by Eq. (7.45):

$$\tilde{t} = 2m \left\{ \sqrt{\frac{r_0}{2m} - 1} \left[\eta + \frac{r_0}{4m} (\eta + \sin \eta) \right] + 2 \ln \left[\cos \frac{\eta}{2} + \left(\frac{r_0}{2m} - 1 \right)^{-1/2} \sin \frac{\eta}{2} \right] \right\}, \quad (14.31)$$

where $\tilde{t} = 0$ is assumed at the start of the collapse. Note that by combining Eqs. (14.30) and (14.31), one can obtain $r_s = r_s(\tilde{t})$, i.e. the evolution law of the stellar surface in terms of the IEF coordinate time \tilde{t} . The right panel in Fig. 7.4 depicts it for various values of m/r_0 .

14.3.2 Oppenheimer-Snyder solution

The metric (14.28) is valid only for $r \geq r_s(\tilde{t})$. The metric inside the star is given by one of the solutions of the Lemaître-Tolman system given in Sec. 14.2.4. Actually only the solutions with $E(\chi) < 0$, i.e. those given by Eq. (14.13), are to be considered because they are the only ones allowing for a collapse starting from rest.

As stressed in Sec. 14.2.4, formula (14.13), in conjunction with Eqs. (14.7b) and (14.6), defines an infinity of solutions depending on the functions $f(\chi)$, $M(\chi)$ and $\tau_0(\chi)$, which can be chosen arbitrarily. We shall actually focus on the simplest of these solutions, which is that describing a *homogeneous* star, i.e. an object whose density ρ in the matter frame, as defined by Eq. (14.2), is constant at any given proper time τ . In other words, ρ , which is a priori a function of (τ, χ) , is actually a function of τ only. This is certainly a crude approximation of a real star (or stellar iron core), which has a non-constant density profile, but it captures the essential features of spherically symmetric collapse giving birth to a black hole. This homogeneous solution

is usually called ***Oppenheimer-Snyder collapse***, although the solution originally presented by J. Robert Oppenheimer and Hartland Snyder in 1939 [391] was slightly different, since it regards the marginally bound case, i.e. the solution $E(\chi) = 0$ [Eq. (14.12)] instead of $E(\chi) < 0$ [Eq. (14.13)] (cf. the historical note below).

As we are going to see, the homogeneous dust collapse is obtained by the following choice of the freely specifiable functions:

$$f(\chi) = \cos \chi, \quad M(\chi) = \frac{a_0}{2} \sin^3 \chi \quad \text{and} \quad \tau_0(\chi) = 0, \quad (14.32)$$

with

$$a_0 := \sqrt{\frac{r_0^3}{2m}}. \quad (14.33)$$

The above choice of $f(\chi)$ leads to $2E(\chi) = -\sin^2 \chi$ [cf. Eq. (14.10)]; hence $E(\chi) < 0$ for $\chi > 0$ and the solution to be considered is (14.13). The prefactors are $M(\chi)/|2E(\chi)|^{3/2} = a_0/2$ and $M(\chi)/|2E(\chi)| = (a_0/2) \sin \chi$, so that we get, taking into account the choice $\tau_0(\chi) = 0$,

$$\begin{cases} \tau = \frac{a_0}{2} (\eta + \sin \eta) \\ r(\tau, \chi) = \frac{a_0}{2} \sin \chi (1 + \cos \eta) \end{cases}$$

$$0 \leq \eta \leq \pi. \quad (14.34)$$

We note that τ depends only on η and not on χ , contrary to what happens for the general solution (14.13). Moreover, thanks to the choice (14.33), τ matches the proper time at the surface of the star as given by Eq. (14.30). The areal radii at the surface must match as well. Let us denote by χ_s the surface value of χ . The matching of the surface areal radii given by Eqs. (14.30) and (14.34) implies $r_0 = \sqrt{r_0^3/2m} \sin \chi_s$, from which we get the relation between χ_s and (m, r_0) :

$$\sin \chi_s = \sqrt{\frac{2m}{r_0}}.$$

$$(14.35)$$

This implies $\sin \chi_s > 0$ and hence $\chi_s \in (0, \pi)$. There are actually two solutions of Eq. (14.35) for χ_s . Given that \arcsin is a one-to-one map $[0, 1] \rightarrow [0, \pi/2]$, they are

$$\chi_s = \arcsin \sqrt{\frac{2m}{r_0}}$$

$$(14.36)$$

and $\chi_s = \pi - \arcsin \sqrt{2m/r_0}$. For the physical scenario we have in mind (gravitational collapse of an “ordinary” stellar core), we shall retain only the solution (14.36), which implies $\chi_s \in (0, \pi/2]$. The second solution, for which $\chi_s \in [\pi/2, \pi)$, will be discussed in Remark 3 on p. 557 below⁴. Equation (14.36) shows that χ_s is a function of the dimensionless ratio m/r_0 , which is the ***compactness*** of the initial star (some numerical values can be found in Table 14.1).

⁴For the moment, we can note that such a solution would lead to the areal radius r being a decreasing function of χ near χ_s , since Eq. (14.34) shows that $r(\tau, \chi)$ behaves as $\sin \chi$, which is decaying on $(\pi/2, \pi)$. Such a behavior is certainly not expected in an “ordinary” star.

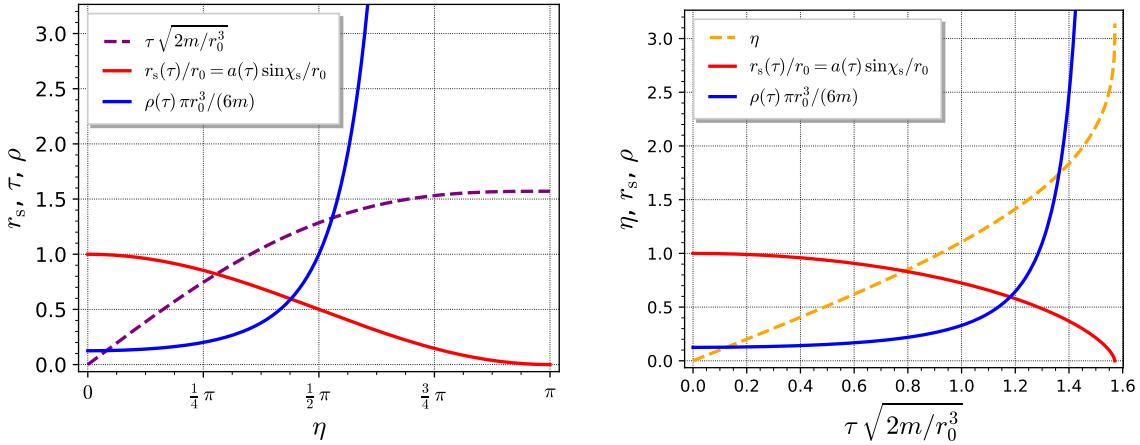


Figure 14.1: Evolution of the areal radius $r_s(\tau)$ the collapsing star [Eq. (14.30); red curve] and of the proper matter density $\rho(\tau)$ [Eq. (14.39); blue curve] in terms of the conformal time η (left) or the matter proper time τ (right). The function $\tau = \tau(\eta)$ is depicted by the purple dashed curve, while its inverse, $\eta = \eta(\tau)$, is depicted by the orange dashed one. As indicated in the legends, the plot of metric factor $a(\tau)$ is the same as that of $r_s(\tau)$ up to a rescaling by $\sin \chi_s$. [Figures generated by the notebook D.6.2]

The upper limit $\pi/2$ can be excluded from the range of χ_s since it would correspond to $r_0 = 2m$, which is the areal radius of the black hole horizon in Schwarzschild geometry. The range of the comoving coordinate χ in the interior of the star is thus

$$0 \leq \chi \leq \chi_s < \frac{\pi}{2}, \quad (14.37)$$

with $\chi = 0$ locating the center, since it corresponds to $r = 0$ by virtue of Eq. (14.34).

By combining Eqs. (14.32), (14.33) and (14.35), we note that $M(\chi)$ is an increasing function of χ in the range (14.37), with $M(0) = 0$ and

$$M(\chi_s) = m. \quad (14.38)$$

The matter density $\rho(\tau, \chi)$ is deduced from Eq. (14.7b), with the above values of $M(\chi)$ and $r(\tau, \chi)$. We get a function of τ only, via $\eta = \eta(\tau)$ [cf. Eq. (14.34)], which we rewrite as $\rho(\tau)$:

$$\rho(\tau) = \frac{6m}{\pi r_0^3 (1 + \cos \eta)^3}. \quad (14.39)$$

The independence of ρ from χ confirms that the choice (14.32) leads to a uniform density object, i.e. a homogeneous star. Let us stress that the homogeneity is maintained during all the collapse: at each instant of matter proper time τ , the density is uniform in all the star. The function $\rho(\tau)$ is plotted in Fig. 14.1.

Remark 2: In view of Eq. (14.30), we may rewrite Eq. (14.39) as

$$\rho(\tau) = \frac{m}{\frac{4}{3} \pi r_s(\tau)^3}. \quad (14.40)$$

This formula happens to be identical to that giving the mass density of a uniform ball of mass m and radius $r_s(\tau)$ in Newtonian physics (flat spacetime), but there is no profound physical significance in this coincidence.

It is instructive to express the initial energy density, $\rho_0 := \rho(\tau = 0)$, in terms of the geometrical mass density unit m^{-2} and the compactness angle χ_s by combining Eqs. (14.39) and (14.35):

$$\rho_0 = \frac{3 \sin^6 \chi_s}{32\pi m^2}. \quad (14.41)$$

The metric factor $a(\tau, \chi)$, which appears in Eq. (14.1), is deduced from Eq. (14.5), with $\partial r/\partial \chi$ computed from expression (14.34); as for ρ , we get a function of τ only, which we rewrite as $a(\tau)$:

$$a(\tau) = \frac{a_0}{2} (1 + \cos \eta). \quad (14.42)$$

We may then rewrite the second equation in (14.34) as $r(\tau, \chi) = a(\tau) \sin \chi$, so that the metric tensor (14.1) inside the star takes the form

$$\boxed{\mathbf{g} = -d\tau^2 + a(\tau)^2 [d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2)]}. \quad (14.43)$$

We recognize the **Friedmann-Lemaître-Robertson-Walker (FLRW) metric** corresponding to a closed universe (cf. e.g. Chap. 3 of Ref. [412] or Chap. 17 of Book 3 of Ref. [162]). This is not so surprising since FLRW metrics are solution of the Einstein equation based on the hypothesis of homogeneity of constant τ slices. We may use η , instead of τ , as the time coordinate inside the star, thanks to the relation

$$d\tau = \frac{a_0}{2}(1 + \cos \eta)d\eta, \quad (14.44)$$

which follows from (14.34). We then deduce from Eq. (14.43) that

$$\boxed{\mathbf{g} = \frac{a_0^2}{4}(1 + \cos \eta)^2 [-d\eta^2 + d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2)]}. \quad (14.45)$$

The term inside square brackets is nothing but the metric $\tilde{\mathbf{g}}$ of the *Einstein cylinder* introduced in Sec. 4.2.3 [compare Eq. (4.18) with the change of notation $\eta \leftrightarrow \tau$]. Hence, inside the star, the metric \mathbf{g} is conformal to $\tilde{\mathbf{g}}$: $\mathbf{g} = \Omega^2 \tilde{\mathbf{g}}$, with the conformal factor $\Omega := a_0/2(1 + \cos \eta) = a(\tau)$. For this reason, η is called the **conformal time**. Note that while $\tilde{\mathbf{g}}$ is a static metric [it obeys Eq. (5.5)], \mathbf{g} is not, since the conformal factor depends on η .

On each hypersurface $\eta = \text{const}$, or equivalently $\tau = \text{const}$, the 3-metric \mathbf{h} induced by the conformal metric $\tilde{\mathbf{g}}$ is the standard metric of the hypersphere \mathbb{S}^3 , expressed in terms of the hyperspherical coordinates (χ, θ, φ) :

$$\mathbf{h} = d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (14.46)$$

On the whole \mathbb{S}^3 , χ ranges from 0 to π . Since in the present case, $\chi \in [0, \chi_s]$ with $\chi_s < \pi/2$ [Eq. (14.37)], we may view each constant η slice of the interior of the collapsing star as a piece

of \mathbb{S}^3 , scaled by the factor $a(\tau)$ given by Eq. (14.42). If one uses the conformal coordinates (η, χ) to draw a 2-dimensional spacetime diagram of the interior, as in the left part of Fig. 14.2, then the radial null geodesics appear as straight lines inclined by $\pm 45^\circ$ with the horizontal, reflecting the Minkowskian part $-d\eta^2 + d\chi^2$ of the metric (14.45).

To show that the FLRW interior (14.45) along with the Schwarzschild exterior (14.28) form a regular spacetime, one shall check that the *Darmois junction conditions* are fulfilled at the interface between the two regions, i.e. on the timelike hypersurface Σ defining the surface of the collapsing star. These conditions, first enounced by Georges Darmois [153] (see also Sec. 21.13 of MTW textbook [370], as well as Ref. [331]), are the continuity of the induced 3-metric of Σ and of its extrinsic curvature, when these two tensors are computed on each side of Σ . These geometric conditions translate the absence of surface layers on Σ . We shall not check them here; the computation can be found in Sec. 3.8 of Poisson's textbook [415].

Historical note : As mentioned above, the solution originally obtained by J. Robert Oppenheimer and Hartland Snyder in 1939 [391] regards pressureless matter evolving on marginally bound timelike geodesics, i.e. matter in free fall from $r \rightarrow +\infty$, where it is at rest at $\tau \rightarrow -\infty$. More precisely, Oppenheimer and Snyder's original solution corresponds to the choice

$$f(\chi) = 1 \quad M(\chi) = m \frac{\chi^3}{\chi_s^3} \quad \text{and} \quad \tau_0(\chi) = \sqrt{\frac{2\chi_s^3}{9m}} \quad (14.47)$$

for the freely specifiable functions determining the solutions of the Lemaître-Tolman system (cf. Sec. 14.2.4), instead of the choice (14.32) adopted in this chapter. Choosing $f(\chi) = 1$ leads to $E(\chi) = 0$ [cf. Eq. (14.10)] and hence to the solution (14.12), instead of (14.13). This solution is still homogeneous, since ρ , computed via Eq. (14.7b), is a function of τ only, but it is less adapted to a star of finite size collapsing from an equilibrium state than the solution (14.13). The latter, which we are using here, has been first presented by David Beckedorff in 1962 [45] and popularized by B. Kent Harrison, Kip Thorne, Masami Wakano and John Archibald Wheeler in their 1965 monograph on gravitational collapse [255], as well as by Misner, Thorne and Wheeler in their 1973 textbook [370].

14.3.3 Spacetime diagrams and maximal extensions

To draw a full spacetime diagram of the collapse, we may extend the ingoing Eddington-Finkelstein (IEF) coordinates $(\tilde{t}, r, \theta, \varphi)$, used to described the Schwarzschild exterior of the star to the interior region by demanding that relation (14.31), which a priori links \tilde{t} to η at the stellar surface, holds everywhere in the interior. Regarding the IEF coordinate r , it is naturally identified with the areal radius $r(\eta, \chi)$ given by Eq. (14.34). The resulting spacetime diagram is shown in the right part of Fig. 14.2. In this diagram, only the ingoing radial null geodesics in the exterior appear as straight lines, as a consequence of the definition of the IEF coordinates (cf. Sec. 6.3.2). The slices of constant proper time τ (dashed red lines) appear as horizontal line segments by virtue of the above choice for \tilde{t} in the interior. In particular, the star at $\tau = 0$ is depicted by a horizontal magenta segment, as in the (η, χ) diagram on the left of the figure.

Figure 14.3 presents another view of the Oppenheimer-Snyder collapse. It is built on Kruskal-Szekeres-like coordinates (T, X) (cf. Sec. 9.2). These coordinates are defined from the IEF-like ones (\tilde{t}, r) via formulas (9.23), with \tilde{t} being replaced there by $\tilde{t} - 5m$; this translation in \tilde{t} ensures a better representation of the final stages of the collapse. One may note that the

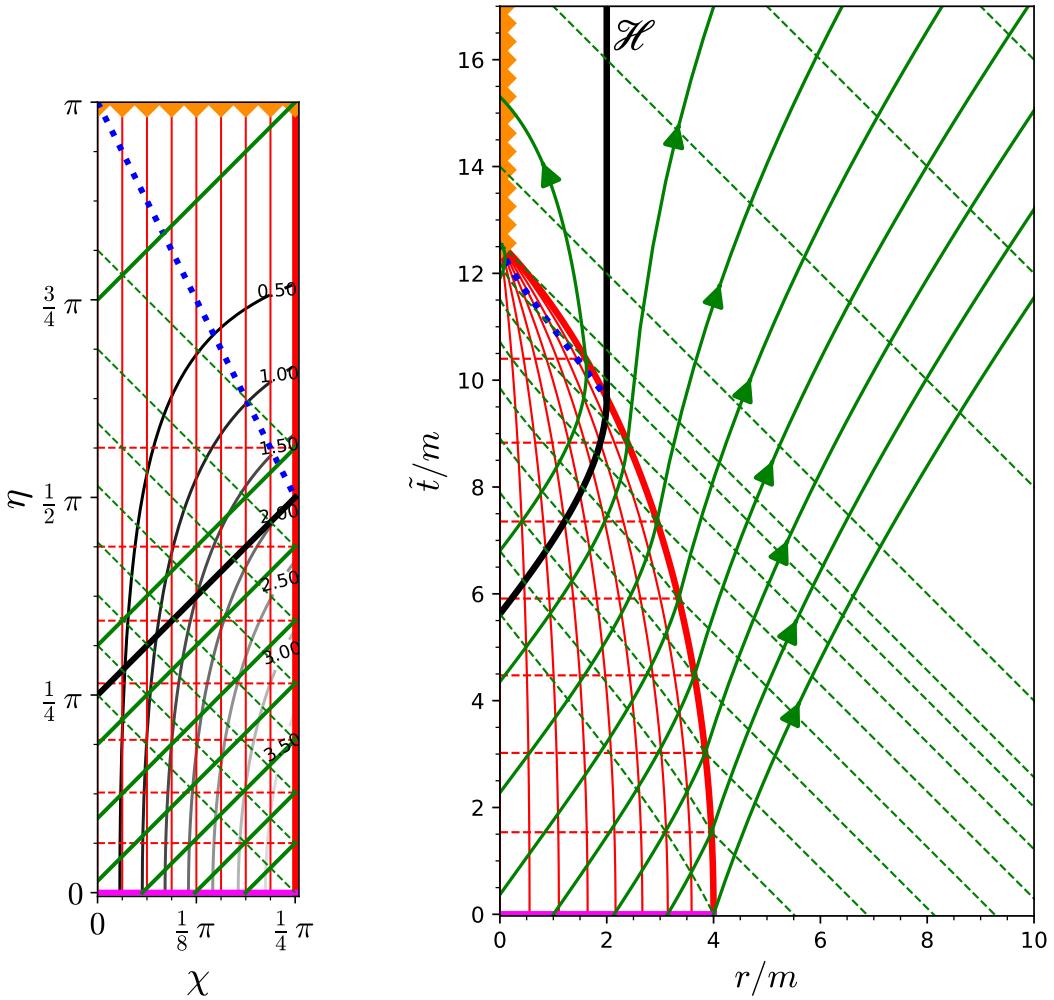


Figure 14.2: Spacetime diagrams of the Oppenheimer-Snyder collapse in conformal coordinates (η, χ) (left figure, depicting only the interior) and in Eddington-Finkelstein coordinates (\tilde{t}, r) (right figure), for the initial compactness $m/r_0 = 1/4$, which corresponds to $\chi_s = \pi/4$. In both figures, solid red lines are worldlines of matter particles, uniformly sampling $[0, \chi_s]$ with $\delta\chi = \pi/32$, while dashed red lines denote isosurfaces of constant matter proper time τ uniformly sampling $[0, \tau_{\text{end}}]$ with $\delta\tau = \tau_{\text{end}}/8$. The thick red line is the star's surface and the thick magenta one is the star at the initial instant $\tau = 0$. The orange zigzag line marks the curvature singularity. Solid (resp. dashed) green lines are outgoing (resp. ingoing) radial null geodesics encountering the star's surface at the above selected values of τ . The thick black line is the black hole event horizon \mathcal{H} and the dotted blue line is the centered future inner trapping horizon \mathcal{T} . In the left figure, thin solid grey to black lines mark isosurfaces of constant values of the areal radius $r(\tau, \chi)$ (labeled in units of m). [Figures generated by the notebook D.6.2]

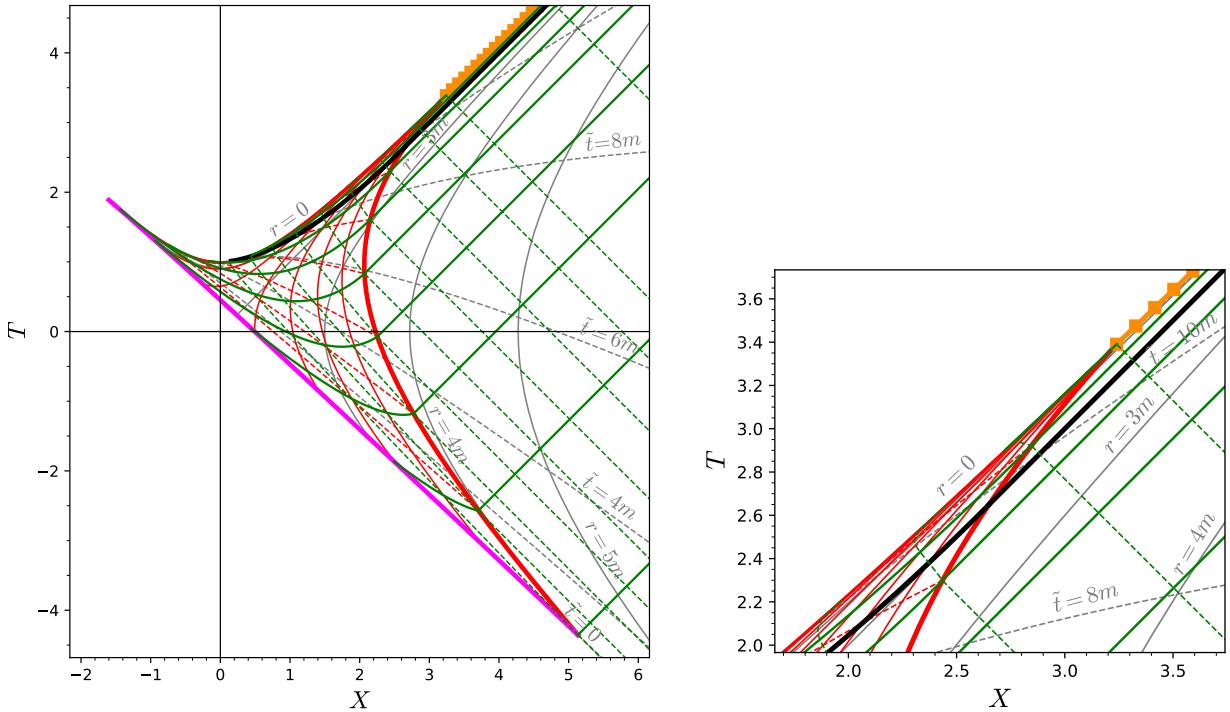


Figure 14.3: Spacetime diagrams of the Oppenheimer-Snyder collapse in Kruskal-Szekeres-type coordinates (T, X) . The colored curves are the same as in Fig. 14.2. In particular, the magenta curve represents the star at $\tau = t = 0$, when the collapse starts. In addition, solid grey lines are hypersurfaces of constant r , uniformly sampling $[0, 8m]$ with $\delta r = m$ and dashed lines are hypersurfaces of constant \tilde{t} , uniformly sampling $[0, 16m]$ with $\delta \tilde{t} = 2m$. The right figure is a zoom on the end of the collapse. [Figures generated by the notebook D.6.2]

surface of the star (thick solid red line) is initially tangent to the hypersurface $r = 4m$ (solid grey line), which is the geometrical translation of the collapse starting from rest. The part of the diagram that lies to the right of the thick solid red line is identical to a part of the diagram shown in Fig. 9.5, which is expected since both represent a piece of Schwarzschild spacetime. A major difference with the Kruskal-Szekeres diagram of Fig. 9.5 is that the spacetime diagram shown in Fig. 14.3 is not conformal in the stellar interior: the radial null geodesics are not straight lines inclined by $\pm 45^\circ$. The outgoing ones even run backwards in T in the left part of the diagram, which means that causal relations cannot be simply inferred there.

Finally let us consider a compactified conformal diagram of the Oppenheimer-Snyder collapse, i.e. a *Carter-Penrose diagram* (cf. Secs. 9.4 and 10.8.1). Thanks to the compactification, such a diagram offers a view of the entire spacetime. It is then natural to consider *maximal* spacetimes, i.e. spacetimes in which each geodesic is inextendible, being complete or terminating at a singularity (cf. Sec. B.3.2). This is clearly not the case of the spacetime depicted in the diagrams of Figs. 14.2 and 14.3: when run in the past direction, causal geodesics in the star interior stop on the hypersurface $\tau = 0$ (the magenta line). In particular, the free-falling matter geodesics, along which τ is an affine parameter, are clearly incomplete and could a priori be extended to $\tau < 0$.

There are actually two ways to maximally extend the Oppenheimer-Snyder spacetime considered up to now. The first one corresponds to a star in hydrostatic equilibrium for $\tau < 0$

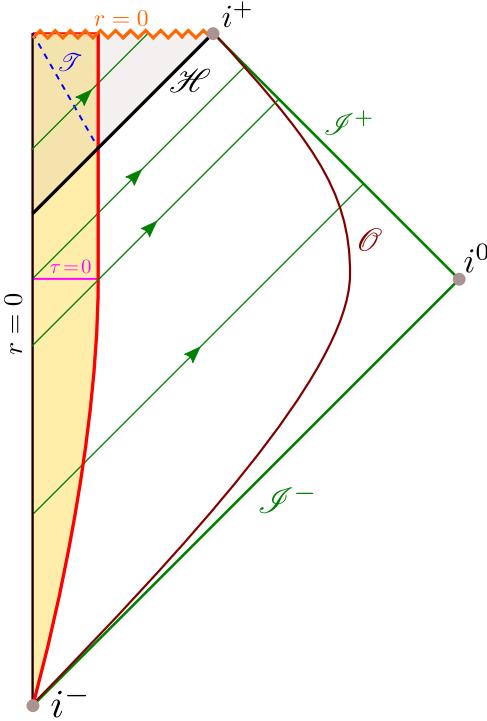


Figure 14.4: Carter-Penrose diagram of the Oppenheimer-Snyder collapse generated by switching off the pressure at $\tau = 0$ in a star in hydrostatic equilibrium for $\tau \in (-\infty, 0)$. The stellar surface is indicated by the red curve and some outgoing radial null geodesics are drawn as green straight lines. The black hole event horizon \mathcal{H} is indicated by the black line, the shaded region being the black hole interior. The dashed blue line marks the trapping horizon \mathcal{T} discussed in Sec. 14.3.6, while the orange zigzag line marks the curvature singularity. The maroon curve is the worldline of the remote static observer \mathcal{O} considered in Sec. 14.4.

and in which the pressure drops suddenly to zero at $\tau = 0$. This extension is “astrophysical” insofar as it models the gravitational collapse of the core of a massive star. Of course the equilibrium for $\tau < 0$ requires some nonzero pressure inside the star. Therefore it cannot be described by the Lemaître-Tolman equations discussed in Sec. 14.2; it should rather obey the *Tolman-Oppenheimer-Volkoff equations*, which govern hydrostatic equilibria in spherical symmetry (see e.g. Chap. 23 of Ref. [370]). The Carter-Penrose diagram corresponding to this extension is drawn schematically (i.e. not by means of explicit coordinates) in Fig. 14.4. The star seems to expand from the past timelike infinity i^- , but this is an artifact of the drawing⁵: the star surface (red line) follows actually a curve $r = \text{const} (= r_0)$ for $\tau < 0$. For $\tau \geq 0$, the surface of the star appears as a vertical line, as in the left part of Fig. 14.2, while the radius is actually decreasing. Actually, one may consider that this part of the Carter-Penrose diagram is drawn by means of the conformal coordinates (η, χ) used in Fig. 14.2. Note that the Carter-Penrose diagram of Fig. 14.4 corresponds to a maximally extended spacetime: all geodesics are inextendible. Note as well that the exterior region is formed by parts of regions \mathcal{M}_I and \mathcal{M}_{II} of Schwarzschild spacetime (compare with Fig. 9.10).

The second maximal extension of the Oppenheimer-Snyder spacetime is based on the

⁵Recall that a Carter-Penrose diagram reflects the causal structure of spacetime but does not respect the metric distances.

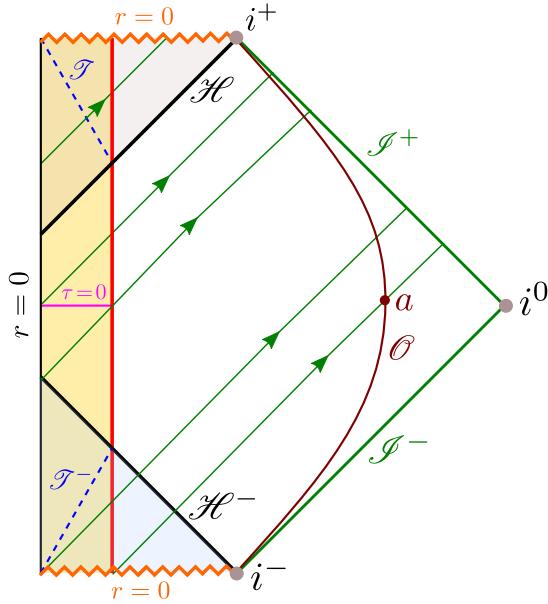


Figure 14.5: Carter-Penrose diagram of the Oppenheimer-Snyder collapse starting at the instant of time symmetry ($\tau = 0$) after an expanding phase from a past singularity. The legend is the same as in Fig. 14.4, with in addition the past curvature singularity (bottom orange zigzag line) and the white hole region (bottom shaded area), bounded by the past event horizon \mathcal{H}^- .

assumption that the proper time $\tau = 0$ is an instant of time symmetry of a ball of pressureless matter that was expanding from an initial past singularity. The spacetime for $\tau < 0$ is then exactly the time-reversed of that for $\tau > 0$, as illustrated on the Carter-Penrose diagram of Fig. 14.5. Such an extension is less “astrophysical” than the one considered above, but it has the advantage to invoke only pressureless matter and hence to be entirely described by an exact solution of the Einstein equation: the Friedmann-Lemaître-Robertson-Walker solution (14.45) extended to $\eta \in (-\pi, \pi)$ inside the star and matched to the Schwarzschild solution outside the star. This maximal solution contains a *white hole* region, i.e. a region in which causal curves arising from the past null infinity \mathcal{I}^- cannot penetrate (cf. Sec. 4.4.2). The white hole (null) boundary is the past event horizon \mathcal{H}^- (cf. Fig. 14.5). The pressureless matter starts its life by expanding from a past curvature singularity where $r = 0$, as a “mini big-bang”, the areal radius being given by Eq. (14.34) with $\eta \in (-\pi, 0)$ in the expansion phase and $\eta \in (0, \pi)$ in the collapse phase. Note that the exterior region is formed by parts of regions \mathcal{M}_I , \mathcal{M}_{II} and \mathcal{M}_{IV} of the maximally extended Schwarzschild spacetime (compare with Fig. 9.10). Note also that the star is not visible from the infinite past to the remote static observer \mathcal{O} : it only appears to \mathcal{O} after the event denoted by a in Fig. 14.5. On the contrary, in the “astrophysical” model discussed above, the star is always visible to \mathcal{O} (cf. Fig. 14.4), despite it fades very rapidly after the collapse phase has started to be observed, as we shall see in Sec. 14.4.

Remark 3: The time-symmetric extended solution considered above opens the path to pressureless solutions with $\chi_s \in (\pi/2, \pi)$, i.e. with $\chi_s = \pi - \arcsin \sqrt{2m/r_0}$ as the solution of the junction equation (14.35). We disregarded these solutions in Sec. 14.3.2, after having noticed that they lead to the areal radius $r(\tau, \chi)$ being a decreasing function of χ for $\chi \in (\pi/2, \chi_s]$. In other words, starting from the center of the star at $\chi = 0$ and keeping τ fixed, $r(\tau, \chi)$ increases up to the maximum value $r_s(\tau)/\sin \chi_s$,

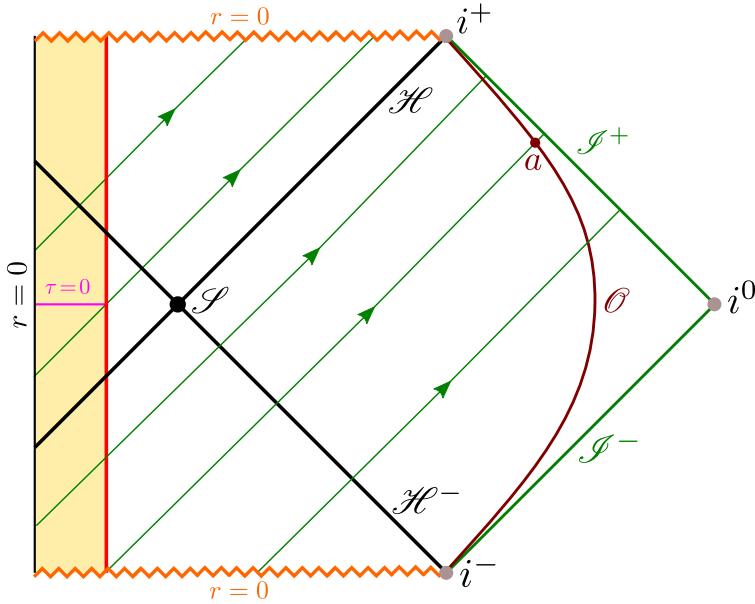


Figure 14.6: Carter-Penrose diagram of a semiclosed world. See Fig. 14.5 for the legend.

achieved for $\chi = \pi/2$, and then decreases to $r_s(\tau)$ for $\chi = \chi_s$. Such a behavior is possible if the star's surface meets Schwarzschild spacetime in a region of the latter where r decreases when moving away from the surface (i.e. “to the right” if one places the star at the left of a Carter-Penrose diagram as in Figs. 14.4-14.5). This happens in the region $\hat{X} < 0$ of the compactified Kruskal-Szekeres coordinates (cf. Fig. 9.10). We may then place the star there; the corresponding Carter-Penrose diagram is shown in Fig. 14.6. We note that the star exterior comprises the entire region \mathcal{M}_1 of the extended Schwarzschild spacetime, as well as parts of regions \mathcal{M}_{II} , \mathcal{M}_{III} and \mathcal{M}_{IV} . Contrary to the solution for $\chi_s < \pi/2$ shown in Fig. 14.5, it contains the bifurcation sphere \mathcal{S} , where \mathcal{H} meets \mathcal{H}^- (cf. Sec. 9.3.3). Such a configuration is called a **semiclosed world** (cf. p. 149 of Ref. [530], Sec. 2.7 of Ref. [386] or Sec. 2.7.2 of Ref. [201]), because at the limit $\chi_s \rightarrow \pi$, the matter region becomes an entire FRLW closed universe. Note that the collapse phase ($\tau > 0$) cannot be seen by a remote static observer, like \mathcal{O} in Fig. 14.6, for it is hidden below the event horizon \mathcal{H} ; only some initial part of the expanding phase is visible to \mathcal{O} on the segment $a - i^+$ of his worldline.

Historical note : The *semiclosed world* solutions have been introduced by Oscar Klein in 1961 [316] and Yakov B. Zeldovich in 1962 [525]. They have been further studied by Igor D. Novikov in 1963 [385].

14.3.4 Final singularity

The collapse ends at $\eta = \pi$, since this value corresponds to the surface areal radius $r_s(\tau) = 0$, by virtue of Eq. (14.30). Equation (14.34) shows that this corresponds as well to $r(\tau, \chi) = 0$ for all particles inside the star. The matter proper time τ at the end of the collapse is given by $\eta = \pi$ in Eq. (14.34):

$$\tau_{\text{end}} = \frac{\pi}{2} a_0 = \pi \sqrt{\frac{r_0^3}{8m}} = \pi \left(\frac{r_0}{2m} \right)^{3/2} m = \pi \frac{m}{\sin^3 \chi_s} = \sqrt{\frac{3\pi}{32\rho_0}}, \quad (14.48)$$

where the last equality has been obtained by using Eq. (14.41) to express $m/\sin^3 \chi_s$ in terms of the initial proper energy density ρ_0 . We note that τ_{end} is a function of ρ_0 only.

Remark 4: The above formulas for the collapse proper time τ_{end} are identical to those giving the collapse time of a spherical ball of dust of mass m , initial radius r_0 and initial mean mass density $\rho_0 = 3m/(4\pi r_0^3)$ in Newtonian gravity. This coincidence arises from the same cycloidal law for $r_s(\tau)$ in both general relativity and Newtonian gravity, as already discussed in Remark 1 on p. 549.

Similarly, setting $\eta = \pi$ in Eq. (14.31), we get the value of the IEF coordinate \tilde{t} at the end of the collapse:

$$\tilde{t}_{\text{end}} = 2m \left[\pi \left(1 + \frac{r_0}{4m} \right) \sqrt{\frac{r_0}{2m} - 1} - \ln \left(\frac{r_0}{2m} - 1 \right) \right]. \quad (14.49)$$

Example 1: For the compactness $m/r_0 = 1/4$, which is that considered in Figs. 14.2-14.3, the above formulas yield $\tau_{\text{end}} = 2\sqrt{2}\pi m \simeq 8.89 m$ and $\tilde{t}_{\text{end}} = 4\pi m \simeq 12.57 m$.

Example 2: Some numerical values of τ_{end} are given in Table 14.1 for various astrophysical objects, if they happen to collapse by a sudden lack of force acting against gravitation. One sees that the Earth would collapse in 15 minutes, while it would take half an hour for the Sun, 3 seconds for a white dwarf and a tenth of millisecond for a neutron star. The similarity of the values for the Earth and the Sun, which have very different masses and radii, is due to the closeness of their values of ρ_0 .

Equation (14.39) with $\eta \rightarrow \pi$ immediately yields an infinite proper energy density at the end of the collapse:

$$\lim_{\tau \rightarrow \tau_{\text{end}}} \rho(\tau) = +\infty. \quad (14.50)$$

Since ρ is a measurable quantity (by comoving observers), this signals some physical singularity. This corresponds actually to the apparition of a *curvature singularity* in the part of spacetime covering the interior of the star (cf. the discussion in Sec. 6.3.4). Indeed, from the interior metric (14.45), one can evaluate⁶ curvature scalars like the Ricci scalar

$$R := R^\mu_\mu = \frac{24}{a_0^2(1 + \cos \eta)^3} = \frac{6m}{r_s(\tau)^3}, \quad (14.51)$$

the “square” of the Ricci tensor

$$R_{\mu\nu}R^{\mu\nu} = \frac{576}{a_0^4(1 + \cos \eta)^6} = \frac{36m^2}{r_s(\tau)^6} \quad (14.52)$$

and the Kretschmann scalar defined by Eq. (6.44):

$$K := R_{\mu\nu\rho\sigma}R^{\mu\nu\rho\sigma} = \frac{960}{a_0^4(1 + \cos \eta)^6} = \frac{60m^2}{r_s(\tau)^6}. \quad (14.53)$$

Remark 5: Although the components of the Riemann and Ricci tensors with respect to the coordinates $(\eta, \chi, \theta, \varphi)$ depend on χ (cf. the notebook D.6.3), the above three scalars are independent of it, in full agreement with the homogeneity of the stellar interior at each instant η .

⁶See the notebook D.6.3 for the computations.

	Earth	Sun	white dwarf	neutron star
$m [M_\odot]$	$3 \cdot 10^{-6}$	1	0.6	1.4
$r_0 [\text{km}]$	$6.37 \cdot 10^3$	$6.96 \cdot 10^5$	$8.0 \cdot 10^3$	12
m/r_0	$7.0 \cdot 10^{-10}$	$2.1 \cdot 10^{-6}$	$1.1 \cdot 10^{-4}$	0.17
χ_s	$3.7 \cdot 10^{-5}$	$2.1 \cdot 10^{-3}$	$1.5 \cdot 10^{-2}$	0.63
$\rho_0 [\text{kg m}^{-3}]$	$5.5 \cdot 10^3$	$1.4 \cdot 10^3$	$5.6 \cdot 10^8$	$3.8 \cdot 10^{17}$
τ_{end}	14 min 55 s	29 min 30 s	2.82 s	$107 \mu\text{s}$
τ_{hb}	14 min 55 s	29 min 30 s	2.82 s	$75 \mu\text{s}$
$\tau_{\text{end}} - \tau_{\text{hb}}$	$6.7 \cdot 10^{-11} \text{ s}$	$22 \mu\text{s}$	$13 \mu\text{s}$	$32 \mu\text{s}$
$\tau_{\text{end}} - \tau_*$	$2.0 \cdot 10^{-11} \text{ s}$	$6.6 \mu\text{s}$	$3.9 \mu\text{s}$	$10.4 \mu\text{s}$
$\rho_{\text{hb}} [\text{kg m}^{-3}]$	$1.8 \cdot 10^{29}$	$1.8 \cdot 10^{18}$	$4.5 \cdot 10^{18}$	$1.4 \cdot 10^{18}$

Table 14.1: Numerical values for the gravitational collapse of various astrophysical bodies, assuming that all forces counterbalancing gravitation suddenly disappear at the proper time $\tau = 0$. m is the gravitational mass (constant during the collapse), r_0 is the initial areal radius, m/r_0 is the initial compactness, χ_s is the (constant) value of the hyperspherical angle χ at the surface of the body [Eq. (14.36)], ρ_0 is the initial proper energy density, assuming that the body is homogeneous [Eq. (14.40) with $\tau = 0$, or Eq. (14.41)], τ_{end} is the matter proper time at the end of the collapse, when the central singularity appears [Eq. (14.48)], τ_{hb} is the matter proper time when the black hole event horizon forms at the center of the body [Eq. (14.58)], τ_* is the proper time when the horizon encounters the infalling surface [Eq. (14.55)] and ρ_{hb} is the proper energy density at $\tau = \tau_{\text{hb}}$ [Eq. (14.59)].

It is clear that the above three curvature invariants diverge when $\eta \rightarrow \pi$, or equivalently, when $r_s(\tau) \rightarrow 0$. Hence a curvature singularity appears in the stellar interior at the end of the collapse. It is depicted by the orange zigzag segment in left spacetime diagram of Fig. 14.2. This singularity is shrunk to the point $(\tilde{t}, r) = (\tilde{t}_{\text{end}}, 0)$ in the IEF diagram on the right part of the figure, because $r_s(\tau_{\text{end}}) = 0$. It is then connected to the curvature singularity at $r = 0$ of Schwarzschild spacetime, depicted as the vertical orange zigzag segment.

Remark 6: The reader might be puzzled by the Kretschmann scalar (14.53) being different from that of Schwarzschild spacetime, as given by Eq. (6.45): $K_{\text{Schwarz}} = 48m^2/r^6$. Both have the same m^2/r^6 structure but they differ by the numerical prefactor: 60 versus 48. This actually reflects the discontinuity of the Riemann tensor between the stellar interior and the Schwarzschild exterior. This discontinuity is enforced on the Ricci part of the Riemann tensor via the Einstein equation and the uniform density of the star, the latter yielding a jump of the energy-momentum tensor \mathbf{T} between the value (14.2) and zero (vacuum exterior). The difference is even more dramatic on the Ricci scalar (14.51) and the square of the Ricci tensor (14.52), since they are identically zero in the Schwarzschild exterior. In other words, the Einstein equation implies that the metric tensor of the Oppenheimer-Snyder solution is not C^2 . For an actual stellar core, there is a density gradient so that $\rho = 0$ at the surface and the metric tensor is at least C^2 .

14.3.5 Black hole formation

Since the surface areal radius $r_s(\tau)$ is a monotonically decreasing function of τ (cf. Fig. 14.1), there exists a unique value τ_* of τ for which $r_s(\tau_*) = 2m$. For $\tau > \tau_*$, $r_s(\tau) < 2m$ and the exterior Schwarzschild spacetime contains a black hole event horizon \mathcal{H} , located at $r = 2m$ in IEF coordinates (the vertical black segment in Fig. 14.2). The value of τ_* is determined through the corresponding conformal time η_* via Eq. (14.30):

$$\frac{r_0}{2}(1 + \cos \eta_*) = 2m \iff \cos \eta_* = 4\frac{m}{r_0} - 1 = 2\sin^2 \chi_s - 1 = -\cos(2\chi_s),$$

where use has been made of Eq. (14.35). Given that $0 < \eta_* < \pi$ and $0 < \chi_s < \pi/2$, the solution is

$$\eta_* = \pi - 2\chi_s. \quad (14.54)$$

Equation (14.30) yields then

$$\tau_* = \frac{\pi - 2\chi_s + \sin(2\chi_s)}{\sin^3 \chi_s} m = \left(1 - \frac{2\chi_s - \sin(2\chi_s)}{\pi}\right) \tau_{\text{end}}.$$

(14.55)

Example 3: For the compactness $1/4$ considered in Figs. 14.2-14.3, one has $\chi_s = \pi/4$, so that $\eta_* = \pi/2$ and $\tau_* = \sqrt{2}(\pi + 2)m \simeq 7.27 m \simeq 0.82 \tau_{\text{end}}$.

Example 4: Values of τ_* for the various objects considered in Example 2 are given in Table 14.1. We note that for the collapsing Earth, Sun and white dwarf, τ_* is very close to τ_{end} , while for the collapsing neutron star, $\tau_* \simeq 0.9 \tau_{\text{end}}$. This reflects the smallness of χ_s for the first three objects, since Eq. (14.55) implies $\tau_*/\tau_{\text{end}} \simeq 1 - 4\chi_s^3/(3\pi)$ for $\chi_s \ll 1$.

For $\tau < \tau_*$, one has $r_s(\tau) > 2m$ and the event horizon must be located inside the collapsing star. To determine it, let us consider the outgoing radial null geodesics, which are drawn as solid green curves in Fig. 14.2. In view of the metric (14.45), the equation of such a geodesic is very simple in terms of the coordinates (η, χ) : $\eta = \chi - \chi_s + \eta_s$, where the constant η_s is the value of η when the geodesic reaches the surface of the star ($\chi = \chi_s$). Radial null geodesics for which $\eta_s < \eta_*$ emerge out of the star in a part of Schwarzschild spacetime where $r > 2m$, so that they can escape to infinity. On the contrary, geodesics having $\eta_s > \eta_*$ emerge in the black hole region of Schwarzschild spacetime; they are thus trapped and eventually end on the curvature singularity at $r = 0$. The event horizon is generated by the null radial geodesics of the limiting case: $\eta_s = \eta_*$. Hence, inside the star, the equation of the black hole event horizon \mathcal{H} is $\eta = \chi - \chi_s + \eta_*$. Given the value (14.54) for η_* , we get

$$\mathcal{H} : \boxed{\eta = \chi + \pi - 3\chi_s}. \quad (14.56)$$

For a fixed value of (θ, φ) , this is the equation of a straight line segment inclined by $+45^\circ$ with respect to the horizontal (cf. Fig. 14.2, left).

The equation of \mathcal{H} in terms of the internal IEF coordinate is obtained in parametric form, with parameter $\chi \in [0, \chi_s]$ by substituting (14.56) for η in Eq. (14.31) providing \tilde{t} and in Eq. (14.34) providing r . One obtains then a curved segment in the (\tilde{t}, r) plane, which is drawn in black in Fig. 14.2, right.

In view of Eq. (14.56), we may state:

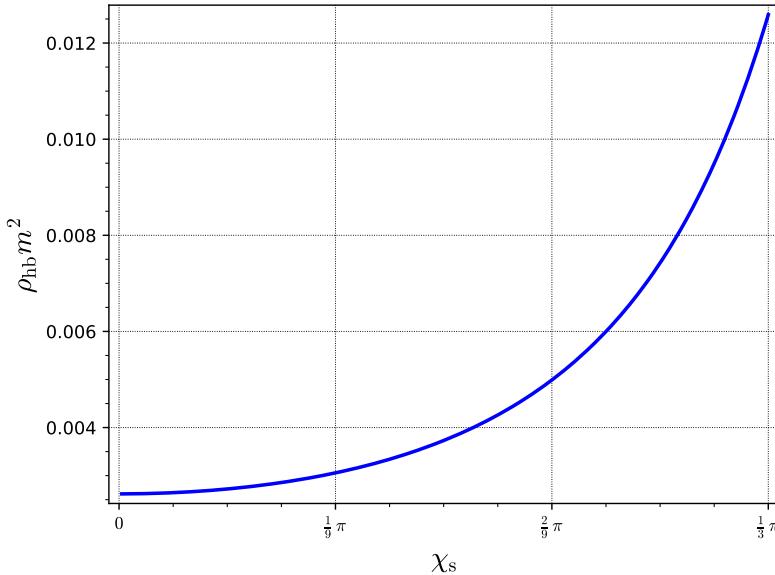


Figure 14.7: Proper energy density ρ_{hb} at the instant of the black hole formation at the center of the star, in units of m^{-2} , as a function of the initial compactness angle χ_s . [Figure generated by the notebook D.6.2]

Property 14.5: birth and growth of the event horizon

- If $\chi_s \leq \pi/3$, or equivalently if $m/r_0 \leq 3/8$, the event horizon \mathcal{H} appears at the center of the star ($\chi = 0$) at the conformal time

$$\eta_{\text{hb}} = \pi - 3\chi_s \quad (14.57)$$

or equivalently at the matter proper time

$$\tau_{\text{hb}} = \frac{\pi - 3\chi_s + \sin(3\chi_s)}{\sin^3 \chi_s} m = \left(1 - \frac{3\chi_s - \sin(3\chi_s)}{\pi}\right) \tau_{\text{end}}, \quad (14.58)$$

the label ‘hb’ standing for “horizon birth”.

- If $\chi_s > \pi/3$, or equivalently if $m/r_0 > 3/8$, the event horizon \mathcal{H} is already present in the star at $\tau = 0$, at the location $\chi = 3\chi_s - \pi$.

In all cases, the horizon expands according to Eq. (14.56) until it reaches the surface of the star ($\chi = \chi_s$) at $\eta = \eta_* = \pi - 2\chi_s$ or $\tau = \tau_*$ given by Eq. (14.55).

Example 5: Continuing with Example 3, i.e. with $m/r_0 = 1/4$, or equivalently with $\chi_s = \pi/4$, we get $\eta_{\text{hb}} = \pi/4$, which corresponds to the origin of the black straight segment drawn in Fig. 14.2, left. Equation (14.58) gives $\tau_{\text{hb}} = (2 + \pi/\sqrt{2})m \simeq 4.22 m \simeq 0.48 \tau_{\text{end}}$.

Example 6: Values of τ_{hb} for the various objects considered in Examples 2 and 4 are given in Table 14.1. As for τ_* , we note that for the collapsing Earth, Sun and white dwarf, τ_{hb} is very close to τ_{end} . This

follows from $\tau_{\text{hb}}/\tau_{\text{end}} \simeq 1 - 9\chi_s^3/(2\pi)$ for $\chi_s \ll 1$. On the contrary, for the collapsing neutron star, $\tau_{\text{hb}} \simeq 0.7 \tau_{\text{end}}$.

For $\chi_s \leq \pi/3$, the matter proper energy density at the horizon birth, i.e. at $\tau = \tau_{\text{hb}}$, is obtained by substituting (14.57) for η into Eq. (14.39); using the identity (14.35) to express m^3/r_0^3 in terms of χ_s , we get

$$\rho_{\text{hb}} := \rho(\tau_{\text{hb}}) = \frac{3}{4\pi m^2} \frac{\sin^6 \chi_s}{(1 - \cos(3\chi_s))^3}. \quad (14.59)$$

It is instructive to take the limit of small initial compactness $m/r_0 \ll 1$, which corresponds to $\chi_s \ll 1$ by virtue of relation (14.36):

$$\rho_{\text{hb}} \simeq \frac{2}{243\pi m^2} \left(1 + \frac{5}{4}\chi_s^2\right) + O(\chi_s^4). \quad (14.60)$$

We deduce from Eqs. (14.59)-(14.60) that

$$\lim_{\chi_s \rightarrow 0} \rho_{\text{hb}} = \frac{2}{243\pi m^2} \simeq 2.62 \cdot 10^{-3} m^{-2} \quad \text{and} \quad \lim_{\chi_s \rightarrow \pi/3} \rho_{\text{hb}} = \frac{81}{2048\pi m^2} \simeq 1.26 \cdot 10^{-2} m^{-2}.$$

Between these two limits, ρ_{hb} has a monotonic behavior, as shown in Fig. 14.7. In particular, one has $\rho_{\text{hb}} < 0.02 m^{-2}$. Hence

Property 14.6: matter density at the horizon's birth

By selecting a sufficiently large mass m , the proper energy density of matter when the black hole forms at the center of the collapsing star can be made arbitrarily small.

Example 7: Let us determine the mass m for which ρ_{hb} is as small as the density of water: $\rho_{\text{hb}} = 10^3 \text{ kg m}^{-3}$. By restoring the G 's and the c 's, the density m^{-2} becomes $c^6/G^3 m^{-2}$. We then get $m \simeq 4 \cdot 10^7 M_\odot$ for small initial compactness ($\chi_s \ll 1$) and $m \simeq 9 \cdot 10^7 M_\odot$ for large initial compactness ($\chi_s \sim \pi/3$).

The relative density increase at the horizon birth with respect to the initial value $\rho_0 = m/(4/3 \pi r_0^3)$ is obtained by setting $\eta = \pi - 3\chi_s$ in Eq. (14.39):

$$\frac{\rho_{\text{hb}}}{\rho_0} = \frac{8}{(1 - \cos(3\chi_s))^3}. \quad (14.61)$$

For small compactness, we get

$$\frac{\rho_{\text{hb}}}{\rho_0} \underset{\chi_s \rightarrow 0}{\sim} \frac{64}{729\chi_s^6}. \quad (14.62)$$

Hence ρ_{hb}/ρ_0 is diverging very rapidly when $\chi_s \rightarrow 0$. On the contrary, for χ_s large, ρ_{hb}/ρ_0 can be moderate. Even, for $\chi_s = \pi/3$, Eq. (14.61) leads to $\rho_{\text{hb}}/\rho_0 = 1$, in agreement with $\tau_{\text{hb}} = 0$ for that value of χ_s .

Example 8: Values of ρ_{hb} for the various objects considered in Examples 2, 4 and 6 are given in Table 14.1. These values are all larger than the nuclear density $\rho_{\text{nuc}} = 2 \cdot 10^{17} \text{ kg m}^{-3}$. For the Earth, ρ_{hb} is even $\sim 10^{12} \rho_{\text{nuc}}$, in agreement with the diverging behavior (14.62). On the contrary, for the neutron star, for which $\chi_s = 0.63$, we have only $\rho_{\text{hb}}/\rho_0 \simeq 3.7$. By comparing with Example 7, we note that all the examples considered in Table 14.1 have a mass much smaller to that required for a low value of ρ_{hb} .

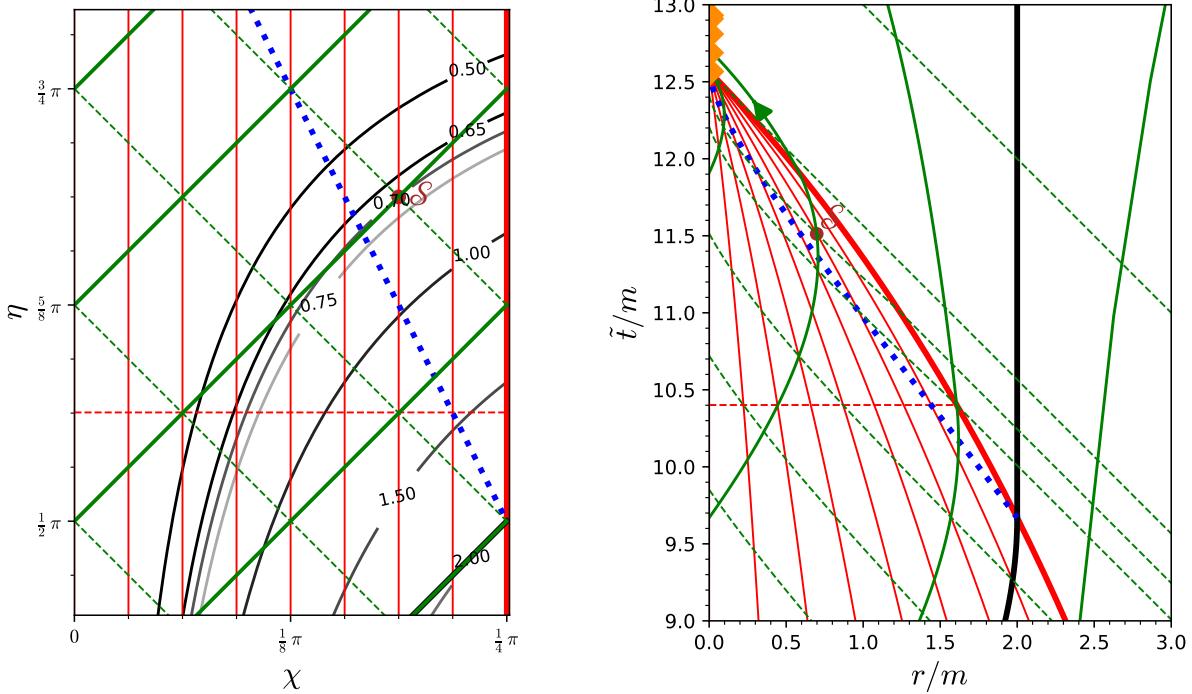


Figure 14.8: Spacetime diagrams of the Oppenheimer-Snyder collapse, as in Fig. 14.2, but zooming around the trapped sphere \mathcal{S} located at $(\eta, \chi) = (11\pi/16, 3\pi/16)$. The left diagram is drawn in terms of the conformal coordinates (η, χ) , while the right one is based on the Eddington-Finkelstein coordinates (\tilde{t}, r) . See the legend of Fig. 14.2 for the meaning of the various curves. [Figures generated by the notebook D.6.2]

14.3.6 Trapped surfaces

The concept of trapped surface has been introduced in Sec. 3.2.3 (cf. Fig. 3.1). Let us show that such surfaces appear at late times in the Oppenheimer-Snyder collapse. To this aim, we consider a 2-dimensional surface \mathcal{S} at constant coordinates (η, χ) inside the collapsing star, with $\eta \neq \pi$ and $\chi \neq 0$. Such a surface appears as a point in the spacetime diagrams of Fig. 14.8. The surface \mathcal{S} has the topology of a 2-sphere and is spanned by the coordinates (θ, φ) . In particular, it is a *closed* manifold (compact without boundary). The metric q induced by g on \mathcal{S} is readily obtained by setting $d\eta = 0$ and $d\chi = 0$ in Eq. (14.45):

$$q = \frac{a_0^2}{4}(1 + \cos \eta)^2 \sin^2 \chi (\mathrm{d}\theta^2 + \sin^2 \theta \mathrm{d}\varphi^2). \quad (14.63)$$

This is clearly a Riemannian metric, i.e. \mathcal{S} is a spacelike 2-surface of (\mathcal{M}, g) . At each point of \mathcal{S} , there are two null directions orthogonal to \mathcal{S} : the outgoing one, represented by a solid green line in Fig. 14.8 and the ingoing one, represented by a dashed green line (see also Fig. 2.10). Future-directed null vectors along these two directions are respectively

$$\ell = \partial_\eta + \partial_\chi \quad \text{and} \quad k = \partial_\eta - \partial_\chi. \quad (14.64)$$

The expansions of \mathcal{S} along ℓ and k are given by Eq. (2.55) [see also Eq. (3.6) and Fig. 3.1]:

$$\theta_{(\ell)} = \mathcal{L}_\ell \ln \sqrt{q} \quad \text{and} \quad \theta_{(k)} = \mathcal{L}_k \ln \sqrt{q}, \quad (14.65)$$

where q is the determinant of the metric q with respect to any coordinate system of \mathcal{S} that is carried along with ℓ or k , such that (θ, φ) , since θ and φ are constant along the integral curves of ℓ and k (the radial null geodesics). From Eq. (14.63), we get $\sqrt{q} = (a_0^2/4)(1+\cos\eta)^2 \sin^2\chi \sin\theta$. Furthermore, in view of the components (14.64), $\mathcal{L}_\ell \ln \sqrt{q} = \ell^\mu \partial \ln \sqrt{q}/\partial x^\mu = \partial \ln \sqrt{q}/\partial \eta + \partial \ln \sqrt{q}/\partial \chi$ and $\mathcal{L}_k \ln \sqrt{q} = k^\mu \partial \ln \sqrt{q}/\partial x^\mu = \partial \ln \sqrt{q}/\partial \eta - \partial \ln \sqrt{q}/\partial \chi$. We then get easily

$$\theta_{(\ell)} = 4 \frac{\cos(\eta/2) \cos(\chi + \eta/2)}{(1 + \cos\eta) \sin\chi} \quad \text{and} \quad \theta_{(k)} = -4 \frac{\cos(\eta/2) \cos(\chi - \eta/2)}{(1 + \cos\eta) \sin\chi}. \quad (14.66)$$

Given that $\eta \in [0, \pi]$ and $\chi \in (0, \chi_s]$, with $\chi_s < \pi/2$, one has $1 + \cos\eta > 0$, $\cos(\eta/2) > 0$, $\sin\chi > 0$ and $\cos(\chi - \eta/2) > 0$. It follows immediately that

$$\theta_{(k)} < 0 \quad (14.67)$$

and the sign of $\theta_{(\ell)}$ is that of $\cos(\chi + \eta/2)$. Given that the above ranges of η and χ imply $0 < \chi + \eta/2 < \pi$, we have $\theta_{(\ell)} < 0 \iff \chi + \eta/2 > \pi/2$, i.e.

$$\theta_{(\ell)} < 0 \iff \eta > \pi - 2\chi. \quad (14.68)$$

Since \mathcal{S} is a closed spacelike 2-surface, we conclude, in view of the definitions given in Sec. 3.2.3:

Property 14.7: trapped surfaces

At late times, i.e. for $\eta > \pi - 2\chi$, the spheres \mathcal{S} defined by $(\eta, \chi) = \text{const}$, with $\eta < \pi$ and $\chi > 0$, are trapped surfaces ($\theta_{(\ell)} < 0$ and $\theta_{(k)} < 0$). These spheres are marginally trapped ($\theta_{(\ell)} = 0$ and $\theta_{(k)} < 0$) for $\eta = \pi - 2\chi$ and untrapped ($\theta_{(\ell)} > 0$ and $\theta_{(k)} < 0$) for $\eta < \pi - 2\chi$.

Example 9: The trapped surface \mathcal{S} corresponding to $(\eta, \chi) = (11\pi/16, 3\pi/16)$ in the Oppenheimer-Snyder collapse with $r_0 = 4m$ is plotted as a brown dot in Fig. 14.8. According to Eq. (14.34), the areal radius of \mathcal{S} is $r_{\mathcal{S}} = 2\sqrt{2} \sin(3\pi/16)(1 + \cos(11\pi/16)) \simeq 0.70\text{ m}$. The trapped character of \mathcal{S} can be read on Fig. 14.8: moving to the future along either the ingoing radial null geodesic (dashed green line) or the outgoing one (solid green line) leads to spheres of smaller areal radius, and hence smaller area. This is clear on the left diagram when noticing that both geodesics encounter the curve $r = 0.65\text{ m}$ in the immediate future of \mathcal{S} . It is even more immediate on the right diagram since r is read directly on the x -axis.

The set formed by all the marginally trapped spheres of fixed (η, χ) is the hypersurface \mathcal{T} defined by $\theta_{(\ell)} = 0$. Its equation is obtained by saturating the inequalities in Eq. (14.68):

$$\mathcal{T} : \quad \eta = \pi - 2\chi. \quad (14.69)$$

As we shall see in Chap. 18, \mathcal{T} is called a **future inner trapping horizon**. It is drawn as a dotted blue line in Figs. 14.2 and 14.8. Since the slope of that line is -2 and (η, χ) are conformal coordinates, it is clear from the left diagrams in these figures that \mathcal{T} is a timelike hypersurface. This is actually a generic feature of future inner trapping horizons, as we shall see in Chap. 18.

We note from Figs. 14.2 and 14.8 that \mathcal{T} forms at the surface of the collapsing star at the very same instant where the event horizon \mathcal{H} reaches the surface. Indeed, searching for the intersection $\mathcal{H} \cap \mathcal{T}$ from the equations defining the hypersurfaces \mathcal{H} and \mathcal{T} , i.e. Eqs. (14.56) and (14.69), we get $(\eta, \chi) = (\pi - 2\chi_s, \chi_s)$, which is the 2-sphere representing the surface of the collapsing star at $\eta = \pi - 2\chi_s$.

The trapped surfaces \mathcal{S} are located “above” \mathcal{T} , i.e. at $\eta > \pi - 2\chi$. The region containing trapped 2-spheres with $r = \text{const}$ actually extend through the collapsing star’s surface up to the event horizon \mathcal{H} . Indeed, the spacetime there is a piece of the region \mathcal{M}_{II} of Schwarzschild spacetime (cf. Sec. 6.2.5), where every 2-sphere at $\tilde{t} = \text{const}$ and $r = \text{const}$ is trapped (this is obvious since both ingoing and outgoing radial null geodesics have decreasing areal radius r towards the future in \mathcal{M}_{II} , cf. Fig. 6.3). Outside \mathcal{H} , i.e. in the \mathcal{M}_{I} region of Schwarzschild spacetime, there is no trapped surface.

Remark 7: The trapping horizon \mathcal{T} is not the boundary of the region where there exist trapped surfaces in the Oppenheimer-Snyder collapse, but only the boundary for trapped 2-spheres \mathcal{S} centered around $\chi = 0$. Indeed, since the collapsing stellar interior is homogeneous, the trapped 2-spheres \mathcal{S} can be translated in χ at fixed η , leading to other trapped surfaces. One can then show that such “translated” trapped 2-spheres exist in all the region $\eta > \pi - \chi_s - \chi$, which extends to the past of \mathcal{T} , since for $\chi \in [0, \chi_s]$, $\pi - \chi_s - \chi < \pi - 2\chi$. Furthermore, there exist more complicated trapped surfaces even below this limit, see Ref. [55] for details.

14.4 Observing the black hole formation

Let us consider a remote static observer \mathcal{O} who receives electromagnetic radiation from the surface of a spherically symmetric collapsing star. By *static*, it is meant that the worldline of \mathcal{O} is a field line of the static Killing vector $\xi = \partial_t = \partial_{\tilde{t}}$ in the exterior Schwarzschild region [cf. Eq. (7.28)]. In particular, \mathcal{O} evolves at a fixed value of $(r_{\mathcal{O}}, \theta_{\mathcal{O}}, \varphi_{\mathcal{O}})$ of the coordinates (r, θ, φ) (cf. Fig. 14.9; see also Figs. 14.4 and 14.5). Without any loss of generality, we choose $\theta_{\mathcal{O}} = \pi/2$ and $\varphi_{\mathcal{O}} = 0$. Furthermore, we assume that \mathcal{O} is *remote*: $r_{\mathcal{O}} \gg m$.

14.4.1 Apparent collapse freezing

Let us suppose that, at the matter comoving proper time τ_1 , an electromagnetic signal (“flash”) is emitted in the radial direction from the surface of the collapsing star at $\theta = \pi/2$ and $\varphi = 0$ (i.e. towards observer \mathcal{O}) and that a similar signal is emitted subsequently at proper time τ_2 . Since each signal is carried by an outgoing radial null geodesic (the green curves in Fig. 14.9), the equation of which is (6.51) with respect to the IEF coordinates, we have

$$\tilde{t}_i^{\text{rec}} - r_{\mathcal{O}} - 4m \ln \left(\frac{r_{\mathcal{O}}}{2m} - 1 \right) = \tilde{t}^{\text{em}}(\tau_i) - r_s(\tau_i) - 4m \ln \left(\frac{r_s(\tau_i)}{2m} - 1 \right), \quad i \in \{1, 2\},$$

where \tilde{t}_i^{rec} is the IEF coordinate \tilde{t} of the reception event of signal no. i by observer \mathcal{O} and $\tilde{t}^{\text{em}}(\tau_i)$ is the IEF coordinate \tilde{t} of the emission event taking place at proper time τ_i ; $\tilde{t}^{\text{em}}(\tau_i)$ is given by Eq. (14.31) with η being the function of τ_i defined implicitly by Eq. (14.30). We deduce from the

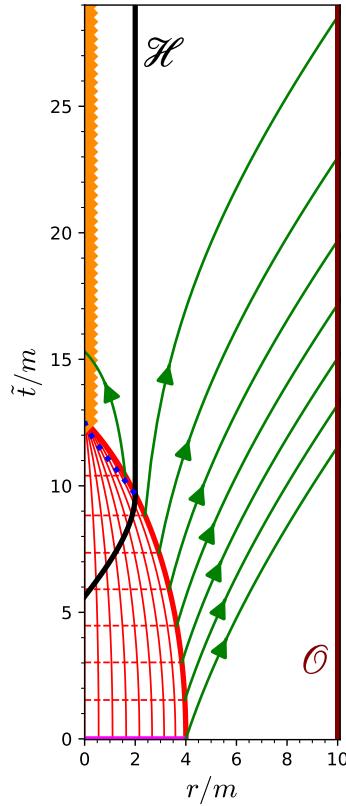


Figure 14.9: Spacetime diagram of the Oppenheimer-Snyder collapse, as in the right panel of Fig. 14.2, illustrating the increasing delay in the reception by a remote static observer \mathcal{O} (maroon worldline at $r = 10 m$) of signals emitted radially from the surface of the collapsing star with a constant interval of matter proper time, $\delta\tau = \tau_{\text{end}}/8$. See the caption of Fig. 14.2 for the legend of the various curves. [Figure generated by the notebook D.6.2]

above relation the \tilde{t} -interval separating the receptions of signal 1 and 2 by \mathcal{O} :

$$\tilde{t}_2^{\text{rec}} - \tilde{t}_1^{\text{rec}} = \tilde{t}^{\text{em}}(\tau_2) - \tilde{t}^{\text{em}}(\tau_1) + r_s(\tau_1) - r_s(\tau_2) + 4m \ln \left(\frac{r_s(\tau_1) - 2m}{r_s(\tau_2) - 2m} \right). \quad (14.70)$$

Since \mathcal{O} is remote ($r_{\mathcal{O}} \gg m$), $\tilde{t}_2^{\text{rec}} - \tilde{t}_1^{\text{rec}}$ is nothing but the lapse of \mathcal{O} 's proper time between the two reception events. The quantities $\tilde{t}^{\text{em}}(\tau_2) - \tilde{t}^{\text{em}}(\tau_1)$ and $r_s(\tau_1) - r_s(\tau_2)$ are always finite; on the other side, the logarithm term diverges when $r_s(\tau_2)$ approaches $2m$, i.e. when τ_2 approaches τ_* [cf. Eq. (14.55)], so that

$$\tilde{t}_2^{\text{rec}} - \tilde{t}_1^{\text{rec}} \rightarrow +\infty \quad \text{when} \quad \tau_2 \rightarrow \tau_*. \quad (14.71)$$

In particular, if a sequence of signals is emitted from the star's surface at a constant rate in terms of the comoving proper time τ , the signals are received by the remote observer \mathcal{O} with an increasing delay as the collapse proceeds. The delay between two successive signals becoming infinite, there exists a very last signal received by \mathcal{O} . All subsequent signals of the emitted sequence do not cross the event horizon, i.e. are captured into the black hole.

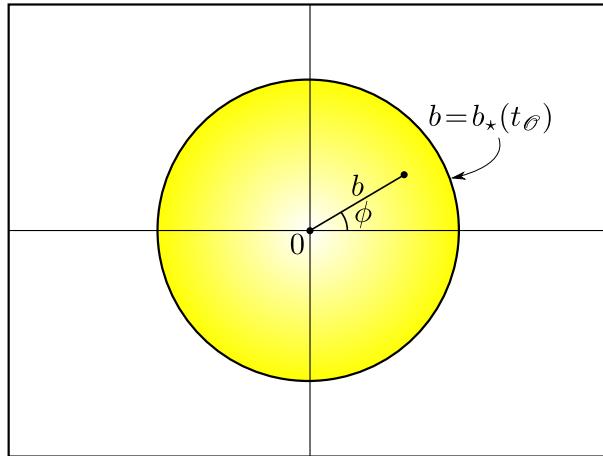


Figure 14.10: Image of the collapsing star on the screen of the remote observer \mathcal{O} , at a given instant $t_{\mathcal{O}}$.

Example 10: Consider Fig. 14.9, which depicts the Oppenheimer-Snyder collapse with $r_0 = 4m$, or equivalently $\chi_s = \pi/4$. Starting from $\tau = 0$, a uniform sequence of 8 signals is emitted from the star's surface, the proper time separation between two successive signals being $\delta\tau = \tau_{\text{end}}/8 = \pi\sqrt{2}m/4$ (cf. Example 1 on p. 559). One sees graphically that the signals are received by \mathcal{O} with an increasing separation $\delta\tilde{t}$. The last received signal is the 7th one. The 8th signal is trapped in the black hole region and ends on the curvature singularity.

We conclude from the above analysis:

Property 14.8: apparent freeze of the collapse

For a remote static observer, the collapse, as monitored by the reception of signals regularly emitted from the star's surface, appears to slow down and even to freeze completely when the areal radius of the star tends towards $2m$. In particular, the remote observer never sees the star's surface crossing the event horizon and does not perceive at all the final singularity.

Historical note : The above behavior explains why the name *frozen star* was given to black holes before the name *black hole* itself was coined (cf. the historical note on p. 109).

14.4.2 Shape and size of the images of the collapse

The image of the collapsing star is formed on the screen of observer's \mathcal{O} by photons emitted from the star's surface and following null geodesics of Schwarzschild spacetime (the spacetime exterior to the star). In Sec. 14.4.1, we have considered only radial null geodesics. Discussing the image requires to consider non-radial null geodesics as well. Such geodesics have been studied in detail in Chap. 8. In particular, it has been shown that they are essentially characterized by a single parameter: the impact parameter $b := |L|/E$ introduced in Sec. 8.2.3. Radial geodesics are those for which $b = 0$. In what follows, we shall work in terms of the Schwarzschild-Droste

coordinates (t, r, θ, φ) , as in Chap. 8. These coordinates are regular in the Schwarzschild exterior where the null geodesics propagate to the remote observer \mathcal{O} and they make computations slightly simpler than IEF coordinates do.

Due to spherical symmetry, the image of the star at a given instant $t_{\mathcal{O}}$ of observer \mathcal{O} 's proper time is a disk and is invariant under rotation around the disk center. Since \mathcal{O} is assumed to lie far away in the weak-field region, $t_{\mathcal{O}}$ coincides with the Schwarzschild-Droste time coordinate at the location of \mathcal{O} . As discussed in Sec. 8.5.1, the distance of a screen pixel from the disk center is proportional to the impact parameter b of the null geodesic carrying the photon hitting this pixel. Accordingly, all the properties of the image (redshift and intensity at a given point) are functions of the pair $(t_{\mathcal{O}}, b)$, where b can be identified with the radial coordinate in the screen plane, centered on the image (cf. Fig. 14.10). The pixel at the disk center corresponds to a null geodesic having $b = 0$, i.e. traveling radially in the direction $\theta = \theta_{\mathcal{O}} = \pi/2$ and $\varphi = \varphi_{\mathcal{O}} = 0$. The image's rim, i.e. the circle bounding the disk, is located at some maximal value of b , $b_{\star}(t_{\mathcal{O}})$ say. A great amount of information on $b_{\star}(t_{\mathcal{O}})$ can be inferred from the effective potential analysis developed in Sec. 8.2.3. Indeed, along a given null geodesic \mathcal{L} of impact parameter b , the following constraint must hold in the exterior of the star:

$$b \leq \frac{1}{\sqrt{U(r)}} = \frac{r}{\sqrt{1 - 2m/r}}, \quad (14.72)$$

where $U(r)$ is the effective potential (8.14): $U(r) := (1 - 2m/r)/r^2$. This follows directly from $(dr/d\tilde{\lambda})^2 \geq 0$ in Eq. (8.11). Prior to the collapse, let us assume that the surface of the star is static and located at $r = r_0$. Let us then denote by $t_{\mathcal{O}}^{**}$ the value of $t_{\mathcal{O}}$ at which the disk image starts to shrink. This corresponds to null geodesics emitted from the star's surface at the matter proper time $\tau = 0$ arriving at the disk boundary. For $t_{\mathcal{O}} \leq t_{\mathcal{O}}^{**}$, one has then $r \geq r_0$ all along \mathcal{L} . If we assume $r_0 > 3m$, then $1/\sqrt{U(r)}$ is increasing with r all along \mathcal{L} , given that $U(r)$ is a decreasing function on the interval $(3m, +\infty)$ (cf. Fig. 8.1). The constraint (14.72) is then equivalent to $b \leq 1/\sqrt{U(r_0)}$, so that the maximal value of b is $b_0 := 1/\sqrt{U(r_0)}$ and we can write:

$$b_{\star}(t_{\mathcal{O}}) = b_0 := \frac{r_0}{\sqrt{1 - 2m/r_0}} \quad \text{for} \quad t_{\mathcal{O}} \leq t_{\mathcal{O}}^{**} \quad \text{and} \quad r_0 > 3m. \quad (14.73)$$

For $t_{\mathcal{O}} \leq t_{\mathcal{O}}^{**}$, a null geodesic at the image's rim, i.e. having $b = b_{\star}(t_{\mathcal{O}})$, fulfills thus $U(r_0) = b^{-2}$, which implies $dr/d\tilde{\lambda} = 0$ at $r = r_0$ according to Eq. (8.11). This means that the emission point ($r = r_0$) is a periastron of the geodesic: it corresponds to the minimal value of r along \mathcal{L} , even if the latter is extended infinitely far in the past. This implies that \mathcal{L} is tangent to the hypersurface $r = r_0$ at the emission point, i.e. that the photon is emitted in a direction tangent to the star's surface.

Remark 1: In terms of the compactness angle χ_s defined by Eq. (14.36), formula (14.73) takes a simple form:

$$b_{\star}(t_{\mathcal{O}}) = r_0 \cos \chi_s \quad \text{for} \quad t_{\mathcal{O}} \leq t_{\mathcal{O}}^{**} \quad \text{and} \quad r_0 > 3m. \quad (14.74)$$

In the non-relativistic limit, $\chi_s \rightarrow 0$ and we get $b_{\star}(t_{\mathcal{O}}) = r_0$, as expected (recall that the truly observed quantity through a telescope is the small angle $\hat{b} = b/r_{\mathcal{O}}$ [Eq. (8.118)], so that $\hat{b} = r_0/r_{\mathcal{O}}$ at the disk's rim).

One can also use the effective potential analysis to get the value of $b_*(t_\mathcal{O})$ at late times, more precisely for images formed by photons emitted after the collapsing star has shrunk below the *photon sphere* introduced in Sec. 8.2.3, which corresponds to $r_s(\tau) < 3m$. Since the function $r \mapsto 1/\sqrt{U(r)}$ is decreasing on $(2m, 3m)$, reaches a minimum at $r = 3m$ and increases on $(3m, +\infty)$ (cf. the plot of $U(r)$ in Fig. 8.1), the constraint (14.72) becomes $b \leq \min(1/\sqrt{U(r)}) = b_c$, where $b_c = 3\sqrt{3} m$ [Eq. (8.21)]. An example of photon emitted at $r < 3m$ and fulfilling this constraint is represented by the trajectory no. 3 in Fig. 8.1. Furthermore, the constraint is actually a sufficient condition for any photon emitted outward to reach infinity (cf. Fig. 8.1 once again). However, for b close to b_c , photons perform a large number of orbits very close to the photon sphere, this number being actually infinite for $b = b_c$ (critical null geodesics), as discussed in Sec. 8.3.2 (cf. Figs. 8.7 to 8.9). As a consequence, the elapsed time coordinate t along the corresponding null geodesic diverges as $b \rightarrow b_c$, so that the image boundary $b_*(t_\mathcal{O}) = b_c$ is reached only for $t_\mathcal{O} \rightarrow +\infty$:

$$\lim_{t_\mathcal{O} \rightarrow +\infty} b_*(t_\mathcal{O}) = 3\sqrt{3} m. \quad (14.75)$$

However, one can show (see e.g. Ref. [523] for details) that $b_*(t_\mathcal{O})$ converges to b_c exponentially fast: $b_*(t_\mathcal{O}) - b_c \propto \exp(-t_\mathcal{O}/b_c)$.

To summarize:

Property 14.9: apparent size of the collapsing star

- If the initial star is not ultracompact, i.e. if its areal radius r_0 is larger than $3m$, the radius of the disk image shrinks from $b_0 = r_0/\sqrt{1 - 2m/r_0}$ [Eq. (14.73)] to $b_\infty := 3\sqrt{3} m$ [Eq. (14.75)]. The latter value is reached only asymptotically, i.e. for $t_\mathcal{O} \rightarrow +\infty$, but the convergence is exponentially fast, so that, $b_*(t_\mathcal{O}) \simeq 3\sqrt{3} m$ is a very good approximation for $t_\mathcal{O} > t_\mathcal{O}^{**} + \alpha m$, where $\alpha \sim 2b_c/m \sim 10$.
- Collapsing stars that are initially ultracompact ($r_0 < 3m$) generate images of constant size: the radius of the disk image remains fixed at $b_* = 3\sqrt{3} m$.

Remark 2: The above results are not specific to pressureless collapse: they hold for any spherically symmetric gravitational collapse.

Example 11: For our favorite example with $r_0 = 4m$ (Figs. 14.2, 14.3, 14.8 and 14.9), the radius of the image disk evolves from $4\sqrt{2} m \simeq 5.66 m$ to $3\sqrt{3} m \simeq 5.20 m$, which constitutes a pretty modest decrease: only 8%. This is of course because such a star has a large compactness. For the neutron star considered in Table 14.1, which has a smaller compactness: $m/r_0 = 0.17$, the shrinking factor is higher: 28%. For a concrete example of images of the gravitational collapse of an unstable neutron star into a black hole computed by means of numerical relativity, see Fig. 4 of Ref. [491], where the modest decrease of the disk image is very apparent.

The photons that form the star's image on \mathcal{O} 's screen at a given instant $t_\mathcal{O}$ have impact parameters b in the range $0 \leq b \leq b_*(t_\mathcal{O})$ and have been emitted from the star's surface at various proper times τ , depending on the value of b . Indeed, let us evaluate the elapsed

Schwarzschild-Droste time t between along a null geodesic \mathcal{L} of impact parameter b between the photon emission at the star surface and its reception by \mathcal{O} . Thanks to spherical symmetry, it suffices to consider photons traveling in the plane $\theta = \pi/2$. The geodesic equations of motion are then (8.9)-(8.11), from which we get

$$\frac{dt}{dr} \Big|_{\mathcal{L}} = \epsilon_r \left(1 - \frac{2m}{r}\right)^{-1} (1 - b^2 U(r))^{-1/2} \quad (14.76a)$$

$$\frac{d\varphi}{dr} \Big|_{\mathcal{L}} = \epsilon_r \epsilon_L \frac{b}{r^2} (1 - b^2 U(r))^{-1/2}, \quad (14.76b)$$

where $\epsilon_L = \pm 1$ indicates the sign of the conserved angular momentum L of \mathcal{L} and $\epsilon_r = \pm 1$ is $+1$ (resp. -1) on portions of \mathcal{L} where r increases (resp. decreases) towards the future. Note that ϵ_L is constant along \mathcal{L} . Regarding ϵ_r , it has to be $+1$ in the vicinity of observer \mathcal{O} . Given that a null geodesic of Schwarzschild spacetime has at most one r -turning point (cf. Sec. 8.2.4), one has either $\epsilon_r = +1$ along all \mathcal{L} or $\epsilon_r = -1$ just around the emission point, switching to $+1$ after some periastron passage.

Let \mathcal{E} be the event of emission at the star's surface of a photon that reaches \mathcal{O} 's screen at the radial position b at the instant $t_{\mathcal{O}}$. Let us denote by respectively $t_{\text{em}}(b, t_{\mathcal{O}})$, $\tau_{\text{em}}(b, t_{\mathcal{O}})$ and $r_s^{\text{em}}(b, t_{\mathcal{O}})$, the Schwarzschild-Droste time coordinate of \mathcal{E} , the infalling matter proper time of \mathcal{E} and the areal coordinate r of \mathcal{E} . For a pressureless collapse, $t_{\text{em}}(b, t_{\mathcal{O}})$, $\tau_{\text{em}}(b, t_{\mathcal{O}})$ and $r_s^{\text{em}}(b, t_{\mathcal{O}})$ are related by the free-fall equations (7.41) [or equivalently (14.30)] and (7.42):

$$t_{\text{em}}(b, t_{\mathcal{O}}) = 2m \left\{ \sqrt{\frac{r_0}{2m} - 1} \left[\eta + \frac{r_0}{4m} (\eta + \sin \eta) \right] + \ln \left| \frac{\sqrt{\frac{r_0}{2m} - 1} + \tan \frac{\eta}{2}}{\sqrt{\frac{r_0}{2m} - 1} - \tan \frac{\eta}{2}} \right| \right\} \quad (14.77a)$$

$$\tau_{\text{em}}(b, t_{\mathcal{O}}) = \sqrt{\frac{r_0^3}{8m}} (\eta + \sin \eta) \quad (14.77b)$$

$$r_s^{\text{em}}(b, t_{\mathcal{O}}) = \frac{r_0}{2} (1 + \cos \eta). \quad (14.77c)$$

Note that we have chosen the origin of the Schwarzschild-Droste coordinate t to coincide with the start of the collapse: $t = 0 \iff \tau = 0 \iff \eta = 0$.

In terms of the star's areal radius function $r_s(\tau)$, one has of course $r_s^{\text{em}}(b, t_{\mathcal{O}}) = r_s(\tau_{\text{em}}(b, t_{\mathcal{O}}))$. The above relations are valid for $\tau_{\text{em}}(b, t_{\mathcal{O}}) \geq 0$, i.e. after the collapse has started. For $\tau_{\text{em}}(b, t_{\mathcal{O}}) < 0$, we have, instead of (14.77),

$$t_{\text{em}}(b, t_{\mathcal{O}}) = \left(1 - \frac{2m}{r_0}\right)^{-1/2} \tau_{\text{em}}(b, t_{\mathcal{O}}) < 0 \quad (14.78a)$$

$$r_s^{\text{em}}(b, t_{\mathcal{O}}) = r_0. \quad (14.78b)$$

By integrating Eq. (14.76a) between the emission and reception points, we get

$$t_{\mathcal{O}} - t_{\text{em}}(b, t_{\mathcal{O}}) = \int_{r_s^{\text{em}}(b, t_{\mathcal{O}})}^{r_{\mathcal{O}}} \left(1 - \frac{2m}{r}\right)^{-1} (1 - b^2 U(r))^{-1/2} dr + \Delta, \quad (14.79)$$

where $\Delta = 0$ if there is no r -turning point (periastron) along the photon geodesic \mathcal{L} and

$$\Delta = 2 \int_{r_p}^{r_s^{\text{em}}(b, t_\theta)} \left(1 - \frac{2m}{r}\right)^{-1} (1 - b^2 U(r))^{-1/2} dr \quad (14.80)$$

if there is one, located at $r = r_p < r_s^{\text{em}}(b, t_\theta)$. For a given observation instant t_θ and a given position on the screen, measured by b , Eq. (14.79) must be solved to get $t_{\text{em}}(b, t_\theta)$. In general, this is a difficult problem since the lower boundary of the integral, $r_s^{\text{em}}(b, t_\theta)$, depends on $t_{\text{em}}(b, t_\theta)$ in a complicated way, via the parametric system (14.77). However, for $b = 0$ (radial geodesic), one can compute the integral in Eq. (14.79) in an elementary way; taking into account that $\Delta = 0$ for $b = 0$ (no r -turning point along a radial geodesic), we obtain

$$t_{\text{em}}(0, t_\theta) = t_\theta - r_\theta + r_s^{\text{em}}(0, t_\theta) + 2m \ln \left(\frac{r_s^{\text{em}}(0, t_\theta) - 2m}{r_\theta - 2m} \right). \quad (14.81)$$

This equation must still be solved to get $t_{\text{em}}(0, t_\theta)$ since $r_s^{\text{em}}(0, t_\theta)$ is a function of $t_{\text{em}}(0, t_\theta)$ given parametrically by the system (14.77).

Another simplification of the general formula (14.79) occurs for $t_{\text{em}}(b, t_\theta) < 0$ (emission before the start of the collapse), since then $r_s^{\text{em}}(b, t_\theta) = r_0$ [Eq. (14.78b)]. Equation (14.79) provides then directly the value of $t_{\text{em}}(b, t_\theta)$ and we see that this is a decreasing function of b . Indeed, at fixed r , $b \mapsto (1 - b^2 U(r))^{-1/2}$ is an increasing function of b , given that $U(r) > 0$. Consequently, the first value of b for which $t_{\text{em}}(b, t_\theta) \geq 0$, i.e. for which the collapse starts to be perceptible, is $b = 0$. In other words, the collapse manifests itself first at the center of the disk image. Let us then denote by t_θ^* the corresponding value of t_θ . It is obtained by setting $t_{\text{em}}(0, t_\theta) = 0$ and $r_s^{\text{em}}(0, t_\theta) = r_0$ in Eq. (14.81):

$$t_\theta^* = r_\theta - r_0 + 2m \ln \left(\frac{r_\theta - 2m}{r_0 - 2m} \right). \quad (14.82)$$

On the other side, the value of t_θ^{**} , defined above as the instant when the disk image starts to shrink, is obtained by setting $b = b_0$ [the value (14.73)], $t_{\text{em}}(b, t_\theta) = 0$ and $r_s^{\text{em}}(b, t_\theta) = r_0$ in Eq. (14.79):

$$t_\theta^{**} = \int_{r_0}^{r_\theta} \left(1 - \frac{2m}{r}\right)^{-1} (1 - b_0^2 U(r))^{-1/2} dr. \quad (14.83)$$

Note that we have set $\Delta = 0$ in Eq. (14.79) because the geodesic delimiting the image's rim is precisely tangent to the surface in the static (pre-collapse) case, which implies that the photon's periastron coincides with the emission point.

By subtracting from Eq. (14.83) the integral expression of t_θ^* deduced from Eq. (14.79) by setting $b = 0$, $t_{\text{em}}(b, t_\theta) = 0$ and $r_s^{\text{em}}(b, t_\theta) = r_0$, we get, after substitution of Eq. (8.14) for $U(r)$,

$$t_\theta^{**} - t_\theta^* = b_0^2 \int_{r_0}^{r_\theta} \frac{dr}{r^2 - b_0^2(1 - 2m/r) + r\sqrt{r^2 - b_0^2(1 - 2m/r)}}. \quad (14.84)$$

Since $r^2 - b_0^2(1 - 2m/r) \geq 0$ for $r \geq r_0$, this expression shows that the time delay $t_\theta^{**} - t_\theta^*$ is positive.

14.4.3 Redshift in the image

Let \mathcal{O}_{em} be an observer at the surface of the star comoving with the infalling matter. In particular the proper time of \mathcal{O} is τ and the r -coordinate of \mathcal{O}_{em} is $r_s(\tau)$, as given by Eq. (14.30) for a pressureless collapse. One defines the **redshift** z of a photon hitting the remote observer screen's \mathcal{O} by $z := (\lambda_\infty - \lambda_{\text{em}})/\lambda_{\text{em}}$, where λ_{em} is the photon wavelength at emission, i.e. with respect to observer \mathcal{O}_{em} , and λ_∞ is the photon wavelength measured by \mathcal{O} . Thanks to the Planck-Einstein relation (1.39): $E = h\nu = h/(c\lambda)$, we have

$$1 + z = \frac{E_{\text{em}}}{E_\infty}, \quad (14.85)$$

where E_{em} (resp. E_∞) is the photon energy with respect to observer \mathcal{O}_{em} (resp. \mathcal{O}).

Frame of the comoving surface observer

The 4-velocity of the surface observer \mathcal{O}_{em} is $\mathbf{u}_s = u_s^t \partial_t + u_s^r \partial_r$. In any pressureless collapse from rest, like the Oppenheimer-Snyder one, \mathcal{O}_{em} is in radial free fall from rest, so that $u_s^t = dt/d\tau$ is given by Eq. (7.34) with $\varepsilon = \sqrt{1 - 2m/r_0}$ [Eq. (7.40)] and $u_s^r = dr/d\tau$ is given by Eq. (7.36) with $\varepsilon^2 - 1 = -2m/r_0$. Hence we get

$$\mathbf{u}_s = \sqrt{1 - \frac{2m}{r_0}} \left(1 - \frac{2m}{r_s}\right)^{-1} \partial_t - \sqrt{\frac{2m}{r_s} - \frac{2m}{r_0}} \partial_r, \quad (14.86)$$

where r_s is a shortcut for $r_s(\tau)$. It is useful to expand \mathbf{u}_s onto the frame of the local static observer whose position coincides with the star surface at a given instant. The orthonormal frame of the static observer is

$$\mathbf{e}_{(t)} = \left(1 - \frac{2m}{r}\right)^{-1/2} \partial_t, \quad \mathbf{e}_{(r)} = \left(1 - \frac{2m}{r}\right)^{1/2} \partial_r, \quad \mathbf{e}_{(\theta)} = \frac{1}{r} \partial_\theta, \quad \mathbf{e}_{(\varphi)} = \frac{1}{r \sin \theta} \partial_\varphi. \quad (14.87)$$

We deduce then from (14.86) that

$$\mathbf{u}_s = \Gamma (\mathbf{e}_{(t)} - V \mathbf{e}_{(r)}), \quad (14.88)$$

with

$$V := \sqrt{\frac{\frac{2m}{r_s} - \frac{2m}{r_0}}{1 - \frac{2m}{r_0}}} \quad \text{and} \quad \Gamma := \frac{1}{\sqrt{1 - V^2}} = \sqrt{\frac{1 - \frac{2m}{r_0}}{1 - \frac{2m}{r_s}}} \quad (14.89)$$

By comparing with the generic formulas (1.35)-(1.36), we conclude that V is the norm of the velocity $\mathbf{V} = -V \mathbf{e}_{(r)}$ of \mathcal{O}_{em} with respect to the local static observer and Γ is the corresponding Lorentz factor.

Remark 3: Formula (14.89) implies $\lim_{r_s \rightarrow 2m} V = 1$. We recover then a well known result: with respect to a static observer located just outside the event horizon, a radially free-falling body is falling at the speed of light.

Observer \mathcal{O}_{em} is endowed with an orthonormal frame $(\mathbf{u}_s, \mathbf{n}_s, \mathbf{e}_2^s, \mathbf{e}_3^s)$, where $\mathbf{n}_s = n_s^{(t)} \mathbf{e}_{(t)} + n_s^{(r)} \mathbf{e}_{(r)}$ is a unit spacelike vector normal to the star surface and oriented towards the exterior,

while the vectors (e_2^s, e_3^s) are unit spacelike vectors tangent to the star surface. The components $(n_s^{(t)}, n_s^{(r)})$ of \mathbf{n}_s are obtained from the two conditions $\mathbf{u}_s \cdot \mathbf{n}_s = 0$ and $\mathbf{n}_s \cdot \mathbf{n}_s = 1$, along with (14.88):

$$\mathbf{n}_s = \Gamma (-V \mathbf{e}_{(t)} + \mathbf{e}_{(r)}) . \quad (14.90)$$

Remark 4: Equations (14.88) and (14.90) simply express that the orthonormal pairs $(\mathbf{u}_s, \mathbf{n}_s)$ and $(\mathbf{e}_{(t)}, \mathbf{e}_{(r)})$ are connected by a *Lorentz boost* of velocity parameter $-V$.

General formula for the redshift

Let us consider a photon emitted from the surface of the collapsing star. The photon's energy-momentum vector \mathbf{p} can be decomposed with respect to \mathcal{O}_{em} 's frame according to Eqs. (1.31) and (1.30):

$$\mathbf{p} = E_{\text{em}} \mathbf{u}_s + \mathbf{P}_{\text{em}}, \quad (14.91)$$

where \mathbf{P}_{em} is a vector orthogonal to \mathbf{u}_s representing the photon's linear momentum as measured by \mathcal{O}_{em} . The energy E_{em} is then obtained as [cf. Eq. (1.22)]

$$E_{\text{em}} = -\mathbf{u}_s \cdot \mathbf{p} \quad (14.92)$$

Due to spherical symmetry, we may restrict the study to photons travelling in the equatorial plane $\theta = \pi/2$. The photon's energy-momentum vector \mathbf{p} is then given by Eq. (8.121), which involves the impact parameter b . Once re-expressed in terms of the static observer frame (14.87), this formula becomes

$$\mathbf{p} = E_\infty \left[\left(1 - \frac{2m}{r} \right)^{-1/2} \left(\mathbf{e}_{(t)} + \epsilon_r \sqrt{1 - b^2 U(r)} \mathbf{e}_{(r)} \right) + \epsilon_L \frac{b}{r} \mathbf{e}_{(\varphi)} \right], \quad (14.93)$$

where $\epsilon_r = \pm 1$ and $\epsilon_L = \pm 1$. Far from the star, one has of course $\epsilon_r = +1$, but, as we shall see below, $\epsilon_r = -1$ may occur close to the star, which corresponds to a photon emitted in the inward direction (decreasing r); in this case, the photon will reach a periastron, where ϵ_r is switched to $+1$.

The scalar product in Eq. (14.92) is easily evaluated from expressions (14.88) and (14.93), owing to the orthonormality of the frame $(\mathbf{e}_{(\alpha)})$; one gets

$$1 + z = \frac{E_{\text{em}}}{E_\infty} = \Gamma \left(1 - \frac{2m}{r_s} \right)^{-1/2} \left(1 + \epsilon_r V \sqrt{1 - b^2 U(r_s)} \right). \quad (14.94)$$

Substituting expressions (14.89) for V and Γ , as well as (8.14) for $U(r)$, yields

$$1 + z = \left(1 - \frac{2m}{r_s} \right)^{-1} \left[\sqrt{1 - \frac{2m}{r_0}} + \epsilon_r \sqrt{\frac{2m}{r_s} - \frac{2m}{r_0}} \sqrt{1 - \frac{b^2}{r_s^2} \left(1 - \frac{2m}{r_s} \right)} \right]. \quad (14.95)$$

This formula gives the redshift of a photon that hits the screen of the remote observer \mathcal{O} at the impact parameter b at the instant $t_{\mathcal{O}}$. The quantity r_s has to be replaced by $r_s^{\text{em}}(b, t_{\mathcal{O}})$ as given by the system (14.77) for the pressureless collapse, where $t_{\text{em}}(b, t_{\mathcal{O}})$ is given by Eq. (14.79).

Remark 5: Equation (14.94) is valid for any spherically symmetric gravitational collapse, provided that V is the velocity of the surface with respect to the local static observer as defined by Eq. (14.88) and Γ is the corresponding Lorentz factor. On the contrary, Eq. (14.95) holds only for a pressureless collapse starting from rest.

Emission angle

The direction of emission in the frame of the comoving observer \mathcal{O}_{em} is characterized by the angle Θ_{em} between the photon's linear momentum \mathbf{P}_{em} and the normal to the surface \mathbf{n}_s . Considering that the collapsing matter is opaque to radiation, the maximum inclination of an emitted photon capable to reach the remote observer \mathcal{O} is $\Theta_{\text{em}} = \pi/2$ (emission tangent to the surface). Hence any photon received by \mathcal{O} has necessarily

$$0 \leq \Theta_{\text{em}} \leq \frac{\pi}{2}. \quad (14.96)$$

By definition, Θ_{em} is given by

$$\cos \Theta_{\text{em}} = \frac{\mathbf{n}_s \cdot \mathbf{P}_{\text{em}}}{\sqrt{\mathbf{P}_{\text{em}} \cdot \mathbf{P}_{\text{em}}}}. \quad (14.97)$$

Now, from Eq. (14.91), we have $\sqrt{\mathbf{P}_{\text{em}} \cdot \mathbf{P}_{\text{em}}} = E_{\text{em}}$ (using that \mathbf{p} is a null vector) and $\mathbf{n}_s \cdot \mathbf{P}_{\text{em}} = \mathbf{n}_s \cdot \mathbf{p}$. Evaluating this last scalar product by means of Eqs. (14.90) and (14.93) and substituting expression (14.94) for E_{em}/E_∞ , we get

$$\boxed{\cos \Theta_{\text{em}} = \frac{V + \epsilon_r \sqrt{1 - b^2 U(r_s)}}{1 + \epsilon_r V \sqrt{1 - b^2 U(r_s)}}}. \quad (14.98)$$

This relation can be easily inverted to get

$$\epsilon_r \sqrt{1 - b^2 U(r_s)} = \frac{\cos \Theta_{\text{em}} - V}{1 - V \cos \Theta_{\text{em}}}, \quad (14.99)$$

which allows us to express b in terms of Θ_{em} :

$$b = \frac{\sqrt{1 - V^2} \sin \Theta_{\text{em}}}{\sqrt{U(r_s)}(1 - V \cos \Theta_{\text{em}})}. \quad (14.100)$$

Remark 6: For $V = 0$, i.e. when the comoving observer coincides with the local static observer (prior to the start of the collapse), Eqs. (14.98) and (14.100) reduce respectively to Eqs. (8.126) and (8.127) taking into account the change of notations $r_s \leftrightarrow r_{\text{em}}$ and $\Theta_{\text{em}} \leftrightarrow |\eta|$.

We may also substitute Eq. (14.99) into Eq. (14.94) to get an expression of the redshift in terms of Θ_{em} :

$$1 + z = \left(1 - \frac{2m}{r_s}\right)^{-1/2} \frac{1}{\Gamma(1 - V \cos \Theta_{\text{em}})}. \quad (14.101)$$

For a pressureless collapse, Γ and V are given by Eq. (14.89) and the above relation simplifies to

$$\boxed{1 + z = \frac{1}{\sqrt{1 - \frac{2m}{r_0}} - \sqrt{\frac{2m}{r_s} - \frac{2m}{r_0}} \cos \Theta_{\text{em}}}}. \quad (14.102)$$

Central redshift

The redshift z_c at the center of the image of a pressureless collapse is obtained by setting $b = 0$ (which implies $\epsilon_r = +1$) in Eq. (14.95):

$$1 + z_c = \left(1 - \frac{2m}{r_s^{\text{em}}(0, t_\mathcal{O})}\right)^{-1} \left[\sqrt{1 - \frac{2m}{r_0}} + \sqrt{\frac{2m}{r_s^{\text{em}}(0, t_\mathcal{O})} - \frac{2m}{r_0}} \right], \quad (14.103)$$

where we have used $r_s^{\text{em}}(0, t_\mathcal{O})$ for r_s . One obtains an equivalent expression if one set $\Theta_{\text{em}} = 0$ (radial emission) in Eq. (14.102), in agreement with the property $b = 0 \iff \Theta_{\text{em}} = 0$, which follows from relation (14.100).

For $t_{\text{em}}(0, t_\mathcal{O}) \leq 0$, $r_s^{\text{em}}(0, t_\mathcal{O}) = r_0$ and formula (14.103) reduces to $1 + z_c = (1 - 2m/r_0)^{-1/2}$, which is the standard expression for the redshift at infinity of a photon emitted by a *static* observer at $r = r_0$ in the Schwarzschild spacetime.

For $r_s^{\text{em}}(0, t_\mathcal{O}) \rightarrow 2m$, formula (14.103) yields

$$1 + z_c \underset{r_s \rightarrow 2m}{\sim} 2\sqrt{1 - \frac{2m}{r_0}} \left(1 - \frac{2m}{r_s(0, t_\mathcal{O})}\right)^{-1}, \quad (14.104)$$

which implies $z \rightarrow +\infty$ for $r_s^{\text{em}}(0, t_\mathcal{O}) \rightarrow 2m$.

The elapsed proper time of observer \mathcal{O} since the reception of the central photon emitted at $\tau = 0$ (start of the collapse) is $t_\mathcal{O} - t_\mathcal{O}^*$. We get, from Eqs. (14.81) and (14.82),

$$t_\mathcal{O} - t_\mathcal{O}^* = t_{\text{em}}(0, t_\mathcal{O}) - r_s^{\text{em}}(0, t_\mathcal{O}) + r_0 + 2m \ln \left(\frac{r_0}{2m} - 1 \right) - 2m \ln \left(\frac{r_s^{\text{em}}(0, t_\mathcal{O})}{2m} - 1 \right).$$

When $r_s^{\text{em}}(0, t_\mathcal{O}) \rightarrow 2m$, two terms in the right-hand side of the above equation diverge: the last logarithm and the first term, $t_{\text{em}}(0, t_\mathcal{O})$, due to the singularity of Schwarzschild-Droste coordinates at $r = 2m$. We can express the latter divergence in terms of a logarithm involving $r_s^{\text{em}}(0, t_\mathcal{O})$ thanks to Eq. (6.32): $t_{\text{em}}(0, t_\mathcal{O}) = \tilde{t}_{\text{em}}(0, t_\mathcal{O}) - 2m \ln [r_s^{\text{em}}(0, t_\mathcal{O})/2m - 1] + K$, where $\tilde{t}_{\text{em}}(0, t_\mathcal{O})$ is the IEF coordinate of the emission event of the photon that reaches the center of \mathcal{O} 's screen at the instant $t_\mathcal{O}$ and K is a constant. $\tilde{t}_{\text{em}}(0, t_\mathcal{O})$ is finite (cf. Fig. 14.9) and we get

$$t_\mathcal{O} - t_\mathcal{O}^* = \underbrace{\tilde{t}_{\text{em}}(0, t_\mathcal{O}) - r_s^{\text{em}}(0, t_\mathcal{O}) + r_0 + 2m \ln \left(\frac{r_0}{2m} - 1 \right) + K}_{A} - 4m \ln \left(\frac{r_s^{\text{em}}(0, t_\mathcal{O})}{2m} - 1 \right).$$

The terms denoted by A are bounded when $r_s^{\text{em}}(0, t_\mathcal{O}) \rightarrow 2m$, so that

$$t_\mathcal{O} - t_\mathcal{O}^* \underset{r_s \rightarrow 2m}{\sim} -4m \ln \left(\frac{r_s^{\text{em}}(0, t_\mathcal{O})}{2m} - 1 \right) \iff 1 - \frac{2m}{r_s^{\text{em}}(0, t_\mathcal{O})} \underset{t_\mathcal{O} \rightarrow +\infty}{\sim} e^{-(t_\mathcal{O} - t_\mathcal{O}^*)/(4m)}.$$

Combining with Eq. (14.104), we obtain the expression of the central redshift at late times:

$$1 + z_c \underset{t_\mathcal{O} \rightarrow +\infty}{\sim} 2\sqrt{1 - \frac{2m}{r_0}} e^{(t_\mathcal{O} - t_\mathcal{O}^*)/(4m)}.$$

(14.105)

Hence we conclude:

Property 14.10: exponential growth of the central redshift

The photons arriving at the center of the image ($b = 0$) suffer an increasing redshift. At late times, the redshift grows exponentially with a characteristic time scale $4m$. Equivalently, the energies of the central photons vanish exponentially fast.

Redshift at the image's rim

We have the following characterization of the image's rim:

Property 14.11: photons forming the image's rim

The image's rim, or *limb*, at $b = b_*(t_\mathcal{O})$ (cf. Sec. 14.4.2), is formed by photons emitted tangentially to the star's surface. In other words, the image's rim is formed by photons emitted in the direction $\Theta_{\text{em}} = \pi/2$ with respect to the normal in the frame of the comoving observer \mathcal{O}_{em} .

We have demonstrated this property for $t_\mathcal{O} \leq t_\mathcal{O}^{**}$, i.e. prior to the collapse, in Sec. 14.4.2. We shall admit that it remains true for $t_\mathcal{O} > t_\mathcal{O}^{**}$.

One has then $\cos \Theta_{\text{em}} = 0$ for $b = b_*(t_\mathcal{O})$ and Eq. (14.98) implies $\epsilon_r = -1$ as soon as $V > 0$, i.e. when the collapse has started. Hence these photons are emitted in the inward direction. Their geodesics have then necessarily a r -turning point (periastron), after which r is increasing monotonically ($\epsilon_r = +1$) until reaching observer \mathcal{O} .

The redshift z_* at the image's rim is obtained by setting $\Theta_{\text{em}} = \pi/2$ in Eq. (14.101):

$$1 + z_* = \Gamma^{-1} \left(1 - \frac{2m}{r_s^{\text{em}}(b_*, t_\mathcal{O})} \right)^{-1/2}, \quad (14.106)$$

where b_* stands for $b_*(t_\mathcal{O})$. For pressureless collapse, Γ can be substituted by expression (14.89); this makes the terms involving $r_s^{\text{em}}(b_*, t_\mathcal{O})$ cancel out. Hence

Property 14.12: constant redshift at the image's rim

For a pressureless collapse starting from rest, the redshift at the image's rim keeps a constant value:

$$\boxed{1 + z_* = \left(1 - \frac{2m}{r_0} \right)^{-1/2}.} \quad (14.107)$$

This rather surprising result can be understood by considering the general expression (14.106): $1 + z_*$ is the product of two competing factors: the “static” gravitational redshift $(1 - 2m/r_s)^{-1/2}$, which is always larger than one and diverges as $r_s \rightarrow 2m$, and the Doppler blueshift factor Γ^{-1} , which is always lower than one and tends to zero as $r_s \rightarrow 2m$ (since $V \rightarrow 1$ in that limit). The Doppler blueshift is of course due to the rim's photons being emitted inward, which is the direction of the matter motion.

Redshift profile

The redshift across the image, i.e. z as a function of b at fixed $t_{\mathcal{O}}$, is given by Eq. (14.94) for a generic collapse and by Eq. (14.95) for a pressureless collapse from rest. However, one has to solve for $r_s(b, t_{\mathcal{O}})$ to make these formulas fully explicit. Doing so (see e.g. Fig. 2 of Ref. [332] or Fig. 6 of Ref. [523]), one obtains a redshift which is maximal at the image's center ($b = 0$), where it is given by formulas (14.103), and increases exponentially with time [Eq. (14.105)]. The redshift decreases with b until a small annulus near the image's rim ($b = b_*(t_{\mathcal{O}})$), where it takes the constant value (14.107). The width of this annulus decays to zero with time.

14.4.4 Image brightness

Beside the redshift, the image is characterized by its brightness. Each point in the image corresponds to a direction in observer \mathcal{O} 's sky. We may then think of a small solid angle $d\Omega$ around this direction as a “pixel” in the image. The brightness of this pixel in the frequency bandwidth $[\nu, \nu + d\nu]$ is represented by the energy flux (energy per unit time per unit area) dF given by

$$dF = I_{\nu} d\Omega d\nu, \quad (14.108)$$

where I_{ν} is the **specific intensity** of the electromagnetic radiation from the collapsing star⁷. The specific intensity I_{ν} depends on the radiation frequency ν as measured by observer \mathcal{O} , as well as on \mathcal{O} 's proper time $t_{\mathcal{O}}$ and the direction in \mathcal{O} 's sky. Taking the origin at the center of the image, the direction is defined by the small angle $\hat{b} = b/r_{\mathcal{O}}$ and some azimuthal angle $\phi \in [0, 2\pi)$. The solid angle increment $d\Omega$ can be then written $d\Omega = \sin \hat{b} d\hat{b} d\phi \simeq \hat{b} d\hat{b} d\phi$. Due to the rotation symmetry of the image, I_{ν} is independent of ϕ . We may then write $I_{\nu} = I_{\nu}(b, t_{\mathcal{O}}, \nu)$, using b to represent $\hat{b} = b/r_{\mathcal{O}}$.

A fundamental property of the specific intensity is that the ratio I_{ν}/ν^3 is constant along a given light ray. In particular, it does not depend on the distance between the source and the observer. This property follows from *Etherington's reciprocity theorem* which basically states that, in any spacetime, identical area elements carried by two observers in arbitrary motion subtend identical solid angles when seen by the other observer, up to some redshift factor $(1+z)^2$. This theorem can be proven in the framework of geometrical optics in curved spacetime (cf. e.g. Sec. 3.6 of Ref. [447], Sec. 6 of Ref. [182] or Sec. 2 of Ref. [411]) or by applying *Liouville's theorem* expressing the constancy of the photon distribution function f along a photon trajectory in phase space, given that $I_{\nu}/\nu^3 = h^4 f$, where h is Planck constant (see e.g. Sec. 22.6 of Ref. [370]). We deduce from the constancy of I_{ν}/ν^3 that

$$I_{\nu}(b, t_{\mathcal{O}}, \nu) = (1+z)^{-3} I_{\nu}^{\text{em}}(\Theta_{\text{em}}, t_{\text{em}}, \nu_{\text{em}}), \quad (14.109)$$

where $z = z(b, t_{\mathcal{O}})$ is the redshift at the location b on the image at time $t_{\mathcal{O}}$ as given by Eq. (14.94) or (14.95), Θ_{em} is the emission angle expressed in terms of $(b, t_{\mathcal{O}})$ by Eq. (14.98), $t_{\text{em}} = t_{\text{em}}(b, t_{\mathcal{O}})$ is the emission coordinate time expressed via Eq. (14.79) and ν_{em} is the emission frequency. By the very definition of the redshift, ν_{em} is related to ν by $\nu_{\text{em}} = (1+z)\nu$ [cf. Eq. (14.85)]. The

⁷The notation I_{ν} follows a standard convention in the literature, the subscript ν being merely an indicator to distinguish this quantity from the *intensity* I , which is integrated over all frequencies.

function $I_\nu^{\text{em}}(\Theta_{\text{em}}, t_{\text{em}}, \nu_{\text{em}})$ — the specific intensity at the source — depends on the model of emission assumed for the surface of the collapsing star.

The **intensity** I is the specific intensity integrated over all the frequencies:

$$I(b, t_\theta) = \int_0^{+\infty} I_\nu(b, t_\theta, \nu) d\nu \quad (14.110)$$

Given expression (14.109) and $\nu = (1+z)^{-1}\nu_{\text{em}}$, we get

$$I(b, t_\theta) = (1+z)^{-4} \int_0^{+\infty} I_\nu^{\text{em}}(\Theta_{\text{em}}, t_{\text{em}}, \nu_{\text{em}}) d\nu_{\text{em}}. \quad (14.111)$$

In view of the central redshift behavior (14.105), we deduce from this formula that the intensity at the center of the image decreases exponentially with a characteristic time equal to m :

$$\boxed{I(0, t_\theta) \underset{t_\theta \rightarrow +\infty}{\sim} I_0 e^{-(t_\theta - t_\theta^*)/m}}. \quad (14.112)$$

The **specific flux** of the image is obtained by integrating the flux element (14.108) over the solid angle $d\Omega$. Using $d\Omega = \hat{b} d\hat{b} d\phi = b/r_\theta^2 db d\phi$, we get

$$F_\nu(t_\theta, \nu) = \frac{2\pi}{r_\theta^2} \int_0^{b_*} I_\nu(b, t_\theta, \nu) b db. \quad (14.113)$$

The **flux** of the image is the specific flux integrated over all the frequencies:

$$F(t_\theta) = \int_0^{+\infty} F_\nu(t_\theta, \nu) d\nu = \frac{2\pi}{r_\theta^2} \int_0^{b_*} I(b, t_\theta) b db. \quad (14.114)$$

It can be shown (see e.g. Refs. [12, 454]) that, independently of the emission model, the flux is decaying exponentially according to

$$\boxed{F(t_\theta) \underset{t_\theta \rightarrow +\infty}{\sim} F_0 e^{-(t_\theta - t_\theta^*)/(3\sqrt{3}m)}}. \quad (14.115)$$

Historical note : The first computation of the appearance of a gravitational collapse as seen by a remote observer has been performed by Mikhail A. Podurets in 1964 [414]. He obtained the exponential decay formula (14.115) for the flux, up to an erroneous factor 2 in factor of $t_\theta - t_\theta^*$. This has been corrected by William L. Ames and Kip S. Thorne in 1968 [12]. Podurets considered only the image flux, which is relevant for a remote observer who cannot resolve the collapsing star. The intensity at the center of the image has been derived by Yakov B. Zeldovich and Igor D. Novikov in 1964 [528], who obtained the exponential decay formula (14.112). The computation of the intensity at each point of the image has been achieved by Ames and Thorne in their 1968 study [12], who showed that (i) the image is brightest at its rim, (ii) the redshift is constant there and (iii) the width of this peripheral bright region is decaying exponentially with time. In 1969, Jack Jaffe [295] pointed out that Ames and Thorne considered only photons with an emission angle $\Theta'_{\text{em}} \leq \pi/2$ in the local *static* observer frame, while the limiting condition should be $\Theta_{\text{em}} \leq \pi/2$ in the *comoving* observer frame [Eq. (14.96)]. This encompasses extra photons: those that are emitted inward, as shown above. In 1979, Kayll Lake and R. C. Roeder [332] recomputed the intensity with this hypothesis and concluded that the main properties found by Podurets, Ames and Thorne are not significantly altered.

14.5 Going further

This chapter dealt with the collapse of a spherically symmetric and homogeneous ball of some pressureless fluid (“dust”). A first generalization amounts to allowing for non-homogeneous density profiles, keeping the vanishing pressure, see e.g. Ref. [413]. More realistic models shall take into account pressure, rotation and possibly other physical processes, like neutrino emission. Although very relevant from an astrophysical point of view, rotation significantly complicates the problem of gravitational collapse, since gravitational waves are emitted in this case. This prevents the exterior spacetime to be described by an exact solution, such as the Schwarzschild one⁸. Only numerical solutions are known; see e.g. Chap. 9 of Ref. [455] or Chaps. 8, 10 and 14 of Ref. [43].

⁸Note that there is no equivalent to the Jepsen-Birkhoff theorem (Property 14.4) for a non-spherically symmetric (e.g. flattened) body; in particular, the vacuum metric outside the body is *not* Kerr metric.

Chapter 15

Black hole formation 2: Vaidya collapse

Contents

15.1 Introduction	581
15.2 The ingoing Vaidya metric	581
15.3 Imploding shell of radiation	586
15.4 Configurations with a naked singularity	600
15.5 Going further	612

15.1 Introduction

Having investigated the gravitational collapse of a star, modeled as a ball of dust, in the preceding chapter, we move to a much less astrophysical scenario: the formation of a black hole by the collapse of a spherical shell of pure radiation (no matter!). Albeit quite academic, this process illustrates various features of black hole birth and dynamics, in a way somewhat complementary to the collapse of a dust ball. The model is based on an exact solution of the Einstein equation sourced by a pure radial electromagnetic flux, known as *Vaidya metric*, which we present first (Sec. 15.2). Then, we introduce the model describing the implosion of a radiation shell and study the black hole formation via such a process (Sec. 15.3). Finally, we focus on a subclass of models giving birth a *naked singularity*, i.e. models in which the central curvature singularity is visible to remote observers located outside the black hole region (Sec. 15.4).

15.2 The ingoing Vaidya metric

15.2.1 General expression

Let us consider a spherically symmetric spacetime $(\mathcal{M}, \mathbf{g})$ described by coordinates (v, r, θ, φ) such that $v \in \mathbb{R}$, $r \in (0, +\infty)$, $\theta \in (0, \pi)$ and $\varphi \in (0, 2\pi)$, (θ, φ) being standard spherical

coordinates on \mathbb{S}^2 and r being the areal radius associated with spherical symmetry (cf. Sec. 6.2.2). The **ingoing Vaidya metric** is the metric tensor

$$\boxed{\mathbf{g} = - \left(1 - \frac{2M(v)}{r}\right) \mathbf{d}v^2 + 2 \mathbf{d}v \mathbf{d}r + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2)}, \quad (15.1)$$

where $M(v)$ is a real-valued function of v . We immediately notice that this expression strongly resembles that of the Schwarzschild metric expressed in the *null ingoing Eddington-Finkelstein* coordinates, as given by Eq. (6.29). Actually, the only difference is the constant m in Eq. (6.29) replaced by the function $M(v)$ in Eq. (15.1). We may even say that the Schwarzschild metric is the special case $M(v) = \text{const}$ of the ingoing Vaidya metric.

A key property of the ingoing Vaidya metric is

Property 15.1: null hypersurfaces $v = \text{const}$ and their normals

The hypersurfaces $v = \text{const}$ are null (i.e. v is a null coordinate); a normal to them is the null vector^a:

$$\mathbf{k} := -\vec{\mathbf{d}}v = -\vec{\nabla}v \iff \underline{\mathbf{k}} := -\mathbf{d}v. \quad (15.2)$$

Moreover, \mathbf{k} is equal to minus the vector ∂_r of coordinates (v, r, θ, φ) :

$$\mathbf{k} = -\left.\frac{\partial}{\partial r}\right|_{v,\theta,\varphi}, \quad (15.3)$$

where we have rewritten ∂_r as $\partial/\partial r|_{v,\theta,\varphi}$ to distinguish it from the vector ∂_r of the IEF coordinates, to be introduced in Sec. 15.2.2.

^ain index notation: $k^\alpha := -g^{\alpha\mu}\partial_\mu v = -\nabla^\alpha v \iff k_\alpha := -\partial_\alpha v$.

Proof. Let Σ_v be a hypersurface defined by $v = \text{const}$. The metric induced by \mathbf{g} on Σ_v is obtained by setting $\mathbf{d}v = 0$ in Eq. (15.1): $\mathbf{g}|_{\Sigma_v} = r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2)$. This metric has clearly the signature $(0, +, +)$, i.e. it is degenerate, hence Σ_v is a null hypersurface (cf. Sec. 2.2.2). By construction, \mathbf{k} is normal to Σ_v . It is thus a null vector. Besides, the metric dual of the coordinate vector field ∂_r is the 1-form $\underline{\partial}_r = g_{\mu r} \mathbf{d}x^\mu = g_{vr} \mathbf{d}v = \mathbf{d}v = -\underline{\mathbf{k}}$, which proves Eq. (15.3). \square

Property 15.2: time orientation of spacetime

Since $\mathbf{d}v$ is nowhere vanishing, \mathbf{k} is a nonzero null vector field on \mathcal{M} . We then set the *time orientation* of $(\mathcal{M}, \mathbf{g})$ by declaring that \mathbf{k} is future-directed (cf. Sec. 1.2.2).

As shown in the notebook D.6.4, the Ricci tensor of \mathbf{g} takes a simple form:

$$\mathbf{R} = \frac{2M'(v)}{r^2} \mathbf{d}v \otimes \mathbf{d}v, \quad (15.4)$$

where $M'(v)$ stands for the derivative of the function $M(v)$. The Ricci scalar $R = g^{\mu\nu}R_{\mu\nu} = (2M'(v)/r^2)\nabla_\mu v \nabla^\mu v = (2M'(v)/r^2)k_\mu k^\mu$ is identically zero, since \mathbf{k} is a null vector. A consequence is

Property 15.3: Vaidya metric as a solution of the Einstein equation

The ingoing Vaidya metric (15.1) is a solution of the Einstein equation (1.40) with $\Lambda = 0$ and with the energy-momentum tensor

$$\boxed{\mathbf{T} = \frac{M'(v)}{4\pi r^2} \underline{\mathbf{k}} \otimes \underline{\mathbf{k}}.} \quad (15.5)$$

Remark 1: We have already noticed that Vaidya metric reduces to Schwarzschild metric for $M(v) = \text{const}$. This corresponds to $M'(v) = 0$, so that Eq. (15.5) reduces to $\mathbf{T} = 0$ and we recover that Schwarzschild metric is a solution of the vacuum Einstein equation.

The tensor (15.5) has the same structure as the energy-momentum tensor of the dust model considered in Chap. 14: $\mathbf{T}_{\text{dust}} = \rho \underline{\mathbf{u}} \otimes \underline{\mathbf{u}}$ [Eq. (14.2)]. The main difference is that \mathbf{u} is a timelike vector (the dust 4-velocity), while \mathbf{k} is a null vector. For this reason, the energy-momentum tensor (15.5) is sometimes referred to as a **null dust** model [415]. It corresponds physically to the energy-momentum tensor of some *monochromatic electromagnetic radiation* in the *geometrical optics* approximation (see Box 22.4 of Ref. [370]): $\mathbf{T}_{\text{rad}} = q \underline{\mathbf{K}} \otimes \underline{\mathbf{K}}$, where $q \geq 0$ and $\mathbf{K} := \vec{\nabla}\Phi$ is the *wave vector*, Φ being the rapidly-varying phase in the geometrical optics decomposition $\mathbf{A} = \text{Re}(e^{i\Phi} \mathbf{a})$ of the electromagnetic potential 1-form \mathbf{A} . The quantity q is related to the energy density ε of the electromagnetic field as measured by an observer \mathcal{O} of 4-velocity \mathbf{u} by $\varepsilon = \omega^2 q$, where $\omega = -\mathbf{K} \cdot \mathbf{u}$ is the frequency of the electromagnetic radiation as measured by \mathcal{O} . Maxwell equations imply that \mathbf{K} is a null vector and that it is geodesic: $\nabla_{\mathbf{K}} \mathbf{K} = 0$. The geodesic integral curves of \mathbf{K} are nothing but the *light rays* of the geometrical optics framework. The geodesic character also holds for \mathbf{k} as defined by Eq. (15.2), since \mathbf{k} is normal to the null hypersurfaces $v = \text{const}$: the integral curves of \mathbf{k} are the null geodesic generators of these hypersurfaces (cf. Sec. 2.3.3); actually, we have exactly

$$\nabla_{\mathbf{k}} \mathbf{k} = 0. \quad (15.6)$$

This follows from Eq. (2.22) with $\ell = \mathbf{k}$ and $\kappa = 0$ by virtue of Eq. (2.21) with $\rho = 0$ implied by Eq. (2.11) given that $u = v$ and $\mathbf{k} = -\vec{\nabla}v$ [Eq. (15.2)]. Equations (15.6) and (15.3) imply that the curves $(v, \theta, \varphi) = \text{const}$ are null geodesics, $\lambda := -r$ is an affine parameter along them and \mathbf{k} is the corresponding tangent vector.

Another condition for the identification of \mathbf{T} with \mathbf{T}_{rad} is that the coefficient in front of $\underline{\mathbf{k}} \otimes \underline{\mathbf{k}}$ in (15.5) is non-negative, since $q \geq 0$ in \mathbf{T}_{rad} . This constraint can also be seen as the *null energy condition* (2.95) introduced in Sec. 2.4.2: for any null vector ℓ , we have $\mathbf{T}(\ell, \ell) = M'(v)/(4\pi r^2)(\mathbf{k} \cdot \ell)^2$ and hence $\mathbf{T}(\ell, \ell) \geq 0 \iff M'(v) \geq 0$. In other words, the function $M(v)$ must be monotonically increasing. Then, we can set $\mathbf{k} = \alpha \mathbf{K}$, where α is a constant, to have a perfect match of (15.5) with the electromagnetic radiation energy-momentum tensor \mathbf{T}_{rad} . To summarize:

Property 15.4: Vaidya metric sourced by electromagnetic radiation

Provided that $M(v)$ is an increasing function^a, the ingoing Vaidya spacetime (\mathcal{M}, g) is generated by a spherical symmetric electromagnetic radiation within the geometrical optics approximation. The corresponding light rays are the *ingoing radial null geodesics* $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$ defined by $v = \text{const}$, $\theta = \text{const}$ and $\varphi = \text{const}$. These geodesics admit k as the tangent vector associated with their affine parameter $\lambda := -r$.

^aby *increasing*, it is meant strictly increasing ($M'(v) > 0$) or locally constant ($M'(v) = 0$).

Historical note : The Vaidya metric has been actually first derived by Henri Mineur in 1933 [366], as the solution of the Einstein equation for the exterior of a spherically symmetric body, the “mass” of which “is varying” due to the radiation of an “energy flux of light”. It thus corresponds to the *outgoing* version of the Vaidya metric, while the version (15.1) considered here is *ingoing*. More precisely, the solution given by Mineur reads¹

$$-ds^2 = 2 dx dr + r^2 \frac{du^2 - dv^2}{u^2} - \left(1 - \frac{2M(x)}{r}\right) dx^2. \quad (15.7)$$

The coordinate x is the opposite of a retarded time ($x = -u$ in our notations), while (u, v) are coordinates on the 2-sphere \mathbb{S}^2 (submanifold $(x, r) = \text{const}$) such that the standard (round) metric of \mathbb{S}^2 is $q = u^{-2} (-du^2 + dv^2)$ (cf. the unnumbered equation at the top of p. 35 of Mineur’s article [366]). This looks quite surprising, given that $u^{-2} (-du^2 + dv^2)$ is rather the metric of the 2-dimensional anti-de Sitter space expressed in a variant of Poincaré coordinates (compare Eq. (13.66) with $T = v$ and $R = 1/u$). In particular the signature is $(-, +)$, while it should be $(+, +)$ for \mathbb{S}^2 ! It turns out that the coordinate u used by Mineur is pure imaginary, i.e. Mineur wrote the metric of \mathbb{S}^2 via a kind of Wick rotation of the metric of the hyperbolic plane² \mathbb{H}^2 , the latter being $(dX^2 + dY^2)/Y^2$: setting $v = X$ and $u = iY$, one gets $-q = u^{-2} (du^2 - dv^2)$. If one restores the standard spherical coordinates (θ, φ) on \mathbb{S}^2 and uses the modern notation $u = -x$ as well as the metric signature $(-, +, +, +)$, one can rewrite Mineur’s solution as³

$$g = - \left(1 - \frac{2M(u)}{r}\right) du^2 - 2 du dr + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (15.8)$$

This ressembles the metric (15.1). The only difference is the $-$ sign in front of $du dr$, while there is a $+$ sign in front $dv dr$ in (15.1). This results from Mineur’s version being *outgoing*, given that he considered a radiating body. On the contrary, the version (15.1) is *ingoing*, since we are interested in gravitational collapse and black hole formation.

Twenty years later, in 1953, Prahalad Chunnilal Vaidya presented the metric that bears his name in the form (15.8) [487]. Previously, in 1943 [485] and in 1951 [486], Vaidya presented the metric outside a

¹This is Mineur’s Eq. (21), p. 47 of Ref. [366]. The minus sign in the left-hand side, which is present Mineur’s article [366], is fortunate for comparison with the current text since Mineur is using the metric signature $(+, -, -, -)$, i.e. the opposite of ours.

²The spaces \mathbb{H}^2 and \mathbb{S}^2 are somehow connected as being the only 2-dimensional Riemannian manifolds of nonzero constant scalar curvature (negative for \mathbb{H}^2 and positive for \mathbb{S}^2).

³As one can infer by comparing with other expressions of ds^2 in Mineur’s article, the $+$ sign in front of r^2 times the \mathbb{S}^2 line element in Eq. (15.7) is certainly a typo and should be replaced by a $-$ sign; this correction is performed to get Eq. (15.8).

spherically symmetric radiating star in an equivalent, but more complicated form:

$$g = -\frac{1}{f(M)^2} \left(\frac{\partial M}{\partial T} \right)^2 \left(1 - \frac{2M}{r} \right) dt^2 + \left(1 - \frac{2M}{r} \right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (15.9)$$

where f is an arbitrary function and $M = M(T, r)$ fulfills the differential equation $\partial M / \partial r (1 - 2M/r) = f(M)$. By introducing the scalar field u such that $du = -f(M)^{-1} dM$ and promoting it as a coordinate, one can bring this solution to Mineur's form (15.8). In a foreword to Vaidya's 1951 article [486], Vishnu Vasudev Narlikar, who was Vaidya's PhD advisor, wrote: “The treatment as given here is essentially different from that of Professor H. Mineur as it appears in Ann. de l'Ecole Normal Supérieure, Ser. 3, 5, 1, 1933 (our Ref. [366]). Our attention was kindly drawn to it by Professor Mineur some years ago.” In a common article by Narlikar and Vaidya published in 1947 [377], one can also read “The line element (5) [our Eq. (15.9)] was first published by one of us some years ago. A line element equivalent to (5) but not obviously so was obtained by Mineur several years earlier.” It would probably be fair to call the metric (15.8) the *Mineur-Vaidya metric*; we shall however keep here the usual name of *Vaidya metric*. The metric (15.8) was popularized and further studied by Richard W. Lindquist, Robert A. Schwartz and Charles W. Misner in 1965 [346], who apparently were not aware of Mineur's work.

15.2.2 Expression in ingoing Eddington-Finkelstein coordinates

To deal with black hole formation in Vaidya spacetime, it is quite convenient to work with the **ingoing Eddington-Finkelstein (IEF) coordinates** (t, r, θ, φ) , which are defined from the null coordinates (v, r, θ, φ) in the same way as in the Schwarzschild case, i.e. by considering that v is the **advanced time** with respect to t [cf. Eq. (6.30)]:

$$t := v - r \iff v = t + r. \quad (15.10)$$

Remark 2: The coordinate t was denoted by \tilde{t} in Chap. 6, where t was reserved for the Schwarzschild-Droste time coordinate.

We have $dv = dt + dr$, from which the IEF expression of the metric tensor is immediately deduced from Eq. (15.1):

$$\boxed{g = - \left(1 - \frac{2M(t+r)}{r} \right) dt^2 + \frac{4M(t+r)}{r} dt dr + \left(1 + \frac{2M(t+r)}{r} \right) dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2).} \quad (15.11)$$

The expression of k in terms of the IEF coordinate frame is deduced from Eq. (15.3) and the chain rule:

$$k = \partial_t - \partial_r. \quad (15.12)$$

Remark 3: The analog equation in Schwarzschild spacetime is Eq. (6.48).

15.2.3 Outgoing radial null geodesics

Let us determine the radial null directions at each point by searching for null vectors of the form $\ell = \partial_t + V\partial_r$. The condition $\mathbf{g}(\ell, \ell) = 0$ with the expression (15.11) for \mathbf{g} yields immediately a quadratic equation for V :

$$\left(1 + \frac{2M(t+r)}{r}\right)V^2 + \frac{4M(t+r)}{r}V + 1 - \frac{2M(t+r)}{r} = 0.$$

The two solutions are $V = -1$ and $V = [r - 2M(t+r)]/[r + 2M(t+r)]$. The first solution gives back the ingoing null vector \mathbf{k} introduced in Sec. 15.2.1 [cf. Eq. (15.12)]. The null vector ℓ corresponding to the second solution is

$$\ell = \partial_t + \frac{r - 2M(t+r)}{r + 2M(t+r)}\partial_r. \quad (15.13)$$

Property 15.5: radial null geodesics

The integral curves of the vector fields \mathbf{k} and ℓ are null geodesics. Those of \mathbf{k} are the *ingoing radial null geodesics* $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$ already discussed in Sec. 15.2.1, while those of ℓ are called the ***outgoing radial null geodesics***.

Proof. A direct computation shows that ℓ is a pregeodesic vector field: $\nabla_\ell\ell = \kappa\ell$, with $\kappa = -4[rM'(t+r) - M(t+r)]/[r + 2M(t+r)]^2$ (cf. the notebook D.6.4). It follows that the integral curves of ℓ are geodesics (cf. Sec. B.2.2). \square

The differential equation governing the outgoing radial null geodesics is obtained by demanding that ℓ is their tangent vector:

$$\boxed{\frac{dr}{dt} = \frac{r - 2M(t+r)}{r + 2M(t+r)}}. \quad (15.14)$$

In what follows, we shall get exact solutions of this equation for $M(v)$ piecewise linear.

Remark 4: As in the Schwarzschild case (cf. Remark 1 on p. 189), the outgoing radial null geodesics are actually *ingoing*, i.e. have r decreasing towards the future, as soon as $r < 2M(t+r)$. This corresponds to the region bounded by the red curve in Fig. 15.1, to be discussed in Sec. 15.3.5.

15.3 Imploding shell of radiation

15.3.1 The imploding shell model

An ***imploding shell of radiation*** is defined by the ingoing Vaidya metric with the function $M(v)$ obeying $M'(v) \neq 0$ only on a finite interval of v . By choosing properly the origin of v , we may consider this interval to be $[0, v_0]$, where the parameter $v_0 > 0$ governs the shell's

thickness. The function $M(v)$ is thus constant outside the interval $[0, v_0]$. In order to describe the formation of a black hole, we choose $M(v) = 0$ for $v < 0$. This corresponds to a piece of Minkowski spacetime, since the metric (15.11) clearly reduces to Minkowski metric (4.3) wherever $M(v) = 0$. Denoting by $m > 0$ the constant value of $M(v)$ for $v > v_0$, we have then

$$M(v) = \begin{cases} 0 & \text{for } v < 0 \\ m S(v/v_0) & \text{for } 0 \leq v \leq v_0 \\ m & \text{for } v > v_0 \end{cases} \quad \begin{array}{l} (\text{Minkowski region, } \mathcal{M}_{\text{Min}}) \\ (\text{radiation region, } \mathcal{M}_{\text{rad}}) \\ (\text{Schwarzschild region, } \mathcal{M}_{\text{Sch}}), \end{array} \quad (15.15)$$

where $S : [0, 1] \rightarrow [0, 1]$, $x \mapsto S(x)$ is an increasing function obeying $S(0) = 0$ and $S(1) = 1$. The region with $v > v_0$ is qualified as *Schwarzschild* since for $M(v) = m = \text{const}$, the Vaidya metric reduces to Schwarzschild metric, as noticed in Sec. 15.2.1. The three regions are shown in terms of coordinates (t, r) on Fig. 15.1: \mathcal{M}_{rad} (the imploding shell) is the yellow region, \mathcal{M}_{Min} lies below it and \mathcal{M}_{Sch} lies above it.

The simplest example of a function $M(v)$ obeying (15.15) is obtained for $S(x) = x$:

$$S(x) = x \iff M(v) = m \frac{v}{v_0} \quad (0 \leq v \leq v_0). \quad (15.16)$$

It is shown as the blue curve in Fig. 15.2. This choice of S makes $M(v)$ piecewise linear. The resulting metric tensor (15.1) is continuous but not C^1 at $v = 0$ and $v = v_0$. A choice of S that yields a C^2 metric tensor is $S(x) = 6x^5 - 15x^4 + 10x^3$. This choice is depicted by the red curve in Fig. 15.2.

Historical note : The imploding shell model (15.15) has been introduced by William A. Hiscock, Leslie G. Williams and Douglas M. Eardley in 1982 [278], as well as by Achilles Papapetrou in 1985 [399]. Both studies regarded the specific case $S(x) = x$. Hiscock et al. considered the ingoing Vaidya metric in the form (15.1) (coordinates (v, r, θ, φ)), while Papapetrou made use of the form (15.11) (coordinates (t, r, θ, φ)).

15.3.2 Solution for $M(v)$ piecewise linear

Let us consider the simplest choice for $M(v)$, i.e. Eq. (15.16). In the radiation region, the metric tensor (15.1) takes the form

$$\mathbf{g} = - \left(1 - \alpha \frac{v}{r}\right) \mathbf{d}v^2 + 2 \mathbf{d}v \mathbf{d}r + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2) \quad (0 \leq v \leq v_0), \quad (15.17)$$

where α is the positive constant defined by

$$\alpha := \frac{2m}{v_0}. \quad (15.18)$$

It is immediately apparent on (15.17) that for any $\lambda > 0$, the homothety $H_\lambda : (v, r) \mapsto (\lambda v, \lambda r)$ maps \mathbf{g} to $\lambda^2 \mathbf{g}$. Hence H_λ is a conformal isometry of $(\mathcal{M}_{\text{rad}}, \mathbf{g})$ with a constant conformal factor λ^2 . The homotheties $(H_\lambda)_{\lambda \in \mathbb{R}_{>0}}$ form a 1-dimensional group, the generator of which is

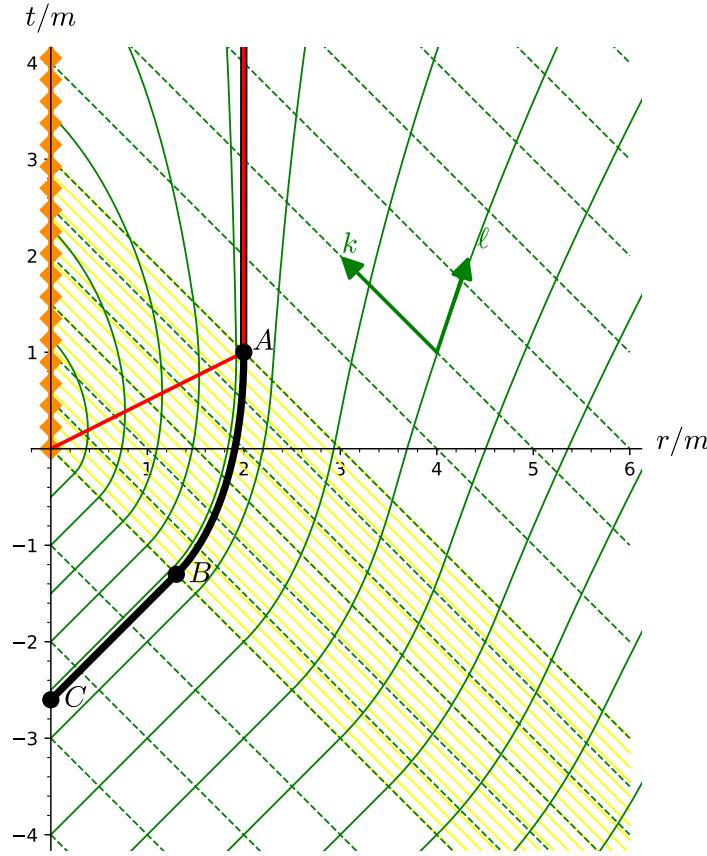


Figure 15.1: Spacetime diagram of the Vaidya collapse based on the IEF coordinates (t, r) and for the linear mass function $M(v) = mv/v_0$ with $v_0 = 3m$ (homothetic model with $\alpha = 2/3$). The yellow area is the radiation region \mathcal{M}_{rad} [cf. Eq. (15.15)], below it lies the Minkowski region \mathcal{M}_{Min} and above it, the Schwarzschild region \mathcal{M}_{Sch} . The solid (resp. dashed) green curves are outgoing (resp. ingoing) radial null geodesics. The thick black line marks the event horizon \mathcal{H} (Sec. 15.3.3) and the red one the future outer trapping horizon \mathcal{T} (Sec. 15.3.5). Note that \mathcal{H} and \mathcal{T} coincide in \mathcal{M}_{Sch} . The curvature singularity is indicated by the orange zigzag line. The part of the figure corresponding to \mathcal{M}_{Min} can be compared with Fig. 4.1, while that corresponding to \mathcal{M}_{Sch} can be compared with Fig. 6.3. [Figure generated by the notebook D.6.4]

obtained by considering infinitesimal transformations, i.e. homotheties of ratio $\lambda = 1 + d\lambda$ where $d\lambda$ is infinitely small. The components of the corresponding displacement vector are $dv = d\lambda v$ and $dr = d\lambda r$, so that formula (3.14) (with $t \leftrightarrow \lambda$) leads to the generator

$$\xi = v \partial_v + r \partial_r|_{v,\theta,\varphi} = t \partial_t + r \partial_r. \quad (15.19)$$

The second equality follows from the change of coordinates (15.10). That ξ has the same expression with respect to (v, r) and (t, r) coordinates should not be surprising since the homothety H_λ has the same expression in both coordinate systems: $H_\lambda : (t, r) \mapsto (\lambda t, \lambda r)$, given that $\lambda v = \lambda(t + r) = \lambda t + \lambda r$. The vector field ξ is called a **homothetic Killing vector**. The Lie derivative of g along ξ is twice g (cf. the notebook D.6.4 for the computation):

$$\mathcal{L}_\xi g = 2g. \quad (15.20)$$

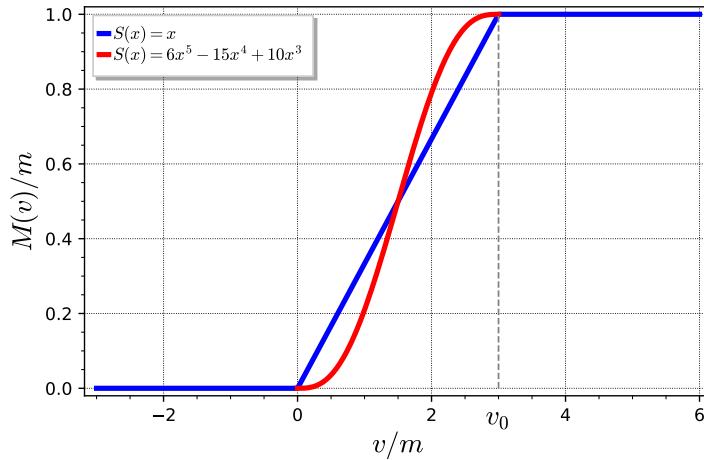


Figure 15.2: Function $M(v)$ for the imploding shell model, for $v_0 = 3m$ and two different choices of $S(x)$ in formula (15.15).

We shall refer to the choice $M(v)$ piecewise linear as the **homothetic radiation shell model**.

Remark 1: The denomination **self-similar**, in place of *homothetic*, is also used in the literature (e.g. [382, 383]).

Remark 2: A *homothetic Killing vector* is not a *Killing vector*, for the right-hand side of Eq. (15.20) would be zero if ξ were a Killing vector [cf. Eq. (3.18)]. In other words, except for $\lambda = 1$, the homotheties H_λ are not isometries, but only *conformal isometries*. Generally, vector fields generating conformal isometries are called **conformal Killing vectors**. They fulfill $\mathcal{L}_\xi g = \sigma g$, where σ is a scalar field. Equation (15.20) constitutes the particular case $\sigma = 2$.

Let us introduce the variable

$$x := \frac{v}{r}, \quad (15.21)$$

which is invariant under the homotheties H_λ . The differential equation governing the outgoing radial null geodesics, Eq. (15.14), can be rewritten as $dt/dr = (1+\alpha x)/(1-\alpha x)$ [cf. Eq. (15.18)]. Given that $t = v - r = r(x-1)$ implies $dt/dr = x-1+r dx/dr$, we get the equivalent form

$$r \frac{dx}{dr} = \frac{\alpha x^2 - x + 2}{1 - \alpha x}. \quad (15.22)$$

Fortunately, this ordinary differential equation is separable, so that its solutions are easily obtained by quadrature. They depend on whether the quadratic polynomial $P_\alpha(x) := \alpha x^2 - x + 2$ admits real roots or not. Let us first focus on the case where P_α has no real root. The discriminant being $1 - 8\alpha$, this occurs if, and only if,

$$\alpha > \frac{1}{8} \iff v_0 < 16m. \quad (15.23)$$

By considering the energy-momentum tensor (15.5) with expression (15.16) substituted for $M(v)$, we get

$$\mathbf{T} = \frac{\alpha}{8\pi r^2} \underline{k} \otimes \underline{k}. \quad (15.24)$$

Consequently, we may say that the case $\alpha > 1/8$ corresponds to a **large radiation energy density**. We shall discuss the low radiation energy density case ($\alpha < 1/8$) in Sec. 15.4. For the moment, assuming (15.23), we have $P_\alpha(x) > 0$ for any $x \in \mathbb{R}$ and we may rewrite Eq. (15.22) as

$$d \ln r = \frac{1 - \alpha x}{\alpha x^2 - x + 2} dx, \quad (15.25)$$

the solution of which is $r = r_0 f_\alpha(x)$, where

$$\ln f_\alpha(x) := \int_0^x \frac{1 - \alpha x'}{\alpha x'^2 - x' + 2} dx' \quad (15.26)$$

and the integration constant $r_0 > 0$ is the value of r at $x = 0$, or equivalently at $v = 0$, since $f_\alpha(0) = 1$. Explicitly,

$$f_\alpha(x) = \frac{\sqrt{2}}{\sqrt{\alpha x^2 - x + 2}} \exp \left\{ \frac{1}{\sqrt{8\alpha - 1}} \left[\arctan \left(\frac{2\alpha x - 1}{\sqrt{8\alpha - 1}} \right) + \arctan \left(\frac{1}{\sqrt{8\alpha - 1}} \right) \right] \right\}, \quad (15.27)$$

Introducing the dimensionless parameter $u := -\ln(r_0/m)$, we conclude:

Property 15.6: outgoing radial null geodesics for the $\alpha > 1/8$ homothetic shell

For the homothetic model with $\alpha > 1/8$, the outgoing radial null geodesics in \mathcal{M}_{rad} form a 3-parameter family of curves $(\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}})$, where the parameter $u \in \mathbb{R}$ is related to the value r_0 of r at the inner edge of the radiation shell ($v = 0$) by $r_0 = me^{-u}$. The parametric equation of $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ in terms of the IEF coordinates is

$$\begin{cases} t = me^{-u}(x - 1)f_\alpha(x) \\ r = me^{-u}f_\alpha(x) & 0 \leq x \leq x_{\max}, \\ \theta = \text{const}, \varphi = \text{const} \end{cases} \quad (15.28)$$

where the function $f_\alpha(x)$ is defined by Eq. (15.27) and either $x_{\max} = +\infty$ ($\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ reaches $r = 0$ for some $v < v_0$) or x_{\max} is the solution of $me^{-u}x_{\max}f_\alpha(x_{\max}) = v_0$ ($\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ reaches the outer edge of the radiation shell).

The function $f_\alpha(x)$ is plotted in Fig. 15.3. It increases from 1 at $x = 0$ to some maximum reached for $x = \alpha^{-1}$ and then decreases to 0 as $x \rightarrow +\infty$. This behavior follows directly from the sign of the numerator $1 - \alpha x$ in Eq. (15.25), given that the denominator $\alpha x^2 - x + 2$ is always positive for $\alpha > 1/8$. Note that it could be that $x_{\max} < \alpha^{-1}$ so that the maximum of $f_\alpha(x)$ is actually not reached along $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$. In that case, r increases monotonically along $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ in \mathcal{M}_{rad} , from r_0 to $r_0 f_\alpha(x_{\max})$.

It appears clearly on Eq. (15.28) that the homothety $H_\lambda : (t, r) \mapsto (\lambda t, \lambda r)$ transforms the geodesic $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ into the geodesic $\mathcal{L}_{(u',\theta,\varphi)}^{\text{out}}$ with $u' = u - \ln \lambda$, in agreement with the homothetic symmetry of $(\mathcal{M}_{\text{rad}}, g)$ discussed above.

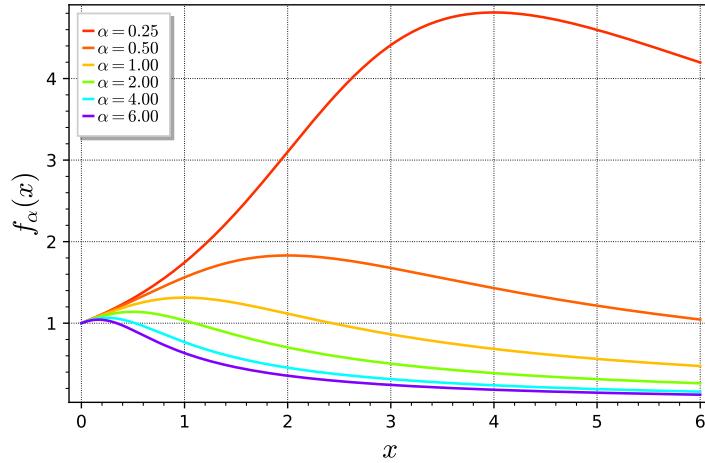


Figure 15.3: Function $f_\alpha(x)$, defined by Eq. (15.27), for some selected values of $\alpha > 1/8$. For each value of α , the maximum of $f_\alpha(x)$ is achieved for $x = \alpha^{-1}$. [Figure generated by the notebook D.6.5]

Some outgoing radial null geodesics are depicted as solid green lines in Fig. 15.1. In \mathcal{M}_{rad} , they obey Eq. (15.28). Note that the homothetic symmetry appears clearly on the figure. If a geodesic $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ has a r -turning point in \mathcal{M}_{rad} , it must be located at $x = \alpha^{-1}$, i.e. at $t/r = \alpha^{-1} - 1$, or equivalently at

$$t = \left(\frac{v_0}{2m} - 1 \right) r. \quad (15.29)$$

The above equation defines a straight line through $(t, r) = (0, 0)$, whose intersection with \mathcal{M}_{rad} is depicted by a red segment in Fig. 15.1.

Remark 3: The turning point value (15.29) can be obtained directly by setting $dr/dt = 0$ in Eq. (15.14) and using the value (15.16) for $M(v)$.

In the Minkowski region, the outgoing radial null geodesics are straight line segments inclined at $+45^\circ$ in Fig. 15.1, while in the Schwarzschild region, they are curves obeying Eq. (6.51) (with the change of notation $\tilde{t} \leftrightarrow t$).

15.3.3 Black hole formation

In spherical symmetry, the inspection of radial null geodesics provides a direct access to the black hole event horizon \mathcal{H} . Since we have at disposal the exact solution (15.28) for the outgoing radial null geodesics, let us determine the location of \mathcal{H} for the homothetic shell collapse. We arrive at the following result:

Property 15.7: black hole formation for the $\alpha > 1/8$ homothetic shell

The homothetic imploding radiation shell with $\alpha := 2m/v_0 > 1/8$ generates a black hole. The black hole event horizon \mathcal{H} is the future light cone of the point of IEF coordinates

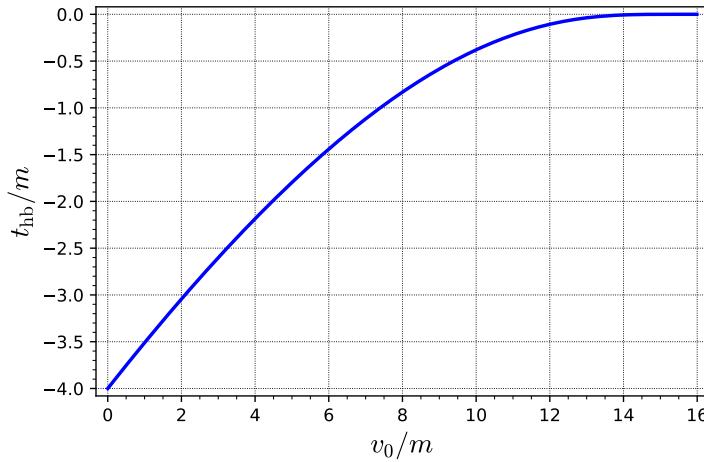


Figure 15.4: Value t_{hb} of the coordinate t at the black hole birth [Eq. (15.30)] as a function of the radiation shell thickness v_0 , for the homothetic shell collapse [Eq. (15.16)].

$(t, r) = (t_{hb}, 0)$, with^a

$$t_{hb} = -4m \exp \left[-\frac{2}{\sqrt{8\alpha - 1}} \arctan \left(\frac{1}{\sqrt{8\alpha - 1}} \right) \right]. \quad (15.30)$$

Outside the radiation shell, \mathcal{H} coincides with the Killing horizon of Schwarzschild space-time located at $r = 2m$.

^aAs in Chap. 14, the subscript ‘hb’ stands for *horizon birth*.

Proof. In the Schwarzschild region \mathcal{M}_{Sch} , the Killing horizon is the hypersurface $r = 2m$. It is generated by the null geodesics $\mathcal{L}_{(\theta, \varphi)}^{\text{out}, \mathcal{H}}$ discussed in Sec. 6.3.5 [cf. Eq. (6.52)]. Let us consider one such geodesic, $\hat{\mathcal{L}}$ say. It has a fixed value of (θ, φ) and, when followed in the past direction, it encounters the outer edge of the radiation region \mathcal{M}_{rad} (hypersurface $v = v_0$) at the point A such that $r_A = 2m$ and $t_A = v_0 - 2m$ (cf. Fig. 15.1, where $\hat{\mathcal{L}}$ can be identified with the black curve). If $\hat{\mathcal{L}}$ is prolonged into \mathcal{M}_{rad} , still in the past direction, it encounters the inner edge of \mathcal{M}_{rad} (hypersurface $v = 0$) at the point B (cf. Fig. 15.1). The portion AB of $\hat{\mathcal{L}}$ coincides with the geodesic $\mathcal{L}_{(u_B, \theta, \varphi)}^{\text{out}}$ of the outgoing radial null family given by Eq. (15.28), with $u_B = -\ln(r_B/m)$ (since $f_\alpha(x_B) = 1$, given that $x_B = 0$). We have then $r_A = r_B f_\alpha(x_A)$. Now, by definition of x [Eq. (15.21)] and α [Eq. (15.18)], $x_A = v_A/r_A = v_0/(2m) = \alpha^{-1}$. We have thus $2m = r_B f_\alpha(\alpha^{-1})$. In view of expression (15.27) for $f_\alpha(x)$, there comes

$$r_B = 2m \exp \left[-\frac{2}{\sqrt{8\alpha - 1}} \arctan \left(\frac{1}{\sqrt{8\alpha - 1}} \right) \right].$$

Since B is located on the hypersurface $v = 0$, we have $t_B = -r_B$. If the radial null geodesic $\hat{\mathcal{L}}$ is prolonged further to the past in the Minkowski region \mathcal{M}_{Min} , it becomes the straight line of equation $t = r - 2r_B$. $\hat{\mathcal{L}}$ thus reaches $r = 0$ at some event C of coordinate $t = t_{hb} = -2r_B$, hence Eq. (15.30).

There remains to prove that the null hypersurface generated by $\hat{\mathcal{L}}$ when (θ, φ) varies, i.e. the future light cone \mathcal{H} of the point $(t, r) = (t_{\text{hb}}, 0)$, is indeed the black hole event horizon. To this aim, let us consider an outgoing radial null geodesic \mathcal{L} in \mathcal{M}_{Min} such that \mathcal{L} crosses $r = 0$ at $t < t_{\text{hb}}$, i.e. outside \mathcal{H} . \mathcal{L} arrives then at the inner edge of \mathcal{M}_{rad} with $r = r_0 > r_B$. In \mathcal{M}_{rad} , according to Eq. (15.28), \mathcal{L} is homothetic to a part of $\hat{\mathcal{L}}$ with a ratio $r_0/r_B > 1$. Hence it emerges at the outer edge of \mathcal{M}_{rad} with $r > r_A = 2m$. \mathcal{L} is there in the exterior of the Schwarzschild black hole, so that it will subsequently reach the future null infinity \mathcal{I}^+ of \mathcal{M}_{Sch} . On the contrary, if \mathcal{L} crosses $r = 0$ with $t > t_{\text{hb}}$, i.e. inside \mathcal{H} , it encounters the inner edge of \mathcal{M}_{rad} with $r = r_0 < r_B$. A part of \mathcal{L} is then homothetic to the segment BA of $\hat{\mathcal{L}}$ with a ratio $r_0/r_B < 1$. Then either (i) \mathcal{L} has a r -turning point and reaches $r = 0$ in \mathcal{M}_{rad} or (ii) \mathcal{L} reaches the outer edge of \mathcal{M}_{rad} ($v = v_0$) at some point A' . Given that $x_A = \alpha^{-1}$ corresponds to the maximum of $f_\alpha(x)$, one has necessarily $f_\alpha(x_{A'}) \leq f_\alpha(x_A)$ and thus $r_0 f_\alpha(x_{A'}) < r_B f_\alpha(x_A)$. By Eq. (15.28), this implies $r_{A'} < r_A = 2m$, so that \mathcal{L} emerges in the black hole region of Schwarzschild spacetime. So none of the two possible cases (i) or (ii) leads to \mathcal{L} reaching \mathcal{I}^+ . We conclude that \mathcal{H} is a black hole horizon. \square

The black hole event horizon \mathcal{H} is depicted as the thick black curve in Fig. 15.1. Since $t_{\text{hb}} < 0$ [cf. Eq. (15.30)], one immediately notices:

Property 15.8: black hole formation in the Minkowski region

The black hole forms in the Minkowski region of Vaidya spacetime, i.e. in a region where the spacetime curvature is zero.

This striking feature reflects the non-local character of black holes and will be discussed further in Chap. 18.

The dependency of t_{hb} on the width v_0 of the radiation shell is shown in Fig. 15.4. One has $-4m < t_{\text{hb}} < 0$, with

$$\lim_{v_0 \rightarrow 0} t_{\text{hb}} = -4m \quad \text{and} \quad \lim_{v_0 \rightarrow 16m} t_{\text{hb}} = 0. \quad (15.31)$$

Remark 4: The first limit in (15.31), which corresponds to $\alpha \rightarrow +\infty$ in Eq. (15.30), is easily recovered by a pure geometric construction: a zero-width shell implies the equality of the two points A and B considered in the proof of (15.30), as well as $t_A = t_B = -2m$, hence $t_C = t_{\text{hb}} = -4m$.

Historical note : The imploding radiation shell with $M(v)$ piecewise linear (homothetic model) has been extensively studied by Achilles Papapetrou in 1985 [399]. He obtained the solution (15.28) for the outgoing radial null geodesics and determined the location of the event horizon, along the lines presented above.

15.3.4 Curvature singularity

The Kretschmann scalar $K := R_{\mu\nu\rho\sigma} R^{\mu\nu\rho\sigma}$ (cf. Sec. 6.3.4) is computed in the notebook D.6.4:

$$K = \frac{48M(t+r)^2}{r^6}. \quad (15.32)$$

K is identically zero in the Minkowski region ($M(r + t) = 0$), as it should!

It diverges at $r = 0$ in the radiation and Schwarzschild regions ($M(r + t) > 0$), tracing the existence of a curvature singularity there. In other words:

Property 15.9: curvature singularity in Vaidya shell collapse

The Vaidya shell collapse introduced in Sec. 15.3.1 generates a spacetime with a curvature singularity located at $r = 0$ and $t \geq 0$.

Remark 5: The Kretschmann scalar of Vaidya metric has the same structural form as that of the Schwarzschild metric, compare Eq. (6.45), while a priori K could have contained some term involving the derivative of $M(v)$. Indeed $M'(v)$ appears in some components of the Riemann tensor, since it is present in the components of the Ricci tensor, as given by Eq. (15.4).

Remark 6: Since the Ricci tensor of the Vaidya metric is not identically zero [cf. Eq. (15.4)], other curvature invariants that one might have thought of for tracking the curvature singularity are the Ricci scalar $R := g^{\mu\nu}R_{\mu\nu}$ and the Ricci “squared” $R_{\mu\nu}R^{\mu\nu}$. However, they are both identically zero for \mathbf{k} is a null vector.

The curvature singularity is depicted as the orange broken line in Fig. 15.1, which regards the homothetic radiation shell with $\alpha > 1/8$. It is clear on this figure that the singularity bounds the future of both ingoing and outgoing radial null geodesics (this will be shown rigorously in Sec. 15.3.6). This implies that, for $\alpha > 1/8$, the curvature singularity is *spacelike*, as in Schwarzschild spacetime. We shall see in Sec. 15.4 that for $\alpha < 1/8$, the curvature singularity has a *null* segment, in addition to the spacelike one.

For the homothetic collapse with $\alpha > 1/8$ ($v_0 < 16m$) considered in Secs. 15.3.2 and 15.3.3 and depicted in Fig. 15.1, one has $t_{\text{hb}} < 0$ [Eq. (15.30)], so that the curvature singularity is entirely located in the black hole region. It is therefore hidden to a remote observer. In Sec. 15.4, we will see that this is no longer the case for $\alpha < 1/8$: the singularity is then naked.

15.3.5 Trapped surfaces

Let us show that there exist trapped surfaces (cf. Sec. 3.2.3) in the radiation and Schwarzschild regions of Vaidya spacetime. To this aim, we consider a 2-surface \mathcal{S} defined by $(t, r) = \text{const.}$ It is a closed surface with the topology of a 2-sphere, spanned by the coordinates (θ, φ) . Moreover, \mathcal{S} is spacelike since the metric induced by \mathbf{g} on it is $q = r^2(d\theta^2 + \sin^2\theta d\varphi^2)$, as readily seen on Eq. (15.11). The area of \mathcal{S} is simply $4\pi r^2$ (r is the areal radius coordinate), so to check whether \mathcal{S} is a trapped surface, it suffices to determine the behavior of r along the two null directions normal to \mathcal{S} , which are nothing but the directions of the ingoing and outgoing radial null geodesics. Along the ingoing geodesics (tangent vector \mathbf{k}), one has $dr/dt = -1$, since $\mathbf{k} = \partial_t - \partial_r$ [Eq. (15.12)], so that the expansion is negative: $\theta_{(k)} < 0$. Along the outgoing radial null geodesics, dr/dt is given by Eq. (15.14). The sign of the expansion $\theta_{(\ell)}$ is then that of $r - 2M(t + r)$. Hence

Property 15.10: trapped surfaces in Vaidya shell collapse

The spherically symmetric surface \mathcal{S} defined by $(t, r) = \text{const}$ obeys

$$\mathcal{S} \text{ is trapped} \iff r < 2M(t + r). \quad (15.33)$$

Obviously, the above criterion cannot be fulfilled in the Minkowski region, where $M(t + r) = 0$. On the contrary, trapped surfaces exist in the central part of the radiation region, since $M(t + r) > 0$ there. They also exist in the black hole part ($r < 2m$) of the Schwarzschild region, where $M(t + r) = m$.

Let us denote by \mathcal{T} the hypersurface formed by all the *marginally* trapped spheres of fixed (t, r) (vanishing expansion along ℓ). The equation of \mathcal{T} is obtained by saturating the inequality in Eq. (15.33):

$$\mathcal{T} : \quad r = 2M(t + r). \quad (15.34)$$

As we shall see in Chap. 18, \mathcal{T} is called a ***future outer trapping horizon***. Equation (15.34) implies $dr/dt = 0$ in Eq. (15.14), hence we get

Property 15.11: trapping horizon and r -turning points of radial null geodesics

If an outgoing radial null geodesic \mathcal{L} crosses \mathcal{T} , the crossing point is a r -turning point of \mathcal{L} .

This feature appears clearly in Figs. 15.1 and 15.5. Furthermore, \mathcal{T} obeys the following properties:

Property 15.12: causal type of the future outer trapping horizon

In the radiation region, the future outer trapping horizon \mathcal{T} is a spacelike hypersurface, while in the Schwarzschild region, \mathcal{T} coincides with the event horizon \mathcal{H} and hence is a null hypersurface there.

Proof. Let us consider (v, θ, φ) as coordinates on the 3-manifold \mathcal{T} . We may rewrite Eq. (15.34) for \mathcal{T} as $r = 2M(v)$, so that $dr = 2M'(v) dv$ along \mathcal{T} . Plugging this relation, as well as $2M(v)/r = 1$, into Eq. (15.1) yields immediately the metric \mathbf{h} induced by \mathbf{g} on \mathcal{T} : $\mathbf{h} = 4M'(v) dv^2 + r^2(d\theta^2 + \sin^2 \theta d\varphi^2)$. In \mathcal{M}_{rad} , $M'(v) > 0$ [cf. Eq. (15.15)], which implies that \mathbf{h} is a positive definite metric. Hence \mathcal{T} is a spacelike hypersurface in \mathcal{M}_{rad} . In \mathcal{M}_{Sch} , $M'(v) = 0$ and \mathbf{h} is a degenerate metric, so that \mathcal{T} is a null hypersurface there. Actually, in \mathcal{M}_{Sch} , Eq. (15.34) reduces to $r = 2m$, so that \mathcal{T} coincides with the event horizon \mathcal{H} there. \square

Remark 7: The spacelike character of \mathcal{T} is actually a generic feature of future outer trapping horizons in non-stationary spacetimes (such as $(\mathcal{M}_{\text{rad}}, \mathbf{g})$), as we shall see in Chap. 18.

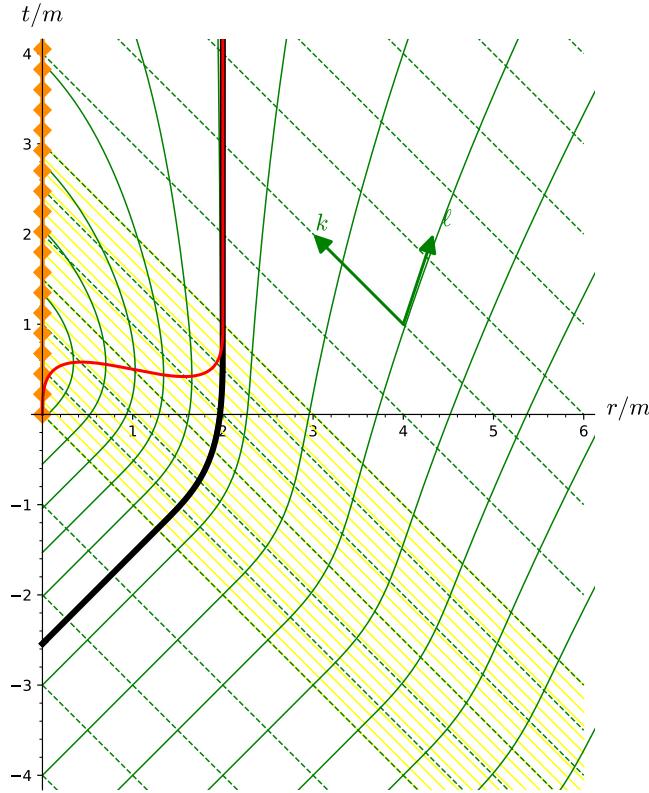


Figure 15.5: Same as Fig. 15.1, but for the C^2 mass function $M(v)$ corresponding to the choice $S(x) = 6x^5 - 15x^4 + 10x^3$ in Eq. (15.15) (cf. the red curve in Fig. 15.2). Note that the trapping horizon \mathcal{T} (red curve) is tangent to the event horizon \mathcal{H} (black curve) at the outer edge of radiation region $((t, r) = (m, 2m))$. [Figure generated by the notebook D.6.4]

For the homothetic shell model considered in Sec. 15.3.2, $M(v) = mv/v_0 = \alpha v/2$ and the equation of \mathcal{T} in \mathcal{M}_{rad} is very simple:

$$\mathcal{T} \cap \mathcal{M}_{\text{rad}} : \quad r = \alpha v. \quad (15.35)$$

Example 1: For $v_0 = 3m$, as in Fig. 15.1, $\alpha = 2/3$ and we get $r = 2v/3 = 2(t+r)/3$ or equivalently $t = r/2$. Hence, in \mathcal{M}_{rad} , \mathcal{T} appears as the straight line segment of slope $1/2$ drawn in red in Fig. 15.1. It appears clearly there that the outgoing radial null geodesics that cross \mathcal{T} do it at a r -turning point. By considering the light cones delineated by the ingoing and outgoing radial null geodesics, it also appears clearly on Fig. 15.1 that \mathcal{T} is a spacelike hypersurface in \mathcal{M}_{rad} .

Remark 8: As shown above, \mathcal{T} coincides with \mathcal{H} in \mathcal{M}_{Sch} , so that the slope of \mathcal{T} in Fig. 15.1 changes abruptly from $1/2$ to $+\infty$ at the outer edge of \mathcal{M}_{rad} . This lack of smoothness of the hypersurface \mathcal{T} reflects actually the lack of smoothness of the function $M(v)$ of the homothetic shell model (cf. the blue curve in Fig. 15.2). Would $M(v)$ be smooth, \mathcal{T} would merge smoothly with \mathcal{H} at $v = v_0$. This is shown in Fig. 15.5, where $M(v)$ is of class C^2 . Besides, we can check on this figure that, while it has a shape more complicated than in Fig. 15.1, \mathcal{T} always lies outside the null cones in \mathcal{M}_{rad} , i.e. \mathcal{T} is spacelike there.

15.3.6 Carter-Penrose diagram

In order to draw a Carter-Penrose diagram of the Vaidya collapse, we need first to introduce a coordinate system (u, v, θ, φ) in the radiation region \mathcal{M}_{rad} such that u is constant along the outgoing radial null geodesics – v being constant along the ingoing ones by construction. For the homothetic model ($M(v) = \alpha v/2$ in \mathcal{M}_{rad}), the relation between coordinates (u, v, θ, φ) and (v, r, θ, φ) is provided by Eq. (15.28):

$$u = -\ln\left(\frac{r}{m}\right) + \ln f_\alpha\left(\frac{v}{r}\right). \quad (15.36)$$

From Eqs. (15.36) and (15.26), we get

$$du = -\frac{1}{r}dr + \frac{1-\alpha x}{\alpha x^2 - x + 2}d\left(\frac{v}{r}\right) = -\frac{1}{r(\alpha x^2 - x + 2)}[(\alpha x - 1)dv + 2dr].$$

By comparing with (15.17), we deduce immediately the expression of the metric tensor in terms of the coordinates (u, v, θ, φ) :

$$\mathbf{g} = -r(\alpha x^2 - x + 2)du\,dv + r^2(d\theta^2 + \sin^2\theta\,d\varphi^2) \quad (0 \leq v \leq v_0), \quad (15.37)$$

where $x := v/r$ and r is to be considered as a function of (u, v) defined implicitly by Eq. (15.36). Since for $\alpha > 1/8$, the polynomial $\alpha x^2 - x + 2$ never vanishes, the metric component g_{uv} read on Eq. (15.37) is nonzero in all \mathcal{M}_{rad} . This shows that the coordinates (u, v, θ, φ) are regular on \mathcal{M}_{rad} (except for the singularities inherent to the spherical coordinates (θ, φ)). Moreover, they constitute a **double-null coordinate system**: both u and v are null coordinates. This has already been shown for v (cf. Property 15.1) and for u , this follows from the component g^{uu} of the inverse metric deduced from (15.37) being identically zero [cf. Eq. (A.56b)]. Besides, we read on Eq. (15.37) that $g_{uu} = 0$ and $g_{vv} = 0$, which means that the coordinate vectors ∂_u and ∂_v are both null. The vector ∂_u is actually tangent to the ingoing radial null geodesics $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$ and the vector ∂_v is tangent to the outgoing ones, $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$. Furthermore, the null vectors ∂_u and ∂_v are both future-directed. Indeed, the time orientation of Vaidya spacetime is given by the vector \mathbf{k} (cf. Sec. 15.2.1) and we deduce from Eq. (15.3) that $\mathbf{k} = -(\partial u / \partial r)\partial_u$. Evaluating $\partial u / \partial r$ from Eqs. (15.36) and (15.26), we get

$$\mathbf{k} = \frac{2}{r(\alpha x^2 - x + 2)}\partial_u. \quad (15.38)$$

The coefficient in front of ∂_u being always positive, we conclude that ∂_u is future-directed. Then, from Eq. (15.37), $\mathbf{g}(\partial_u, \partial_v) = g_{uv} < 0$, so that, according to Lemma 1.2 (Sec. 1.2.2), ∂_v is future-directed as well.

Remark 9: The reader may have noticed a slight asymmetry between the coordinates u and v : u is dimensionless, while v has the dimension of a time. We could of course make u have the same dimension as v by introducing an overall factor m in the right-hand side of Eq. (15.36). However, this would make the formulas slightly more complicate, without any real benefit.

Let us discuss the boundary of \mathcal{M}_{rad} in terms of the (u, v) coordinates. By definition of \mathcal{M}_{rad} , a part of the boundary consists in the hypersurfaces $v = 0$ and $v = v_0$. Another part

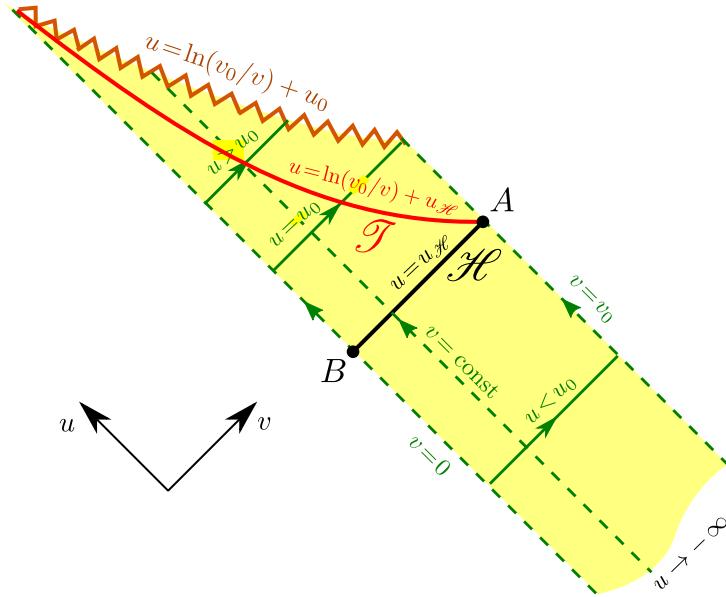


Figure 15.6: Radiation region \mathcal{M}_{rad} for the homothetic model ($M(v) = mv/v_0$) with $v_0 < 16m$, depicted in terms of the double-null coordinates (u, v) . Solid (resp. dashed) green lines are outgoing (resp. ingoing) radial null geodesics. The orange zigzag curve corresponds to the curvature singularity at $r = 0$. The black segment marks the event horizon \mathcal{H} , while the red curve marks the trapping horizon \mathcal{T} [Eq. (15.42)]. The points A and B are the same as in Fig. 15.1.

corresponds to the limit $r \rightarrow +\infty$. In view of Eq. (15.36) and $f_\alpha(0) = 1$ (cf. Eq. (15.26) or Fig. 15.3), this corresponds to $u \rightarrow -\infty$. The last part of the boundary of \mathcal{M}_{rad} is set by the curvature singularity at $r = 0$ (cf. Sec. 15.3.4). Taking the limit $r \rightarrow 0$ in Eq. (15.36), we found the equation ruling this boundary in terms of (u, v) :

$$u = \ln\left(\frac{v_0}{v}\right) + u_0, \quad u_0 := \ln\sqrt{\frac{\alpha}{2}} + \frac{1}{\sqrt{8\alpha-1}} \left[\frac{\pi}{2} + \arctan\left(\frac{1}{\sqrt{8\alpha-1}}\right) \right]. \quad (15.39)$$

We conclude that the range of the coordinates (u, v) on \mathcal{M}_{rad} is (cf. Fig. 15.6)

$$\mathcal{M}_{\text{rad}} : \quad \begin{cases} 0 \leq v \leq v_0 \\ -\infty < u < \ln\left(\frac{v_0}{v}\right) + u_0. \end{cases} \quad (15.40)$$

In the last inequality, the right-hand side must be replaced by $+\infty$ if $v = 0$.

Along a given ingoing radial null geodesic $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$, u can be considered as a (non-affine) parameter, ∂_u being tangent to $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$. The geodesic hits the curvature singularity for $u \rightarrow \ln(v_0/v) + u_0$. Similarly, along an outgoing radial null geodesic $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$, v can be considered as a (non-affine) parameter, ∂_v being tangent to $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$. If $u < u_0$, $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ reaches the outer boundary of the radiation shell for $v = v_0$ and is extendible to a null geodesic of the Schwarzschild region, while if $u \geq u_0$, $\mathcal{L}_{(u,\theta,\varphi)}^{\text{out}}$ hits the curvature singularity for $v \rightarrow v_0 e^{u_0-u}$ (cf. Fig. 15.6).

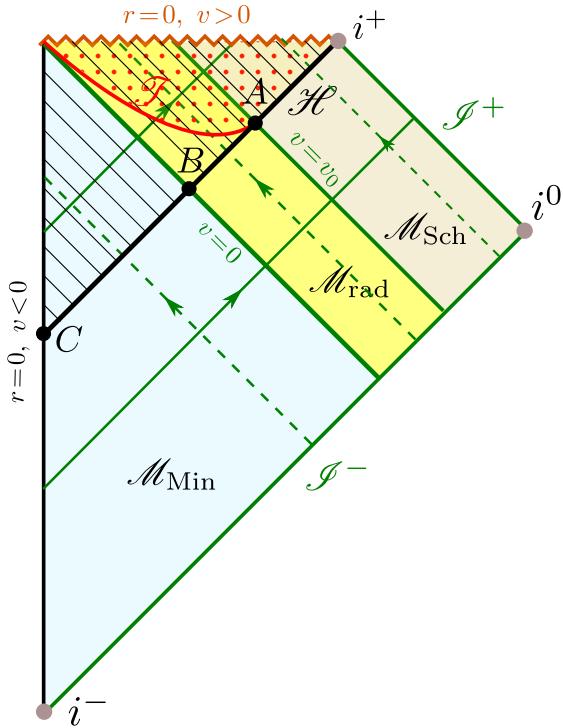


Figure 15.7: Carter-Penrose diagram of Vaidya collapse for the homothetic shell model ($M(v) = mv/v_0 = \alpha v/2$) with $\alpha > 1/8$ ($v_0 < 16m$). The Minkowski region \mathcal{M}_{Min} , radiation region \mathcal{M}_{rad} and Schwarzschild region \mathcal{M}_{Sch} are depicted in respectively pale blue, yellow and grey. Solid (resp. dashed) green lines are outgoing (resp. ingoing) radial null geodesics. The orange zigzag segment corresponds to the curvature singularity at $r = 0$. The hatched area is the black hole region, delimited by the event horizon \mathcal{H} . The red dot-filled area, delimited by the future outer trapping horizon \mathcal{T} (red curve), is the region where the spheres $(t, r) = \text{const}$ are trapped surfaces. The points A , B and C along \mathcal{H} are the same as in Fig. 15.1.

Since it is generated by outgoing null geodesics, the event horizon \mathcal{H} crosses \mathcal{M}_{rad} at a fixed value of u , $u_{\mathcal{H}}$ say. The value of $u_{\mathcal{H}}$ is found by setting $r = 2m$ and $v = v_0$ in Eq. (15.36):

$$u_{\mathcal{H}} = \frac{2}{\sqrt{8\alpha - 1}} \arctan \left(\frac{1}{\sqrt{8\alpha - 1}} \right) - \ln 2. \quad (15.41)$$

Given expression (15.39) for u_0 , it is easy to check that $u_{\mathcal{H}} < u_0$, as drawn in Fig. 15.6.

Regarding the future outer trapping horizon \mathcal{T} introduced in Sec. 15.3.5, it obeys $r = \alpha v$ in \mathcal{M}_{rad} [Eq. (15.35)], so that the equation of \mathcal{T} in terms of the double-null coordinates is easily obtained by substituting α^{-1} for v/r in Eq. (15.36) and making use of expressions (15.27) and (15.41); one gets

$$\mathcal{T} : \quad u = \ln \left(\frac{v_0}{v} \right) + u_{\mathcal{H}}. \quad (15.42)$$

Note that this implies $u \rightarrow +\infty$ for $v \rightarrow 0$ and $u \rightarrow u_{\mathcal{H}}$ for $v \rightarrow v_0$; the latter property agrees with \mathcal{T} coinciding with \mathcal{H} in \mathcal{M}_{Sch} , i.e. for $v > v_0$. \mathcal{T} is drawn as a red curve in Fig. 15.6.

The spacetime diagram of \mathcal{M}_{rad} in Fig. 15.6 is already conformal since all radial null geodesics are straight lines inclined at $\pm 45^\circ$. To integrate it into a Carter-Penrose diagram, one needs

to compactify it along the u direction by introducing a coordinate of the type $U = \arctan u$ – the v direction being already compactified, given the finite range of v in \mathcal{M}_{rad} . One can then match the obtained diagram to Carter-Penrose diagrams of \mathcal{M}_{Min} (cf. Fig. 4.6) and \mathcal{M}_{Sch} (cf. Fig. 9.10 or 9.11), thereby getting a Carter-Penrose diagram of the whole Vaidya spacetime, as shown in Fig. 15.7.

Historical note : The double-null coordinates (u, v) , leading to expression (15.37) of the metric tensor in \mathcal{M}_{rad} , have been introduced by B. Waugh and Kayll Lake in 1986 [515]. A Carter-Penrose diagram similar to that of Fig. 15.7 except for the radiation region extending to \mathcal{I}^+ (no pure Schwarzschild exterior but $M'(v) \rightarrow 0$ for $v \rightarrow +\infty$) has been exhibited by Yuhji Kuroda in 1984 [329] (cf. his Fig. 2b). The case of a radiation region bounded by $v = v_0$, as here, can be found in Fig. 2 of a 2007 article by Brien C. Nolan [383].

15.4 Configurations with a naked singularity

15.4.1 The low radiation density case

In Sec. 15.3, we have focused on homothetic radiation shells ($M(v) = \alpha v/2$) with $\alpha > 1/8$ [Eq. (15.23)]. Let us now discuss the opposite case⁴:

$$\alpha < \frac{1}{8} \iff v_0 > 16m. \quad (15.43)$$

In view of the form (15.24) of the energy momentum tensor, this corresponds to a low energy density of the radiation field. When (15.43) is fulfilled, the polynomial $P_\alpha(x) := \alpha x^2 - x + 2$, which appears in the numerator of the ODE (15.22) ruling outgoing radial null geodesics in \mathcal{M}_{rad} , admits two real roots:

$$x_1 := \frac{1 - \sqrt{1 - 8\alpha}}{2\alpha} \quad \text{and} \quad x_2 := \frac{1 + \sqrt{1 - 8\alpha}}{2\alpha}. \quad (15.44)$$

Useful identities are $x_1 + x_2 = \alpha^{-1}$ and $x_1 x_2 = 2\alpha^{-1}$. Note also that $2 < x_1 < 4 < x_2 < \alpha^{-1}$. Equation (15.22) can be recast as

$$r \frac{dx}{dr} = \frac{(x - x_1)(x - x_2)}{x_1 + x_2 - x}. \quad (15.45)$$

This differential equation admits two special solutions, corresponding to outgoing radial null geodesics with constant value of x :

$$x = x_1 \quad \text{and} \quad x = x_2. \quad (15.46)$$

Let us denote these geodesics by respectively $\mathcal{L}_{(\theta,\varphi)}^{*1}$ and $\mathcal{L}_{(\theta,\varphi)}^{*2}$. Since $x := v/r = 1 + t/r$, their equations in terms of the IEF coordinates (t, r, θ, φ) and in \mathcal{M}_{rad} is simply

$$\mathcal{L}_{(\theta,\varphi)}^{*1} : t = (x_1 - 1)r \quad \text{and} \quad \mathcal{L}_{(\theta,\varphi)}^{*2} : t = (x_2 - 1)r. \quad (15.47)$$

Thus, in \mathcal{M}_{rad} , $\mathcal{L}_{(\theta,\varphi)}^{*1}$ and $\mathcal{L}_{(\theta,\varphi)}^{*2}$ are straight line segments through the origin $(t, r) = (0, 0)$. They are depicted as the segments OC and OB respectively in Fig. 15.8.

⁴The marginal case $\alpha = 1/8$ will not be discussed here; it is actually qualitatively similar to the case $\alpha < 1/8$ insofar as it leads to a naked singularity as well [399].

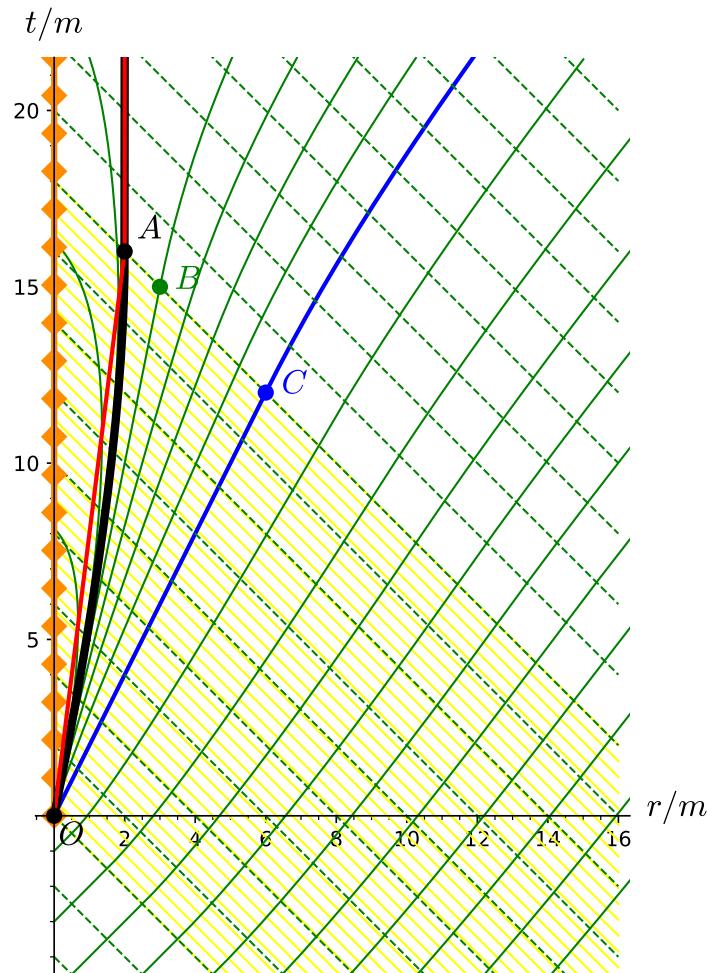


Figure 15.8: Spacetime diagram of the Vaidya collapse based on the IEF coordinates (t, r) and for the linear mass function $M(v) = mv/v_0$ with $v_0 = 18 m$ (homothetic shell model with $\alpha = 1/9$, which corresponds to $x_1 = 3$ and $x_2 = 6$). The legend is the same as in Fig. 15.1, with in addition the blue line marking the Cauchy horizon \mathcal{H}_C induced by the naked singularity at $(t, r) = (0, 0)$ (Sec. 15.4.4). In the radiation region, the straight line segments OB and OC are the traces in the (t, r) plane of the homothetic Killing horizons $x = x_2$ and $x = x_1$, the last one being a part of the Cauchy horizon. [Figure generated by the notebook [D.6.4](#)]

Property 15.13: homothetic Killing horizons

When (θ, φ) varies, $\mathcal{L}_{(\theta,\varphi)}^{*1}$ and $\mathcal{L}_{(\theta,\varphi)}^{*2}$ generate null hypersurfaces, \mathcal{H}_1 and \mathcal{H}_2 respectively, such that the homothetic Killing vector ξ [cf. Eq. (15.19)] is normal to them in \mathcal{M}_{rad} . One says that \mathcal{H}_1 and \mathcal{H}_2 are **homothetic Killing horizons**^a.

^aRecall from Sec. 3.3.2 that a *Killing horizon* is a null hypersurface such that a Killing vector is normal to it.

Proof. Their equations being $t = (x_1 - 1)r$ and $t = (x_2 - 1)r$, $\mathcal{L}_{(\theta,\varphi)}^{*1}$ and $\mathcal{L}_{(\theta,\varphi)}^{*2}$ are clearly orbits of the homothety group $(H_\lambda)_{\lambda \in \mathbb{R}_{>0}}$ discussed in Sec. 15.3.2. The group generator ξ is thus tangent to the null curves $\mathcal{L}_{(\theta,\varphi)}^{*1}$ on \mathcal{H}_1 and $\mathcal{L}_{(\theta,\varphi)}^{*2}$ on \mathcal{H}_2 . Hence, ξ is null there. Given that the only null direction in a null hypersurface is the normal one, we conclude that ξ is normal to \mathcal{H}_1 and \mathcal{H}_2 . \square

Let us now consider *generic* outgoing radial null geodesics, i.e. geodesics with $x \neq x_1$ and $x \neq x_2$; we may then rewrite Eq. (15.45) as

$$d \ln r = \frac{x_1 + x_2 - x}{(x - x_1)(x - x_2)} dx.$$

This equation is easily integrated to

$$r = r_0 \frac{|x/x_2 - 1|^{x_1/(x_2-x_1)}}{|x/x_1 - 1|^{x_2/(x_2-x_1)}},$$

(15.48)

where r_0 is constant along the considered geodesic (but may vary from one geodesic to the other). If the geodesic intersects \mathcal{M}_{rad} 's inner edge ($x = 0$), then r_0 represents the value of r at the intersection point.

Example 2: For $\alpha = 1/9$, one has $x_1 = 3$ and $x_2 = 6$ [cf. Eq. (15.44)] and Eq. (15.48) reduces to $r = r_0|x/6 - 1|/(x/3 - 1)^2$. Such a function of x is plotted in Fig. 15.9.

Having determined the outgoing radial null geodesics, we are in position to establish the following results:

Property 15.14: black hole and naked singularity for the $\alpha < 1/8$ model

The homothetic imploding radiation shell with $\alpha := 2m/v_0 < 1/8$ generates a black hole, the event horizon of which originates at $(t, r) = (0, 0)$, with a slope $t/r = x_2 - 1$. Moreover, the curvature singularity at $(t, r) = (0, 0)$ is **naked**: it is connected to remote observers by the outgoing radial null geodesics $\mathcal{L}_{(\theta,\varphi)}^{*1}$, as well as by the family of outgoing radial null geodesics that cross the outer edge of the radiation shell ($v = v_0$) with $x_1 < x < \alpha^{-1}$; all geodesics of this family, which contains $\mathcal{L}_{(\theta,\varphi)}^{*2}$, emanate from $(t, r) = (0, 0)$ with a slope $t/r = x_2 - 1$.

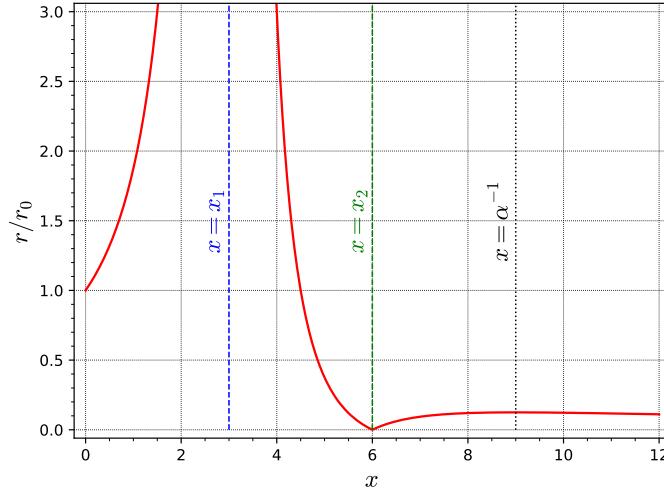


Figure 15.9: The coordinate r as a function of $x := v/r$ along outgoing radial null geodesics for the homothetic Vaidya collapse with $\alpha = 1/9$ (which implies $x_1 = 3$ and $x_2 = 6$). The graph of $r(x)$ admits a local maximum for $x = \alpha^{-1}$, i.e. $x = 9$ in the present case, although this is barely noticeable on the figure. [Figure generated by the notebook D.6.5]

Proof. As in Sec. 15.3.3, let us consider a null generator $\hat{\mathcal{L}}$ of the Killing horizon of the Schwarzschild region \mathcal{M}_{Sch} , which is located at $r = 2m$. $\hat{\mathcal{L}}$ intersects the outer edge $v = v_0$ of \mathcal{M}_{rad} at a point A such that $r_A = 2m$ and hence $x_A = \alpha^{-1}$ (cf. Fig. 15.8). When prolonged to the past in \mathcal{M}_{rad} , i.e. to $x < \alpha^{-1}$, $\hat{\mathcal{L}}$ has r decreasing (cf. the graph of $r(x)$ at the left of $x = \alpha^{-1}$ in Fig. 15.9), until $r = 0$ is reached for $x = x_2$. Since the Kretschmann scalar (15.32) is $K = 12\alpha^2 x^2/r^4$ for the homothetic model, we get $K \rightarrow \infty$ for $r \rightarrow 0$ when $x \rightarrow x_2 > 0$. We thus conclude that $\hat{\mathcal{L}}$ hits the curvature singularity at $x = x_2$ and cannot be extended further in the past, contrary to the case $\alpha > 1/8$ dealt with in Sec. 15.3.3.

Let us now consider any outgoing radial null geodesic \mathcal{L} in \mathcal{M}_{Sch} that intersects the outer edge $v = v_0$ of \mathcal{M}_{rad} with x such that $x_1 < x < \alpha^{-1}$, i.e. between the points C and A in Fig. 15.8. \mathcal{L} has $r > 2m$ in \mathcal{M}_{Sch} , hence it reaches the future null infinity \mathcal{I}^+ . When \mathcal{L} is prolonged backward in \mathcal{M}_{rad} , it starts with r decaying since Eq. (15.14) can be rewritten $dr/dt = (1 - \alpha x)/(1 + \alpha x)$ for the homothetic model, implying $dr/dt < 0$. The entire past of \mathcal{L} in \mathcal{M}_{rad} can then be read on Fig. 15.9: if one follows the decaying r direction either for $x_1 < x < x_2$ or $x_2 < x < \alpha^{-1}$, one ends up to $r = 0$ for $x = x_2$, as for $\hat{\mathcal{L}}$. Hence the same conclusion holds: \mathcal{L} hits the curvature singularity in the past direction at $x = x_2$. Given that \mathcal{L} extends to \mathcal{I}^+ in the future, we conclude that the curvature singularity is naked.

Finally, let us consider an outgoing radial null geodesic \mathcal{L} in \mathcal{M}_{Sch} that intersects the outer edge $v = v_0$ of \mathcal{M}_{rad} with $x < x_1$, i.e. below the point C in Fig. 15.8. When \mathcal{L} is prolonged backward in \mathcal{M}_{rad} , r decreases along it and we read on the left part of Fig. 15.9 that r reaches the finite nonzero value r_0 at $x = 0$, i.e. at the inner edge of \mathcal{M}_{rad} . It can then be extended backward to the Minkowski region \mathcal{M}_{Min} , where it reaches $r = 0$ at $t = t_0 := -2r_0$. Since for $x \rightarrow x_1^-$, $r/r_0 \rightarrow +\infty$ (cf. Fig. 15.9 or Eq. (15.48)), we conclude that r_0 , and hence $|t_0|$, can be made arbitrarily small by having x close enough to x_1 at the outer edge of \mathcal{M}_{rad} . This proves that the whole of \mathcal{M}_{Min} can be connected to \mathcal{I}^+ by this type of null geodesics. It follows that

the black hole event horizon is the null hypersurface generated by $\hat{\mathcal{L}}$. \square

Historical note : The formation of naked singularities in spacetimes with a Vaidya region has been put forward first by B. Steinmüller, Andrew R. King, and Jean-Pierre Lasota in 1975 [462], but this regards bodies radiating away all their masses, the exterior of which is described by the *outgoing* Vaidya metric (15.8). In the context of imploding radiating shells (*ingoing* Vaidya metric) considered here, the appearance of naked singularities has been shown first by William A. Hiscock, Leslie G. Williams and Douglas M. Eardley in 1982 [278] and has been further studied by Yuhji Kuroda in 1984 [329] and Achilles Papapetrou in 1985 [399]. In particular, the solution (15.48) for the outgoing radial null geodesics has been exhibited by Papapetrou: compare Eq. (18) in Ref. [399], where $X_1 = x_2$ and $X_2 = x_1$.

15.4.2 Analysis in double-null coordinate systems

The above result contains something puzzling at first glance: for a given value of (θ, φ) , there are distinct radial null geodesics emanating from the “point” $(t, r) = (0, 0)$, namely all the geodesics that cross the outer edge of the radiation shell with a slope $t/r = x - 1$, $x \in [x_1, \alpha^{-1}]$. This appears clearly on Fig. 15.8. It looks like there is an infinity of future light cones emanating from a single spacetime point! This state of affairs results actually from a bad behavior of the coordinates (t, r) near $(0, 0)$. To clarify this, let us introduce new coordinates (u, v, θ, φ) such that u is constant along the outgoing radial null geodesics, v being constant along the ingoing ones by construction. Let us start by substituting v/r for x in the geodesic equation (15.48); we get

$$\frac{|v/x_2 - r|^{x_1/(x_2-x_1)}}{|v/x_1 - r|^{x_2/(x_2-x_1)}} = \frac{1}{r_0}. \quad (15.49)$$

Contrary to the case $\alpha > 1/8$ dealt with in Sec. 15.3.6, two coordinate patches (u, v, θ, φ) and (u', v, θ, φ) are actually required to get regular double-null coordinates on \mathcal{M}_{rad} . They are defined on two overlapping subregions of \mathcal{M}_{rad} :

$$\mathcal{N}_I : r > \frac{v}{x_2} \quad \text{and} \quad \mathcal{N}_{II} : r < \frac{v}{x_1}. \quad (15.50)$$

Since $x_2 > x_1$, one has $\mathcal{M}_{\text{rad}} = \mathcal{N}_I \cup \mathcal{N}_{II}$. Furthermore, \mathcal{N}_I (resp. \mathcal{N}_{II}) contains the homothetic Killing horizon \mathcal{H}_1 (resp. \mathcal{H}_2) and $\mathcal{N}_I \cap \mathcal{N}_{II}$ is the region between \mathcal{H}_1 and \mathcal{H}_2 .

The \mathcal{N}_I region

On \mathcal{N}_I , let us define the parameter u so that

$$r_0 = |u|^{x_2/(x_2-x_1)}. \quad (15.51)$$

As r_0 , u is constant along a given outgoing radial null geodesic. From Eq. (15.49), we get⁵ $|u| = |v/x_1 - r|/(r - v/x_2)^{x_1/x_2}$. This equation determines u in terms of v and r up to some

⁵We have made use of the identity $|v/x_2 - r| = r - v/x_2$, which holds on \mathcal{N}_I .

sign. We choose the latter so that $u \rightarrow +\infty$ at the inner boundary of \mathcal{N}_I ($r \rightarrow v/x_2$). This yields

$$\boxed{u = \frac{v/x_1 - r}{(r - v/x_2)^{x_1/x_2}}}, \quad -\infty < u < +\infty, \quad (15.52)$$

with $\lim_{r \rightarrow +\infty} u = -\infty$ (cf. Fig. 15.10). We may use (u, v, θ, φ) as a coordinate system on \mathcal{N}_I , instead of the IEF coordinates (t, r, θ, φ) .

Example 3: As in Example 2, let us consider the case $\alpha = 1/9$, for which $x_1 = 3$ and $x_2 = 6$ [cf. Eq. (15.44)]. Equation (15.52) reduces then to

$$u = \frac{v/3 - r}{\sqrt{r - v/6}}. \quad (15.53)$$

This relation can be inverted, yielding an explicit expression for $r(u, v)$:

$$r = \frac{v}{3} + \frac{u}{2} \left(u - \sqrt{u^2 + \frac{2}{3}v} \right). \quad (15.54)$$

The metric components in coordinates (u, v, θ, φ) are deduced from those in coordinates (v, r, θ, φ) , i.e. Eq. (15.17), via the identity $\alpha = 2/(x_1 x_2)$. We get (see the notebook D.6.6 for the computation):

$$\mathbf{g} = -\frac{2x_2}{(x_2 - x_1)r} \left(r - \frac{v}{x_2} \right)^{x_1/2} \mathbf{d}u \mathbf{d}v + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2) \quad \left(r > \frac{v}{x_2} \right). \quad (15.55)$$

In this expression, r shall be considered as the function of (u, v) defined implicitly by Eq. (15.52). The metric component g_{uv} is regular and non-vanishing in all \mathcal{N}_I . Moreover we have a double-null coordinate system, since v is a null coordinate (Property 15.1) and it can be inferred from (15.55) that $g^{uu} = 0$, so that u is null as well by Eq. (A.56b). The coordinate vectors ∂_u and ∂_v are both null, given that $g_{uu} = 0$ and $g_{vv} = 0$. The vector ∂_u is actually tangent to the ingoing radial null geodesics, which are defined by $(v, \theta, \varphi) = \text{const}$, and the vector ∂_v is tangent to the outgoing ones, which are defined by $(u, \theta, \varphi) = \text{const}$. Furthermore, the homothetic Killing horizon \mathcal{H}_1 , which is defined by $v/r = x_1$, is the hypersurface $u = 0$ of \mathcal{N}_I .

Let us investigate the structure of the inner boundary $v = 0$ of the radiation shell in the double-null coordinates (u, v, θ, φ) . For $r \neq 0$, the limit $v \rightarrow 0$ in Eq. (15.52) leads to $u = -r^{1-x_1/x_2}$. This implies $u < 0$. In this part, which is the future boundary of the Minkowski region of Vaidya spacetime, this relation defines a bijection between $r \in (0, +\infty)$ and $u \in (0, -\infty)$. So we may say that r is a regular coordinate of \mathcal{N}_I for $u < 0$. On the other side, the limit $r \rightarrow 0$ in \mathcal{N}_I implies $v \rightarrow 0$ since $r > v/x_2$ in \mathcal{N}_I . Actually, the following relation holds:

$$v \underset{r \rightarrow 0}{\sim} x_2 r \quad \text{in } \mathcal{N}_I. \quad (15.56)$$

This follows directly from Eq. (15.48), which implies the equivalence

$$r \rightarrow 0 \iff (x \rightarrow x_2 \text{ or } x \rightarrow +\infty) \quad (15.57)$$

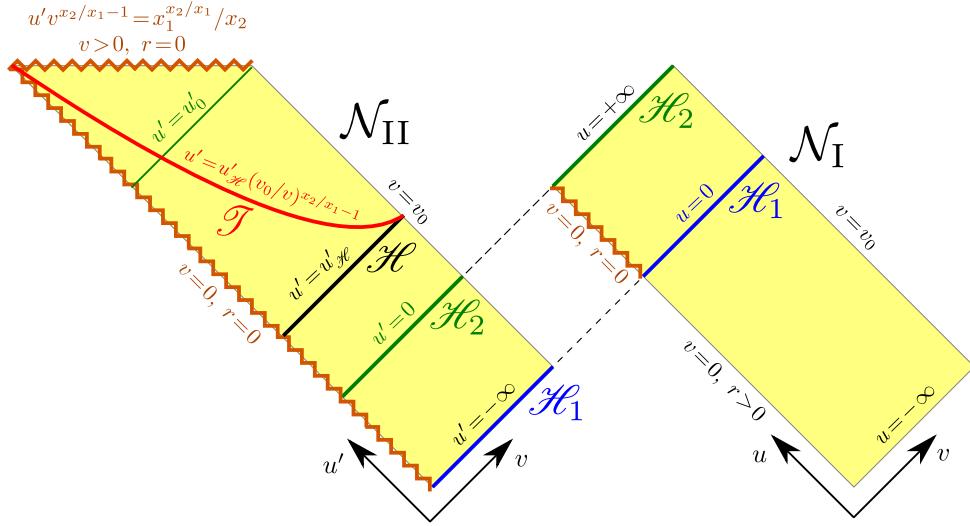


Figure 15.10: Carter-Penrose diagrams of the subregions \mathcal{N}_I (right) and \mathcal{N}_{II} (left) of the radiation region \mathcal{M}_{rad} . The region between the homothetic Killing horizons \mathcal{H}_1 and \mathcal{H}_2 is common to both \mathcal{N}_I and \mathcal{N}_{II} . Zigzag lines indicate a curvature singularity. The red curve in \mathcal{N}_{II} is the trapping horizon \mathcal{T} [Eq. (15.72)].

along any outgoing radial null geodesic, irrespective of the value of u (see also Fig. 15.9). Given that in \mathcal{N}_I , $x \rightarrow +\infty$ is excluded by the constraint $x < x_2$ [Eq. (15.50)], there remains $x \rightarrow x_2$, which yields the equivalence (15.56). Now, Eq. (15.56) implies that the numerator of Eq. (15.52) is equivalent to $(x_2/x_1 - 1)r$ at the limit $r \rightarrow 0$ and thus is positive, so that $u > 0$ on the part of the boundary of \mathcal{N}_I where $r \rightarrow 0$. We conclude that the “point” $(t, r) = (v, r) = (0, 0)$ in IEF coordinates becomes the hypersurface $v = 0, u > 0$ in the double-null coordinates. This means that the IEF coordinates are not adapted to describe the vicinity of $(t, r) = (0, 0)$ in Vaidya spacetime when $\alpha < 1/8$.

Example 4: Let us perform a first order expansion in v of the explicit expression of $r(u, v)$ found in Example 3 [Eq. (15.54)]. Taking into account the identity $\sqrt{u^2} = |u|$, we get

$$\begin{cases} r = u^2 + \frac{1}{3}v + O(v^2) & \text{for } u < 0 \\ r = \frac{1}{6}v + O(v^2) & \text{for } u > 0. \end{cases} \quad (15.58)$$

The first equation agrees with the relation $u = -r^{1-x_1/x_2}$ found above for $r \neq 0$ and $v = 0$, given that $x_1/x_2 = 1/2$ in the present case, while the last equation is exactly (15.56), given that $x_2 = 6$.

As discussed above, the double null coordinates (u, v, θ, φ) are regular coordinates on \mathcal{N}_I , because the components (15.55) of the metric tensor are regular (except for the standard singularities of the spherical coordinates (θ, φ)). Another advantage of these coordinates is that they allow for an immediate drawing of Carter-Penrose diagrams. It suffices to bring u to a finite range by setting e.g. $U = \arctan u$ and use the coordinates $T = U + v$ and $X = v - U$ to get such a diagram of \mathcal{N}_I , as depicted in Fig. 15.10.

Let us describe the (vicinity of the) curvature singularity in the double null coordinates. In

the present case, where $M(v) = \alpha v/2$, the Kretschmann scalar (15.32) reduces to

$$K = 12\alpha^2 \frac{v^2}{r^6}, \quad (15.59)$$

where r is the function of (u, v) given implicitly by Eq. (15.52) (or explicitly by Eq. (15.54) in the case $\alpha = 1/9$). The above expression shows that K diverges only if $r \rightarrow 0$. This limit occurs only for $v \rightarrow 0$ in the subregion $u \geq 0$ of \mathcal{N}_I . There, Eq. (15.56) leads to $K \sim 12\alpha^2 x_2^6/v^4$, which actually diverges as $v \rightarrow 0$. We conclude:

Property 15.15: curvature singularity in region \mathcal{N}_I

In \mathcal{N}_I , the curvature singularity corresponds to $u \geq 0$ and $v \rightarrow 0$.

This curvature singularity corresponds to the zigzag segment in right part of Fig. 15.10.

The \mathcal{N}_{II} region

On \mathcal{N}_{II} , let us introduce the parameter u' such that

$$r_0 = |u'|^{-x_1/(x_2-x_1)}. \quad (15.60)$$

Then Eq. (15.49) yields $|u'| = |v/x_2 - r|/(v/x_1 - r)^{x_2/x_1}$. Choosing the sign of u' such that $u' \rightarrow -\infty$ at the outer boundary of \mathcal{N}_{II} ($r \rightarrow v/x_1$), we get

$$\boxed{u' = \frac{v/x_2 - r}{(v/x_1 - r)^{x_2/x_1}}}. \quad (15.61)$$

We may then consider (u', v, θ, φ) as a coordinate system on \mathcal{N}_{II} .

Remark 1: Contrary to the range of u on \mathcal{N}_I , which is $(-\infty, +\infty)$, the range of u' on \mathcal{N}_{II} is bounded from above by a function of v , which is given by Eq. (15.68) below.

Example 5: For $\alpha = 1/9$, i.e. $x_1 = 3$ and $x_2 = 6$, Eq. (15.61) reduces to

$$u' = \frac{v/6 - r}{(v/3 - r)^2}. \quad (15.62)$$

This relation can be inverted, yielding an explicit expression for $r(u', v)$:

$$r = \frac{v}{3} + \frac{1}{2u'} \left(\sqrt{1 - \frac{2}{3}u'v} - 1 \right). \quad (15.63)$$

As above, the metric components in coordinates (u', v, θ, φ) are deduced from those in coordinates (v, r, θ, φ) [Eq. (15.17)] (cf. the notebook D.6.6 for the computation):

$$\mathbf{g} = -\frac{2x_1}{(x_2 - x_1)r} \left(\frac{v}{x_1} - r \right)^{x_2/2} \mathbf{d}u' \mathbf{d}v + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2) \quad \left(r < \frac{v}{x_1} \right), \quad (15.64)$$

where $r = r(u', v)$ is defined implicitly by Eq. (15.61). Again, we get a double-null coordinate system adapted to the ingoing and outgoing radial null geodesics. Moreover, the coordinates (u', v, θ, φ) are regular since $g_{u'v}$ is finite and non-vanishing in all \mathcal{N}_{II} . In addition, the homothetic Killing horizon \mathcal{H}_2 , which is defined by $v/r = x_2$, is the hypersurface $u' = 0$ of \mathcal{N}_{II} . In $\mathcal{N}_{\text{I}} \cap \mathcal{N}_{\text{II}}$, we have $u > 0$ and $u' < 0$. The relation between the two coordinates is obtained by equating the right-hand sides of Eqs. (15.51) and (15.60); one gets

$$u' = -u^{-x_2/x_1} \quad \text{in } \mathcal{N}_{\text{I}} \cap \mathcal{N}_{\text{II}}, \quad (15.65)$$

along with the following characterization of the two homothetic Killing horizons (cf. Fig. 15.10):

$$\mathcal{H}_1 : \quad u = 0, \quad u' \rightarrow -\infty \quad (15.66a)$$

$$\mathcal{H}_2 : \quad u \rightarrow +\infty, \quad u' = 0. \quad (15.66b)$$

Example 6: For $\alpha = 1/9$ ($x_1 = 3$ and $x_2 = 6$), Eq. (15.65) reduces to

$$u' = -\frac{1}{u^2} \quad \text{in } \mathcal{N}_{\text{I}} \cap \mathcal{N}_{\text{II}}. \quad (15.67)$$

Substituting this expression for u' in Eq. (15.63), one recovers Eq. (15.54).

Let us locate the curvature singularity at the boundary of \mathcal{N}_{II} . A necessary condition for the Kretschmann scalar (15.59) to diverge is $r(u', v) \rightarrow 0$. According to Eq. (15.57), this occurs along an outgoing radial null geodesic if, and only if, $x \rightarrow x_2$ or $x \rightarrow +\infty$. Now, let us rewrite Eq. (15.61) by substituting xr for v :

$$u' = \frac{x/x_2 - 1}{(x/x_1 - 1)^{x_2/x_1} r^{x_2/x_1 - 1}}.$$

Since u' is fixed along the outgoing radial null geodesic, we recover from this expression the two limits $x \rightarrow x_2$ or $x \rightarrow +\infty$ in which $r(u', v) \rightarrow 0$ may occur. The first limit corresponds to $v \sim x_2 r$ (actually the same behavior as (15.56) in \mathcal{N}_{I}), which implies $v \rightarrow 0$. This leads to $K \sim 12\alpha^2 x_2^6/v^4$, which does diverge for $v \rightarrow 0$. For the second limit, $x \rightarrow +\infty$, we get

$$u' \sim \frac{x_1^{x_2/x_1}}{x_2(xr)^{x_2/x_1 - 1}}.$$

This implies that $xr = v$ tends to a finite value. Then $K = 12\alpha^2 v^2/r^6$ diverges as r^{-6} there. We conclude:

Property 15.16: curvature singularity in region \mathcal{N}_{II}

In \mathcal{N}_{II} , the limit $r \rightarrow 0$ always corresponds to a curvature singularity. This singularity is reached in two places, which constitute parts of the boundary of \mathcal{N}_{II} : (i) $v \rightarrow 0$ (the zigzag segment inclined at 45° in Fig. 15.10 left) and (ii) $u'v^{x_2/x_1 - 1} \rightarrow x_1^{x_2/x_1}/x_2$ (the horizontal zigzag segment in Fig. 15.10 left). In particular, the ranges of the coordinates (u', v) in \mathcal{N}_{II}

are

$$0 < v \leq v_0 \quad \text{and} \quad -\infty < u' < \frac{x_1^{x_2/x_1}}{x_2 v^{x_2/x_1-1}}. \quad (15.68)$$

Example 7: For $\alpha = 1/9$, i.e. $x_1 = 3$ and $x_2 = 6$, Eq. (15.63) implies

$$\lim_{v \rightarrow 0} r = 0 \quad \text{and} \quad \lim_{u'v \rightarrow 3/2} r = 0. \quad (15.69)$$

The second limit agrees with the upper boundary for u' in (15.68), since the latter can be rewritten as $u' < 3/(2v)$ for $x_1 = 3$ and $x_2 = 6$.

As a particular case of Eq. (15.68), the maximal value of u' along an ingoing radial null geodesic with $v = v_0$ (geodesic on the outer edge of the radiation shell, cf. Fig. 15.10) is

$$u'_0 = \frac{x_1^{x_2/x_1}}{x_2 v_0^{x_2/x_1-1}}. \quad (15.70)$$

15.4.3 Carter-Penrose diagram of the whole spacetime

A Carter-Penrose diagram of the whole Vaidya spacetime can be constructed by assembling the Carter-Penrose diagrams obtained for the radiation region \mathcal{M}_{rad} (Fig. 15.10) with Carter-Penrose diagrams of the Minkowski region ($v < 0$) and of the Schwarzschild region ($v > v_0$). To achieve this, one should locate the black hole event horizon \mathcal{H} in \mathcal{M}_{rad} . Given that $x_A = \alpha^{-1} > x_1$ (cf. Sec. 15.4.1), one has $\mathcal{H} \cap \mathcal{M}_{\text{rad}} \subset \mathcal{N}_{\text{II}}$. Since \mathcal{H} is generated by outgoing radial null geodesics, $\mathcal{H} \cap \mathcal{M}_{\text{rad}}$ lies a constant value of u' , which we shall denote by $u'_{\mathcal{H}}$. The value of $u'_{\mathcal{H}}$ is obtained by setting $v = v_0$ and $r = 2m$ in Eq. (15.61). Using the identities $\alpha = 2m/v_0$ and $\alpha = 2/(x_1 x_2)$, we get

$$u'_{\mathcal{H}} = u'_0 \left(1 - \frac{2}{x_1}\right) \left(1 - \frac{2}{x_2}\right)^{-x_2/x_1}. \quad (15.71)$$

From the expressions of x_1 and x_2 in terms of α [Eq. (15.44)], it is easy to show that $u'_{\mathcal{H}} < u'_0$. Taking this property into account, one can draw \mathcal{H} in the diagram of \mathcal{N}_{II} of Fig. 15.10 and finally obtain the Carter-Penrose diagram of \mathcal{M} shown in Fig. 15.11.

Example 8: For $v_0 = 18m$, one has $\alpha = 1/9$, $x_1 = 3$ and $x_2 = 6$, so that Eqs. (15.70) and (15.72) yield respectively $u'_0 = 3/(2v_0) = 1/(12m)$ and $u'_{\mathcal{H}} = 3u'_0/4 = 9/(8v_0) = 1/(16m)$.

The Carter-Penrose diagram of Fig. 15.11 enables us to fully solve the puzzle of multiple outgoing radial null geodesics emanating from the point $(t, r) = (0, 0)$ raised at the beginning of Sec. 15.4.2: $(t, r) = (0, 0)$, or equivalently, $(v, r) = (0, 0)$, does not define a single point, but a full segment, denoted by $O_1 O_3$ in Fig. 15.11. From each point of this segment, there emanates a single outgoing radial null geodesic (cf. the green solid lines in Fig. 15.11).

It is instructive to add the future outer trapping horizon \mathcal{T} introduced in Sec. 15.3.5 to the Carter-Penrose diagram. To do so, we need the equation of \mathcal{T} in terms of the double null coordinates. First of all, we notice that in \mathcal{M}_{rad} , \mathcal{T} is entirely contained in \mathcal{N}_{II} . Indeed, \mathcal{T}

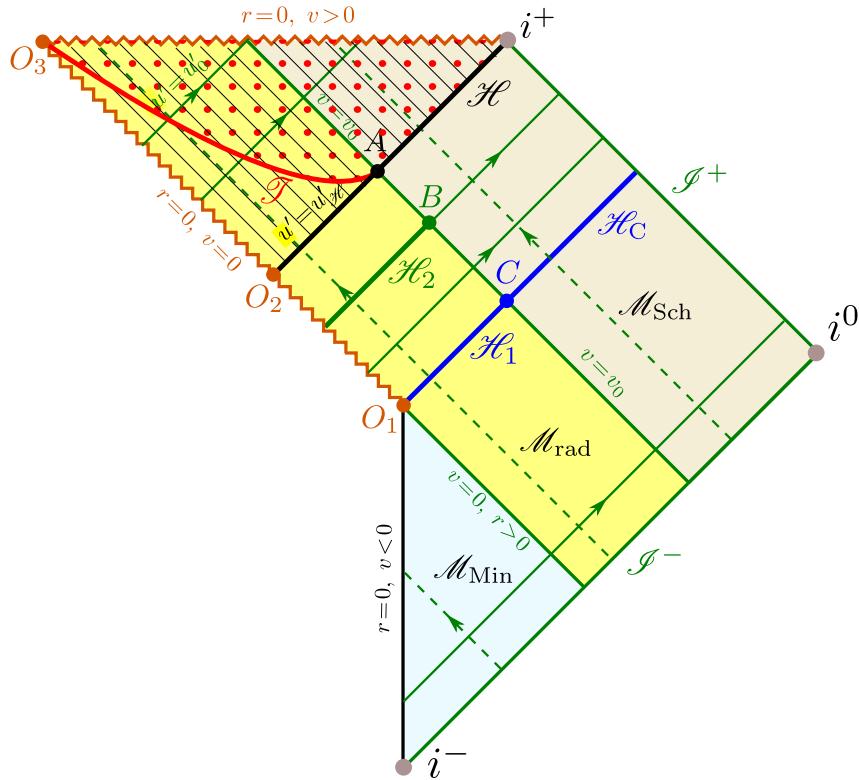


Figure 15.11: Carter-Penrose diagram of the collapsing Vaidya spacetime with $M(v) = \alpha v/2$ (homothetic radiation region) and $\alpha < 1/8$ (low energy density case). The Minkowski region \mathcal{M}_{Min} , radiation region \mathcal{M}_{rad} and Schwarzschild region \mathcal{M}_{Sch} are depicted in respectively pale blue, yellow and grey. Solid (resp. dashed) green lines are outgoing (resp. ingoing) radial null geodesics. Zigzag lines indicate curvature singularities. The hatched area is the black hole region, delimited by the event horizon \mathcal{H} . The red dot-filled area, delimited by the future outer trapping horizon \mathcal{T} (red curve), is the region where the spheres $(t, r) = \text{const}$ are trapped surfaces. The points A , B and C are the same as in Fig. 15.8, while the point O of Fig. 15.8 corresponds to the whole segment O_1O_3 .

obeying $r = \alpha v$ [Eq. (15.35)], the property $\alpha < x_2^{-1}$ excludes \mathcal{T} from \mathcal{N}_I , by the definition (15.50) of the latter. It suffices then to express \mathcal{T} in terms of the coordinates (u', v) of \mathcal{N}_{II} . This is achieved by substituting αv for r in Eq. (15.61) and making use of successively the identity $\alpha = 2/(x_1 x_2)$, Eq. (15.71) and Eq. (15.70); one gets

$$\mathcal{T} \cap \mathcal{M}_{\text{rad}} : \quad u' = u'_{\mathcal{H}} \left(\frac{v_0}{v} \right)^{x_2/x_1 - 1}. \quad (15.72)$$

Note that this implies $u' \rightarrow +\infty$ for $v \rightarrow 0$ and $u' \rightarrow u'_{\mathcal{H}}$ for $v \rightarrow v_0$ (cf. Fig. 15.10).

Example 9: For our favorite example $\alpha = 1/9$ ($x_1 = 3$ and $x_2 = 6$), the equation of $\mathcal{T} \cap \mathcal{M}_{\text{rad}}$ reduces to $u' = u'_{\mathcal{H}} v_0/v$, which corresponds to a branch of hyperbola in the (u', v) plane.

Historical note : The split of \mathcal{M}_{rad} in two regions, denoted here \mathcal{N}_I and \mathcal{N}_{II} , in order to get regular double-null coordinate systems, has been performed by B. Waugh and Kayll Lake in 1986 [515]. The double-null coordinates (u, v) and (u', v) introduced above are theirs, except for some constant factors. Four years before, William A. Hiscock, Leslie G. Williams and Douglas M. Eardley [278] exhibited a

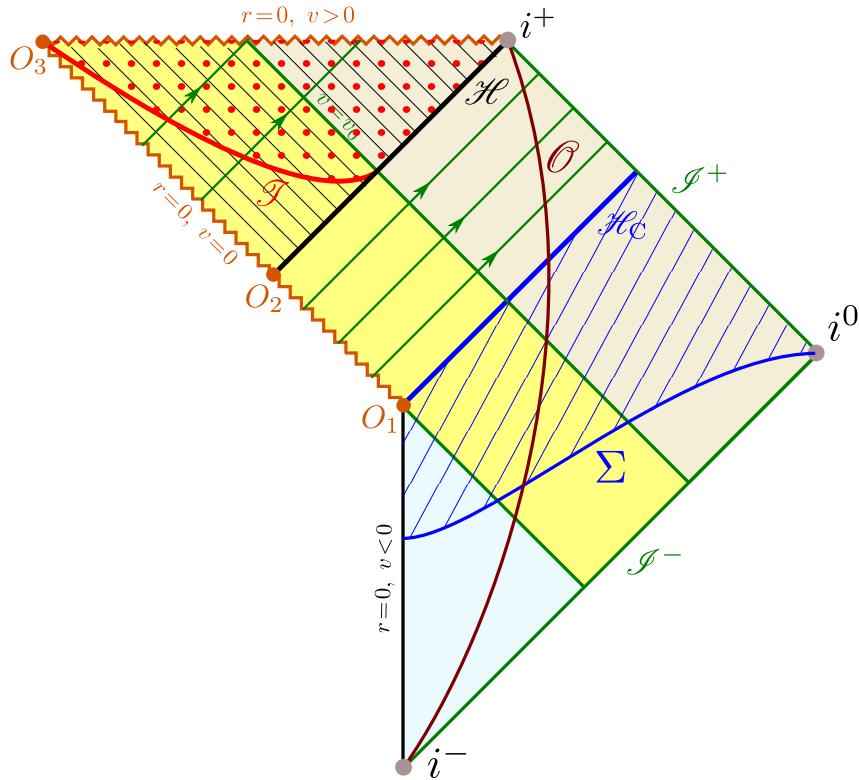


Figure 15.12: Same as Fig. 15.11, with in addition the remote observer \mathcal{O} , the partial Cauchy surface Σ and its future Cauchy development $D^+(\Sigma)$ (blue hatched area, cf. Sec. 10.8.3 for the definition), which is bounded in the future by the Cauchy horizon \mathcal{H}_C and the future null infinity \mathcal{I}^+ .

Carter-Penrose diagram similar to that of Fig. 15.11 except that the Cauchy horizon (to be discussed below) and the event horizon coincide (Fig. 2 in Ref. [278]). Such a configuration is obtained by adding a radiation surface layer (Dirac delta) at $v = v_0$. However, as discussed in Ref. [515], these authors have not set up regular double-null coordinates in \mathcal{M}_{rad} , so that their construction can be seen as heuristic. In 1984, Yuhji Kuroda [329] exhibited a Carter-Penrose diagram again similar to that of Fig. 15.7, except for the radiation region extending to \mathcal{I}^+ (no pure Schwarzschild exterior but $M'(v) \rightarrow 0$ for $v \rightarrow +\infty$) (Fig. 3 in Ref. [329]). The case of a radiation region bounded by $v = v_0$, as here, can be found in Fig. 1 of a 2007 article by Brien C. Nolan [383].

15.4.4 Naked singularity and Cauchy horizon

We recover on the Carter-Penrose diagram of Fig. 15.11 the naked singularity feature noticed in Sec. 15.4.1: light rays emitted by the curvature singularity between O_1 and O_2 reach the future null infinity \mathcal{I}^+ of the Schwarzschild exterior. This means that the singularity is visible to remote observers, such as observer \mathcal{O} in Fig. 15.12. This contrasts with the curvature singularity of the Vaidya collapse with $\alpha > 1/8$ (compare Fig. 15.7 with Fig. 15.11). The latter is indeed located at the “future boundary” of spacetime, so that no causal geodesic can originate from it.

Remark 2: The part O_2O_3 of the curvature singularity (cf. Fig. 15.11) is visible to observers that fall into the black hole, but remains hidden to remote observers. For this reason, the singularity O_2O_3 is

called **locally naked**, while O_1O_2 is called **globally naked**. For remote observers, one may say that the singularity O_2O_3 is “clothed” by the event horizon \mathcal{H} .

Remark 3: The curvature singularity of the Vaidya spacetime with $v_0 < 16m$ (Fig. 15.7) is not even locally naked: observers falling into the black hole region never see it until they hit it at the end of their lives.

The part $(v, r) = (0, 0)$ of the curvature singularity, i.e. the segment O_1O_3 in the diagram of Fig. 15.11 corresponds to a past null boundary of spacetime. Such a singularity is generically called a **shell-focusing singularity**. This term has been coined by Eardley & Smarr [175], who discovered this type of singularity in spherically symmetric inhomogeneous dust collapse. Due to its past null character, a shell-focusing singularity is always locally naked: close observers can see it.

Remote observers see the naked singularity as soon as they cross the null hypersurface \mathcal{H}_C generated by the outgoing radial null geodesics emanating from the birth of the singularity at $u = 0$ in \mathcal{N}_1 (point O_1 in Figs. 15.11 and 15.12). \mathcal{H}_C can be seen as the null hypersurface extending the homothetic Killing horizon \mathcal{H}_1 introduced in Sec. 15.4.1 beyond the radiation region, i.e. one has $\mathcal{H}_C \cap \mathcal{M}_{\text{rad}} = \mathcal{H}_1$. The hypersurface \mathcal{H}_C is actually the *future Cauchy horizon*⁶ of any partial Cauchy surface Σ that encounters the Minkowski region⁷. In other words, the region of spacetime in the future of \mathcal{H}_C cannot be entirely predicted from initial data on Σ and the Einstein equation. In particular, data from the naked curvature singularity O_1O_3 is required. Since a curvature singularity marks a validity limit of general relativity, one cannot unambiguously prescribe data there. This seriously hampers the predictability of general relativity alone for describing the region beyond \mathcal{H}_C . For instance, a burst of radiation could come out at any time from the naked singularity and change the fate of observers travelling in this region. This regards all remote observers at a sufficiently advanced amount of their proper time, since any such observer necessarily crosses \mathcal{H}_C at some point (cf. observer \mathcal{O} in Fig. 15.12, noticing that the worldlines of all remote observers terminate at the future timelike infinity i^+).

15.5 Going further

For simplicity, we have mostly restricted ourselves to $M(v)$ linear in the radiation region, except in Secs. 15.3.4 and 15.3.5. This allowed us to benefit from an extra symmetry (homothetic Killing vector) and to get exact solutions for the outgoing radial null geodesics. The obtained results are nevertheless representative of those for more general functions $M(v)$. In particular, the criterion for having a shell-focusing singularity with a generic function $M(v)$ remains $\alpha < 1/8$, provided that α is defined as

$$\alpha := \lim_{v \rightarrow 0^+} \frac{2M(v)}{v}. \quad (15.73)$$

⁶Cf. Sec. 10.8.3 for the definition.

⁷Since it has to be acausal without edge (cf. Sec. 10.8.3), Σ has to meet the spacelike infinity i^0 in the Carter-Penrose diagram of Fig. 15.12.

See Ref. [329, 193] for details. Of course, for $M(v) = mv/v_0$, the above definition of α reduces to (15.18). For $\alpha < 1/8$, a new possibility may occur: that of a shell-focusing singularity that lies entirely in the black hole region, i.e. that is naked locally but not globally, contrary to what happens for $M(v)$ linear (Sec. 15.4). See e.g. Fig. 4 of Ref. [26] or Fig. 9.20 of Ref. [242] for the corresponding Carter-Penrose diagram. Actually, the Vaidya collapse depicted in Fig. 15.5 lies in this category: it has $M(v) = m(v/v_0)^3 [6(v/v_0)^2 - 15v/v_0 + 10]$, so that Eq. (15.73) yields $\alpha = 0$. The singularity is entirely located in the black hole region and appears to be locally naked (note the outgoing radial null geodesic emerging from it near $(t, r) = (0, 0)$ in Fig. 15.5).

It is worth to stress that in the literature, many studies regard the *outgoing* Vaidya metric, starting from the original studies (cf. the historical note on p. 584), because they are motivated either by the modelling of radiating bodies or the study of black hole evaporation. Nonetheless, most results are applicable to Vaidya collapse (which is based on the *ingoing* version of the metric) by performing some time reversal. For instance, one can infer from the recent study [142] that for an imploding Vaidya spacetime without any Minkowski region (v spanning the range $(-\infty, v_0)$ in the radiation region), but such that $\lim_{v \rightarrow -\infty} M(v) = 0$, there is a curvature singularity at the null boundary $v \rightarrow -\infty$ of spacetime, whatever the choice of the non-decreasing mass function $M(v)$.

Chapter 16

Evolution and thermodynamics of black holes

Contents

16.1 Introduction	615
16.2 Towards the first law of black hole dynamics	615
16.3 Evolution of the black hole area	633
16.4 The laws of black hole dynamics	640
16.5 Hawking radiation	644
16.6 Black hole thermodynamics	648
16.7 Concluding remarks	651

16.1 Introduction

This chapter is in a draft stage.

16.2 Towards the first law of black hole dynamics

16.2.1 Mass variation formula for Kerr black holes

In this section, we assume 4-dimensional general relativity. Let us consider an initially isolated Kerr black hole of mass and spin parameters (m, a) that is perturbed by the arrival of some external body or some gravitational radiation. After some transitory dynamical regime (e.g. absorption of the incoming body and emission of gravitational waves), the black hole relaxes to a new equilibrium configuration. According to the no-hair theorem (Property 5.38), the final state has to be a Kerr black hole, of parameters $(m + \delta m, a + \delta a)$ say. All global properties of the black hole (cf. Sec. 10.6) are changed accordingly and we are going to express the change in the

Komar mass $M = m$ [Eq. (10.78)] in terms of the change in the area $A = 8\pi m(m + \sqrt{m^2 - a^2})$ [Eq. (10.83)] and in the angular momentum $J = am$ [Eq. (10.80)].

Rewriting formula (10.83) as $A = 8\pi(M^2 + \sqrt{M^4 - J^2})$ and differentiating, we get, for small variations δM and δJ of M and J ,

$$\frac{1}{8\pi} \delta A = 2M \delta M + \frac{2M^3}{\sqrt{M^4 - J^2}} \delta M - \frac{J}{\sqrt{M^4 - J^2}} \delta J,$$

or equivalently

$$\delta M = \underbrace{\frac{1}{8\pi} \frac{\sqrt{M^4 - J^2}}{2M(M^2 + \sqrt{M^4 - J^2})} \delta A}_{\kappa} + \underbrace{\frac{J}{2M(M^2 + \sqrt{M^4 - J^2})} \delta J}_{\Omega_{\mathcal{H}}},$$

where the identifications of the black hole's surface gravity κ and rotation velocity $\Omega_{\mathcal{H}}$ result from Eqs. (10.77) and (10.69) respectively. Hence we get

$$\boxed{\delta M = \frac{\kappa}{8\pi} \delta A + \Omega_{\mathcal{H}} \delta J}. \quad (16.1)$$

16.2.2 General mass variation formula

The mass variation formula (16.1) can be derived in a much more general framework, without assuming that it corresponds to changes between two nearby Kerr solutions and without restricting the spacetime dimension to 4 or assuming that the event horizon is connected. It even holds for other gravity theories than general relativity, more specifically for any theory based on a diffeomorphism-covariant Lagrangian, provided that the area A is replaced by a quantity named the *Wald entropy*, which reduces to A for general relativity, as we shall discuss in Chap. 17.

Here we establish the mass variation formula for two nearby black hole equilibrium configurations in general relativity, from integral mass formulas obtained in Chap. 5. More precisely, we consider a stationary spacetime (\mathcal{M}, g) of dimension $n \geq 4$ containing a black hole and a “nearby” black hole spacetime $(\mathcal{M}, g + \delta g)$, which has the same symmetries (stationarity and possible axisymmetries) as (\mathcal{M}, g) . We shall call $(\mathcal{M}, g + \delta g)$ the **perturbed spacetime**, although we do not require that $(\mathcal{M}, g + \delta g)$ is obtained from (\mathcal{M}, g) by some specific physical perturbation. Note that the same manifold \mathcal{M} is used for both spacetimes. There is no loss of generality in doing so, since the perturbed manifold must be diffeomorphic to the original one, which allows one to identify the two manifolds. In particular, both manifolds have the same topology: one would certainly not qualify as “nearby” a manifold with a distinct topology.

Property 16.1: mass variation formula for generic black holes

Let (\mathcal{M}, g) be a stationary spacetime of dimension $n \geq 4$ (stationary Killing vector ξ) that contains a black hole, the event horizon of which has $K \geq 1$ connected components $\mathcal{H}_1, \dots, \mathcal{H}_K$. We shall assume that each \mathcal{H}_k is a Killing horizon with respect to a Killing vector χ_k . This is guaranteed with $\chi_k = \xi$ if \mathcal{H}_k is non-rotating (ξ null on all \mathcal{H}_k ; Property 5.23),

while if \mathcal{H}_k is rotating (ξ spacelike on some parts of \mathcal{H}_k), this holds under the hypotheses of the strong rigidity theorem (Property 5.25), with

$$\chi_k = \xi + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \eta_{\mathcal{H}_k(i)}, \quad (16.2)$$

where $1 \leq L_k \leq [(n-1)/2]$, the $\Omega_{\mathcal{H}_k}^{(i)}$'s are constants and the $\eta_{\mathcal{H}_k(i)}$'s are axisymmetric Killing vectors (Property 5.26). We shall encompass both the non-rotating case and the rotating one by allowing L_k to take the value 0 in Eq. (16.2) so that $\chi_k = \xi$ is recovered if \mathcal{H}_k is non-rotating. Let $J_{\mathcal{H}_k}^{(i)}$ be the Komar angular momentum with respect to $\eta_{\mathcal{H}_k(i)}$ over any cross-section of \mathcal{H}_k [Eq. (5.63)]. Being a non-expanding horizon, each \mathcal{H}_k has a well defined area A_k (Property 3.1). Let κ_k be the surface gravity of \mathcal{H}_k , i.e. the coefficient such that $\nabla_{\chi_k} \chi_k = \kappa_k \chi_k$ on \mathcal{H}_k [cf. Eq. (3.29)]. Let us assume that the null dominance condition (3.43) is fulfilled on \mathcal{H}_k or that \mathcal{H}_k is part of a bifurcate Killing horizon; by the zeroth law of black hole dynamics (Property 3.10 or 3.16), this implies that κ_k is constant. Finally let us assume that (\mathcal{M}, g) is asymptotically flat with g having the same asymptotic behavior (5.31) as in general relativity^a. Let $(\mathcal{M}, g + \delta g)$ be a nearby stationary spacetime sharing the same characteristics as (\mathcal{M}, g) . Then, the change δM_∞ in Komar mass at infinity (cf. Property 5.16) between the two spacetimes is related to the changes δA_k in area and to the changes $\delta J_{\mathcal{H}_k}^{(i)}$ in Komar angular momentum by

$$\begin{aligned} \delta M_\infty &= \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) \\ &\quad + \frac{1}{16\pi} \int_{\Sigma} G^{\mu\nu} \delta g_{\mu\nu} \xi_\rho dV^\rho - \frac{1}{8\pi} \delta \int_{\Sigma} G_{\mu\nu} \xi^\mu dV^\nu, \end{aligned} \quad (16.3)$$

where (i) Σ is any asymptotically flat spacelike hypersurface, the inner boundary of which is an axisymmetric cross-section \mathcal{S} of the event horizon, i.e. $\mathcal{S} = \cup_{k=1}^K \mathcal{S}_k$ with $\mathcal{S}_k := \Sigma \cap \mathcal{H}_k$ and $\eta_{\mathcal{H}_k(i)}$ tangent to \mathcal{S}_k for $i \in \{1, \dots, L_k\}$, (ii) G is the Einstein tensor of g and (iii) dV is the normal volume element vector of Σ , as defined by Eq. (5.48).

^aNote that g is not required to obey the Einstein equation in \mathcal{M} .

Proof. We may consider that the black hole event horizon \mathcal{H} is the same hypersurface in both spacetimes (\mathcal{M}, g) and $(\mathcal{M}, g + \delta g)$. If the horizons would differ, we could find a diffeomorphism $\mathcal{M} \rightarrow \mathcal{M}$ that would map the original horizon to the perturbed spacetime's horizon. Similarly, there is no loss of generality in considering that the Killing vectors ξ and $\eta_{\mathcal{H}_k(i)}$ generating the stationarity and axisymmetries are identical, as vector fields on \mathcal{M} :

$$\delta \xi = 0 \quad \text{and} \quad \delta \eta_{\mathcal{H}_k(i)} = 0. \quad (16.4)$$

This amounts to identifying the orbits of the isometry group actions in the two spacetimes. Given that \mathcal{H} is globally invariant by these group actions, this requirement is compatible with the identification of \mathcal{H} in both spacetimes. Let us introduce the short-hand notation $h := \delta g$, or in index notation $h_{\alpha\beta} = \delta g_{\alpha\beta}$. In what follows, indices are raised or lowered with

the unperturbed metric \mathbf{g} . In particular, $h^{\alpha\beta} := g^{\alpha\mu}g^{\beta\nu}h_{\mu\nu}$. Note that $h^{\alpha\beta} \neq \delta g^{\alpha\beta}$. Actually, by variation of the identity $g^{\alpha\mu}g_{\mu\beta} = \delta^\alpha_\beta$, one gets $\delta g^{\alpha\beta} = -h^{\alpha\beta}$. The starting point for proving (16.3) is the variation of the generalized Smarr formula (5.84):

$$\begin{aligned} \frac{2(n-3)}{n-2} \delta M_\infty &= \sum_{k=1}^K \left[\frac{1}{4\pi} (A_k \delta \kappa_k + \kappa_k \delta A_k) + 2 \sum_{i=1}^{L_k} \left(J_{\mathcal{H}_k}^{(i)} \delta \Omega_{\mathcal{H}_k}^{(i)} + \Omega_{\mathcal{H}_k}^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) \right] \\ &\quad - \frac{1}{4\pi} \delta \int_{\Sigma} R_{\mu\nu} \xi^\nu dV^\mu, \end{aligned} \quad (16.5)$$

where Σ is any asymptotically flat spacelike hypersurface, the inner boundary of which is a cross-section of the event horizon, i.e. some union of cross-sections \mathcal{S}_k of the connected components \mathcal{H}_k and dV^μ is the normal volume element vector of Σ defined by Eq. (5.48). Let us start by evaluating the term $A_k \delta \kappa_k$ in the above formula. For the sake of brevity, we shall drop the index k in what follows, given that we temporarily focus on a single connected component \mathcal{H}_k of the event horizon \mathcal{H} . The surface gravity κ ($= \kappa_k$) is given by Eq. (3.31): $2\kappa = k^\mu \partial_\mu (\chi_\nu \chi^\nu)$, where χ ($= \chi_k$) is the Killing vector normal to the Killing horizon \mathcal{H}_k and \mathbf{k} is the future-directed null vector field defined on \mathcal{H}_k , transverse to \mathcal{H}_k , normal to the cross-section $\mathcal{S}_k = \mathcal{H}_k \cap \Sigma$ and normalized by $\chi \cdot \mathbf{k} = -1$. Note that (χ, \mathbf{k}) is a null basis of the normal plane $T_p^\perp \mathcal{S}_k$ at each point $p \in \mathcal{S}_k$, χ playing the role of the vector ℓ in Fig. 2.10. Varying the above expression of κ yields

$$\begin{aligned} 2\delta\kappa &= \delta k^\mu \partial_\mu (\chi_\nu \chi^\nu) + k^\mu \partial_\mu (\delta \chi_\nu \chi^\nu + \chi_\nu \delta \chi^\nu) \\ &= \delta k^\mu \nabla_\mu (\chi_\nu \chi^\nu) + k^\mu \nabla_\mu (\delta \chi_\nu \chi^\nu + \chi_\nu \delta \chi^\nu) \\ &= 2\delta k^\mu \chi^\nu \nabla_\mu \chi_\nu + k^\mu (\chi^\nu \nabla_\mu \delta \chi_\nu + \delta \chi_\nu \nabla_\mu \chi^\nu + \delta \chi^\nu \nabla_\mu \chi_\nu + \chi_\nu \nabla_\mu \delta \chi^\nu) \\ &= 2\delta k^\mu \chi^\nu \nabla_\mu \chi_\nu + k^\mu [\chi^\nu (\nabla_\mu \delta \chi_\nu + \nabla_\nu \delta \chi_\mu) + 2\delta \chi_\nu \nabla_\mu \chi^\nu + \delta \chi^\nu \nabla_\mu \chi_\nu + \chi_\nu \nabla_\mu \delta \chi^\nu] \\ &= 2\delta k^\mu \chi_\nu \nabla_\mu \chi^\nu + (\chi^\mu k^\nu + k^\mu \chi^\nu) \nabla_\mu \delta \chi_\nu + k^\mu (2\delta \chi_\nu \nabla_\mu \chi^\nu + \delta \chi^\nu \nabla_\mu \chi_\nu + \chi_\nu \nabla_\mu \delta \chi^\nu). \end{aligned}$$

To get the last but one line, we have used the identity $\chi^\nu \nabla_\nu \delta \chi_\mu + \delta \chi_\nu \nabla_\mu \chi^\nu = 0$, which expresses the vanishing of the Lie derivative of the 1-form $\underline{\chi}$ [Eq. (A.86) with $(k, \ell) = (0, 1)$]:

$$\mathcal{L}_\chi \delta \underline{\chi} = 0. \quad (16.6)$$

The invariance property (16.6) holds because $\underline{\chi}$ is a normal 1-form to \mathcal{H}_k (i.e. a vector \mathbf{v} is tangent to \mathcal{H}_k iff $\langle \underline{\chi}, \mathbf{v} \rangle = 0$) and since \mathcal{H}_k is the same hypersurface in the original and the perturbed spacetime, $\underline{\chi} + \delta \underline{\chi}$ is a normal 1-form of \mathcal{H}_k as well. Two normal 1-forms to a given hypersurface are necessarily collinear: there exists a scalar field λ such that $\underline{\chi} + \delta \underline{\chi} = \lambda \underline{\chi}$. Setting $\delta \lambda := \lambda - 1$, we get

$$\delta \underline{\chi} = \delta \lambda \underline{\chi}. \quad (16.7)$$

Then $\mathcal{L}_\chi \delta \underline{\chi} = (\mathcal{L}_\chi \delta \lambda) \underline{\chi} + \delta \lambda \mathcal{L}_\chi \underline{\chi}$. But $\mathcal{L}_\chi \underline{\chi} = 0$ for χ is a Killing vector of $(\mathcal{M}, \mathbf{g})$ and, thanks to Eq. (16.2), $\mathcal{L}_\chi \delta \lambda = \mathcal{L}_\xi \delta \lambda + \sum_{i=1}^L \Omega_{\eta_{(i)}}^{(i)} \delta \lambda = 0 + 0 = 0$ because ξ and $\eta_{(i)}$ are symmetry generators of both $(\mathcal{M}, \mathbf{g})$ and $(\mathcal{M}, \mathbf{g} + \delta \mathbf{g})$ [cf. Eq. (16.4)]; this establishes (16.6). On the other side, we have

$$\delta \chi^\alpha = \sum_{i=1}^L \delta \Omega_{\eta_{(i)}}^{(i)} \eta_{(i)\alpha}^\alpha \quad \text{and} \quad \delta \chi_\alpha = h_{\alpha\mu} \chi^\mu + \sum_{i=1}^L \delta \Omega_{\eta_{(i)}}^{(i)} \eta_{(i)\alpha}. \quad (16.8)$$

The first formula readily follows from Eqs. (16.2) and (16.4), while the second one follows from $\chi_\alpha = g_{\alpha\mu}\chi^\mu$, which implies $\delta\chi_\alpha = \delta g_{\alpha\mu}\chi^\mu + g_{\alpha\mu}\delta\chi^\mu$, where $\delta g_{\alpha\mu} =: h_{\alpha\mu}$. By means of Eq. (16.8), we can rewrite the last two terms in the above expression of $\delta\kappa$ as¹

$$\delta\chi^\nu\nabla_\mu\chi_\nu + \chi_\nu\nabla_\mu\delta\chi^\nu = \sum_{i=1}^L \delta\Omega^{(i)} (\eta_{(i)}^\nu\nabla_\mu\chi_\nu + \chi_\nu\nabla_\mu\eta_{(i)}^\nu) = 2 \sum_{i=1}^L \delta\Omega^{(i)} \chi_\nu\nabla_\mu\eta_{(i)}^\nu.$$

The last equality follows from $\eta_{(i)}^\nu\nabla_\mu\chi_\nu = -\eta_{(i)}^\nu\nabla_\nu\chi_\mu$ (Killing equation for χ) $= -g_{\mu\sigma}\eta_{(i)}^\nu\nabla_\nu\chi^\sigma = -g_{\mu\sigma}\chi^\nu\nabla_\nu\eta_{(i)}^\sigma$ (χ and $\eta_{(i)}$ commute, cf. Property 5.26) $= -\chi^\nu\nabla_\nu\eta_{(i)\mu} = \chi^\nu\nabla_\mu\eta_{(i)\nu}$ (Killing equation for $\eta_{(i)}$). Accordingly, the formula for $\delta\kappa$ becomes

$$\delta\kappa = \frac{1}{2}(k^\mu\chi^\nu + \chi^\mu k^\nu)\nabla_\mu\delta\chi_\nu + \sum_{i=1}^L \delta\Omega^{(i)} k^\mu\chi^\nu\nabla_\mu\eta_{(i)\nu} + \underbrace{(\delta k^\mu\chi_\nu + k^\mu\delta\chi_\nu)\nabla_\mu\chi^\nu}_{\mathcal{A}}. \quad (16.9)$$

Let us show that $\mathcal{A} = 0$. Thanks to Eq. (16.7), we have $\mathcal{A} = (\delta k^\mu\chi_\nu + \delta\lambda k^\mu\chi_\nu)\nabla_\mu\chi^\nu = (\delta k^\mu + \delta\lambda k^\mu)\nabla_\mu(\chi_\nu\chi^\nu)/2$. Now, since \mathcal{H}_k is a Killing horizon, we have $\nabla_\mu(\chi_\nu\chi^\nu) = 2\kappa\chi_\mu$ on \mathcal{H}_k [Eq. (3.31)]. This yields $\mathcal{A} = \kappa\chi_\mu(\delta k^\mu + \delta\lambda k^\mu) = \kappa(\chi_\mu\delta k^\mu - \delta\lambda)$, since $\chi_\mu k^\mu = -1$ from the definition of k . Now, the variation of $\chi_\mu k^\mu = -1$ gives $\delta\chi_\mu k^\mu + \chi_\mu \delta k^\mu = 0$, which in view of Eq. (16.7), can be rewritten as $\delta\lambda\chi_\mu k^\mu + \chi_\mu \delta k^\mu = 0$, or equivalently $-\delta\lambda + \chi_\mu \delta k^\mu = 0$, hence $\mathcal{A} = 0$. Let us now consider the first term in the right-hand side of Eq. (16.9); we may rewrite it as

$$(\chi^\mu k^\nu + k^\mu\chi^\nu)\nabla_\mu\delta\chi_\nu = (q^{\mu\nu} - g^{\mu\nu})\nabla_\mu\delta\chi_\nu,$$

where $q^{\mu\nu}$ stands for the double metric dual of the metric q induced by g on the cross-section \mathcal{S}_k of \mathcal{H}_k : $q = g + \underline{\chi} \otimes \underline{k} + \underline{k} \otimes \underline{\chi}$ (cf. Eq. (2.34) with ℓ standing for χ). Now, thanks to Eq. (16.7),

$$q^{\mu\nu}\nabla_\mu\delta\chi_\nu = q^{\mu\nu}\nabla_\mu(\delta\lambda\chi_\nu) = \nabla_\mu\delta\lambda \underbrace{q^{\mu\nu}\chi_\nu}_0 + \delta\lambda \underbrace{q^{\mu\nu}\nabla_\mu\chi_\nu}_0 = 0,$$

where $q^{\mu\nu}\chi_\nu = 0$ holds for χ is normal to \mathcal{S}_k (cf. Eq. (2.48) with ℓ standing for χ) and $q^{\mu\nu}\nabla_\mu\chi_\nu = 0$ follows from $\nabla_\mu\chi_\nu = (a_\mu\chi_\nu - a_\nu\chi_\mu)/2$ on \mathcal{H}_k [Eq. (3.34)] combined with $q^{\mu\nu}\chi_\nu = 0$. Hence

$$\begin{aligned} (\chi^\mu k^\nu + k^\mu\chi^\nu)\nabla_\mu\delta\chi_\nu &= -g^{\mu\nu}\nabla_\mu\delta\chi_\nu = -\nabla^\mu\delta\chi_\mu = -\nabla^\mu \left(h_{\mu\nu}\chi^\nu + \sum_{i=1}^L \delta\Omega^{(i)}\eta_{(i)\mu} \right) \\ &= -\chi^\nu\nabla^\mu h_{\mu\nu} - \underbrace{h_{\mu\nu}\nabla^\mu\chi^\nu}_0 - \sum_{i=1}^L \delta\Omega^{(i)} \underbrace{\nabla^\mu\eta_{(i)\mu}}_0 = -\chi^\nu\nabla^\mu h_{\mu\nu}, \end{aligned}$$

where Eq. (16.8) has been used to express $\delta\chi_\mu$ and $h_{\mu\nu}\nabla^\mu\chi^\nu = 0$ and $\nabla^\mu\eta_{(i)\mu} = 0$ follow from the Killing equation for respectively χ and $\eta_{(i)}$, given the symmetry of h . Together with $\mathcal{A} = 0$, this result allows us rewrite Eq. (16.9) as

$$\delta\kappa = -\frac{1}{2}\chi_\mu\nabla^\nu h^\mu_\nu + \sum_{i=1}^L \delta\Omega^{(i)} k^\mu\chi^\nu\nabla_\mu\eta_{(i)\nu}. \quad (16.10)$$

¹Note that $\nabla_\mu\delta\Omega^{(i)} = 0$ since both $\Omega^{(i)}$ and $\Omega^{(i)} + \delta\Omega^{(i)}$ are constant.

Let us integrate this relation over the cross-section \mathcal{S}_k , setting $dS := \sqrt{q} d^{n-2}x$ (the area element of \mathcal{S}_k). Since $\delta\kappa$ and $\delta\Omega^{(i)}$ are constant, we get

$$\underbrace{\delta\kappa \int_{\mathcal{S}_k} dS}_{A_k} = -\frac{1}{2} \underbrace{\int_{\mathcal{S}_k} \chi_\mu \nabla^\nu h^\mu_\nu dS}_{I_{\mathcal{S}_k}} + \sum_{i=1}^L \delta\Omega^{(i)} \underbrace{\int_{\mathcal{S}_k} \nabla_\mu \eta_{(i)\nu} k^\mu \chi^\nu dS}_{-8\pi J_{\mathcal{H}_k}^{(i)}}. \quad (16.11)$$

The identification of the last integral with $-8\pi J_{\mathcal{H}_k}^{(i)}$ readily follows from formula (5.65) for the Komar angular momentum, once the area element normal bivector to \mathcal{S}_k is expressed as $dS^{\mu\nu} = (\chi^\mu k^\nu - k^\mu \chi^\nu) dS$ [Eq. (5.86)]. As for the integral denoted $I_{\mathcal{S}_k}$ in Eq. (16.11), it can be rewritten as an integral involving \mathbf{h} and the stationary Killing vector ξ instead of the Killing vector χ (which depends on \mathcal{H}_k , contrary to ξ), namely

$$I_{\mathcal{S}_k} = \int_{\mathcal{S}_k} \omega_{\mu\nu} dS^{\mu\nu}, \quad \text{with} \quad \omega_{\alpha\beta} := \frac{1}{2} (\xi \wedge \underline{\mathbf{H}})_{\alpha\beta} = \frac{1}{2} (\xi_\alpha H_\beta - H_\alpha \xi_\beta), \quad (16.12)$$

where \mathbf{H} is the vector field defined by

$$H^\alpha := \nabla^\mu h^\alpha_\mu - \nabla^\alpha h^\mu_\mu = g^{\alpha\rho} g^{\mu\nu} (\nabla_\nu \delta g_{\rho\mu} - \nabla_\rho \delta g_{\mu\nu}). \quad (16.13)$$

To prove (16.12), let us evaluate $\omega_{\mu\nu} dS^{\mu\nu}$, using successively Eq. (5.86), $\xi_\mu \chi^\mu = 0$ (ξ tangent to \mathcal{H}_k) and $\eta_{(i)\mu} k^\mu = 0$ ($\eta_{(i)}$ tangent to \mathcal{S}_k):

$$\begin{aligned} \omega_{\mu\nu} dS^{\mu\nu} &= \frac{1}{2} (\xi_\mu H_\nu - H_\mu \xi_\nu) (\chi^\mu k^\nu - k^\mu \chi^\nu) dS = (\underbrace{\xi_\mu \chi^\mu}_{0} H_\nu k^\nu - \xi_\mu k^\mu H_\nu \chi^\nu) dS \\ &= -\left(\underbrace{\chi_\mu k^\mu}_{-1} - \sum_{i=1}^L \Omega^{(i)} \underbrace{\eta_{(i)\mu} k^\mu}_{0}\right) H_\nu \chi^\nu dS = \chi_\mu H^\mu dS \\ &= (\chi_\mu \nabla^\nu h^\mu_\nu - \chi^\mu \nabla_\mu h^\nu_\nu) dS \\ &= (\chi_\mu \nabla^\nu h^\mu_\nu - \underbrace{\mathcal{L}_\xi h^\nu_\nu}_{0} - \sum_{i=1}^L \Omega^{(i)} \underbrace{\mathcal{L}_{\eta_{(i)}} h^\nu_\nu}_{0}) dS = \chi_\mu \nabla^\nu h^\mu_\nu dS. \end{aligned}$$

In the last line, $\mathcal{L}_\xi h^\nu_\nu = 0$ follows from $h^\nu_\nu = g^{\mu\nu} h_{\mu\nu}$, $\mathcal{L}_\xi g^{\mu\nu} = 0$ (ξ Killing vector of \mathbf{g}) and $\mathcal{L}_\xi h_{\mu\nu} = 0$, the latter being a consequence of ξ being a Killing vector of both \mathbf{g} and $\mathbf{g} + \delta\mathbf{g} = \mathbf{g} + \mathbf{h}$ (cf. Eq. (16.4)), so that $\mathcal{L}_\xi \mathbf{h} = \mathcal{L}_\xi (\mathbf{g} + \delta\mathbf{g}) - \mathcal{L}_\xi \mathbf{g} = 0 - 0 = 0$. Similarly, $\mathcal{L}_{\eta_{(i)}} h^\nu_\nu = 0$. In view of the above result and the definition of $I_{\mathcal{S}_k}$ in Eq. (16.11), we conclude that Eq. (16.12) holds. Therefore, Eq. (16.11) can be rewritten as

$$A_k \delta\kappa_k = -\frac{1}{2} \int_{\mathcal{S}_k} \omega_{\mu\nu} dS^{\mu\nu} - 8\pi \sum_{i=1}^L J_{\mathcal{H}_k}^{(i)} \delta\Omega_{\mathcal{H}_k}^{(i)}, \quad (16.14)$$

where we have restored the index k on κ and the subscript \mathcal{H}_k on $\Omega^{(i)}$. If we plug this relation

into Eq. (16.5), the terms $J_{\mathcal{H}_k}^{(i)} \delta \Omega_{\mathcal{H}_k}^{(i)}$ cancel out and we are left with

$$\begin{aligned} \frac{(n-3)}{n-2} \delta M_\infty &= \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) - \frac{1}{16\pi} \sum_{k=1}^K \int_{\mathcal{S}_k} \omega_{\mu\nu} dS^{\mu\nu} \\ &\quad - \frac{1}{8\pi} \delta \int_{\Sigma} R_{\mu\nu} \xi^\nu dV^\mu. \end{aligned}$$

We note that the inner boundary of the hypersurface Σ is $\mathcal{S}_{\text{int}} = \mathcal{S} = \cup_{k=1}^K \mathcal{S}_k$, so that the sum of the integrals over the \mathcal{S}_k 's is actually an integral over \mathcal{S}_{int} and we may apply formula (5.49) regarding the flux of a 2-form (here ω) to write

$$\begin{aligned} \frac{(n-3)}{n-2} \delta M_\infty &= \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) - \frac{1}{16\pi} \int_{\mathcal{S}_\infty} \omega_{\mu\nu} dS^{\mu\nu} \\ &\quad + \frac{1}{8\pi} \int_{\Sigma} \nabla^\nu \omega_{\mu\nu} dV^\mu - \frac{1}{8\pi} \delta \int_{\Sigma} R_{\mu\nu} \xi^\nu dV^\mu, \end{aligned} \quad (16.15)$$

where \mathcal{S}_∞ stands for the outer boundary of Σ , i.e. the limit $r \rightarrow +\infty$ of a sphere \mathcal{S} of constant value of (x^0, r) , (x^α) being an asymptotically Minkowskian coordinate system such that Σ is a hypersurface $x^0 = \text{const}$, $\xi = \partial_0$ and $r := \sqrt{(x^1)^2 + \dots + (x^{n-1})^2}$. In view of Eqs. (16.12)-(16.13) and (5.42), we may write

$$\int_{\mathcal{S}_\infty} \omega_{\mu\nu} dS^{\mu\nu} = \int_{\mathcal{S}_\infty} \xi_\mu H_\nu dS^{\mu\nu} = \int_{\mathcal{S}_\infty} \xi_\mu (\nabla^\sigma h_{\nu\sigma} - \nabla_\nu h^\sigma_\sigma) (s^\mu n^\nu - n^\mu s^\nu) \sqrt{q} d^{n-2}y,$$

where $(y_a)_{1 \leq a \leq n-2}$ is a coordinate system of \mathcal{S}_∞ and q is the determinant with respect to it of the metric \mathbf{g} induced by \mathbf{g} on \mathcal{S}_∞ . Now, on \mathcal{S}_∞ , $\mathbf{n} = \xi$ and $\mathbf{s} = (x^i/r)\partial_i$, with $\xi_\mu n^\mu = \xi_\mu \xi^\mu = -1$ and $\xi_\mu s^\mu = 0$. Moreover, since (x^α) is asymptotically Minkowskian, we may substitute the covariant derivatives with partial ones. Hence we get, accounting for $\partial_0 h_{i0} = 0$ ($\mathbf{h} = \delta \mathbf{g}$ is stationary, since both \mathbf{g} and $\mathbf{g} + \delta \mathbf{g}$ are),

$$\int_{\mathcal{S}_\infty} \omega_{\mu\nu} dS^{\mu\nu} = \int_{\mathcal{S}_\infty} \frac{x^i}{r} (\partial_j h_{ij} - \partial_i h^\sigma_\sigma) \sqrt{q} d^{n-2}y.$$

By hypothesis, the metric tensor has the asymptotic behavior (5.31); it follows that, for $r \rightarrow +\infty$,

$$h_{00} = \alpha_n \frac{\delta M_\infty}{r^{n-3}} + \mathcal{O}\left(\frac{1}{r^{n-2}}\right) \quad \text{and} \quad h_{ij} = \frac{\alpha_n}{n-3} \frac{\delta M_\infty}{r^{n-3}} \delta_{ij} + \mathcal{O}\left(\frac{1}{r^{n-2}}\right),$$

where α_n is given by Eq. (5.32). According to Property 5.11, δM_∞ in the above formulas is the same variation of the Komar mass at infinity as in the left-hand side of Eq. (16.15). The trace of \mathbf{h} takes the asymptotic value $h^\sigma_\sigma = -h_{00} + \sum_{i=1}^{n-1} h_{ii} = 2\alpha_n \delta M_\infty / ((n-3)r^{n-3}) + \mathcal{O}(1/r^{n-2})$. Hence we get

$$\int_{\mathcal{S}_\infty} \omega_{\mu\nu} dS^{\mu\nu} = -\frac{\alpha_n \delta M_\infty}{n-3} \underbrace{\int_{\mathcal{S}_\infty} \frac{x^i}{r} \frac{\partial}{\partial x^i} \left(\frac{1}{r^{n-3}} \right) \underbrace{\sqrt{q}}_{r^{n-2} \sqrt{\bar{q}}} d^{n-2}y}_{-(n-3)/r^{n-2}} = \alpha_n \delta M_\infty \underbrace{\int_{\mathcal{S}_\infty} \sqrt{\bar{q}} d^{n-2}y}_{\Omega_{n-2}},$$

where \bar{q} stands for the determinant of the round metric of the unit sphere \mathbb{S}^{n-2} with respect to the coordinates (y^a) . Since $\alpha_n \Omega_{n-2} = 16\pi/(n-2)$ [cf. Eq. (5.32)], there comes

$$\int_{\mathcal{S}_\infty} \omega_{\mu\nu} dS^{\mu\nu} = \frac{16\pi}{n-2} \delta M_\infty.$$

Accordingly, Eq. (16.15) simplifies to

$$\delta M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) + \frac{1}{8\pi} \int_{\Sigma} \nabla^\nu \omega_{\mu\nu} dV^\mu - \frac{1}{8\pi} \delta \int_{\Sigma} R_{\mu\nu} \xi^\nu dV^\mu.$$

Now, from Eq. (16.12),

$$\begin{aligned} 2\nabla^\nu \omega_{\mu\nu} dV^\mu &= \nabla_\nu (\xi^\mu H^\nu - H^\mu \xi^\nu) dV_\mu = (\xi^\mu \nabla_\nu H^\nu + \underbrace{H^\nu \nabla_\nu \xi^\mu - \xi^\nu \nabla_\nu H^\mu}_{0} - H^\mu \underbrace{\nabla_\nu \xi^\nu}_{0}) dV_\mu \\ &= \nabla_\nu H^\nu \xi^\mu dV_\mu, \end{aligned}$$

where $\nabla_\nu \xi^\nu = 0$ holds because ξ is a Killing vector and we have used $H^\nu \nabla_\nu \xi^\mu - \xi^\nu \nabla_\nu H^\mu = [\mathbf{H}, \xi]^\mu = -[\xi, \mathbf{H}]^\mu = -\mathcal{L}_\xi H^\mu = 0$ because $\mathcal{L}_\xi h = 0$ and Eq. (16.13) implies $\mathcal{L}_\xi \mathbf{H} = 0$, for ∇ and \mathcal{L}_ξ commute, ξ being a Killing vector². Hence

$$\delta M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) + \frac{1}{16\pi} \int_{\Sigma} \nabla_\nu H^\nu \xi^\mu dV_\mu - \frac{1}{8\pi} \delta \int_{\Sigma} R^\mu_\nu \xi^\nu dV_\mu.$$

Let us make the Einstein tensor \mathbf{G} appear in the last integral via Eq. (A.111): $R^\mu_\nu \xi^\nu dV_\mu = G^\mu_\nu \xi^\nu dV_\mu + R \xi^\mu dV_\mu / 2$. Given that $\delta \xi^\mu = 0$ [Eq. (16.4)], we have $\delta(R \xi^\mu dV_\mu) = \xi^\mu (\delta R dV_\mu + R \delta(dV_\mu))$, so that we may write

$$\delta M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) + \frac{1}{8\pi} I - \frac{1}{8\pi} \delta \int_{\Sigma} G^\mu_\nu \xi^\nu dV_\mu, \quad (16.16)$$

with

$$I := \frac{1}{2} \int_{\Sigma} \xi^\mu [\nabla_\nu H^\nu dV_\mu - (\delta R dV_\mu + R \delta(dV_\mu))].$$

Now from Eq. (5.48), $dV_\alpha = -n_\alpha \sqrt{\gamma} d^{n-1}x$, where $(x^i)_{1 \leq i \leq n-1}$ are coordinates on Σ , γ is the determinant with respect to (x^i) of the metric γ induced by \mathbf{g} on Σ and the normal 1-form n_α is collinear to the differential of t , the latter being a scalar field defining Σ as a hypersurface $t = \text{const}$: $n_\alpha = -N(\mathbf{d}t)_\alpha = -N\partial_\alpha t$. Given the identity $\sqrt{-g} = N\sqrt{\gamma}$ (see e.g. Eq. (5.55) of Ref. [226]), we have $dV_\alpha = \partial_\alpha t \sqrt{-g} d^{n-1}x$. Since obviously $\delta(\partial_\alpha t) = 0$ and $\delta(d^{n-1}x) = 0$, we get

$$\delta(dV_\alpha) = \partial_\alpha t \delta \sqrt{-g} d^{n-1}x = \frac{1}{\sqrt{-g}} \delta \sqrt{-g} dV_\alpha. \quad (16.17)$$

²This is easy to see in a coordinate system (x^α) such that $\xi = \partial_0$, cf. Eq. (A.85); see also Appendix E2 of Ref. [423] for a covariant proof.

Hence

$$I = \frac{1}{2} \int_{\Sigma} \left[\nabla_{\nu} H^{\nu} - \frac{1}{\sqrt{-g}} \delta(R\sqrt{-g}) \right] \xi^{\mu} dV_{\mu}.$$

Now, a standard identity³, which is at the root of the derivation of the Einstein equation by extremalizing the Einstein-Hilbert action, is

$$\delta(R\sqrt{-g}) = (G_{\mu\nu} \delta g^{\mu\nu} + \nabla_{\mu} H^{\mu}) \sqrt{-g}. \quad (16.18)$$

It's rather straightforward to establish (16.18) from the **Palatini identity**⁴:

$$\delta R_{\alpha\beta} = \nabla_{\mu} \delta \Gamma^{\mu}_{\alpha\beta} - \nabla_{\beta} \delta \Gamma^{\mu}_{\alpha\mu}, \quad (16.19)$$

where the variations $\delta \Gamma^{\gamma}_{\alpha\beta}$ of the Christoffel symbols are actually tensor fields on \mathcal{M} (contrary to the Christoffel symbols themselves), being expressible via the manifestly tensorial relation⁵

$$\delta \Gamma^{\gamma}_{\alpha\beta} = \frac{1}{2} g^{\gamma\mu} (\nabla_{\alpha} h_{\mu\beta} + \nabla_{\beta} h_{\alpha\mu} - \nabla_{\mu} h_{\alpha\beta}). \quad (16.20)$$

Using $R := g^{\mu\nu} R_{\mu\nu}$, we have indeed

$$\delta(R\sqrt{-g}) = \delta g^{\mu\nu} R_{\mu\nu} \sqrt{-g} + g^{\mu\nu} \delta R_{\mu\nu} \sqrt{-g} + g^{\mu\nu} R_{\mu\nu} \delta \sqrt{-g},$$

where the second term can be expressed by combining Eqs. (16.19), (16.20) and (16.13):

$$g^{\mu\nu} \delta R_{\mu\nu} = \nabla_{\mu} (\nabla^{\nu} h^{\mu}_{\nu} - \nabla^{\mu} h^{\nu}_{\nu}) = \nabla_{\mu} H^{\mu}.$$

On the other side, formula (A.73) for the variation of the determinant g leads to

$$\delta \sqrt{-g} = \frac{1}{2} \sqrt{-g} \delta \ln |g| = \frac{1}{2} \sqrt{-g} g^{\mu\nu} \delta g_{\mu\nu} = -\frac{1}{2} \sqrt{-g} g_{\mu\nu} \delta g^{\mu\nu}. \quad (16.21)$$

The above three equations establish (16.18). It immediately follows that

$$I = -\frac{1}{2} \int_{\Sigma} G_{\mu\nu} \delta g^{\mu\nu} \xi^{\rho} dV_{\rho} = \frac{1}{2} \int_{\Sigma} G^{\mu\nu} \delta g_{\mu\nu} \xi_{\rho} dV^{\rho},$$

where the second equality results from $\delta g^{\mu\nu} = -g^{\mu\rho} g^{\nu\sigma} \delta g_{\rho\sigma}$. In view of Eq. (16.16), this completes the proof of the mass variation formula (16.3). \square

Remark 1: Contrary to the generalized Smarr formula (5.84), the spacetime dimension n does not appear in the mass variation formula (16.3).

Let us apply the general mass variation formula (16.3) to a solution of the vacuum Einstein equation (1.44). In this case, the Ricci tensor vanishes identically, and so does the Einstein tensor: $G = 0$. Accordingly one gets rid of the integrals in the right-hand side of formula (16.3). Moreover, the null dominance condition is automatically fulfilled by $G = 0$, which guarantees that the surface gravities κ_k are constant on each connected component of the event horizon. We may therefore state:

³See e.g. Eq. (E.1.18) in Wald's textbook [498], where $v = H$ or Eqs. (4.18)-(4.19) and the unnumbered equation below (4.20) on p. 437-438 of Deruelle & Uzan's textbook [162], where $V = H$.

⁴See e.g. Eq. (21.21) in MTW [370], Eq. (3.86) in Straumann's textbook [463] or Eq. (4.62) in Carroll's textbook [86].

⁵See e.g. Eq. (3.93) in Straumann's textbook [463].

Property 16.2: mass variation formula for black holes in vacuum general relativity

Let $(\mathcal{M}, \mathbf{g})$ be a stationary spacetime of dimension $n \geq 4$ such that \mathbf{g} obeys the vacuum Einstein equation $\mathbf{R} = 0$ [Eq. (1.44)]. Let us suppose that $(\mathcal{M}, \mathbf{g})$ contains a black hole, the event horizon of which has $K \geq 1$ connected components $\mathcal{H}_1, \dots, \mathcal{H}_K$, each of them being assumed to be a Killing horizon. Let $(\mathcal{M}, \mathbf{g} + \delta\mathbf{g})$ be a nearby stationary spacetime sharing the same characteristics. Using the same notations as in Property 16.1, the change δM in Komar mass between the two spacetimes is given by^a

$$\boxed{\delta M = \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^{L_k} \Omega_{\mathcal{H}_k}^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right)}. \quad (16.22)$$

In the case of a connected event horizon ($K = 1$) with at most one axisymmetry ($0 \leq L_1 \leq 1$, which is the only possibility if $n = 4$), this formula reduces to

$$\boxed{\delta M = \frac{\kappa}{8\pi} \delta A + \Omega_{\mathcal{H}} \delta J}. \quad (16.23)$$

In particular, one recovers formula (16.1), which was obtained for a Kerr black hole.

^aWe have dropped the subscript ∞ on δM because, in the vacuum case, the Komar mass is independent of the integration surface (Property 5.15) and hence needs not to be taken at the asymptotically flat end of the stationary spacetime.

16.2.3 Mass variation formula for charged black holes

For electrically charged black holes in general relativity (electrovacuum spacetimes, cf. Secs. 1.5.2 and 5.5.2), the Einstein tensor \mathbf{G} is proportional to the energy-momentum tensor \mathbf{T} of the electromagnetic field \mathbf{F} , as given by Eq. (1.52), and the mass variation formula (16.3) can be transformed into a simple formula:

Property 16.3: mass variation formula for charged black holes in general relativity

Let $(\mathcal{M}, \mathbf{g})$ be a stationary spacetime of dimension $n \geq 4$, with stationary Killing vector ξ , endowed with a source-free electromagnetic field \mathbf{F} such that (\mathbf{g}, \mathbf{F}) obeys the electrovacuum Einstein equation (1.54). Let us assume that $(\mathcal{M}, \mathbf{g})$ contains a black hole, the event horizon of which has $K \geq 1$ connected components $\mathcal{H}_1, \dots, \mathcal{H}_K$. Each \mathcal{H}_k ($1 \leq k \leq K$) is assumed to be a Killing horizon with respect to the Killing vector $\chi = \xi + \sum_{i=1}^L \Omega^{(i)} \eta_{(i)}$, where $0 \leq L \leq [(n-1)/2]$, the $\Omega^{(i)}$ are constants and the $\eta_{(i)}$ are axisymmetric Killing vectors^a. One may have $L = 0$, i.e. $\chi = \xi$ (static configuration). Let $(\mathcal{M}, \mathbf{g} + \delta\mathbf{g})$ be a nearby stationary spacetime sharing the same characteristics. The change δM_∞ in Komar mass at infinity (cf. Property 5.16) between the two spacetimes is related to the changes δA_k in area, to the changes $\delta J_\infty^{(i)}$ in Komar angular momentum at infinity (cf. Property 5.22)

and to the changes $\delta Q_{\mathcal{H}_k}$ in electric charge of each Killing horizon (cf. Property 5.30) by

$$\delta M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \Phi_{\mathcal{H}_k} \delta Q_{\mathcal{H}_k} \right) + \sum_{i=1}^L \Omega^{(i)} \delta J_\infty^{(i)}, \quad (16.24)$$

where κ_k is the surface gravity of \mathcal{H}_k and $\Phi_{\mathcal{H}_k}$ is the electric potential of \mathcal{H}_k (cf. Property 5.29).

For a connected event horizon ($K = 1$) and a single axisymmetry ($L = 1$, which is required if $n = 4$), formula (16.24) reduces to

$$\delta M_\infty = \frac{\kappa}{8\pi} \delta A + \Omega_{\mathcal{H}} \delta J_\infty + \Phi_{\mathcal{H}} \delta Q_{\mathcal{H}}, \quad (16.25)$$

where we have dropped the indices k and i on the various quantities and have denoted $\Omega_{\mathcal{H}}$ the angular velocity of \mathcal{H} with respect to the single axisymmetry.

^aNote that contrary to the more general setting of Property 16.1, all the connected components \mathcal{H} are assumed to be Killing horizons with respect to the same Killing vector χ ; in other words, the rotation velocities $\Omega^{(i)}$ and axisymmetric Killing vectors $\eta_{(i)}$ are the same for all the \mathcal{H}_k 's.

Proof. First of all, we note that the null dominant energy condition (3.46) is fulfilled by the electromagnetic field energy-momentum tensor \mathbf{T} [Eq. (1.52)] [320]. By virtue of the zeroth law (Property 3.10), this guarantees that the surface gravities κ_k are constant over each Killing horizon \mathcal{H}_k . All the hypotheses of Property 16.1 are thus fulfilled. We shall use the same setup as in the proof of this property; in particular Eq. (16.4) holds: $\delta\xi = 0$ and $\delta\eta_{(i)} = 0$. By virtue of Einstein's equation (1.41) (with $\Lambda = 0$), we may substitute \mathbf{G} in the last integral of the general formula (16.3) by $8\pi\mathbf{T}$, where \mathbf{T} is given by Eq. (1.52); we get

$$\delta M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^{L_k} \Omega^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) + \frac{1}{2} \int_{\Sigma} T^{\mu\nu} h_{\mu\nu} \xi^\rho dV_\rho - \delta I, \quad (16.26)$$

with $h_{\mu\nu} := \delta g_{\mu\nu}$ and

$$I := \frac{1}{\mu_0} \int_{\Sigma} \left(F_{\rho\mu} F_{\nu}^{\rho} \xi^{\nu} - \frac{1}{4} F_{\rho\sigma} F^{\rho\sigma} \xi_{\mu} \right) dV^{\mu}.$$

Let us introduce an electromagnetic potential \mathbf{A} ($\mathbf{F} = \mathbf{d}\mathbf{A}$) obeying all the spacetime symmetries; in particular, $\mathcal{L}_{\xi} \mathbf{A} = 0$. Thanks to the Cartan identity (A.95), this is equivalent to $\xi \cdot \mathbf{F} + \mathbf{d}\langle \mathbf{A}, \xi \rangle = 0$ or $F_{\alpha\nu} \xi^{\nu} = \nabla_{\alpha} (A_{\nu} \xi^{\nu})$. It follows then that $F_{\rho\mu} F_{\nu}^{\rho} \xi^{\nu} = F_{\rho\mu} \nabla^{\rho} (A_{\nu} \xi^{\nu}) = \nabla^{\rho} (A_{\nu} \xi^{\nu} F_{\rho\mu})$, where the last equality stems from the source-free Maxwell equation $\nabla^{\rho} F_{\rho\mu} = 0$ [Eq. (1.48) with $j^{\alpha} = 0$]. We may then use Lemma 5.12 to turn the first term in I into a surface integral:

$$I = \frac{1}{2\mu_0} \int_{\mathcal{S}} A_{\sigma} \xi^{\sigma} F_{\mu\nu} dS^{\mu\nu} - \frac{1}{4\mu_0} \int_{\Sigma} F_{\mu\nu} F^{\mu\nu} \xi^{\rho} dV_{\rho}, \quad (16.27)$$

where $\mathcal{S} := \cup_{k=1}^K \mathcal{S}_k$ is the inner boundary of Σ ($\mathcal{S}_k := \mathcal{H}_k \cup \Sigma$) and we have set the surface integral at infinity to zero, thanks to the decay of \mathbf{F} and \mathbf{A} , at least as $1/r^{n-2}$ and $1/r^{n-3}$

respectively. Let us evaluate the variation of the integral over Σ . We have, using $\delta\xi^\rho = 0$ [Eq. (16.4)],

$$\delta(F_{\mu\nu}F^{\mu\nu}\xi^\rho dV_\rho) = \delta(F_{\mu\nu}F^{\mu\nu})\xi^\rho dV_\rho + F_{\mu\nu}F^{\mu\nu}\xi^\rho\delta(dV_\rho).$$

Now, $\delta(dV_\rho) = (1/2)g^{\mu\nu}h_{\mu\nu}dV_\rho$ [cf. Eqs. (16.17) and (16.21)] and

$$\begin{aligned} \delta(F_{\mu\nu}F^{\mu\nu}) &= \delta F_{\mu\nu}F^{\mu\nu} + F_{\mu\nu}\delta(g^{\mu\rho}g^{\nu\sigma}F_{\rho\sigma}) \\ &= \delta F_{\mu\nu}F^{\mu\nu} + F_{\mu\nu}\underbrace{\delta g^{\mu\rho}g^{\nu\sigma}}_{-h^{\mu\rho}}F_{\rho\sigma} + F_{\mu\nu}g^{\mu\rho}\underbrace{\delta g^{\nu\sigma}}_{-h^{\nu\sigma}}F_{\rho\sigma} + \underbrace{F_{\mu\nu}g^{\mu\rho}g^{\nu\sigma}}_{F^{\rho\sigma}}\delta F_{\rho\sigma} \\ &= 2F^{\mu\nu}\delta F_{\mu\nu} - 2F_{\rho\mu}F_{\nu}^{\rho}h^{\mu\nu} \\ &= 2F^{\mu\nu}\delta(\partial_\mu A_\nu - \partial_\nu A_\mu) - 2F_{\rho}^{\mu}F^{\rho\nu}h_{\mu\nu} \\ &= 2F^{\mu\nu}(\partial_\mu\delta A_\nu - \partial_\nu\delta A_\mu) - 2F_{\sigma}^{\mu}F^{\sigma\nu}h_{\mu\nu} \\ &= 2F^{\mu\nu}(\nabla_\mu\delta A_\nu - \nabla_\nu\delta A_\mu) - 2F_{\sigma}^{\mu}F^{\sigma\nu}h_{\mu\nu} \\ &= 4F^{\mu\nu}\nabla_\mu\delta A_\nu - 2F_{\sigma}^{\mu}F^{\sigma\nu}h_{\mu\nu}. \end{aligned}$$

Hence

$$\begin{aligned} \delta(F_{\mu\nu}F^{\mu\nu}\xi^\rho dV_\rho) &= \left(4F^{\mu\nu}\nabla_\mu\delta A_\nu - 2F_{\sigma}^{\mu}F^{\sigma\nu}h_{\mu\nu} + \frac{1}{2}F_{\sigma\tau}F^{\sigma\tau}g^{\mu\nu}h_{\mu\nu}\right)\xi^\rho dV_\rho \\ &= (4F^{\mu\nu}\nabla_\mu\delta A_\nu - 2\mu_0 T^{\mu\nu}h_{\mu\nu})\xi^\rho dV_\rho \\ &= 4\nabla_\nu\Omega^{\mu\nu}dV_\mu - 2\mu_0 T^{\mu\nu}h_{\mu\nu}\xi^\rho dV_\rho, \end{aligned} \quad (16.28)$$

where

$$\Omega^{\mu\nu} := \delta A_\sigma(F^{\sigma\mu}\xi^\nu - F^{\sigma\nu}\xi^\mu). \quad (16.29)$$

To prove (16.28), we need to show that $F^{\nu\sigma}\nabla_\nu\delta A_\sigma\xi^\mu = \nabla_\nu\Omega^{\mu\nu}$. This is easily done by a direct computation of $\nabla_\nu\Omega^{\mu\nu}$ from (16.29), using $\nabla_\nu\xi^\nu = 0$ (Killing equation) and $\nabla_\nu F^{\sigma\nu} = 0$ (source-free Maxwell equation):

$$\nabla_\nu\Omega^{\mu\nu} = \nabla_\nu\delta A_\sigma F^{\sigma\mu}\xi^\nu - \nabla_\nu\delta A_\sigma\underbrace{F^{\sigma\nu}\xi^\mu}_{-F^{\nu\sigma}} + \delta A_\sigma(\nabla_\nu F^{\sigma\mu}\xi^\nu + F^{\sigma\mu}\underbrace{\nabla_\nu\xi^\nu}_0 - \underbrace{\nabla_\nu F^{\sigma\nu}\xi^\mu}_0 - F^{\sigma\nu}\nabla_\nu\xi^\mu).$$

Expressing the symmetry properties $\mathcal{L}_\xi\vec{F} = 0$ and $\mathcal{L}_\xi\delta A = 0$ as respectively $\xi^\nu\nabla_\nu F^{\sigma\mu} - F^{\sigma\nu}\nabla_\nu\xi^\mu = F^{\nu\mu}\nabla_\nu\xi^\sigma$ and $\xi^\nu\nabla_\nu\delta A_\sigma = -\delta A_\nu\nabla_\sigma\xi^\nu$ [cf. Eq. (A.86)], we get

$$\nabla_\nu\Omega^{\mu\nu} = -F^{\sigma\mu}\delta A_\nu\nabla_\sigma\xi^\nu + F^{\nu\sigma}\nabla_\nu\delta A_\sigma\xi^\mu + \delta A_\sigma F^{\nu\mu}\nabla_\nu\xi^\sigma = F^{\nu\sigma}\nabla_\nu\delta A_\sigma\xi^\mu.$$

Hence Eq. (16.28) holds. We may then use expression (16.27) to write the variation of I as

$$\delta I = \frac{1}{2\mu_0}\delta\int_{\mathcal{S}}A_\sigma\xi^\sigma F_{\mu\nu}dS^{\mu\nu} - \frac{1}{\mu_0}\int_{\Sigma}\nabla^\nu\Omega_{\mu\nu}dV^\mu + \frac{1}{2}\int_{\Sigma}T^{\mu\nu}h_{\mu\nu}\xi^\rho dV_\rho.$$

In view of the definition (16.29), $\Omega^{\mu\nu}$ is antisymmetric, so that $\Omega_{\mu\nu}$ describes a 2-form and we may invoke Lemma 5.12 with $\mathcal{S}_{\text{int}} := \mathcal{S}$ to express the second integral as a flux integral through \mathcal{S} (setting to zero the flux through \mathcal{S}_∞ due to the fast decay of A and F and hence of Ω). We thus get

$$\delta I = \frac{1}{2\mu_0}\delta\int_{\mathcal{S}}A_\sigma\xi^\sigma F_{\mu\nu}dS^{\mu\nu} + \frac{1}{\mu_0}\int_{\mathcal{S}}\delta A_\sigma F_{\mu}^{\sigma}\xi_\nu dS^{\mu\nu} + \frac{1}{2}\int_{\Sigma}T^{\mu\nu}h_{\mu\nu}\xi^\rho dV_\rho,$$

where we have used Eq. (16.29) and the antisymmetry of $dS^{\mu\nu}$ to write $\Omega_{\mu\nu}dS^{\mu\nu} = \delta A_\sigma(F^\sigma_\mu\xi_\nu - F^\sigma_\nu\xi_\mu)dS^{\mu\nu} = 2\delta A_\sigma F^\sigma_\mu\xi_\nu dS^{\mu\nu}$. Substituting the above expression for δI into Eq. (16.26) leads to

$$\delta M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \sum_{i=1}^L \Omega^{(i)} \delta J_{\mathcal{H}_k}^{(i)} \right) + (\delta M_\infty)^{\text{EM}}, \quad (16.30)$$

where the electromagnetic part is defined by

$$(\delta M_\infty)^{\text{EM}} := -\frac{1}{2\mu_0} \delta \int_{\mathcal{S}} A_\sigma \xi^\sigma F_{\mu\nu} dS^{\mu\nu} - \frac{1}{\mu_0} \int_{\mathcal{S}} \delta A_\sigma F^\sigma_\mu \xi_\nu dS^{\mu\nu}.$$

Let us use $\xi = \chi - \sum_{i=1}^L \Omega^{(i)} \eta_{(i)}$ to let appear the null normal χ of the event horizon. Taking into account that $\eta_{(i)\nu} dS^{\mu\nu} = 0$ for $\eta_{(i)}$ is tangent to \mathcal{S} , we get

$$(\delta M_\infty)^{\text{EM}} = -\frac{1}{2\mu_0} \delta \int_{\mathcal{S}} \left(A_\sigma \chi^\sigma - \sum_{i=1}^L \Omega^{(i)} A_\sigma \eta_{(i)}^\sigma \right) F_{\mu\nu} dS^{\mu\nu} - \frac{1}{\mu_0} \int_{\mathcal{S}} \delta A_\sigma F^\sigma_\mu \chi_\nu dS^{\mu\nu}.$$

Now, on each connected component \mathcal{S}_k of \mathcal{S} , Eq. (5.99) gives $F_{\mu\nu} dS^{\mu\nu} = 2\mu_0 \sigma_k dS$, where σ_k is the effective electric charge density of \mathcal{S}_k (cf. Property 5.29) and $dS := \sqrt{q} d^{n-2}x$ is the area element of \mathcal{S}_k . Furthermore, thanks to Eq. (5.86), $\chi_\nu dS^{\mu\nu} = \chi_\nu (\chi^\mu k^\nu - k^\nu \chi^\mu) dS = -\chi^\mu dS$. Hence

$$(\delta M_\infty)^{\text{EM}} = \sum_{k=1}^K \left[-\delta \int_{\mathcal{S}_k} A_\sigma \chi^\sigma \sigma_k dS + \sum_{i=1}^L \delta \int_{\mathcal{S}_k} \Omega^{(i)} A_\sigma \eta_{(i)}^\sigma \sigma_k dS + \frac{1}{\mu_0} \int_{\mathcal{S}_k} \delta A_\sigma F^\sigma_\mu \chi^\mu dS \right].$$

But, $F^\sigma_\mu \chi^\mu = E^\sigma = \mu_0 \sigma_k \chi^\sigma$ on \mathcal{S}_k [Eqs. (5.93) and (5.96)]. Consequently, the last integral cancels out with the part $\delta A_\sigma \chi^\sigma \sigma_k dS$ arising from the first integral and we are left with

$$(\delta M_\infty)^{\text{EM}} = \sum_{k=1}^K \left[- \int_{\mathcal{S}_k} A_\sigma \delta \chi^\sigma \sigma_k dS - \int_{\mathcal{S}_k} A_\sigma \chi^\sigma \delta (\sigma_k dS) + \sum_{i=1}^L \delta \int_{\mathcal{S}_k} \Omega^{(i)} A_\sigma \eta_{(i)}^\sigma \sigma_k dS \right].$$

Now, from Eq. (16.8), $\delta \chi^\sigma = \sum_{i=1}^L \delta \Omega^{(i)} \eta_{(i)}^\sigma$, so that the first integral cancels out with the $\delta \Omega^{(i)}$ part arising from the last integral. Moreover, on \mathcal{S}_k , $A_\sigma \chi^\sigma = -\Phi_{\mathcal{H}_k} = \text{const}$ [Eq. (5.95) with the constant set to zero by a proper choice of \mathbf{A}], so that the second integral can be written as $-\Phi_{\mathcal{H}_k} \delta Q_{\mathcal{H}_k}$, where use has been made of Eq. (5.98) to let appear the electric charge $Q_{\mathcal{H}_k}$. Hence, we get

$$(\delta M_\infty)^{\text{EM}} = \sum_{k=1}^K \left[\Phi_{\mathcal{H}_k} \delta Q_{\mathcal{H}_k} + \sum_{i=1}^L \Omega^{(i)} \delta \int_{\mathcal{S}_k} A_\sigma \eta_{(i)}^\sigma \sigma_k dS \right].$$

Consequently, we may rewrite Eq. (16.30) as

$$\delta M_\infty = \sum_{k=1}^K \left(\frac{\kappa_k}{8\pi} \delta A_k + \Phi_{\mathcal{H}_k} \delta Q_{\mathcal{H}_k} \right) + \sum_{i=1}^L \Omega^{(i)} \sum_{k=1}^K \left(\delta J_{\mathcal{H}_k}^{(i)} + \delta \int_{\mathcal{S}_k} A_\sigma \eta_{(i)}^\sigma \sigma_k dS \right). \quad (16.31)$$

Besides, the Komar angular momentum at infinity with respect to the axisymmetry generated by $\eta_{(i)}$ is given by formula (5.70):

$$J_{\infty}^{(i)} = J_{\mathcal{S}}^{(i)} + \int_{\Sigma} T_{\mu\nu} \eta_{(i)}^{\nu} dV^{\mu},$$

where we have taken (5.48) into account and have set to zero the term involving the trace T in Eq. (5.70) since it is always possible to choose Σ so that all the $\eta_{(i)}$'s are tangent to it, ensuring $\eta_{(i)}^{\mu} dV_{\mu} = 0$. Substituting the electromagnetic form (1.52) for T and using again $\eta_{(i)}^{\mu} dV_{\mu} = 0$, we get

$$J_{\infty}^{(i)} = J_{\mathcal{S}}^{(i)} + \frac{1}{\mu_0} \int_{\Sigma} F_{\rho\mu} F_{\nu}^{\rho} \eta_{(i)}^{\nu} dV^{\mu}.$$

Now since A obeys the spacetime symmetries, we have $\mathcal{L}_{\eta_{(i)}} A = 0$, from which we deduce that $F_{\rho\mu} F_{\nu}^{\rho} \eta_{(i)}^{\nu} = \nabla^{\rho} (A_{\nu} \eta_{(i)}^{\nu} F_{\rho\mu})$, in exactly the same way as that used above to establish $F_{\rho\mu} F_{\nu}^{\rho} \xi^{\nu} = \nabla^{\rho} (A_{\nu} \xi^{\nu} F_{\rho\mu})$ when evaluating the integral I . Hence

$$J_{\infty}^{(i)} = J_{\mathcal{S}}^{(i)} - \frac{1}{\mu_0} \int_{\Sigma} \nabla^{\nu} (A_{\rho} \eta_{(i)}^{\rho} F_{\mu\nu}) dV^{\mu} = J_{\mathcal{S}}^{(i)} + \frac{1}{2\mu_0} \int_{\mathcal{S}} A_{\rho} \eta_{(i)}^{\rho} F_{\mu\nu} dS^{\mu\nu},$$

where the second equality stems from Lemma 5.12. Since $\mathcal{S} = \cup_{k=1}^K \mathcal{S}_k$, $J_{\mathcal{S}}^{(i)} = \sum_{k=1}^K J_{\mathcal{H}_k}^{(i)}$ and $F_{\mu\nu} dS^{\mu\nu} = 2\mu_0 \sigma_k dS$ on \mathcal{S}_k [Eq. (5.99)], we may rewrite the above expression as

$$J_{\infty}^{(i)} = \sum_{k=1}^K \left(J_{\mathcal{H}_k}^{(i)} + \int_{\mathcal{S}_k} A_{\rho} \eta_{(i)}^{\rho} \sigma_k dS \right). \quad (16.32)$$

Then obviously, the term in factor of $\Omega^{(i)}$ in Eq. (16.31) is $\delta J_{\infty}^{(i)}$, which ends the proof of formula (16.24). \square

Remark 2: When contrasted with the electrovacuum Smarr formula (5.100), the mass variation formula (16.24) has the distinctive feature of being independent of the spacetime dimension n .

Remark 3: In the limit of a vanishing electric charge ($Q_{\mathcal{H}} = 0$), formula (16.25) reduces to the vacuum formula (16.23), given that $M_{\infty} = M$ and $J_{\infty} = J$ are independent of the integration surface in vacuum.

Remark 4: An alternative derivation of the mass variation formulas (16.23) and (16.25) can be performed within the ADM Hamiltonian formulation of general relativity, as shown by Sudarsky and Wald (1992) [466, 467] (see also Sec. 6.2 of Ref. [502]). This derivation does not rely on the generalized Smarr formula (used in the proof of Property 16.1, from which formulas (16.23) and (16.25) are derived) but it requires that the black hole event horizon is part of a bifurcate Killing horizon. This last assumption has been removed by Gao and Wald in 2001 [211].

16.2.4 Mass variation formula for a black hole surrounded by a fluid

Another interesting application of the general mass variation formula (16.3) regards a black hole surrounded by some perfect-fluid matter in 4-dimensional general relativity. We first need some preliminaries about the perfect fluid model:

Property 16.4: simple perfect fluid in axisymmetric spacetimes

In a given 4-dimensional spacetime $(\mathcal{M}, \mathbf{g})$, the energy-momentum tensor of a perfect fluid of 4-velocity \mathbf{u} , proper energy density ε and pressure p is

$$\mathbf{T} = (\varepsilon + p)\mathbf{u} \otimes \mathbf{u} + p\mathbf{g}. \quad (16.33)$$

A *simple fluid*^a is defined by an equation of state of the type

$$\varepsilon = \varepsilon(s, n_b), \quad (16.34)$$

where s and n_b are respectively the entropy density and the baryon density, both measured in the fluid's frame. The fluid's **thermodynamic temperature** T and the baryon **relativistic chemical potential**^b μ_b are then defined as the partial derivatives $T := (\partial\varepsilon/\partial s)_{n_b}$ and $\mu_b := (\partial\varepsilon/\partial n_b)_s$. The resulting identity

$$d\varepsilon = Tds + \mu_b dn_b \quad (16.35)$$

expresses the *first law of thermodynamics* for a small *fixed* comoving volume. The fluid's pressure p is not an independent function of (s, n_b) but rather is given by the thermodynamic identity

$$p = Ts + \mu_b n_b - \varepsilon, \quad (16.36)$$

which, along with (16.35), implies the Gibbs-Duhem relation: $dp = s dT + n_b d\mu_b$.

The law of baryon number conservation asserts that $n_b \mathbf{u}$ is a conserved current: $\nabla \cdot (n_b \mathbf{u}) = 0$. Furthermore, if we assume that \mathbf{g} obeys the Einstein equation (1.41), the energy-momentum conservation (1.45) holds: $\nabla \cdot \overrightarrow{\mathbf{T}} = 0$. Its projection along \mathbf{u} lead to the conservation of the entropy current $s\mathbf{u}$, i.e. $\nabla \cdot (s\mathbf{u}) = 0$ (no heat exchange between neighboring elements of a perfect fluid). Given a spacelike hypersurface Σ , the conserved currents $n_b \mathbf{u}$ and $s\mathbf{u}$ give rise to the concepts of *number of baryons* \mathcal{N}_b in Σ and *fluid entropy* S_f in Σ :

$$\mathcal{N}_b := \int_{\Sigma} n_b u^{\mu} dV_{\mu} \quad \text{and} \quad S_f = \int_{\Sigma} s u^{\mu} dV_{\mu}, \quad (16.37)$$

where dV is Σ 's normal volume element vector [Eq. (5.48)]. If the spacetime $(\mathcal{M}, \mathbf{g})$ is axisymmetric (Killing vector $\boldsymbol{\eta}$), a third conserved current is $\mathbf{j} := -1/(16\pi)\mathcal{J}[\boldsymbol{\eta}]$, where $\mathcal{J}[\boldsymbol{\eta}]$ is the Komar current of $\boldsymbol{\eta}$, as given by Eq. (5.69): $\mathcal{J}[\boldsymbol{\eta}] = -2\overrightarrow{\mathbf{R}} \cdot \boldsymbol{\eta}$. Thanks to the Einstein equation (1.42), $\mathbf{j} = \overrightarrow{\mathbf{T}} \cdot \boldsymbol{\eta} - (T^{\mu}_{\mu}/2)\boldsymbol{\eta} = (\varepsilon + p)(\mathbf{u} \cdot \boldsymbol{\eta})\mathbf{u} + (\varepsilon - p)/2\boldsymbol{\eta}$. The *fluid angular momentum* in Σ is the integral of $\mathbf{j} \cdot dV$; assuming that $\boldsymbol{\eta}$ is tangent to Σ (i.e. $\boldsymbol{\eta} \cdot dV = 0$), we get

$$J_f = \int_{\Sigma} (\varepsilon + p) u_{\nu} \eta^{\nu} u^{\mu} dV_{\mu}. \quad (16.38)$$

The three quantities \mathcal{N}_b , S_f and J_f do not depend upon the choice of the hypersurface Σ as long as the latter is large enough to encompass all the fluid, i.e. Σ intersects each fluid line exactly once (cf. Fig. ??).

^aSee Refs. [370] (Chap. 22), [96, 225] or [227] (Chap. 21).

^bBy *relativistic* it is meant that the rest mass is included in the chemical potential.

Proof. The entropy conservation $\nabla \cdot (s\mathbf{u}) = 0$ follows straightforwardly from $\nabla \cdot \vec{T} = 0$ projected along \mathbf{u} , using the identities (16.35), (16.36) and $\nabla \cdot (n_b \mathbf{u}) = 0$ (see e.g. Ref. [225] for details). Because they are fluxes of conserved currents, the quantities \mathcal{N}_b , S_f and J_f take the same values on two hypersurfaces Σ_1 and Σ_2 that are connected by a (timelike or null) hypersurface \mathcal{W} with all the currents vanishing on \mathcal{W} , which happens if \mathcal{W} is located outside the fluid worldtube. \square

From an astrophysical point of view, it is natural to consider that the matter surrounding the black hole forms an accretion disk or torus, so that it can be modeled by a perfect fluid in circular motion around the black hole rotation axis. One then derives from the general law (16.3) the following mass variation formula:

Property 16.5: mass variation formula for a black hole surrounded by a fluid

Let $(\mathcal{M}, \mathbf{g})$ be a 4-dimensional stationary (Killing vector ξ) and axisymmetric (Killing vector η) spacetime that contains a black hole with a connected event horizon \mathcal{H} . We assume that \mathcal{H} is a Killing horizon with respect to the Killing vector $\chi = \xi + \Omega_{\mathcal{H}}\eta$. If $\Omega_{\mathcal{H}} \neq 0$, this requires the hypotheses of the strong rigidity theorem (Property 5.25). Let us suppose that \mathbf{g} obeys the Einstein equation (1.41) with $\Lambda = 0$: $\mathbf{G} = 8\pi\mathbf{T}$, where \mathbf{T} is the energy-momentum tensor of a perfect fluid located in the black hole exterior and whose features are described in Property 16.4. Moreover, let us assume that the fluid is in **circular motion**, i.e. the fluid's 4-velocity \mathbf{u} is a linear combination of the two Killing vectors:

$$\mathbf{u} = u^t (\xi + \Omega\eta), \quad (16.39)$$

where u^t and Ω are two scalar fields. Ω is called the **fluid angular velocity** because $\Omega = d\varphi/dt$ along a fluid line, where t and φ are adapted coordinates: $\partial_t = \xi$ and $\partial_\varphi = \eta$. The notation u^t reminds that this factor is the component of \mathbf{u} along ∂_t in adapted coordinates. It is fully determined by Ω and the three scalar products formed from the two Killing vectors:

$$u^t = (-\xi \cdot \xi - 2\Omega\xi \cdot \eta - \Omega^2\eta \cdot \eta)^{-1/2}. \quad (16.40)$$

Let $(\mathcal{M}, \mathbf{g} + \delta\mathbf{g})$ be a nearby stationary spacetime sharing the same characteristics. The change δM_∞ in Komar mass at infinity between the two spacetimes is given by

$$\delta M_\infty = \frac{\kappa}{8\pi} \delta A + \Omega_{\mathcal{H}} \delta J_{\mathcal{H}} + \int_{\Sigma} \frac{T}{u^t} \delta(dS_f) + \int_{\Sigma} \frac{\mu_b}{u^t} \delta(d\mathcal{N}_b) + \int_{\Sigma} \Omega \delta(dJ_f), \quad (16.41)$$

where κ is \mathcal{H} 's surface gravity, A is \mathcal{H} 's area, $J_{\mathcal{H}}$ is \mathcal{H} 's Komar angular momentum, Σ is any spacelike hypersurface that intersects each fluid line exactly once and that admits $\boldsymbol{\eta}$ as a tangent vector, T is the fluid temperature, μ_b is the baryon chemical potential and dS_f , $d\mathcal{N}_b$ and dJ_f are the infinitesimal increments in the integrals (16.37) and (16.38) defining the fluid's entropy, baryon number and angular momentum:

$$dS_f := su^\mu dV_\mu, \quad d\mathcal{N}_b := n_b u^\mu dV_\mu \quad \text{and} \quad dJ_f := (\varepsilon + p)u_\nu \eta^\nu u^\mu dV_\mu. \quad (16.42)$$

Proof. Let us specialize the general formula (16.3) to the case $n = 4$, $K = 1$, $L_1 = 1$ and \mathbf{g} obeying the Einstein equation (so that the Einstein tensor \mathbf{G} can be replaced by $8\pi\mathbf{T}$):

$$\delta M_\infty = \frac{\kappa}{8\pi} \delta A + \Omega_{\mathcal{H}} \delta J_{\mathcal{H}} + \frac{1}{2} \int_{\Sigma} T^{\mu\nu} \delta g_{\mu\nu} \xi^\rho dV_\rho - \delta \int_{\Sigma} T^\mu_\nu \xi^\nu dV_\mu. \quad (16.43)$$

Let us rewrite the integrand of the last integral by means of the perfect fluid expression (16.33) for \mathbf{T} :

$$T^\mu_\nu \xi^\nu dV_\mu = [(\varepsilon + p)u^\mu u_\nu \xi^\nu + p\xi^\mu] dV_\mu.$$

Substituting $(u^t)^{-1}\mathbf{u} - \Omega\boldsymbol{\eta}$ for $\boldsymbol{\xi}$ [Eq. (16.39)] and using $u_\nu u^\nu = -1$ and $\eta^\mu dV_\mu = 0$ (since $\boldsymbol{\eta}$ is tangent to Σ), we get

$$T^\mu_\nu \xi^\nu dV_\mu = -\frac{\varepsilon}{u^t} u^\mu dV_\mu - \Omega dJ_f,$$

where use has been made of Eq. (16.42) to let appear dJ_f . Let us perform the variation of this expression and use $\delta\varepsilon = T\delta s + \mu_b\delta n_b$ in agreement with Eq. (16.35):

$$\delta(T^\mu_\nu \xi^\nu dV_\mu) = -\frac{1}{u^t}(T\delta s + \mu_b\delta n_b)u^\mu dV_\mu + \frac{\varepsilon\delta u^t}{(u^t)^2}u^\mu dV_\mu - \frac{\varepsilon}{u^t}\delta(u^\mu dV_\mu) - \delta\Omega dJ_f - \Omega\delta(dJ_f).$$

Using Eq. (16.42) to let appear $\delta(dS_f)$ and $\delta(d\mathcal{N}_b)$ and substituting p for $Ts + \mu_b n_b - \varepsilon$ [Eq. (16.36)], we get

$$\delta(T^\mu_\nu \xi^\nu dV_\mu) = -\frac{T}{u^t}\delta(dS_f) - \frac{\mu_b}{u^t}\delta(d\mathcal{N}_b) - \Omega\delta(dJ_f) + \frac{p}{u^t}\delta(u^\mu dV_\mu) + \frac{\varepsilon\delta u^t}{(u^t)^2}u^\mu dV_\mu - \delta\Omega dJ_f.$$

Now $\delta(u^\mu dV_\mu) = \delta u^\mu dV_\mu + u^\mu \delta(dV_\mu)$, with, according to Eqs. (16.39) and (16.4),

$$\delta u^\mu = (\delta u^t/u^t)u^\mu + u^t\delta\Omega\eta^\mu \quad (16.44)$$

and, according to Eqs. (16.17) and (16.21), $\delta(dV_\mu) = (1/2)g^{\rho\sigma}\delta g_{\rho\sigma}dV_\mu$. Hence, using $\eta^\mu dV_\mu = 0$,

$$\delta(u^\mu dV_\mu) = \left(\frac{\delta u^t}{u^t} + \frac{1}{2}g^{\rho\sigma}\delta g_{\rho\sigma} \right) u^\mu dV_\mu.$$

The above expression of $\delta(T^\mu_\nu \xi^\nu dV_\mu)$ becomes then

$$\begin{aligned} \delta(T^\mu_\nu \xi^\nu dV_\mu) &= -\frac{T}{u^t}\delta(dS_f) - \frac{\mu_b}{u^t}\delta(d\mathcal{N}_b) - \Omega\delta(dJ_f) \\ &\quad + \left[\left(\frac{\delta u^t}{(u^t)^2} - \delta\Omega u_\nu \eta^\nu \right) (\varepsilon + p) + \frac{1}{2}\frac{p}{u^t}g^{\rho\sigma}\delta g_{\rho\sigma} \right] u^\mu dV_\mu. \end{aligned}$$

Note that we have used Eq. (16.42) as $\delta\Omega dJ_f = (\varepsilon + p)\delta\Omega u_\nu\eta^\nu u^\mu dV_\mu$. Now, from Eq. (16.44), $\delta u^t/u^t - u^t\delta\Omega u_\nu\eta^\nu = -u_\nu\delta u^\nu$. Hence

$$\delta(T_\nu^\mu \xi^\nu dV_\mu) = -\frac{T}{u^t} \delta(dS_f) - \frac{\mu_b}{u^t} \delta(d\mathcal{N}_b) - \Omega \delta(dJ_f) + \left[\frac{p}{2} g^{\rho\sigma} \delta g_{\rho\sigma} - (\varepsilon + p) u_\nu \delta u^\nu \right] \frac{1}{u^t} u^\mu dV_\mu.$$

Since $g_{\rho\sigma} u^\rho u^\sigma = -1$ implies $u_\rho \delta u^\rho = -(1/2) \delta g_{\rho\sigma} u^\rho u^\sigma$, we may rewrite the above relation as

$$\delta(T_\nu^\mu \xi^\nu dV_\mu) = -\frac{T}{u^t} \delta(dS_f) - \frac{\mu_b}{u^t} \delta(d\mathcal{N}_b) - \Omega \delta(dJ_f) + \frac{1}{2} T^{\rho\sigma} \delta g_{\rho\sigma} \xi^\mu dV_\mu,$$

where we have used Eq. (16.33) to let appear $T^{\rho\sigma}$ and Eq. (16.39) along with $\eta^\mu dV_\mu = 0$ to write $(u^t)^{-1} u^\mu dV_\mu = \xi^\mu dV_\mu$. Integrating the above expression on Σ and substituting into Eq. (16.43) yields formula (16.41). \square

Historical note : The mass-variation formula (16.1) for Kerr black holes has been first derived by Jacob Bekenstein in 1972 [50, 51], using the same differentiation procedure as in Sec. 16.2.1. Actually, the formula derived by Bekenstein applies to Kerr-Newman black holes, being a special case of the generic formula (16.25) for electrically charged black holes. The extension to generic stationary 4-dimensional black holes surrounded by a perfect fluid in circular motion [Eq. (16.41)] has been obtained by James Bardeen, Brandon Carter and Stephen Hawking in 1973 [39]. A different derivation of the same formula (16.41) can also be found in Bardeen's lecture at the famous 1972 Les Houches Summer School [38]. The generalization to include an electromagnetic field has been presented by Carter at the same summer school [95]. In the pure electrovacuum case (no fluid, no electric current), Carter's formula (Eq. (9.64) in Ref. [95]) reduces to the case $K = 1, L = 1$ of formula (16.24). As for the general form (16.3) involving the Einstein tensor, it has been exhibited by Carter in 1979 [98] in the case $n = 4$ and $K = 1$ (connected event horizon). The mass variation formula with a fluid (16.41) has been extended to non-circular flows by Carter in 1979 [97].

16.2.5 A first law?

At this stage, it would be premature to call any of the formulas (16.22), (16.23), (16.24) or (16.41) the first law of black hole dynamics by analogy with the first law of thermodynamics $dE = TdS - PdV$. One can reasonably interpret δM as some energy variation, the term $\Omega_{\mathcal{H}} \delta J$ as the work⁶ performed by the torque that is changing by δJ the angular momentum J of a body rotating at the angular velocity $\Omega_{\mathcal{H}}$ and the term $\Phi_{\mathcal{H}} \delta Q_{\mathcal{H}}$ as the work performed to change by an amount $\delta Q_{\mathcal{H}}$ the electric charge of a body at the electric potential $\Phi_{\mathcal{H}}$. However, we have not got any argument yet to identify the term $(\kappa/8\pi) \delta A$ with the classical heat exchange term TdS . For this, we need first to identify the entropy S with the black hole area A . This is performed in the next section.

⁶In Newtonian mechanics, the work done by a torque τ on a body that is rotating by $d\varphi$ is $dW = \tau d\varphi$. Given that $\tau := dJ/dt$, one gets $dW = \Omega dJ$, where $\Omega = d\varphi/dt$ is the body's angular velocity.

16.3 Evolution of the black hole area

16.3.1 Irreversibility in the evolution of a Kerr black hole

Let us consider a particle \mathcal{P} in free fall from infinity into a Kerr black hole of mass M and spin $J = aM$, with $a < M$ (non-extremal case). Let us denote by E and L the conserved energy and conserved angular momentum of \mathcal{P} (cf. Property 11.1), with $E \ll M$ and $|L| \ll J$. We shall assume that either \mathcal{P} falls directly into the black hole or \mathcal{P} splits in the ergoregion into two particles, \mathcal{P}' and \mathcal{P}_* , such that \mathcal{P}_* falls into the black hole and \mathcal{P}' escapes to infinity – as in the Penrose process described in Sec. 11.3.2. Let (E', L') and (E_*, L_*) be the conserved energy and angular momentum of \mathcal{P}' and \mathcal{P}_* , respectively. As shown in Sec. 11.3.2, these quantities fulfill $E = E' + E_*$ and $L = L' + L_*$. If \mathcal{P} falls directly into the black hole, we shall set $\mathcal{P}_* = \mathcal{P}$, $E' = 0$, $E_* = E$, $L' = 0$ and $L_* = L$. By virtue of the no-hair theorem 5.38, after \mathcal{P}_* has been captured by the black hole and \mathcal{P}' has gone sufficiently far, the black hole appears to the exterior world as a Kerr black hole of mass $M + \delta M$ and angular momentum $J + \delta J$, say. Initially, when \mathcal{P} was far apart from the black hole, the total energy (ADM mass) of the spacetime was $M + E$ (recall that E is the energy at infinity of \mathcal{P} , cf. Sec. 11.2.2) and similarly the total angular momentum was $J + L$. To the first order in E/M , we may neglect the energy and angular momentum of the gravitational radiation generated in the process, so that the final spacetime energy is $M + \delta M + E'$ and the final angular momentum is $J + \delta J + L'$. By conservation of the total energy and angular momentum, we then get $\delta M = E - E'$ and $\delta J = L - L'$, i.e.

$$\delta M = E_* \quad \text{and} \quad \delta J = L_*. \quad (16.45)$$

The evolution of the black hole area A in the process is given by Eq. (16.1):

$\delta A = 8\pi/\kappa (\delta M - \Omega_{\mathcal{H}} \delta J)$. Given the identities (16.45), one gets

$$\delta A = \frac{8\pi}{\kappa} (E_* - \Omega_{\mathcal{H}} L_*). \quad (16.46)$$

Note that the above equation is well-posed since $\kappa > 0$ (the black hole is not extremal). Now, we have established in Sec. 11.3.2 that $E_* > \Omega_{\mathcal{H}} L_*$ [Eq. (11.60)] as a necessary condition for \mathcal{P}_* to cross the event horizon. It follows immediately that $\delta A > 0$:

Property 16.6: area increase for accreting Kerr black holes

Under the accretion of particles of energy sufficiently small to neglect gravitational radiation, the area A of a Kerr black hole necessarily increases:

$$\boxed{\delta A > 0}. \quad (16.47)$$

Remark 1: The area increase holds even if the mass M decreases, which occurs for a Penrose process ($E' > E$), since then $E_* < 0$ and thus $\delta M < 0$ by virtue of Eq. (16.45).

Any transformation of a black hole with $\delta A > 0$ is called an **irreversible transformation**. For particle accretion, the irreversibility traces back to the inequality $E_* > \Omega_{\mathcal{H}} L_*$ [Eq. (11.60)],

which must hold if \mathcal{P}_* is to cross the event horizon, i.e. is to be actually accreted. Indeed, in view of Eq. (16.45), one might think naively that the transformation $(M, J) \rightarrow (M + \delta M, J + \delta J)$ could be reverted by throwing into the black hole a particle of conserved energy $E'_* = -\delta M = -E_*$ and conserved angular momentum $L'_* = -\delta J = -L_*$. However this is not permitted because the constraint $E'_* > \Omega_{\mathcal{H}} L'_*$ is violated if $E_* > \Omega_{\mathcal{H}} L_*$ holds.

A **reversible transformation** is a transformation such that $\delta A = 0$. For particle accretion, it is achieved in the limit case $E_* = \Omega_{\mathcal{H}} L_*$ [cf. Eq. (16.46)]. The particle \mathcal{P}_* reaches then the event horizon \mathcal{H} with a vanishing radial affine velocity $dr/d\lambda$, λ being the affine parameter along \mathcal{P}_* 's worldline associated with \mathcal{P}_* 's 4-momentum (cf. Sec. 11.2.1). Indeed, the equation of geodesic motion (11.43b) yields, on \mathcal{H} ($r = r_+$),

$$\frac{dr}{d\lambda} \Big|_{r=r_+} = -\frac{r_+^2 + a^2}{r_+^2 + a^2 \cos^2 \theta} (E_* - \Omega_{\mathcal{H}} L_*). \quad (16.48)$$

Note that we have set $\epsilon_r = -1$ in Eq. (11.43b) to account for the infall motion of \mathcal{P}_* and have specialized expression (11.44) for $R(r)$ to $\Delta = 0$ ($r = r_+$), which yields $\sqrt{R(r_+)} = (r_+^2 + a^2)|E_* - \Omega_{\mathcal{H}} L_*| = (r_+^2 + a^2)(E_* - \Omega_{\mathcal{H}} L_*)$ thanks to Eq. (11.60) and the identity $a/(r_+^2 + a^2) = \Omega_{\mathcal{H}}$ [Eq. (10.69)]. Equation (16.48) clearly shows that $dr/d\lambda \rightarrow 0$ in the limit $\Omega_{\mathcal{H}} L_* \rightarrow E_*$. Moreover, from expression (11.29) for $R(r)$, one gets, for a geodesic with $\Omega_{\mathcal{H}} L_* = E_*$, $R'(r_+) = -2\sqrt{M^2 - a^2}(\mu_*^2 r_+^2 + \mathcal{K}_*)$, where μ_* is the mass of \mathcal{P}_* and \mathcal{K}_* is the Carter constant of \mathcal{P}_* . Given that $M^2 - a^2 \neq 0$ (non-extremal Kerr) and $\mathcal{K}_* \geq 0$ [Eq. (11.30)], we obtain $R'(r_+) \neq 0$, except if $\mu_* = 0$ and $\mathcal{K}_* = 0$. According to Property 11.2, the last case would occur only if \mathcal{P}_* followed a principal null geodesic of Kerr spacetime. However, this is excluded since those geodesics do not fulfill $\Omega_{\mathcal{H}} L_* = E_*$ in the black hole exterior: they obey $L_* = aE_* \sin^2 \theta$ [Eqs. (11.6) and (11.9b)], so that $\Omega_{\mathcal{H}} L_* = E_*$ would imply $\Omega_{\mathcal{H}} a \sin^2 \theta = 1$, i.e. $r_+^2 + a^2 \cos^2 \theta = 0$, which is impossible. Hence one has always $R'(r_+) \neq 0$. We conclude that \mathcal{P}_* has necessarily a r -turning point on \mathcal{H} (cf. Sec. 11.2.6). Hence a reversible transformation corresponds to the idealized case where the particle \mathcal{P}_* is on the verge of being accreted by the black hole but is actually not. This case is approached when \mathcal{P}_* crosses \mathcal{H} with an affine velocity $dr/d\lambda$ as small as possible.

The square root of the horizon area A (rescaled by $1/(16\pi)$) is called the black hole's **irreducible mass**:

$$M_{\text{irr}}^2 := \frac{A}{16\pi}. \quad (16.49)$$

M_{irr} has the dimension of a mass and Eq. (16.47) is equivalent to $\delta M_{\text{irr}} > 0$, i.e. M_{irr} cannot decrease in any accretion process, hence its name. This contrasts with M , which may decrease (Penrose process). Formula (10.83) for A can be turned into an expression for M_{irr} in terms of M and $J = aM$ (cf. Fig. 16.1):

$$M_{\text{irr}} = M \sqrt{\frac{1}{2} \left(1 + \sqrt{1 - \frac{J^2}{M^4}} \right)}. \quad (16.50)$$

This relation is easily inverted to express M in terms of M_{irr} and J :

$$M^2 = M_{\text{irr}}^2 + \frac{J^2}{4M_{\text{irr}}^2}. \quad (16.51)$$

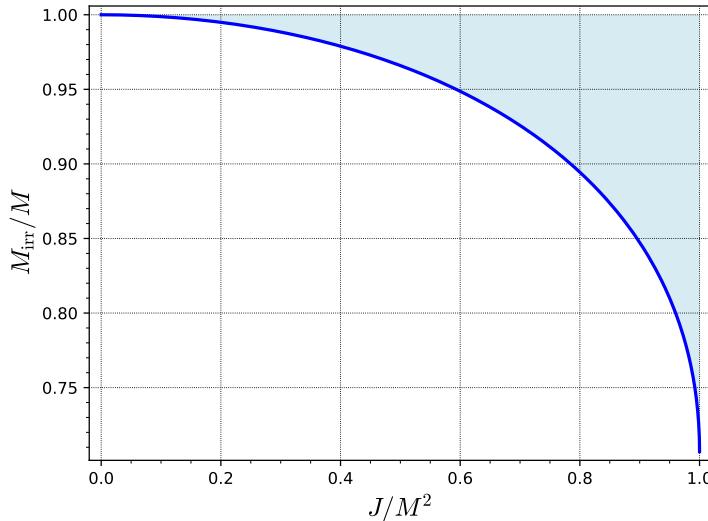


Figure 16.1: Irreducible mass M_{irr} of a Kerr black hole of mass M , as a function of the black hole angular momentum J [Eq. (16.50)] ($J = 0$ corresponds to a Schwarzschild black hole and $J/M^2 = 1$ to an extremal Kerr one). The shaded area depicts the maximal energy that can be extracted from the black hole by a Penrose process bringing J down to 0.

Suppose some energy is extracted by a sequence of Penrose processes (cf. Sec. 11.3.2) from a Kerr black hole of initial mass M and angular momentum J , leading to a final state corresponding to a Kerr black hole of mass M_0 and angular momentum J_0 . The maximal energy $\Delta M = M - M_0$ that can be extracted is achieved when the final state is a Schwarzschild black hole ($J_0 = 0$), since then the Penrose process necessarily stops, due to the lack of an outer ergoregion (cf. Sec. 11.3.2). Since the irreducible mass of a Schwarzschild black hole coincides with its mass (set $J = 0$ in Eq. (16.50)), the irreducible mass of the final state is $M_{\text{irr},0} = M_0$ and we have $\Delta M = M - M_{\text{irr},0}$. The evolution law $M_{\text{irr},0} \geq M_{\text{irr}}$ provides then a upper bound on the extracted energy:

$$\Delta M \leq M - M_{\text{irr}}, \quad (16.52)$$

with equality only for a sequence of reversible transformations. Figure 16.1 shows that $\Delta M/M$ is maximal when the initial state is (close to) an extremal Kerr black hole ($J = M^2$). From Eq. (16.50), one gets

$$\max \frac{\Delta M}{M} = 1 - \frac{1}{\sqrt{2}} \simeq 0.293. \quad (16.53)$$

We conclude:

Property 16.7: maximal energy extraction by a Penrose process

A sequence of Penrose processes can extract up to 29% of the mass of a Kerr black hole. This upper bound is achieved for a maximally spinning black hole (extremal Kerr), by bringing it down to a non-rotating state (Schwarzschild) via reversible transformations.

Historical note : Demetrios Christodoulou introduced the concepts of *irreversible transformation* and *irreducible mass* of a Kerr black hole in 1970 [111] and established formula (16.51). Actually, he did not

make any argument about the black hole area; rather he integrated the relation $dM = \Omega_{\mathcal{H}} dJ$, which holds when the black hole is accreting a particle with $E_* = \Omega_{\mathcal{H}} L_*$ (cf. Eq. (16.45), i.e. at the saturation limit of the inequality (11.60). Indeed, by expressing $\Omega_{\mathcal{H}}$ as a function of (M, J) via Eq. (10.69), one gets $dM = J dJ / (2M(M^2 + \sqrt{M^4 - J^2}))$, which is equivalent to $d(M^2 + \sqrt{M^4 - J^2}) = 0$. The irreducible mass M_{irr} then appears when writing the integration constant for this equation as $2M_{\text{irr}}^2$, namely $M^2 + \sqrt{M^4 - J^2} = 2M_{\text{irr}}^2$, which is equivalent to Eq. (16.50). The connection with the black hole area [Eq. (16.49)] has been performed in a second study, with Remo Ruffini, published in 1971 [113] and which extends Christodoulou's results to Kerr-Newmann black holes. Meanwhile, Roger Penrose and Roger Floyd noticed at the end of 1970 (published 1971 [408]) that the area of a Kerr black hole necessarily increases under particle accretion, i.e. they established Property 16.6. In 1972, Brandon Carter [93] pointed out that the relation $dM = \Omega_{\mathcal{H}} dJ$ also holds for a rigidly rotating fluid star undergoing a thermodynamically reversible transformation (in that case, $\Omega_{\mathcal{H}}$ stands for the star's angular velocity). This strengthened the analogy between irreversibility in black hole processes and standard thermodynamics.

16.3.2 Hawking's area theorem

The area increase law turns out to be far more general than established in Property 16.6, which regards only a Kerr black hole accreting particles: the law actually applies to any black hole evolving under any process (not necessarily a small perturbation) such that the null convergence condition [Eq. (2.94)] is satisfied in the vicinity of the event horizon. This is the essence of the famous *area theorem*, first established by Hawking in 1971.

The first step towards the area theorem is:

Property 16.8: non-negative expansion of a black hole horizon

Let (\mathcal{M}, g) be a n -dimensional spacetime containing a black hole of event horizon \mathcal{H} . If the Ricci tensor \mathbf{R} obeys the null convergence condition (2.94) on \mathcal{H} , i.e. if $\mathbf{R}(\ell, \ell) \geq 0$ for any null vector ℓ on \mathcal{H} – which holds in general relativity if the null energy condition (2.95) is fulfilled –, then the expansion $\theta_{(\ell)}$ of \mathcal{H} along any future-directed null normal ℓ , as defined in Sec. 2.3.5, is positive or zero:

$$\theta_{(\ell)} \geq 0. \quad (16.54)$$

Proof. Let ℓ be a future-directed null normal vector field of \mathcal{H} . ℓ is necessarily tangent to the null geodesic generators of \mathcal{H} (Property 2.1) and is thus a pregeodesic vector, i.e. it obeys Eq. (2.22): $\nabla_\ell \ell = \kappa \ell$. If ℓ is not geodesic ($\kappa \neq 0$), it is always possible to rescale it to $\ell' = \alpha \ell$ with $\alpha > 0$ so that ℓ' is a future-directed geodesic vector field: $\nabla_{\ell'} \ell' = 0$ [Eq. (2.24)]. We have then $\theta_{(\ell')} = \alpha \theta_{(\ell)}$ [cf. Eq. (2.68)], so that $\theta_{(\ell)} \geq 0 \iff \theta_{(\ell')} \geq 0$. Accordingly, for proving (16.54), there is no loss of generality in assuming that ℓ is a geodesic vector field. Let us consider a null geodesic generator \mathcal{L} of \mathcal{H} . Up to some additive constant, there is a unique affine parameter λ of \mathcal{L} associated to ℓ , i.e. such that $\ell = dx/d\lambda$ along \mathcal{L} . The evolution of $\theta_{(\ell)}$ along \mathcal{L} is measured by $d\theta_{(\ell)}/d\lambda = \nabla_\ell \theta_{(\ell)}$ and is given by the null Raychaudhuri

equation (2.88). Owing to $\kappa = 0$ (for ℓ is assumed to be geodesic), it simplifies to

$$\frac{d\theta_{(\ell)}}{d\lambda} = -\frac{1}{n-2}\theta_{(\ell)}^2 - \underbrace{\sigma_{ab}\sigma^{ab}}_{\geq 0} - \underbrace{\mathbf{R}(\ell, \ell)}_{\geq 0},$$

where $\sigma_{ab}\sigma^{ab} \geq 0$ has been established in Sec. 2.4.2 [Eq. (2.92)] and $\mathbf{R}(\ell, \ell) \geq 0$ holds by virtue of the null convergence condition on \mathcal{H} . Hence

$$\frac{d\theta_{(\ell)}}{d\lambda} \leq -\frac{1}{n-2}\theta_{(\ell)}^2. \quad (16.55)$$

Let us assume the negation of (16.54), i.e. that there exists a point $p \in \mathcal{L} \cap \mathcal{H}$ where $\theta_{(\ell)} = \theta_0 < 0$. By choosing the additive constant in the definition of the affine parameter λ , we may ensure $\lambda(p) = 0$. Equation (16.55) implies then

$$\forall \lambda \geq 0, \quad \theta_{(\ell)}(\lambda) \leq \bar{\theta}(\lambda), \quad (16.56)$$

where $\bar{\theta}(\lambda)$ obeys

$$\frac{d\bar{\theta}}{d\lambda} = -\frac{1}{n-2}\bar{\theta}^2 \quad \text{and} \quad \bar{\theta}(0) = \theta_0.$$

The unique solution of this differential equation is

$$\bar{\theta}(\lambda) = \frac{\theta_0}{1 + \theta_0\lambda/(n-2)}.$$

It follows that $\bar{\theta} \rightarrow -\infty$ as $\lambda \rightarrow -(n-2)/\theta_0 > 0$. The inequality (16.56) then implies $\theta_{(\ell)} \rightarrow -\infty$ as $\lambda \rightarrow \lambda_*$ with $0 < \lambda_* \leq -(n-2)/\theta_0$. Hence the point $p_* \in \mathcal{L}$ of parameter λ_* is a *focusing point*, i.e. a point where neighboring null geodesic generators of \mathcal{H} intersect. But according to Property 4.8 of black hole event horizons (cf. Sec. 4.4.3), this can happen only if p_* is a crossover point, i.e. a point at which the null geodesic \mathcal{L} enters \mathcal{H} ; however, this situation is excluded since $\lambda_* > 0$ implies that p_* lies in the future of p , where \mathcal{L} is already in \mathcal{H} . Hence the hypothesis $\theta_0 < 0$ leads to a contradiction. It follows that $\theta_0 \geq 0$, i.e. at any point $p \in \mathcal{H}$, $\theta_{(\ell)} \geq 0$. \square

Property 16.9: area theorem (Hawking 1971 [258], Chruściel, Delay, Galloway and Howard 2001 [119])

Let (\mathcal{M}, g) be a spacetime of dimension $n \geq 2$ that contains a black hole of event horizon \mathcal{H} . Let us assume that the Ricci tensor \mathbf{R} fulfills the null convergence condition (2.94), i.e. $\mathbf{R}(\ell, \ell) \geq 0$ for any null vector ℓ – which holds in general relativity if the null energy condition (2.95) is fulfilled. Let Σ_1 and Σ_2 be spacelike hypersurfaces such that Σ_2 lies in the causal future of Σ_1 : $\Sigma_2 \subset J^+(\Sigma_1)$ (cf. Sec. 4.4.1). Let $\mathcal{S}_1 = \mathcal{H} \cap \Sigma_1$ and $\mathcal{S}_2 = \mathcal{H} \cap \Sigma_2$. If

- (i) \mathcal{H} is smooth between \mathcal{S}_1 and \mathcal{S}_2 , and \mathcal{S}_1 and \mathcal{S}_2 are cross-sections^a of \mathcal{H} that are intersected by the same null geodesic generators of \mathcal{H} (cf. Fig. 16.2)

or

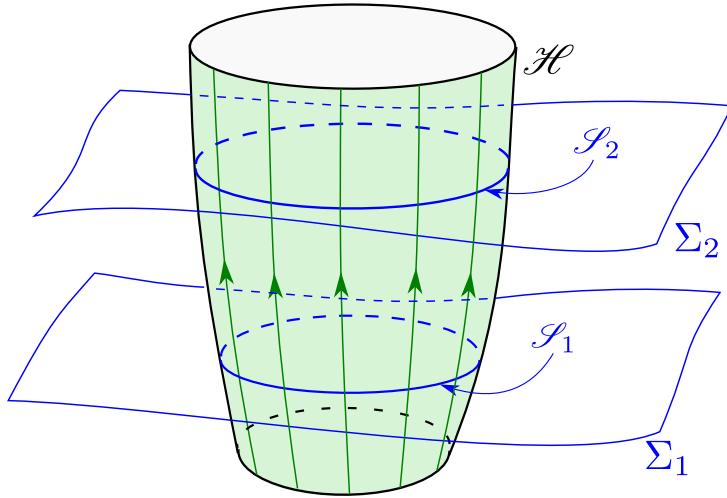


Figure 16.2: Cross-sections \mathcal{S}_1 and \mathcal{S}_2 induced by the spacelike hypersurfaces Σ_1 and Σ_2 in a smooth part of the event horizon \mathcal{H} , such that \mathcal{S}_1 and \mathcal{S}_2 are intersected by the same null geodesic generators of \mathcal{H} (green curves).

- (ii) the closure of $J^-(\mathcal{I}^+) \cup \mathcal{H}$ in the conformal manifold $\tilde{\mathcal{M}} \supset \mathcal{M}$ defining \mathcal{I}^+ is included in a globally hyperbolic region \mathcal{V} of $(\tilde{\mathcal{M}}, \tilde{g})$ and Σ_1 and Σ_2 are Cauchy surfaces of \mathcal{V} ,

then the areas $A(\mathcal{S}_1)$ and $A(\mathcal{S}_2)$ obey

$$A(\mathcal{S}_2) \geq A(\mathcal{S}_1). \quad (16.57)$$

^aIf \mathcal{H} is smooth between \mathcal{S}_1 and \mathcal{S}_2 , it is necessarily a null hypersurface there (Property 4.11 on p. 118), so that the concept of cross-section as defined in Sec. 2.3.4 makes sense.

Proof. For $n = 2$, the theorem is trivial since \mathcal{S}_1 and \mathcal{S}_2 are each reduced to a single point, so that $A(\mathcal{S}_1) = A(\mathcal{S}_2) = 0$ and Eq. (16.57) is fulfilled. Let us then assume $n \geq 3$. We consider first the case (i) (\mathcal{H} smooth between \mathcal{S}_1 and \mathcal{S}_2). \mathcal{S}_1 and \mathcal{S}_2 are then cross-sections of \mathcal{H} that are connected by null geodesic generators of \mathcal{H} (cf. Fig. 16.2). One may introduce a 1-parameter foliation $(S_t)_{t \in [0,1]}$ of \mathcal{H} by cross-sections S_t such that $S_0 = \mathcal{S}_1$ and $S_1 = \mathcal{S}_2$ (cf. the proof of Property 3.1 for details, λ playing the role of t there). The label t can then be considered as a parameter along the null geodesic generators of \mathcal{H} connecting \mathcal{S}_1 to \mathcal{S}_2 . Let $\ell = dx/dt$ be the associated tangent vector. Given that \mathcal{S}_2 lies in the future of \mathcal{S}_1 , ℓ is future-directed. By the very definition of the expansion $\theta_{(\ell)}$ along ℓ [Eq. (2.49) combined with Eq. (2.53)], one has

$$\frac{d}{dt} A(S_t) = \int_{S_t} \theta_{(\ell)} \sqrt{q} \, dx^2 \cdots dx^{n-1},$$

where (x^2, \dots, x^{n-1}) is a coordinate system on S_t and q is the determinant with respect to these coordinates of the metric induced by g on S_t . Now, if the null convergence condition holds, Property 16.8 implies $\theta_{(\ell)} \geq 0$; hence

$$\frac{d}{dt} A(S_t) \geq 0.$$

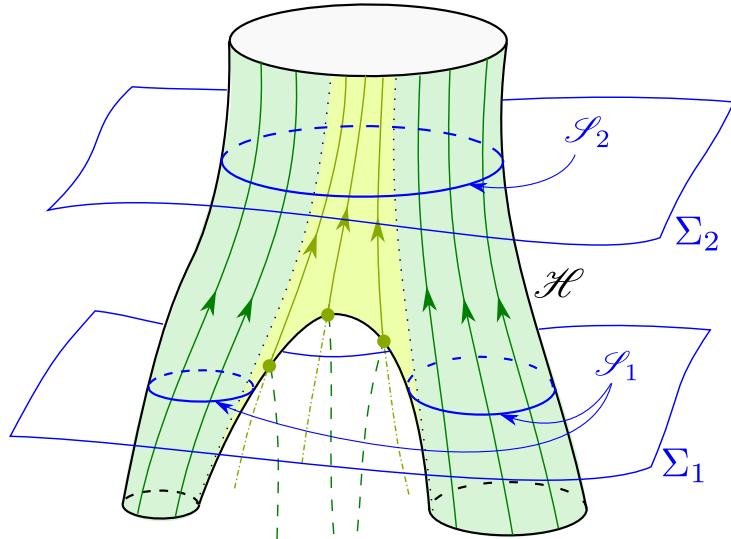


Figure 16.3: Surfaces \mathcal{S}_1 and \mathcal{S}_2 induced by the spacelike hypersurfaces Σ_1 and Σ_2 on the event horizon \mathcal{H} corresponding to a black hole merger. All solid dark green and light green curves are null geodesic generators of \mathcal{H} . The light green part of \mathcal{H} is generated by null geodesics that entered \mathcal{H} at some caustic points (three of them are indicated as light green dots); the parts of these geodesics outside \mathcal{H} are drawn as light green dot-dashed curves. The dark green dashed curves depict other null geodesics that enter \mathcal{H} at the same caustic points.

It follows that $t \mapsto A(S_t)$ is a nondecreasing function. Consequently, $A(S_1) \geq A(S_0)$ and Eq. (16.57) holds.

If \mathcal{H} is not smooth between \mathcal{S}_1 and \mathcal{S}_2 , this is due to the crease set, i.e. the subset of \mathcal{H} where new null geodesics enter \mathcal{H} (cf. Sec. 4.4.3). Naively, this reinforces the inequality $A(\mathcal{S}_2) > A(\mathcal{S}_1)$ since the new geodesics are generating new parts of \mathcal{H} and therefore parts of \mathcal{S}_2 distinct from those that can be connected to \mathcal{S}_1 by null geodesic generators (cf. Fig. 16.3). More precisely, let us assume (ii) and let us consider a point $p \in \mathcal{S}_1$. Let \mathcal{L} be the null geodesic generator of \mathcal{H} through p . By property 4.8, \mathcal{L} stays in \mathcal{H} for all its future after p . Since Σ_2 is a Cauchy surface lying in the causal future of Σ_1 , \mathcal{L} necessarily intersects Σ_2 at a unique point $q \in \Sigma_2 \cap \mathcal{H} = \mathcal{S}_2$. Hence every point of \mathcal{S}_1 is mapped to a point of \mathcal{S}_2 by a null geodesic generator. Let \mathcal{S}_2^* be the part of \mathcal{S}_2 covered by this map. If we assume that the part of \mathcal{H} between \mathcal{S}_1 and \mathcal{S}_2^* is smooth, we may apply (i) to the pair $(\mathcal{S}_1, \mathcal{S}_2^*)$ and get $A(\mathcal{S}_2^*) \geq A(\mathcal{S}_1)$. Given that $\mathcal{S}_2^* \subset \mathcal{S}_2$, one has $A(\mathcal{S}_2) \geq A(\mathcal{S}_2^*)$ and (16.57) follows. We refer to the article [119] for the proof in the case where \mathcal{H} is not assumed smooth between \mathcal{S}_1 and \mathcal{S}_2^* (70 pages!). \square

Example 1 (Oppenheimer-Snyder collapse): Let us consider the black hole formed by the collapse of a ball of pressureless matter, as described by the Oppenheimer-Snyder model studied in Sec. 14.3. Let $\mathcal{S}_{\tilde{t}}$ be a cross-section of the event horizon at constant ingoing Eddington-Finkelstein coordinate \tilde{t} . The area of $\mathcal{S}_{\tilde{t}}$ is simply $A = 4\pi r^2$, where r is the areal radius (cf. Sec. 14.3.5). Its evolution can thus be read directly on Fig. 14.2 (right): it is increasing from $A = 0$ (at $\tilde{t} \simeq 5.6$ m) to the Schwarzschild value $A = 16\pi m^2$ (at $\tilde{t} \simeq 9.7$ m). This fully agrees with the area theorem, given that the null energy condition (2.95) is fulfilled by the energy-momentum tensor (14.2) of the collapsing matter: $\mathbf{T}(\ell, \ell) = \rho(\mathbf{u} \cdot \ell)^2 \geq 0$ since $\rho \geq 0$.

Example 2 (Vaidya collapse): Similarly, we check on Figs. 15.1 and 15.5 that for the radiation shell collapse studied in Chap. 15, the area of cross-sections of the event horizon at constant ingoing Eddington-Finkelstein coordinate t is increasing towards the future, for r is again the areal radius [cf. the metric (15.11)]. As we have already noticed in Sec. 15.2.1, the radiation energy-momentum tensor (15.5) fulfills the null energy condition, given that $M'(v) \geq 0$ (cf. Fig. 15.2).

Historical note : As mentioned in the historical note on p. 635, Roger Penrose and Roger Floyd [408] have shown in 1970 (published 1971) that the area of a Kerr black hole always increases during particle accretion, even if its mass is decreasing, as in a Penrose process (Property 16.6). Soon after, in 1971, Stephen Hawking [258] established the area theorem for generic dynamical black holes and a detailed proof was given in Hawking & Ellis' 1973 textbook (Proposition 9.2.7 in [265]). In 2001, Piotr Chruściel, Erwann Delay, Gregory Galloway and Ralph Howard [119] pointed out that the Hawking & Ellis proof is valid only for \mathcal{H} piecewise smooth (cf. the discussion in Sec. 3.5.1 of Chruściel's textbook [118]); they constructed a new proof that does not rely on the smoothness of \mathcal{H} .

16.3.3 A second law?

Basically the area theorem 16.9 states that the area of cross-sections of a black hole event horizon cannot decrease from the past to the future. By its irreversible character, this property bears some resemblance with the second law of thermodynamics. By itself, this is of course not sufficient to identify the black hole area with some entropy (any nondecaying physical quantity has not to be an entropy!). However, we have seen in Sec. 16.2.5 that the candidate TdS term for a possible first law could be $\kappa/(8\pi)dA$. It is thus tempting to identify A with an entropy S , up to some constant factor α . Then the temperature T would be $\kappa/(8\pi\alpha)$:

$$S = \alpha A \quad (16.58a)$$

$$T = \frac{1}{8\pi\alpha} \kappa, \quad (16.58b)$$

so that $\kappa/(8\pi)dA = TdS$, which makes the mass variation formula (16.23) look pretty much like the first law of thermodynamics.

16.4 The laws of black hole dynamics

In Secs. 16.2 and 16.3, we have established black hole properties that are quite analog to the first and second laws of thermodynamics. Let us now examine the zeroth and third laws of thermodynamics.

16.4.1 Zeroth law

We have already established a so-called *zeroth law of black hole dynamics* in Chap. 3, namely the surface gravity κ of a Killing horizon is constant, provided that the null dominance condition is fulfilled (Property 3.10) or that the Killing horizon is part of a bifurcate Killing horizon (Property 3.16). Now, by Properties 5.23 and 5.25, (a connected component of) the event horizon of a stationary black hole must be a Killing horizon. Hence, we may extend the zeroth law to black holes:

Property 16.10: zeroth law of black hole dynamics

Let (\mathcal{M}, g) be a stationary spacetime of dimension $n \geq 4$ containing a black hole. Let \mathcal{H} be a connected component of the black hole event horizon. If \mathcal{H} is non-rotating (i.e. if the stationary Killing vector ξ is null on \mathcal{H}), \mathcal{H} is necessarily a Killing horizon (Property 5.23). If \mathcal{H} is rotating (ξ spacelike on some parts of \mathcal{H}), we shall assume that the hypotheses of the strong rigidity theorem (Property 5.25) hold, so that \mathcal{H} is a Killing horizon as well. If the null dominance condition (3.43) is fulfilled on \mathcal{H} — which is guaranteed in general relativity if the null dominant energy condition (3.46) holds on \mathcal{H} — or if \mathcal{H} is part of a bifurcate Killing horizon, then the surface gravity κ is constant over \mathcal{H} .

Given that a black hole “in equilibrium” is modeled by a black hole in a stationary spacetime, this property bears a strong resemblance with the *zeroth law of thermodynamics*, which states that the temperature of a body in equilibrium is uniform over the entire body. This strengthens the interpretation of the surface gravity κ as the temperature T (up to some factor) performed in Eq. (16.58b).

16.4.2 What about the third law?

The classical Nernst formulation of the *third law of thermodynamics* states that the entropy of a system must approach zero (or a universal constant) as its temperature tends to zero. In this form and with the identification (16.58), this law certainly does not hold for black holes, because extremal black holes, which have $\kappa = 0$, have a non-vanishing area. For instance, the area of the extremal Kerr black hole of mass m (Chap. 13) is $A = 8\pi m^2$ (take the limit $a \rightarrow m$ in Eq. (10.83)), which is neither zero, nor a universal constant. Similarly, the area of the extremal Reissner-Nordström black hole of mass m is $A = 4\pi m^2$ (this is immediate from the metric (18.1) below). Another formulation of the third law of thermodynamics states that it is impossible to bring any system to zero temperature by a finite number of operations. This was the formulation adopted for black holes, as a conjecture, in 1973 by Carter [95] and Bardeen, Carter and Hawking [39]. It was then reformulated, with a tentative proof, by Israel in 1986 [290], as

No continuous process in which the energy-momentum tensor of accreted matter remains bounded and satisfies the weak energy condition in a neighborhood of the apparent horizon can reduce the surface gravity of a black hole to zero within a finite advanced time.

However, as pointed out recently by Kehle and Unger [310], Israel’s proof is restricted to the case where outermost trapped surfaces evolve smoothly, while, as we shall discuss in Chap. 18, their evolution generically presents discontinuous jumps, even if the spacetime and all the matter fields remain perfectly smooth. In particular, Kehle and Unger [310] have exhibited a regular spacetime where the collapse of a charged scalar field (obeying the dominant, and hence the weak, energy condition) turns a Schwarzschild black hole into an extremal Reissner-Nordström

black hole within a finite advanced time⁷. Let us note that this result has been recently mitigated by Reall [426], who has shown that the collapse to a supersymmetric extremal black hole (such as the extremal Reissner-Nordström one, but not the extremal Kerr one) cannot occur in theories with an upper bound on the local charge to mass ratio.

From an astrophysical point of view, Bardeen has shown in 1970 [35] that a Schwarzschild black hole of mass m_0 , and hence of surface gravity $\kappa = 1/(4m_0) > 0$ [Eq. (2.29)], can in principle be spun up to an extremal Kerr black hole ($\kappa = 0$) by accreting matter from an accretion disk terminating at the innermost stable circular orbit (ISCO, cf. Sec. 11.5.3). The total rest-mass of accreted matter is then $1.86 m_0$ and the mass of the final black hole is $m \simeq 2.45 m_0$. However, if one takes into account the electromagnetic emission of the accretion disk, the extremal Kerr state cannot be achieved. The reason is that a substantial part of the emitted photons carry a negative angular momentum ℓ and the black hole absorption cross-section for photons with $\ell < 0$ is larger than for those with $\ell > 0$. This clearly appears on the shadow picture of Fig. 12.20, where the part $\alpha > 0$, which is due to photons with $\ell < 0$ (cf. Eq. 12.88a), is much larger than the part $\alpha < 0$ corresponding to $\ell > 0$. Consequently, the black hole absorbs more photons with $\ell < 0$ than with $\ell > 0$; this diminishes the increase of the black hole spin a induced by accretion of matter from the disk (which has a positive angular momentum). Thorne has shown in 1974 [477] that the balance between the two processes leads to a final spin parameter $a \simeq 0.998 m$, quite insensitive to the details of the emission mechanism in the accretion disk. This is close to, but strictly lower than, the critical value $a = m$ that would have yield a zero surface gravity. Hence, in this respect, we may say that astrophysical black holes absorbing matter from an accretion disk obey a kind of third law. But this does not appear to arise from some fundamental properties of black holes; it rather results from the properties of their environment.

It must be stressed that in standard thermodynamics as well, the third law has not the same fundamental status as the other laws. In particular it can be violated by some rather simple systems (see Ref. [503] for a discussion and an example involving a boson (or fermion) gas confined to a circular ring).

For all the above reasons, we shall no longer discuss the third law here.

16.4.3 Summary: the three laws

Let us collect the results obtained so far, in the form of three laws similar to the laws of thermodynamics. To stress the analogy, we are using the phrase *black hole in equilibrium* for *black hole in a stationary spacetime* (cf. Chap. 5). Besides, for the sake of brevity, we shall not repeat the various assumptions underlying these laws, referring to previously stated properties for all the details. Furthermore, we consider a connected event horizon for simplicity.

Property 16.11: laws of black hole dynamics

Zeroth law: Under the hypotheses of Property 16.10, the surface gravity κ of the event

⁷As shown in Sec. 5.5.6 of Poisson's textbook [415], this cannot happen with a charged null dust (i.e. a charged generalization of Vaidya collapse discussed in Chap. 15) that obeys the weak energy condition.

horizon of a black hole in equilibrium is uniform over the horizon:

$$\kappa = \text{const.} \quad (16.59)$$

First law: Under the hypotheses of Property 16.3 (in particular, general relativity), the change δM in total mass between two nearby electrovacuum configurations of a black hole in equilibrium is related to the change δA in horizon area, the change δJ in total angular momentum and the change $\delta Q_{\mathcal{H}}$ in electric charge by^a

$$\delta M = \frac{\kappa}{8\pi} \delta A + \Omega_{\mathcal{H}} \delta J + \Phi_{\mathcal{H}} \delta Q_{\mathcal{H}}, \quad (16.60)$$

where $\Omega_{\mathcal{H}}$ and $\Phi_{\mathcal{H}}$ are the horizon's angular velocity and electric potential respectively.

Second law: Under the hypotheses of Property 16.9, the area of cross-sections of a black hole event horizon cannot decrease in time:

$$A(\mathcal{S}_2) \geq A(\mathcal{S}_1), \quad (16.61)$$

as soon as the cross-section \mathcal{S}_2 lies in the causal future of the cross-section \mathcal{S}_1 .

^aFor simplicity, we are considering black holes with a connected horizon and a single angular momentum; see Eq. (16.24) for the general formula, where M is denoted by M_∞ and J by J_∞ .

Remark 1: The three laws are valid for any spacetime dimension $n \geq 4$, and even $n \geq 2$ for the second law. If one assumes that the event horizon is a Killing horizon, then the zeroth law is valid for $n \geq 2$ as well.

Historical note : The above ensemble of laws, plus a tentative third law, has been first stated in 1973 by James Bardeen, Brandon Carter and Stephen Hawking in a seminal article entitled *The Four Laws of Black Hole Mechanics* [39]. This article, which was mostly written during the famous 1972 Les Houches Summer School [163], exhibits proofs of the zeroth and first laws for $n = 4$, relies on Hawking's article [259] for the second law, and presents the third law as a conjecture, saying that “*there are strong reasons for believing in it*” (cf. Sec. 16.4.2).

The three laws listed in Property 16.11 are named the laws of black hole *dynamics* and not *thermodynamics*, because they still appear as a mathematical analogy without any definite physical content. In particular, if we stick to a pure classical (i.e. non-quantum) point of view and would like to attribute some thermodynamic temperature T to a black hole, it should be $T = 0$. Indeed, if the black hole is placed in contact with a thermal reservoir of temperature $T_0 > 0$, energy will necessarily flow from the reservoir to the black hole and not in the reverse way; this implies $T < T_0$, whatever T_0 . For a stationary non-extremal black hole ($\kappa \neq 0$), $T = 0$ contradicts the tentative identification (16.58b) between T and κ . It is striking that quantum field theory in curved spacetime restores the identification, thereby opening the path to a genuine thermodynamics of black holes. The identification $T \propto \kappa$ arises from Hawking radiation, which we discuss in the next section.

16.5 Hawking radiation

16.5.1 The Hawking radiation phenomenon

Property 16.12: Hawking radiation

Let (\mathcal{M}, g) be a n -dimensional ($n \geq 2$) asymptotically flat spacetime that is stationary and contains a black hole, the event horizon of which is a Killing horizon of constant surface gravity κ . Then, quantum field theory in the curved background (\mathcal{M}, g) predicts that any quantum field gives birth to a thermal radiation from the black hole to infinity, called **Hawking radiation**. The radiation temperature as measured by asymptotic inertial observers at rest with respect to the black hole is

$$T_H = \frac{\hbar}{k_B} \frac{\kappa}{2\pi}, \quad (16.62)$$

where $\hbar = 1.054571817 \cdot 10^{-34}$ J s is the reduced Planck constant and $k_B = 1.380649 \cdot 10^{-23}$ J K $^{-1}$ is the Boltzmann constant. T_H is called the **Hawking temperature** of the black hole.

We shall not establish this property here; this is done in many textbooks, e.g. [498, 502, 201, 86, 189], as well in some review articles, e.g. [74, 152, 292]. Rather, we shall limit ourselves to a few remarks:

Remark 1: Hawking radiation is an outcome of quantum field theory within a fixed *classical* curved background (see e.g. [502]). It has not been derived from any quantum theory of gravity, i.e. the background spacetime (\mathcal{M}, g) is not quantized. One says that Hawking radiation is a *semiclassical effect*.

Remark 2: The phenomenon of Hawking radiation is essentially kinematic and does not depend on any field equation for the metric g . In particular it does not rely on the Einstein equation (see e.g. Ref. [495]). Moreover, the Hawking temperature (16.62) is independent from the spacetime dimension n (see e.g. Ref. [306]).

Remark 3: Formula (16.62) is given for $c = 1$ units, which are used throughout the text. If one restores c and considers that κ has the dimension of an acceleration, the formula should read $T_H = \frac{\hbar}{k_B} \frac{\kappa}{2\pi c}$.

Remark 4: The Hawking temperature T_H is the radiation temperature measured infinitely far from the black hole; it is not a “local” temperature measured in the vicinity of the event horizon \mathcal{H} . More precisely, let us assume for simplicity that the black hole is non-rotating (cf. Property 5.2) and that its exterior is strictly static. The stationary Killing vector ξ is then null on \mathcal{H} and timelike in the black hole exterior. Let us consider static observers at a finite distance from \mathcal{H} , i.e. observers of 4-velocity $u = \xi/V$, where $V := \sqrt{-\xi \cdot \xi}$ [cf. Eqs. (3.70) and (3.61)]. The temperature of the Hawking radiation

measured by these observers is ⁸

$$T = \frac{T_H}{V}. \quad (16.63)$$

For infinitely distant observers, $V \rightarrow 1$ [cf. Eq. (5.1)], and we recover $T = T_H$. For static observers infinitely close to the event horizon, $V \rightarrow 0$ and we get $T \rightarrow +\infty$. Note that those observers are also experiencing an infinite acceleration [cf. Eq. (3.75)]. On the contrary, a freely-falling observer crossing the horizon measures $T = 0$, i.e. she does not detect the Hawking radiation (cf. Remark 5 hereafter).

Remark 5: The Hawking radiation is closely related to the ***Unruh effect***, which is another prediction of quantum field theory: in vacuum Minkowski spacetime, a uniformly accelerated observer measures a black-body radiation of all particle species at the temperature $T_U = (\hbar/k_B) a/(2\pi)$, where $a = \sqrt{\mathbf{a} \cdot \mathbf{a}}$ is the norm of the observer's 4-acceleration [483, 502, 86]. A static observer \mathcal{O} close to the black hole horizon \mathcal{H} , as considered in Remark 4, feels an acceleration given by Eq. (3.74): $a \sim \kappa/V$. By combining formulas (16.63) and (16.62), we may then rewrite the radiation temperature T measured by \mathcal{O} as $T \sim (\hbar/k_B) a/(2\pi)$. Hence, it coincides with the Unruh temperature T_U of the observer in Minkowski spacetime sharing the same acceleration a as \mathcal{O} . This is somehow expected if one assumes that freely-falling observers near \mathcal{H} perceive quantum states identical to the Minkowski vacuum (no radiation).

Remark 6: In terms of elementary particles, the Hawking radiation is composed of all existing species of massless and massive particles. For the latter ones, the contribution is significant only if $k_B T_H > m_p$, where m_p is the particle's mass. For Schwarzschild black holes of mass $M \gg 2 \cdot 10^{13}$ kg (the threshold for $k_B T_H$ being lower than the electron mass, see below), the Hawking radiation is composed quasi-exclusively of neutrinos (87% of the radiated energy), photons (12%) and possibly gravitons (1%) [393, 480].

Remark 7: The observed radiation spectrum at infinity is not exactly that of a black body of temperature T_H : it is corrected by greybody factors, which originate from the interaction of the radiation with the spacetime curvature (backscattering).

For a Kerr black hole, the surface gravity κ is given by formula (10.77), so that the Hawking temperature can be expressed in terms of the black hole mass $M = m$ and the dimensionless spin parameter $\bar{a} := a/m = J/M^2$ ($0 \leq \bar{a} \leq 1$):

$$T_H = \boxed{\frac{\hbar}{k_B} \frac{c^3}{8\pi GM} \frac{2}{1 + (1 - \bar{a}^2)^{-1/2}}} \Big|_{\text{Kerr}}. \quad (16.64)$$

Note that we have restored the gravitational constant G and the speed of light c (cf. Remark 3). Numerically, we get

$$T_H = 6.17 \cdot 10^{-8} \left(\frac{M_\odot}{M} \right) \frac{2}{1 + (1 - \bar{a}^2)^{-1/2}} \text{ K}, \quad (16.65)$$

or, in units more adapted to small black holes,

$$k_B T_H = 1.057 \cdot 10^{10} \left(\frac{1 \text{ kg}}{M} \right) \frac{2}{1 + (1 - \bar{a}^2)^{-1/2}} \text{ GeV}. \quad (16.66)$$

⁸Cf. e.g. Eq. (7.2.10) in [502]; in view of Eq. (16.63), we may say that Hawking radiation obeys ***Tolman-Ehrenfest law***, which states that the temperature T of a medium in thermal equilibrium in a stationary gravitational field obeys $TV = \text{const}$ [443].

Hence, for a Schwarzschild black hole ($\bar{a} = 0$), $T_H = 6.17 \cdot 10^{-8} (M_\odot/M)$ K. For astrophysical black holes, this leads to tiny temperatures: $T_H \simeq 4 \cdot 10^{-9}$ K for a $15 M_\odot$ stellar black hole (e.g. Cyg X-1, cf. Table 7.1), $T_H \simeq 1.6 \cdot 10^{-14}$ K for the Milky Way central black hole Sgr A* ($M = 4.1 \cdot 10^6 M_\odot$) and $T_H \simeq 9 \cdot 10^{-18}$ K for M87* ($M = 6.5 \cdot 10^9 M_\odot$). The value $T_H = 1$ K is achieved for a black hole of mass $6.17 \cdot 10^{-8} M_\odot$, i.e. 2% of the Earth's mass, while the value $k_B T_H = m_e c^2$ ($m_e = 9.11 \cdot 10^{-31}$ kg being the electron mass) is achieved for $M = 2.07 \cdot 10^{13}$ kg.

Another conclusion that one can draw from formula (16.64) is that $T_H = 0$ for an extremal Kerr black hole ($\bar{a} = 1$): such an object does not emit any Hawking radiation.

Remark 8: Formula (16.64) shows that a Schwarzschild black hole ($\bar{a} = 0$) has a *negative heat capacity*: its temperature increases while its energy (M) diminishes. This feature is shared with Newtonian self-gravitating systems, if one defines the temperature of such systems from the mean kinetic energy of their constituents.

16.5.2 Black hole evaporation

Since it emits some radiation to infinity, the black hole loses energy, which makes its mass decrease. For a Schwarzschild black hole, this results in a temperature increase (cf. Remark 8) and hence an enhanced Hawking radiation. Furthermore, the higher the temperature, the more quantum fields can join the radiation (cf. Remark 6), which enhances the rate of energy loss as well. We are thus facing a runaway process, in which the whole black hole eventually disappears. This phenomenon is called **Hawking evaporation** or **black hole evaporation**. However, the evaporation occurs on extremely long time scales for astrophysical black holes, as we are going to see.

To get a rough estimate of the evaporation time, let us consider that the total power (energy radiated per unit time, also called *luminosity*) L of Hawking radiation received at infinity obeys the Stefan-Boltzmann law: $L = \mathcal{A}\sigma T_H^4$, where σ is Stefan-Boltzmann constant: $\sigma = \pi^2 k_B^4 / (60\hbar^3)$ and \mathcal{A} is some “emission area”. For simplicity, let us restrict to a Schwarzschild black hole of mass M . Then, a suitable value of \mathcal{A} for an estimate is $\mathcal{A} = 4\pi b_c^2$, where $b_c = 3\sqrt{3}M$ is the critical impact parameter for null geodesics introduced in Chap. 8 [cf. Eq. (8.21)]. Indeed, $4\pi b_c^2$ can be viewed as the black hole's area “seen from infinity” (cf. Fig. 8.24). Moreover, it can be shown that the spectrum of Hawking radiation at high frequencies (i.e. $\nu \gg M^{-1}$, so that quantum fields can be treated within geometrical optics) is that of a black body of temperature T_H and total area $4\pi b_c^2$ (see Fig. 1 of Ref. [393]). Hence $\mathcal{A} = 108\pi M^2$. Given that $k_B T_H = \hbar/(8\pi M)$ [Eq. (16.64) with $\bar{a} = 0$], Stefan-Boltzmann law yields

$$L = C \frac{\hbar}{M^2}, \quad (16.67)$$

where $C = 9/(20480\pi) \simeq 1.40 \cdot 10^{-4}$. Actually, a precise computation, taking into account the propagation of quantum fields in the Schwarzschild geometry and various particle species, leads to the same formula (16.67) with [480, 393, 394]

$$C = 2.83 \cdot 10^{-4}. \quad (16.68)$$

This value is valid for $M \gg 2 \cdot 10^{13}$ kg (T_H below the electron threshold, cf. Remark 6), so that the Hawking radiation contains only photons, neutrinos and gravitons⁹. The black hole mass decreases according to $dM/dt = -L$, i.e.

$$\frac{dM}{dt} = -C \frac{\hbar}{M^2}. \quad (16.69)$$

Here t is the time measured by an inertial observer at infinity. This differential equation is easily integrated into

$$M(t) = (M_0^3 - 3Ch\bar{t})^{1/3}, \quad (16.70)$$

where M_0 is the mass at $t = 0$. Defining the evaporation time t_{evap} as the time to reach $M = 0$ from $M = M_0$, we deduce from Eq. (16.70) the following property:

Property 16.13: Hawking evaporation of a Schwarzschild black hole

As seen by an inertial observer at infinity, a Schwarzschild black hole of initial mass M_0 fully evaporates via Hawking radiation within a time

$$t_{\text{evap}} = \frac{M_0^3}{3Ch} = 1.54 \cdot 10^{66} \left(\frac{M_0}{M_\odot} \right)^3 \text{ yr}, \quad (16.71)$$

where the value (16.68) of C has been used to get the numerical estimate.

For an astrophysical black hole ($M_0 > 1M_\odot$), t_{evap} is huge — more than 56 orders of magnitude larger than the age of the Universe ($t_{\text{Univ}} \simeq 1.38 \cdot 10^{10}$ yr)! We conclude that Hawking radiation is ultra-negligible for astrophysical black holes and does not play any role in their dynamics. Hawking radiation becomes relevant only when $t_{\text{evap}} < t_{\text{Univ}}$, say. From Eq. (16.71), one obtains $t_{\text{evap}} < t_{\text{Univ}} \iff M_0 < 4.15 \cdot 10^{11}$ kg. Actually, this value is below $2 \cdot 10^{13}$ kg, so that electrons, and even muons, should have been included into the Hawking radiation (cf. Remark 6). However, the above estimate is pretty good, a precise computation leading to [351]

$$t_{\text{evap}} < t_{\text{Univ}} \iff M_0 < 5.00 \cdot 10^{11} \text{ kg}. \quad (16.72)$$

The upper bound is roughly the mass of an asteroid of half-kilometer size. Note that a black hole of that mass has a Schwarzschild radius of only 0.74 fm (around the proton size!); it would pertain to the hypothetical class of the so-called **primordial black holes**, i.e. black holes of subsolar mass that could have been formed in the primordial universe (see e.g. [85] for a review).

Remark 9: For a Schwarzschild black hole, the horizon area A is proportional to the square of the mass: $A = 16\pi M^2$ [Eq. (3.4)], so that the evaporation law (16.69) implies $dA/dt < 0$. At first glance, this seems

⁹The original computation, performed by Page in 1976 [393], resulted in $C = 2.01 \cdot 10^{-4}$ but it took into account only the four species of neutrinos known at that time (ν_e , $\bar{\nu}_e$, ν_μ and $\bar{\nu}_\mu$), in addition to photons and gravitons. It has been updated by adding the tau neutrinos (ν_τ and $\bar{\nu}_\tau$) by Thorne, Zurek & Price in 1986 [480]. The value (16.68) is also recovered by setting $n_{1/2} = 3$, $n_1 = 1$ and $n_2 = 1$ in Eq. (19) of Page's review article [394].

to contradict the second law of black hole dynamics (Property 16.11), i.e. the area theorem 16.9. Actually there is no contradiction since one of the hypotheses of the theorem is not fulfilled for evaporating black holes, namely the null convergence condition. Indeed, the effective energy-momentum tensor¹⁰ of the quantum fields generating the Hawking radiation does not obey the null energy condition (2.95).

Historical note : In 1971, while introducing the superradiant scattering of classical fields by a Kerr black hole (cf. Sec. 11.3.2 and the historical note on p. 390), Yakov Zeldovich [526] pointed out that the quantum analog of the field superradiance would be the spontaneous creation of particle/antiparticle pairs, with one particle falling into the black hole and the other one escaping to infinity. Zeldovich and his doctoral student Alexei Starobinsky presented a computation of this effect to Stephen Hawking while he was visiting Moscow in September 1973, accompanied by Kip Thorne (see Hawking's account in Chap. 7 of [263] and Thorne's one in Chap. 12 of [478]). Hawking tried to rederive the effect via a different approach of quantum field theory in curved spacetime and he realized that superradiance is not at all required: the spontaneous radiation takes place even for a non-rotating black hole, which has no outer ergoregion and thus cannot trigger any Penrose/superradiance process. This fantastic discovery was presented in short article published in early 1974 [261] and the computations were detailed in a subsequent one published in 1975 [262]. Both articles give formula (16.62) relating the radiation temperature to the surface gravity and discuss the black hole evaporation.

16.6 Black hole thermodynamics

16.6.1 Bekenstein-Hawking entropy

The phenomenon of Hawking radiation puts on firm grounds the identification of the surface gravity as the black hole temperature and fully fixes the coefficient α appearing in Eq. (16.58). Indeed, comparing Eq. (16.58b) with formula (16.62) for $T = T_H$ yields $\alpha = k_B/(4\hbar)$. Accordingly, Eq. (16.58a) becomes $S = k_B A/(4\hbar)$ and we may state:

Property 16.14: Bekenstein-Hawking entropy

A n -dimensional ($n \geq 3$) black hole in equilibrium is endowed with the **Bekenstein-Hawking entropy** S_{BH} , which is proportional to the area A of the event horizon according to the formula:

$$S_{BH} = k_B \frac{A}{4\hbar}. \quad (16.73)$$

For $n \geq 4$ and Einstein-Maxwell theory (see Property 16.3 for detailed assumptions), the first law of black hole dynamics (16.60) writes then

$$\delta M = T_H \delta S_{BH} + \Omega_{\mathcal{H}} \delta J + \Phi_{\mathcal{H}} \delta Q_{\mathcal{H}}, \quad (16.74)$$

where T_H is the Hawking temperature.

¹⁰See e.g. Ref. [202] for the computation of the renormalized expectation value of the energy-momentum tensor of a quantum field near the black hole horizon.

Remark 1: The subscript “BH” in S_{BH} is meant for *Bekenstein-Hawking*, not for *black hole*.

Remark 2: As for the Hawking temperature (16.62), formula (16.73) for the Bekenstein-Hawking entropy does not depend on the spacetime dimension n . It is often summarized by $S_{\text{BH}} = A/4$, which holds in units such that $k_{\text{B}} = 1$ and $\hbar = 1$.

Remark 3: For $n = 4$, if one restores the G ’s and c ’s, formula (16.73) reads $S_{\text{BH}} = k_{\text{B}} \frac{c^3 A}{4G\hbar}$. One recognizes the square of the **Planck length**: $\ell_{\text{P}} := \sqrt{G\hbar/c^3} \simeq 1.62 \cdot 10^{-35}$ m, so that formula (16.73) can be rewritten in terms of the dimensionless ratio A/ℓ_{P}^2 as

$$S_{\text{BH}} = k_{\text{B}} \frac{A}{4\ell_{\text{P}}^2}. \quad (16.75)$$

For a Kerr black hole, the area A is related to the mass M and to the spin parameter $a = \bar{a}M$ via Eq. (10.83), so that the Bekenstein-Hawking entropy (16.73) can be expressed as

$$\boxed{S_{\text{BH}} = 4\pi k_{\text{B}} \frac{GM^2}{\hbar c} \frac{1 + \sqrt{1 - \bar{a}^2}}{2}}_{\text{Kerr}}. \quad (16.76)$$

Note that, as in formula (16.64) for the Hawking temperature, we have restored G and c in the above formula. Numerically, we get

$$S_{\text{BH}} = 1.05 \cdot 10^{77} \left(\frac{M}{M_{\odot}} \right)^2 \frac{1 + \sqrt{1 - \bar{a}^2}}{2} k_{\text{B}}. \quad (16.77)$$

Hence, for a Schwarzschild black hole ($\bar{a} = 0$), $S_{\text{BH}} = 1.05 \cdot 10^{77} (M/M_{\odot})^2 k_{\text{B}}$. This yields huge values for astrophysical black holes: $S_{\text{BH}} \simeq 2 \cdot 10^{79} k_{\text{B}}$ for a $15 M_{\odot}$ stellar black hole, $S_{\text{BH}} \simeq 2 \cdot 10^{90} k_{\text{B}}$ for Sgr A* ($M = 4.1 \cdot 10^6 M_{\odot}$) and $S_{\text{BH}} \simeq 4 \cdot 10^{96} k_{\text{B}}$ for M87* ($M = 6.5 \cdot 10^9 M_{\odot}$). These numbers are terribly large: besides black holes, the total entropy of the observable Universe is about $1.1 \cdot 10^{90} k_{\text{B}}$. It originates mostly from the cosmic microwave background and the cosmic neutrino background (roughly one half each), largely dominating the entropy of the interstellar and intergalactic medium ($\sim 10^{82} k_{\text{B}}$) and that of all the stars ($\sim 10^{81} k_{\text{B}}$) [177]. It follows that the entropy of a single massive black hole, such as Sgr A* or M 87*, is larger than the entropy of the whole observable Universe!

16.6.2 The generalized second law

In the presence of a black hole, the standard second law of thermodynamics cannot hold in the Universe outside the black hole: the total entropy of matter and radiation located there can be decreased by dropping some matter (and its entropy) into the black hole. One then has to replace the standard second law by

Property 16.15: generalized second law of thermodynamics (GSL) (Bekenstein 1973 [51, 52])

Let us consider a “closed system” formed by a black hole and some matter and electromag-

netic radiation in the black hole exterior. Define the ***generalized entropy*** S_{gen} by

$$S_{\text{gen}} := S_{\text{mat}} + S_{\text{BH}}, \quad (16.78)$$

where S_{mat} is the ordinary entropy of the matter and radiation and S_{BH} is the Bekenstein-Hawking entropy (16.73) of the black hole. Then, any physical process makes S_{gen} increase, or at the very least, stay constant:

$$\Delta S_{\text{gen}} \geq 0. \quad (16.79)$$

The GSL should be considered more as a postulate ruling thermodynamics rather than a theorem that can be established. At most it can be proven in some specific cases. For instance, if one regards the process of Hawking radiation itself, one has $\Delta S_{\text{BH}} < 0$, since the black hole area diminishes during Hawking evaporation, as discussed at the end of Sec. 16.5.1. But one can show that the entropy of the Hawking radiation compensates for this loss, leading to $\Delta S_{\text{mat}} \geq |\Delta S_{\text{BH}}|$, so that $\Delta S_{\text{gen}} \geq 0$ is fulfilled (see Ref. [504] for details). Besides, one can also show that the GSL is actually obeyed in various thought experiments naively devised to violate it [84, 502, 504, 509].

Historical note : In 1970-71, the discovery of the irreversible increase of the area of a Kerr black hole interacting with particles by Roger Penrose, Roger Floyd and Demetrios Christodoulou [408, 111], along with the general area theorem established by Stephen Hawking [258] (cf. the historical notes on p. 635 and 640), suggested an analogy between the black hole area and the entropy in thermodynamics. In 1972, Jacob Bekenstein [49, 51] was the first to go beyond this formal analogy by considering that a black hole is endowed with a genuine entropy S proportional to its area A , “*as a measure of the inaccessibility of information (to an exterior observer) as to which particular internal configuration of the black hole is actually realized in a given case*” [51]. Moreover, by means of some heuristic arguments, Bekenstein found that the relation between S and A should be $S = \eta k_B A / (4\pi\hbar)$, where η is a dimensionless coefficient of order unity [49, 51]. This is exactly formula (16.73) provided that $\eta = \pi$. This value of η was determined only after the black hole temperature was set to the Hawking temperature (16.62). This was done by Stephen Hawking in the 1975 article detailing the computation of Hawking radiation [262] (cf. historical note on p. 648). As for the generalized second law of thermodynamics, it was proposed by Bekenstein himself in the 1972-73 articles introducing the Bekenstein-Hawking entropy [49, 51]. Further arguments for its validity have been provided in a subsequent article [52]. Previously, it was said that the (standard) second law of thermodynamics was *transcended* in presence of a black hole [39, 95], for it could not hold when some matter (and its entropy!) is thrown into the black hole.

16.6.3 Black hole entropy from quantum gravity

By superseding the standard second law of thermodynamics, the GSL provides a genuine thermodynamic meaning to the Bekenstein-Hawking entropy. In particular, by adding S_{BH} to the “ordinary” entropy S_{mat} , formula (16.78) puts S_{BH} on the same footing as S_{mat} . It would be fully satisfactory though to get a “statistical mechanics” origin for S_{BH} , i.e. to interpret S_{BH} as a measure of the number \mathcal{N} of (quantum) *microscopic states* corresponding to a given macroscopic state of the black hole, according to Boltzmann’s formula:

$$S_{\text{BH}} = k_B \ln \mathcal{N}. \quad (16.80)$$

Defining the microscopic states a priori requires a quantum theory of gravity, which we do not have yet. However some partial results have been obtained, based on the two main current approaches to quantum gravity: string theory (cf. e.g. [401] and Chap. 9 of [245]) and loop quantum gravity (cf. e.g. [409, 441]). Regarding the former, Strominger and Vafa (1996) [464] have first established the Bekenstein-Hawking formula (16.73) by counting the microstates of an extremal supersymmetric black hole in dimension $n = 5$. By *extremal*, it is meant that the event horizon is a degenerate Killing horizon (cf. Sec. 5.4.3), i.e. it has $\kappa = 0$, or equivalently, a vanishing Hawking temperature. By *supersymmetric*, it is meant that these black hole solutions have been obtained within a $N = 4$ supersymmetric Yang-Mills theory, which corresponds to a low energy limit of some string theory. Further developments are discussed in the review articles [84, 152, 284, 504, 512, 524]; they are still limited to extremal black holes or near-extremal ones. Computations are indeed easier for such black holes because their near-horizon geometries have additional symmetries, as we have seen for the extremal Kerr case (cf. Sec. 13.4). For the latter, the Bekenstein-Hawking entropy of has been recovered via some holographic duality¹¹ [246, 136]. Let us stress that the Bekenstein-Hawking entropy of a Schwarzschild black hole, which is non-extremal, has not been recovered by string theory yet.

Regarding loop quantum gravity, first approaches, devised by Rovelli in 1996 [440] and by Ashtekar, Baez, Corichi and Krasnov in 1997 [21], resulted in a statistical entropy proportional to the black hole area, as for the Bekenstein-Hawking entropy (16.73). Subsequent studies lead to the following formula for the entropy of a spherically symmetric macroscopic black hole of area A (see e.g. Ref. [187] and Refs. [32, 84, 409] for reviews):

$$S = k_B \left[\frac{\gamma_0}{\gamma} \frac{A}{4\hbar} - \frac{3}{2} \ln \left(\frac{A}{\hbar} \right) \right], \quad (16.81)$$

where $\gamma_0 \simeq 0.274$ and the dimensionless quantity γ is the so-called *Barbero-Immirzi parameter*. Its value is not set by the theory (yet). This unknown parameter is not expected to appear at the classical limit of loop quantum gravity, which should be general relativity. On the contrary, in the quantum regime, γ plays a significant role by determining the quantum of area a_0 in terms of the Planck length $\ell_P := \sqrt{G\hbar/c^3}$ (cf. Remark 3) as $a_0 = 4\sqrt{3}\pi\gamma\ell_P^2$ [441]. For large values of A , i.e. for $A \gg \hbar$, the logarithmic term can be neglected in formula (16.81) and one recovers the Bekenstein-Hawking entropy (16.73), provided that $\gamma = \gamma_0$.

Other attempts to derive the Bekenstein-Hawking entropy from statistical mechanics within quantum gravity are described in Sec. 7 of Carlip's review [84]. See also Ref. [244] for an historical review up to 2013.

16.7 Concluding remarks

16.7.1 Black hole entropy and the holographic principle

...

¹¹This is the *Kerr/CFT correspondence* mentioned in Sec. 13.4.6.

16.7.2 Dependency on general relativity

The first law of black hole dynamics (16.60) strongly depends on the theory of gravity being general relativity. Indeed, it is derived from the general mass variation formula (16.3) by substituting the electromagnetic field energy-momentum tensor for the Einstein tensor G (cf. the proof of Property 16.3), which amounts to assuming the electrovacuum Einstein equation. On the other hand, the second law of black hole dynamics (cf. Property 16.11), depends less on general relativity for it requires only the null convergence condition (2.94) to be fulfilled (cf. the area theorem 16.9). This is of course guaranteed by general relativity with matter and fields obeying the null energy condition (2.95), but this might hold in some other theories as well. Similarly, the zeroth law (16.59) depends weakly on general relativity; it is even fully independent of it if the event horizon is part of a bifurcate Killing horizon (Property 3.16) or if the spacetime is axisymmetric with orthogonal transitivity of the $\mathbb{R} \times \text{SO}(2)$ group action (cf. Remark 6 on p. 74); if none of these properties is guaranteed, then the zeroth law requires the null dominance condition (3.43) (cf. Property 3.10), which is fulfilled by general relativity combined with the null dominant energy condition (3.46), but may be obeyed in other theories as well.

Hence, among the three laws listed in Property 16.11, the first one is the most sensitive to a deviation from general relativity. Contemplating its expression (16.60), one may wonder whether the term $\frac{\kappa}{8\pi} \delta A$ could be changed to something else in an alternative theory of gravity. Would the new term be of the form $\kappa \delta(\dots)$, we could still interpret it as a “ $T dS$ ” term, given that the identification $T \leftrightarrow \kappa$ remains valid beyond general relativity, thanks to the universal character of Hawking radiation (cf. Remark 2 on p. 644). This would then question the identification of the entropy S with the black hole area A , i.e. the Bekenstein-Hawking formula (16.73). We shall discuss this in the next chapter.

Chapter 17

Black hole thermodynamics beyond general relativity

Contents

17.1 Introduction	653
17.2 Diffeomorphism-invariant theories of gravity	653
17.3 Wald entropy	664
17.4 Covariant phase space formalism	668
17.5 First law for diffeomorphism-invariant theories	682
17.6 What about the second law?	691

17.1 Introduction

This chapter is in a draft stage.

This chapter relies heavily on exterior calculus, i.e. on calculus on differential forms. The unfamiliar reader is urged to take a look at Sec. A.4.3 of Appendix A first.

17.2 Diffeomorphism-invariant theories of gravity

17.2.1 Framework

We consider that the n -dimensional spacetime (\mathcal{M}, g) is ruled by a metric theory of gravity built upon a least action principle, with the following action:

$$\mathcal{S} = \int_{\mathcal{M}} L = \int_{\mathcal{M}} L \epsilon = \int_{\mathcal{M}} L \sqrt{-g} d^n x. \quad (17.1)$$

Here \mathbf{L} is the **Lagrangian n -form**¹, L is the **Lagrangian scalar** and ϵ is the Levi-Civita tensor (cf. Sec. A.3.4). The second equality reflects the identity² $\mathbf{L} = L \epsilon$ and the third one provides an expression of the integral in terms of a coordinate system (x^α) on \mathcal{M} , g being the determinant of g with respect to these coordinates [cf. Eq. (A.52)]. For simplicity we disregard any scalar or matter field, so that \mathbf{L} is a function of the metric tensor g only. Actually all that follows can be easily generalized to account for extra fields, as in scalar-tensor theories for instance, where \mathbf{L} is a function of g and some scalar field ϕ . We limit ourselves here to g , mostly in order to keep notations simple. More precisely, we assume that at a given point $p \in \mathcal{M}$, \mathbf{L} depends only on the values of g and its derivatives at p , up to a certain order. By *derivative*, it is meant derivative with respect to some affine connection on \mathcal{M} , for instance the one formed by the partial derivatives with respect to given coordinates. We shall write $\mathbf{L}(g)$ to stress this dependency of \mathbf{L} on g , according to the following convention:

Notation: Boldface parentheses are used to denote the local (pointwise) dependency of a tensor field on \mathcal{M} with respect to other ones: $A(B, C)$ means that at a given point $p \in \mathcal{M}$, A depends only on the values of the tensor fields B and C and their derivatives at p .

We shall consider only gravity theories whose formulation does not depend upon any specific coordinates on \mathcal{M} , nor upon any extra-structure on \mathcal{M} , like a privileged vector field. This amounts to demanding that the Lagrangian n -form \mathbf{L} is covariant under the group of diffeomorphisms of \mathcal{M} . More precisely, under the action of a diffeomorphism $\Phi : \mathcal{M} \rightarrow \mathcal{M}$, the metric tensor is transformed to its pullback $(\Phi^{-1})^* g$ by Φ^{-1} (cf. Sec. A.2.8 for the definition of the pullback). The inverse of Φ is invoked here because, as g is a tensor of type $(0, 2)$, Φ^* would carry it from³ $\Phi(\mathcal{M})$ to \mathcal{M} , while we are considering the transformation from \mathcal{M} to $\Phi(\mathcal{M})$ (cf. Remark 12 on p. 719). Let $\mathbf{L}((\Phi^{-1})^* g)$ be the Lagrangian n -form constructed by keeping the same functional dependence in $(\Phi^{-1})^* g$ as $\mathbf{L}(g)$ has in g . The invariance of the theory under diffeomorphisms amounts to demanding that, as a n -form on \mathcal{M} , $\mathbf{L}((\Phi^{-1})^* g)$ is nothing but the n -form $\mathbf{L}(g)$ transformed under the action of Φ , i.e. the pullback of $\mathbf{L}(g)$ by Φ^{-1} : $\mathbf{L}((\Phi^{-1})^* g) = (\Phi^{-1})^* \mathbf{L}(g)$. Given that the spacetimes (\mathcal{M}, g) and $(\Phi(\mathcal{M}), (\Phi^{-1})^* g)$ are physically equivalent, this means that $\mathbf{L}(g)$ and $\mathbf{L}((\Phi^{-1})^* g)$ are basically the same object. In particular, the action \mathcal{S} given by Eq. (17.1) takes the same value if $\mathbf{L}((\Phi^{-1})^* g)$ is substituted for $\mathbf{L}(g)$. If one interprets the diffeomorphism action in a “passive” way, i.e. as a change of coordinates on \mathcal{M} , the diffeomorphism-invariance condition simply means that the functional form of L expressed in Eq. (17.3) below does not depend upon the choice of coordinates. Since the condition $\mathbf{L}((\Phi^{-1})^* g) = (\Phi^{-1})^* \mathbf{L}(g)$ must hold for any diffeomorphism, we may simplify it by introducing $\Phi' := \Phi^{-1}$ and renaming Φ' to Φ :

¹As the integral of a n -form over a n -dimensional manifold, \mathcal{S} is well-posed and coordinate independent.

²Since the space of n -forms at any point $p \in \mathcal{M}$ is 1-dimensional, \mathbf{L} and ϵ are necessarily proportional, so that there is no loss of generality in the writing $\mathbf{L} = L \epsilon$.

³Since Φ is a diffeomorphism $\mathcal{M} \rightarrow \mathcal{M}$, we have of course $\Phi(\mathcal{M}) = \mathcal{M}$; however we keep the notation $\Phi(\mathcal{M})$ to stress the image character by Φ .

A gravity theory is said to be **diffeomorphism-invariant** iff its Lagrangian n -form \mathbf{L} is **covariant** under the group of diffeomorphisms of \mathcal{M} , i.e. iff for any diffeomorphism $\Phi : \mathcal{M} \rightarrow \mathcal{M}$,

$$\mathbf{L}(\Phi^* \mathbf{g}) = \Phi^* \mathbf{L}(\mathbf{g}), \quad (17.2)$$

where Φ^* stands for the pullback by Φ , as defined in Sec. A.2.8.

We have then

Property 17.1: Lagrangian of diffeomorphism-invariant theories (Iyer & Wald 1994 [291])

Any diffeomorphism-invariant theory of gravity based on a Lagrangian involving only the metric tensor \mathbf{g} and a finite number of its derivatives (with respect to an arbitrary affine connection on \mathcal{M}) can be recast in terms of a Lagrangian involving only \mathbf{g} , its Riemann tensor and a finite number of symmetrized^a covariant derivatives of the latter:

$$\mathbf{L}(\mathbf{g}) = L(g_{\alpha\beta}, R_{\alpha\beta\gamma\delta}, \nabla_\lambda R_{\alpha\beta\gamma\delta}, \dots, \nabla_{(\lambda_1} \cdots \nabla_{\lambda_m)} R_{\alpha\beta\gamma\delta}). \quad (17.3)$$

^aThere is no loss of generality in considering only symmetrized derivatives, since any antisymmetrized combination can be reexpressed in term of the Riemann tensor itself and lower order derivatives.

The reader is referred to the original article [291] for the proof.

Example 1 (Lagrangian of general relativity): For general relativity, L is the Einstein-Hilbert Lagrangian:

$$L = (16\pi)^{-1} R = (16\pi)^{-1} g^{\mu\nu} R_{\mu\nu} = (16\pi)^{-1} g^{\mu\nu} R^\sigma_{\mu\sigma\nu} = (16\pi)^{-1} g^{\mu\nu} g^{\rho\sigma} R_{\rho\mu\sigma\nu}. \quad (17.4)$$

The last expression is clearly of the form (17.3), given that the components ($g^{\alpha\beta}$) are functions of the components ($g_{\alpha\beta}$) only.

17.2.2 Lagrangian variation and presymplectic potential form

A standard computation (see e.g. [137, 344]) leads to

Property 17.2: variation of the Lagrangian of a pure metric theory

Let \mathbf{L} be a Lagrangian n -form that depends only of the metric tensor \mathbf{g} and its derivatives, as in Eq. (17.3). The variation of \mathbf{L} resulting from a generic variation $\delta\mathbf{g}$ of \mathbf{g} is expressible as

$$\delta\mathbf{L} = E^{\mu\nu} \delta g_{\mu\nu} \epsilon + d\theta, \quad (17.5)$$

where $E^{\mu\nu}$ stands for the components of a symmetric type-(2, 0) tensor field \mathbf{E} , called the **Euler-Lagrange tensor**, and θ is a $(n - 1)$ -form, called the **presymplectic potential** ($n - 1$)-**form** (this name will be justified below). The n -form $d\theta$ is the exterior derivative of θ (cf. Sec. A.4.3). \mathbf{E} depends on \mathbf{g} but *not* on $\delta\mathbf{g}$, while θ depends on both \mathbf{g} and $\delta\mathbf{g}$, in a linear way in terms of $\delta\mathbf{g}$. All these dependencies are *local*, so if we want to stress them,

we shall write $\mathbf{E} = \mathbf{E}(\mathbf{g})$ and $\boldsymbol{\theta} = \boldsymbol{\theta}(\mathbf{g}, \delta\mathbf{g})$ according to the notation introduced on p. 654.

Remark 1: The term $d\boldsymbol{\theta}$ collects all the derivatives of $\delta\mathbf{g}$, which have been gathered by means of the Leibniz rule, so that the term $E^{\mu\nu}\delta g_{\mu\nu}\epsilon$ in Eq. (17.5) does not contain any derivative of $\delta\mathbf{g}$.

It must be stressed immediately that the presymplectic potential $(n - 1)$ -form $\boldsymbol{\theta}$ is not unique. First of all, since d is nilpotent [Eq. (A.92)], $\boldsymbol{\theta}$ in formula (17.5) can be substituted by $\boldsymbol{\theta}' = \boldsymbol{\theta} + d\mathbf{Y}$, where $\mathbf{Y} = \mathbf{Y}(\mathbf{g}, \delta\mathbf{g})$ is an arbitrary $(n - 2)$ -form. Secondly, for a given gravity theory, the Lagrangian n -form itself is not unique. Indeed the n -form $\mathbf{L}' = \mathbf{L} + d\boldsymbol{\mu}$, where $\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{g})$ is an arbitrary $(n - 1)$ -form with compact support, leads to the same action (17.1) by virtue of Stokes' theorem (A.94). The corresponding change in the presymplectic potential is $\boldsymbol{\theta} \rightarrow \boldsymbol{\theta} + \delta\boldsymbol{\mu}$. From these two facts, we conclude:

Property 17.3: ambiguities in the presymplectic potential form

For a given gravity theory, the presymplectic potential $(n - 1)$ -form $\boldsymbol{\theta}$ is not unique: it can be substituted by

$$\boldsymbol{\theta}' = \boldsymbol{\theta} + d\mathbf{Y} + \delta\boldsymbol{\mu}, \quad (17.6)$$

where (i) $\mathbf{Y} = \mathbf{Y}(\mathbf{g}, \delta\mathbf{g})$ is an arbitrary $(n - 2)$ -form that is a local function of \mathbf{g} and $\delta\mathbf{g}$, with a linear dependence on $\delta\mathbf{g}$ and (ii) $\delta\boldsymbol{\mu}$ is the variation triggered by $\delta\mathbf{g}$ of an arbitrary $(n - 1)$ -form $\boldsymbol{\mu} = \boldsymbol{\mu}(\mathbf{g})$ that is a local function of \mathbf{g} and has compact support. The $\boldsymbol{\mu}$ term in Eq. (17.6) arises from the allowed change

$$\mathbf{L}' = \mathbf{L} + d\boldsymbol{\mu} \quad (17.7)$$

in the Lagrangian n -form.

Example 2 (presymplectic potential form of general relativity): For general relativity, one has $\mathbf{L} = (16\pi)^{-1}R\epsilon$ [Eq. (17.4)]. Given a coordinate system (x^α) on \mathcal{M} , this can be rewritten as $\mathbf{L} = (16\pi)^{-1}R\sqrt{-g}\mathbf{e}$, where g is the determinant of \mathbf{g} with respect to (x^α) and \mathbf{e} is the n -form whose components with respect to (x^α) are $e_{\alpha_1\dots\alpha_n} = 1$ (resp. -1) if $(\alpha_1, \dots, \alpha_n)$ is an even (resp. odd) permutation of $(0, \dots, n)$ and 0 otherwise [cf. Eq. (A.52)]. One has then $\delta\mathbf{L} = (16\pi)^{-1}\delta(R\sqrt{-g})\mathbf{e}$, since obviously (x^α) , and thus \mathbf{e} , is not altered by the variation $\delta\mathbf{g}$. From the expression of $\delta(R\sqrt{-g})$ given by Eq. (16.18) and using $G_{\mu\nu}\delta g^{\mu\nu} = -G^{\mu\nu}\delta g_{\mu\nu}$, we get $16\pi\delta\mathbf{L} = -G^{\mu\nu}\delta g_{\mu\nu}\epsilon + \nabla_\mu H^\mu\epsilon$, where $G^{\mu\nu}$ is the double metric dual of the Einstein tensor and \mathbf{H} is the vector field defined in terms of $\delta\mathbf{g}$ by Eq. (16.13), where $h^\alpha_\beta = g^{\alpha\mu}\delta g_{\mu\beta}$ since $h_{\alpha\beta} := \delta g_{\alpha\beta}$. Now, the identity (A.93) yields $\nabla_\mu H^\mu\epsilon = \mathbf{d}(\mathbf{H}\cdot\epsilon)$, so that $16\pi\delta\mathbf{L} = -G^{\mu\nu}\delta g_{\mu\nu}\epsilon + \mathbf{d}(\mathbf{H}\cdot\epsilon)$. We conclude that the variation of the Einstein-Hilbert Lagrangian is indeed of the form (17.5) with

$$E^{\alpha\beta} = -(16\pi)^{-1}G^{\alpha\beta} \quad \text{and} \quad \boldsymbol{\theta} = (16\pi)^{-1}\mathbf{H}\cdot\epsilon = (16\pi)^{-1}\star\mathbf{H}, \quad (17.8)$$

where the last equality stems from the definition of the Hodge dual of a 1-form [Eq. (5.38) with $p = 1$]. In index notation, $\theta_{\alpha_1\dots\alpha_{n-1}} = (16\pi)^{-1}H^\mu\epsilon_{\mu\alpha_1\dots\alpha_{n-1}}$. Substituting Eq. (16.13) for H^μ , we get the explicit expression of $\boldsymbol{\theta}(\mathbf{g}, \delta\mathbf{g})$:

$$\theta_{\alpha_1\dots\alpha_{n-1}} = (16\pi)^{-1}g^{\mu\nu}g^{\rho\sigma}(\nabla_\sigma\delta g_{\nu\rho} - \nabla_\nu\delta g_{\rho\sigma})\epsilon_{\mu\alpha_1\dots\alpha_{n-1}}. \quad (17.9)$$

We check that $\boldsymbol{\theta}(\mathbf{g}, \delta\mathbf{g})$ is linear in $\delta\mathbf{g}$, as stated in Property 17.2.

Property 17.4: field equations from the least action principle

The action \mathcal{S} defined by (17.1) is extremal with respect to any metric variation $\delta\mathbf{g}$ that has compact support if, and only if,

$$E(\mathbf{g}) = 0. \quad (17.10)$$

This constitutes the so-called **Euler-Lagrange equations** for the considered gravity theory, also called the **field equations** or **equations of motion**.

Proof. Since $\delta\mathbf{g}$ is assumed to have compact support, let \mathcal{D} be a n -dimensional submanifold with boundary of \mathcal{M} such that $\delta\mathbf{g} = 0$ in the neighborhood of the boundary $\partial\mathcal{D}$ and outside \mathcal{D} . In view of Eqs. (17.1) and (17.5), we have then

$$\delta\mathcal{S} = \int_{\mathcal{M}} \delta\mathbf{L} = \int_{\mathcal{D}} E^{\mu\nu} \delta g_{\mu\nu} \epsilon + \int_{\mathcal{D}} d\boldsymbol{\theta} = \int_{\mathcal{D}} E^{\mu\nu} \delta g_{\mu\nu} \epsilon + \underbrace{\int_{\partial\mathcal{D}} \boldsymbol{\theta}}_0 = \int_{\mathcal{D}} E^{\mu\nu} \delta g_{\mu\nu} \epsilon,$$

where Stokes' theorem (A.94) has been used to let appear the integral of $\boldsymbol{\theta}$ on $\partial\mathcal{D}$, which has been set to zero due to $\boldsymbol{\theta}(\mathbf{g}, \delta\mathbf{g})|_{\partial\mathcal{D}} = 0$. This last property results from the linearity of $\boldsymbol{\theta}$ in $\delta\mathbf{g}$, given that $\delta\mathbf{g}$ and all its derivatives are vanishing on $\partial\mathcal{D}$. From the above expression of $\delta\mathcal{S}$ it is clear that having $\delta\mathcal{S} = 0$ for any variation $\delta\mathbf{g}$ is equivalent to Eq. (17.10). \square

Example 3 (field equations of general relativity): For general relativity, \mathbf{E} is given by Eq. (17.8). Hence the field equations (17.10) become $\mathbf{G} = 0$, which is equivalent to $\mathbf{R} = 0$ – the vacuum Einstein equation (1.44).

17.2.3 Noether charges related to diffeomorphism invariance

Any smooth vector field ξ on \mathcal{M} generates a 1-parameter family of infinitesimal diffeomorphisms, namely the flow maps along ξ : $\Phi_t : \mathcal{M} \rightarrow \mathcal{M}$, where t is an infinitesimal parameter. As described in Sec. A.4.2, each Φ_t displaces any point $p \in \mathcal{M}$ by the infinitesimal vector $t \xi|_p$. The flow map Φ_t is clearly a diffeomorphism, whose inverse is $\Phi_t^{-1} = \Phi_{-t}$. Let us consider the variation $\delta\mathbf{g}$ induced by Φ_t , or more precisely by Φ_t^{-1} in order to simplify expressions; indeed choosing Φ_t^{-1} makes the pullback $\Phi_t^*\mathbf{g}$ appear instead of $(\Phi_t^{-1})^*\mathbf{g}$, as discussed in Sec. 17.2.1. The variation of \mathbf{g} is then

$$\delta\mathbf{g} = \Phi_t^*\mathbf{g} - \mathbf{g}. \quad (17.11)$$

Since t is infinitesimal, the very definition of the Lie derivative [Eq. (A.82)] leads to $\Phi_t^*\mathbf{g} - \mathbf{g} = t\mathcal{L}_\xi\mathbf{g}$. It follows immediately that

$$\delta\mathbf{g} = t\mathcal{L}_\xi\mathbf{g}. \quad (17.12)$$

The variation of the Lagrangian n -form induced by $\delta\mathbf{g}$ is $\delta\mathbf{L} = \mathbf{L}(\Phi_t^*\mathbf{g}) - \mathbf{L}(\mathbf{g})$. Since the theory is assumed to be diffeomorphism-invariant, the covariance property (17.2) holds: $\mathbf{L}(\Phi_t^*\mathbf{g}) = \Phi_t^*\mathbf{L}(\mathbf{g})$. This yields $\delta\mathbf{L} = \Phi_t^*\mathbf{L}(\mathbf{g}) - \mathbf{L}(\mathbf{g})$. This is the same form as \mathbf{g} 's variation expressed in

Eq. (17.11) and, again, the definition (A.82) of the Lie derivative (which is applicable here since \mathbf{L} is a tensor field of type $(0, n)$) leads to

$$\delta \mathbf{L} = t \mathcal{L}_\xi \mathbf{L}. \quad (17.13)$$

Remark 2: The variation $\delta \mathbf{g}$ is often defined as a *variation rate* rather than a genuine variation, namely as the *derivative* of a 1-parameter family of metric tensors $\hat{\mathbf{g}}(t)$ such that $\hat{\mathbf{g}}(0) = \mathbf{g}$. Using the symbol $\hat{\delta}$ to distinguish from the variation considered here, the definition is $\hat{\delta} \mathbf{g} := d\hat{\mathbf{g}}/dt|_{t=0}$. Since the metric family induced by the flow maps along ξ is $\hat{\mathbf{g}}(t) := \Phi_t^* \mathbf{g}$, one gets $\hat{\delta} \mathbf{g} = \lim_{t \rightarrow 0} t^{-1} (\Phi_t^* \mathbf{g} - \mathbf{g})$. In view of the definition (A.82) of the Lie derivative, this yields

$$\hat{\delta} \mathbf{g} = \mathcal{L}_\xi \mathbf{g}. \quad (17.14)$$

Similarly, one defines $\hat{\delta} \mathbf{L} := d\mathbf{L}(\hat{\mathbf{g}}(t))/dt|_{t=0}$, which yields $\hat{\delta} \mathbf{L} = \mathcal{L}_\xi \mathbf{L}$. When comparing with Eqs. (17.12) and (17.13), we see that the derivative-based definition has the advantage of getting rid of the (not particularly significant) infinitesimal factor t .

If one plugs Eqs. (17.12) and (17.13) into the generic Lagrangian variation formula (17.5), one gets $t \mathcal{L}_\xi \mathbf{L} = t E^{\mu\nu} \mathcal{L}_\xi g_{\mu\nu} \epsilon + d(\theta(g, t \mathcal{L}_\xi g))$. Since θ is linear with respect to its second argument (cf. Property 17.2), one has $d(\theta(g, t \mathcal{L}_\xi g)) = t d(\theta(g, \mathcal{L}_\xi g))$, so that there comes

$$\mathcal{L}_\xi \mathbf{L} = E^{\mu\nu} \mathcal{L}_\xi g_{\mu\nu} \epsilon + d(\theta(g, \mathcal{L}_\xi g)).$$

Now, the Cartan identity (A.95) gives $\mathcal{L}_\xi \mathbf{L} = \xi \cdot d\mathbf{L} + d(\xi \cdot \mathbf{L}) = d(\xi \cdot \mathbf{L})$ since $d\mathbf{L} = 0$, as for any $(n+1)$ -form on a n -dimensional manifold. Accordingly, we may rewrite the above equation as

$$d(\theta(g, \mathcal{L}_\xi g) - \xi \cdot \mathbf{L}) = -E^{\mu\nu} \mathcal{L}_\xi g_{\mu\nu} \epsilon. \quad (17.15)$$

If the field equations are fulfilled, i.e. if $E = 0$ [Eq. (17.10)], the right-hand side of Eq. (17.15) is zero, which implies that the argument of the exterior derivative in the left-hand side is a closed $(n-1)$ -form. This defines a conserved current, which is nothing but the expression of Noether's theorem for the invariance of the Lagrangian by the 1-dimensional group of diffeomorphisms generated by the vector field ξ :

Property 17.5: Noether theorem for diffeomorphisms generated by a vector field

Given a smooth vector field ξ on \mathcal{M} , the **Noether current** $(n-1)$ -form of ξ is defined by [501, 291]

$$J(\xi) := \theta(g, \mathcal{L}_\xi g) - \xi \cdot \mathbf{L}, \quad (17.16)$$

where $\theta(g, \mathcal{L}_\xi g)$ is the $(n-1)$ -form obtained by substituting $\mathcal{L}_\xi g$ for δg in the presymplectic potential form $\theta(g, \delta g)$ (cf. Property 17.2) and $\xi \cdot \mathbf{L}$ is the $(n-1)$ -form obtained by setting the first argument of the n -form \mathbf{L} to ξ : $\xi \cdot \mathbf{L} := \mathbf{L}(\xi, \dots)$. $J(\xi)$ is a local function of ξ and its derivatives, and depends linearly on ξ . If the gravity theory is diffeomorphism-invariant, then the $(n-1)$ -form $J(\xi)$ is closed as soon as the equations of motion (17.10) are fulfilled or ξ is a Killing vector:

$$dJ(\xi) = 0. \quad (17.17)$$

The Noether current $(n - 1)$ -form is not unique, some ambiguity $\mathbf{J} \rightarrow \mathbf{J}'$ resulting from (17.6) and (17.7), with

$$\mathbf{J}'(\xi) = \mathbf{J}(\xi) + \mathbf{d} [\mathbf{Y}(g, \mathcal{L}_\xi g) + \xi \cdot \boldsymbol{\mu}(g)], \quad (17.18)$$

where $\mathbf{Y}(g, \mathcal{L}_\xi g)$ is any $(n - 2)$ -form that is linear in $\mathcal{L}_\xi g$, $\boldsymbol{\mu}(g)$ is any $(n - 1)$ -form with compact support and $\xi \cdot \boldsymbol{\mu}$ is the $(n - 2)$ -form $\boldsymbol{\mu}(\xi, \dots)$.

Proof. That $\mathbf{J}(\xi)$ is a linear function of ξ follows immediately from the linearity of $\theta(\cdot, \cdot)$ with respect to its second argument. Similarly the locality of $\mathbf{J}(\xi)$ with respect of ξ is inherited from that of $\theta(\cdot, \cdot)$. If the theory diffeomorphism-invariant, Eq. (17.15) holds. Given the definition of $\mathbf{J}(\xi)$, Eq. (17.15) is equivalent to $\mathbf{d}\mathbf{J}(\xi) = -E^{\mu\nu}\mathcal{L}_\xi g_{\mu\nu}\epsilon$, so that $\mathbf{d}\mathbf{J}(\xi) = 0$ if $E = 0$ is fulfilled or ξ is a Killing vector ($\mathcal{L}_\xi g = 0$). Finally, both $\mathbf{L}'(g)$ and $\mathbf{L}(g)$ in Eq. (17.7) must be covariant under diffeomorphisms. It follows that $\boldsymbol{\mu}(g)$ must be covariant as well. Then the variation of $\boldsymbol{\mu}$ resulting from a diffeomorphism Φ_t generated by ξ is $\delta\boldsymbol{\mu} = t\mathcal{L}_\xi\boldsymbol{\mu}$, by the very same argument that led to formula (17.13) for $\delta\mathbf{L}$. Accordingly, Eq. (17.6) yields $\theta'(g, \mathcal{L}_\xi g) = \theta(g, \mathcal{L}_\xi g) + \mathbf{d}\mathbf{Y}(g, \mathcal{L}_\xi g) + \mathcal{L}_\xi\boldsymbol{\mu}(g)$. By combining this identity with $\mathbf{L}' = \mathbf{L} + \mathbf{d}\boldsymbol{\mu}$ [Eq. (17.7)], the definition (17.16) applied to \mathbf{J}' leads to

$$\mathbf{J}'(\xi) = \mathbf{J}(\xi) + \mathbf{d}\mathbf{Y}(g, \mathcal{L}_\xi g) + \mathcal{L}_\xi\boldsymbol{\mu}(g) - \xi \cdot \mathbf{d}\boldsymbol{\mu}(g).$$

The Cartan identity (A.95) gives $\mathcal{L}_\xi\boldsymbol{\mu}(g) - \xi \cdot \mathbf{d}\boldsymbol{\mu}(g) = \mathbf{d}(\xi \cdot \boldsymbol{\mu}(g))$ and we obtain (17.18). \square

Remark 3: From the definition (17.16), it is clear that $\mathbf{J}(\xi)$ is a local function of g (and its derivatives), in addition to ξ . So, following the notation introduced in Sec. 17.2.1, we could have written $\mathbf{J}(g, \xi)$. However for the sake of brevity, we use $\mathbf{J}(\xi)$, leaving the dependency on g implicit.

Example 4 (Noether current form of general relativity): From expression (17.9) for θ in general relativity with $\delta g_{\mu\nu} = t\mathcal{L}_\xi g_{\mu\nu} = t(\nabla_\mu\xi_\nu + \nabla_\nu\xi_\mu)$ [Eq. (A.87)], we get

$$\begin{aligned} 16\pi\theta(g, \mathcal{L}_\xi g)_{\alpha_1 \dots \alpha_{n-1}} &= g^{\mu\nu} (\nabla_\sigma\nabla_\nu\xi^\sigma + \nabla_\sigma\nabla^\sigma\xi_\nu - 2\nabla_\nu\nabla_\rho\xi^\rho) \epsilon_{\mu\alpha_1 \dots \alpha_{n-1}} \\ &= g^{\mu\nu} (2R_{\nu\sigma}\xi^\sigma - \nabla_\sigma\nabla_\nu\xi^\sigma + \nabla_\sigma\nabla^\sigma\xi_\nu) \epsilon_{\mu\alpha_1 \dots \alpha_{n-1}} \\ &= (2R^\mu_\nu\xi^\nu + \nabla_\nu\nabla^\nu\xi^\mu - \nabla_\nu\nabla^\mu\xi^\nu) \epsilon_{\mu\alpha_1 \dots \alpha_{n-1}}, \end{aligned}$$

where the contracted Ricci identity (A.107) has been used to get the second line. Plugging this formula, as well as $\mathbf{L} = R\epsilon$ (cf. Example 1) into the definition (17.16) of $\mathbf{J}(\xi)$ yields

$$\mathbf{J}(\xi)_{\alpha_1 \dots \alpha_{n-1}} = (16\pi)^{-1} (2G^\mu_\nu\xi^\nu + \nabla_\nu\nabla^\nu\xi^\mu - \nabla_\nu\nabla^\mu\xi^\nu) \epsilon_{\mu\alpha_1 \dots \alpha_{n-1}}.$$

If the field equations are obeyed, then $\mathbf{G} = 0$ (cf. Example 3) and the above expression reduces to

$$\mathbf{J}(\xi)_{\alpha_1 \dots \alpha_{n-1}} = (16\pi)^{-1} (\nabla_\nu\nabla^\nu\xi^\mu - \nabla_\nu\nabla^\mu\xi^\nu) \epsilon_{\mu\alpha_1 \dots \alpha_{n-1}}. \quad (17.19)$$

We recognize in the term within parentheses the *Komar current* of ξ as defined by Eq. (5.50): $\mathcal{J}(\xi)^\mu := \nabla_\nu\nabla^\nu\xi^\mu - \nabla_\nu\nabla^\mu\xi^\nu$. Accordingly, we may write

$$\mathbf{J}(\xi) = (16\pi)^{-1} \mathcal{J}(\xi) \cdot \epsilon = (16\pi)^{-1} \star \underline{\mathcal{J}}(\xi), \quad (17.20)$$

where the last equality stems from Eq. (5.38) with $p = 1$. Hence, for general relativity, the Noether current $(n - 1)$ -form of ξ is nothing but the Hodge dual of the (metric dual of the) Komar current of ξ , up to the factor $(16\pi)^{-1}$. That $\mathbf{J}(\xi)$ is closed in this case follows readily from the identity (A.93): $16\pi\mathbf{d}\mathbf{J}(\xi) = \mathbf{d}(\mathcal{J}(\xi) \cdot \epsilon) = (\nabla \cdot \underline{\mathcal{J}}(\xi))\epsilon$, along with the conservation law (5.51): $\nabla \cdot \underline{\mathcal{J}}(\xi) = 0$.

Property 17.6: Noether potential form

Given a vector field ξ on \mathcal{M} and a diffeomorphism-invariant theory described by a Lagrangian n -form L , there exists a $(n-2)$ -form $Q(\xi)$ on \mathcal{M} , which is a linear local function of ξ , such that the Noether current $(n-1)$ -form of ξ is the exterior derivative of $Q(\xi)$:

$$J(\xi) = dQ(\xi). \quad (17.21)$$

$Q(\xi)$ is called the **Noether potential** $(n-2)$ -form of ξ . It is not unique, being subject to the ambiguity $Q \rightarrow Q'$, with

$$Q'(\xi) = Q(\xi) + Y(g, \mathcal{L}_\xi g) + \xi \cdot \mu(g) + dZ(g, \xi), \quad (17.22)$$

where $Y(g, \mathcal{L}_\xi g)$ is any $(n-2)$ -form that is linear in $\mathcal{L}_\xi g$, $\mu(g)$ is any $(n-1)$ -form with compact support and $Z(g, \xi)$ is any $(n-3)$ -form that is linear in ξ .

Proof. Since $J(\xi)$ is a closed differential form [Eq. (17.17)], the Poincaré lemma states that it is *locally* exact (on a contractible open set, cf. Sec. A.4.3). However, because $dJ(\xi) = 0$ holds for any vector field ξ , it can be shown that $J(\xi)$ is *globally* exact, i.e. takes the form (17.21) with $Q(\xi)$ being a local function of ξ [499]. The linearity of $Q(\xi)$ in terms of ξ arises from the linearity of $J(\xi)$ (cf. Property 17.5). Finally, the Y and μ ambiguities in (17.22) are inherited from the ambiguities of J [Eq. (17.18)], while the Z ambiguity stems from Eq. (17.21) defining Q up to the addition of a closed $(n-2)$ -form (cf. Lemma 1 in Ref. [499]). \square

Remark 4: As for $J(\xi)$ (cf. Remark 3 on p. 659), $Q(\xi)$ is a local function of g , in addition to ξ , although the dependency on g is not explicated in the notation $Q(\xi)$.

Remark 5: The $(n-2)$ -form $Q(\xi)$ has been introduced by Wald (1993) [501] under the name *Noether charge* $(n-2)$ -form. However, as argued by other authors [293, 254], the name *Noether potential* $(n-2)$ -form is more relevant, since $Q(\xi)$ is not conserved: a priori $dQ(\xi) \neq 0$ and Eq. (17.21) shows clearly that $Q(\xi)$ plays the role of a potential of a conserved current. Another name given to $Q(\xi)$ in the literature is the *Noether-Wald charge density* [245, 249].

Example 5 (Noether potential form of general relativity): For general relativity, $J(\xi)$ is given by Eq. (17.19), or Eq. (17.20) in terms of the Komar current of ξ . Let us show that $Q(\xi)$ is then nothing but the Hodge dual of the exterior derivative of the 1-form $\underline{\xi}$ (the metric dual to ξ), up to the factor $-(16\pi)^{-1}$:

$$Q(\xi) = -(16\pi)^{-1} \star (d\underline{\xi}). \quad (17.23)$$

Given the expression of the Hodge dual of a 2-form [Eq. (5.38) with $p = 2$], the identity $(d\underline{\xi})_{\alpha\beta} = \nabla_\alpha \xi_\beta - \nabla_\beta \xi_\alpha$ [Eq. (A.91b)] and the antisymmetry of ϵ , a formula equivalent to (17.23) is

$$Q(\xi)_{\alpha_1 \dots \alpha_{n-2}} = -(16\pi)^{-1} \nabla^\mu \xi^\nu \epsilon_{\mu\nu\alpha_1 \dots \alpha_{n-2}}. \quad (17.24)$$

To prove that $Q(\xi)$ as given by (17.23) obeys $dQ(\xi) = J(\xi)$ [Eq. (17.21)], let us establish the Hodge dual of this relation: $\star dQ(\xi) = \star J(\xi)$, which is an equality between 1-forms. From Eq. (17.23), we have $\star dQ(\xi) = -(16\pi)^{-1} \star d \star (d\underline{\xi})$. Now, the operator $\star d \star$ is, up to a sign, the *codifferential* δ , which

maps a p -form to a $(p - 1)$ -form and is defined by⁴ $\delta := (-1)^{n(p+1)+1} \star d\star$. Since $d\xi$ is a 2-form, we get $\star dQ(\xi) = (-1)^n (16\pi)^{-1} \delta(d\xi)$. From the expression of the codifferential in a coordinate chart (x^α) (cf. e.g. Eq. (14.37) in Ref. [463]), we may write, for any 2-form A :

$$(\delta A)^\alpha = \frac{1}{\sqrt{-g}} \frac{\partial}{\partial x^\mu} (\sqrt{-g} A^{\mu\alpha}) = -\nabla_\mu A^{\mu\alpha},$$

where the last equality follows from the identity (A.75). Applying the above formula to $A := d\xi$, we get $\delta(d\xi)^\alpha = -\nabla_\mu (\nabla^\alpha \xi^\mu - \nabla^\mu \xi^\alpha) = J(\xi)^\alpha$, i.e. $\delta(d\xi) = J(\xi)$, where $J(\xi)$ is the Komar current of ξ defined by Eq. (5.50). Hence $\star dQ(\xi) = (-1)^n (16\pi)^{-1} J(\xi)$. On the other hand, we have $\star J(\xi) = (16\pi)^{-1} \star \star J(\xi)$ from Eq. (17.20). Since $\star \star = (-1)^n \text{Id}$ for a 1-form (cf. e.g. formula (14.26) in Ref. [463]), we immediately get $\star J(\xi) = \star dQ(\xi)$, which is equivalent to Eq. (17.21).

Property 17.7: Noether charge of a spacelike hypersurface

Let us consider a diffeomorphism-invariant theory described by a Lagrangian n -form L . Given a vector field ξ on \mathcal{M} and a spacelike hypersurface Σ , the **Noether charge of Σ with respect to ξ** is

$$\mathcal{D}_{\Sigma,\xi} := \int_{\Sigma} J(\xi), \quad (17.25)$$

where $J(\xi)$ is the Noether current $(n - 1)$ -form of ξ (Property 17.5) and the integral is taken with respect to Σ endowed with the **future orientation**, i.e. the orientation defined by $n \cdot \epsilon$, where n is the future-directed timelike unit normal to Σ . The Noether charge is a priori not unique, since any change $J \rightarrow J'$ allowed by (17.18) leads to $\mathcal{D}_{\Sigma,\xi} \rightarrow \mathcal{D}'_{\Sigma,\xi}$ with

$$\mathcal{D}'_{\Sigma,\xi} = \mathcal{D}_{\Sigma,\xi} + \int_{\partial\Sigma} [Y(g, \mathcal{L}_\xi g) + \xi \cdot \mu(g)]. \quad (17.26)$$

However, if ξ is a Killing vector and (i) ξ is tangent to $\partial\Sigma$ or (ii) $\partial\Sigma$ is located outside the support of $\mu(g)$, the Noether charge is uniquely defined: $\mathcal{D}'_{\Sigma,\xi} = \mathcal{D}_{\Sigma,\xi}$.

Let Σ_1 and Σ_2 be two spacelike hypersurfaces, such that Σ_2 lies in the future of Σ_1 and both hypersurfaces are connected by a hypersurface \mathcal{W} , i.e. $\Sigma_1 \cup \Sigma_2 \cup \mathcal{W}$ is a closed hypersurface constituting the boundary of a n -dimensional manifold $\mathcal{D} \subset \mathcal{M}$ (cf. Fig. ??). Then the following conservation law holds:

$$\boxed{\mathcal{D}_{\Sigma_2,\xi} = \mathcal{D}_{\Sigma_1,\xi} + \int_{\mathcal{W}} J(\xi)}, \quad (17.27)$$

where the integral over \mathcal{W} is taken with respect to the *inward* orientation (indicated by the arrow), i.e. the orientation defined by $v \cdot \epsilon$, where v is any vector field pointing into \mathcal{D} .

Proof. Equation (17.26) readily follows from (17.18) and Stokes' theorem (A.94). If ξ is a Killing vector, then $\mathcal{L}_\xi g = 0$ by definition (cf. Property 3.6); this implies $Y(g, \mathcal{L}_\xi g) = 0$ in the

⁴See e.g. Eq. (14.35) of Straumann's book [463], whose sign convention is followed here. Since the same letter δ is used to denote the codifferential and the variation operator, we are using a boldface symbol to distinguish the former.

right-hand side of Eq. (17.26), given that \mathbf{Y} is linear with respect to $\mathcal{L}_\xi g$. Moreover, if ξ is tangent to $\partial\Sigma$ (case (i)), the $(n-2)$ -form $\xi \cdot \mu = \mu(\xi, \dots)$ restricted to vectors tangent to $\partial\Sigma$ is necessarily zero, for μ is a $(n-1)$ -form. Obviously, this holds as well in the case (ii). Hence $\mathcal{Q}'_{\Sigma,\xi} = \mathcal{Q}_{\Sigma,\xi}$. As for Eq. (17.27), given that $\partial\mathcal{D} = \Sigma_1 \cup \Sigma_2 \cup \mathcal{W}$, Stokes' theorem (A.94) and $dJ(\xi) = 0$ [Eq. (17.17)] lead to

$$\int_{\Sigma_1^\rightarrow} J(\xi) + \int_{\Sigma_2^\rightarrow} J(\xi) + \int_{\mathcal{W}^\rightarrow} J(\xi) = 0,$$

where the arrows mean that Σ_1 , Σ_2 and \mathcal{W} are oriented according to the outward convention (with respect to \mathcal{D}). Noticing that $\int_{\Sigma_1^\rightarrow} J(\xi) = -\mathcal{Q}_{\Sigma_1,\xi}$, $\int_{\Sigma_2^\rightarrow} J(\xi) = \mathcal{Q}_{\Sigma_2,\xi}$ and $\int_{\mathcal{W}^\rightarrow} J(\xi) = -\int_{\mathcal{W}^\leftarrow} J(\xi)$ yields (17.27). \square

The integral over \mathcal{W} in Eq. (17.27) can be interpreted as an “inward flux”, especially if \mathcal{W} is a timelike hypersurface (cf. the electromagnetic analogy detailed in Remark 7 below). If this inward flux vanishes, Eq. (17.27) expresses the invariance of the Noether charge with respect to the choice of the hypersurface (conservation during “time evolution”). This occurs if $J(\xi) = 0$ on \mathcal{W} or if \mathcal{W} is at some asymptotically flat end of \mathcal{M} with $J(\xi)$ decaying sufficiently fast so that its integral on \mathcal{W} is zero.

Property 17.8: Noether charge of a closed spacelike $(n-2)$ -surface

Let us consider a diffeomorphism-invariant theory described by a Lagrangian n -form \mathbf{L} and let assume that (\mathcal{M}, g) is asymptotically flat. Given a vector field ξ on \mathcal{M} and a closed spacelike $(n-2)$ -surface $\mathcal{S} \subset \mathcal{M}$, the **Noether charge of \mathcal{S} with respect to ξ** is

$$\mathcal{Q}_{\mathcal{S},\xi} := \int_{\mathcal{S}} Q(\xi), \quad (17.28)$$

where $Q(\xi)$ is the Noether potential $(n-2)$ -form of ξ (Property 17.6) and the orientation of \mathcal{S} for the above integral is the *future-outward* orientation defined by Eq. (5.39). The Noether charge is a priori not unique, since any change $Q \rightarrow Q'$ obeying (17.22) leads to $\mathcal{Q}_{\mathcal{S},\xi} \rightarrow \mathcal{Q}'_{\mathcal{S},\xi}$ with

$$\mathcal{Q}'_{\mathcal{S},\xi} = \mathcal{Q}_{\mathcal{S},\xi} + \int_{\mathcal{S}} [\mathbf{Y}(g, \mathcal{L}_\xi g) + \xi \cdot \mu(g)]. \quad (17.29)$$

However, if ξ is a Killing vector and (i) ξ is tangent to \mathcal{S} or (ii) \mathcal{S} is located outside the support of $\mu(g)$, the Noether charge is uniquely defined: $\mathcal{Q}'_{\mathcal{S},\xi} = \mathcal{Q}_{\mathcal{S},\xi}$.

Let Σ be a compact spacelike hypersurface bounded by two closed $(n-2)$ -surfaces: an “internal” one, \mathcal{S}_{int} , and an “external” one, \mathcal{S}_{ext} , the exterior direction being that of the asymptotic flat end of (\mathcal{M}, g) . Then

$$\boxed{\mathcal{Q}_{\mathcal{S}_{\text{ext}},\xi} = \mathcal{Q}_{\mathcal{S}_{\text{int}},\xi} + \mathcal{Q}_{\Sigma,\xi}}. \quad (17.30)$$

Proof. Equation (17.29) readily follows from (17.22) along with $\int_{\mathcal{S}} dZ = 0$ as a result of Stokes' theorem (A.94), given that \mathcal{S} is a closed manifold: $\partial\mathcal{S} = \emptyset$. The proof that $\mathcal{Q}'_{\mathcal{S},\xi} = \mathcal{Q}_{\mathcal{S},\xi}$ in the considered cases is identical to the proof of $\mathcal{Q}'_{\Sigma,\xi} = \mathcal{Q}_{\Sigma,\xi}$ in Property 17.7. Finally, to prove (17.30), let us substitute dQ for J [cf. Eq. (17.21)] in expression (17.25) of $\mathcal{Q}_{\Sigma,\xi}$. Applying Stokes' theorem (A.94) and using $\partial\Sigma = \mathcal{S}_{\text{int}} \cup \mathcal{S}_{\text{ext}}$, we get

$$\mathcal{Q}_{\Sigma,\xi} = \int_{\mathcal{S}_{\text{int}}^{\nearrow}} Q(\xi) + \int_{\mathcal{S}_{\text{ext}}^{\nearrow}} Q(\xi),$$

where the arrows indicate that \mathcal{S}_{int} and \mathcal{S}_{ext} are outward-oriented with respect to Σ (which is itself future-oriented). Noticing that $\int_{\mathcal{S}_{\text{int}}^{\nearrow}} Q(\xi) = -\mathcal{Q}_{\mathcal{S}_{\text{int}},\xi}$ and $\int_{\mathcal{S}_{\text{ext}}^{\nearrow}} Q(\xi) = \mathcal{Q}_{\mathcal{S}_{\text{ext}},\xi}$ yields (17.30). \square

Remark 6: We are using the same letter \mathcal{Q} to denote the Noether charge of a hypersurface (Property 17.7) and the Noether charge of a $(n-2)$ -surface (Property 17.8). The distinction between the two quantities occurs only through the subscript: $\mathcal{Q}_{\Sigma,\xi}$ versus $\mathcal{Q}_{\mathcal{S},\xi}$. This common notation is justified in so far as both quantities are strongly related by Eq. (17.30). In particular, if the hypersurface Σ has a single (external) boundary \mathcal{S} , i.e. $\mathcal{S}_{\text{int}} = \emptyset$, Eq. (17.30) reduces to $\mathcal{Q}_{\mathcal{S},\xi} = \mathcal{Q}_{\Sigma,\xi}$.

Remark 7: The name *charge* in the expression *Noether charge* comes of course from the analogy with electromagnetism: considering $n=4$ for simplicity, the analog of the Noether current 3-form J is the Hodge dual of the 1-form j , metric dual of the electric 4-current j : $J = \star j$, while the analog of the Noether potential 2-form Q is the Hodge dual of the electromagnetic field tensor F divided by μ_0 : $Q = \mu_0^{-1} \star F$. The relation $dQ = J$ [Eq. (17.21)] is then nothing but Maxwell equation (1.50): $d\star F = \mu_0 \star j$. Then $\mathcal{Q}_{\Sigma,\xi}$ is the total electric charge within the volume Σ (cf. e.g. Eq. (18.2) in Ref. [227]) and $\mathcal{Q}_{\mathcal{S},\xi}$ is μ_0^{-1} times the flux of the electromagnetic field through the surface \mathcal{S} . Accordingly, Eq. (17.27) expresses the *conservation of electric charge* and Eq. (17.30) with $\mathcal{S}_{\text{int}} = \emptyset$ is nothing but *Gauss's law* (cf. e.g. Eq. (18.40) in Ref. [227]).

From Example 5, we note that the $(n-2)$ -form $\star(d\xi)$ giving $Q(\xi)$ for general relativity [cf. Eq. (17.23)] is exactly the differential form that constitutes the integrand in formula (5.36) defining the Komar mass or in formula (5.63) defining the Komar angular momentum ($\xi := \eta$ in that case). In view of the definition (17.28) of the Noether charge, we conclude that

Property 17.9: Komar mass and angular momentum as Noether charges

In a stationary spacetime governed by general relativity, the Komar mass $M_{\mathcal{S}}$ over a $(n-2)$ -dimensional closed spacelike surface \mathcal{S} [Eq. (5.36)] is $(n-2)/(n-3)$ times the Noether charge^a of \mathcal{S} with respect to the stationary Killing vector ξ :

$$M_{\mathcal{S}} = \frac{n-2}{n-3} \mathcal{Q}_{\mathcal{S},\xi}. \quad (17.31)$$

Similarly, in an axisymmetric spacetime governed by general relativity, the Komar angular momentum $J_{\mathcal{S}}$ over \mathcal{S} [Eq. (5.63)] is minus the Noether charge of \mathcal{S} with respect to the axisymmetric Killing vector η :

$$J_{\mathcal{S}} = -\mathcal{Q}_{\mathcal{S},\eta}. \quad (17.32)$$

^aSince ξ is a Killing vector, the Noether charge is uniquely defined in the present case, provided \mathcal{S} is chosen outside the support of μ if one allows for a Lagrangian change of the type (17.7), cf. Property 17.8.

17.3 Wald entropy

17.3.1 Wald entropy as a Noether charge

Having introduced the concept of Noether charge related to diffeomorphism invariance, we are in position to define the Wald entropy of a stationary black hole.

Property 17.10: Wald entropy as a Noether charge

Let (\mathcal{M}, g) be a stationary spacetime of dimension $n \geq 3$ governed by a diffeomorphism-invariant theory and containing a black hole, whose event horizon \mathcal{H} is part of a bifurcate Killing horizon (cf. Sec. 3.4). Let χ be the Killing vector generating \mathcal{H} and κ the corresponding surface gravity: $\nabla_\chi \chi \stackrel{\mathcal{H}}{=} \kappa \chi$ [Eq. (3.29)]. By virtue of the zeroth law for bifurcate Killing horizons (Property 3.16), κ is constant over \mathcal{H} and $\kappa \neq 0$. Let \mathcal{S} be a complete cross-section of \mathcal{H} . The **Wald entropy** of the black hole is then defined as κ^{-1} times the Noether charge of \mathcal{S} with respect to χ , up to the constant $2\pi k_B/\hbar$:

$$S_W := \frac{2\pi k_B}{\hbar\kappa} Q_{\mathcal{S}, \chi} = \frac{2\pi k_B}{\hbar\kappa} \int_{\mathcal{S}} Q(\chi). \quad (17.33)$$

Here $Q(\chi)$ is the Noether potential $(n-2)$ -form of χ , as defined in Property 17.6. S_W is independent of the choice of the cross-section \mathcal{S} of \mathcal{H} . Moreover, the Noether charge $Q_{\mathcal{S}, \chi}$ does not suffer from the ambiguity (17.29), so that S_W is uniquely defined.

Proof. Let \mathcal{S}' be another complete cross-section of \mathcal{H} lying entirely in the future of \mathcal{S} . Then $\mathcal{S} \cup \mathcal{S}'$ constitutes the boundary of a portion of \mathcal{H} , which we shall denote by $\Delta\mathcal{H}$. Stokes' theorem (A.94) and the identity $dQ(\chi) = J(\chi)$ [Eq. (17.21)] give

$$\int_{\mathcal{S} \nearrow} Q(\chi) + \int_{\mathcal{S}' \nearrow} Q(\chi) = \int_{\Delta\mathcal{H}} J(\chi), \quad (17.34)$$

where the arrows indicate that \mathcal{S} and \mathcal{S}' are outward-oriented with respect to $\Delta\mathcal{H}$, the orientation of the latter being set by a future-directed null vector field k transverse to \mathcal{H} . From the definition (17.16), $J(\chi) := \theta(g, \mathcal{L}_\chi g) - \chi \cdot L = -\chi \cdot L$, since $\mathcal{L}_\chi g = 0$ (χ Killing) along with θ being linear in $\mathcal{L}_\chi g$ imply $\theta(g, \mathcal{L}_\chi g) = 0$. Now, when restricted to $\Delta\mathcal{H}$ (i.e. applied to $(n-1)$ -tuples of vectors tangent to $\Delta\mathcal{H}$, or equivalently, considered as the pullback $\iota^*(\chi \cdot L)$ of $\chi \cdot L$ by the inclusion map $\iota : \Delta\mathcal{H} \rightarrow \mathcal{M}$, cf. Sec. A.2.8), $\chi \cdot L$ is a vanishing $(n-1)$ -form, since $\chi \cdot L := L(\chi, \dots)$, χ is tangent to $\Delta\mathcal{H}$ and L is a n -form. Accordingly, the integral of $J(\chi) = -\chi \cdot L$ in the right-hand side of Eq. (17.34) is zero. Given that $\int_{\mathcal{S} \nearrow} Q(\chi) = -Q_{\mathcal{S}, \chi}$ and $\int_{\mathcal{S}' \nearrow} Q(\chi) = Q_{\mathcal{S}', \chi}$, it follows that $Q_{\mathcal{S}', \chi} = Q_{\mathcal{S}, \chi}$, which

shows that S_W does not depend on the choice of \mathcal{S} . Regarding the ambiguity (17.29), the term $Y(g, \mathcal{L}_\chi g)$ does not contribute since $Y(g, \mathcal{L}_\chi g)$ is linear in $\mathcal{L}_\chi g$ (Property 17.3), so that $\mathcal{L}_\chi g = 0$ (χ Killing) implies $Y(g, \mathcal{L}_\chi g) = 0$. The term $\chi \cdot \mu(g)$ does not contribute either because, taking advantage of the independence of $\mathcal{Q}_{\mathcal{S}, \chi}$ from \mathcal{S} , we are free to choose \mathcal{S} as the bifurcation surface, where $\chi = 0$ (Property 3.12). \square

If the diffeomorphism-invariant theory is general relativity, the Wald entropy reduces to the Bekenstein-Hawking entropy defined in Sec. 16.6.1:

Property 17.11: Wald entropy for general relativity

For general relativity, the Wald entropy of a stationary black hole coincides with its Bekenstein-Hawking entropy, as given by formula (16.73):

$$S_W = S_{\text{BH}} = k_B \frac{A}{4\hbar} \quad (\text{general relativity}), \quad (17.35)$$

A being the black hole area.

Proof. For general relativity, the Noether potential $(n-2)$ -form is given by Eq. (17.23): $Q(\chi) = -(16\pi)^{-1} \star (d\underline{\chi})$. We have then

$$16\pi \mathcal{Q}_{\mathcal{S}, \chi} = - \int_{\mathcal{S}} \star (d\underline{\chi}) = -\frac{1}{2} \int_{\mathcal{S}} (d\underline{\chi})_{\mu\nu} dS^{\mu\nu} = - \int_{\mathcal{S}} \nabla_\mu \chi_\nu dS^{\mu\nu},$$

where the second equality follows from Lemma 5.10 and the third one from expression (A.91b) of the exterior derivative and the antisymmetric character of $dS^{\mu\nu}$. Now, from Eq. (5.87), $\nabla_\mu \chi_\nu dS^{\mu\nu} = -2\kappa dS$, where $dS := \sqrt{q} d^{n-2}x$ is the area element of \mathcal{S} . Since κ is constant, there comes $\mathcal{Q}_{\mathcal{S}, \chi} = \kappa A / (8\pi)$. In view of the definition (17.33) of S_W , we get (17.35). \square

17.3.2 Geometric expression of the Wald entropy

The Bekenstein-Hawking entropy does not involve directly the horizon Killing vector χ , but a geometrical quantity related to a cross-section \mathcal{S} of the horizon, namely the area A of \mathcal{S} , which is independent of \mathcal{S} for a Killing horizon. Similarly, one can recast the Wald entropy in a form that gets rid of χ and κ , referring only to geometrical quantities related to the bifurcation surface (and to the Lagrangian scalar):

Property 17.12: geometric expression of the Wald entropy (Iyer & Wald 1994 [291])

Let us consider a stationary black hole in a diffeomorphism-invariant theory, whose Lagrangian scalar L is recast as (17.3). Assuming that the black hole event horizon is part of a bifurcate Killing horizon, with bifurcation surface $\hat{\mathcal{S}}$, the Wald entropy is expressible as

$$S_W = -2\pi \frac{k_B}{\hbar} \int_{\hat{\mathcal{S}}} \frac{\delta L}{\delta R_{\mu\nu\rho\sigma}} \hat{\epsilon}_{\rho\sigma}^\perp dS_{\mu\nu}, \quad (17.36)$$

where

- $\delta L / \delta R_{\alpha\beta\gamma\delta}$ stands for the type-(4, 0) tensor field that is the **Euler-Lagrange derivative** of L with respect to $R_{\alpha\beta\gamma\delta}$, if the latter were considered as a field independent of $g_{\alpha\beta}$ in expression (17.3) of L :

$$\frac{\delta L}{\delta R_{\alpha\beta\gamma\delta}} := \frac{\partial L}{\partial R_{\alpha\beta\gamma\delta}} - \nabla_\mu \frac{\partial L}{\partial \nabla_\mu R_{\alpha\beta\gamma\delta}} + \nabla_{(\mu} \nabla_{\nu)} \frac{\partial L}{\partial \nabla_{(\mu} \nabla_{\nu)} R_{\alpha\beta\gamma\delta}} - \dots \quad (17.37)$$

- $dS_{\alpha\beta}$ is the metric dual of the area element normal bivector $dS^{\alpha\beta}$ to $\hat{\mathcal{S}}$, as defined by Eq. (5.42);
- $\hat{\epsilon}^\perp$ is the **binormal 2-form** to $\hat{\mathcal{S}}$, i.e. for each point $p \in \hat{\mathcal{S}}$, $\hat{\epsilon}_{\alpha\beta}^\perp|_p$ is the volume 2-form of the metric induced by g in the vector space $T_p^\perp \hat{\mathcal{S}}$ normal to $\hat{\mathcal{S}}$ equipped with the future-outward orientation^a; $\hat{\epsilon}^\perp$ can be expressed as the exterior product $\underline{s} \wedge \underline{n}$ of any pair of vector fields $(\underline{n}, \underline{s})$ along $\hat{\mathcal{S}}$ forming an orthonormal basis of $T_p^\perp \hat{\mathcal{S}}$ at each $p \in \hat{\mathcal{S}}$, with \underline{n} future-directed unit timelike and \underline{s} outward-directed unit spacelike:

$$\hat{\epsilon}^\perp = \underline{s} \wedge \underline{n} \iff \hat{\epsilon}_{\alpha\beta}^\perp = s_\alpha n_\beta - n_\alpha s_\beta. \quad (17.38)$$

Note that Eqs. (5.42) and (17.38) yield the following relation between the area element normal bivector dS and the binormal 2-form $\hat{\epsilon}^\perp$:

$$dS_{\alpha\beta} = \hat{\epsilon}_{\alpha\beta}^\perp dS, \quad (17.39)$$

where $dS = \sqrt{q} d^{n-2}x$ is the area element of $\hat{\mathcal{S}}$.

^aSee item (iv) on p. 134 for details.

Partial proof. $Q(\chi)$ has a linear dependence on χ (cf. Property 17.6). It may contain high order covariant derivatives of χ but since the latter is a Killing vector, one may use iteratively Kostant formula (3.83) to express all derivatives of order higher than one in terms of χ and $\nabla\chi$, so that $Q(\chi)$ is expressible as

$$Q(\chi)_{\alpha_1 \dots \alpha_{n-2}} = \chi^\mu W_{\mu\alpha_1 \dots \alpha_{n-2}} + \nabla_\mu \chi_\nu X^{\mu\nu}_{\alpha_1 \dots \alpha_{n-2}},$$

where $W = W(g)$ and $X = X(g)$ are tensor fields of type $(0, n-1)$ and $(2, n-2)$ respectively, which are antisymmetric with respect to their last $n-2$ arguments. For instance, in the case of general relativity, we read on (17.24) that $W = 0$ and $X^{\mu\nu}_{\alpha_1 \dots \alpha_{n-2}} = -(16\pi)^{-1} \epsilon^{\mu\nu}_{\alpha_1 \dots \alpha_{n-2}}$. Let us evaluate the Wald entropy S_W by selecting the bifurcation surface $\hat{\mathcal{S}}$ for \mathcal{S} in formula (17.33). Given that $\chi = 0$ on $\hat{\mathcal{S}}$ (Property 3.12), the term in W does not contribute to S_W . Furthermore, the following identity holds on $\hat{\mathcal{S}}$:

$$\nabla_\alpha \chi_\beta \stackrel{\hat{\mathcal{S}}}{=} \kappa \hat{\epsilon}_{\alpha\beta}^\perp. \quad (17.40)$$

Indeed, since $\chi = 0$ on $\hat{\mathcal{S}}$, we have, for any vector v tangent to $\hat{\mathcal{S}}$, $v^\mu \nabla_\mu \chi_\alpha = 0$. This shows that, at each point $p \in \hat{\mathcal{S}}$, the 2-form $\nabla_\alpha \chi_\beta = \nabla_{[\alpha} \chi_{\beta]}$ leaves in the vector space $T_p^\perp \hat{\mathcal{S}}$ normal to $\hat{\mathcal{S}}$, as the 2-form $\hat{\mathcal{S}}\epsilon_{\alpha\beta}^\perp$. Since $T_p^\perp \hat{\mathcal{S}}$ is 2-dimensional, these 2-forms are necessarily proportional: $\nabla_\alpha \chi_\beta = \lambda \hat{\mathcal{S}}\epsilon_{\alpha\beta}^\perp$ for some scalar field λ on \mathcal{S} . By combining Eqs. (5.86) and (17.39), we get $\hat{\mathcal{S}}\epsilon_{\alpha\beta}^\perp = \chi_\alpha k_\beta - k_\alpha \chi_\beta$, where k is the future-directed null vector of $T_p^\perp \hat{\mathcal{S}}$ such that $k \cdot \chi = -1$. Then, Eq. (3.29) yields $\kappa \chi_\alpha = \chi^\mu \nabla_\mu \chi_\alpha = \lambda(\chi^\mu \chi_\mu k_\alpha - \chi^\mu k_\mu \chi_\alpha) = \lambda(0 + \chi_\alpha)$, from which $\lambda = \kappa$, thereby establishing (17.40). We have then $Q(\chi)_{\alpha_1 \dots \alpha_{n-2}} \stackrel{\hat{\mathcal{S}}}{=} \kappa \hat{\mathcal{S}}\epsilon_{\rho\sigma}^\perp X^{\rho\sigma}_{\alpha_1 \dots \alpha_{n-2}}$. Then Eq. (17.33) yields

$$S_W = 2\pi \frac{k_B}{\hbar} \int_{\hat{\mathcal{S}}} \hat{\mathcal{S}}\epsilon_{\rho\sigma}^\perp X^{\rho\sigma}_{\alpha_1 \dots \alpha_{n-2}}. \quad (17.41)$$

This expression of S_W does no longer involve χ , nor κ , but only the geometrical quantity $\hat{\mathcal{S}}\epsilon^\perp$ defined on $\hat{\mathcal{S}}$ and X , which depends on the Lagrangian n -form L . To go further, we must invoke a result of Wald and Iyer (cf. Eqs (31) and (52) of Ref. [291]), namely

$$X^{\rho\sigma}_{\alpha_1 \dots \alpha_{n-2}} = -\frac{\delta L}{\delta R_{\mu\nu\rho\sigma}} \epsilon_{\mu\nu\alpha_1 \dots \alpha_{n-2}}. \quad (17.42)$$

Accordingly, given the definition (5.38) of the Hodge dual, Eq. (17.41) becomes

$$S_W = -4\pi \frac{k_B}{\hbar} \int_{\hat{\mathcal{S}}} \star A, \quad \text{with} \quad A_{\alpha\beta} := g_{\alpha\mu} g_{\beta\nu} \frac{\delta L}{\delta R_{\mu\nu\rho\sigma}} \hat{\mathcal{S}}\epsilon_{\rho\sigma}^\perp.$$

Lemma 5.10 leads then to (17.36). \square

Let us stress that the integral in expression (17.36) of the Wald entropy has to be taken on the bifurcation surface $\hat{\mathcal{S}}$, while in the Noether charge expression (17.33), the integral can be taken on any complete cross-section \mathcal{S} of the event horizon.

Example 6 (Wald entropy from formula (17.36) for general relativity): For general relativity, $L = (16\pi)^{-1} g^{\mu\nu} g^{\rho\sigma} R_{\rho\mu\sigma\nu}$ [Eq. (17.4)], so that Eq. (17.37) results in

$$\frac{\delta L}{\delta R_{\alpha\beta\gamma\delta}} = \frac{\partial L}{\partial R_{\alpha\beta\gamma\delta}} = \frac{1}{16\pi} g^{\alpha\gamma} g^{\beta\delta}.$$

Accordingly, formula (17.36), combined with Eq. (17.39), yields

$$\frac{\hbar}{k_B} S_W = -\frac{1}{8} \int_{\hat{\mathcal{S}}} g^{\mu\rho} g^{\nu\sigma} \hat{\mathcal{S}}\epsilon_{\rho\sigma}^\perp \hat{\mathcal{S}}\epsilon_{\mu\nu}^\perp dS = -\frac{1}{8} \int_{\hat{\mathcal{S}}} \underbrace{(\hat{\mathcal{S}}\epsilon^\perp)^{\mu\nu} \hat{\mathcal{S}}\epsilon_{\mu\nu}^\perp}_{-2} dS = \frac{1}{4} \int_{\hat{\mathcal{S}}} dS = \frac{A}{4},$$

where A is the area of $\hat{\mathcal{S}}$ and $(\hat{\mathcal{S}}\epsilon^\perp)^{\mu\nu} \hat{\mathcal{S}}\epsilon_{\mu\nu}^\perp = -2$ follows from Eq. (17.38) along with $n_\mu n^\mu = -1$, $n_\mu s^\mu = 0$ and $s_\mu s^\mu = 1$. Hence we recover formula (17.35) – the Bekenstein-Hawking entropy.

Of course, being defined by a formula that reduces to the Bekenstein-Hawking entropy for general relativity (Property 17.11) is not sufficient by itself to assert that the Wald entropy is the correct concept of black hole entropy for all theories of gravity. This holds because the Wald entropy naturally enters in the generalization of the first law to all diffeomorphism-invariant theories that admit a Hamiltonian formulation, which we will discuss in Sec. 17.5, after having introduced the covariant phase space formalism.

17.4 Covariant phase space formalism

The extension of the first law beyond general relativity is naturally achieved within a Hamiltonian framework. Indeed, such a framework provides the concept of total mass M — the variation of which constitutes the left-hand side of the first law — as the *canonical energy* of the system (value of the Hamiltonian), thereby generalizing the concept of Kormar or ADM masses beyond general relativity. The relevant Hamiltonian framework is the co-called *covariant phase space* approach, which is adapted to field theories, preserving general covariance. On the contrary, the standard approach relies on the choice of canonical coordinates and of a time function t , with respect to which the evolution is governed by Hamilton's equations.

17.4.1 Hamiltonian formalism without canonical coordinates

In standard mechanics, a system with N degrees of freedom is described by some configuration coordinates $(q^i)_{1 \leq i \leq N}$ and a Lagrangian $L = L(q^i, \dot{q}^i)$ (with $\dot{q}^i := dq^i/dt$). The system's Hamiltonian is then $H(q^i, p_i) = \sum_{i=1}^N p_i \dot{q}^i - L$, where p_i 's are the conjugate momenta: $p_i := \partial L / \partial \dot{q}^i$, and the system's motion is governed by **Hamilton's equations**: $\dot{q}^i = \partial H / \partial p_i$ and $\dot{p}_i = -\partial H / \partial q^i$, or equivalently,

$$\frac{dq^i}{dt} = \{q^i, H\} \quad \text{and} \quad \frac{dp_i}{dt} = \{p_i, H\}, \quad 1 \leq i \leq N, \quad (17.43)$$

where $\{., .\}$ stands for the **Poisson bracket**:

$$\{f, g\} := \sum_{i=1}^N \left(\frac{\partial f}{\partial q^i} \frac{\partial g}{\partial p_i} - \frac{\partial g}{\partial q^i} \frac{\partial f}{\partial p_i} \right). \quad (17.44)$$

More generally, for any scalar function $f = f(q^i, p_i)$, one has

$$\frac{df}{dt} = \{f, H\}. \quad (17.45)$$

The coordinates $(y^a) := (q^1, \dots, q^N, p_1, \dots, p_N)$ span a smooth manifold \mathcal{F} of dimension $2N$ called the system's **phase space**. The Poisson bracket can be rewritten covariantly, i.e. in a form independent of the coordinates (y^a) , by introducing the following 2-form on \mathcal{F} :

$$\Omega := \sum_{i=1}^N dp_i \wedge dq^i, \quad (17.46)$$

where d stands for the exterior derivative⁵ on \mathcal{F} , so that dp_i (resp. dq^i) is the differential 1-form of p_i (resp. q^i) considered as a scalar field on \mathcal{F} . Given that $q^i = y^i$ and $p_i = y^{i+N}$, the

⁵On phase space manifolds, we are not using boldface notations for operators and tensor fields, keeping them for the spacetime manifold only. Hence the exterior derivative on \mathcal{F} is denoted by d , and not \mathbf{d} , as defined in Sec. A.4.3, the 2-form (17.46) is denoted Ω and not $\mathbf{\Omega}$, etc.

components of Ω with respect to the coordinates (y^a) are $\Omega_{ab} = \sum_{i=1}^N (\delta_a^{i+N} \delta_b^i - \delta_a^i \delta_b^{i+N}) = \delta_{a,b+N} - \delta_{b,a+N}$. Hence, we may write, using block matrix notation,

$$\Omega_{ab} = \begin{pmatrix} 0 & -I_N \\ I_N & 0 \end{pmatrix}, \quad (17.47)$$

where a and b are respectively the row and column indices and I_N stands for the identity matrix of size N . Any coordinate system on \mathcal{F} in which the 2-form Ω is expressible as (17.46), or equivalently, in which the components of Ω are the matrix (17.47), is called a system of **canonical coordinates**. The matrix (Ω_{ab}) is clearly invertible, its inverse being its opposite:

$$(\Omega^{-1})^{ab} = \begin{pmatrix} 0 & I_N \\ -I_N & 0 \end{pmatrix}. \quad (17.48)$$

This implies that the 2-form Ω is non-degenerate and allows one to define an antisymmetric type-(2, 0) tensor field on \mathcal{F} , called the **Poisson bivector**, by

$$\Omega^{-1} := (\Omega^{-1})^{ab} \partial_a \otimes \partial_b. \quad (17.49)$$

As a type-(2, 0) tensor, Ω^{-1} acts bilinearly on 1-forms (cf. Sec. A.2.5) and one may define the **Poisson bracket** of two scalar fields f and g on \mathcal{F} by letting Ω^{-1} act on the pair of 1-forms (df, dg) :

$$\{f, g\} := \Omega^{-1}(df, dg). \quad (17.50)$$

Given the components (17.48) of Ω^{-1} in canonical coordinates, it is easy to check that this definition reduces to (17.44) for these type of coordinates.

Because $dd = 0$ (nilpotence of the exterior derivative, cf. Eq. (A.92)), it follows immediately from Eq. (17.46) that the 2-form Ω is closed: $d\Omega = 0$. This is also obvious from the coordinate expression (A.90c) of the exterior derivative, given that the components Ω_{ab} are constant in canonical coordinates [Eq. (17.47)]. Generically, a closed non-degenerate 2-form Ω on a smooth manifold \mathcal{F} is called a **symplectic form** and the pair (\mathcal{F}, Ω) is called a **symplectic manifold** (see e.g. Chap. 22 of Lee's textbook [342]). Not only Ω is closed, but it is exact. Indeed, it follows from expression (17.46) that

$$\Omega = d\Theta, \quad (17.51)$$

where Θ is the 1-form defined by

$$\Theta := \sum_{i=1}^N p_i dq^i. \quad (17.52)$$

Because of (17.51), the 1-form Θ is called the **symplectic potential**. It is also called the **canonical 1-form** on \mathcal{F} , because its expression (17.52) is independent of the choice of the canonical coordinates (q^i, p_i) .

To any scalar field f on \mathcal{F} , one associates a unique vector field X_f on \mathcal{F} via the Poisson bivector Ω^{-1} :

$$X_f := \Omega^{-1}(., df). \quad (17.53)$$

Specifically, for any 1-form σ on \mathcal{F} , one has

$$\langle \sigma, X_f \rangle = X_f(\sigma) = \Omega^{-1}(\sigma, df). \quad (17.54)$$

It follows that the components of X_f with respect to any coordinate system (y^a) are

$$X_f^a = (\Omega^{-1})^{ab} \frac{\partial f}{\partial y^b}. \quad (17.55)$$

In particular, for canonical coordinates $(y^a) = (q^i, p_i)$, Eq. (17.48) leads to

$$X_f = \frac{\partial f}{\partial p_i} \partial_{q^i} - \frac{\partial f}{\partial q^i} \partial_{p_i}. \quad (17.56)$$

As a vector field, X_f is a derivative operator on scalar fields [cf. Eq. (A.10)], which we may express in terms of the Poisson bracket by combining Eqs. (17.53) and (17.50):

$$X_f = \{., f\}. \quad (17.57)$$

The above identity means that for any scalar field g , $X_f(g) = \{g, f\}$, with $X_f(g) = X_f^a \partial_a g$.

Formula (17.53) can be inverted to

$$df = \Omega(., X_f). \quad (17.58)$$

Proof. $\Omega(., X_f)$ is a 1-form whose components are $\Omega_{ab} X_f^b$. Thanks to Eq. (17.55), we get $\Omega_{ab} X_f^b = \Omega_{ab} (\Omega^{-1})^{bc} \partial_c f = \delta_a^c \partial_c f = \partial_a f$, which gives (17.58). \square

Remark 1: The vector field X_f and the 1-form df can be considered as being dual to each other through the symplectic form Ω in the very same way as vector fields and 1-forms are dual through the metric tensor g on pseudo-Riemannian manifolds: compare formula (17.58) with (A.44) (rewritten as $\underline{u} = g(., u)$) and formula (17.55) with (A.47). These dualities occur because both Ω and g are *non-degenerate* bilinear forms.

The Poisson bracket (17.50) can be expressed in terms of the symplectic form Ω acting of the pair of vector fields (X_f, X_g) :

$$\{f, g\} = -\Omega(X_f, X_g). \quad (17.59)$$

Proof. Starting from Eq. (17.50) and using Eq. (17.58), we have $\{f, g\} = (\Omega^{-1})^{ab} \partial_a f \partial_b g = (\Omega^{-1})^{ab} \Omega_{ac} X_f^c \Omega_{bd} X_g^d = \delta_a^c \Omega_{ac} X_f^c X_g^d = \Omega_{dc} X_f^c X_g^d = \Omega(X_g, X_f) = -\Omega(X_f, X_g)$. \square

Remark 2: Notice the minus sign in Eq. (17.59), as compared to Eq. (17.50); of course, since Ω is antisymmetric, Eq. (17.59) can be recast as $\{f, g\} = \Omega(X_g, X_f)$.

By combining Eqs. (17.58) and (17.59) there comes $\langle df, X_g \rangle = \Omega(X_g, X_f) = \{f, g\}$. Hence, we get various equivalent expressions for the Poisson bracket:

$$\{f, g\} := \Omega^{-1}(df, dg) = \Omega(X_g, X_f) = \langle df, X_g \rangle = X_g(f). \quad (17.60)$$

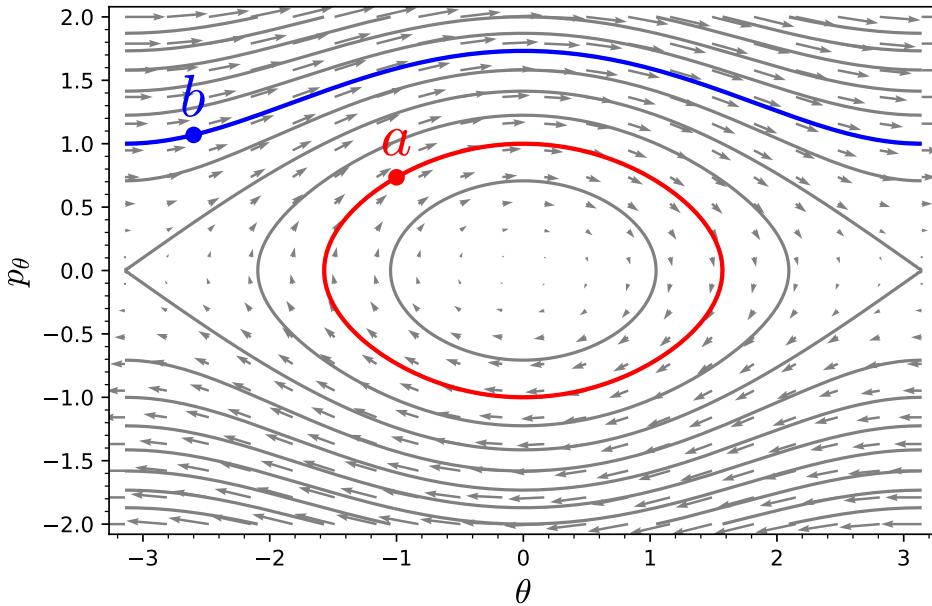


Figure 17.1: Phase space \mathcal{F} of a simple gravity pendulum, depicted in terms of the canonical coordinates (θ, p_θ) (cf. Example 7). The vertical lines $\theta = -\pi$ and $\theta = \pi$ must be identified, since $\mathcal{F} = \mathbb{S}^1 \times \mathbb{R}$. The grey arrows show the Hamiltonian vector field X_H and the grey lines are some of its integral curves. Two of them are highlighted: the red one corresponds to oscillating solutions (θ having a limited range and p_θ taking positive and negative values), while the blue one corresponds to solutions rotating indefinitely in the anticlockwise direction ($p_\theta > 0$). The points a and b each represent the initial data of a given solution in these families; these points can also be viewed as defining the solution itself (covariant phase space approach).

The **Hamiltonian vector field** is the vector field on \mathcal{F} associated to the Hamiltonian differential dH by symplectic duality, i.e. the vector field X_H . It obeys Eq. (17.58):

$$dH = \Omega(., X_H). \quad (17.61)$$

Remark 3: In the mathematical literature (e.g. [342]), the vector field X_f associated to any scalar function f on \mathcal{F} is called the *Hamiltonian vector field of f* . This reflects the fact that any scalar function f can be considered as the Hamiltonian of some theory, even if the latter is unphysical.

Example 7 (phase space of a pendulum): A simple gravity pendulum is a system with $N = 1$, the unique degree of freedom being the angle $\theta \in (-\pi, \pi]$ with respect to the vertical direction. For a unit mass, the Lagrangian is $L(\theta, \dot{\theta}) = (\ell\dot{\theta})^2/2 + g\ell \cos \theta$, where ℓ is the length of the (massless) rod and g the magnitude of the vertical gravity field. The momentum conjugate to θ is $p_\theta = \partial L / \partial \dot{\theta} = \ell^2 \dot{\theta}$. Since the latter can take any real value, the phase space \mathcal{F} is the 2-dimensional infinite cylinder $\mathbb{S}^1 \times \mathbb{R}$, spanned by the canonical coordinates (θ, p_θ) . The Hamiltonian is deduced from L by the Legendre transform $H = p_\theta \dot{\theta} - L$, yielding $H(\theta, p_\theta) = p_\theta^2/(2\ell^2) - g\ell \cos \theta$. The Hamiltonian vector field is obtained via Eq. (17.56): $X_H = \ell^{-2} p_\theta \partial_\theta - g\ell \sin \theta \partial_{p_\theta}$. It is depicted in Fig. 17.1.

Equation (17.57) with $f = H$ reads $\{., H\} = X_H$. Hence, for any scalar field f on \mathcal{F} , $\{f, H\} = X_H(f)$ and the evolution law (17.45) can be rewritten as

$$\boxed{\frac{df}{dt} = X_H(f)}. \quad (17.62)$$

In particular, Hamilton's equations (17.43) become

$$\frac{dq^i}{dt} = X_H(q^i) \quad \text{and} \quad \frac{dp_i}{dt} = X_H(p_i), \quad 1 \leq i \leq N. \quad (17.63)$$

It follows that any solution $(q^i(t), p_i(t))$ to the equations of motion is an integral curve of the Hamiltonian vector field in the phase space (cf. Fig. 7). H is constant along such a curve, since, from Eq. (17.61), $X_H(H) = \langle dH, X_H \rangle = \Omega(X_H, X_H) = 0$.

In view of Eq. (17.62), one may consider the vector field X_H as defining the “time evolution” without having to refer explicitly to the time coordinate t . Of course, we recover from Eq. (17.62) that $dH/dt = 0$ since $X_H(H) = 0$. For tensor fields on \mathcal{F} of valence higher than a scalar field, one defines the “time evolution” as the *Lie derivative* \mathcal{L}_{X_H} along the Hamiltonian vector field X_H . This is quite natural since the vanishing of $\mathcal{L}_{X_H} T$ means that the tensor field T is transported identically to itself along the flow defined by X_H (cf. Sec. A.4.2). In particular, this encompasses Eq. (17.62), since $\mathcal{L}_{X_H} f = X_H(f)$ for any scalar field f . An important feature of the evolution thus defined is that the symplectic form is conserved:

$$\boxed{\mathcal{L}_{X_H} \Omega = 0}. \quad (17.64)$$

Proof. The proof relies on the fact that Ω is closed, i.e. $d\Omega = 0$. Indeed, using successively the Cartan identity (A.95), Eq. (17.61) and the nilpotence property (A.92), we get

$$\mathcal{L}_{X_H} \Omega = X_H \cdot \underbrace{d\Omega}_0 + d(\underbrace{X_H \cdot \Omega}_{\Omega(X_H,.)}) = -d(\underbrace{\Omega(., X_H)}_{dH}) = -ddH = 0.$$

□

Let us summarize the above results as follows:

Property 17.13: Hamiltonian system

A **Hamiltonian system** is a triplet (\mathcal{F}, Ω, H) where \mathcal{F} is a smooth manifold, called the **phase space** of the system, Ω is a **symplectic form** on \mathcal{F} , i.e. a closed non-degenerate 2-form, and H is a smooth function $\mathcal{F} \rightarrow \mathbb{R}$ called the **Hamiltonian**. The differential dH of H is a 1-form on \mathcal{F} , whose symplectic dual defines the **Hamiltonian vector field**, namely the unique vector field X_H on \mathcal{F} such that

$$\boxed{dH = \Omega(., X_H)}. \quad (17.65)$$

By construction, X_H preserves the symplectic form: $\mathcal{L}_{X_H} \Omega = 0$ [Eq. (17.64)]. The evolution of the underlying physical system is entirely determined by the Hamiltonian vector field: any solution to the equations of motion is an integral curve of X_H . A smooth function

$f : \mathcal{F} \rightarrow \mathbb{R}$ is called a **conserved quantity** iff f is constant on every integral curve of X_H , i.e. iff $X_H(f) = 0$. In view of Eq. (17.60), this is equivalent to $\{f, H\} = 0$. A conserved quantity is H itself. Its value for a given solution to the equations of motion is called the **energy** of the solution. An **equilibrium point** is a point of \mathcal{F} where $X_H = 0$, or equivalently, given Eq. (17.65), a point where $dH = 0$.

Example 8 (solutions for the gravity pendulum): For the gravity pendulum considered in Example 7, the integral curves of X_H are depicted in Fig. 17.1. Two of them are represented by the red and the blue curves. It is clear from the figure that X_H vanishes at two points: $(\theta, p_\theta) = (0, 0)$ and $(\theta, p_\theta) = (\pi, 0)$, which are thus two equilibrium points (respectively stable and unstable).

Remark 4: The above formulation of Hamiltonian systems does not refer to any system of canonical coordinates (q^i, p_i) on the phase space \mathcal{F} and, more generally, to any kind of coordinates. It does not even assume that \mathcal{F} is the cotangent bundle $T^*\mathcal{C}$ of some “configuration space” \mathcal{C} , while the cotangent bundle structure is the standard arena for the Hamiltonian mechanics with a finite number of degrees of freedom [1].

17.4.2 Covariant phase space

Despite Remark 4, the Hamiltonian formulation sketched in Property 17.13 is not fully covariant. Indeed, via the Hamiltonian vector field X_H , it describes the evolution of the system with respect to some “privileged time” t [cf. Eq. (17.62)]. To treat covariantly relativistic systems, it would be desirable to get rid of the choice of a time function. This can be achieved as follows.

There is one, and exactly one, integral curve of X_H through each point⁶ of \mathcal{F} . It follows that a solution is entirely determined by a single point $a \in \mathcal{F}$, which can be thought of as representing the state of the system at some initial instant $t = t_0$, i.e. the coordinates of a are $(q^i(t_0), p_i(t_0))$ (cf. Fig. 17.1). If the initial value problem is well posed, there is a one-to-one correspondence between a point of \mathcal{F} and a solution. Consequently, one may view \mathcal{F} as the *space of solutions* to the equations of motion. This is the essence of the so-called *covariant phase space* approach:

Property 17.14: covariant phase space

Let (\mathcal{F}, Ω, H) be a Hamiltonian system whose equations of motion constitute a well posed initial value problem. Each solution to the equations of motion is a curve $S : \mathbb{R} \rightarrow \mathcal{F}$, $t \mapsto S(t)$, which admits the Hamiltonian vector field X_H as tangent vector. The set \mathcal{F}_{sol} of all solutions can be given the structure of a smooth manifold diffeomorphic to \mathcal{F} , and for each $t_0 \in \mathbb{R}$, an explicit diffeomorphism is the “initial data” map

$$\begin{aligned}\Phi_{t_0} : \quad \mathcal{F}_{\text{sol}} &\longrightarrow \mathcal{F} \\ S &\longmapsto S(t_0).\end{aligned}\tag{17.66}$$

⁶At an equilibrium point, the integral curve is reduced to the point itself.

Moreover, the pullback of the symplectic form Ω by Φ_{t_0} ,

$$\Omega_{\text{sol}} := \Phi_{t_0}^* \Omega, \quad (17.67)$$

defines a symplectic form on \mathcal{F}_{sol} , which is actually independent of t_0 . This makes $(\mathcal{F}_{\text{sol}}, \Omega_{\text{sol}})$ a well-defined symplectic manifold, called the **covariant phase space** of the system. Infinitesimal displacement vectors on the manifold \mathcal{F}_{sol} correspond to infinitesimal variations δS between solutions, so that Eq. (17.67) is equivalent to

$$\Omega_{\text{sol}}(\delta_1 S, \delta_2 S) := \Omega \left(\overrightarrow{S(t_0)S_1(t_0)}, \overrightarrow{S(t_0)S_2(t_0)} \right), \quad (17.68)$$

where $S, S_1 = S + \delta_1 S$ and $S_2 = S + \delta_2 S$ stand for three nearby solutions.

Similarly, the pullback of the Hamiltonian H by Φ_{t_0} , i.e. $H_{\text{sol}} := H \circ \Phi_{t_0}$, or equivalently,

$$H_{\text{sol}}(S) := H(S(t_0)), \quad (17.69)$$

defines a scalar field H_{sol} on \mathcal{F}_{sol} , which is independent of the choice of t_0 .

Proof. The equivalence between Eqs. (17.67) and (17.68) follows readily from the definition of the pullback map given in Sec. A.2.8. The independence of $\Omega_{\text{sol}}(\delta_1 S, \delta_2 S)$ from the choice of t_0 follows from the constancy of Ω along the Hamiltonian flow, i.e. Eq. (17.64). Indeed $\Omega_{\text{sol}}(\delta_1 S, \delta_2 S)$, as defined by Eq. (17.68), is independent of t_0 iff $df/dt = 0$, where $f(t) := \Omega(Y_1(t), Y_2(t))$ and we have introduced the separation vectors $Y_1(t) := \overrightarrow{S(t)S_1(t)}$ and $Y_2(t) := \overrightarrow{S(t)S_2(t)}$. Now, from Eq. (17.62) and the Leibniz rule,

$$\frac{df}{dt} = \mathcal{L}_{X_H} f = \mathcal{L}_{X_H} [\Omega(Y_1, Y_2)] = \underbrace{\mathcal{L}_{X_H} \Omega}_0 (Y_1, Y_2) + \Omega(\underbrace{\mathcal{L}_{X_H} Y_1}_0, Y_2) + \Omega(Y_1, \underbrace{\mathcal{L}_{X_H} Y_2}_0) = 0,$$

where $\mathcal{L}_{X_H} \Omega = 0$ is the invariance property (17.64) and $\mathcal{L}_{X_H} Y_1 = 0$ and $\mathcal{L}_{X_H} Y_2 = 0$ follows from the very definition of a Lie derivative (cf. Sec. A.4.2), the separation vectors Y_1 and Y_2 being naturally Lie-dragged by X_H . Finally, the independence of H_{sol} from the choice of t_0 is immediate from Eq. (17.69), which shows that $H_{\text{sol}}(S)$ is the energy of the solution S (cf. Property 17.13). \square

The triplet $(\mathcal{F}_{\text{sol}}, \Omega_{\text{sol}}, H_{\text{sol}})$ clearly constitutes a Hamiltonian system, as defined in Property 17.13, but without reference to any “time evolution”, given that each point $S \in \mathcal{F}_{\text{sol}}$ is a fully evolved solution by itself. The fully covariant picture is thus achieved. The Hamiltonian vector field $X_{H_{\text{sol}}}$ is a vector field on \mathcal{F}_{sol} , which generates 1-dimensional families of solutions. Two solutions S_1 and S_2 that lie on the same integral curve of $X_{H_{\text{sol}}}$ are then solutions to the original equations of motion that only differ by a time translation, i.e. $S_2(t) = S_1(t + t_0)$ for some $t_0 \in \mathbb{R}$. In particular S_1 and S_2 have the same energy.

17.4.3 Covariant phase space for gravity theories

In order to apply the covariant phase space approach to a gravity theory governed by a Lagrangian of the type (17.3), let us consider the space \mathcal{F}_{sol} of solutions \mathbf{g} to the field equations

(17.10). Since no confusion may arise, for simplicity, we shall drop the index “sol” and denote this space simply by⁷ \mathcal{F} . A difference with what precedes is that we are dealing with a field theory, which, by essence, has an infinite number of degrees of freedom. In other words, the space of solutions \mathcal{F} is an *infinite-dimensional manifold*. More precisely, it can be given the structure of an infinite-dimensional **Banach manifold** [344, 254], i.e. a topological space that locally resembles⁸ a Banach space of infinite dimension, in the same way as a n -dimensional manifold locally resembles the vector space \mathbb{R}^n . Since any point of \mathcal{F} is a metric field \mathbf{g} on \mathcal{M} , a variation $\delta\mathbf{g}$ can be viewed as an infinitesimal displacement vector between two neighboring points of the manifold \mathcal{F} , similar to the displacement vector $d\mathbf{x}$ on \mathcal{M} introduced in Secs. 1.2.1 and A.2.3. As in the finite dimensional case (cf. Sec. A.2.3), (non-infinitesimal) vectors on \mathcal{F} are defined⁹ as tangent vectors along *curves of* \mathcal{F} , the latter being the images of smooth maps $P : I \subset \mathbb{R} \rightarrow \mathcal{F}, \lambda \mapsto P(\lambda) =: \mathbf{g}(\lambda)$. Here $\mathbf{g}(\lambda)$ stands for a 1-parameter family of metric fields on \mathcal{M} , this family constituting a parametrization of a given curve \mathcal{L} of \mathcal{F} . Any vector X at a point $\mathbf{g}_0 \in \mathcal{F}$ can then be written as in Eq. (A.14):

$$X = \frac{\delta \mathbf{g}}{\delta \lambda}, \quad (17.70)$$

where $\delta\mathbf{g}$ is the infinitesimal displacement vector connecting $\mathbf{g}_0 =: \mathbf{g}(\lambda)$ and $\mathbf{g}(\lambda + \delta\lambda)$ along the curve \mathcal{L} to which X is tangent to; λ is then the parameter of \mathcal{L} associated to X and we denote by $\delta\lambda$ its small increment, instead of $d\lambda$ as in Sec. A.2.3.

Property 17.15: space of asymptotically flat solutions

In what follows, we assume that the n -dimensional spacetime manifold \mathcal{M} is diffeomorphic to $\mathbb{R} \times \hat{\Sigma}$, where $\hat{\Sigma}$ is an orientable manifold of dimension $(n - 1)$ with an **asymptotically Euclidean end**, i.e. there exists a compact subset K of $\hat{\Sigma}$ such that $\hat{\Sigma}_{\text{ext}} := \hat{\Sigma} \setminus K$ is diffeomorphic to $\mathbb{R}^{n-1} \setminus \mathbb{B}^{n-1}$, where \mathbb{B}^{n-1} is the unit ball of \mathbb{R}^{n-1} . Furthermore, given two points p and p' in \mathcal{M} , we shall say that p' lies in the **background future** of p iff for any orientation-preserving diffeomorphism $\Psi : \mathcal{M} \rightarrow \mathbb{R} \times \hat{\Sigma}, \Psi(p') = (t', q')$ and $\Psi(p) = (t, q)$ with $t' > t$.

We shall restrict the space \mathcal{F} of solutions \mathbf{g} of the considered gravity theory to those that are asymptotically flat, with the asymptotically flat end corresponding to $\hat{\Sigma}_{\text{ext}}$. The precise definition of **asymptotically flat** depends on the gravity theory and should be such that some integrals indicated below converge and that the asymptotic limit of some surface integrals is zero.

We are going to see that, from the action (17.1), we can endow \mathcal{F} with a (pre)symplectic form Ω . Let Σ and Σ' be two non-intersecting hypersurfaces of \mathcal{M} , both diffeomorphic to $\hat{\Sigma}$ and such that Σ' lies in the background future of Σ . Let Σ_r and Σ'_r be compact subparts of Σ and

⁷Be aware that in some part of the literature [24, 344, 506], \mathcal{F} denotes instead the space of all fields \mathbf{g} that are required to only obey some asymptotic flatness conditions, but not necessarily the field equations; the subspace of solutions to the field equations – our \mathcal{F} – is then denoted by $\bar{\mathcal{F}}$ [344, 506] or Γ [24].

⁸Cf. Lang’s book [335] for a precise definition of a Banach manifold.

⁹Notice that the definition (A.8) of a tangent vector does not require the manifold to be of finite dimension.

Σ' respectively, extending in the asymptotic flat end and corresponding to balls of Euclidean radius r in $\hat{\Sigma}_{\text{ext}}$. Let \mathcal{W}_r be the hypercylinder connecting Σ_r to Σ'_r (cf. Fig. ??) and \mathcal{D} be the part of \mathcal{M} bounded by Σ_r , Σ'_r and \mathcal{W}_r . Let us then consider the action (17.1) restricted to \mathcal{D} :

$$\mathcal{S}_{\mathcal{D}}[\mathbf{g}] := \int_{\mathcal{D}} L(\mathbf{g}). \quad (17.71)$$

We are using the *functional* notation $\mathcal{S}_{\mathcal{D}}[\mathbf{g}]$, but, when restricted to solutions \mathbf{g} of the field equations, $\mathcal{S}_{\mathcal{D}}$ is merely a scalar field on the manifold \mathcal{F} , i.e. a *function* $\mathcal{F} \rightarrow \mathbb{R}$. Accordingly, the variation $\delta \mathcal{S}_{\mathcal{D}}$ of $\mathcal{S}_{\mathcal{D}}$ due to a small displacement $\delta \mathbf{g}$ on \mathcal{F} can be viewed as the action of the differential 1-form $d\mathcal{S}_{\mathcal{D}}$ on the infinitesimal vector $\delta \mathbf{g}$: $\delta \mathcal{S}_{\mathcal{D}} = \langle d\mathcal{S}_{\mathcal{D}}, \delta \mathbf{g} \rangle$ [cf. Eq. (A.21)]. Now, from the definition (17.71), we get $\delta \mathcal{S}_{\mathcal{D}} = \int_{\mathcal{D}} \delta L(\mathbf{g})$, with $\delta L(\mathbf{g})$ given by Eq. (17.5), in which \mathbf{E} is set to zero for elements of \mathcal{F} are solutions of the field equations. Hence $\langle d\mathcal{S}_{\mathcal{D}}, \delta \mathbf{g} \rangle = \int_{\mathcal{D}} d\theta(\mathbf{g}, \delta \mathbf{g})$. By linearity¹⁰, we get for any vector $X = \delta \mathbf{g}/\delta \lambda$ at a point $\mathbf{g} \in \mathcal{F}$,

$$\langle d\mathcal{S}_{\mathcal{D}}, X \rangle = \int_{\mathcal{D}} d\theta\left(\mathbf{g}, \frac{\delta \mathbf{g}}{\delta \lambda}\right).$$

Thanks to Stokes' theorem (A.94), we rewrite the right-hand side as a surface integral, taking into account that $\partial \mathcal{D} = \Sigma'_r \cup \Sigma_r \cup \mathcal{W}_r$:

$$\langle d\mathcal{S}_{\mathcal{D}}, X \rangle = \int_{\Sigma'_r \nearrow} \theta\left(\mathbf{g}, \frac{\delta \mathbf{g}}{\delta \lambda}\right) + \int_{\Sigma_r \nearrow} \theta\left(\mathbf{g}, \frac{\delta \mathbf{g}}{\delta \lambda}\right) + \int_{\mathcal{W}_r \nearrow} \theta\left(\mathbf{g}, \frac{\delta \mathbf{g}}{\delta \lambda}\right), \quad (17.72)$$

where the arrows indicate the outward orientation of the various hypersurfaces with respect to \mathcal{D} . Since Eq. (17.72) is valid for any vector X on \mathcal{F} , we may rewrite it as an identity between 1-forms:

$$d\mathcal{S}_{\mathcal{D}} = \Theta_{\Sigma'_r} - \Theta_{\Sigma_r} + \Upsilon_{\Sigma_r, \Sigma'_r}. \quad (17.73)$$

Here Θ_{Σ_r} stands the 1-form on \mathcal{F} defined by the following action on a vector X in the tangent space $T_g \mathcal{F}$:

$$\langle \Theta_{\Sigma_r}, X \rangle := \int_{\Sigma_r} \theta\left(\mathbf{g}, \frac{\delta \mathbf{g}}{\delta \lambda}\right), \quad (17.74)$$

where $\delta \mathbf{g}/\delta \lambda$ is the relative variation of \mathbf{g} defining the vector X by Eq. (17.70) and the orientation of Σ_r is that provided by any background-future directed vector transverse to Σ_r . Note that Θ_{Σ_r} is well defined, i.e. the integral in Eq. (17.74) is finite, because Σ_r is compact and θ is smooth; moreover, Θ_{Σ_r} is linear for $\theta(\cdot, \cdot)$ is linear in its second argument. The same definition holds for $\Theta_{\Sigma'_r}$. Similarly, $\Upsilon_{\Sigma_r, \Sigma'_r}$ stands for the 1-form on \mathcal{F} defined by

$$\langle \Upsilon_{\Sigma_r, \Sigma'_r}, X \rangle := \int_{\mathcal{W}_r \nearrow} \theta\left(\mathbf{g}, \frac{\delta \mathbf{g}}{\delta \lambda}\right), \quad (17.75)$$

with \mathcal{W}_r defined as above. Note that the minus sign in front of Θ_{Σ_r} in Eq. (17.73) occurs because of the change of orientation of Σ_r between Eqs. (17.72) and (17.73).

In an attempt to endow \mathcal{F} with a symplectic form, thereby turning it into a phase space, let us consider the 2-form Ω that is the limit $\Sigma_r \rightarrow \Sigma$ (i.e. the limit $r \rightarrow +\infty$) of the exterior derivative of Θ_{Σ_r} . This defines a closed 2-form, which turns out to be independent of the choice of Σ :

¹⁰Recall that $\theta(\mathbf{g}, \delta \mathbf{g})$ is linear with respect to $\delta \mathbf{g}$, cf. Property 17.2.

Property 17.16: presymplectic form in the space of solutions of a gravity theory

Let \mathcal{F} be the (Banach) manifold of solutions g to the field equations that are asymptotically flat in the asymptotically Euclidean end of the n -dimensional manifold $\mathcal{M} \sim \mathbb{R} \times \hat{\Sigma}$. Let Σ be a hypersurface of \mathcal{M} diffeomorphic to $\hat{\Sigma}$ and Σ_r a compact subpart of Σ corresponding to a ball of Euclidean radius r in $\hat{\Sigma}_{\text{ext}}$. Given the 1-form Θ_{Σ_r} defined on \mathcal{F} by Eq. (17.74), the asymptotically flatness conditions of the gravity theory must insure that the limit

$$\Omega := \lim_{r \rightarrow +\infty} d\Theta_{\Sigma_r}, \quad (17.76)$$

is well defined, i.e. yields a finite 2-form on \mathcal{F} . This 2-form is closed: $d\Omega = 0$. One then says that Ω is a **presymplectic form**^a on \mathcal{F} . Moreover, Ω does not depend on the choice of the hypersurface Σ . The pair (\mathcal{F}, Ω) is called the **prephase space** of the considered gravity theory (and asymptotic flatness conditions).

^aRecall that a *symplectic form* is a 2-form that is closed *and* non-degenerate; as we shall discuss below, Ω is actually degenerate.

Proof. For each value of r , $d\Theta_{\Sigma_r}$ is a 2-form on \mathcal{F} , so if the limit (17.76) exists, it clearly defines a 2-form Ω on \mathcal{F} . Let us evaluate its exterior derivative, $d\Omega$, by means of formula (A.88) with $p = 2$: for any 3-tuple (X_1, X_2, X_3) of vector fields on \mathcal{F} , we have $d\Omega(X_1, X_2, X_3) = X_1(\Omega(X_2, X_3)) - X_2(\Omega(X_1, X_3)) + X_3(\Omega(X_1, X_2)) - \Omega([X_1, X_2], X_3) + \Omega([X_1, X_3], X_2) - \Omega([X_2, X_3], X_1)$. Now

$$X_1(\Omega(X_2, X_3)) = X_1 \left(\lim_{r \rightarrow +\infty} d\Theta_{\Sigma_r}(X_2, X_3) \right) = \lim_{r \rightarrow +\infty} X_1(d\Theta_{\Sigma_r}(X_2, X_3)),$$

with similar formulas for $X_2(\Omega(X_1, X_3))$ and $X_3(\Omega(X_1, X_2))$. There comes then $d\Omega(X_1, X_2, X_3) = \lim_{r \rightarrow +\infty} dd\Theta_{\Sigma_r}(X_1, X_2, X_3)$. But $dd\Theta_{\Sigma_r} = 0$ (nilpotence of the exterior derivative, cf. Eq. (A.92)). We conclude that $d\Omega = 0$. As for the independence of Ω from Σ , it arises from the following identity obtained by taking the exterior derive of Eq. (17.73) and using $ddS_{\mathcal{D}} = 0$ in the left-hand side:

$$0 = d\Theta_{\Sigma'_r} - d\Theta_{\Sigma_r} + d\Upsilon_{\Sigma_r, \Sigma'_r}.$$

Taking the limit $r \rightarrow +\infty$ yields $\Omega_{\Sigma'} - \Omega_{\Sigma} + \lim_{r \rightarrow +\infty} d\Upsilon_{\Sigma_r, \Sigma'_r} = 0$. The asymptotic flatness conditions ensure that $\lim_{r \rightarrow +\infty} d\Upsilon_{\Sigma_r, \Sigma'_r} = 0$ (Lemma 17.18 below), leaving us with $\Omega_{\Sigma'} = \Omega_{\Sigma}$. \square

Remark 5: Instead of the limit (17.76), it could have seemed simpler to define the presymplectic form as $\Omega := d\Theta_{\Sigma}$, where Θ_{Σ} is defined as in Eq. (17.74) but with the integral extended to the whole of Σ . However, as noticed in Ref. [24] (p. 429), standard asymptotic flatness conditions do not guarantee that this integral is convergent; hence the 1-form Θ_{Σ} would be ill-defined.

Property 17.17: presymplectic current $(n - 1)$ -form

The action of the presymplectic form Ω on a pair of commuting vector fields (X_1, X_2)

of \mathcal{F} is expressible as follows. According to Eq. (17.70), each vector field describes the variation of \mathbf{g} along some curves of \mathcal{F} . Since they commute, the vector fields (X_1, X_2) can be associated to a 2-parameter family $\mathbf{g}(\lambda_1, \lambda_2)$ of metric tensors, spanning a 2-surface $\mathcal{S} \subset \mathcal{F}$, so that X_1 (resp. X_2) corresponds to variations $\delta\mathbf{g}$ along curves $\lambda_2 = \text{const}$ (resp. $\lambda_1 = \text{const}$) in the 2-surface \mathcal{S} , i.e. to partial derivatives:

$$X_1 = \left. \frac{\delta \mathbf{g}}{\delta \lambda_1} \right|_{\lambda_2=\text{const}} =: \frac{\partial \mathbf{g}}{\partial \lambda_1} \quad \text{and} \quad X_2 = \left. \frac{\delta \mathbf{g}}{\delta \lambda_2} \right|_{\lambda_1=\text{const}} =: \frac{\partial \mathbf{g}}{\partial \lambda_2}. \quad (17.77)$$

We have then

$$\Omega(X_1, X_2) = \int_{\Sigma} \omega\left(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_1}, \frac{\partial \mathbf{g}}{\partial \lambda_2}\right), \quad (17.78)$$

where ω is the $(n - 1)$ -form on \mathcal{M} defined from the presymplectic potential $(n - 1)$ -form θ of the considered gravity theory (cf. Property 17.2) by

$$\omega\left(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_1}, \frac{\partial \mathbf{g}}{\partial \lambda_2}\right) := \frac{\partial}{\partial \lambda_1} \left(\theta\left(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_2}\right) \right) - \frac{\partial}{\partial \lambda_2} \left(\theta\left(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_1}\right) \right). \quad (17.79)$$

ω is called the **presymplectic current** $(n - 1)$ -form [506]^a. It has a local dependence on the metric tensor \mathbf{g} and on the derivatives $\partial\mathbf{g}/\partial\lambda_1$ and $\partial\mathbf{g}/\partial\lambda_2$ associated to the 2-parameter family $\mathbf{g}(\lambda_1, \lambda_2)$, and is linear with respect to these derivatives.

^aIt is called the *symplectic current* $(n - 1)$ -form in articles by Wald and Iyer [501, 291] preceding [506], but, as discussed below, *presymplectic* is a better qualifier than *symplectic*.

Proof. Let us compute $d\Theta_{\Sigma_r}(X_1, X_2)$ via the exterior derivative formula (A.89), noticing that (i) this formula does not involve the manifold dimensionality (contrary to (A.90b), which applies to finite dimensions only) and (ii) the last term in the right-hand side of (A.89) vanishes since $[X_1, X_2] = 0$. We get

$$\begin{aligned} d\Theta_{\Sigma_r}(X_1, X_2) &= X_1(\langle \Theta_{\Sigma_r}, X_2 \rangle) - X_2(\langle \Theta_{\Sigma_r}, X_1 \rangle) \\ &= \frac{\partial}{\partial \lambda_1} \langle \Theta_{\Sigma_r}, X_2 \rangle - \frac{\partial}{\partial \lambda_2} \langle \Theta_{\Sigma_r}, X_1 \rangle \\ &= \frac{\partial}{\partial \lambda_1} \int_{\Sigma_r} \theta\left(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_2}\right) - \frac{\partial}{\partial \lambda_2} \int_{\Sigma_r} \theta\left(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_1}\right) \\ &= \int_{\Sigma_r} \left\{ \frac{\partial}{\partial \lambda_1} \left(\theta\left(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_2}\right) \right) - \frac{\partial}{\partial \lambda_2} \left(\theta\left(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_1}\right) \right) \right\}. \end{aligned}$$

The second equality results from $X_1(f) = \partial f / \partial \lambda_1$ for the scalar field $f := \langle \Theta_{\Sigma_r}, X_2 \rangle$ [cf. Eq. (A.8)], the third equality makes use of Eqs. (17.74) and (17.77) and the last equality stems from the linearity of the integral. Since $\Omega(X_1, X_2) = \lim_{r \rightarrow +\infty} d\Theta_{\Sigma_r}(X_1, X_2)$ and $\Sigma_r \rightarrow \Sigma$ for $r \rightarrow +\infty$, this proves Eq. (17.78), provided that ω is defined by (17.79). The $(n - 1)$ -form ω does not depend on the second derivative $\frac{\partial^2 \mathbf{g}}{\partial \lambda_1 \partial \lambda_2}$ because the second derivatives, which appear in each of the two terms of the right-hand side of Eq. (17.79), cancel out thanks to the identity $\frac{\partial^2 \mathbf{g}}{\partial \lambda_1 \partial \lambda_2} = \frac{\partial^2 \mathbf{g}}{\partial \lambda_2 \partial \lambda_1}$. Hence the notation $\omega = \omega(\mathbf{g}, \partial \mathbf{g} / \partial \lambda_1, \partial \mathbf{g} / \partial \lambda_2)$. \square

Example 9 (presymplectic current form of general relativity): For general relativity, one deduces from expression (17.9) for θ and the definition (17.79) that the presymplectic current $(n - 1)$ -form is

$$\omega(\mathbf{g}, \delta_1\mathbf{g}, \delta_2\mathbf{g})_{\alpha_1 \dots \alpha_{n-1}} = (32\pi)^{-1} P^{\lambda\mu\nu\tau\rho\sigma} (\delta_2 g_{\mu\nu} \nabla_\tau \delta_1 g_{\rho\sigma} - \delta_1 g_{\mu\nu} \nabla_\tau \delta_2 g_{\rho\sigma}) \epsilon_{\lambda\alpha_1 \dots \alpha_{n-1}}, \quad (17.80)$$

with

$$P^{\lambda\mu\nu\tau\rho\sigma} := 2g^{\lambda\rho}g^{\sigma\mu}g^{\nu\tau} - g^{\lambda\tau}g^{\mu\rho}g^{\sigma\nu} - g^{\lambda\mu}g^{\nu\tau}g^{\rho\sigma} - g^{\mu\nu}g^{\lambda\rho}g^{\sigma\tau} + g^{\mu\nu}g^{\lambda\tau}g^{\rho\sigma}. \quad (17.81)$$

We shall not detail the computation, which is straightforward but rather lengthy. Let us simply mention that one shall start by expressing the covariant derivatives in formula (17.9) for θ in terms of \mathbf{g} 's Christoffel symbols [Eq. (A.64) with $(k, \ell) = (0, 2)$] and replace ϵ by $\sqrt{-g} e$ as in Example 2, plug the result in Eq. (17.79) and make use of the identities (16.20) and (16.21). As a check, one can compare with Eqs. (21)-(23) of Ref. [282] or Eqs. (41)-(43) of Ref. [506].

Lemma 17.18: asymptotic vanishing of $d\Upsilon_{\Sigma_r, \Sigma'_r}$

The 1-form $\Upsilon_{\Sigma_r, \Sigma'_r}$ defined by Eq. (17.75) obeys $\lim_{r \rightarrow +\infty} d\Upsilon_{\Sigma_r, \Sigma'_r} = 0$.

Proof. By exactly the same computation as that leading to Eq. (17.78), we have

$$d\Upsilon_{\Sigma_r, \Sigma'_r}(X_1, X_2) = \int_{\mathcal{W}_r^\nearrow} \omega(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_1}, \frac{\partial \mathbf{g}}{\partial \lambda_2}).$$

As argued in Ref. [291], if the asymptotic flatness conditions ensure that the integral (17.78) on the unbounded hypersurface Σ is convergent, then the integral of the same integrand ω over \mathcal{W}_r must tend to zero for $r \rightarrow +\infty$. \square

Property 17.19: uniqueness of the presymplectic form

The presymplectic form Ω is unique for a given gravity theory with appropriate asymptotic flatness conditions. Not only it does not depend on the choice of the hypersurface Σ (Property 17.16), but it is unaffected by the ambiguities $\theta \rightarrow \theta + d\mathbf{Y} + \delta\boldsymbol{\mu}$ in the definition of θ [cf. Eq. (17.6)]. Moreover, the presymplectic current $(n - 1)$ -form ω itself is unaffected by the ambiguity $\theta \rightarrow \theta + \delta\boldsymbol{\mu}$.

Proof. Let us first show the last point, by evaluating ω from Eq. (17.79) for $\theta'(\mathbf{g}, \delta\mathbf{g}) = \theta(\mathbf{g}, \delta\mathbf{g}) + \delta\boldsymbol{\mu}(\mathbf{g})$. For $\delta\mathbf{g} = (\partial\mathbf{g}/\partial\lambda_2) \delta\lambda_2$, we get

$$\theta'(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_2}) = \theta(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_2}) + \frac{\partial}{\partial \lambda_2} \boldsymbol{\mu}(\mathbf{g}),$$

so that

$$\frac{\partial}{\partial \lambda_1} \left(\theta'(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_2}) \right) = \frac{\partial}{\partial \lambda_1} \left(\theta(\mathbf{g}, \frac{\partial \mathbf{g}}{\partial \lambda_2}) \right) + \frac{\partial^2}{\partial \lambda_1 \partial \lambda_2} \boldsymbol{\mu}(\mathbf{g}).$$

Thanks to the commutativity of partial derivatives, the terms involving μ cancel each other when substituting θ' for θ in Eq. (17.79). Hence the $(n - 1)$ -form ω computed from θ' is the same as that computed from θ . It follows then from Eq. (17.78) that the presymplectic form Ω is independent of μ as well. On the other side, the change $\theta'(g, \delta g) = \theta(g, \delta g) + d(Y(g, \delta g))$, as permitted by Eq. (17.6), leads to the following change in the presymplectic form:

$$\Omega'(X_1, X_2) = \Omega(X_1, X_2) + \int_{\mathcal{S}_\infty} \left[\frac{\partial}{\partial \lambda_1} \left(Y(g, \frac{\partial g}{\partial \lambda_2}) \right) - \frac{\partial}{\partial \lambda_2} \left(Y(g, \frac{\partial g}{\partial \lambda_1}) \right) \right],$$

where use has been made of Eqs. (17.78)-(17.79), as well as the Stokes' theorem (A.94), \mathcal{S}_∞ standing for the “outer boundary” of Σ , i.e. the limit $r \rightarrow +\infty$ of a $(n - 2)$ -sphere \mathcal{S}_r of radius r in the asymptotically Euclidean end of Σ . Now, given that the $(n - 2)$ form Y must be constructed from $g, \delta g$ and their derivatives in a covariant manner, the fall-off conditions from standard asymptotic flatness are sufficient to make the above integral over \mathcal{S}_∞ vanish [291]. Hence $\Omega' = \Omega$. \square

Property 17.20: closedness of the presymplectic current $(n - 1)$ -form

Provided that the 2-parameter family $g(\lambda_1, \lambda_2)$ involved in the definition (17.79) regards only solutions to the field equations (i.e. $g(\lambda_1, \lambda_2) \in \mathcal{F}$), the presymplectic current ω is a closed $(n - 1)$ -form:

$$d\omega = 0. \quad (17.82)$$

Proof. The 2-parameter family of metric tensors $g(\lambda_1, \lambda_2)$ gives birth to a 2-parameter family of Lagrangian n -forms $L(\lambda_1, \lambda_2) := L(g(\lambda_1, \lambda_2))$. The variation $\delta_1 L$ of L corresponding to a variation of λ_1 is given by Eq. (17.5) with $E^{\mu\nu} = 0$ since $g(\lambda_1, \lambda_2)$ fulfills the field equations. Hence $\delta_1 L = d(\theta(g, \delta_1 g))$. Given that $\delta_1 L = \frac{\partial L}{\partial \lambda_1} \delta \lambda_1$, $\delta_1 g = \frac{\partial g}{\partial \lambda_1} \delta \lambda_1$ and $\theta(., .)$ is linear with respect to its second argument, we get $\partial L / \partial \lambda_1 = d(\theta(g, \partial g / \partial \lambda_1))$. Taking the derivative of this expression with respect to λ_2 yields

$$\frac{\partial^2 L}{\partial \lambda_2 \partial \lambda_1} = \frac{\partial}{\partial \lambda_2} d\left(\theta(g, \frac{\partial g}{\partial \lambda_1})\right) = d\frac{\partial}{\partial \lambda_2} \left(\theta(g, \frac{\partial g}{\partial \lambda_1})\right).$$

Swapping λ_1 and λ_2 and subtracting the above formula yields, in view of ω 's definition (17.79), $\frac{\partial^2 L}{\partial \lambda_1 \partial \lambda_2} - \frac{\partial^2 L}{\partial \lambda_2 \partial \lambda_1} = d\omega$. But the left-hand side of this equation is trivially zero, hence the result (17.82). \square

It turns out that in general, the presymplectic form Ω is *degenerate*: there exists nonzero vectors X on \mathcal{F} , such that $\Omega(X, .) = 0$. Hence Ω fails to be a symplectic form on \mathcal{F} . For instance, one may conceive a nonzero variation δg of g such that $\delta g = 0$ on the hypersurface Σ involved in Eq. (17.78). Then, by linearity of $\omega(g, \delta_1 g, \delta_2 g)$ with respect to $\delta_1 g$ and $\delta_2 g$ (cf. Property 17.17), one has $\omega(g, \delta g, .) = 0$ on Σ and Eq. (17.78) shows that δg defines a degeneracy direction of Ω . More generally, the degeneracy of Ω is related to the gauge freedom of the gravity theory, i.e. to the invariance by diffeomorphism. For general relativity, it has been shown [24] that the degeneracy directions of Ω are exactly those generated by diffeomorphisms

of \mathcal{M} [cf. Eqs. (17.11)-(17.12)] that reduce to the identity at infinity (i.e. such that ξ in Eq. (17.12) tends to zero for $r \rightarrow +\infty$). The gauge freedom is intimately related to the initial value problem of the theory being *not* well posed (it becomes well posed only in a fixed gauge, see e.g. Chap. VI of Ref. [107]). Recall that in the covariant phase space construction of Property 17.14, it was required that the initial value problem is well posed to endow the space of solutions with a (non-degenerate) symplectic form.

It is however possible to introduce a quotient space $\bar{\mathcal{F}} = \mathcal{F}/\mathcal{R}$ on which Ω gives birth to a true symplectic form. The idea is to “factor out” the degeneracy directions of Ω . This is possible because these directions are integrable into submanifolds of \mathcal{F} . According to Frobenius theorem (Theorem B.3.1 in Ref. [498]), the integrability is equivalent to the commutator $[X_1, X_2]$ being a degeneracy vector field for any two degeneracy vector fields X_1 and X_2 . To show this property, let us use the identity (A.79) and the Leibniz rule to write $\Omega([X_1, X_2], .) = \Omega(\mathcal{L}_{X_1} X_2, .) = \mathcal{L}_{X_1}[\Omega(X_2, .)] - \mathcal{L}_{X_2}\Omega(X_1, .) = -\mathcal{L}_{X_1}\Omega(X_2, .)$ since $\Omega(X_2, .) = 0$. Now, thanks to Cartan identity (A.95), $\mathcal{L}_{X_1}\Omega = d\Omega(X_1, ., .) + d[\Omega(X_1, .)] = 0$, given that $d\Omega = 0$ (Property 17.16) and $\Omega(X_1, .) = 0$. Hence $\Omega([X_1, X_2], .) = 0$ and the degeneracy directions of Ω are integrable into submanifolds of \mathcal{F} . Assuming that these submanifolds provide a regular foliation of \mathcal{F} (cf. the discussion in Ref. [344]), one defines an equivalence relation \mathcal{R} on \mathcal{F} by setting $\mathbf{g}_1 \mathcal{R} \mathbf{g}_2$ iff the points \mathbf{g}_1 and \mathbf{g}_2 of \mathcal{F} belong to the same degeneracy submanifold. The set of equivalence classes $\bar{\mathcal{F}} := \mathcal{F}/\mathcal{R}$ can be given the structure of Banach manifold, such that the canonical projection $\pi : \mathcal{F} \rightarrow \bar{\mathcal{F}}$ is a smooth map. One may then define a 2-form $\bar{\Omega}$ on $\bar{\mathcal{F}}$ by setting, for any pair (\bar{X}_1, \bar{X}_2) of tangent vectors at a point $\bar{\mathbf{g}} \in \bar{\mathcal{F}}$, $\bar{\Omega}(\bar{X}_1, \bar{X}_2) := \Omega(X_1, X_2)$, where $(X_1, X_2) \in (T_{\mathbf{g}}\mathcal{F})^2$, such that $\pi(\mathbf{g}) = \bar{\mathbf{g}}$, $\pi_* X_1 = \bar{X}_1$ and $\pi_* X_2 = \bar{X}_2$ (pushforward by π , cf. Sec. A.2.8). Because Ω is preserved along its degeneracy directions (cf. $\mathcal{L}_{X_1}\Omega = 0$ in the above proof), the definition of $\bar{\Omega}(\bar{X}_1, \bar{X}_2)$ is independent of the choice of \mathbf{g} in the fiber $\pi^{-1}(\bar{\mathbf{g}})$. It is also independent of the choice of X_1 and X_2 in $T_{\mathbf{g}}\mathcal{F}$ because two vectors X_1 and X'_1 such that $\pi_* X_1 = \pi_* X'_1$ differ only by a vector along a degeneracy direction of Ω . Hence $\bar{\Omega}(\bar{X}_1, \bar{X}_2)$ is unambiguously defined. It can be shown that $\bar{\Omega}$ is a closed non-degenerate 2-form, i.e. a symplectic form (cf. Refs. [254, 344] for details). The genuine covariant phase space of the considered gravity theory is then $(\bar{\mathcal{F}}, \bar{\Omega})$. However, in what follows, we will deal only with the prephase space (\mathcal{F}, Ω) .

17.4.4 Fundamental identity of the covariant phase space formalism

Property 17.20 tells that the $(n-1)$ -form $\omega(\mathbf{g}, \delta_1\mathbf{g}, \delta_2\mathbf{g})$ is closed for any pair of variations $(\delta_1\mathbf{g}, \delta_2\mathbf{g})$ among the solutions to the field equations. Actually, ω is *globally*¹¹ exact as soon as $\delta_2\mathbf{g}$ results from an infinitesimal diffeomorphism generated by a vector field ξ : ω is then the exterior derivative of a $(n-2)$ -form, which involves the variation $\delta_1 Q(\xi)$ of ξ 's Noether potential $(n-2)$ -form. More precisely, we have:

Property 17.21: fundamental identity of the covariant phase space formalism

Let (\mathcal{M}, g) be a n -dimensional spacetime governed by a diffeomorphism-invariant theory

¹¹Let us recall that thanks to Poincaré lemma (cf. Sec. A.4.3), ω closed $\implies \omega$ locally exact.

and let ξ be a vector field on \mathcal{M} . Any variation δg among the solutions to the field equations^a generates a variation $\delta Q(\xi)$ of the Noether potential $(n - 2)$ -form of ξ (cf. Property 17.6 and Remark 4 on p. 660), such that

$$\omega(g, \delta g, \mathcal{L}_\xi g) = d[\delta Q(\xi) - \xi \cdot \theta(g, \delta g)], \quad (17.83)$$

where ω is the presymplectic current $(n - 1)$ -form (Property 17.17) and θ is the presymplectic potential $(n - 1)$ -form (Property 17.2).

^aThis means that both g and $g + \delta g$ are assumed to be solutions to the field equations, which implies that δg is a solution to the linearized field equation.

Proof. The variation of the Noether current $(n - 1)$ -form $J(\xi)$ induced by δg is obtained from Eq. (17.16): $\delta J(\xi) = \delta \theta(g, \mathcal{L}_\xi g) - \xi \cdot \delta L$. Note that we have set $\delta \xi = 0$, for the vector field ξ is independent of g . Here δL is given by Eq. (17.5) with $E = 0$ since we are dealing with solutions to the field equations only: $\delta L = d\theta(g, \delta g)$. Hence

$$\delta J(\xi) = \delta \theta(g, \mathcal{L}_\xi g) - \xi \cdot d\theta(g, \delta g) = \delta \theta(g, \mathcal{L}_\xi g) - \mathcal{L}_\xi \theta(g, \delta g) + d[\xi \cdot \theta(g, \delta g)],$$

where use has been made of Cartan identity (A.95). Now, there is no loss of generality in considering that there exists a 2-parameter family $g(\lambda_1, \lambda_2)$ of metric fields such that $\delta g = (\partial g / \partial \lambda_1) \delta \lambda_1$ and $\partial g / \partial \lambda_2 = \mathcal{L}_\xi g$ [cf. Eq. (17.12), rewritten as $\delta_2 g = \delta \lambda_2 \mathcal{L}_\xi g$]. In view of ω 's definition (17.79), we have then $\delta \theta(g, \mathcal{L}_\xi g) - \mathcal{L}_\xi \theta(g, \delta g) = \omega(g, \delta g, \mathcal{L}_\xi g)$, so that the above equation becomes

$$\delta J(\xi) = \omega(g, \delta g, \mathcal{L}_\xi g) + d[\xi \cdot \theta(g, \delta g)].$$

Finally, restoring the explicit dependence of J on g (cf. Remark 3 on p. 659), we have $\delta J(\xi) := J(g + \delta g, \xi) - J(g, \xi)$. Since both g and $g + \delta g$ are solutions to the field equations, we may invoke Eq. (17.21) to write $J(g + \delta g, \xi) = dQ(g + \delta g, \xi)$ and $J(g, \xi) = dQ(g, \xi)$. Hence there comes $\delta J(\xi) = d[Q(g + \delta g, \xi) - Q(g, \xi)] = d\delta Q(\xi)$, from which Eq. (17.83) follows. \square

Remark 6: The denomination *fundamental identity of the covariant phase space formalism* for Eq. (17.83) is used in some part of the literature [137, 245, 249], as here, while in another part [282, 283], it is given to a generalization of it, which holds even if g and $g + \delta g$ do not satisfy the field equations.

17.5 First law for diffeomorphism-invariant theories

17.5.1 Hamiltonian conjugate to an asymptotic symmetry

A vector field ξ on \mathcal{M} is said to be a *generator of an asymptotic symmetry* iff the diffeomorphisms of \mathcal{M} generated by ξ preserve the asymptotic flatness conditions stated in Property 17.15. Equivalently, $\Phi_t^* g \in \mathcal{F}$ for any $g \in \mathcal{F}$ and any diffeomorphism $\Phi_t : \mathcal{M} \rightarrow \mathcal{M}$ generated by ξ . One may then associate to ξ a vector field X_ξ on the prephase space \mathcal{F} : X_ξ is the vector field defined by Eq. (17.70): $X_\xi = \delta g / \delta \lambda$, with δg given by Eq. (17.12) with $t = \delta \lambda$:

$\delta\mathbf{g} = \delta\lambda \mathcal{L}_\xi \mathbf{g}$. The presymplectic product of a generic infinitesimal displacement $\delta\mathbf{g}$ on \mathcal{F} by X_ξ is then given by Eq. (17.78):

$$\Omega(\delta\mathbf{g}, X_\xi) = \int_{\Sigma} \omega(\mathbf{g}, \delta\mathbf{g}, \mathcal{L}_\xi \mathbf{g}).$$

If we substitute the fundamental identity (17.83) for $\omega(\mathbf{g}, \delta\mathbf{g}, \mathcal{L}_\xi \mathbf{g})$ and use Stokes' theorem (A.94), we arrive at

$$\Omega(\delta\mathbf{g}, X_\xi) = \delta \int_{\mathcal{S}_\infty} Q(\xi) - \int_{\mathcal{S}_\infty} \xi \cdot \theta(\mathbf{g}, \delta\mathbf{g}), \quad (17.84)$$

where $\mathcal{S}_\infty = \lim_{r \rightarrow +\infty} \mathcal{S}_r$, \mathcal{S}_r being the boundary of the hypersurface $\Sigma_r \subset \Sigma$ considered in Sec. 17.4.3. Let us suppose that the right-hand side of Eq. (17.84) is a pure variation, i.e. that there exists a function $H_\xi : \mathcal{F} \rightarrow \mathbb{R}$, $\mathbf{g} \mapsto H_\xi[\mathbf{g}]$ such that

$$\delta H_\xi = \delta \int_{\mathcal{S}_\infty} Q(\xi) - \int_{\mathcal{S}_\infty} \xi \cdot \theta(\mathbf{g}, \delta\mathbf{g}). \quad (17.85)$$

Then we can rewrite Eq. (17.84) as $\delta H_\xi = \Omega(\delta\mathbf{g}, X_\xi)$. Given that $\delta H_\xi = \langle dH_\xi, \delta\mathbf{g} \rangle$ [cf. Eq. (A.21)] and $\delta\mathbf{g}$ is a generic displacement on \mathcal{F} , this is actually equivalent to

$$dH_\xi = \Omega(., X_\xi). \quad (17.86)$$

We recognize Eq. (17.65), which defines the Hamiltonian vector field X_H in terms of the Hamiltonian H . The difference is that Eq. (17.65) involves a true symplectic form, while Ω in Eq. (17.86) is only presymplectic. In particular, we could not invert Ω to define X_ξ in terms of H_ξ , but this is not a problem here since X_ξ is given a priori and Eq. (17.86) defines H_ξ from it (assuming that the right-hand side of Eq. (17.84) is a pure variation). One says that H_ξ is the **Hamiltonian conjugate to the vector field ξ** [506]. Hence, if there exists a function H_ξ on \mathcal{F} fulfilling (17.85), the dynamics generated on the prephase space (\mathcal{F}, Ω) by the vector field X_ξ , i.e. the action of the diffeomorphism group generated by ξ on the solutions \mathbf{g} to the field equations, is a Hamiltonian dynamics.

A natural question is whether or not a given vector field ξ admits a conjugate Hamiltonian. A sufficient condition is that there exists a $(n-1)$ -form $B(\mathbf{g})$ on \mathcal{M} such that

$$\int_{\mathcal{S}_\infty} \xi \cdot \theta(\mathbf{g}, \delta\mathbf{g}) = \delta \int_{\mathcal{S}_\infty} \xi \cdot B(\mathbf{g}). \quad (17.87)$$

In particular, this holds if the asymptotic behavior of the presymplectic potential $(n-1)$ -form is $\theta(\mathbf{g}, \delta\mathbf{g}) \underset{r \rightarrow +\infty}{\sim} \delta B(\mathbf{g})$ for some $(n-1)$ -form $B(\mathbf{g})$. This is the case for general relativity:

Example 10 (asymptotic behavior of the presymplectic potential form in general relativity): Thanks to the asymptotic flatness conditions, let us introduce a flat metric f on \mathcal{M} (or at least in the asymptotic region of \mathcal{M}) such that $\mathbf{g} \sim f$ for $r \rightarrow +\infty$. In general relativity, θ is given by Eq. (17.8): $\theta(\mathbf{g}, \delta\mathbf{g}) = (16\pi)^{-1} H(\delta\mathbf{g}) \cdot \epsilon(\mathbf{g})$. We may then write

$$\theta(\mathbf{g}, \delta\mathbf{g}) \underset{r \rightarrow +\infty}{\sim} (16\pi)^{-1} \bar{H}(\delta\mathbf{g}) \cdot e, \quad (17.88)$$

where e is the Levi-Civita tensor of f and $\bar{H}(\delta g)$ is the vector defined from δg in the same way as $H(\delta g)$ in Eq. (16.13), but substituting f for g : $\bar{H}^\alpha := f^{\alpha\rho} f^{\mu\nu} (\mathcal{D}_\nu \delta g_{\rho\mu} - \mathcal{D}_\rho \delta g_{\mu\nu})$, where \mathcal{D} stands for the covariant derivative associated to f . Since $\delta f = 0$ and $\mathcal{D}\delta = \delta\mathcal{D}$, for f is independent of g , we may write $\bar{H}(\delta g) = \delta K(g)$, where $K(g)$ is the vector field defined by

$$K^\alpha := f^{\alpha\rho} f^{\mu\nu} (\mathcal{D}_\nu g_{\rho\mu} - \mathcal{D}_\rho g_{\mu\nu}). \quad (17.89)$$

Since $\delta e = 0$ (for f is independent of g), we deduce from Eq. (17.88) that

$$\theta(g, \delta g) \underset{r \rightarrow +\infty}{\sim} \delta B(g), \quad \text{with} \quad B(g) := (16\pi)^{-1} K(g) \cdot e. \quad (17.90)$$

The $(n-1)$ -form $B(g)$ defined above fulfills Eq. (17.87), thereby ensuring that a conjugate Hamiltonian H_ξ exists for any asymptotic symmetry generator ξ .

Remark 1: As the above example shows, the $(n-1)$ -form $B(g)$ does not need to be covariantly defined: Eq. (17.90) involves the flat metric f , which is arbitrarily chosen. In particular, $B(g)$ is highly non-unique; what matters is only its asymptotic behavior.

If condition (17.87) is satisfied, then a function H_ξ fulfilling (17.85) is

$$H_\xi[g] := \int_{\mathcal{S}_\infty} (Q(\xi) - \xi \cdot B(g)). \quad (17.91)$$

More generally, a necessary and sufficient criterion for the existence of H_ξ is

Property 17.22: existence of a Hamiltonian conjugate to a vector field (Wald & Zoupas 2000 [506])

A necessary condition for the existence of a Hamiltonian $H_\xi : \mathcal{F} \rightarrow \mathbb{R}$ conjugate to the vector field ξ is that

$$\int_{\mathcal{S}_\infty} \xi \cdot \omega(g, \delta_1 g, \delta_2 g) = 0 \quad (17.92)$$

for any pair of variations $(\delta_1 g, \delta_2 g)$ among the solutions to the field equations. Furthermore, if the prephase space \mathcal{F} is simply connected, this condition is sufficient.

Proof. Let us suppose that H_ξ exists. Then, by taking the variation δ_1 of Eq. (17.85) written for $\delta = \delta_2$, we get

$$\delta_1 \delta_2 H_\xi = \delta_1 \delta_2 \mathcal{Q}_{\mathcal{S}_\infty, \xi} - \int_{\mathcal{S}_\infty} \xi \cdot \delta_1 (\theta(g, \delta_2 g)),$$

where $\mathcal{Q}_{\mathcal{S}_\infty, \xi}$ is the Noether charge of \mathcal{S}_∞ with respect to ξ [cf. Eq. (17.28)]. By subtracting the term $1 \leftrightarrow 2$ and using $(\delta_1 \delta_2 - \delta_2 \delta_1) H_\xi = 0$, as well as $(\delta_1 \delta_2 - \delta_2 \delta_1) \mathcal{Q}_{\mathcal{S}_\infty, \xi} = 0$, we obtain (17.92), given the definition (17.79) of ω . Hence (17.92) appears as a necessary condition for H_ξ to exist. To prove that it is sufficient when \mathcal{F} is simply connected, let us select a point $g_0 \in \mathcal{F}$ and define for any point $g \in \mathcal{F}$ and any curve \mathcal{L} from g_0 to g ,

$$\mathcal{H}(g, \mathcal{L}) := \int_{\mathcal{L}} \alpha, \quad (17.93)$$

where α is the 1-form on \mathcal{F} defined by the following action on a vector $X = \partial g / \partial \lambda$:

$$\langle \alpha, X \rangle := \int_{\mathcal{S}_\infty} \left(\frac{\partial}{\partial \lambda} Q(\xi) - \xi \cdot \theta(g, \frac{\partial g}{\partial \lambda}) \right). \quad (17.94)$$

Let \mathcal{L}' be a curve of \mathcal{F} , distinct from \mathcal{L} , connecting g_0 to g . If \mathcal{F} is simply connected, there exists a homotopy from \mathcal{L} to \mathcal{L}' . This defines a compact 2-dimensional surface $\mathcal{S} \subset \mathcal{F}$, the boundary of which is $\mathcal{L} \cup \overset{\leftarrow}{\mathcal{L}'}$, where $\overset{\leftarrow}{\mathcal{L}'}$ is same curve as \mathcal{L}' , but oriented from g to g_0 . Stokes' theorem (A.94) yields then

$$\mathcal{H}(g, \mathcal{L}) - \mathcal{H}(g, \mathcal{L}') = \int_{\mathcal{S}} d\alpha.$$

To evaluate the 2-form $d\alpha$ on \mathcal{S} , let us compute $d\alpha(X_1, X_2)$ for a pair of vectors (X_1, X_2) associated to a 2-parameter family $\bar{g}(\lambda_1, \lambda_2)$ spanning \mathcal{S} [cf. Eq. (17.77)]. We may choose $(\lambda_1, \lambda_2) \in [0, 1]^2$, with $\bar{g}(\lambda_1, 0)$ spanning \mathcal{L} , $\bar{g}(\lambda_1, 1)$ spanning \mathcal{L}' , $\bar{g}(0, 0) = \bar{g}(0, 1) = g_0$ and $\bar{g}(1, 0) = \bar{g}(1, 1) = g$. By a computation similar to that of $d\Theta_{\Sigma_r}(X_1, X_2)$ in the proof of Property 17.17, we arrive at

$$d\alpha(X_1, X_2) = \int_{\mathcal{S}_\infty} \left\{ \underbrace{\frac{\partial^2 Q(\xi)}{\partial \lambda_1 \partial \lambda_2} - \frac{\partial^2 Q(\xi)}{\partial \lambda_2 \partial \lambda_1}}_0 - \xi \cdot \left[\underbrace{\frac{\partial}{\partial \lambda_1} \left(\theta(\bar{g}, \frac{\partial \bar{g}}{\partial \lambda_2}) \right) - \frac{\partial}{\partial \lambda_2} \left(\theta(\bar{g}, \frac{\partial \bar{g}}{\partial \lambda_1}) \right)}_{\omega\left(\bar{g}, \frac{\partial \bar{g}}{\partial \lambda_1}, \frac{\partial \bar{g}}{\partial \lambda_2}\right)} \right] \right\}.$$

Hence if Eq. (17.92) holds, $d\alpha(X_1, X_2) = 0$. We conclude that $d\alpha = 0$ on \mathcal{S} and thus that $\mathcal{H}(g, \mathcal{L}) - \mathcal{H}(g, \mathcal{L}') = 0$. It follows that $\mathcal{H}(g, \mathcal{L})$ is a function of g only, which we shall denote by $\mathcal{H}(g)$. From the definitions (17.93)-(17.94), as well as the definition of the integral of a 1-form along a curve, we get

$$\delta \mathcal{H} = \mathcal{H}(g + \delta g) - \mathcal{H}(g) = \langle \alpha, \delta g \rangle = \int_{\mathcal{S}_\infty} (\delta Q(\xi) - \xi \cdot \theta(g, \delta g)) = \delta \int_{\mathcal{S}_\infty} Q(\xi) - \int_{\mathcal{S}_\infty} \xi \cdot \theta(g, \delta g).$$

Comparing with (17.85), we conclude that $\mathcal{H}(g)$ is a Hamiltonian conjugate to ξ . \square

If the vector field ξ corresponds to an asymptotic time translation, it is natural to define the energy of the system as the value of the conjugate Hamiltonian (cf. Property 17.13):

Property 17.23: canonical energy relative to an asymptotic time translation

Let (\mathcal{M}, g) be an asymptotically flat spacetime governed by some diffeomorphism-invariant theory of gravity. Let ξ be a vector field on \mathcal{M} , generator of an asymptotic symmetry corresponding to a time translation. Let us assume that the asymptotic flatness conditions guarantee that a Hamiltonian H_ξ conjugate to ξ exists, as well as a $(n-1)$ -form B such that H_ξ takes the shape (17.91). The **canonical energy** of the spacetime (\mathcal{M}, g) relative to

ξ is then the value of H_ξ at g :

$$\mathcal{E} := \int_{\mathcal{S}_\infty} (Q(\xi) - \xi \cdot \mathbf{B}). \quad (17.95)$$

Example 11 (canonical energy for general relativity is ADM energy): From expression (17.23) for $Q(\xi)$ in general relativity, we get

$$\int_{\mathcal{S}_\infty} Q(\xi) = -\frac{1}{16\pi} \int_{\mathcal{S}_\infty} \star(d\xi) = -\frac{1}{32\pi} \int_{\mathcal{S}_\infty} (d\xi)_{\mu\nu} dS^{\mu\nu} = -\frac{1}{16\pi} \int_{\mathcal{S}_\infty} \partial_\mu \xi_\nu dS^{\mu\nu},$$

where we have used Lemma 5.10 to express the integral in terms of the area element normal bivector dS to \mathcal{S}_∞ and have used formula (A.90b) and the antisymmetry of dS to get the last equality. Now, since ξ generates an asymptotic time translation, we have $\xi = \partial_t$, where t is the first coordinate of an *asymptotically Minkowskian system* $(x^\alpha) = (t, x^1, \dots, x^{n-1})$, i.e. a f -Minkowskian coordinate system (cf. Sec. 5.3.1) for a flat metric f such that $g \sim f$ for $r := \sqrt{(x^1)^2 + \dots + (x^{n-1})^2} \rightarrow +\infty$. \mathcal{S}_∞ can then be considered as the limit $r \rightarrow +\infty$ of a $(n-2)$ -sphere $\mathcal{S}_{t,r}$ defined by $(t, r) = \text{const}$. From Eq. (5.42), we have $dS^{\mu\nu} := (s^\mu n^\nu - n^\mu s^\nu) dS$, with $dS := \sqrt{q} d^{n-2}y$ (the area element of $\mathcal{S}_{t,r}$) and, at the limit $r \rightarrow +\infty$, $n^\mu = \xi^\mu = \delta^\mu_t$ and $s^\mu = (0, s^i)$, with $s^i := x^i/r$ for $i \in \{1, \dots, n-1\}$. Furthermore, $\xi_\nu = g_{\nu\sigma} \xi^\sigma = g_{t\nu}$. Hence, there comes

$$\int_{\mathcal{S}_\infty} Q(\xi) = -\frac{1}{16\pi} \int_{\mathcal{S}_\infty} \partial_\mu g_{t\nu} (s^\mu \delta^\nu_t - \delta^\mu_t s^\nu) dS = -\frac{1}{16\pi} \int_{\mathcal{S}_\infty} (\partial_t g_{tt} - \partial_t g_{ti}) s^i dS. \quad (17.96)$$

On the other hand, from the $(n-1)$ -form \mathbf{B} obtained in Eq. (17.90), we have $\xi \cdot \mathbf{B} = (16\pi)^{-1} \xi \cdot (\mathbf{K} \cdot \mathbf{e}) = (16\pi)^{-1} \mathbf{e}(\mathbf{K}, \xi, \dots)$. For $r \rightarrow +\infty$, $e \sim \epsilon$ and we may write $\xi \cdot \mathbf{B} \sim (16\pi)^{-1} \epsilon(\mathbf{K}, \xi, \dots) = (16\pi)^{-1} \star(\underline{\mathbf{K}} \wedge \underline{\xi})$, where $\underline{\mathbf{K}} \wedge \underline{\xi}$ stands for the 2-form of components $(\underline{\mathbf{K}} \wedge \underline{\xi})_{\alpha\beta} = K_\alpha \xi_\beta - \xi_\alpha K_\beta$ and we have used Eq. (5.38) with $p=2$ to let appear the Hodge dual of $\underline{\mathbf{K}} \wedge \underline{\xi}$. Using again Lemma 5.10 and the antisymmetry property $(K_\mu \xi_\nu - \xi_\mu K_\nu) dS^{\mu\nu} = 2K_\mu \xi_\nu dS^{\mu\nu}$, we arrive at

$$\begin{aligned} \int_{\mathcal{S}_\infty} \xi \cdot \mathbf{B} &= \frac{1}{16\pi} \int_{\mathcal{S}_\infty} K_\mu \xi_\nu dS^{\mu\nu} = \frac{1}{16\pi} \int_{\mathcal{S}_\infty} K_\mu g_{t\nu} (s^\mu \delta^\nu_t - \delta^\mu_t s^\nu) dS \\ &= \frac{1}{16\pi} \int_{\mathcal{S}_\infty} (K_i \underbrace{g_{tt}}_{-1} - K_t \underbrace{g_{ti}}_0) s^i dS = -\frac{1}{16\pi} \int_{\mathcal{S}_\infty} K_i s^i dS. \end{aligned}$$

Now, from Eq. (17.89), we have, on \mathcal{S}_∞ , $K_i = f^{\mu\nu} (\mathcal{D}_\nu g_{i\mu} - \mathcal{D}_i g_{\mu\nu}) = -(\partial_t g_{it} - \partial_i g_{tt}) + \delta^{kl} (\partial_l g_{ik} - \partial_i g_{kl})$. Hence

$$\int_{\mathcal{S}_\infty} \xi \cdot \mathbf{B} = -\frac{1}{16\pi} \int_{\mathcal{S}_\infty} (\partial_t g_{tt} - \partial_t g_{ti}) s^i dS - \frac{1}{16\pi} \int_{\mathcal{S}_\infty} \delta^{kl} (\partial_l g_{ik} - \partial_i g_{kl}) s^i dS.$$

We notice that the first integral on the right-hand side is exactly that appearing in Eq. (17.96), so that they cancel each other in formula (17.95) for the canonical energy, leaving

$$\mathcal{E} = \frac{1}{16\pi} \int_{\mathcal{S}_\infty} \delta^{kl} (\partial_l g_{ik} - \partial_i g_{kl}) s^i dS. \quad (17.97)$$

This is exactly the formula giving the ADM energy (cf. Sec. 5.3.5): compare Eq. (17.97) with e.g. Eq. (8.11) in Ref. [226].

Remark 2: The integral of $Q(\xi)$ over \mathcal{S}_∞ , which appears in formula (17.95) for \mathcal{E} , is the Noether charge $\mathcal{Q}_{\mathcal{S}_\infty, \xi}$ of \mathcal{S}_∞ with respect to ξ (cf. Property 17.8). For a stationary spacetime in general relativity, we have noticed that $\mathcal{Q}_{\mathcal{S}_\infty, \xi}$ is $(n-3)/(n-2)$ times the Komar mass at infinity (Property 17.9). The latter is identical to the ADM energy (Property 5.18). Hence, for a stationary spacetime ruled by general relativity, the relative contributions to \mathcal{E} of the terms $Q(\xi)$ and $-\xi \cdot B$ in Eq. (17.95) are respectively $(n-3)/(n-2)$ and $1/(n-2)$, i.e. $1/2$ and $1/2$ for $n=4$.

Let us now consider a vector field η on \mathcal{M} that is the generator of an asymptotic symmetry corresponding to a spatial rotation, i.e. $\eta \sim_{r \rightarrow +\infty} x\partial_y - y\partial_x$ in some asymptotically Minkowskian coordinate system $(t, x^1 =: x, x^2 =: y, x^3, \dots, x^{n-1})$. Without any loss of generality, we may choose the $(n-2)$ -surfaces $\mathcal{S}_{t,r}$ converging to \mathcal{S}_∞ for $r \rightarrow +\infty$ such that η is tangent to each $\mathcal{S}_{t,r}$ and thus to \mathcal{S}_∞ . Then, in Eq. (17.84) with $\xi := \eta$, we have $\int_{\mathcal{S}_\infty} \eta \cdot \theta(g, \delta g) = 0$. Indeed, $\eta \cdot \theta := \theta(\eta, ., ., .)$ is a $(n-2)$ -form that vanishes on \mathcal{S}_∞ : for any $(n-2)$ -tuple of vectors (v_1, \dots, v_{n-2}) tangent to \mathcal{S}_∞ , $(\eta, v_1, \dots, v_{n-2})$ is $(n-1)$ -tuple of vectors all tangent to the $(n-2)$ -dimensional manifold \mathcal{S}_∞ ; these vectors are thus necessarily linearly dependent, so that $\theta(\eta, v_1, \dots, v_{n-2}) = 0$ since θ is fully antisymmetric. It follows that a Hamiltonian conjugate to η always exists and is given by $H_\eta[g] = \int_{\mathcal{S}_\infty} Q(\eta)$. The negative of its value for a given solution g to the field equation defines the canonical angular momentum of that solution relative to η :

Property 17.24: canonical angular momentum relative to an asymptotic spatial rotation

Let (\mathcal{M}, g) be an asymptotically flat spacetime governed by some diffeomorphism-invariant theory of gravity. Let η be a vector field on \mathcal{M} , generator of an asymptotic symmetry corresponding to a spatial rotation. Then a Hamiltonian H_η conjugate to η always exists and the **canonical angular momentum** of the spacetime (\mathcal{M}, g) relative to η is the value of $-H_\eta$ at g :

$$\mathcal{J} := - \int_{\mathcal{S}_\infty} Q(\eta). \quad (17.98)$$

\mathcal{J} is actually the opposite of the Noether charge $\mathcal{Q}_{\mathcal{S}_\infty, \eta}$ of \mathcal{S}_∞ with respect to η , as defined in Property 17.8.

Example 12 (canonical angular momentum of an axisymmetric spacetime in general relativity): Let (\mathcal{M}, g) be an axisymmetric spacetime ruled by general relativity and let η be the Killing vector generating the axisymmetry. According to Eq. (17.32), \mathcal{J} is nothing than the Komar angular momentum relative to η at infinity.

17.5.2 First law for diffeomorphism-invariant theories

Finally, we arrive at the first law:

Property 17.25: first law of black hole thermodynamics for diffeomorphism-invariant theories (Iyer & Wald 1994 [291, 501])

Let $(\mathcal{M}, \mathbf{g})$ be a n -dimensional stationary spacetime governed by a diffeomorphism-invariant gravity theory. Let us assume that $(\mathcal{M}, \mathbf{g})$ contains a black hole, whose event horizon \mathcal{H} is part of a bifurcate Killing horizon (cf. Sec. 3.4) with respect to the Killing vector

$$\chi = \xi + \sum_{i=1}^L \Omega^{(i)} \eta_{(i)}, \quad (17.99)$$

where ξ is the stationary Killing vector, the $\eta_{(i)}$'s are L axisymmetric Killing vectors ($1 \leq L \leq [(n-1)/2]$) and the $\Omega^{(i)}$'s are L real constants. By virtue of Property 3.16, the surface gravity κ of \mathcal{H} is constant, so that the Hawking temperature of \mathcal{H} is well defined: $T_H = \hbar\kappa/(2\pi k_B)$ [Eq. (16.62)]. Let us consider an asymptotically flat solution $\delta\mathbf{g}$ to the linearized field equations around \mathbf{g} , i.e. a variation $\delta\mathbf{g}$ in the space \mathcal{F} of asymptotically flat solutions to the field equations ($\delta\mathbf{g}$ is not required to be stationary). Let us define the **Wald entropy** of the black hole in the perturbed spacetime $(\mathcal{M}, \mathbf{g} + \delta\mathbf{g})$ by the same formula as in the stationary case, namely Eq. (17.33), but with the integral taken on the bifurcation surface $\hat{\mathcal{S}}$ of the unperturbed configuration^a:

$$S_W := \frac{2\pi k_B}{\hbar\kappa} \int_{\hat{\mathcal{S}}} Q(\chi), \quad (17.100)$$

where χ is the same vector field on \mathcal{M} as in the unperturbed configuration (in other words, $\delta\chi = 0$). Finally, let us assume that the asymptotic flatness conditions ensure that a Hamiltonian conjugate to ξ exists. Then $\delta\mathbf{g}$ generates a variation $\delta\mathcal{E}$ of the canonical energy relative to ξ [Eq. (17.95)] obeying

$$\boxed{\delta\mathcal{E} = T_H \delta S_W + \sum_{i=1}^L \Omega^{(i)} \delta \mathcal{J}_{(i)}}, \quad (17.101)$$

where $\mathcal{J}_{(i)}$ is the canonical angular momenta relative to $\eta_{(i)}$ [Eq. (17.98)].

^a As discussed in the proof of Property 16.1, there is no loss of generality in considering that the event horizons of the stationary spacetime $(\mathcal{M}, \mathbf{g})$ and the perturbed one $(\mathcal{M}, \mathbf{g} + \delta\mathbf{g})$ are described by the same hypersurface \mathcal{H} of \mathcal{M} . A priori \mathcal{H} is not a Killing horizon in $(\mathcal{M}, \mathbf{g} + \delta\mathbf{g})$, since the latter is not assumed stationary. The submanifold $\hat{\mathcal{S}}$ of \mathcal{H} is thus not a bifurcation surface in $(\mathcal{M}, \mathbf{g} + \delta\mathbf{g})$. However, this does not prevent ones to define the integral of the $(n-2)$ -form $Q(\chi)$ in $(\mathcal{M}, \mathbf{g} + \delta\mathbf{g})$ (i.e. $Q(\chi) = Q(\mathbf{g} + \delta\mathbf{g}, \chi)$, if the dependence of Q on \mathbf{g} is made explicit) on the $(n-2)$ -surface $\hat{\mathcal{S}}$, as in Eq. (17.100). Similarly, in Eq. (17.100), κ is the surface gravity of \mathcal{H} in $(\mathcal{M}, \mathbf{g})$, given that there is no well-defined concept of surface gravity if \mathcal{H} is not a Killing horizon.

Proof. Let us consider the fundamental identity (17.83) with $\xi = \chi$, which holds since we assume that χ is a fixed vector field on \mathcal{M} ($\delta\chi = 0$). We may set $\omega(\mathbf{g}, \delta\mathbf{g}, \mathcal{L}_\chi \mathbf{g}) = 0$ in its left-hand side since $\omega(\mathbf{g}, \delta\mathbf{g}, \mathcal{L}_\chi \mathbf{g})$ is linear with respect to $\mathcal{L}_\chi \mathbf{g}$ (Property 17.17) and $\mathcal{L}_\chi \mathbf{g} = 0$ for χ is a Killing vector of \mathbf{g} . There remains then

$$d[\delta Q(\chi) - \chi \cdot \theta(\mathbf{g}, \delta\mathbf{g})] = 0.$$

The left-hand side is a $(n-1)$ -form. Let us integrate it on a hypersurface Σ extending from

the bifurcation surface $\hat{\mathcal{S}}$ of the bifurcate Killing horizon to the asymptotically flat end. Using Stokes' theorem (A.94) with $\partial\Sigma = \hat{\mathcal{S}} \cup \mathcal{S}_\infty$, we get

$$\delta \int_{\hat{\mathcal{S}}} Q(\chi) = \int_{\mathcal{S}_\infty} [\delta Q(\chi) - \chi \cdot \theta(g, \delta g)]. \quad (17.102)$$

Note that we have used the vanishing of χ on $\hat{\mathcal{S}}$ (Property 3.12) to write the integral in the left-hand side as $\int_{\hat{\mathcal{S}}} \delta Q(\chi)$, which is equal to $\delta \int_{\hat{\mathcal{S}}} Q(\chi)$. Since $Q(\chi)$ is linear with respect to χ (Property 17.6), Eq. (17.99) leads to

$$\underbrace{\delta \int_{\hat{\mathcal{S}}} Q(\chi)}_{\frac{\hbar\kappa}{2\pi k_B} S_W} = \underbrace{\int_{\mathcal{S}_\infty} [\delta Q(\xi) - \xi \cdot \theta(g, \delta g)]}_{\delta H_\xi} + \sum_{i=1}^L \Omega^{(i)} \underbrace{\int_{\mathcal{S}_\infty} [\delta Q(\eta_{(i)}) - \eta_{(i)} \cdot \theta(g, \delta g)]}_{\delta H_{\eta_{(i)}}},$$

where we have used Eq. (17.100) to let appear the Wald entropy S_W and Eq. (17.85) to let appear the variations of the Hamiltonians H_ξ and $H_{\eta_{(i)}}$ conjugate to ξ and $\eta_{(i)}$ respectively. Given that $\delta\kappa = 0$ (for κ in Eq. (17.100) is set to the surface gravity in the unperturbed spacetime, cf. footnote *a*, $\hbar\kappa/(2\pi k_B) = T_H$ [Eq. (16.62)], $H_\xi[g] = \mathcal{E}$ (Property 17.23) and $H_{\eta_{(i)}}[g] = -\mathcal{J}_{(i)}$ (Property 17.24), we get Eq. (17.101). \square

At first glance, Property 17.25 generalizes the first law of black hole thermodynamics obtained for general relativity in Chap. 16 to any diffeomorphism-invariant theory of gravity: compare Eq. (17.101) with Eq. (16.74), taking into account that, for general relativity, $\mathcal{E} = M$ (cf. Example 11 on p. 686). However, one should notice that δS_W in Eq. (17.101) is the variation of the entropy defined by Eq. (17.100), while in Eq. (16.74), δS_{BH} was the variation of entropy between two nearby *stationary* configurations. Given that for a stationary configuration, the Wald entropy $S_W =: S_W^{\text{eq}}$ is defined by Eq. (17.33), where κ is the surface gravity corresponding to that configuration and not of some reference configuration as in Eq. (17.100), one should prove that $\delta S_W = \delta S_W^{\text{eq}}$ to claim that Eq. (17.101) truly generalizes the first law (16.74). Specifically, let us use the same framework as in Chap. 16, namely the perturbed configuration is a stationary black hole, the horizon of which is a Killing horizon with respect to the Killing vector

$$\chi + \delta\chi_{\text{eq}} = \xi + \sum_{i=1}^L (\Omega^{(i)} + \delta\Omega^{(i)}) \eta_{(i)}, \quad (17.103)$$

and the corresponding surface gravity is $\kappa + \delta\kappa$. Note that we are using the same gauge freedom as in Chap. 16, namely the horizon is the same hypersurface of \mathcal{M} and the stationary and axisymmetric Killing vectors are held fixed: $\delta\xi = 0$ and $\delta\eta_{(i)} = 0$ [Eq. (16.4)]. We are using the subscript “eq” in $\delta\chi_{\text{eq}}$ to avoid any confusion with $\delta\chi = 0$ assumed in Property 17.25. We have then, from Eq. (17.33) with $\mathcal{S} = \hat{\mathcal{S}}$,

$$\delta S_W^{\text{eq}} = \frac{2\pi k_B}{\hbar(\kappa + \delta\kappa)} \int_{\hat{\mathcal{S}}} Q(g + \delta g, \chi + \delta\chi_{\text{eq}}) - \frac{2\pi k_B}{\hbar\kappa} \int_{\hat{\mathcal{S}}} Q(g, \chi).$$

Using the linearity of $Q(g, \chi)$ with respect to χ (cf. Property 17.6) and keeping the linear order in the perturbation, we get

$$\delta S_W^{\text{eq}} = \frac{2\pi k_B}{\hbar\kappa} \int_{\hat{\mathcal{S}}} [Q(g + \delta g, \chi) + Q(g, \delta\chi_{\text{eq}}) - Q(g, \chi)] - \frac{2\pi k_B \delta\kappa}{\hbar\kappa^2} \int_{\hat{\mathcal{S}}} Q(g, \chi).$$

Now, the difference between the first and third terms in the first integral corresponds exactly to δS_W with S_W defined by Eq. (17.100). Hence there comes

$$\delta S_W^{\text{eq}} = \delta S_W + \frac{2\pi k_B}{\hbar\kappa} \left[\int_{\hat{\mathcal{S}}} \mathbf{Q}(\delta\chi_{\text{eq}}) - \frac{\delta\kappa}{\kappa} \int_{\hat{\mathcal{S}}} \mathbf{Q}(\chi) \right],$$

where we have suppressed the explicit dependence of \mathbf{Q} on \mathbf{g} . From Eqs. (17.99) and (17.103), one has $\delta\chi_{\text{eq}} = \sum_{i=1}^L \delta\Omega^{(i)} \boldsymbol{\eta}_{(i)}$ (in agreement with Eq. (16.8)). Given that $\mathbf{Q}(\chi)$ is linear in χ , this yields $\mathbf{Q}(\delta\chi_{\text{eq}}) = \sum_{i=1}^L \delta\Omega^{(i)} \mathbf{Q}(\boldsymbol{\eta}_{(i)})$. On the other hand, we may express $\delta\kappa$ in terms of the $\delta\Omega^{(i)}$'s as well, thanks to Eq. (16.11), which is purely kinematic and thus independent of the gravity theory:

$$\delta\kappa = -\frac{1}{2A} \sum_{i=1}^L \delta\Omega^{(i)} \int_{\hat{\mathcal{S}}} \nabla_\mu \eta_{(i)\nu} dS^{\mu\nu}, \quad (17.104)$$

where A is the horizon area. Note that we have set to zero the term $I_{\mathcal{S}_k}$ in Eq. (16.11) since $\chi = 0$ on $\hat{\mathcal{S}}$ and $\nabla^\nu h_{\mu\nu} := \nabla^\nu \delta g_{\mu\nu}$ is assumed regular on $\hat{\mathcal{S}}$. Hence we obtain

$$\delta S_W^{\text{eq}} = \delta S_W + \frac{2\pi k_B}{\hbar\kappa} \sum_{i=1}^L \delta\Omega^{(i)} \left[\underbrace{\int_{\hat{\mathcal{S}}} \mathbf{Q}(\boldsymbol{\eta}_{(i)})}_{-J_{\hat{\mathcal{S}}}^{(i)}} + \underbrace{\frac{1}{2\kappa A} \int_{\hat{\mathcal{S}}} \nabla_\mu \eta_{(i)\nu} dS^{\mu\nu}}_{16\pi J_{\hat{\mathcal{S}}}^{(i)}} \times \underbrace{\int_{\hat{\mathcal{S}}} \mathbf{Q}(\chi)}_{\kappa A/(8\pi)} \right].$$

The values indicated below the underbraces are those obtained for general relativity, resulting respectively from Eqs. (17.32), (5.65) and (17.35). The terms between the square brackets then cancel each other, leaving us with $\delta S_W^{\text{eq}} = \delta S_W$. We thus conclude:

Property 17.26: first law for general relativity

For vacuum general relativity and for a variation between two stationary black holes, the first law (17.101) reduces to the “comparison” first law obtained in Property 16.2: Eq. (17.101) is identical to Eq. (16.22) with $K = 1$. Moreover, still within general relativity, Property 17.25 extends Property 16.2 to non-stationary variations $\delta\mathbf{g}$.

Remark 3: We have already noticed that the Hawking temperature T_H is independent of the gravity theory (cf. Remark 2 on p. 644). Thus, contrary to the entropy, there was no need to amend it to get the first law in a generic diffeomorphism-invariant gravity theory.

Remark 4: In the original work by Iyer & Wald [291], the black hole entropy in the perturbed configuration is defined by formula (17.36) instead of formula (17.100). Formula (17.36) makes sense for a perturbed black hole as well, provided that $\hat{\mathcal{S}}$ is the bifurcation surface in the unperturbed configuration. Let us denote by S'_W the entropy resulting from that definition, keeping S_W for (17.100). For a stationary black hole, both definitions are equivalent: $S_W = S'_W$ (Property 17.12), but, for a nonstationary one, this is not obvious. However, it has been proven by Iyer & Wald that $\delta S_W = \delta S'_W$ (cf. Eq. (94) in Ref. [291]), so that the first law (17.101) holds if δS_W is substituted by $\delta S'_W$ (cf. Eq. (93) in Ref. [291]). Given the definition of S'_W , it is obvious that, when considering two nearby stationary configurations,

$\delta S'_W^{\text{eq}} = \delta S'_W$, so that using Iyer & Wald's definition, Property 17.26 would have been immediate. We nevertheless stick to the definition (17.100) since it gives the quantity that naturally appears when deriving the first law (cf. the proof of Property 17.25), while the derivation for S'_W is more complicated, cf. the proof in Ref. [291]. Furthermore, the definition (17.100) is connected with the recently proposed entropy definitions for dynamical black holes [283, 505, 497], to be discussed below.

17.6 What about the second law?

We have seen above that, for theories distinct from general relativity, the black hole entropy is a priori not proportional to the black hole area. In particular, to show that $S_W \propto A$ in Property 17.11, we have used the value (17.23) of $Q(\chi)$, which is specific to general relativity. Thus even in theories for which the null convergence condition holds, the area theorem (Property 16.9) cannot be viewed as the second law of black hole thermodynamics.

Establishing a second law of black hole thermodynamics for diffeomorphism-invariant gravity theories requires first to identify some entropy candidate for dynamical black holes. The Wald entropy (17.100) cannot play this role, even for linear perturbations of a stationary black hole, for it requires to be evaluated at the bifurcation surface, which is a specific “time” in the black hole history, and which may even not exist for black holes born in gravitational collapse, as studied in Chap. 14. The entropy of dynamical black holes has to be defined for arbitrary cross-sections of the event horizon. In the proof of the first law (Property 17.25), sticking to the bifurcation surface was crucial to set to zero the term $\chi \cdot \theta(g, \delta g)$ in the left-hand side of Eq. (17.102), so that it becomes the variation of the Noether charge of χ and hence the variation of the Wald enthalpy. If the integration would have been performed on an arbitrary cross-section, this would a priori not have lead to an exact variation, contrary to what happens for the right-hand side of Eq. (17.102), where the integration over \mathcal{S}_∞ leads to exact variations, thanks to the existence of Hamiltonians conjugate to ξ and the $\eta_{(i)}$'s.

However, recently, Hollands, Wald and Zhang [283] (hereafter HWZ) have shown that, for *first-order perturbations δg of a stationary black hole*, one can define a $(n - 1)$ -form $B_{\mathcal{H}}$ on the event horizon \mathcal{H} , such the pullback $\iota^* \theta(g, \delta g)$ of $\theta(g, \delta g)$ on \mathcal{H} (i.e. the restriction of the $(n - 1)$ -form θ to vectors tangent to \mathcal{H}) writes $\iota^* \theta(g, \delta g) \stackrel{\mathcal{H}}{=} \delta B_{\mathcal{H}}$. This allows one to define the entropy of an arbitrary cross-section \mathcal{S} by

$$S_{\text{HWZ}} := \frac{2\pi k_B}{\hbar\kappa} \int_{\mathcal{S}} (Q(\chi) - \chi \cdot B_{\mathcal{H}}), \quad (17.105)$$

where κ is the surface gravity of the reference stationary black hole. For $\mathcal{S} = \hat{\mathcal{S}}$, this formula reduces to the Wald entropy expression (17.100), given that $\chi = 0$ on $\hat{\mathcal{S}}$. Moreover $B_{\mathcal{H}} = B_{\mathcal{H}}(g, \chi, \mathcal{L}_\chi g)$ and is linear with respect to $\mathcal{L}_\chi g$, so that $B_{\mathcal{H}} = 0$ on a stationary horizon and S_{HWZ} reduces then to S_W as defined by Eq. (17.33). In the proof of Property 17.25, if one use a hypersurface Σ with inner boundary \mathcal{S} , instead of $\hat{\mathcal{S}}$, one obtains readily the first law (17.101) with δS_{HWZ} substituted for δS_W . Given that δE and the $\delta J_{(i)}$'s are the same (for they depend only on values on \mathcal{S}_∞), it follows that $\delta S_{\text{HWZ}} = \delta S_W$ for first order *vacuum* perturbations of a stationary black hole. For non-vacuum perturbations, represented by an energy-momentum tensor δT , Hollands, Wald and Zhang [283] have shown that $\delta S_{\text{HWZ}}(\mathcal{S}_2) - \delta S_{\text{HWZ}}(\mathcal{S}_1) \geq 0$

if $\delta\mathbf{T}$ obeys the null energy condition (2.95) and \mathcal{S}_2 lies in the future of \mathcal{S}_1 . This constitutes a perturbative version of the second law.

For general relativity, the HWZ entropy does *not* reduce to the Bekenstein-Hawking entropy S_{BH} [Eq. (16.73)], except at the bifurcation surface $\hat{\mathcal{S}}$ of the unperturbed spacetime, since its writes

$$S_{\text{HWZ}} = S_{\text{BH}} + \frac{k_{\text{B}}}{4\hbar} \int_{\mathcal{S}} v\theta_{(\ell)} dS \quad (\text{general relativity}), \quad (17.106)$$

where v is an affine parameter along the null geodesic generators of \mathcal{H} such that $v = 0$ at $\hat{\mathcal{S}}$, $\ell := dx/dv$ is the associated tangent vector (ℓ is thus a null normal to \mathcal{H}), and $\theta_{(\ell)}$ is the expansion of \mathcal{H} along ℓ . (cf. Property 2.9). Note that the product $v\theta_{(\ell)}$ is independent of the choice of the affine parameter v . Indeed, for an affine parameter $v' := \alpha v$, one has $\ell' = \alpha^{-1}\ell$ and $\theta_{(\ell')} = \alpha^{-1}\theta_{(\ell)}$ (Property 2.10 with $\alpha \leftrightarrow \alpha^{-1}$), so that $v'\theta_{(\ell')} = v\theta_{(\ell)}$. We refer to [505, 497] for extended discussions about the HWZ entropy, especially in connection with the extension of the generalized second law (GSL, cf. Property 16.15) beyond general relativity.

Other proposals for the entropy of dynamical black holes in diffeomorphism-invariant theories of gravity have been put forward in the literature, notably by Dong [170] and Wall [510]. As for HWZ, they regard perturbations of a stationary black hole (see Ref. [283, 497] for a comparison with HWZ entropy). Recently, a non-perturbative approach has been introduced by Davies & Reall [157], leading to a second law for effective field theories.

Chapter 18

The quasi-local approach: trapping horizons

Contents

18.1 Introduction	693
18.2 Trapped surfaces and singularity theorems	693
18.3 Trapping horizons	705

18.1 Introduction

This chapter is in a draft stage.

18.2 Trapped surfaces and singularity theorems

18.2.1 Trapped surfaces

The concept of trapped surfaces has been introduced in Sec. 3.2.3 (see in particular Fig. 3.1). Let us recall that a submanifold \mathcal{S} of a n -dimensional spacetime (\mathcal{M}, g) ($n \geq 3$) is a **trapped surface** iff (i) \mathcal{S} is a compact $(n - 2)$ -dimensional manifold (without boundary), (ii) \mathcal{S} is spacelike (positive definite induced metric) and the two systems of null geodesics emerging orthogonally from \mathcal{S} towards the future locally converge, i.e. they have negative expansions at \mathcal{S} : $\theta_{(\ell)} < 0$ and $\theta_{(k)} < 0$ [Eq. (3.5)], ℓ and k being any future-directed vectors tangent to these geodesics.

Remark 1: Trapped surfaces are sometimes called *closed trapped surfaces* (e.g. [404, 265]), to stress their closed manifold aspect (compact without boundary). We follow here the textbooks [370, 498] and call them merely *trapped surfaces*.

Example 1 (trapped surfaces in the Schwarzschild spacetime): Let (\mathcal{M}, g) be the Schwarzschild spacetime as defined in Sec. 6.3.6 [Eq. (6.54)]; it is entirely covered by the ingoing Eddington-Finkelstein

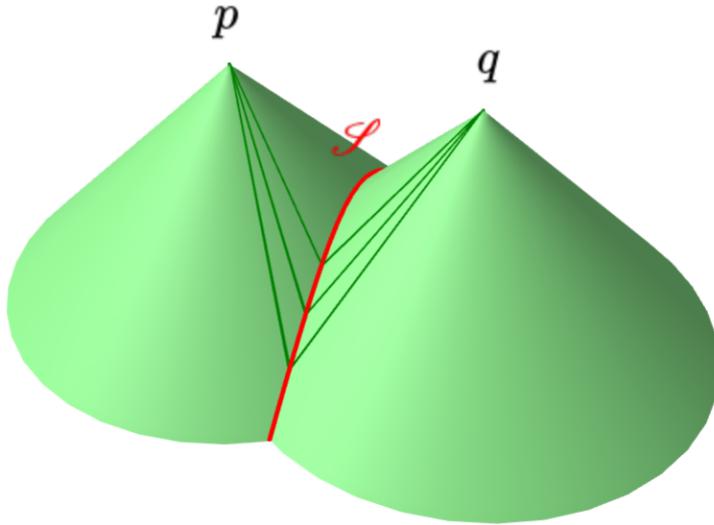


Figure 18.1: Intersection \mathcal{S} (red curve) of the past light cones of two points p and q in Minkowski spacetime. Some light rays emerging from \mathcal{S} in the two orthogonal null directions are depicted; both sets of light rays are converging. Yet \mathcal{S} is not trapped because it is not compact. [Figure generated by the notebook D.6.7]

coordinates $(\tilde{t}, r, \theta, \varphi)$. Let \mathcal{S} be any surface $(\tilde{t}, r) = \text{const}$. \mathcal{S} is diffeomorphic to \mathbb{S}^2 and thus compact. Moreover, from Eq. (6.33), the metric induced by \mathbf{g} on \mathcal{S} is $q = r^2(\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2)$, which is clearly positive definite; hence \mathcal{S} is spacelike. One also reads on the expression of q that the area of \mathcal{S} is simply $A = 4\pi r^2$ (i.e. r is the areal radius, cf. Sec. 6.2.2). If \mathcal{S} is located in the black hole interior, i.e. if $0 < r < 2m$, Property 6.9 implies that A is decreasing along any future-directed null geodesic. It follows that \mathcal{S} is trapped. On the contrary, if located in the black hole exterior ($r > 2m$), \mathcal{S} is untrapped. This can be checked on Fig. 6.3, where it appears clearly that, in the region $r > 2m$, r is increasing along the outgoing radial null geodesics, which are normal to \mathcal{S} .

Example 2 (trapped surfaces in the Oppenheimer-Snyder collapse): Spherically symmetric trapped surfaces in the Oppenheimer-Snyder collapse have been investigated in Sec. 14.3.6 (in particular, see Fig. 14.8): in terms of the FLRW coordinates $(\eta, \chi, \theta, \varphi)$ spanning the collapsing star [cf. Eq. (14.45)], they are all the spheres $(\eta, \chi) = \text{const}$ such that $\eta > \pi - 2\chi$ (Property 14.7). Outside the collapsing star, i.e. in the Schwarzschild region [cf. Eq. (14.28)], they are all the spheres $(\tilde{t}, r) = \text{const}$ with $r < 2m$ (cf. Example 1).

Example 3 (trapped surfaces in the Vaidya collapse): Spherically symmetric trapped surfaces in the Vaidya radiation shell collapse have been investigated in Sec. 15.3.5: in terms of the IEF coordinates (t, r, θ, φ) [cf. Eq. (15.11)], they are the spheres $(t, r) = \text{const}$ such that $r < 2M(t+r)$ (Property 15.10). The region containing them is depicted as red dot-filled in the Carter-Penrose diagrams of Figs. 15.7, 15.11 and 15.12.

The existence of a trapped surface characterizes strong gravity. More precisely it indicates that gravity is strong enough to focus all light rays emitted from the surface. This characterization is *quasi-local*, in the sense that it involves a *surface*. The surface can be “small” but it cannot be reduced to a point, to which the qualifier *local* would apply. Note that the hypothesis of *compact* manifold is crucial in the definition of a trapped surface. Without it, trapped surfaces would exist in Minkowski spacetime, as the following example shows.

Example 4 (a counter-example in Minkowski spacetime): Let us consider the intersection \mathcal{S} of the past light cones \mathcal{C}_p and \mathcal{C}_q of two points p and q in Minkowski spacetime (cf. Fig. 18.1). Being a cross-section of the null hypersurface \mathcal{C}_p (or \mathcal{C}_q), \mathcal{S} is a spacelike surface (Property 2.2). Moreover the null directions \mathbf{k} and ℓ normal to it are given by the null generators of \mathcal{C}_p (since \mathcal{S} is a cross-section of \mathcal{C}_p) and of \mathcal{C}_q (since \mathcal{S} is a cross-section of \mathcal{C}_q). Because \mathcal{C}_p and \mathcal{C}_q are past light cones, one has clearly $\theta_{(\ell)} < 0$ and $\theta_{(\mathbf{k})} < 0$ (cf. Fig. 18.1). However, \mathcal{S} fails to be a trapped surface for it is not compact. Truncating the light cones would not help, because this would make \mathcal{S} a manifold with boundary and hence not a closed one.

It is worth noticing that some (peculiar) black hole spacetimes do not contain any trapped surface in the strict sense ($\theta_{(\mathbf{k})} < 0$ and $\theta_{(\ell)} < 0$), but only *marginally trapped surfaces* ($\theta_{(\mathbf{k})} < 0$ and $\theta_{(\ell)} = 0$). This is demonstrated by the following example:

Example 5 (no trapped surface in the extremal Reissner-Nordström spacetime): The *extremal Reissner-Nordström black hole* is defined by setting $Q = \sqrt{4\pi/\mu_0} M$ and $P = 0$ in Eq. (5.107). Performing the change of notation $M \rightarrow m$, this yields the metric

$$\mathbf{g} = - \left(1 - \frac{m}{r}\right)^2 dt^2 + \left(1 - \frac{m}{r}\right)^{-2} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2). \quad (18.1)$$

The coordinates (t, r, θ, φ) are singular at $r = m$. One can introduce Eddington-Finkelstein-like coordinates $(\tilde{t}, r, \theta, \varphi)$ via

$$\tilde{t} := t + 2m \ln \left| \frac{r}{m} - 1 \right| + \frac{m}{1 - r/m} \quad (18.2)$$

(compare with Eq. (6.32) in the Schwarzschild case) to get a regular coordinate system in the whole range $r \in (0, +\infty)$. The hypersurface \mathcal{H} defined by $r = m$ appears then as a black hole event horizon. \mathcal{H} is a *degenerate* Killing horizon with respect to the Killing vector $\xi = \partial_t = \partial_{\tilde{t}}$ (cf. Sec. 3.3.6) (Exercise: prove it!). This is similar to the event horizon of the extremal Kerr spacetime (cf. Chap. 13), which is a degenerate Killing horizon as well (Property 13.3). From the metric (18.1), it appears that the vector field ∂_t is timelike in both regions $r < m$ (black hole interior) and $r > m$ (black hole exterior), since $g_{tt} < 0$ there. Moreover, for the standard time orientation given by $\partial_{\tilde{t}}$, ∂_t is future-directed in these two regions. Let us consider a 2-sphere \mathcal{S} defined by $(t, r) = \text{const}$ with $r \neq m$. The two families of null geodesics \mathcal{L}_{\pm} normal to \mathcal{S} are ruled by the equations $(\theta, \varphi) = \text{const}$ and

$$\frac{dr}{dt} \Big|_{\mathcal{L}_{\pm}} = \pm \left(1 - \frac{m}{r}\right)^2. \quad (18.3)$$

This last equation is deduced from Eq. (18.1) by setting $ds^2 := \mathbf{g}(dx, dx) = 0$ with $dx = dt \partial_t + dr \partial_r$ along \mathcal{L}_{\pm} . One reads on the metric (18.1) that the area of \mathcal{S} is $A = 4\pi r^2$, as in the Schwarzschild case (Example 1). Now, since ∂_t is future-directed timelike, it appears from Eq. (18.3) that r increases (resp. decreases) towards the future along \mathcal{L}_+ (resp. \mathcal{L}_-), in both regions $r < m$ and $r > m$. It follows immediately that \mathcal{S} is not trapped. Actually, there exists only *marginally* trapped surfaces in this spacetime: the cross-sections of the event horizon \mathcal{H} .

18.2.2 Penrose's singularity theorem

Trapped surfaces are the key ingredient of Penrose's singularity theorem, which basically states that, under some rather generic assumptions, if gravity is strong enough so that a trapped

surface occurs, then some singularity will appear in the future of it. Before stating the theorem, let us recall a few concepts on which it relies.

First of all, a **Cauchy surface** Σ of a spacetime (\mathcal{M}, g) is a spacelike hypersurface such that every inextendible timelike curve of \mathcal{M} intersects Σ exactly once (cf. Sec. 5.2.3). If (\mathcal{M}, g) admits a Cauchy surface, it is said to be a **globally hyperbolic** spacetime. On such a spacetime, general relativity can be formulated as a well posed Cauchy problem [110], i.e. there exists a unique solution g of the Einstein equation in \mathcal{M} from initial data prescribed on Σ , provided that the initial data fulfill four components of the Einstein equation known as the *constraint equations* (see e.g. Ref. [226]).

The second concept involved in Penrose's theorem is that of an *inextendible incomplete* geodesic. As defined in Appendix B (Sec. B.3.2), a geodesic \mathcal{L} of a spacetime (\mathcal{M}, g) is **incomplete** if some affine parameter λ does not span the whole of \mathbb{R} along \mathcal{L} . Given that any two affine parameters are related by an affine transformation, if this holds for a given affine parameter, this holds for all. More precisely, if \mathcal{L} is a causal (i.e. timelike or null) geodesic and λ is increasing to the future, one says that \mathcal{L} is **future-incomplete** if, and only if, $\lambda < \lambda_{\max}$ along \mathcal{L} , for some $\lambda_{\max} \in \mathbb{R}$. Similarly, one says that \mathcal{L} is **past-incomplete** if, and only if, $\lambda > \lambda_{\min}$ along \mathcal{L} , for some $\lambda_{\min} \in \mathbb{R}$. Furthermore, a geodesic \mathcal{L} is said **inextendible** if, and only if, there does not exist any geodesic \mathcal{L}' of (\mathcal{M}, g) distinct from \mathcal{L} and such that $\mathcal{L} \subset \mathcal{L}'$. An inextendible incomplete causal geodesic marks the existence of a singularity of some kind. For instance a free-falling observer having an inextendible future-incomplete timelike geodesic as worldline has his proper time¹ abruptly stopping at some finite value! Actually, the existence of an inextendible incomplete geodesic is the nowadays adopted definition of a **spacetime singularity**. We shall discuss the connection with the concept of *curvature singularity* introduced in Sec. 6.3.4 later on. For the moment, we have enough material to state the famous theorem:

Property 18.1: Penrose's singularity theorem (Penrose 1965 [404])

Let (\mathcal{M}, g) be a n -dimensional time-orientable spacetime ($n \geq 3$) such that

- (\mathcal{M}, g) admits a non-compact Cauchy surface Σ ;
- the null convergence condition (2.94) holds on \mathcal{M} : for any null vector ℓ , $\mathbf{R}(\ell, \ell) \geq 0$, where \mathbf{R} is g 's Ricci tensor – for general relativity, such a condition is equivalent to the null energy condition (2.95): $\mathbf{T}(\ell, \ell) \geq 0$;
- there exists a trapped surface \mathcal{S} .

Then at least one inextendible null geodesic emerging orthogonally from \mathcal{S} is future-incomplete.

Proof. We shall prove the theorem by contradiction. Hence let us assume that all inextendible null geodesics emerging orthogonally from \mathcal{S} are future-complete. Let \mathcal{F} be the boundary

¹Recall that the proper time is an affine parameter along a timelike geodesic (Property B.7).

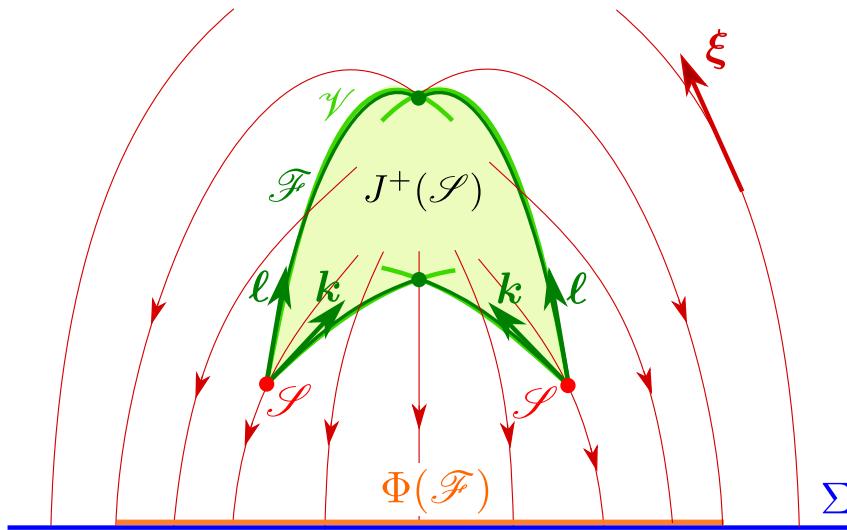


Figure 18.2: Spacetime diagram representing various objects involved in the proof of Penrose's singularity theorem (Property 18.1): the trapped surface \mathcal{S} (the two red dots), the two null directions ℓ and k normal to \mathcal{S} , the causal future of \mathcal{S} , $J^+(\mathcal{S})$ (pale green region), the boundary \mathcal{F} of $J^+(\mathcal{S})$ (dark green curve), the compact space \mathcal{V} containing \mathcal{F} (light green curve) and the image $\Phi(\mathcal{F})$ of \mathcal{F} (orange segment) into the Cauchy surface Σ (blue line) by the topological embedding Φ constructed from the field lines of a global timelike vector ξ (maroon lines). The two dark green dots are crossover points, where null geodesics leave \mathcal{F} to the interior of $J^+(\mathcal{S})$. The compact $(n - 2)$ -dimensional surface \mathcal{S} appears disconnected and reduced to two points because this 2D figure is the cut by a vertical plane Π of the 3D view of Fig. 3.1, where \mathcal{S} is drawn as an horizontal circle, whose intersection with Π is two points.

of the causal future of \mathcal{S} (cf. Sec. 4.4.1 and Fig. 18.2): $\mathcal{F} := \partial J^+(\mathcal{S})$. Since (\mathcal{M}, g) is globally hyperbolic (it admits Σ as a Cauchy surface), it is necessarily causally simple (cf. Proposition 6.6.1 of HE [265]). By the very definition of *causally simple* (cf. Sec. 4.4.1) and the compacity of \mathcal{S} , it follows that \mathcal{F} coincides with the future horismos of \mathcal{S} : $\mathcal{F} = E^+(\mathcal{S})$. This implies that points of \mathcal{F} are connected to \mathcal{S} by null geodesics. More precisely, we note that \mathcal{F} is an *achronal boundary*, namely the boundary of the causal future or past of a given set. As mentioned in Sec. 4.4.3 (Remark 6), the properties of achronal boundaries are the same as those established for a black hole event horizon $\mathcal{H} = \partial J^-(\mathcal{S}^+) \cap \mathcal{M}$ (Properties 4.5 to 4.11), provided one changes *future* to *past* when appropriate. It follows that \mathcal{F} is an achronal $(n - 1)$ -dimensional topological manifold (Properties 4.5 and 4.7) that is ruled by null geodesics, called the *generators* of \mathcal{F} and obeying the following properties: when followed in the past direction, a generator never leaves \mathcal{F} until it encounters \mathcal{S} and there is a unique generator through each point of \mathcal{F} , except at special points, named *crossovers*, where the generators leave \mathcal{F} in the future direction, i.e. are future-extended to null geodesics of \mathcal{M} that do not belong² to \mathcal{F} (Property 4.8). Note that two crossovers are shown in Fig. 18.2.

We actually need an additional property of \mathcal{F} , which has not been established in Sec. 4.4.3: on \mathcal{S} (which is part of \mathcal{F} , cf. Sec. 4.4.1), the generators of \mathcal{F} are orthogonal to \mathcal{S} (cf. e.g. Theorem 9.3.11 in Wald's textbook [498]). They therefore start from \mathcal{S} along the two null

²They actually belong to the interior of $J^+(\mathcal{S})$, i.e. to the chronological future $I^+(\mathcal{S})$ [Eq. (4.33)] (see e.g. Proposition 4.5.14 in HE [265]).

directions ℓ and k normal to \mathcal{S} (cf. Fig. 18.2). Let us denote by \mathcal{L} a generic null generator along ℓ and by λ the affine parameter of \mathcal{L} such that, on \mathcal{S} , $\lambda = 0$ and $dx/d\lambda|_{\mathcal{L}} = \ell$. We shall use the latter relation to extend the definition of ℓ along \mathcal{L} away from \mathcal{S} . Similarly, let us denote by $\bar{\mathcal{L}}$ a generic null generator along k and by $\bar{\lambda}$ the affine parameter of $\bar{\mathcal{L}}$ such that $\bar{\lambda} = 0$ on \mathcal{S} and $dx/d\bar{\lambda}|_{\bar{\mathcal{L}}} = k$. By hypothesis, \mathcal{L} and $\bar{\mathcal{L}}$ are future-complete. Since \mathcal{S} is a trapped surface, $\theta_0 := \theta_{(\ell)}|_{\mathcal{S}} < 0$ and $\bar{\theta}_0 := \theta_{(k)}|_{\mathcal{S}} < 0$. Then, the null Raychaudhuri equation (2.88) applied to the null hypersurface generated by the null geodesics of the \mathcal{L} family and the null convergence condition $R(\ell, \ell) \geq 0$ implies that there exists $\lambda_* \in (0, (n-2)/|\theta_0|]$ such that $\theta_{(\ell)} \rightarrow -\infty$ for $\lambda \rightarrow \lambda_*$ along \mathcal{L} . The reasoning is exactly the same at that used in the proof of Property 16.8 and we shall not repeat it here. The point $\lambda = \lambda_*$ on \mathcal{L} is a caustic point, where all nearby geodesics starting from \mathcal{S} along ℓ converge. Given the properties of \mathcal{F} , we conclude that for $\lambda > \lambda_*$, \mathcal{L} ceases to be a null generator of \mathcal{F} , i.e. $\mathcal{L}_{\lambda > \lambda_*} \notin \mathcal{F}$. Similarly, along $\bar{\mathcal{L}}$, $\theta_{(k)} \rightarrow -\infty$ for $\bar{\lambda} \rightarrow \bar{\lambda}_*$ with $\bar{\lambda}_* \in (0, (n-2)/|\bar{\theta}_0|]$, so that $\bar{\mathcal{L}}_{\bar{\lambda} > \bar{\lambda}_*} \notin \mathcal{F}$. Let

$$\lambda_{\max} := \sup_{p \in \mathcal{S}} \left(\frac{n-2}{|\theta_{(\ell)}(p)|} \right) \quad \text{and} \quad \bar{\lambda}_{\max} := \sup_{p \in \mathcal{S}} \left(\frac{n-2}{|\theta_{(k)}(p)|} \right).$$

Since \mathcal{S} is compact, λ_{\max} and $\bar{\lambda}_{\max}$ are finite. Let us then consider the map $f : \mathcal{S} \times [0, \lambda_{\max}] \rightarrow \mathcal{M}, (p, \lambda) \mapsto q$, where q is the point of affine parameter λ on the null geodesic of the \mathcal{L} -family through p . Since all the \mathcal{L} geodesics are assumed to be future-complete, f is well defined. Moreover, f is clearly continuous. Then, given that $\mathcal{S} \times [0, \lambda_{\max}]$ is a compact set, the image set $f(\mathcal{S} \times [0, \lambda_{\max}])$ is necessarily compact. Similarly, the image set $\bar{f}(\mathcal{S} \times [0, \bar{\lambda}_{\max}])$ is compact, where $\bar{f} : \mathcal{S} \times [0, \bar{\lambda}_{\max}] \rightarrow \mathcal{M}, (p, \lambda) \mapsto \bar{q}$ — the point of affine parameter $\bar{\lambda}$ on the null geodesic of the $\bar{\mathcal{L}}$ -family through p . The set

$$\mathcal{V} := f(\mathcal{S} \times [0, \lambda_{\max}]) \cup \bar{f}(\mathcal{S} \times [0, \bar{\lambda}_{\max}])$$

is then a compact subset of \mathcal{M} , as the union of two such sets. By construction, for each $p \in \mathcal{S}$, the point $q = f(p, \lambda_{\max})$ (resp. $\bar{q} = \bar{f}(p, \lambda_{\max})$) lies beyond the caustic point on the geodesic \mathcal{L} (resp. $\bar{\mathcal{L}}$) through p . It follows that $\mathcal{F} \subset \mathcal{V}$ (cf. Fig. 18.2). Given that \mathcal{F} is closed, being a topological boundary ($\mathcal{F} := \partial J^+(\mathcal{S})$), and \mathcal{V} is compact, we conclude that \mathcal{F} is compact.

Because (\mathcal{M}, g) is time-orientable, there exists a nonvanishing timelike vector field ξ on \mathcal{M} . The field lines of ξ form a congruence of timelike curves \mathcal{C} , each of them intersecting Σ in a single point, for Σ is a Cauchy surface. We may then define a map $\Phi : \mathcal{F} \rightarrow \Sigma, p \mapsto \mathcal{C}_p \cap \Sigma$, where \mathcal{C}_p is the unique timelike curve of the \mathcal{C} congruence through p (cf. Fig. 18.2). The map Φ is clearly continuous. Moreover, it is injective given that \mathcal{F} is achronal (if a point $q \in \Sigma$ was the image of two distinct points p and p' of \mathcal{F} , this would mean that p and p' are connected by a timelike curve, namely a segment of \mathcal{C}_p). Now, there cannot be any injective continuous map from a compact manifold (without boundary) (here \mathcal{F}) into a non-compact manifold of the same dimension (here Σ). For instance, there is no injective continuous map from the sphere \mathbb{S}^2 into the plane \mathbb{R}^2 . To show this impossibility in the present case, and hence to reach the sought contradiction, let us invoke a standard topology theorem (see e.g. Lemma 4.50 in Lee's textbook [341]), according to which an injective continuous map from a compact space to a Hausdorff space is a topological embedding, i.e. a homeomorphism onto its image. This theorem applies here since Σ is Hausdorff, being a manifold (cf. Sec. A.2.1). Hence

$\Phi(\mathcal{F})$ is homeomorphic to \mathcal{F} . $\Phi(\mathcal{F})$ is therefore a compact subset of Σ . Another standard result of topology (see e.g. Proposition 4.36b in Lee's textbook [341]) states that a compact subset of a Hausdorff space is closed. Hence $\Phi(\mathcal{F})$ is a closed subset of Σ . On the other side, since \mathcal{F} is a $(n - 1)$ -dimensional topological manifold, each point of \mathcal{F} has a neighborhood homeomorphic to an open subset of \mathbb{R}^{n-1} . The same feature holds then for $\Phi(\mathcal{F})$. Since Σ is a $(n - 1)$ -dimensional manifold, it follows that $\Phi(\mathcal{F})$ is an open subset of Σ . Hence $\Phi(\mathcal{F})$ is both open and closed in Σ . Given that Σ is connected, being a Cauchy surface in a connected spacetime, this implies that $\Phi(\mathcal{F}) = \Sigma$. Here we reach a contradiction because Σ is assumed non-compact. \square

It is worth to stress that Penrose's singularity theorem is *not* telling that a *curvature* singularity, as described in Sec. 6.3.4, must exist in spacetime (or at some “boundary” of it, since curvature singularities are usually excluded from the spacetime manifold). The theorem only stipulates that some inextendible null geodesic is future-incomplete. However, a good reason for a null geodesic to be future-incomplete is to hit a curvature singularity as the examples below are going to illustrate.

Another cause of incompleteness could be that the geodesic, \mathcal{L} say, has reached some regular “external boundary” \mathcal{H} of the spacetime (\mathcal{M}, g) and both \mathcal{M} and g can be smoothly extended beyond this boundary, i.e. there exists a spacetime $(\tilde{\mathcal{M}}, \tilde{g})$ such that $\mathcal{M} \subset \tilde{\mathcal{M}}$ and $\tilde{g}|_{\mathcal{M}} = g$. The boundary \mathcal{H} , which is a part of $\tilde{\mathcal{M}}$, not of \mathcal{M} (for \mathcal{M} is a manifold without boundary), is then the *future Cauchy horizon* of the hypersurface Σ in $\tilde{\mathcal{M}}$ (cf. Sec. 10.8.3). Note that since Σ is a Cauchy surface of (\mathcal{M}, g) , Σ has no Cauchy horizon within \mathcal{M} . With respect to $(\tilde{\mathcal{M}}, \tilde{g})$, Σ is only a *partial* Cauchy surface (cf. Sec. 10.8.3 and Fig. 10.12). The geodesic \mathcal{L} , which was inextendible within \mathcal{M} , can then be extended beyond \mathcal{H} in a possibly complete geodesic of $(\tilde{\mathcal{M}}, \tilde{g})$ and there might be no curvature singularity at all in the spacetime $(\tilde{\mathcal{M}}, \tilde{g})$.

Example 6 (Schwarzschild spacetime): Let us consider the maximally extended Schwarzschild spacetime (\mathcal{M}, g) discussed in Chap. 9. The hypersurface Σ defined by $T = 0$ in terms of the Kruskal-Szekeres coordinates (T, X, θ, φ) is a non-compact Cauchy surface of (\mathcal{M}, g) (cf. Fig. 9.8 or Fig. 9.11). The null convergence condition is trivially fulfilled since the Ricci tensor \mathbf{R} is identically zero, g being a solution of the vacuum Einstein equation (1.44). As detailed in Example 1 (p. 693), any sphere \mathcal{S} defined by $(\tilde{t}, r) = \text{const}$ is a trapped surface in the region $r < 2m$ of the ingoing Eddington-Finkelstein patch, i.e. the region \mathcal{M}_{II} . All the hypotheses of Penrose's singularity theorem are then fulfilled, so there must exist an incomplete null geodesic emerging from \mathcal{S} . Actually, all null geodesics emerging orthogonally from \mathcal{S} are incomplete, for they all reach the curvature singularity $r = 0$ within a finite affine parameter, as it appears clearly in Fig. 9.7 (remember that r is an affine parameter along these geodesics, cf. Property 6.3).

Example 7 (Oppenheimer-Snyder collapse): The gravitational collapse of a homogeneous pressureless ball, as described by the Oppenheimer-Snyder model discussed in Sec. 14.3, provides an example of a dynamical spacetime illustrating Penrose's theorem. The hypersurface Σ defined by $\tilde{t} = 0$ in terms of the global IEF coordinates $(\tilde{t}, r, \theta, \varphi)$ is a non-compact Cauchy surface of the underlying spacetime (\mathcal{M}, g) (cf. the Carter-Penrose diagram of Fig. 14.4). Since g is a solution of the Einstein equation, the null convergence condition reduces to the null energy condition: $\mathbf{T}(\ell, \ell) \geq 0$ for any null vector ℓ . Given the form (14.2) of the energy-momentum tensor \mathbf{T} , we get, inside the collapsing star, $\mathbf{T}(\ell, \ell) = \rho(\mathbf{u} \cdot \ell)^2 > 0$ since $\rho > 0$, while $\mathbf{T}(\ell, \ell) = 0$ outside the star. Hence the null convergence condition is fulfilled. Finally, as detailed in Example 2 (p. 694), this spacetime contains trapped surfaces. All the hypotheses of

Penrose's singularity theorem are thus satisfied. Let us check its predictions on the two families \mathcal{L} and $\bar{\mathcal{L}}$ of null geodesics emanating orthogonally from a trapped sphere \mathcal{S} that is defined in terms of the conformal coordinates $(\eta, \chi, \theta, \varphi)$ by $(\eta, \chi) = (\eta_0, \chi_0) = \text{const}$, with $\pi - \eta_0 < \chi_0 < \chi_s - (\pi - \eta_0)$. This constraint implies that both \mathcal{L} and $\bar{\mathcal{L}}$ stay in the interior of the collapsing star until they reach the curvature singularity $\eta = \pi$ (cf. the conformal diagram in Fig. 14.2 left). Let us check that this happens for a finite value of their affine parameters. The equation of these geodesics is easily deduced from the conformal form (14.45) of the metric tensor: we get $\eta - \eta_0 = \pm(\chi - \chi_0)$ with $\pm = +$ for \mathcal{L} and $\pm = -$ for $\bar{\mathcal{L}}$. Moreover, η is an affine parameter along both \mathcal{L} and $\bar{\mathcal{L}}$ with respect to the conformal metric $\mathbf{h} := -d\eta^2 + d\chi^2 + \sin^2 \chi (d\theta^2 + \sin^2 \theta d\varphi^2)$. According to Eq. (4.13) any affine parameter λ of \mathcal{L} or $\bar{\mathcal{L}}$ with respect to the physical metric \mathbf{g} is related to the \mathbf{h} -affine parameter η by $d\lambda/d\eta = a\Omega^2$, where a is a positive constant and the conformal factor Ω is read on Eq. (14.45): $\Omega^2 = (a_0/2)^2(1 + \cos \eta)^2$. The integration leads to

$$\lambda = \alpha(6\eta + 8\sin \eta + \sin 2\eta) + \beta,$$

where α and β are two constants such that $\alpha > 0$. We have then $\lim_{\eta \rightarrow \pi} \lambda = 6\alpha\pi + \beta$, which is obviously finite. We conclude that both \mathcal{L} and $\bar{\mathcal{L}}$ are future-incomplete.

Example 8 (Vaidya collapse): Let us consider the collapse of a radiation shell as studied in Sec. 15.3, assuming that the shell is thin (large radiation density): $\alpha := 2m/v_0 > 1/8$ [cf. Eqs. (15.18) and (15.23)]. As it can be inferred from the Carter-Penrose diagram of Fig. 15.7, a hypersurface $t = \text{const}$ in the black hole exterior is a non-compact Cauchy surface (for instance the hypersurface $t = -3m$ in the $\alpha = 2/3$ model considered in Fig. 15.1). Besides, we have already noticed in Sec. 15.2.1 that the null energy condition is fulfilled, for $M(v)$ is an increasing function (cf. Fig. 15.2). Finally, as detailed in Example 3 (p. 694), trapped surfaces forms in the late stages of the collapse. Penrose's singularity theorem thus applies and there should be an incomplete inextendible null geodesic emerging from each trapped sphere \mathcal{S} described in Example 3. It is easy to show that this is indeed the case for all ingoing null geodesics³ $\bar{\mathcal{L}} = \mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$, i.e. geodesics emerging from \mathcal{S} along the vector \mathbf{k} normal to the null hypersurfaces $v = \text{const}$ (cf. Property 15.1). As it is clear on Fig. 15.7, all the geodesics $\mathcal{L}_{(v,\theta,\varphi)}^{\text{in}}$ terminate at the curvature singularity $r = 0, v > 0$. Since $-r$ is an affine parameter along them (Property 15.4), we conclude that they are future-incomplete.

Example 9 (Bardeen regular black hole as a counter-example): There exist solutions of the Einstein equation (with some rather exotic matter sources) corresponding to a black hole without any curvature singularity. They are called **regular black holes** (see e.g. the recent book [30]). A well known example (historically the first one) is the **Bardeen black hole** [34] (see also Sec. V of Ref. [68]). The underlying spacetime $(\mathcal{M}, \mathbf{g})$ is static and spherically symmetric; in some part of it, the metric is expressed in terms of the Schwarzschild-Droste-like coordinates (t, r, θ, φ) by

$$\mathbf{g} = -\left(1 - \frac{2M(r)}{r}\right) dt^2 + \left(1 - \frac{2M(r)}{r}\right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \quad (18.4)$$

where

$$M(r) := m \frac{r^3}{(r^2 + \ell^2)^{3/2}}, \quad (18.5)$$

m and ℓ being two positive constants obeying $\ell < 4m/(3\sqrt{3})$. Would $M(r)$ be a constant function, one would get the Schwarzschild metric (6.14). Actually the Schwarzschild metric is recovered asymptotically,

³All outgoing null geodesics from \mathcal{S} are also incomplete, but this is harder to show for there is no simple expression of an affine parameter along them.

since $M(r) \rightarrow m$ for $r \rightarrow +\infty$. For small values of r , one has instead

$$M(r) \underset{r \rightarrow 0}{\sim} \frac{m}{\ell^3} r^3. \quad (18.6)$$

This implies that \mathbf{g} is regular at $r = 0$, contrary to the Schwarzschild metric. More precisely, setting $\Lambda := 6m/\ell^3$, we get

$$\mathbf{g} \underset{r \rightarrow 0}{\sim} -\left(1 - \frac{\Lambda}{3} r^2\right) \mathbf{dt}^2 + \left(1 - \frac{\Lambda}{3} r^2\right)^{-1} \mathbf{dr}^2 + r^2 (\mathbf{d}\theta^2 + \sin^2 \theta \mathbf{d}\varphi^2). \quad (18.7)$$

We recognize in the right-hand side the metric of de Sitter spacetime of cosmological constant Λ expressed in the so-called static coordinates (see e.g. Sec. 4.3 of Ref. [242]). On the other side, the metric components (18.4) are singular at values of r for which $2M(r) = r$. This equation admits two solutions: $r = r_{\pm}$ with $0 < r_- < r_+$. By introducing coordinates $(\tilde{t}, r, \theta, \varphi)$ of Eddington-Finkelstein type, one can show that \mathbf{g} is regular at $r = r_{\pm}$. The hypersurfaces $r = r_{\pm}$ turn out to be two Killing horizons; the outermost one ($r = r_+$) is the black hole event horizon, while the innermost one is a Cauchy horizon in the maximally extended spacetime (see Fig. 1 of Ref. [15]). The Bardeen metric \mathbf{g} is a solution of the Einstein equation (1.40) sourced by a magnetic monopole in some nonlinear electrodynamics [28]. By **nonlinear electrodynamics**, it is meant a theory involving an electromagnetic 2-form \mathbf{F} and for which the Lagrangian density \mathcal{L} is a nonlinear function of the electromagnetic invariant $\mathcal{F} := F_{\mu\nu}F^{\mu\nu}/4$, the standard Maxwell theory corresponding to the linear case: $\mathcal{L} = -\mathcal{F}/\mu_0$. Specifically, for the Bardeen black hole, the nonlinear electrodynamics is defined by [28]

$$\mathcal{L} = -\frac{3m}{\mu_0 |\ell| \ell^2} \left(\frac{\sqrt{2\ell^2 \mathcal{F}}}{1 + \sqrt{2\ell^2 \mathcal{F}}} \right)^{5/2} \quad (18.8)$$

and the electromagnetic field is

$$\mathbf{F} = \frac{\mu_0}{4\pi} \ell \sin \theta \mathbf{d}\theta \wedge \mathbf{d}\varphi, \quad (18.9)$$

which corresponds to a magnetic monopole ℓ located at the center of symmetry (cf. Remark 2 on p. 172).

The energy-momentum tensor⁴ \mathbf{T} derived from the Lagrangian density (18.8) fulfills the weak energy condition (2.96), which implies the null energy condition and thus the null convergence condition. Moreover, the Bardeen spacetime contains trapped surfaces in the region $r_- < r < r_+$. Yet all inextendible null geodesics are complete. The Bardeen spacetime evades Penrose's singularity theorem because it does not admit any non-compact Cauchy surface (actually no Cauchy surface at all). More precisely, in the internal region it admits a *partial* Cauchy surface Σ to which the boundary \mathcal{F} of the causal future of a given trapped surface \mathcal{S} can be mapped injectively. However Σ is compact, having the topology of \mathbb{S}^3 (see [69] for details), so that one cannot reach the contradiction completing the proof of Penrose's theorem.

18.2.3 Hawking & Penrose's singularity theorem

Among the three hypotheses of Penrose's singularity theorem (Property 18.1), the first one, namely there exists a non-compact Cauchy surface, is an assumption about the *global* structure

⁴For a Lagrangian density $\mathcal{L} = \mathcal{L}(\mathcal{F})$, the energy-momentum tensor is $T_{\alpha\beta} = -\mathcal{L}'(\mathcal{F})F_{\mu\alpha}F^\mu_\beta + \mathcal{L}(\mathcal{F})g_{\alpha\beta}$, which generalizes the Maxwellian expression (1.52) to nonlinear electrodynamics.

of spacetime. It cannot be checked by any local physical experiment. As the Bardeen black hole shows (Example 9), one cannot get rid of this hypothesis without replacing it by other ones. It turns out that it is possible find more local hypotheses: a stronger convergence condition and the non-existence of closed timelike curve. More precisely, we have:

Property 18.2: Hawking-Penrose theorem (1970 [266])

Let (\mathcal{M}, g) be a n -dimensional time-orientable spacetime ($n \geq 3$) such that

1. \mathcal{M} does not contain any closed timelike curve;
2. the **causal convergence condition** holds on \mathcal{M} :

$$\mathbf{R}(v, v) \geq 0 \quad \text{for any timelike or null vector } v, \quad (18.10)$$

where \mathbf{R} is g 's Ricci tensor – for general relativity with a cosmological constant Λ , such a condition is equivalent to the **strong energy condition**:

$$\mathbf{T}(v, v) - \frac{1}{n-2} \left(T - \frac{\Lambda}{4\pi} \right) v \cdot v \geq 0 \quad \text{for any timelike or null vector } v, \quad (18.11)$$

where \mathbf{T} is the matter energy-momentum tensor and $T := g^{\mu\nu}T_{\mu\nu}$ is its trace;

3. the following **generic condition** holds: every causal geodesic contains at least one point at which the (nonzero) tangent vectors v fulfill

$$v^{[\alpha} R^{\beta]}_{\mu\nu[\gamma} v_{\delta]} v^\mu v^\nu \neq 0, \quad (18.12)$$

where $R^{\alpha}_{\beta\gamma\delta}$ is g 's Riemann curvature tensor;

4. \mathcal{M} contains at least one of the following:

- (a) a trapped surface,
- (b) a point p such that the expansion of all the future-directed null geodesics, or of all the past-directed null geodesics, emanating from p becomes negative at some point along each geodesic,
- (c) a spacelike compact hypersurface without boundary (i.e. an embedded compact manifold of dimension $n-1$).

Then \mathcal{M} contains at least one incomplete inextendible causal geodesic.

The proof is quite long and we refer to HE [265], p. 267, where it is given for $n = 4$, the generalization to any dimension being straightforward (see also Sec. 5 of Ref. [450] and Sec. 5.1.4 of Ref. [453]). Here, we shall limit ourselves to a few comments:

- The non-existence of closed timelike curves is a much weaker condition than the existence

of a Cauchy surface required in Penrose's theorem (Property 18.1), the latter condition implying the former one.

- On the other side, the causal convergence condition (18.10) is stronger than the null convergence condition (2.94) involved in Penrose's theorem, for it has to hold for timelike vectors, in addition to null ones.
- For general relativity, the causal convergence condition (18.10) is equivalent to the strong energy condition (18.11) thanks to the Einstein equation (1.42). For a perfect fluid of proper energy density ρ and pressure p , the energy-momentum is $\mathbf{T} = (\rho + p)\underline{\mathbf{u}} \otimes \underline{\mathbf{u}} + p\mathbf{g}$ and the strong energy condition (18.11) is equivalent to

$$(n - 3)\rho + (n - 1)p - \frac{\Lambda}{4\pi} \geq 0. \quad (18.13)$$

Note that, even for $\Lambda = 0$, the strong energy condition does *not* imply the weak energy condition introduced in Sec. 2.4.2 [Eq. (2.96)], contrary to what the names may suggest. The only logical connection between the two conditions is that both imply the null energy condition (2.95).

- For timelike geodesics, the generic condition (18.12) is equivalent to⁵

$$R^\alpha_{\mu\nu\beta} v^\mu v^\nu \neq 0. \quad (18.14)$$

In view of the geodesic deviation equation (B.49), this implies that the relative acceleration in any timelike geodesic congruence cannot vanish identically along the geodesics. In physical terms, this means that tidal forces are necessarily felt sometime along the worldline of any inertial observer. Apart from flat (Minkowski) spacetime, this condition is violated only in peculiar cases, involving a high degree of symmetry, hence the qualifier *generic*. An example is the Einstein static universe considered in Sec. 4.2.3 [Eq. (4.18)]: it can be shown [450] that the curves $(\chi, \theta, \varphi) = \text{const}$ are timelike geodesics that have $R^\alpha_{\mu\nu\beta} v^\mu v^\nu = 0$ all along them.

- Hypothesis 4 of the Hawking-Penrose theorem is more general than the requirement of a trapped surface in Penrose's theorem (Property 18.1), since the trapped surface can be replaced by a light cone that reconverges in the future (case (b)) or a spacelike compact hypersurface (case (c)). The latter possibility arises in spatially closed cosmological models, such as the de Sitter spacetime (cf. Example 3 on p. 103), which admits spacelike hypersurfaces of \mathbb{S}^3 topology. This does not occur in Minkowski spacetime: embedded hypersurfaces with \mathbb{S}^3 topology are of course possible (the unit sphere of \mathbb{R}^4 !), but they cannot be everywhere spacelike.
- The Hawking-Penrose theorem does not specify whether the incomplete geodesic is timelike or null and whether it is future-incomplete or past-incomplete; it is therefore less precise than Penrose's singularity theorem, which tells that the incomplete geodesic is null and future-incomplete. It can be argued however that in the case (4a) (trapped surface), the geodesic is future-incomplete (cf. Ref. [450], p. 792).

⁵To see it, it suffices to contract (18.12) by $v_\alpha v^\delta$ and use the symmetries of the Riemann tensor (cf. Sec. A.5).

Remark 2: Hypothesis (4c) (existence of a spacelike compact hypersurface) is that stated in the original article by Hawking & Penrose [266]. It can be generalized to the existence of an achronal compact set without edge (cf. Theorem 2 on p. 266 of HE [265] and Theorem 5.6 of Ref. [450]).

Example 10 (strong energy condition): Let us check that the strong energy condition (18.11) holds for the classical spacetimes involved in Examples 6 (Schwarzschild), 7 (Oppenheimer-Snyder) and 8 (Vaidya). For the Schwarzschild black hole, this is trivial since $\mathbf{T} = 0$ and $\Lambda = 0$. For the Oppenheimer-Snyder collapse, still $\Lambda = 0$ and we note that the matter is a perfect fluid with $\rho \geq 0$ and $p = 0$, so that the fluid strong energy condition (18.13) is fulfilled. For the Vaidya collapse, the energy-momentum tensor \mathbf{T} is given by Eq. (15.5), which yields $\mathbf{T}(\mathbf{v}, \mathbf{v}) = M'(v)(\mathbf{k} \cdot \mathbf{v})^2/(4\pi r^2)$ and $T = 0$. Given that $M'(v) \geq 0$ and $\Lambda = 0$, it is clear that (18.11) is satisfied.

Example 11 (Bardeen regular black hole as a counter-example): We have seen in Example 9 that the Bardeen black hole, which has no singularity and is geodesically complete, evades Penrose's theorem 18.1 because it has no Cauchy surface. It must evade the Hawking-Penrose theorem 18.2 as well. Hypothesis 1 of that theorem is fulfilled: the Bardeen spacetime does not contain any closed timelike curve. Hypotheses 3 and 4a are fulfilled as well. The only reason for which the Hawking-Penrose theorem does not apply is actually hypothesis 2: the strong energy condition is violated by the nonlinear-electrodynamics energy-momentum tensor sourcing the Bardeen solution (albeit the weak energy condition is not).

Remark 3: The scope of the Hawking-Penrose theorem goes beyond gravitational collapse and black hole physics. It encompasses cosmology and can be used to predict the Big-Bang singularity, via the hypothesis (4b) with the point p being our current position in the Universe and considering the past-directed null geodesics (our past light-cone) (see Sec. 10.1 of HE [265] for details).

Remark 4: Both Penrose and Hawking-Penrose theorems have been generalized to n -dimensional spacetimes containing a “trapped” submanifold of any codimension (the codimension is 2 in Penrose's theorem 18.1 and 2, n and 1 for respectively cases (4a), (4b) and (4c) of Hawking-Penrose theorem 18.2) by Gregory Galloway and José Senovilla in 2010 [210]. In this work, the concept of trapped surface, a priori defined for codimension 2 only (cf. Sec. 18.2.1), is generalized to any codimension by demanding that the mean curvature vector of the submanifold is everywhere future-directed timelike; moreover, the null and causal convergence conditions are generalized to inequalities involving the Riemann tensor.

Historical note : The singularity theorem established by Roger Penrose in 1965 [404] (Property 18.1) was a major breakthrough in general relativity and mathematical physics. It introduced a brand new concept, that of trapped surfaces, which revealed to be extremely fruitful in relativistic gravity. It also brought in the characterization of spacetime singularities by geodesic incompleteness. This novel view point on singularities was subsequently advocated by Robert Geroch in 1968 [215] and Stephen Hawking and George Ellis in their 1973 textbook [265]; it is nowadays the standard one. From an astrophysical perspective, Penrose's theorem showed that once a gravitational collapse has reached a certain stage, characterized by the appearance of trapped surfaces, then the formation of a singularity is inevitable, whatever the geometry of the collapse and the stiffness of the matter's equation of state may be (provided the very mild null energy condition is obeyed). Previously, it could have been argued that the central singularity obtained in exact solutions, like that of the Oppenheimer-Snyder collapse (Chap. 14), is an artifact of spherical symmetry or the assumption of zero pressure, so that e.g. rotation or high pressures could prevent the singularity from appearing. For this achievement, Roger Penrose was awarded the Nobel Prize in Physics in 2020. More details about the history of Penrose's theorem and its impact can be

found in the review articles [451, 452, 453, 334]. As for the theorem proven in 1970 by Stephen Hawking and Roger Penrose [266] (Property 18.2), it is still considered today as the best singularity theorem.

18.3 Trapping horizons

Part V

Appendices

Appendix A

Basic differential geometry

Contents

A.1	Introduction	709
A.2	Differentiable manifolds	709
A.3	Pseudo-Riemannian manifolds	719
A.4	The three basic derivatives	724
A.5	Curvature	732

A.1 Introduction

We recall in this appendix basic definitions and results of differential geometry that are used in the main text. The reader who has some knowledge of general relativity should be familiar with most of them. It should be clear that this appendix is not intended to be a monograph on differential geometry. 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 [330, 342, 343, 389, 58, 109, 188], as well as to the differential geometry sections of the general relativity textbooks [107, 463, 498].

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 \mathcal{M} obeying the following properties:

1. \mathcal{M} is a **separated space** (also called **Hausdorff space**): any two distinct points of \mathcal{M} admit disjoint open neighborhoods.

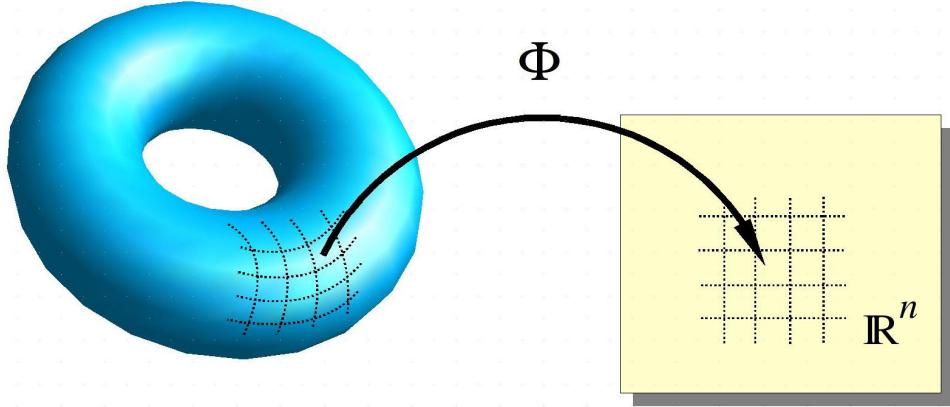


Figure A.1: Locally a manifold resembles \mathbb{R}^n ($n = 2$ on the figure), but not necessarily at the global level.

2. \mathcal{M} has a **countable base**¹: there exists a countable family $(\mathcal{U}_k)_{k \in \mathbb{N}}$ of open sets of \mathcal{M} such that any open set of \mathcal{M} can be written as the union (possibly infinite) of some members of the above family.
3. Around each point of \mathcal{M} , there exists a neighborhood 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 \mathcal{M} in a continuous way by n real numbers $(x^\alpha)_{\alpha \in \{0, \dots, n-1\}}$, which are called **coordinates** (cf. Fig. A.1). More precisely, given an open subset $\mathcal{U} \subset \mathcal{M}$, a **coordinate system** or **chart** on \mathcal{U} is a homeomorphism²

$$\begin{aligned} \Phi : \quad \mathcal{U} \subset \mathcal{M} &\longrightarrow \mathcal{V} \subset \mathbb{R}^n \\ p &\longmapsto (x^0, \dots, x^{n-1}), \end{aligned} \tag{A.1}$$

where \mathcal{V} is an open subset of \mathbb{R}^n .

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 \mathcal{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.

¹In the language of topology, one says that \mathcal{M} is a *second-countable space*.

²Let us recall that a **homeomorphism** between two topological spaces (here \mathcal{U} and \mathcal{V}) is a bijective map Φ such that both Φ and Φ^{-1} are continuous.

Usually, one needs more than one coordinate system to cover \mathcal{M} . An **atlas** on \mathcal{M} is a set of pairs $(\mathcal{U}_i, \Phi_i)_{i \in I}$, where I is a set (not necessarily finite), \mathcal{U}_i an open set of \mathcal{M} and Φ_i a chart on \mathcal{U}_i , such that the union of all \mathcal{U}_i covers \mathcal{M} :

$$\bigcup_{i \in I} \mathcal{U}_i = \mathcal{M}. \quad (\text{A.2})$$

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 \mathcal{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 map

$$\Phi_i \circ \Phi_j^{-1} : \Phi_j(\mathcal{U}_i \cap \mathcal{U}_j) \subset \mathbb{R}^n \longrightarrow \Phi_i(\mathcal{U}_i \cap \mathcal{U}_j) \subset \mathbb{R}^n \quad (\text{A.3})$$

is smooth (i.e. C^∞). Note that the above map is from an open set of \mathbb{R}^n to an open set of \mathbb{R}^n , so that the invoked C^∞ differentiability is nothing but that of \mathbb{R}^n . Such a map is called a **change of coordinates** or, in the mathematical 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: Strictly speaking a smooth manifold is a pair $(\mathcal{M}, \mathcal{A})$ where \mathcal{A} is a (maximal) smooth atlas on \mathcal{M} . Indeed, a given (topological) manifold \mathcal{M} can have non-equivalent differentiable structures, as shown by Milnor (1956) [365] for \mathbb{S}^7 , the unit sphere of dimension 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 standard spacetime case!) there exist uncountably many non-equivalent smooth structures, the so-called *exotic \mathbb{R}^4* [473].

Remark 4: When discussing the no-hair theorem in Chap. 5, one refers to the concept of *analytic manifold*, which is a special case of smooth manifold. Indeed, an **analytic manifold** is defined as a manifold equipped with an atlas for which all the changes of coordinates $\Phi_i \circ \Phi_j^{-1}$ are real 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, \mathcal{M} and \mathcal{M}' , of respective dimensions n and n' , a map $\phi : \mathcal{M} \rightarrow \mathcal{M}'$ is called a **smooth map** iff for every $p \in \mathcal{M}$, there exist a chart (\mathcal{U}, Φ) around p in the smooth atlas of \mathcal{M} and a chart (\mathcal{U}', Φ') around $\phi(p)$ in the smooth atlas of \mathcal{M}' such that $\phi(\mathcal{U}) \subset \mathcal{U}'$ and $\Phi' \circ \phi \circ \Phi^{-1}$ is a smooth map $\Phi(\mathcal{U}) \subset \mathbb{R}^n \rightarrow \mathbb{R}^{n'}$. The map ϕ is said to be a **diffeomorphism** iff it is bijective and both ϕ and ϕ^{-1} are smooth. This implies $n = n'$.

A.2.2 Manifold with boundary

A (topological) **manifold with boundary** \mathcal{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 \mathcal{M} , there exists a neighborhood which is homeomorphic to an open subset³ of the closed half-space

$$\mathbb{H}^n := \{(x^1, \dots, x^n) \in \mathbb{R}^n, \quad x^n \geq 0\}. \quad (\text{A.4})$$

A point $p \in \mathcal{M}$ is said to be a **boundary point** of \mathcal{M} iff there exists a homeomorphism Φ from an open neighborhood of p to an open subset of \mathbb{H}^n such that $\Phi(p) \in \partial\mathbb{H}^n$, where

$$\partial\mathbb{H}^n := \{(x^1, \dots, x^n) \in \mathbb{R}^n, \quad x^n = 0\}. \quad (\text{A.5})$$

This definition is independent from the choice of Φ (cf. Theorem 1.37 of Ref. [342]). The set of all boundary points of \mathcal{M} is naturally called the **boundary of \mathcal{M}** and is denoted by $\partial\mathcal{M}$.

Remark 5: The boundary $\partial\mathcal{M}$ should not be confused with the *topological boundary* of \mathcal{M} , i.e. the boundary of \mathcal{M} as a topological space, which is the closure of \mathcal{M} minus the interior of \mathcal{M} ; since both sets coincide with \mathcal{M} , the topological boundary of \mathcal{M} is obviously \emptyset .

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 smooth manifold, vectors are defined as tangent vectors to a curve. Given an interval $I \subset \mathbb{R}$, a **curve** is a subset $\mathcal{L} \subset \mathcal{M}$ that is the image of a smooth map $I \rightarrow \mathcal{M}$:

$$\begin{aligned} P : I &\longrightarrow \mathcal{M} \\ \lambda &\longmapsto p = P(\lambda) \in \mathcal{L}. \end{aligned} \quad (\text{A.6})$$

Hence $\mathcal{L} = P(I) := \{P(\lambda) \mid \lambda \in I\}$. The function P is called a **parametrization** of \mathcal{L} and the real variable λ is called a **parameter along \mathcal{L}** . Given a coordinate system (x^α) in a neighborhood of a point $p \in \mathcal{L}$, the parametrization P is defined by n functions $X^\alpha : I \rightarrow \mathbb{R}$ such that

$$x^\alpha(P(\lambda)) = X^\alpha(\lambda). \quad (\text{A.7})$$

Remark 6: In the literature, a curve is often defined as a map $P : I \rightarrow \mathcal{M}$ and not as the image of P . According to that definition, different parametrizations give birth to different curves.

A **scalar field** on \mathcal{M} is a function $f : \mathcal{M} \rightarrow \mathbb{R}$. Unless specified, we shall always consider *smooth* scalar fields, i.e. smooth maps as defined in Sec. A.2.1. At a point $p = P(\lambda) \in \mathcal{L}$, the

³By *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 $A = \mathbb{H}^n$).

tangent vector to \mathcal{L} associated with the parametrization P is the operator v which maps every scalar field f defined around p to the real number

$$v(f) := \left. \frac{df}{d\lambda} \right|_{\mathcal{L}} = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} [f(P(\lambda + \varepsilon)) - f(P(\lambda))]. \quad (\text{A.8})$$

Given a coordinate system (x^α) around some point $p \in \mathcal{M}$, there are n curves \mathcal{L}_α through p associated with (x^α) and called the **coordinate lines**: for each $\alpha \in \{0, \dots, n-1\}$, \mathcal{L}_α is defined as the curve through p parameterized by $\lambda = x^\alpha$ and having constant coordinates x^β for all $\beta \neq \alpha$. The tangent vector to \mathcal{L}_α parameterized by x^α is denoted ∂_α . From the definition (A.8), its action on a scalar field f is

$$\partial_\alpha(f) = \left. \frac{df}{dx^\alpha} \right|_{\mathcal{L}_\alpha} = \left. \frac{df}{dx^\alpha} \right|_{\substack{x^\beta = \text{const} \\ \beta \neq \alpha}}.$$

Considering f as a function of the coordinates (x^0, \dots, x^{n-1}) (whereas strictly speaking it is a function of the points on \mathcal{M}) we recognize in the last term the partial derivative of f with respect to x^α . Hence

$$\boxed{\partial_\alpha(f) = \frac{\partial f}{\partial x^\alpha}}. \quad (\text{A.9})$$

Similarly, we may rewrite (A.8) as

$$v(f) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} [f(X^0(\lambda + \varepsilon), \dots, X^{n-1}(\lambda + \varepsilon)) - f(X^0(\lambda), \dots, X^{n-1}(\lambda))] = \frac{\partial f}{\partial x^\alpha} \frac{dX^\alpha}{d\lambda},$$

i.e., in view of Eq. (A.9)

$$\boxed{v(f) = v^\alpha \partial_\alpha(f)}, \quad (\text{A.10})$$

where

$$v^\alpha := \frac{dX^\alpha}{d\lambda}, \quad 0 \leq \alpha \leq n-1. \quad (\text{A.11})$$

In Eq. (A.10), we are using the **Einstein summation convention**: a repeated index implies a summation over all the values taken by this index (here, from $\alpha = 0$ to $\alpha = n-1$). The identity (A.10) being valid for any scalar field f , we conclude that

$$\boxed{v = v^\alpha \partial_\alpha}. \quad (\text{A.12})$$

Since every tangent vector to a curve at p is expressible as (A.12), we conclude that the set of all tangent vectors to a curve at p is a vector space of dimension n and that (∂_α) constitutes a basis of it. This vector space is called the **tangent space to \mathcal{M} at p** and is denoted $T_p \mathcal{M}$. The elements of $T_p \mathcal{M}$ are simply called **vectors at p** . The basis (∂_α) is called the **natural basis** associated with the coordinates (x^α) and the coefficients v^α in (A.12) are called the **components of the vector 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 define a vector connecting two distinct points p and q on a generic manifold, except if p and q are infinitesimally close to each

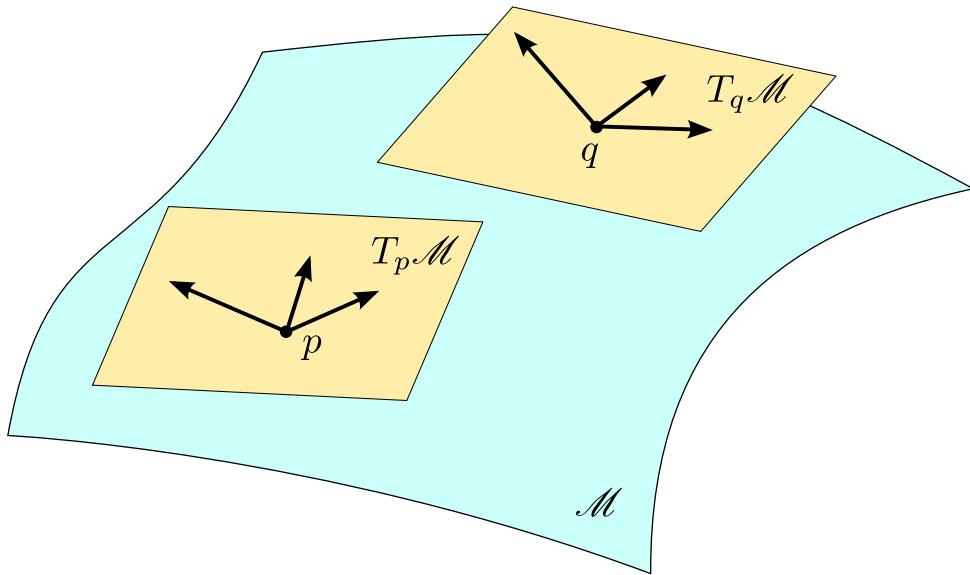


Figure A.2: The vectors at two points p and q on the manifold \mathcal{M} belong to two different vector spaces: the tangent spaces $T_p \mathcal{M}$ and $T_q \mathcal{M}$.

other. Indeed, in the latter case, one defines the ***infinitesimal displacement vector from p to q*** as the vector $d\mathbf{x} \in T_p \mathcal{M}$ whose action on a scalar field f is

$$d\mathbf{x}(f) := df|_{p \rightarrow q} = f(q) - f(p). \quad (\text{A.13})$$

Since p and q are infinitesimally close, there is a unique (piece of) curve \mathcal{L} going from p to q and one has

$$d\mathbf{x} = \mathbf{v} d\lambda, \quad (\text{A.14})$$

where λ is a parameter along \mathcal{L} , \mathbf{v} the associated tangent vector at p and $d\lambda$ the parameter increment from p to q : $p = P(\lambda)$ and $q = P(\lambda + d\lambda)$. The relation (A.14) follows immediately from the definitions (A.8) and (A.13) for respectively \mathbf{v} and $d\mathbf{x}$. Given a coordinate system, let (x^α) be the coordinates of p and $(x^\alpha + dx^\alpha)$ those of q . Then from Eqs. (A.13) and (A.9), $d\mathbf{x}(f) = df = \partial f / \partial x^\alpha dx^\alpha = dx^\alpha \partial_\alpha(f)$. The scalar field f being arbitrary, we conclude that

$$d\mathbf{x} = dx^\alpha \partial_\alpha. \quad (\text{A.15})$$

In other words, the components of the infinitesimal displacement vector $d\mathbf{x}$ with respect to the coordinates (x^α) are nothing but the infinitesimal coordinate increments dx^α .

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 $p \in \mathcal{M}$, one defines a *linear form* as a mapping⁴

$$\begin{aligned}\omega : T_p \mathcal{M} &\longrightarrow \mathbb{R} \\ v &\longmapsto \langle \omega, v \rangle\end{aligned}\tag{A.16}$$

that is linear: $\langle \omega, \lambda v + u \rangle = \lambda \langle \omega, v \rangle + \langle \omega, u \rangle$ for all $v, u \in T_p \mathcal{M}$ 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 $T_p \mathcal{M}$** and denoted by $T_p^* \mathcal{M}$. Given the natural basis (∂_α) of $T_p \mathcal{M}$ associated with some coordinates (x^α) , there is a unique basis of $T_p^* \mathcal{M}$, denoted by (dx^α) , such that

$$\boxed{\langle dx^\alpha, \partial_\beta \rangle = \delta^\alpha_\beta},\tag{A.17}$$

where δ^α_β is the **Kronecker symbol**: $\delta^\alpha_\beta = 1$ if $\alpha = \beta$ and 0 otherwise. The basis (dx^α) is called the **dual basis** of the basis (∂_α) . The notation dx^α stems from the fact that if we apply the linear form dx^α to the infinitesimal displacement vector (A.15), we get nothing but the number dx^α :

$$\langle dx^\alpha, dx \rangle = \langle dx^\alpha, dx^\beta \partial_\beta \rangle = dx^\beta \underbrace{\langle dx^\alpha, \partial_\beta \rangle}_{\delta^\alpha_\beta} = dx^\alpha.\tag{A.18}$$

Remark 7: Do not confuse the linear form dx^α with the infinitesimal increment dx^α of the coordinate x^α .

The dual basis can be used to expand any linear form ω , thereby defining its **components ω_α with respect to the coordinates (x^α)** :

$$\omega = \omega_\alpha dx^\alpha.\tag{A.19}$$

In terms of components, the action of a linear form on a vector takes then a very simple form:

$$\boxed{\langle \omega, v \rangle = \omega_\alpha v^\alpha}.\tag{A.20}$$

This follows immediately from (A.19), (A.12) and (A.17).

A field of linear forms, i.e. a (smooth) map which associates to each point $p \in \mathcal{M}$ an element of the dual space $T_p^* \mathcal{M}$ is called a **1-form**. Given a smooth scalar field f on \mathcal{M} , there exists a 1-form canonically associated with it, called the **differential of f** and denoted df or ∇f . At each point $p \in \mathcal{M}$, df is the unique linear form which, once applied to the infinitesimal displacement vector dx from p to a nearby point q , gives the change in f between points p and q :

$$df := f(q) - f(p) = \langle df, dx \rangle.\tag{A.21}$$

⁴We are using the same bra-ket notation as in quantum mechanics to denote the action of a linear form on a vector.

Since $\mathrm{d}f = \partial f / \partial x^\alpha \mathrm{d}x^\alpha$, Eq. (A.18) 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) :

$$\boxed{\mathrm{d}f = \frac{\partial f}{\partial x^\alpha} \mathrm{d}x^\alpha}. \quad (\text{A.22})$$

Remark 8: 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 $\vec{\omega}$ to any 1-form ω via the Euclidean scalar product of \mathbb{R}^3 : $\forall \vec{v} \in \mathbb{R}^3$, $\langle \omega, \vec{v} \rangle = \vec{\omega} \cdot \vec{v}$. Accordingly, formula (A.21) is rewritten as $\mathrm{d}f = \vec{\nabla} f \cdot \mathrm{d}x$. But we should keep in mind that, at the fundamental level, the key quantity is the differential 1-form $\nabla f = \mathrm{d}f$, for Eq. (A.21) does not require any metric on the manifold \mathcal{M} to be meaningful. The gradient $\vec{\nabla} f$ is a derived quantity, obtained from the differential ∇f by metric duality.

Remark 9: For a fixed value of α , the coordinate x^α can be considered as a scalar field on \mathcal{M} . If we apply (A.22) to $f = x^\alpha$, we then get $\mathrm{d}x^\alpha = \mathrm{d}x^\alpha$. Hence the dual basis to the natural basis (∂_α) is formed by the differentials of the coordinates. This justifies the notation $\mathrm{d}x^\alpha$ used for its elements.

By combining Eqs. (A.22), (A.12) and (A.17), it is easy to see that the 1-form $\mathrm{d}f$ acting on a vector v is nothing but v acting on the scalar field f :

$$\boxed{\langle \mathrm{d}f, v \rangle = v(f)}. \quad (\text{A.23})$$

Natural bases are of course not the only bases in the vector space $T_p \mathcal{M}$. One may use a basis (e_α) that is not related to any coordinate system on \mathcal{M} , for instance an orthonormal basis with respect to some metric. There exists then a unique basis (e^α) of the dual space $T_p^* \mathcal{M}$ such that⁵

$$\boxed{\langle e^\alpha, e_\beta \rangle = \delta^\alpha_\beta}. \quad (\text{A.24})$$

(e^α) is called the **dual basis** to (e_α) . The relation (A.17) is a special case of (A.24), for which $e_\alpha = \partial_\alpha$ and $e^\alpha = \mathrm{d}x^\alpha$.

A.2.5 Tensors

Tensors are generalizations of both vectors and linear forms. Let $(k, \ell) \in \mathbb{N}^2$ with $(k, \ell) \neq (0, 0)$. At a point $p \in \mathcal{M}$, a **tensor of type** (k, ℓ) , also called **tensor k times contravariant and ℓ times covariant**, is a map

$$\begin{aligned} \mathbf{T} : & \underbrace{T_p^* \mathcal{M} \times \cdots \times T_p^* \mathcal{M}}_{k \text{ times}} \times \underbrace{T_p \mathcal{M} \times \cdots \times T_p \mathcal{M}}_{\ell \text{ times}} & \longrightarrow & \mathbb{R} \\ & (\omega_1, \dots, \omega_k, v_1, \dots, v_\ell) & \longmapsto & \mathbf{T}(\omega_1, \dots, \omega_k, v_1, \dots, v_\ell) \end{aligned} \quad (\text{A.25})$$

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 $T_p^{**} \mathcal{M} = T_p \mathcal{M}$,

⁵Notice that, according to the standard usage, the symbol for the vector e_α and that for the linear form e^α differ only by the position of the index α .

which means that every vector \mathbf{v} can be considered as a linear form on the space $T_p^*\mathcal{M}$, via $T_p^*\mathcal{M} \rightarrow \mathbb{R}, \omega \mapsto \langle \omega, \mathbf{v} \rangle$. 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 (e_α) of $T_p\mathcal{M}$ and the corresponding dual basis (e^α) in $T_p^*\mathcal{M}$, any tensor \mathbf{T} of type (k, ℓ) can be expanded as

$$\boxed{\mathbf{T} = T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} e_{\alpha_1} \otimes \dots \otimes e_{\alpha_k} \otimes e^{\beta_1} \otimes \dots \otimes e^{\beta_\ell}}, \quad (\text{A.26})$$

where the **tensor product** $e_{\alpha_1} \otimes \dots \otimes e_{\alpha_k} \otimes e^{\beta_1} \otimes \dots \otimes e^{\beta_\ell}$ is the tensor of type (k, ℓ) for which the image of $(\omega_1, \dots, \omega_k, v_1, \dots, v_\ell)$ as in (A.25) is the real number

$$\prod_{i=1}^k \langle \omega_i, e_{\alpha_i} \rangle \times \prod_{j=1}^\ell \langle e^{\beta_j}, v_j \rangle.$$

Notice that all the products in the above formula are simply products in \mathbb{R} . The $n^{k+\ell}$ real coefficients $T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell}$ in (A.26) are called the **components of the tensor \mathbf{T} with respect to the basis (e_α)** . These components are unique and fully characterize the tensor \mathbf{T} .

Remark 10: The notations v^α and ω_α already introduced for the components of a vector \mathbf{v} [Eq. (A.12)] or a linear form ω [Eq. (A.19)] are of course the particular cases $(k, \ell) = (1, 0)$ or $(k, \ell) = (0, 1)$ of (A.26), with, in addition, $e_\alpha = \partial_\alpha$ and $e^\alpha = dx^\alpha$.

A.2.6 Tensor fields on a manifold

A **tensor field of type** (k, ℓ) is a map \mathbf{T} which associates to each point $p \in \mathcal{M}$ a tensor of type (k, ℓ) on $T_p\mathcal{M}$. A **vector field** is naturally a tensor field of type $(1, 0)$. By convention, a scalar field is considered as a tensor field of type $(0, 0)$. We shall consider only smooth fields. We shall denote by $\mathbf{T}|_p$ the tensor representing the value of the tensor field \mathbf{T} at a point $p \in \mathcal{M}$.

Given a non-negative integer p , a **differential form of degree** 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 (e_α) such that at each point $p \in \mathcal{M}$, $(e_\alpha|_p)$ is a basis of the tangent space $T_p\mathcal{M}$. If $n = 4$, a frame field is also called a **tetrad** and if $n = 3$, it is called a **triad**.

Given a vector field \mathbf{v} and a scalar field f , the function $\mathcal{M} \rightarrow \mathbb{R}, p \mapsto \mathbf{v}|_p(f)$ clearly defines a scalar field on \mathcal{M} , which we denote naturally by $\mathbf{v}(f)$. We may then define the **commutator of two vector fields** \mathbf{u} and \mathbf{v} as the vector field $[\mathbf{u}, \mathbf{v}]$ whose action on a scalar field f is

$$[\mathbf{u}, \mathbf{v}](f) := \mathbf{u}(\mathbf{v}(f)) - \mathbf{v}(\mathbf{u}(f)). \quad (\text{A.27})$$

With respect to a coordinate system (x^α) , it is not difficult, via (A.12), to see that the components of the commutator are

$$\boxed{[\mathbf{u}, \mathbf{v}]^\alpha = u^\mu \frac{\partial v^\alpha}{\partial x^\mu} - v^\mu \frac{\partial u^\alpha}{\partial x^\mu}}. \quad (\text{A.28})$$

A.2.7 Immersions, embeddings and submanifolds

Let \mathcal{M} and \mathcal{N} be two smooth manifolds and $\Phi : \mathcal{M} \rightarrow \mathcal{N}$ be a smooth map, as defined in Sec. A.2.1. At a given point $p \in \mathcal{M}$, the **differential** of Φ is the linear map

$$d\Phi|_p : T_p \mathcal{M} \rightarrow T_{\Phi(p)} \mathcal{N} \quad (\text{A.29})$$

that “approximates” Φ in the following sense: if $dx \in T_p \mathcal{M}$ is the infinitesimal displacement vector from p to some (infinitesimally close) point q (cf. Sec. A.2.3), then

$$d\Phi|_p(dx) = dy, \quad (\text{A.30})$$

where dy is the infinitesimal displacement vector of $T_{\Phi(p)} \mathcal{N}$ connecting $\Phi(p)$ to $\Phi(q)$ (cf. Fig. ??). If (x^α) is a coordinate chart of \mathcal{M} around p and (y^β) a chart of \mathcal{N} around $\Phi(p)$, such that Φ is expressed as $y^\beta = Y^\beta(x^\alpha)$, it follows from Eqs. (A.15) and (A.30) that the matrix of the linear map $d\Phi|_p$ with respect to the bases (∂_α) of $T_p \mathcal{M}$ and (∂_β) of $T_{\Phi(p)} \mathcal{N}$ is the Jacobian matrix $(\partial Y^\beta / \partial x^\alpha)$. Using the characterization of vectors by their action on scalar fields [Eq. (A.8)], it is then easy to see that

$$\forall v \in T_p \mathcal{M}, \forall f \in C^\infty(\mathcal{N}, \mathbb{R}), \quad d\Phi|_p(v)(f) = v(f \circ \Phi). \quad (\text{A.31})$$

This property could be taken as an alternative definition of $d\Phi|_p$.

The smooth map Φ is called an **immersion** iff the differential $d\Phi|_p$ is injective at any point $p \in \mathcal{M}$. Moreover, Φ is called an **embedding** iff (i) Φ is an immersion and (ii) Φ is a homeomorphism $\mathcal{M} \rightarrow \Phi(\mathcal{M})$. Note that an embedding is necessarily injective, contrary to an immersion.

A **submanifold** of \mathcal{M} is a subset $\mathcal{S} \subset \mathcal{M}$ such that (i) \mathcal{S} is a manifold in the subspace topology and (ii) \mathcal{S} has a smooth structure with respect to which the inclusion map $\iota : \mathcal{S} \rightarrow \mathcal{M}$ is an embedding. One can show that \mathcal{S} is a submanifold of \mathcal{M} iff there exists a manifold \mathcal{S}_0 (a priori not a subset of \mathcal{M}) and an embedding $\Phi : \mathcal{S}_0 \rightarrow \mathcal{M}$, such that $\mathcal{S} = \Phi(\mathcal{S}_0)$.

Remark 11: Strictly speaking, the above definition regards an **embedded submanifold**; there is also the wider concept of **immersed submanifold** (see e.g. Chap 5 of [342]).

One has necessarily $\dim \mathcal{S} \leq \dim \mathcal{M}$. The non-negative integer $m = \dim \mathcal{M} - \dim \mathcal{S}$ is called the **codimension** of the submanifold \mathcal{S} . A submanifold of codimension 1 is called a **hypersurface**. A submanifold of dimension 1 is (the image of) a curve in \mathcal{M} , but note that not all curves are submanifolds: a curve with self-crossing points is not a submanifold.

A.2.8 Pushforwards and pullbacks

Given a smooth map $\Phi : \mathcal{M} \rightarrow \mathcal{N}$ and a vector v in the tangent space to \mathcal{M} at a point p , the image of v by the differential of Φ at p is called the **pushforward of v on \mathcal{N} by Φ** and is denoted by $\Phi_* v$:

$$\Phi_* v := d\Phi|_p(v). \quad (\text{A.32})$$

The pushforward can be geometrically interpreted as follows: consider v as the tangent vector to a curve \mathcal{L} through p associated to some parameter λ (cf. Sec. A.2.3) and let q be the point

of \mathcal{L} separated from p by the infinitesimal parameter increase $d\lambda$. One has then $\mathbf{v} = d\mathbf{x}/d\lambda$, where $d\mathbf{x}$ is the infinitesimal vector connecting p to q . Thanks to the linearity of $d\Phi|_p$ and the defining relation (A.30), one has $\Phi_* \mathbf{v} = d\mathbf{y}/d\lambda$, i.e. $\Phi_* \mathbf{v}$ is the finite vector obtained by rescaling the infinitesimal vector connecting $\Phi(p)$ to $\Phi(q)$ by $d\lambda$. Note that $\Phi_* \mathbf{v}$ is a vector tangent to the curve $\Phi(\mathcal{L}) \subset \mathcal{N}$: it is precisely the tangent vector to $\Phi(\mathcal{L})$ at $\Phi(p)$ associated with the parametrization of $\Phi(\mathcal{L})$ by λ .

Let now \mathbf{T} be a fully covariant tensor field on \mathcal{N} , i.e. a tensor field of type $(0, \ell)$ on \mathcal{N} for some $\ell \geq 0$. The **pullback of \mathbf{T} on \mathcal{M} by Φ** is the tensor field of type $(0, \ell)$ on \mathcal{M} denoted by $\Phi^* \mathbf{T}$ and defined by

$$\forall p \in \mathcal{M}, \forall (\mathbf{v}_1, \dots, \mathbf{v}_\ell) \in (T_p \mathcal{M})^\ell, \quad \Phi^* \mathbf{T}|_p (\mathbf{v}_1, \dots, \mathbf{v}_\ell) := \mathbf{T}|_{\Phi(p)} (\Phi_* \mathbf{v}_1, \dots, \Phi_* \mathbf{v}_\ell). \quad (\text{A.33})$$

For a scalar field $f : \mathcal{N} \rightarrow \mathbb{R}$, the above definition with $\ell = 0$ reduces to $\Phi^* f|_p := f|_{\Phi(p)}$, i.e. to $\Phi^* f(p) := f(\Phi(p))$, so that the pullback is nothing but the map composition: $\Phi^* f = f \circ \Phi$.

Remark 12: Via the pushforward operation, a smooth map $\Phi : \mathcal{M} \rightarrow \mathcal{N}$ naturally carries tangent vectors from \mathcal{M} to \mathcal{N} (essentially because it carries curves in \mathcal{M} to curves in \mathcal{N} , as discussed above), but it carries covariant tensors, among which linear forms, in the reverse way, i.e. from \mathcal{N} to \mathcal{M} (pullback operation). Another difference is that the pushforward is a pointwise operation and does not extend a priori to vector fields (cf. Fig. ??), while the pullback is well defined for any covariant tensor field. A case where the pushforward extends to vector fields though is when Φ is a diffeomorphism.

A.3 Pseudo-Riemannian manifolds

A.3.1 Metric tensor

A **pseudo-Riemannian manifold** is a pair $(\mathcal{M}, \mathbf{g})$ where \mathcal{M} is a smooth manifold and \mathbf{g} is a **metric tensor** on \mathcal{M} , i.e. a tensor field obeying the following properties:

1. \mathbf{g} is a tensor field of type $(0, 2)$: at each point $p \in \mathcal{M}$, $\mathbf{g}|_p$ is a bilinear form acting on pairs of vectors in the tangent space $T_p \mathcal{M}$:

$$\begin{aligned} \mathbf{g}|_p : T_p \mathcal{M} \times T_p \mathcal{M} &\longrightarrow \mathbb{R} \\ (\mathbf{u}, \mathbf{v}) &\longmapsto \mathbf{g}|_p(\mathbf{u}, \mathbf{v}) =: \mathbf{g}(\mathbf{u}, \mathbf{v}) \end{aligned} \quad (\text{A.34})$$

2. \mathbf{g} is **symmetric**: $\mathbf{g}(\mathbf{u}, \mathbf{v}) = \mathbf{g}(\mathbf{v}, \mathbf{u})$.
3. \mathbf{g} is **non-degenerate**: at any point $p \in \mathcal{M}$, a vector \mathbf{u} such that $\forall \mathbf{v} \in T_p \mathcal{M}, \mathbf{g}(\mathbf{u}, \mathbf{v}) = 0$ 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 \mathbf{u} and \mathbf{v} are **\mathbf{g} -orthogonal** (or simply **orthogonal** if there is no ambiguity) iff $\mathbf{g}(\mathbf{u}, \mathbf{v}) = 0$. 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 \mathbf{g} :

$$\forall (\mathbf{u}, \mathbf{v}) \in T_p \mathcal{M} \times T_p \mathcal{M}, \quad \boxed{\mathbf{u} \cdot \mathbf{v} := \mathbf{g}(\mathbf{u}, \mathbf{v})}. \quad (\text{A.35})$$

In a given basis (e_α) of $T_p\mathcal{M}$, the components of \mathbf{g} is the matrix $(g_{\alpha\beta})$ defined by formula (A.26) with $(k, \ell) = (0, 2)$:

$$\boxed{\mathbf{g} = g_{\alpha\beta} e^\alpha \otimes e^\beta}. \quad (\text{A.36})$$

For any pair (u, v) of vectors we have then $\mathbf{g}(u, v) = g_{\alpha\beta} u^\alpha v^\beta$. In particular, considering the natural basis associated with some coordinate system (x^α) , the scalar square of an infinitesimal displacement vector $dx = dx^\alpha \partial_\alpha$ [cf. Eqs. (A.13) and (A.15)] is

$$\boxed{ds^2 := \mathbf{g}(dx, dx) = g_{\alpha\beta} dx^\alpha dx^\beta}. \quad (\text{A.37})$$

This formula is called the **line element** expression on the pseudo-Riemannian manifold $(\mathcal{M}, \mathbf{g})$. It is often used to define the metric tensor in general relativity texts. Note that contrary to what the notation may suggest, ds^2 can be negative.

For the dual basis associated with the coordinates (x^α) , one has $e^\alpha = dx^\alpha$ (cf. Sec. A.2.4), so that Eq. (A.36) can be rewritten as

$$\mathbf{g} = g_{\alpha\beta} dx^\alpha \otimes dx^\beta. \quad (\text{A.38})$$

One can recast this relation in a form which reminds the line element (A.37) by introducing the symmetric product notation (cf. e.g. Refs. [343] or [463]):

$$dx^\alpha dx^\beta := \frac{1}{2} (dx^\alpha \otimes dx^\beta + dx^\beta \otimes dx^\alpha) \quad \text{and} \quad (dx^\alpha)^2 := dx^\alpha \otimes dx^\alpha. \quad (\text{A.39})$$

Formula (A.38) then becomes

$$\boxed{\mathbf{g} = g_{\alpha\beta} dx^\alpha dx^\beta}. \quad (\text{A.40})$$

Applying this relation to the pair of infinitesimal vectors (dx, dx) , one gets the line element (A.37), by virtue of the identity $\langle dx^\alpha, dx \rangle = dx^\alpha$ [Eq. (A.18)].

A.3.2 Signature and orthonormal bases

An important feature of the metric tensor is its **signature**: in all bases of $T_p\mathcal{M}$ 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:

$$\text{sign } \mathbf{g} = (\underbrace{-, \dots, -}_{s \text{ times}}, \underbrace{+, \dots, +}_{n-s \text{ times}}). \quad (\text{A.41})$$

If $s = 0$, \mathbf{g} is called a **Riemannian metric** and $(\mathcal{M}, \mathbf{g})$ a **Riemannian manifold**. In this case, \mathbf{g} is **positive-definite**, which means that $\mathbf{g}(v, v) \geq 0$ for all $v \in T_p\mathcal{M}$ and $\mathbf{g}(v, v) = 0$ iff $v = 0$. A standard example of Riemannian metric is of course the scalar product of the Euclidean space \mathbb{R}^n .

If $s = 1$, \mathbf{g} is called a **Lorentzian metric** and $(\mathcal{M}, \mathbf{g})$ a **Lorentzian manifold**. One may then have $\mathbf{g}(v, v) < 0$; vectors for which this occurs are called **timelike**, whereas vectors for

which $\mathbf{g}(\mathbf{v}, \mathbf{v}) > 0$ are called **spacelike**, and those for which $\mathbf{g}(\mathbf{v}, \mathbf{v}) = 0$ are called **null**. The subset of $T_p\mathcal{M}$ formed by all null vectors is termed the **null cone** of \mathbf{g} at p .

A coordinate x^α of a coordinate system (x^0, \dots, x^{n-1}) is said to be a **timelike coordinate** (resp. **spacelike coordinate** or **null coordinate**) iff the hypersurfaces defined by $x^\alpha = \text{const}$ are spacelike⁶ (resp. timelike or null).

Remark 1: Being timelike, spacelike or null is a property of the coordinate x^α per se (i.e. considering x^α as a scalar field $\mathcal{U} \subset \mathcal{M} \rightarrow \mathbb{R}$); on the contrary the causal type of the coordinate vector ∂_α depends on the coordinate system (x^0, \dots, x^{n-1}) to which x^α belongs. More precisely, whatever the causal type of x^α , the vector ∂_α can be made spacelike, timelike or null by a proper choice of the complementary coordinates $(x^\beta)_{\beta \neq \alpha}$.

A basis (e_α) of $T_p\mathcal{M}$ is called a **\mathbf{g} -orthonormal basis** (or simply **orthonormal basis**) if there is no ambiguity on the metric) iff⁷

$$\begin{aligned}\mathbf{g}(e_\alpha, e_\alpha) &= -1 && \text{for } 0 \leq \alpha \leq s-1 \\ \mathbf{g}(e_\alpha, e_\alpha) &= 1 && \text{for } s \leq \alpha \leq n-1 \\ \mathbf{g}(e_\alpha, e_\beta) &= 0 && \text{for } \alpha \neq \beta.\end{aligned}\tag{A.42}$$

A.3.3 Metric duality

Since the bilinear form \mathbf{g} is non-degenerate, its matrix $(g_{\alpha\beta})$ in any basis (e_α) is invertible and the inverse is denoted by $(g^{\alpha\beta})$:

$$g^{\alpha\mu} g_{\mu\beta} = \delta^\alpha_\beta.\tag{A.43}$$

The metric \mathbf{g} induces an isomorphism between $T_p\mathcal{M}$ (vectors) and $T_p^*\mathcal{M}$ (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 $T_p\mathcal{M} \rightarrow T_p^*\mathcal{M}$ and by an arrow the reverse isomorphism $T_p^*\mathcal{M} \rightarrow T_p\mathcal{M}$:

1. For any vector \mathbf{u} in $T_p\mathcal{M}$, $\underline{\mathbf{u}}$ stands for the unique linear form such that

$$\forall \mathbf{v} \in T_p\mathcal{M}, \quad \langle \underline{\mathbf{u}}, \mathbf{v} \rangle = \mathbf{g}(\mathbf{u}, \mathbf{v}).\tag{A.44}$$

However, we will omit the underbar on the components of $\underline{\mathbf{u}}$, since the position of the index allows us to distinguish between vectors and linear forms, following the standard usage: if the components of \mathbf{u} in a given basis (e_α) are denoted by u^α , the components of $\underline{\mathbf{u}}$ in the dual basis (e^α) are then denoted by u_α and are given by

$$u_\alpha = g_{\alpha\mu} u^\mu.\tag{A.45}$$

⁶Cf. Sec. 2.2.2 for the definition of spacelike, timelike and null hypersurfaces.

⁷No summation on α is implied.

2. For any linear form ω in $T_p^*\mathcal{M}$, $\overrightarrow{\omega}$ stands for the unique vector of $T_p\mathcal{M}$ such that

$$\forall v \in T_p\mathcal{M}, \quad g(\overrightarrow{\omega}, v) = \langle \omega, v \rangle. \quad (\text{A.46})$$

As for the underbar, we will omit the arrow on the components of $\overrightarrow{\omega}$ by denoting them ω^α ; they are given by

$$\omega^\alpha = g^{\alpha\mu} \omega_\mu. \quad (\text{A.47})$$

3. We extend the arrow notation to *bilinear* forms on $T_p\mathcal{M}$ (type-(0, 2) tensor): for any bilinear form T , we denote by \overrightarrow{T} the tensor of type (1, 1) such that

$$\forall (u, v) \in T_p\mathcal{M} \times T_p\mathcal{M}, \quad T(u, v) = \overrightarrow{T}(\underline{u}, v) = u \cdot \overrightarrow{T}(v), \quad (\text{A.48})$$

and by $\overrightarrow{\overrightarrow{T}}$ the tensor of type (2, 0) defined by

$$\forall (u, v) \in T_p\mathcal{M} \times T_p\mathcal{M}, \quad T(u, v) = \overrightarrow{\overrightarrow{T}}(\underline{u}, \underline{v}). \quad (\text{A.49})$$

Note that in the second equality of (A.48), we have considered \overrightarrow{T} as an endomorphism $T_p\mathcal{M} \rightarrow T_p\mathcal{M}$, which is always possible for a tensor of type (1, 1). If $T_{\alpha\beta}$ are the components of T in some basis (e_α) , the components of \overrightarrow{T} and $\overrightarrow{\overrightarrow{T}}$ are respectively

$$(\overrightarrow{T})^\alpha{}_\beta = T^\alpha{}_\beta = g^{\alpha\mu} T_{\mu\beta} \quad (\text{A.50a})$$

$$(\overrightarrow{\overrightarrow{T}})^{\alpha\beta} = T^{\alpha\beta} = g^{\alpha\mu} g^{\beta\nu} T_{\mu\nu}. \quad (\text{A.50b})$$

Remark 2: In mathematical literature, the isomorphism induced by g between $T_p\mathcal{M}$ and $T_p^*\mathcal{M}$ is called the **musical isomorphism**, because a flat and a sharp symbols are used instead of, respectively, the underbar and the arrow introduced above:

$$u^\flat = \underline{u} \quad \text{and} \quad \omega^\sharp = \overrightarrow{\omega}.$$

A.3.4 Levi-Civita tensor

Let us assume that \mathcal{M} is an **orientable manifold**, i.e. that there exists a n -form⁸ on \mathcal{M} (n being \mathcal{M} 's dimension) that is continuous on \mathcal{M} and nowhere vanishing. Then, given a metric g on \mathcal{M} , one can show that there exist only two n -forms ϵ such that for any g -orthonormal basis (e_α) ,

$$\epsilon(e_0, \dots, e_{n-1}) = \pm 1. \quad (\text{A.51})$$

Picking one of these two n -forms amounts to choosing an **orientation** for \mathcal{M} . The chosen ϵ is then called the **Levi-Civita tensor** associated with the metric g . It is also called the **volume form** of g (cf. Sec. A.4.3). The sign in the right-hand side of (A.51) gives then the orientation of the basis (e_α) . More generally, given a (not necessarily orthonormal) basis (e_α) of $T_p\mathcal{M}$, one

⁸Cf. Sec. A.2.6 for the definition of a n -form.

has $\epsilon(e_0, \dots, e_{n-1}) \neq 0$ and one says that the basis is **right-handed** (resp. **left-handed**) iff $\epsilon(e_0, \dots, e_{n-1}) > 0$ (resp. < 0). The components of ϵ with respect to (e_α) are

$$\epsilon_{\alpha_1 \dots \alpha_n} = \pm \sqrt{|g|} [\alpha_1, \dots, \alpha_n], \quad (\text{A.52})$$

where \pm must be $+$ (resp. $-$) for a right-handed (resp. left-handed) basis, g stands for the determinant of the matrix of g 's components with respect to the basis (e_α) :

$$g := \det(g_{\alpha\beta}) \quad (\text{A.53})$$

and the symbol $[\alpha_1, \dots, \alpha_n]$ takes the value 0 if any two indices $(\alpha_1, \dots, \alpha_n)$ are equal and takes the value 1 (resp. -1) if $(\alpha_1, \dots, \alpha_n)$ is an even (resp. odd) permutation of $(0, \dots, n-1)$.

A.3.5 Vector normal to a hypersurface

In a pseudo-Riemannian manifold, one can associate to any hypersurface Σ (cf. Sec. A.2.7) a unique normal direction, which can be represented by a nonvanishing vector field \mathbf{n} defined on Σ as follows. Locally the hypersurface Σ can be considered as a level set, i.e. there exists a smooth scalar field $f : \mathcal{M} \rightarrow \mathbb{R}$, such that $\mathbf{d}f \neq 0$ on Σ and for any point p in the considered local region of \mathcal{M} , the following equivalence holds

$$p \in \Sigma \iff f(p) = 0. \quad (\text{A.54})$$

Then, a vector field \mathbf{v} on \mathcal{M} is tangent to Σ iff the value of f stays to 0 for a small displacement $d\lambda$ along \mathbf{v} ; thanks to Eqs. (A.13), (A.14) and (A.12), this is equivalent to $\mathbf{v}(f) = v^\mu \partial f / \partial x^\mu = 0$, or to $\mathbf{g}(\mathbf{n}, \mathbf{v}) = 0$, where we have let appear the gradient vector $\mathbf{n} := \vec{\nabla} f$; in terms of components with respect to a coordinate system (x^α) :

$$n^\alpha = \nabla^\alpha f = g^{\alpha\mu} \frac{\partial f}{\partial x^\mu}. \quad (\text{A.55})$$

The vector field \mathbf{n} is called a **normal** to Σ . All normal vectors to Σ are collinear to each other.

It follows from the definitions given in Sec. A.3.2 that for any coordinate x^α of a given chart (x^0, \dots, x^{n-1}) ,

$$x^\alpha \text{ timelike coordinate} \iff \vec{\nabla} x^\alpha \text{ timelike vector} \iff g^{\alpha\alpha} < 0, \quad (\text{A.56a})$$

$$x^\alpha \text{ null coordinate} \iff \vec{\nabla} x^\alpha \text{ null vector} \iff g^{\alpha\alpha} = 0, \quad (\text{A.56b})$$

$$x^\alpha \text{ spacelike coordinate} \iff \vec{\nabla} x^\alpha \text{ spacelike vector} \iff g^{\alpha\alpha} > 0, \quad (\text{A.56c})$$

where $g^{\alpha\alpha}$ is the component (α, α) of the inverse metric (no summation on α). It appears here because, according to Eq. (A.55), $\mathbf{g}(\vec{\nabla} x^\alpha, \vec{\nabla} x^\alpha) = g_{\mu\nu} g^{\mu\rho} \partial x^\alpha / \partial x^\rho g^{\nu\sigma} \partial x^\alpha / \partial x^\sigma = \delta^\rho_\nu \delta^\alpha_\rho g^{\nu\sigma} \delta^\alpha_\sigma = g^{\alpha\alpha}$.

A.4 The three basic derivatives

Three derivative operators acting on tensor fields can be defined on a pseudo-Riemannian manifold (\mathcal{M}, g) . One of them depends on the metric g : the *covariant derivative* ∇ (Sec. A.4.1). Another one depends on the choice of a reference vector field: the *Lie derivative* \mathcal{L} (Sec. A.4.2). The third one, called the *exterior derivative* and denoted by d , depends only on the smooth-manifold structure, i.e. it is independent of any (metric or vector) field; on the other side, it is applicable only to a specific kind of tensor fields, namely differential forms (Sec. A.4.3).

A.4.1 Covariant derivative

Affine connection on a manifold

Let us denote by $\mathfrak{X}(\mathcal{M})$ the space of smooth vector fields on \mathcal{M} . $\mathfrak{X}(\mathcal{M})$ is an infinite-dimensional vector space over \mathbb{R} . Given a vector field $v \in \mathfrak{X}(\mathcal{M})$, it is not possible from the manifold structure alone to define its variation between two neighboring points p and q . Indeed a formula like $dv := v|_q - v|_p$ is meaningless because the vectors $v|_q$ and $v|_p$ belong to two distinct vector spaces, $T_q \mathcal{M}$ and $T_p \mathcal{M}$ respectively (cf. Fig. A.2), so that their subtraction is a priori not defined. Note that this issue does not arise for a scalar field [cf. Eq. (A.21)]. The solution is to introduce an extra-structure on \mathcal{M} , called an *affine connection*. The term *connection* arises because, by defining the variation of vector fields, this structure *connects* the various tangent spaces on the manifold. More precisely, an *affine connection* on \mathcal{M} is a map

$$\begin{aligned}\nabla : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) &\longrightarrow \mathfrak{X}(\mathcal{M}) \\ (\mathbf{u}, \mathbf{v}) &\longmapsto \nabla_{\mathbf{u}} \mathbf{v}\end{aligned}\tag{A.57}$$

that satisfies the following properties:

1. ∇ is bilinear (considering $\mathfrak{X}(\mathcal{M})$ as a vector space over \mathbb{R}).
2. For any scalar field f and any pair (\mathbf{u}, \mathbf{v}) of vector fields:

$$\nabla_{f\mathbf{u}} \mathbf{v} = f \nabla_{\mathbf{u}} \mathbf{v}.\tag{A.58}$$

3. For any scalar field f and any pair (\mathbf{u}, \mathbf{v}) of vector fields, the following Leibniz rule holds:

$$\nabla_{\mathbf{u}} (f\mathbf{v}) = \langle \mathbf{d}f, \mathbf{u} \rangle \mathbf{v} + f \nabla_{\mathbf{u}} \mathbf{v},\tag{A.59}$$

where $\mathbf{d}f$ is the differential of f as defined in Sec. A.2.4.

The vector $\nabla_{\mathbf{u}} \mathbf{v}$ is called the *covariant derivative of v along 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 the value of $\nabla_{\mathbf{u}} \mathbf{v}$ at a given point $p \in \mathcal{M}$ depends only on the vector $\mathbf{u}|_p \in T_p \mathcal{M}$ and not on the behavior of \mathbf{u} around p ; therefore the role of \mathbf{u} is only to give the direction of the derivative of \mathbf{v} .

Given an affine connection, the variation of a vector field v between two neighboring points, p and q say, is defined as

$$dv := \nabla_{dx} v, \quad (\text{A.60})$$

where dx is the infinitesimal displacement vector connecting p and q [cf Eq. (A.13)]. One says that v is **parallelly transported from p to q with respect to the connection ∇** iff $dv = 0$.

Given a frame field (e_α) on \mathcal{M} , the **connection coefficients** of an affine connection ∇ with respect to (e_α) are the n^3 scalar fields $\Gamma^\gamma{}_{\alpha\beta}$ defined by the expansion, at each point $p \in \mathcal{M}$, of the vector $\nabla_{e_\beta} e_\alpha|_p$ onto the basis $(e_\alpha|_p)$:

$$\boxed{\nabla_{e_\beta} e_\alpha =: \Gamma^\mu{}_{\alpha\beta} e_\mu}. \quad (\text{A.61})$$

An affine connection is entirely defined by its connection coefficients in a given frame field. 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 $v \in \mathfrak{X}(\mathcal{M})$, the **covariant derivative of v with respect to the affine connection ∇** is the tensor field ∇v of type $(1, 1)$ defined by the following action at each point $p \in \mathcal{M}$:

$$\begin{aligned} \nabla v|_p : T_p^* \mathcal{M} \times T_p \mathcal{M} &\longrightarrow \mathbb{R} \\ (\omega, u) &\longmapsto \langle \omega, \nabla_{\tilde{u}} v|_p \rangle, \end{aligned} \quad (\text{A.62})$$

where \tilde{u} is any vector field which performs some extension of u around p : $\tilde{u}|_p = u$. As already noted (cf. Remark 1), $\nabla_{\tilde{u}} v|_p$ is independent of the choice of \tilde{u} , so that the mapping (A.62) is well-defined. By comparing with (A.25), we verify that $\nabla v|_p$ is a tensor of type $(1, 1)$.

The **covariant derivative** is extended to all tensor fields by (i) demanding that for a scalar field it coincides with the differential: $\nabla f := df$ and (ii) using the Leibniz rule. As a result, the covariant derivative of a tensor field T of type (k, ℓ) is a tensor field ∇T of type $(k, \ell + 1)$. Its components with respect a given frame field (e_α) are denoted

$$\nabla_\gamma T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} := (\nabla T)^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell \gamma} \quad (\text{A.63})$$

and are given by

$$\begin{aligned} \nabla_\gamma T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} &= e_\gamma (T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell}) + \sum_{i=1}^k \Gamma^{\alpha_i}_{\sigma \gamma} T^{\overset{i}{\underset{\sigma}{\downarrow}} \dots \alpha_k}_{\beta_1 \dots \beta_\ell} \\ &\quad - \sum_{i=1}^\ell \Gamma^\sigma_{\beta_i \gamma} T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \overset{i}{\underset{\sigma}{\uparrow}} \dots \beta_\ell}, \end{aligned} \quad (\text{A.64})$$

where $e_\gamma (T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell})$ stands for the action of the vector e_γ on the scalar field $T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell}$ resulting from the very definition of a vector (cf. Sec. A.2.3). In particular, if (e_α) is a natural frame associated with some coordinate system (x^α) , then $e_\alpha = \partial_\alpha$ and $e_\gamma (T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell}) = \partial T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} / \partial x^\gamma$ [cf. Eq. (A.9)].

Remark 2: Notice the position of the index γ in Eq. (A.63): it is the last one on the right-hand side. According to (A.26), ∇T is then expressed as

$$\nabla T = \nabla_\gamma T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} e_{\alpha_1} \otimes \dots \otimes e_{\alpha_k} \otimes e^{\beta_1} \otimes \dots \otimes e^{\beta_\ell} \otimes e^\gamma. \quad (\text{A.65})$$

Because e^γ is the *last* 1-form of the tensorial product on the right-hand side, the notation $T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell; \gamma}$ instead of $\nabla_\gamma T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell}$ would have been more appropriate. The index convention (A.65) agrees with that of MTW [370] [cf. their Eq. (10.17)].

The **covariant derivative of the tensor field T along a vector v** is defined by

$$\nabla_v T := \nabla T(\underbrace{\dots}_{k+\ell \text{ slots}}, u). \quad (\text{A.66})$$

The components of $\nabla_v T$ are then $v^\mu \nabla_\mu T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell}$. Note that $\nabla_v T$ is a tensor field of the same type as T . In the particular case of a scalar field f , one has $\nabla_v f = \langle \nabla f, v \rangle = v(f)$.

The **divergence** with respect to the affine connection ∇ of a tensor field T of type (k, ℓ) with $k \geq 1$ is the tensor field denoted $\nabla \cdot T$ of type $(k-1, \ell)$ and whose components with respect to any frame field are given by

$$(\nabla \cdot T)^{\alpha_1 \dots \alpha_{k-1}}_{\beta_1 \dots \beta_\ell} = \nabla_\mu T^{\alpha_1 \dots \alpha_{k-1} \mu}_{\beta_1 \dots \beta_\ell}. \quad (\text{A.67})$$

Remark 3: For the divergence, the contraction is performed on the *last* upper index of T .

Levi-Civita connection

On a pseudo-Riemannian manifold (\mathcal{M}, g) there is a unique affine connection ∇ such that

1. ∇ is **torsion-free**, i.e. for any scalar field f , $\nabla \nabla f$ is a field of *symmetric* bilinear forms:

$$\nabla_\alpha \nabla_\beta f = \nabla_\beta \nabla_\alpha f. \quad (\text{A.68})$$

2. The covariant derivative of the metric tensor vanishes identically:

$$\boxed{\nabla g = 0}. \quad (\text{A.69})$$

∇ is called the **Levi-Civita connection associated with g** .

With respect to the Levi-Civita connection, the Levi-Civita tensor ϵ introduced in Sec. A.3.4 shares the same property as g :

$$\boxed{\nabla \epsilon = 0}. \quad (\text{A.70})$$

Given a coordinate system (x^α) on \mathcal{M} , the connection coefficients of the Levi-Civita connection with respect to the natural basis (∂_α) are called the **Christoffel symbols** of g ; they can be evaluated from the partial derivatives of the metric components with respect to the coordinates:

$$\Gamma^\gamma_{\alpha\beta} = \frac{1}{2} g^{\gamma\mu} \left(\frac{\partial g_{\mu\beta}}{\partial x^\alpha} + \frac{\partial g_{\alpha\mu}}{\partial x^\beta} - \frac{\partial g_{\alpha\beta}}{\partial x^\mu} \right). \quad (\text{A.71})$$

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. Indeed, combining Eqs. (A.67) and (A.64), we get for $\mathbf{v} \in \mathfrak{X}(\mathcal{M})$, $\nabla \cdot \mathbf{v} = \nabla_\mu v^\mu = \partial v^\mu / \partial x^\mu + \Gamma^\mu_{\sigma\mu} v^\sigma$. Now, from (A.71), we have

$$\Gamma^\mu_{\alpha\mu} = \frac{1}{2} g^{\mu\nu} \frac{\partial g_{\mu\nu}}{\partial x^\alpha} = \frac{1}{2} \frac{\partial}{\partial x^\alpha} \ln |g| = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\alpha} \sqrt{|g|}, \quad (\text{A.72})$$

where $g := \det(g_{\alpha\beta})$ [Eq. (A.53)]. The last but one equality follows from the general law of variation of the determinant of any invertible matrix A :

$$\delta(\ln |\det A|) = \text{tr}(A^{-1} \times \delta A), \quad (\text{A.73})$$

where δ denotes any variation that fulfills the Leibniz rule (i.e. a derivation), tr stands for the trace and \times for the matrix product. We conclude that

$$\boxed{\nabla \cdot \mathbf{v} = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\mu} \left(\sqrt{|g|} v^\mu \right).} \quad (\text{A.74})$$

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 type $(2, 0)$:

$$\boxed{\nabla_\mu A^{\alpha\mu} = \frac{1}{\sqrt{|g|}} \frac{\partial}{\partial x^\mu} \left(\sqrt{|g|} A^{\alpha\mu} \right).} \quad (\text{A.75})$$

A.4.2 Lie derivative

As discussed in Sec. A.4.1, the concept of a derivative of a vector field on a manifold \mathcal{M} requires the introduction of some extra-structure on \mathcal{M} . The extra-structure considered in Sec. A.4.1 is an affine connection, possibly provided by some metric tensor (case of the Levi-Civita connection). 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 \mathbf{u} on \mathcal{M} , which shall serve as our “reference flow”. Let \mathbf{v} be another vector field on \mathcal{M} , the variation of which is to be studied. One can use \mathbf{u} to transport \mathbf{v} from one point p to a neighboring one q and then define the variation of \mathbf{v} as the difference between the actual value $\mathbf{v}|_q$ of \mathbf{v} at q and the transported value via \mathbf{u} . More precisely, given an infinitesimal parameter ε , we define the *flow map along \mathbf{u}* as the map $\Phi_\varepsilon : \mathcal{M} \rightarrow \mathcal{M}$

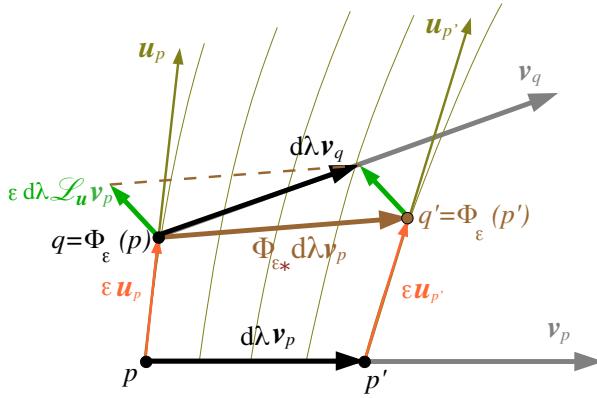


Figure A.3: Geometrical construction of the Lie derivative of a vector field v along a vector field u : given an infinitesimal parameter $d\lambda$, each extremity of the arrow $d\lambda v|_p$ is dragged by some small parameter ε along u , to form the vector denoted by $\Phi_{\varepsilon*} d\lambda v|_p$ (cf. Sec. A.2.8). The latter is then compared with the actual value of $d\lambda v$ at the point q , the difference (divided by $\varepsilon d\lambda$) defining the Lie derivative $\mathcal{L}_u v$ at p .

such that $\Phi_\varepsilon(p) = q$, where q is the point connected to p by the infinitesimal displacement vector $\vec{pq} = \varepsilon u|_p$ (cf. Sec. A.2.3 and Fig. A.3). The natural reference vector to compare $v|_q$ with is then the pushforward of $v|_p$ by Φ_ε , i.e. the vector $\Phi_{\varepsilon*} v|_p$ defined in Sec. A.2.8. Since $v|_q = v|_{\Phi_\varepsilon(p)}$ and $\Phi_{\varepsilon*} v|_p$ belong to the same vector space $T_q \mathcal{M}$, one may subtract the latter from the former and define the **Lie derivative** of v along u at p by

$$\boxed{\mathcal{L}_u v|_p := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left(v|_{\Phi_\varepsilon(p)} - \Phi_{\varepsilon*} v|_p \right)}. \quad (\text{A.76})$$

Remark 4: The term between parentheses in the right-hand side of Eq. (A.76) is the difference between two vectors of $T_q \mathcal{M}$ and thus a vector of $T_q \mathcal{M}$. One gets a vector of $T_p \mathcal{M}$ in the left-hand side only because $q = \Phi_\varepsilon(p)$ tends to p in the limit $\varepsilon \rightarrow 0$. Some authors (e.g. [342, 498]) prefer to define the Lie derivative from the difference of vectors in $T_p \mathcal{M}$, by carrying $v|_q$ to $T_p \mathcal{M}$ via the pushforward by $\Phi_\varepsilon^{-1} = \Phi_{-\varepsilon}$ and comparing it with $v|_p$; this results in the formula

$$\mathcal{L}_u v|_p := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left((\Phi_\varepsilon^{-1})_* v|_{\Phi_\varepsilon(p)} - v|_p \right). \quad (\text{A.77})$$

Thanks to the limit $\varepsilon \rightarrow 0$, formulas (A.76) and (A.77) are equivalent. Formula (A.76) is preferred here in so far as it corresponds to the geometrical construction displayed in Fig. A.3.

Let (x^α) be a coordinate system adapted to the field u in the sense that the first vector of the natural basis (∂_α) associated with (x^α) is $\partial_0 = u$. The coordinate expression of the flow map Φ_ε is then $(x^\alpha) \rightarrow (x^\alpha + \varepsilon \delta_0^\alpha)$. The corresponding Jacobian matrix is the identity matrix so that the coordinate components of $\Phi_{\varepsilon*} v|_p$ are identical to those of $v|_p$ (cf. Sec. A.2.7). It

follows that the component expression of Eq. (A.76) is

$$\begin{aligned} (\mathcal{L}_u v|_p)^\alpha &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \left[(v|_{\Phi_\varepsilon(p)})^\alpha - (v|_p)^\alpha \right] \\ &= \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} [v^\alpha(x^0 + \varepsilon, x^1, \dots, x^{n-1}) - v^\alpha(x^0, x^1, \dots, x^{n-1})]. \end{aligned}$$

Hence, in adapted coordinates, the Lie derivative is simply obtained by taking the partial derivative of the vector components with respect to x^0 :

$$\mathcal{L}_u v^\alpha = \frac{\partial v^\alpha}{\partial x^0}, \quad (\text{A.78})$$

where we have used the standard notation for the components of a Lie derivative: $\mathcal{L}_u v^\alpha := (\mathcal{L}_u v)^\alpha$. Besides, using the fact that the components of u are $u^\alpha = (1, 0, \dots, 0)$ in the adapted coordinate system, we notice that the components of the commutator of u and v , as given by (A.28), are $[u, v]^\alpha = \partial v^\alpha / \partial x^0$. This is exactly (A.78): $[u, v]^\alpha = \mathcal{L}_u v^\alpha$. We conclude that the Lie derivative of a vector with respect to another one is actually nothing but the commutator of these two vectors:

$$\boxed{\mathcal{L}_u v = [u, v].} \quad (\text{A.79})$$

Thanks to formula (A.28), we may then express the components of the Lie derivative in an arbitrary coordinate system:

$$\boxed{\mathcal{L}_u v^\alpha = u^\mu \frac{\partial v^\alpha}{\partial x^\mu} - v^\mu \frac{\partial u^\alpha}{\partial x^\mu}.} \quad (\text{A.80})$$

In view of the symmetry of the Christoffel symbols, the partial derivatives in Eq. (A.80) can be replaced by the covariant derivative ∇ , yielding

$$\mathcal{L}_u v^\alpha = u^\mu \nabla_\mu v^\alpha - v^\mu \nabla_\mu u^\alpha. \quad (\text{A.81})$$

Generalization to any tensor field

Let T be a tensor field of type $(0, \ell)$ on \mathcal{M} , with $\ell \geq 0$. The **Lie derivative** of T along u is then defined by comparing the pullback of T by the flow map Φ_ε at some point p (cf. Sec. A.2.8) to the actual value of T at the same point:

$$\boxed{\mathcal{L}_u T := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\Phi_\varepsilon^* T - T).} \quad (\text{A.82})$$

For a scalar field f , one has $\Phi_\varepsilon^* f = f \circ \Phi_\varepsilon$ (cf. Sec. A.2.8), so that the above definition with $\ell = 0$ yields $\mathcal{L}_u f = u(f) = \langle df, u \rangle = u^\mu \partial_\mu f$ [cf. Eq. (A.8)].

Finally, the Lie derivative is extended to any tensor field by means of the Leibniz rule. As a result, the Lie derivative $\mathcal{L}_u T$ of a tensor field 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

$$\begin{aligned} \mathcal{L}_u T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} &= u^\mu \frac{\partial}{\partial x^\mu} T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} - \sum_{i=1}^k T^{\alpha_1 \dots \overset{i}{\underset{\downarrow}{\sigma}} \dots \alpha_k}_{\beta_1 \dots \beta_\ell} \frac{\partial u^{\alpha_i}}{\partial x^\sigma} \\ &\quad + \sum_{i=1}^\ell T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \overset{i}{\underset{\uparrow}{\sigma}} \dots \beta_\ell} \frac{\partial u^\sigma}{\partial x^{\beta_i}}. \end{aligned} \quad (\text{A.83})$$

In particular, for a 1-form,

$$\mathcal{L}_u \omega_\alpha = u^\mu \frac{\partial \omega_\alpha}{\partial x^\mu} + \omega_\mu \frac{\partial u^\mu}{\partial x^\alpha}. \quad (\text{A.84})$$

In coordinates adapted to the vector field \mathbf{u} , we have $u^\alpha = (1, 0, \dots, 0)$, so that $u^\mu \partial/\partial x^\mu = \partial/\partial x^0$ and $\partial u^\alpha/\partial x^\beta = 0$. Accordingly, Eq. (A.83) reduces to

$$\mathcal{L}_u T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} = \frac{\partial}{\partial x^0} T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} \quad (\text{coordinates adapted to } \mathbf{u}). \quad (\text{A.85})$$

This generalizes Eq. (A.78) obtained for vector fields.

As for the vector case [Eq. (A.80)], the partial derivatives in Eq. (A.83) can be replaced by the covariant derivative ∇ (or any other connection without torsion), yielding

$$\begin{aligned} \mathcal{L}_u T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} &= u^\mu \nabla_\mu T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \beta_\ell} - \sum_{i=1}^k T^{\alpha_1 \dots \overset{i}{\underset{\downarrow}{\sigma}} \dots \alpha_k}_{\beta_1 \dots \beta_\ell} \nabla_\sigma u^{\alpha_i} \\ &\quad + \sum_{i=1}^\ell T^{\alpha_1 \dots \alpha_k}_{\beta_1 \dots \overset{i}{\underset{\uparrow}{\sigma}} \dots \beta_\ell} \nabla_{\beta_i} u^\sigma. \end{aligned} \quad (\text{A.86})$$

A special case of this formula is worth considering, namely $T = g$ (the metric tensor). Since $\nabla_\mu g_{\alpha\beta} = 0$, and $g_{\sigma\beta} \nabla_\alpha u^\sigma = \nabla_\alpha u_\beta$ (both thanks to Eq. (A.69)), one gets the so-called **Killing expression** of the Lie derivative of the metric tensor:

$$\boxed{\mathcal{L}_u g_{\alpha\beta} = \nabla_\alpha u_\beta + \nabla_\beta u_\alpha}. \quad (\text{A.87})$$

A.4.3 Exterior derivative

In Sec. A.2.6, we have introduced the *differential forms* or *p-forms* as tensor fields of type $(0, p)$, with $p \geq 0$, that are antisymmetric in all their arguments as soon as $p \geq 2$. Differential forms play a special role in the theory of integration on manifolds. Indeed, the primary definition of an integral over a manifold \mathcal{M} of dimension n is the integral of a n -form. The Levi-Civita tensor ϵ introduced in Sec. A.3.4 is a n -form, whose integral over a n -dimensional submanifold $\mathcal{D} \subset \mathcal{M}$ gives the volume of \mathcal{D} with respect to the metric g . The electromagnetic field is a 2-form (cf. Sec. 1.5.2) and the vorticity of a fluid is described by a 2-form as well in relativistic hydrodynamics (cf. e.g. Ref. [227]).

Being tensor fields, differential forms are subject to the covariant and Lie derivatives discussed above. In addition, they are subject to a third kind of derivative defined as follows. Given any p -form ω , its **exterior derivative** is the $(p+1)$ -form $d\omega$ whose action on a $(p+1)$ -tuple of vector fields (v_1, \dots, v_{p+1}) is given by

$$\begin{aligned} d\omega(v_1, \dots, v_{p+1}) &:= \sum_{i=1}^{p+1} (-1)^{i+1} v_i (\omega(v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_{p+1})) \\ &\quad + \sum_{\substack{i,j=1 \\ i < j}}^{p+1} (-1)^{i+j} \omega([v_i, v_j], v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_{j-1}, v_{j+1}, \dots, v_{p+1}), \end{aligned} \quad (\text{A.88})$$

where in the first line $\mathbf{v}_i(\cdots)$ stands for the action of the vector field \mathbf{v}_i on the scalar field $\omega(\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_{p+1})$ [cf. Eq. (A.8)]. For $p = 0$ (scalar field), Eq. (A.88) with $\omega \rightarrow f$ reduces to $\langle \mathbf{d}f, \mathbf{v}_1 \rangle := \mathbf{v}_1(f)$, i.e. one recovers Eq. (A.23). Hence, for a scalar field, the exterior derivative is nothing but the differential. For $p = 1$, Eq. (A.88) becomes

$$\mathbf{d}\omega(\mathbf{v}_1, \mathbf{v}_2) := \mathbf{v}_1(\langle \omega, \mathbf{v}_2 \rangle) - \mathbf{v}_2(\langle \omega, \mathbf{v}_1 \rangle) - \langle \omega, [\mathbf{v}_1, \mathbf{v}_2] \rangle. \quad (\text{A.89})$$

In terms of components with respect to a coordinate system (x^α) , the definition (A.88) yields

$$0\text{-form: } (\mathbf{d}\omega)_\alpha = \frac{\partial \omega}{\partial x^\alpha} \quad (\text{A.90a})$$

$$1\text{-form: } (\mathbf{d}\omega)_{\alpha\beta} = \frac{\partial \omega_\beta}{\partial x^\alpha} - \frac{\partial \omega_\alpha}{\partial x^\beta} \quad (\text{A.90b})$$

$$2\text{-form: } (\mathbf{d}\omega)_{\alpha\beta\gamma} = \frac{\partial \omega_{\beta\gamma}}{\partial x^\alpha} - \frac{\partial \omega_{\alpha\gamma}}{\partial x^\beta} + \frac{\partial \omega_{\alpha\beta}}{\partial x^\gamma} \quad (\text{A.90c})$$

$$p\text{-form: } (\mathbf{d}\omega)_{\alpha_1 \dots \alpha_{p+1}} = \sum_{i=1}^{p+1} (-1)^{i+1} \frac{\partial}{\partial x^{\alpha_i}} \omega_{\alpha_1 \dots \alpha_{i-1} \alpha_{i+1} \dots \alpha_{p+1}}. \quad (\text{A.90d})$$

Remark 5: The exterior derivative appeals only to the manifold structure; it does not depend upon the metric tensor g , nor upon any other extra structure on \mathcal{M} . Nevertheless, one may replace all partial derivatives in formulas (A.90) by covariant derivatives taken with respect to the Levi-Civita connection ∇ of g , thanks to the symmetry of the Christoffel symbols on their last two indices:

$$0\text{-form: } (\mathbf{d}\omega)_\alpha = \nabla_\alpha \omega \quad (\text{A.91a})$$

$$1\text{-form: } (\mathbf{d}\omega)_{\alpha\beta} = \nabla_\alpha \omega_\beta - \nabla_\beta \omega_\alpha \quad (\text{A.91b})$$

$$2\text{-form: } (\mathbf{d}\omega)_{\alpha\beta\gamma} = \nabla_\alpha \omega_{\beta\gamma} - \nabla_\beta \omega_{\alpha\gamma} + \nabla_\gamma \omega_{\alpha\beta} \quad (\text{A.91c})$$

$$p\text{-form: } (\mathbf{d}\omega)_{\alpha_1 \dots \alpha_{p+1}} = \sum_{i=1}^{p+1} (-1)^{i+1} \nabla_{\alpha_i} \omega_{\alpha_1 \dots \alpha_{i-1} \alpha_{i+1} \dots \alpha_{p+1}}. \quad (\text{A.91d})$$

A fundamental property of the exterior derivative is to be nilpotent:

$$\boxed{\mathbf{d}\mathbf{d}\omega = 0}. \quad (\text{A.92})$$

A p -form ω is said to be **closed** iff $\mathbf{d}\omega = 0$, and **exact** iff there exists a $(p-1)$ -form σ such that $\omega = \mathbf{d}\sigma$. From property (A.92), any exact p -form is closed. The **Poincaré lemma** states that the converse is true, at least locally, on a contractible open set.

Given a vector field \mathbf{v} on an oriented pseudo-Riemannian manifold (\mathcal{M}, g) of dimension n , the tensor field formed by setting the first argument of the Levi-Civita tensor ϵ to \mathbf{v} , while leaving free the other slots, i.e. $\epsilon(\mathbf{v}, ., ., .)$, is a $(n-1)$ -form denoted⁹ $\mathbf{v} \cdot \epsilon$. Its exterior derivative is thus a n -form, which is expressible in terms of the divergence of \mathbf{v} [cf. Eq. (A.74)] as

$$\boxed{\mathbf{d}(\mathbf{v} \cdot \epsilon) = (\nabla \cdot \mathbf{v}) \epsilon}. \quad (\text{A.93})$$

⁹More generally, the dot notation stands for the contraction on adjacent indices; here: $(\mathbf{v} \cdot \epsilon)_{\alpha_1 \dots \alpha_{n-1}} := v^\mu \epsilon_{\mu \alpha_1 \dots \alpha_{n-1}}$.

The exterior derivative enters in the famous ***Stokes' theorem***: if \mathcal{D} is an oriented d -dimensional manifold with boundary (cf. Sec. A.2.2), then for any $(d-1)$ -form ω on \mathcal{D} ,

$$\boxed{\int_{\partial\mathcal{D}} \omega = \int_{\mathcal{D}} d\omega}, \quad (\text{A.94})$$

where $\partial\mathcal{D}$ is \mathcal{D} 's boundary oriented according to the ***outward convention***: if $\epsilon_{\mathcal{D}}$ is the d -form defining the orientation of \mathcal{D} (cf. Sec. A.3.4), the $(d-1)$ -form defining the orientation of $\partial\mathcal{D}$ is chosen to be $v \cdot \epsilon_{\mathcal{D}}$, where v is an outward-pointing¹⁰ vector field along $\partial\mathcal{D}$. Note that each side of (A.94) is (of course!) well-defined, as the integral of a p -form over an oriented p -dimensional manifold.

Remark 6: If \mathcal{D} is a submanifold of an oriented pseudo-Riemannian manifold (\mathcal{M}, g) of dimension n , the orientation of \mathcal{D} is naturally set by choosing $\epsilon_{\mathcal{D}} = \epsilon$ if $d = n$, where ϵ is the Levi-Civita tensor of g (cf. Sec. A.3.4). If $d < n$, a free family of $n-d$ vector fields (u_1, \dots, u_{n-d}) nowhere tangent to \mathcal{D} has to be chosen first and $\epsilon_{\mathcal{D}}$ is then defined by $\epsilon_{\mathcal{D}} = \epsilon(u_1, \dots, u_{n-d}, \dots, \dots)$. Accordingly, the orientation of $\partial\mathcal{D}$ is set by $\epsilon_{\partial\mathcal{D}} = v \cdot \epsilon_{\mathcal{D}} = \epsilon(u_1, \dots, u_{n-d}, v, \dots, \dots)$.

Another important formula involving the exterior derivative is the ***Cartan identity***, which expresses the Lie derivative of a p -form ω along a vector field u as

$$\boxed{\mathcal{L}_u \omega = u \cdot d\omega + d(u \cdot \omega)}. \quad (\text{A.95})$$

As above, a dot denotes the contraction on adjacent indices: $u \cdot d\omega$ is the p -form $d\omega(u, \dots, \dots)$ and $u \cdot \omega$ is the $(p-1)$ -form $\omega(u, \dots, \dots)$. Notice that if ω is a 1-form, Eq. (A.95) is readily obtained by combining Eqs. (A.84), (A.90a) and (A.90b). An immediate consequence of the Cartan identity and the nilpotence property (A.92) is that the Lie derivative and the exterior derivative commute, i.e. for any vector field u and any p -form ω :

$$\mathcal{L}_u d\omega = d \mathcal{L}_u \omega. \quad (\text{A.96})$$

A.5 Curvature

A.5.1 General definition

The ***Riemann curvature tensor*** of an affine connection ∇ (cf. Sec. A.4.1) is defined by

$$\begin{aligned} \text{Riem} : \quad \mathfrak{X}^*(\mathcal{M}) \times \mathfrak{X}(\mathcal{M})^3 &\longrightarrow C^\infty(\mathcal{M}, \mathbb{R}) \\ (\omega, w, u, v) &\longmapsto \left\langle \omega, \nabla_u \nabla_v w - \nabla_v \nabla_u w - \nabla_{[u,v]} w \right\rangle, \end{aligned} \quad (\text{A.97})$$

where $\mathfrak{X}^*(\mathcal{M})$ stands for the space of 1-forms on \mathcal{M} , $\mathfrak{X}(\mathcal{M})$ for the space of vector fields on \mathcal{M} and $C^\infty(\mathcal{M}, \mathbb{R})$ for the space of smooth scalar fields on \mathcal{M} . The above formula does define a tensor field on \mathcal{M} , i.e. the value of $\text{Riem}(\omega, w, u, v)$ at a given point $p \in \mathcal{M}$ depends only

¹⁰By ***outward-pointing***, it is meant that at each point $p \in \partial\mathcal{D}$, $v|_p$ is not tangent to $\partial\mathcal{D}$ and there exists a curve in \mathcal{D} starting from p whose tangent vector at p is $-v|_p$.

upon the values of the fields ω , w , u and v at p and not upon their behaviors away from p , as the covariant derivatives in Eq. (A.97) might suggest. We denote the components of this tensor in a given basis (e_α) by $R^\gamma_{\delta\alpha\beta}$, instead of $\text{Riem}^\gamma_{\delta\alpha\beta}$. The definition (A.97) leads to the following formula, named the **Ricci identity**:

$$\forall \mathbf{w} \in \mathfrak{X}(\mathcal{M}), \quad (\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha) w^\gamma = R^\gamma_{\mu\alpha\beta} w^\mu. \quad (\text{A.98})$$

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

$$R^\alpha_{\beta\mu\nu} = \frac{\partial \Gamma^\alpha_{\beta\nu}}{\partial x^\mu} - \frac{\partial \Gamma^\alpha_{\beta\mu}}{\partial x^\nu} + \Gamma^\alpha_{\sigma\mu} \Gamma^\sigma_{\beta\nu} - \Gamma^\alpha_{\sigma\nu} \Gamma^\sigma_{\beta\mu}. \quad (\text{A.99})$$

From the definition (A.97), the Riemann tensor is clearly antisymmetric with respect to its last two arguments:

$$\text{Riem}(\cdot, \cdot, \mathbf{u}, \mathbf{v}) = -\text{Riem}(\cdot, \cdot, \mathbf{v}, \mathbf{u}). \quad (\text{A.100})$$

In addition, it satisfies the cyclic property

$$\text{Riem}(\cdot, \mathbf{u}, \mathbf{v}, \mathbf{w}) + \text{Riem}(\cdot, \mathbf{w}, \mathbf{u}, \mathbf{v}) + \text{Riem}(\cdot, \mathbf{v}, \mathbf{w}, \mathbf{u}) = 0. \quad (\text{A.101})$$

The covariant derivatives of the Riemann tensor obey the **Bianchi identity**

$$\nabla_\rho R^\alpha_{\beta\mu\nu} + \nabla_\mu R^\alpha_{\beta\nu\rho} + \nabla_\nu R^\alpha_{\beta\rho\mu} = 0. \quad (\text{A.102})$$

A.5.2 Case of a pseudo-Riemannian manifold

The Riemann tensor of the Levi-Civita connection obeys the additional antisymmetry:

$$\text{Riem}(\omega, \mathbf{w}, \cdot, \cdot) = -\text{Riem}(\underline{\mathbf{w}}, \overrightarrow{\omega}, \cdot, \cdot). \quad (\text{A.103})$$

Combined with (A.100) and (A.101), this implies the symmetry property

$$\text{Riem}(\omega, \mathbf{w}, \mathbf{u}, \mathbf{v}) = \text{Riem}(\underline{\mathbf{u}}, \underline{\mathbf{v}}, \overrightarrow{\omega}, \mathbf{w}). \quad (\text{A.104})$$

A pseudo-Riemannian manifold (\mathcal{M}, g) with a vanishing Riemann tensor is called a **flat manifold**; in this case, g is said to be a **flat metric**. If in addition, g has a Riemannian (resp. Lorentzian) signature, it is called an **Euclidean metric** (resp. **Minkowski metric**).

A.5.3 Ricci tensor

The **Ricci tensor** of the affine connection ∇ is the field of bilinear forms \mathbf{R} defined by

$$\begin{aligned} \mathbf{R} : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) &\longrightarrow C^\infty(\mathcal{M}, \mathbb{R}) \\ (\mathbf{u}, \mathbf{v}) &\longmapsto \text{Riem}(e^\mu, \mathbf{u}, e_\mu, \mathbf{v}), \end{aligned} \quad (\text{A.105})$$

where (e_α) is a vector frame on \mathcal{M} and (e^α) its dual counterpart. This definition is independent of the choice of (e_α) . In terms of components:

$$R_{\alpha\beta} := R^\mu{}_{\alpha\mu\beta}. \quad (\text{A.106})$$

Remark 2: Following 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 \mathbf{R} , when employing the index-free notation.

The Ricci tensor naturally appears in the **contracted Ricci identity**:

$$\forall \mathbf{w} \in \mathfrak{X}(\mathcal{M}), \quad [\nabla_\mu \nabla_\alpha w^\mu - \nabla_\alpha \nabla_\mu w^\mu = R_{\mu\alpha} w^\mu], \quad (\text{A.107})$$

which is obtained by taking the trace of the Ricci identity (A.98) on the indices (α, γ) and relabelling $\beta \rightarrow \alpha$.

For the Levi-Civita connection associated with the metric g , property (A.104) implies that the Ricci tensor is symmetric:

$$\mathbf{R}(\mathbf{u}, \mathbf{v}) = \mathbf{R}(\mathbf{v}, \mathbf{u}). \quad (\text{A.108})$$

In addition, one defines the **Ricci scalar** (also called **scalar curvature**) as the trace of the Ricci tensor with respect to the metric g :

$$R := g^{\mu\nu} R_{\mu\nu}. \quad (\text{A.109})$$

The Bianchi identity (A.102) implies the divergence-free property

$$[\nabla \cdot \vec{G} = 0], \quad (\text{A.110})$$

where \vec{G} is the type-(1, 1) tensor associated by metric duality [cf. (A.48)] to the **Einstein tensor**:

$$\boxed{\mathbf{G} := \mathbf{R} - \frac{1}{2} R \mathbf{g}}. \quad (\text{A.111})$$

Equation (A.110) is called the **contracted Bianchi identity**.

A.5.4 Weyl tensor

Let (\mathcal{M}, g) be a pseudo-Riemannian manifold of dimension n .

For $n = 1$, the Riemann tensor vanishes identically, i.e. (\mathcal{M}, g) is necessarily flat. The reader having 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 curved line \mathcal{L} , the curvature that is not vanishing is the *extrinsic* curvature, i.e. the curvature resulting from the embedding of \mathcal{L} in \mathbb{R}^2 .

For $n = 2$, the Riemann tensor is entirely determined by the Ricci scalar R , according to the formula:

$$R^\gamma_{\delta\alpha\beta} = \frac{R}{2} (\delta^\gamma_\alpha g_{\delta\beta} - \delta^\gamma_\beta g_{\delta\alpha}) \quad (n = 2). \quad (\text{A.112})$$

For $n = 3$, the Riemann tensor is entirely determined by the Ricci tensor, according to

$$\begin{aligned} R^\gamma_{\delta\alpha\beta} &= R^\gamma_\alpha g_{\delta\beta} - R^\gamma_\beta g_{\delta\alpha} + \delta^\gamma_\alpha R_{\delta\beta} - \delta^\gamma_\beta R_{\delta\alpha} \\ &\quad + \frac{R}{2} (\delta^\gamma_\beta g_{\delta\alpha} - \delta^\gamma_\alpha g_{\delta\beta}) \quad (n = 3). \end{aligned} \quad (\text{A.113})$$

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 \mathbf{R} [Eq. (A.106)], and (iii) a “traceless” part, which is constituted by the **Weyl conformal curvature tensor**, \mathbf{C} :

$$\begin{aligned} R^\gamma_{\delta\alpha\beta} &= C^\gamma_{\delta\alpha\beta} + \frac{1}{n-2} (R^\gamma_\alpha g_{\delta\beta} - R^\gamma_\beta g_{\delta\alpha} + \delta^\gamma_\alpha R_{\delta\beta} - \delta^\gamma_\beta R_{\delta\alpha}) \\ &\quad + \frac{1}{(n-1)(n-2)} R (\delta^\gamma_\beta g_{\delta\alpha} - \delta^\gamma_\alpha g_{\delta\beta}). \end{aligned} \quad (\text{A.114})$$

The above relation may be taken as the definition of \mathbf{C} . It implies that \mathbf{C} is traceless: $C^\mu_{\alpha\mu\beta} = 0$. The other possible traces are zero thanks to the symmetries of the Riemann tensor.

Remark 3: The decomposition (A.114) is meaningful for $n = 3$ as well, but then it implies that the Weyl tensor vanishes identically [compare with (A.113)].

Appendix B

Geodesics

Contents

B.1	Introduction	737
B.2	Definition and first properties	737
B.3	Existence and uniqueness of geodesics	742
B.4	Geodesics and extremal lengths	748
B.5	Geodesics and symmetries	755
B.6	Geodesics and curvature	757

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 $\mathcal{L} = P(I)$ of a map (called a **parametrization** of the curve) $P : I \rightarrow \mathcal{M}$, $\lambda \mapsto P(\lambda)$, 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 \mathcal{L} is reduced to a single point of \mathcal{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 \mathcal{L} of a pseudo-Riemannian manifold (\mathcal{M}, g) is called a **geodesic** iff it admits a parametrization P whose associated tangent vector field v is transported parallelly to itself along \mathcal{L} :

$$\boxed{\nabla_v v = 0}, \quad (\text{B.1})$$

where ∇ is the Levi-Civita connection of the metric g . The parametrization P is then called an **affine parametrization** and the corresponding parameter λ is called an **affine parameter** of the geodesic \mathcal{L} . Note that the relation between v and λ is

$$v = \frac{dx}{d\lambda}, \quad (\text{B.2})$$

where dx is the infinitesimal displacement along \mathcal{L} corresponding to the change $d\lambda$ in the parameter λ (cf. Eq. (A.14)).

The qualifier *affine* in the above definition stems from:

Property B.1: affine parametrizations of a geodesic

Any two affine parametrizations of a geodesic \mathcal{L} are necessarily related by an affine transformation:

$$\lambda' = a\lambda + b, \quad (\text{B.3})$$

where a and b are two real constants such that $a \neq 0$.

Proof. Let $P : I \rightarrow \mathcal{L} \subset \mathcal{M}$, $\lambda \mapsto P(\lambda)$ and $P' : I' \rightarrow \mathcal{L}$, $\lambda' \mapsto P'(\lambda')$ be two parametrizations of \mathcal{L} . They are necessarily related by a diffeomorphism $I \rightarrow I'$, $\lambda \mapsto \lambda'(\lambda)$. It follows from Eq. (B.2) that the tangent vector fields v and v' associated with these two parametrizations are related by

$$v = \frac{d\lambda'}{d\lambda} v'. \quad (\text{B.4})$$

Using the rules 2 and 3 governing the connection ∇ (cf. Sec. A.4.1), we get then

$$\nabla_v v = \frac{d^2\lambda'}{d\lambda^2} v' + \left(\frac{d\lambda'}{d\lambda} \right)^2 \nabla_{v'} v'. \quad (\text{B.5})$$

¹As already noticed (cf. Remark 6 p. 712), in the mathematical literature, it is common to define a curve as the parametrization itself, and not as its image.

If both parametrizations are affine, then $\nabla_v v = 0$ and $\nabla_{v'} v' = 0$, so that the above identity reduces to $d^2\lambda'/d\lambda^2 = 0$, which implies the affine law (B.3). \square

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 $\nabla_v v$ being considered as the “acceleration” of (the parametrization P of) the curve \mathcal{L} ; this is of course by extension of the concept of 4-acceleration $a := \nabla_u u$ of a timelike worldline with 4-velocity 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

Property B.2: constancy of the scalar square of affine tangent vector fields

Let \mathcal{L} be a geodesic of (\mathcal{M}, g) and v a tangent vector field associated with an affine parametrization of \mathcal{L} . Then the scalar square of v with respect to the metric g is constant along \mathcal{L} :

$$g(v, v) = \text{const.} \quad (\text{B.6})$$

Proof. The variation of $g(v, v)$ along \mathcal{L} is given by

$$\begin{aligned} \frac{d}{d\lambda}(g(v, v)) &= v(g(v, v)) = \nabla_v(g(v, v)) = v^\mu \nabla_\mu(g_{\rho\sigma} v^\rho v^\sigma) \\ &= v^\mu \underbrace{\nabla_\mu g_{\rho\sigma}}_0 v^\rho v^\sigma + g_{\rho\sigma} v^\mu \underbrace{\nabla_\mu v^\rho}_0 v^\sigma + g_{\rho\sigma} v^\rho \underbrace{\nabla_\mu v^\sigma}_0 = 0, \end{aligned}$$

where we have used the fact that ∇ is the Levi-Civita connection of g [Eq. (A.69)] and v obeys the geodesic equation (B.1). \square

The constancy of $g(v, v)$ has an interesting corollary: the tangent vector v cannot change its causal type along \mathcal{L} . Hence:

Property B.3: type of geodesics on pseudo-Riemannian manifolds

In a pseudo-Riemannian manifold (\mathcal{M}, g) , a geodesic \mathcal{L} necessarily belongs to one of the following three categories:

- **timelike geodesic:** tangent vectors are timelike at all points of \mathcal{L} ;
- **null geodesic:** tangent vectors are null at all points of \mathcal{L} ;
- **spacelike geodesic:** tangent vectors are spacelike at all points of \mathcal{L} .

This is in sharp contrast with generic curves, which, for instance, can be timelike on some portions and null or spacelike on other ones.

In the timelike case, or the spacelike one, the tangent vector field v can be rescaled by the constant $\sqrt{|g(v, v)|}$ to get a unit tangent vector field, i.e. a tangent vector field u which obeys $g(u, u) = -1$ (timelike geodesic) or $g(u, u) = 1$ (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 $a = \sqrt{|g(v, v)|}$. Thus, we have

Property B.4: affine parametrizations with unit tangent vectors

A timelike or spacelike geodesic of a pseudo-Riemannian manifold has an affine parametrization, the tangent vector of which is a unit vector (i.e. of scalar square ± 1 with respect to 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:

Property B.5: generic tangent vector field along a geodesic

A curve \mathcal{L} is a geodesic iff the tangent vector field v associated with any parametrization of \mathcal{L} obeys

$$\nabla_v v = \kappa v, \quad (\text{B.7})$$

where κ is a scalar field along \mathcal{L} .

Proof. Let $P : I \rightarrow \mathcal{L}, \lambda \mapsto P(\lambda)$ be the parametrization of \mathcal{L} corresponding to the tangent vector field v : $v = dx/d\lambda$. If \mathcal{L} is a geodesic, then there exists a parametrization $\lambda' \mapsto P'(\lambda')$ whose tangent vector, v' say, obeys $\nabla_{v'} v' = 0$. Since the accelerations of any two parametrizations of \mathcal{L} are related by Eq. (B.5), we deduce that v obeys (B.7) with

$$\kappa = \left(\frac{d\lambda'}{d\lambda} \right)^{-1} \frac{d^2\lambda'}{d\lambda^2}.$$

Conversely, if v obeys (B.7) with $\kappa = \kappa(\lambda)$, then Eq. (B.5) implies that $\nabla_{v'} v' = 0$, i.e. that \mathcal{L} is a geodesic, provided that the change of parametrization $\lambda' = \lambda'(\lambda)$ fulfills

$$\kappa(\lambda) \frac{d\lambda'}{d\lambda} - \frac{d^2\lambda'}{d\lambda^2} = 0.$$

This differential equation has the following solution:

$$\lambda' = a \int_{\lambda_1}^{\lambda} \left[\exp \left(\int_{\lambda_0}^{\tilde{\lambda}} \kappa(\tilde{\lambda}) d\tilde{\lambda} \right) d\tilde{\lambda} \right] + b,$$

where a, b, λ_0 and λ_1 are constants, with $a \neq 0$ and $\lambda_0, \lambda_1 \in I$. □

Property B.5 motivates the following definitions:

A vector field v obeying (B.7) is called a **pregeodesic vector field**. The scalar field κ is then called the **non-affinity coefficient** of v . If $\kappa = 0$, v is naturally called a **geodesic vector field**.

Note that Property B.5 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:

Property B.6: evolution of the scalar square of a tangent vector field

Along a geodesic \mathcal{L} , the scalar square $g(v, v)$ of the tangent vector v associated with any parametrization of \mathcal{L} evolves according to

$$\nabla_v [g(v, v)] = 2\kappa g(v, v), \quad (\text{B.8})$$

where κ is the non-affinity coefficient of v .

Proof. One has, using $\nabla g = 0$ [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 Eq. (B.6) in the special case $\kappa = 0$ (v geodesic vector).

Remark 2: If λ is the parameter associated with v , i.e. $v = dx/d\lambda$, we may introduce the scalar function $V(\lambda) := g(v, v)$ and rewrite (B.8) as a first-order ordinary differential equation for it [cf. Eq. (A.8)]:

$$\frac{dV}{d\lambda} = 2\kappa(\lambda)V(\lambda). \quad (\text{B.9})$$

A consequence of (B.8) is

Property B.7: proper time and arc length as affine parameters

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 u (cf. Sec. 1.3.3), which is of constant scalar square: $g(u, u) = -1$, 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

Property B.8: geodesic equation

Let \mathcal{L} be a curve in a pseudo-Riemannian manifold $(\mathcal{M}, \mathbf{g})$ of dimension n , such that \mathcal{L} is contained in the domain of a coordinate chart $(x^\alpha)_{0 \leq \alpha \leq n-1}$. Then any parametrization of \mathcal{L} , $P : I \rightarrow \mathcal{L}$, $\lambda \mapsto P(\lambda)$, can be described by n functions $X^\alpha : I \rightarrow \mathbb{R}$ according to Eq. (A.7): $x^\alpha(P(\lambda)) = X^\alpha(\lambda)$. The curve \mathcal{L} is a geodesic iff there exists a parametrization of \mathcal{L} for which the functions X^α fulfills the following system of n second-order differential equations, called the **geodesic equation**:

$$\boxed{\frac{d^2X^\alpha}{d\lambda^2} + \Gamma_{\mu\nu}^\alpha \frac{dX^\mu}{d\lambda} \frac{dX^\nu}{d\lambda} = 0}, \quad 0 \leq \alpha \leq n-1, \quad (\text{B.10})$$

where the $\Gamma_{\mu\nu}^\alpha$'s are the Christoffel symbols of the metric \mathbf{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 \mathbf{v} associated with the parameter λ are [cf. Eq. (A.11)]

$$v^\alpha = \frac{dX^\alpha}{d\lambda}. \quad (\text{B.11})$$

On the other side, the components of $\nabla_{\mathbf{v}} \mathbf{v}$ are

$$v^\mu \nabla_\mu v^\alpha = v^\mu \frac{\partial v^\alpha}{\partial x^\mu} + \Gamma_{\mu\nu}^\alpha v^\mu v^\nu = \mathbf{v}(v^\alpha) + \Gamma_{\mu\nu}^\alpha v^\mu v^\nu = \frac{dv^\alpha}{d\lambda} + \Gamma_{\mu\nu}^\alpha v^\mu v^\nu,$$

where we have used successively Eqs. (A.64), (A.12) 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 $\nabla_{\mathbf{v}} \mathbf{v}$. The conclusion follows from the very definition of a geodesic given in Sec. B.2.1. \square

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 $\kappa dX^\alpha/d\lambda$, which is the coordinate expression of the right-hand side $\kappa \mathbf{v}$ in Eq. (B.7). Hence, we have

Property B.9: pregeodesic equation

A curve \mathcal{L} in the domain of a chart (x^α) is a geodesic iff some (actually all) coordinate expression $x^\alpha = X^\alpha(\lambda)$ of \mathcal{L} fulfills the following system of n second-order differential equations, usually called the **pregeodesic equation**,

$$\boxed{\frac{d^2X^\alpha}{d\lambda^2} + \Gamma_{\mu\nu}^\alpha \frac{dX^\mu}{d\lambda} \frac{dX^\nu}{d\lambda} = \kappa(\lambda) \frac{dX^\alpha}{d\lambda}}, \quad 0 \leq \alpha \leq n-1. \quad (\text{B.12})$$

for some real-valued function $\kappa(\lambda)$.

B.3.2 Existence and uniqueness

We may use the geodesic equation to prove the following theorem:

Property B.10: existence and uniqueness of geodesic from a vector at a point

Given a point p in a pseudo-Riemannian manifold (\mathcal{M}, g) and a vector \mathbf{V} in the tangent space to \mathcal{M} at p , i.e. $\mathbf{V} \in T_p \mathcal{M}$, there exists a geodesic \mathcal{L} through p such that \mathbf{V} is the value at p of the tangent vector of some affine parametrization λ of \mathcal{L} :

$$\mathbf{V} = \left. \frac{d\mathbf{x}}{d\lambda} \right|_p. \quad (\text{B.13})$$

Moreover, this geodesic is unique, in the sense that any geodesic \mathcal{L}' sharing the same property coincides with \mathcal{L} in some open neighborhood of p .

Proof. Let (x^α) be a coordinate chart of \mathcal{M} around p . Let (V^α) be the components of \mathbf{V} in the basis of $T_p \mathcal{M}$ induced by the coordinate frame (∂_α) associated with (x^α) :

$$\mathbf{V} = V^\alpha \partial_\alpha|_p.$$

A geodesic through p having \mathbf{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)]

$$X^\alpha(0) = x^\alpha(p) \quad \text{and} \quad \left. \frac{dX^\alpha}{d\lambda} \right|_0 = V^\alpha. \quad (\text{B.14})$$

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. \square

A few definitions follow naturally:

A geodesic \mathcal{L} is said to be **inextendible** or **maximal** iff there does not exist any geodesic $\mathcal{L}' \neq \mathcal{L}$ such that $\mathcal{L} \subset \mathcal{L}'$.

A geodesic \mathcal{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:

Property B.11: inextendibility of complete geodesics

Any complete geodesic is inextendible.

Proof. Let \mathcal{L} be a complete geodesic. Let us consider any geodesic \mathcal{L}' such that $\mathcal{L} \subset \mathcal{L}'$. Let λ and λ' be affine parameters of respectively \mathcal{L} and \mathcal{L}' . Since $\mathcal{L} \subset \mathcal{L}'$, λ' is also an affine parameter of \mathcal{L} and we must have, along \mathcal{L} , $\lambda' = a\lambda + b$ with $a \neq 0$ [Eq. (B.3)]. Since the range of λ is $(-\infty, +\infty)$, for \mathcal{L} is complete, this implies that the range of λ' on \mathcal{L} is $(-\infty, +\infty)$ as well, which make impossible to have points in $\mathcal{L}' \setminus \mathcal{L}$. Hence $\mathcal{L}' = \mathcal{L}$, i.e. \mathcal{L} is inextendible. \square

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 \mathcal{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 \mathcal{L} at some (curvature) singularity. A spacetime in which this does not occur is called **timelike geodesically complete**. More generally:

The pseudo-Riemannian manifold (\mathcal{M}, 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 length² of all curves from p to q (see e.g. Ref. [343]). However, there is no such theorem for a *Lorentzian* manifold, for the metric does not induce any distance function which would make the manifold a metric space.

The following proposition strengthens Property B.10:

Property B.12: existence and uniqueness of inextendible geodesic from a vector

Given a point p in a pseudo-Riemannian manifold (\mathcal{M}, g) and a nonzero vector V in the tangent space to \mathcal{M} at p , i.e. $V \in T_p \mathcal{M}$, there exists a unique inextendible geodesic through p , which we shall denote by \mathcal{L}_V , such that V is the value at p of the tangent vector of some affine parametrization of \mathcal{L}_V . We shall then denote by P_V the unique affine parametrization of \mathcal{L}_V such that

$$P_V(0) = p \quad \text{and} \quad v|_p = V, \quad (\text{B.15})$$

where v is the tangent vector field of P_V .

We refer to O’Neill’s textbook [389], p. 68 for the proof.

²The length of a curve is defined by Eq. (B.24) below.

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 (\mathcal{M}, g) , let \mathcal{E}_p be the subset of the tangent space $T_p \mathcal{M}$ defined by $\mathbf{V} \in \mathcal{E}_p$ iff either $\mathbf{V} = 0$ or the affine parametrization $P_{\mathbf{V}}$ of the geodesic $\mathcal{L}_{\mathbf{V}}$ associated to \mathbf{V} by Property B.12 has a domain large enough to include the interval $[0, 1]$. The **exponential map at p** is then defined as

$$\begin{aligned} \exp_p : \mathcal{E}_p \subset T_p \mathcal{M} &\longrightarrow \mathcal{M} \\ \mathbf{V} &\longmapsto \begin{cases} p & \text{if } \mathbf{V} = 0 \\ P_{\mathbf{V}}(1) \in \mathcal{L}_{\mathbf{V}} & \text{if } \mathbf{V} \neq 0 \end{cases} \end{aligned} \quad (\text{B.16})$$

In other words, \exp_p maps a nonzero vector \mathbf{V} in the tangent space to \mathcal{M} at p to the point of \mathcal{M} of affine parameter $\lambda = 1$ along the unique geodesic through p such that (i) $\lambda = 0$ corresponds to p and (ii) the tangent vector $d\mathbf{x}/d\lambda$ to the geodesic at p is \mathbf{V} .

Note that if (\mathcal{M}, g) is geodesically complete, $\mathcal{E}_p = T_p \mathcal{M}$ for every point $p \in \mathcal{M}$.

An immediate property of the exponential map is

Property B.13

If $\mathbf{V} \in \mathcal{E}_p \setminus \{0\}$, for any $t \in [0, 1]$, $\exp_p(t\mathbf{V})$ lies on the same geodesic $\mathcal{L}_{\mathbf{V}}$ as $\exp_p(\mathbf{V})$, at the parameter $\lambda = t$ of the parametrization $P_{\mathbf{V}}$:

$$\forall t \in [0, 1], \quad \exp_p(t\mathbf{V}) = P_{\mathbf{V}}(t). \quad (\text{B.17})$$

Proof. For $t = 0$, the property follows from the definition of \exp_p , since $P_{\mathbf{V}}(0) = p$. If $t \neq 0$, the nonzero vector $t\mathbf{V}$ is collinear to \mathbf{V} and the uniqueness property of geodesics (Property B.12) implies that $\mathcal{L}_{t\mathbf{V}} = \mathcal{L}_{\mathbf{V}}$. By virtue of the transformation law (B.4), $t\mathbf{V}$ is the tangent vector to $\mathcal{L}_{\mathbf{V}}$ corresponding to the affine parameter $\lambda' = t^{-1}\lambda$, where λ is the affine parameter whose tangent vector field obeys $v|_p = \mathbf{V}$. From the definition of \exp_p , we have then $\exp_p(t\mathbf{V}) = P_{t\mathbf{V}}(\lambda' = 1) = P_{\mathbf{V}}(\lambda = t \times 1)$, hence (B.17). \square

The exponential map realizes a local identification of the manifold with its tangent space at a given point:

Property B.14: exponential map as a local diffeomorphism

For any $p \in \mathcal{M}$, there exists a neighborhood U of 0 in the tangent space $T_p \mathcal{M}$ and a neighborhood \mathcal{U} of p in the manifold \mathcal{M} such that the exponential map \exp_p is a diffeomorphism from U to \mathcal{U} .

Proof. It is clear from its definition that \exp_p is a smooth map, at least on some neighborhood U' of 0 in $T_p\mathcal{M}$. We may then consider the differential of \exp_p at 0, $d\exp_p|_0$. By virtue of the inverse function theorem for manifolds (see e.g. Theorem 4.5 in Ref. [342]), it suffices to show that $d\exp_p|_0$ is invertible to complete the proof. By definition of the differential of a map (cf. Sec. A.2.7) and since $\exp_p : \mathcal{E}_p \subset T_p\mathcal{M} \rightarrow \mathcal{M}$ and $\exp_p(0) = p$, $d\exp_p|_0$ carries an infinitesimal displacement vector of³ $T_0(T_p\mathcal{M})$, ε say, connecting 0 to a nearby element of $T_p\mathcal{M}$, ε' say, to the infinitesimal vector $\mathbf{E} \in T_p\mathcal{M}$ connecting $p = \exp_p(0)$ to $q = \exp_p(\varepsilon')$:

$$\begin{aligned} d\exp_p|_0 : T_0(T_p\mathcal{M}) &\longrightarrow T_p\mathcal{M} \\ \varepsilon &\longmapsto \mathbf{E} = \overrightarrow{pq}. \end{aligned}$$

Now, since $T_p\mathcal{M}$ is a vector space, we have the canonical identification $T_0(T_p\mathcal{M}) \simeq T_p\mathcal{M}$, from which $\varepsilon' = \varepsilon$. Without any loss of generality, we may write $\varepsilon = \varepsilon\mathbf{V}$, where ε is infinitesimal small and $\mathbf{V} \in T_p\mathcal{M}$. Then $q = \exp_p(\varepsilon\mathbf{V}) = P_{\mathbf{V}}(\varepsilon)$, where the second identity results from (B.17). We have thus

$$d\exp_p|_0(\varepsilon\mathbf{V}) = \mathbf{E} = \overrightarrow{pq} = \overrightarrow{P_{\mathbf{V}}(0)P_{\mathbf{V}}(\varepsilon)}.$$

According to the definition of $P_{\mathbf{V}}$, the infinitesimal vector $\overrightarrow{P_{\mathbf{V}}(0)P_{\mathbf{V}}(\varepsilon)}$ along the geodesic $\mathcal{L}_{\mathbf{V}}$ is $\varepsilon\mathbf{V}$, hence

$$d\exp_p|_0(\varepsilon\mathbf{V}) = \varepsilon\mathbf{V}.$$

Since the differential $d\exp_p|_0$ is a linear map, we get $d\exp_p|_0(\mathbf{V}) = \mathbf{V}$. The vector \mathbf{V} being arbitrary, we conclude that $d\exp_p|_0$ is nothing but the identity map of the vector space $T_p\mathcal{M}$:

$$d\exp_p|_0 = \text{id}_{T_p\mathcal{M}}.$$

In particular, $d\exp_p|_0$ is invertible. □

B.3.4 Normal coordinates

Given $p \in \mathcal{M}$, a **normal neighborhood** of p is a neighborhood \mathcal{U} of p that is the image of a starshaped neighborhood of 0 in $T_p\mathcal{M}$ under the local diffeomorphism \exp_p given by Property B.14. By **starshaped neighborhood** of 0, it is meant a neighborhood U of 0 such that $\mathbf{V} \in U$ implies $t\mathbf{V} \in U$ for any $t \in [0, 1]$. A **convex normal neighborhood** \mathcal{U} is an open subset of \mathcal{M} that is a normal neighborhood of each of its points. It follows that any two points p and q of a convex normal neighborhood \mathcal{U} can be connected by a geodesic lying entirely in \mathcal{U} .

On a normal neighborhood, one may define coordinates linked to geodesics as follows.

³Here the vector space $T_p\mathcal{M}$ is considered as a n -dimensional smooth manifold, and $T_0(T_p\mathcal{M})$ stands for its tangent space at 0 (the zero vector of $T_p\mathcal{M}$).

Let \mathcal{U} be a normal neighborhood of $p \in \mathcal{M}$ and $(\mathbf{E}_\alpha)_{0 \leq \alpha \leq n-1}$ be a basis of $T_p \mathcal{M}$. If (\mathbf{E}^α) stands for the basis of $T_p^* \mathcal{M}$ dual to (\mathbf{E}_α) , the map

$$\begin{aligned}\Phi : \quad \mathcal{U} &\longrightarrow \mathbb{R}^n \\ q &\longmapsto (\langle \mathbf{E}^0, \exp_p^{-1}(q) \rangle, \dots, \langle \mathbf{E}^{n-1}, \exp_p^{-1}(q) \rangle)\end{aligned}\tag{B.18}$$

is a coordinate chart on \mathcal{U} , which is called **geodesic normal coordinates** (often shorten to **normal coordinates**) **associated with the basis** (\mathbf{E}_α) .

In other words, normal coordinates (x^α) on \mathcal{U} are such that the tangent vector $x^\mu(q) \mathbf{E}_\mu \in T_p \mathcal{M}$ has precisely q as image by the exponential map:

$$\forall q \in \mathcal{U}, \quad \exp_p(x^\mu(q) \mathbf{E}_\mu) = q.\tag{B.19}$$

Remark 2: Some authors, e.g. [389], add the condition that the basis (\mathbf{E}_α) is orthonormal (with respect to the metric g) in the definition of normal coordinates. We follow here the more general definition of [317, 62, 265, 463]. The name **Riemann normal coordinates** is also commonly encountered in the literature, either for normal coordinates as defined here (e.g. [370, 498]) or for those with the basis orthonormality requirement (e.g. [417]).

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 :

Property B.15: geodesics in terms of normal coordinates

In a normal coordinate system (x^α) , the equation of the unique geodesic \mathcal{L}_V through p admitting $\mathbf{V} \in T_p \mathcal{M}$ as tangent vector at p is (as long as \mathcal{L}_V remains in the normal neighborhood \mathcal{U})

$$x^\alpha = X^\alpha(\lambda) = \lambda V^\alpha,\tag{B.20}$$

where the V^α 's are the components of \mathbf{V} with respect to the basis (\mathbf{E}_α) defining the normal coordinates: $\mathbf{V} = V^\mu \mathbf{E}_\mu$.

Proof. Let λ be the affine parameter of \mathcal{L}_V corresponding to the parametrization P_V . The coordinate equation of \mathcal{L}_V is then $x^\alpha = X^\alpha(\lambda)$ with [cf. Eq. (B.18)]

$$X^\alpha(\lambda) = x^\alpha(P_V(\lambda)) = \langle \mathbf{E}^\alpha, \exp_p^{-1}(P_V(\lambda)) \rangle.$$

Now, according to Eq. (B.17), $P_V(\lambda) = \exp_p(\lambda \mathbf{V})$. Hence

$$X^\alpha(\lambda) = \langle \mathbf{E}^\alpha, \exp_p^{-1} \circ \exp_p(\lambda \mathbf{V}) \rangle = \langle \mathbf{E}^\alpha, \lambda \mathbf{V} \rangle = \lambda \langle \mathbf{E}^\alpha, \mathbf{V} \rangle = \lambda V^\alpha.$$

□

Property B.16: metric components and Christoffel symbols in terms of normal coordinates

Let (x^α) be a normal coordinate system around $p \in \mathcal{M}$ associated with a basis (E_α) of $T_p\mathcal{M}$. Then

- the coordinate frame (∂_α) associated with (x^α) coincides with (E_α) at p :

$$\partial_\alpha|_p = E_\alpha; \quad (\text{B.21})$$

- the values at p of the components $(g_{\alpha\beta})$ of the metric tensor \mathbf{g} with respect to (x^α) are

$$g_{\alpha\beta}(p) = \mathbf{g}|_p(E_\alpha, E_\beta); \quad (\text{B.22})$$

- the Christoffel symbols of \mathbf{g} with respect to the coordinates (x^α) vanish at p :

$$\Gamma^\alpha_{\beta\gamma}(p) = 0. \quad (\text{B.23})$$

Proof. Let \mathcal{U} be the normal neighborhood covered by (x^α) and $\mathbf{V} \in \exp_p^{-1}(\mathcal{U}) \subset T_p\mathcal{M}$. The tangent vector field to the geodesic \mathcal{L}_V corresponding to the parametrization P_V is $\mathbf{v} = \dot{X}^\mu \partial_\mu$ with \dot{X}^μ obtained by deriving (B.20) with respect to λ : $\dot{X}^\mu = V^\mu$. Hence $\mathbf{v} = V^\mu \partial_\mu$. Now, from the very definition of P_V , $\mathbf{v}|_p = \mathbf{V} = V^\mu E_\mu$. We have therefore

$$V^\mu \partial_\mu|_p = V^\mu E_\mu.$$

This identity being fulfilled for any V^μ , Eq. (B.21) follows. Equation. (B.22) is an immediate consequence of Eq. (B.21), since $g_{\alpha\beta} = \mathbf{g}(\partial_\alpha, \partial_\beta)$. 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 $\mathbf{V} \mapsto \Gamma^\alpha_{\mu\nu}(p) V^\mu V^\nu$ is identically zero on $T_p\mathcal{M}$. 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 $(\mathbf{U}, \mathbf{V}) \mapsto \Gamma^\alpha_{\mu\nu}(p) U^\mu V^\nu$ is identically zero, which amounts to $\Gamma^\alpha_{\mu\nu}(p) = 0$, i.e. Eq. (B.23). \square

B.4 Geodesics and extremal lengths

B.4.1 Length of a curve

Geodesics in a pseudo-Riemannian manifold $(\mathcal{M}, \mathbf{g})$ 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 \mathbf{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 positive definite,

one cannot use the integral of the norm of infinitesimal displacements along the curve, i.e. $ds := \sqrt{\mathbf{g}(dx, dx)}$, since $\mathbf{g}(dx, dx)$ can be negative. Rather, it is quite natural to employ instead $ds := \sqrt{|\mathbf{g}(dx, dx)|}$. Using $dx = v d\lambda$ [Eq. (A.14)], we end up with the following definition:

The **length** of a curve \mathcal{L} connecting two points p and q in a pseudo-Riemannian manifold $(\mathcal{M}, \mathbf{g})$ is the real number

$$L_{(p,q)}(\mathcal{L}) := \int_{\lambda_p}^{\lambda_q} \sqrt{|\mathbf{g}(v, v)|} d\lambda, \quad (\text{B.24})$$

where λ is some parameter along \mathcal{L} , λ_p (resp. λ_q) being its value at p (resp. q), $v = dx/d\lambda$ 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 $L_{(p,q)}(\mathcal{L})$ is independent from the choice of the parametrization of \mathcal{L} , i.e. for a fixed pair of points (p, q) , it is a function of \mathcal{L} only.

When \mathcal{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)]

$$L_{(p,q)}(\mathcal{L}) := \int_{\lambda_p}^{\lambda_q} \sqrt{\left| g_{\mu\nu}(X^\rho(\lambda)) \dot{X}^\mu \dot{X}^\nu \right|} d\lambda, \quad (\text{B.25})$$

where $\dot{X}^\alpha := dX^\alpha/d\lambda$ and $g_{\mu\nu}(X^\rho(\lambda))$ stands for the components of the metric tensor \mathbf{g} with respect to the coordinates (x^α) at the point of coordinates $X^\rho(\lambda)$.

From the very definition of $L_{(p,q)}(\mathcal{L})$, it is obvious that

$$L_{(p,q)}(\mathcal{L}) \geq 0. \quad (\text{B.26})$$

Moreover, if it exists, any null curve from p to q achieves the absolute minimum of the length, without having to be a geodesic:

$$\mathcal{L} \text{ null} \implies L_{(p,q)}(\mathcal{L}) = 0. \quad (\text{B.27})$$

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 \mathbf{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 behavior of the length as a variational problem, i.e. to consider $L_{(p,q)}(\mathcal{L})$ as an “action” and to write the Euler-Lagrange equation for the “Lagrangian” defined as the integrand of (B.25):

$$\mathcal{L}(X^\alpha, \dot{X}^\alpha) = \sqrt{\left| g_{\mu\nu}(X^\rho) \dot{X}^\mu \dot{X}^\nu \right|}. \quad (\text{B.28})$$

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 $L_{(p,q)}(\mathcal{L})$), 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 \mathcal{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 \mathcal{L} (i.e. \mathcal{L} is spacelike) or always negative (i.e. \mathcal{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

$$\mathcal{L}(X^\alpha, \dot{X}^\alpha) = \sqrt{\epsilon g_{\mu\nu}(X^\rho)\dot{X}^\mu\dot{X}^\nu}. \quad (\text{B.29})$$

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:

$$\frac{d}{d\lambda} \left(\frac{\partial \mathcal{L}}{\partial \dot{X}^\alpha} \right) - \frac{\partial \mathcal{L}}{\partial X^\alpha} = 0. \quad (\text{B.30})$$

We have

$$\frac{\partial}{\partial X^\alpha} \left(g_{\mu\nu}(X^\rho)\dot{X}^\mu\dot{X}^\nu \right) = \frac{\partial g_{\mu\nu}}{\partial x^\alpha} \dot{X}^\mu\dot{X}^\nu, \quad (\text{B.31})$$

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),

$$\frac{\partial \mathcal{L}}{\partial X^\alpha} = \frac{\epsilon}{2\mathcal{L}} \frac{\partial g_{\mu\nu}}{\partial x^\alpha} \dot{X}^\mu\dot{X}^\nu. \quad (\text{B.32})$$

Besides,

$$\frac{\partial}{\partial \dot{X}^\alpha} \left(g_{\mu\nu}(X^\rho)\dot{X}^\mu\dot{X}^\nu \right) = g_{\alpha\nu}\dot{X}^\nu + g_{\mu\alpha}\dot{X}^\mu = 2g_{\alpha\mu}\dot{X}^\mu. \quad (\text{B.33})$$

Hence

$$\frac{\partial \mathcal{L}}{\partial \dot{X}^\alpha} = \frac{\epsilon}{\mathcal{L}} g_{\alpha\mu}\dot{X}^\mu,$$

from which,

$$\frac{d}{d\lambda} \left(\frac{\partial \mathcal{L}}{\partial \dot{X}^\alpha} \right) = -\frac{\epsilon}{\mathcal{L}^2} \frac{d\mathcal{L}}{d\lambda} g_{\alpha\mu}\dot{X}^\mu + \frac{\epsilon}{\mathcal{L}} \frac{\partial g_{\alpha\mu}}{\partial x^\nu} \dot{X}^\nu \dot{X}^\mu + \frac{\epsilon}{\mathcal{L}} g_{\alpha\mu}\ddot{X}^\mu. \quad (\text{B.34})$$

In view of (B.32) and (B.34), the Euler-Lagrange equation (B.30) becomes, after multiplication by \mathcal{L}/ϵ ,

$$-\frac{1}{\mathcal{L}} \frac{d\mathcal{L}}{d\lambda} g_{\alpha\mu}\dot{X}^\mu + \frac{\partial g_{\alpha\mu}}{\partial x^\nu} \dot{X}^\mu \dot{X}^\nu + g_{\alpha\mu}\ddot{X}^\mu - \frac{1}{2} \frac{\partial g_{\mu\nu}}{\partial x^\alpha} \dot{X}^\mu \dot{X}^\nu = 0.$$

Now, playing with the names of repeated indices and using the symmetry of $g_{\alpha\beta}$, we can rewrite the second term as

$$\frac{\partial g_{\alpha\mu}}{\partial x^\nu} \dot{X}^\mu \dot{X}^\nu = \frac{1}{2} \left(\frac{\partial g_{\alpha\nu}}{\partial x^\mu} \dot{X}^\nu \dot{X}^\mu + \frac{\partial g_{\mu\alpha}}{\partial x^\nu} \dot{X}^\mu \dot{X}^\nu \right) = \frac{1}{2} \left(\frac{\partial g_{\alpha\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\alpha}}{\partial x^\nu} \right) \dot{X}^\mu \dot{X}^\nu. \quad (\text{B.35})$$

Accordingly, we get

$$g_{\alpha\mu} \ddot{X}^\mu + \frac{1}{2} \left(\frac{\partial g_{\alpha\nu}}{\partial x^\mu} + \frac{\partial g_{\mu\alpha}}{\partial x^\nu} - \frac{\partial g_{\mu\nu}}{\partial x^\alpha} \right) \dot{X}^\mu \dot{X}^\nu = \kappa g_{\alpha\mu} \dot{X}^\mu, \quad (\text{B.36})$$

where we have introduced

$$\kappa := \frac{1}{\mathcal{L}} \frac{d\mathcal{L}}{d\lambda}. \quad (\text{B.37})$$

If we multiply Eq. (B.36) by the matrix $g^{\alpha\beta}$ (the components of the inverse metric) and use $g^{\alpha\beta} g_{\alpha\mu} = \delta_\mu^\beta$ as well as the expression (A.71) of the Christoffel symbols, we get exactly the pregeodesic equation (B.12). Hence we conclude:

Property B.17: timelike and spacelike geodesics as stationary points of the length

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 neighborhood), the stationary length corresponds actually to a *maximum*:

Property B.18: timelike geodesics as maximal length curves

Let (\mathcal{M}, g) be a Lorentzian manifold, $p \in \mathcal{M}$ and \mathcal{U} some normal neighborhood^a of p . For any point $q \in \mathcal{U}$ such that there exists a timelike curve in \mathcal{U} from p to q , the geodesic from p to q is the unique timelike curve of largest length in \mathcal{U} connecting p to q .

^aSee 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 [389]. We shall only illustrate the property on a specific example in flat spacetime (Example 1 below).

If one interprets timelike curves as worldlines and length as proper time (cf. Sec. 1.3.3), Property B.18 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 or her sister, who made a round trip.

Example 1 (Timelike geodesic in Minkowski spacetime): Let us suppose that (\mathcal{M}, 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 \mathcal{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

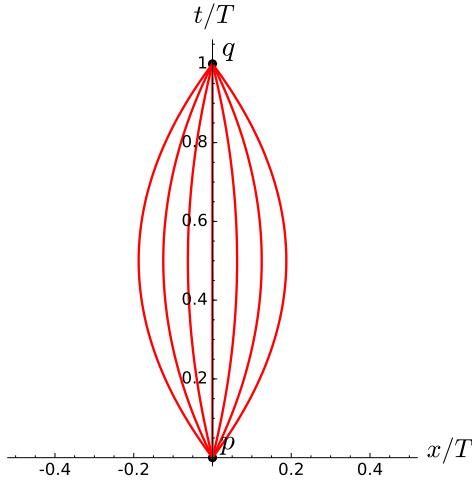


Figure B.1: Timelike curves \mathcal{L}_h connecting the point p of coordinates $(0, 0, 0, 0)$ to the point q of coordinates $(T, 0, 0, 0)$ in Minkowski spacetime. From the left to right, the depicted curves correspond to h spanning $[-3/4, 3/4]$, with the step $\delta h = 1/4$.

time along \mathcal{L} and $L_{(p,q)}(\mathcal{L}) = T$. Let us consider the one-parameter family of curves $(\mathcal{L}_h)_{h \in (-1,1)}$ 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 $\mathcal{L}_0 = \mathcal{L}$ and for $h \neq 0$, \mathcal{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 $x_{\max} = hT/4$. 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 $(g_{\alpha\beta}) = \text{diag}(-1, 1, 1, 1)$, 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 \mathcal{L}_h is a timelike curve as long as $-1 \leq h \leq 1$. Its length is

$$L_{(p,q)}(\mathcal{L}_h) = \int_0^T \sqrt{1 - h^2 \left(1 - 2\frac{\lambda}{T}\right)^2} d\lambda = \frac{T}{2h} \int_{-h}^h \sqrt{1 - u^2} du = \frac{T}{2h} \int_{-\arcsin h}^{\arcsin h} \cos^2 \theta d\theta.$$

Evaluating the integral leads to

$$L_{(p,q)}(\mathcal{L}_h) = \frac{T}{2} \left(\sqrt{1 - h^2} + \frac{\arcsin h}{h} \right).$$

Note that $\arcsin h/h$ is well-defined at $h = 0$, since $\lim_{h \rightarrow 0} \arcsin h/h = 1$. The graph of $L_{(p,q)}(\mathcal{L}_h)$ as a function h is plotted in Fig. B.2. We see clearly that $h = 0$, i.e. the geodesic \mathcal{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.

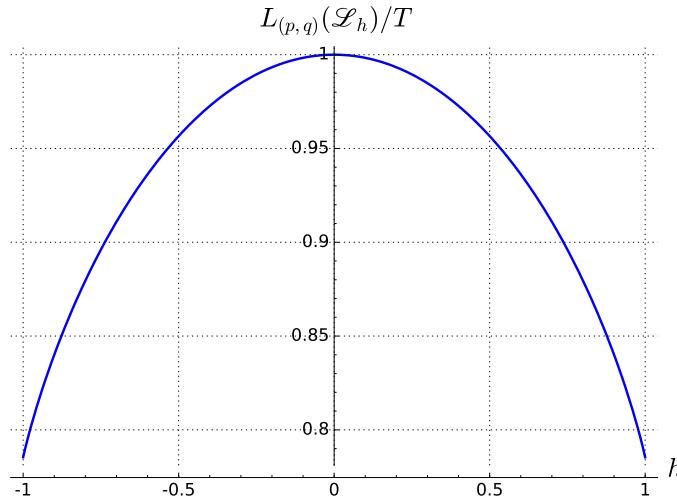


Figure B.2: Length of the timelike curve \mathcal{L}_h connecting the point p of coordinates $(0, 0, 0, 0)$ to the point q of coordinates $(T, 0, 0, 0)$ in Minkowski spacetime, as a function of the parameter h measuring the deviation from the timelike geodesic $\mathcal{L} = \mathcal{L}_0$.

Example 2 (Spacelike geodesic in Minkowski spacetime): As in Example 1, we consider Minkowski spacetime, but this time, \mathcal{L} is assumed to be a spacelike geodesic from p to q . Since \mathcal{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 $L_{(p,q)}(\mathcal{L})$ of the geodesic \mathcal{L} . Any spacelike curve \mathcal{L}' connecting p and q and lying in the hyperplane Σ defined by $t = 0$ obeys $L_{(p,q)}(\mathcal{L}') \geq L_{(p,q)}(\mathcal{L})$ since Σ , equipped with the metric induced by \mathbf{g} , is a 3-dimensional Euclidean space.

Let us consider some one-parameter family of curves $(\mathcal{L}_h)_{h \in (-1,1)}$ 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 $\mathcal{L}_0 = \mathcal{L}$ and for $h \neq 0$, the \mathcal{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

$$L_{(p,q)}(\mathcal{L}_h) = \frac{L}{2} \left(\sqrt{1 - h^2} + \frac{\arcsin h}{h} \right).$$

$L_{(p,q)}(\mathcal{L}_h)/L$ is exactly the same of function of h as $L_{(p,q)}(\mathcal{L}_h)/T$ in Example 1. In view of Fig. B.2, we therefore assert that $L_{(p,q)}(\mathcal{L}_h) \leq L_{(p,q)}(\mathcal{L})$.

We conclude that the spacelike geodesic \mathcal{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

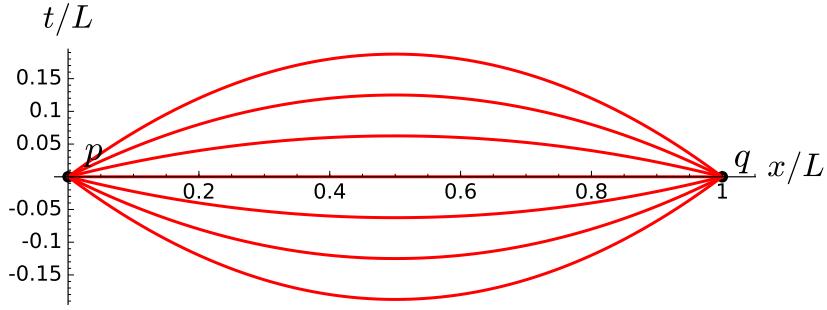


Figure B.3: Spacelike curves \mathcal{L}_h connecting the point p of coordinates $(0, 0, 0, 0)$ to the point q of coordinates $(0, L, 0, 0)$ in Minkowski spacetime. From the bottom to the top, the depicted curves correspond to h spanning $[-3/4, 3/4]$, with the step $\delta h = 1/4$.

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

$$S_{(p,q)}(P) := \frac{1}{2} \int_{\lambda_p}^{\lambda_q} g_{\mu\nu}(X^\rho) \dot{X}^\mu \dot{X}^\nu d\lambda, \quad (\text{B.38})$$

where P is a parametrization of the curve \mathcal{L} , λ the corresponding parameter and $x^\alpha = X^\alpha(\lambda)$ the coordinate expression of P .

The Lagrangian in (B.38) is

$$\mathcal{L}_2(X^\alpha, \dot{X}^\alpha) = \frac{1}{2} g_{\mu\nu}(X^\rho) \dot{X}^\mu \dot{X}^\nu. \quad (\text{B.39})$$

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 $S_{(p,q)}(\mathcal{L})$. For this reason, $S_{(p,q)}(P)$ is not expected to have any significant physical meaning, contrary to $L_{(p,q)}(\mathcal{L})$, 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

$$\frac{\partial \mathcal{L}_2}{\partial X^\alpha} = \frac{1}{2} \frac{\partial g_{\mu\nu}}{\partial x^\alpha} \dot{X}^\mu \dot{X}^\nu \quad \text{and} \quad \frac{\partial \mathcal{L}_2}{\partial \dot{X}^\alpha} = g_{\alpha\mu} \dot{X}^\mu,$$

so that

$$\frac{d}{d\lambda} \left(\frac{\partial \mathcal{L}_2}{\partial \dot{X}^\alpha} \right) = \frac{\partial g_{\alpha\mu}}{\partial x^\nu} \dot{X}^\nu \dot{X}^\mu + g_{\alpha\mu} \ddot{X}^\mu.$$

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:

Property B.19: geodesics as stationary points of the action (B.38)

In a pseudo-Riemannian manifold (\mathcal{M}, g) , a curve \mathcal{L} equipped with a parametrization P is a stationary point of the action (B.38) iff \mathcal{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 conjugate to X^α takes a simple form:

$$\Pi_\alpha := \frac{\partial \mathcal{L}}{\partial \dot{X}^\alpha} = g_{\alpha\mu} \dot{X}^\mu. \quad (\text{B.40})$$

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.

$$\mathcal{H}(X^\alpha, \Pi_\alpha) = \frac{1}{2} g^{\mu\nu}(X^\rho) \Pi_\mu \Pi_\nu. \quad (\text{B.41})$$

Such a Hamiltonian has been used by Carter [89] 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 isometry groups 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:

Property B.20: constancy of the scalar product of an affine tangent vector with a Killing vector

If the pseudo-Riemannian manifold (\mathcal{M}, g) admits a 1-dimensional isometry group of generator ξ , i.e. if ξ is a Killing vector field of (\mathcal{M}, g) , then along any geodesic \mathcal{L} , the g -scalar product of ξ by any tangent vector field $v = dx/d\lambda$ associated with an affine parameter λ of \mathcal{L} is constant:

$$g(\xi, v) = \text{const.} \quad (\text{B.42})$$

Proof. The variation of $\mathbf{g}(\xi, v)$ along \mathcal{L} is, according to Eq. (A.8),

$$\begin{aligned}\frac{d}{d\lambda} \mathbf{g}(\xi, v) &= \mathbf{v}(\mathbf{g}(\xi, v)) = \nabla_{\mathbf{v}}(\mathbf{g}(\xi, v)) \\ &= v^\sigma \nabla_\sigma(g_{\mu\nu} \xi^\mu v^\nu) = v^\sigma \nabla_\sigma(\xi_\nu v^\nu) = \nabla_\sigma \xi_\nu v^\sigma v^\nu + \xi_\nu v^\sigma \nabla_\sigma v^\nu \\ &= \frac{1}{2} (\underbrace{\nabla_\sigma \xi_\nu + \nabla_\nu \xi_\sigma}_0) v^\sigma v^\nu + \xi_\nu \underbrace{v^\sigma \nabla_\sigma v^\nu}_0 = 0,\end{aligned}\tag{B.43}$$

where the first zero holds because ξ obeys the Killing equation (3.19) and the second one holds thanks to Eq. (B.1), which expresses that \mathcal{L} is a geodesic and v the tangent vector associated with some affine parameter. \square

Remark 1: If the tangent vector v is associated with a generic (not necessarily affine) parameter of \mathcal{L} , the second zero in Eq. (B.43) must be replaced by κv^ν , where κ is the non-affinity coefficient of v [cf. Eq. (B.7)]. Accordingly $\mathbf{g}(\xi, v)$ is no longer constant along \mathcal{L} but rather evolves according to

$$\frac{d}{d\lambda} \mathbf{g}(\xi, v) = \kappa \mathbf{g}(\xi, v).\tag{B.44}$$

Note that κ a priori varies along \mathcal{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 (3.19) 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** [94]) of rank $p \geq 1$ in the pseudo-Riemannian manifold (\mathcal{M}, g) is a tensor field \mathbf{K} of type $(0, p)$ that is fully symmetric and whose covariant derivative obeys

$$\nabla_{(\alpha_1} K_{\alpha_2 \dots \alpha_{p+1})} = 0.\tag{B.45}$$

Example 3: A trivial example of Killing tensor is the metric tensor g itself. If (\mathcal{M}, g) admits a Killing vector ξ , another example is $\mathbf{K} = \underline{\xi}$ (the 1-form associated to ξ by metric duality), since for $p = 1$, Eq. (B.45) reduces to the Killing equation (3.19). An example for $p = 2$ is then $\mathbf{K} = \underline{\xi} \otimes \underline{\xi}$ (by the Leibniz rule + the Killing equation). Similarly, $\mathbf{K} = \underline{\xi} \otimes \underline{\xi} \otimes \underline{\xi}$ 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 g nor any Killing vector as in Example 3, one often says that this spacetime has a **hidden symmetry** (see e.g. the review article [200] for an extended discussion). This is because, as Killing vectors, Killing tensors give birth to conserved quantities along geodesics:

Property B.21: conserved quantity induced by a Killing tensor along a geodesic

If \mathbf{K} is a Killing tensor of rank p on the pseudo-Riemannian manifold $(\mathcal{M}, \mathbf{g})$, then along any geodesic \mathcal{L} , the scalar $\mathbf{K}(\mathbf{v}, \dots, \mathbf{v})$, where $\mathbf{v} = dx/d\lambda$ is any tangent vector field associated with an affine parameter λ of \mathcal{L} , is constant:

$$\mathbf{K}(\mathbf{v}, \dots, \mathbf{v}) = \text{const.} \quad (\text{B.46})$$

Proof. The variation of $\mathbf{K}(\mathbf{v}, \dots, \mathbf{v})$ along \mathcal{L} is given by

$$\begin{aligned} \frac{d}{d\lambda} (\mathbf{K}(\mathbf{v}, \dots, \mathbf{v})) &= \nabla_{\mathbf{v}} (\mathbf{K}(\mathbf{v}, \dots, \mathbf{v})) = v^\mu \nabla_\mu (K_{\nu_1 \dots \nu_p} v^{\nu_1} \dots v^{\nu_p}) \\ &= \underbrace{\nabla_\mu K_{\nu_1 \dots \nu_p} v^\mu v^{\nu_1} \dots v^{\nu_p}}_0 + K_{\nu_1 \dots \nu_p} \underbrace{v^\mu \nabla_\mu v^{\nu_1} \dots v^{\nu_p}}_0 + \dots + K_{\nu_1 \dots \nu_p} v^{\nu_1} \dots \underbrace{v^\mu \nabla_\mu v^{\nu_p}}_0 \\ &= 0, \end{aligned}$$

where the first zero results from (B.45), while the zeros in the line below arise from the geodesic equation (B.1) obeyed by \mathbf{v} . \square

Example 4: Since we have already noticed that the metric tensor \mathbf{g} is a Killing tensor (Example 3), Property B.2 (constancy of $\mathbf{g}(\mathbf{v}, \mathbf{v})$) appears as a special case of Property B.21. For the Walker-Penrose Killing tensor \mathbf{K} of Kerr spacetime, the conserved quantity $\mathbf{K}(\mathbf{v}, \mathbf{v})$ leads to the Carter constant (cf. Chap. 11).

B.6 Geodesics and curvature

Let $(\mathcal{L}_s)_{s \in \mathbb{R}}$ be a smooth 1-parameter family of non-intersecting geodesics in a pseudo-Riemannian manifold $(\mathcal{M}, \mathbf{g})$. Let λ be an affine parameter along each geodesic \mathcal{L}_s , such that λ varies smoothly from one geodesic to the next one, i.e. such that the map $\Phi : I \times \mathbb{R} \rightarrow \mathcal{M}$, $(\lambda, s) \mapsto P_s(\lambda)$ is smooth, $P_s : I \rightarrow \mathcal{M}$ being the parametrization of \mathcal{L}_s associated to λ (cf. Sec. B.2.1). Note that $\mathcal{S} := \bigcup_{s \in \mathbb{R}} \mathcal{L}_s$ is a 2-dimensional surface of \mathcal{M} , which is parametrized by (λ, s) (cf. Fig. B.4); \mathcal{S} is actually the image of the map Φ defined above. On \mathcal{S} , let us denote by \mathbf{v} and \mathbf{z} the vector fields of the natural basis associated to the coordinates (λ, s) :

$$\mathbf{v} := \partial_\lambda \quad \text{and} \quad \mathbf{z} := \partial_s. \quad (\text{B.47})$$

By construction, \mathbf{v} coincides with the tangent vector field $dx/d\lambda$ associated to the affine parameter λ along each geodesic \mathcal{L}_s [Eq. (B.2)]. The vector field \mathbf{z} is called the **separation vector at fixed λ** of the family $(\mathcal{L}_s)_{s \in \mathbb{R}}$ because at a fixed value of λ , two infinitely close geodesics \mathcal{L}_s and \mathcal{L}_{s+ds} are connected by the vector $dx = ds \mathbf{z}$ (cf. Fig. B.4). Note that the separation vector depends upon the choice of the affine parameter λ .

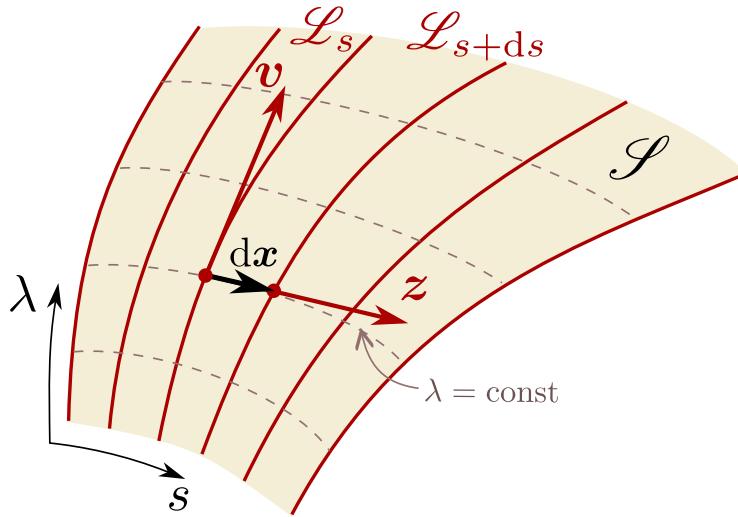


Figure B.4: 1-parameter family of geodesics $(\mathcal{L}_s)_{s \in \mathbb{R}}$ affinely parametrized by λ . The 2-surface \mathcal{S} formed by the geodesic family is spanned by the coordinates (λ, s) . The infinitesimal vector connecting a point of \mathcal{L}_s to the point of \mathcal{L}_{s+ds} having the same value of λ is $dx = ds z$, where z is the separation vector at fixed λ .

When the geodesics \mathcal{L}_s are timelike or null, v is the evolution vector along \mathcal{L}_s associated to the affine parameter λ . One may then consider the covariant derivative $\nabla_v z$ as the *relative velocity* of neighboring geodesics and the second order derivative $\nabla_v \nabla_v z$ as the *relative acceleration* of neighboring geodesics. Even if \mathcal{L}_s is spacelike, we keep the terms *relative velocity* and *relative acceleration*.

Property B.22: geodesic deviation equation

The relative acceleration of nearby members of a 1-parameter family $(\mathcal{L}_s)_{s \in \mathbb{R}}$ of geodesics in a pseudo-Riemannian manifold (\mathcal{M}, g) is given by

$$\nabla_v \nabla_v z = \text{Riem}(., v, v, z), \quad (\text{B.48})$$

where v is the evolution vector along \mathcal{L}_s associated to some affine parameter λ , z is the separation vector at fixed λ , ∇ is the Levi-Civita connection of g and Riem is the associated Riemann curvature tensor (cf. Sec. A.5). Equation (B.48) is called the **geodesic deviation equation** or the **Jacobi equation**. In index notation, Eq. (B.48) becomes

$$v^\mu \nabla_\mu (v^\nu \nabla_\nu z^\alpha) = R^\alpha_{\rho\mu\nu} v^\rho v^\mu z^\nu. \quad (\text{B.49})$$

Proof. Because they are vector fields associated to the coordinate system (λ, s) of \mathcal{S} , it is obvious that the vector fields v and z commute: $[v, z] = 0$. This can also be seen from the identity $[v, z] = \mathcal{L}_v z$ [Eq. (A.79)] and the very definition of the Lie derivative $\mathcal{L}_v z$ (cf. Sec. A.4.2), which in the present case yields $\mathcal{L}_v z = 0$. Hence we have $v^\nu \nabla_\nu z^\alpha = z^\nu \nabla_\nu v^\alpha$.

Taking another derivative along v , we get successively

$$\begin{aligned} v^\mu \nabla_\mu (v^\nu \nabla_\nu z^\alpha) &= v^\mu \nabla_\mu (z^\nu \nabla_\nu v^\alpha) = v^\mu \nabla_\mu z^\nu \nabla_\nu v^\alpha + z^\nu v^\mu \nabla_\mu \nabla_\nu v^\alpha \\ &= v^\mu \nabla_\mu z^\nu \nabla_\nu v^\alpha + z^\nu v^\mu (\nabla_\nu \nabla_\mu v^\alpha + R^\alpha_{\rho\mu\nu} v^\rho) \\ &= z^\mu \nabla_\mu v^\nu \nabla_\nu v^\alpha + z^\mu v^\nu \nabla_\mu \nabla_\nu v^\alpha + R^\alpha_{\rho\mu\nu} v^\rho v^\mu z^\nu \\ &= z^\mu \nabla_\mu \underbrace{(v^\nu \nabla_\nu v^\alpha)}_0 + R^\alpha_{\rho\mu\nu} v^\rho v^\mu z^\nu = R^\alpha_{\rho\mu\nu} v^\rho v^\mu z^\nu, \end{aligned}$$

where the Ricci identity (A.98) has been used to get the second line and again the commutation property $v^\mu \nabla_\mu z^\nu = z^\mu \nabla_\mu v^\nu$ to get the third line. Finally, $v^\nu \nabla_\nu v^\alpha = 0$ in the last line follows from the geodesic character of v [Eq. (B.1)], given that it is the tangent vector corresponding to an affine parametrization of \mathcal{L}_s . \square

The geodesic deviation equation (B.48) is a second order linear differential equation for the vector field z . Thanks to it, the curvature of the pseudo-Riemannian manifold (\mathcal{M}, g) can be interpreted as the obstruction for neighboring geodesics that are initially parallel (zero relative velocity $\nabla_v z$ at $\lambda = 0$) to stay parallel for $\lambda \neq 0$. Actually, Eq. (B.48) can even be used to *define* the Riemann curvature tensor Riem , instead of Eq. (A.97) (see e.g. Chap. 11 of MTW [370] for details).

Appendix C

Kerr-Schild metrics

Contents

C.1 Generic Kerr-Schild spacetimes	761
C.2 Case of Kerr spacetime	763

C.1 Generic Kerr-Schild spacetimes

C.1.1 Definition

A spacetime $(\mathcal{M}, \mathbf{g})$ is said to have a **Kerr-Schild metric** iff the metric tensor \mathbf{g} can be written

$$\boxed{\mathbf{g} = \mathbf{f} + 2H\mathbf{k} \otimes \mathbf{k}}, \quad (\text{C.1})$$

or equivalently (in index notation):

$$\boxed{g_{\alpha\beta} = f_{\alpha\beta} + 2Hk_\alpha k_\beta}, \quad (\text{C.2})$$

where \mathbf{f} is a *flat* Lorentzian metric on \mathcal{M} (Minkowski metric), H is a scalar field on \mathcal{M} and \mathbf{k} is a 1-form on \mathcal{M} such that the vector associated to it via \mathbf{f} is a null vector of the metric \mathbf{f} :

$$f^{\mu\nu} k_\mu k_\nu = 0, \quad (\text{C.3})$$

where $f^{\mu\nu}$ stands for the components of the inverse of the metric \mathbf{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:

$$g^{\alpha\beta} = f^{\alpha\beta} - 2Hk^\alpha k^\beta, \quad (\text{C.4})$$

where

$$k^\alpha := f^{\alpha\mu} k_\mu. \quad (\text{C.5})$$

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). \square

Given (C.4), it is easy to see that the vector field \mathbf{k} associated to the 1-form \underline{k} by \mathbf{g} -duality (cf. Sec. A.3.3) is the same as the vector field obtained by \mathbf{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 \mathbf{k} simply as k^α without having to specify whether the index has been raised with \mathbf{g} or \mathbf{f} :

$$k^\alpha = f^{\alpha\mu} k_\mu = g^{\alpha\mu} k_\mu. \quad (\text{C.6})$$

It follows immediately that \mathbf{k} is a null vector field for both metrics:

$$\boxed{\mathbf{g}(\mathbf{k}, \mathbf{k}) = \mathbf{f}(\mathbf{k}, \mathbf{k}) = 0}. \quad (\text{C.7})$$

If $(\mathcal{M}, \mathbf{g})$ is a spacetime of Kerr-Schild type, then **Kerr-Schild coordinates** are coordinates $(x^\alpha) = (t, x, y, z)$ that are *Minkowskian* for \mathbf{f} , i.e. coordinates in which the flat metric \mathbf{f} takes the form

$$\mathbf{f} = -dt^2 + dx^2 + dy^2 + dz^2. \quad (\text{C.8})$$

C.1.2 Basic property

Property C.1: null vector of Kerr-Schild metric as a geodesic vector

Let \mathbf{g} be a Kerr-Schild metric. If \mathbf{g} obeys the vacuum Einstein equation (1.44), i.e. if the Ricci tensor of \mathbf{g} vanishes identically, then the scalar field H appearing in Eq. (C.1) can be chosen so that \mathbf{k} is a geodesic vector field^a:

$$\boxed{k^\mu \nabla_\mu k^\alpha = 0}, \quad (\text{C.9})$$

where ∇ stands for the covariant derivative associated with \mathbf{g} .

^aSee Sec. B.2.2 for the definition of a geodesic vector field.

The proof of the above proposition can be found in Ref. [313].

C.2 Case of Kerr spacetime

C.2.1 Kerr-Schild form

Let consider the Kerr spacetime $(\mathcal{M}, \mathbf{g})$, where \mathcal{M} is the manifold (10.29): $\mathcal{M} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathcal{R}$ and \mathbf{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 \mathbf{g} is a Kerr-Schild metric, with the associated null vector field \mathbf{k} being nothing but the vector field generating the ingoing principal null geodesics $\mathcal{L}_{(v, \theta, \tilde{\varphi})}^{\text{in}}$ discussed in Sec. 10.4. Its expression in terms of the Kerr coordinates is given by Eq. (10.43):

$$\mathbf{k} = \partial_{\tilde{t}} - \partial_{\tilde{r}}. \quad (\text{C.10})$$

In other words, the components of \mathbf{k} with respect to the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ are $k^\alpha = (1, -1, 0, 0)$. The 1-form $\underline{\mathbf{k}}$ associated to \mathbf{k} by \mathbf{g} -duality is given by Eq. (10.44):

$$\underline{\mathbf{k}} = -d\tilde{t} - dr + a \sin^2 \theta d\tilde{\varphi}. \quad (\text{C.11})$$

Equivalently, $k_\alpha = (-1, -1, 0, a \sin^2 \theta)$. Let us then introduce the symmetric bilinear form

$$\mathbf{f} := \mathbf{g} - 2H\underline{\mathbf{k}} \otimes \underline{\mathbf{k}}, \quad (\text{C.12})$$

where H is the following scalar field on \mathcal{M} :

$$H := \frac{mr}{\rho^2}, \quad (\text{C.13})$$

with $\rho^2 := r^2 + a^2 \cos^2 \theta$ [Eq. (10.9)]. The expression of \mathbf{f} in terms of Kerr coordinates is deduced from that of \mathbf{g} [Eq. (10.36)] and that of $\underline{\mathbf{k}}$ [Eq. (C.11)]:

$$\mathbf{f} = -d\tilde{t}^2 + dr^2 - 2a \sin^2 \theta dr d\tilde{\varphi} + \rho^2 d\theta^2 + (r^2 + a^2) \sin^2 \theta d\tilde{\varphi}^2. \quad (\text{C.14})$$

It is easy to check that $f^{\alpha\beta} := g^{\alpha\beta} + 2Hk^\alpha k^\beta$ defines an inverse of \mathbf{f} : $f^{\alpha\mu} f_{\mu\beta} = \delta^\alpha_\beta$ (computation similar to that in Sec. C.1.1). Hence the symmetric bilinear form \mathbf{f} is non-degenerate; this implies that \mathbf{f} is a *metric tensor* on \mathcal{M} (cf. Sec. A.3.1). Given the components (C.11), it is immediate to check that \mathbf{k} is a null vector for \mathbf{f} as well: $\mathbf{f}(\mathbf{k}, \mathbf{k}) = 0$. Moreover, \mathbf{f} is a *flat* metric, since a direct computation of its Riemann tensor (cf. the notebook D.5.4) reveals that

$$\text{Riem}(\mathbf{f}) = 0. \quad (\text{C.15})$$

In view of the definition given in Sec. C.1.1, we conclude:

Property C.2: Kerr metric as a Kerr-Schild metric

The Kerr metric \mathbf{g} is a Kerr-Schild metric, i.e. it can be written in the form (C.1) with the flat metric \mathbf{f} given by Eq. (C.14), the scalar field H given by Eq. (C.13) and the null vector \mathbf{k} given by Eq. (C.10), \mathbf{k} being the tangent vector field to the ingoing principal null geodesics.

In Sec. 10.4, we have already noticed that \mathbf{k} is a geodesic vector: $\nabla_{\mathbf{k}} \mathbf{k} = 0$ [Eq. (10.45)], in agreement with (C.9).

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 f given by Eq. (C.14) is a flat Lorentzian metric, except at the limit $a = 0$. Let us introduce coordinates in which f takes a manifest 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 $\mathcal{M} = \mathbb{R}^2 \times \mathbb{S}^2 \setminus \mathcal{R}$ as defined by Eq. (10.29). One has to split it in two parts:

$$\mathcal{M} := \mathcal{M}_+ \cup \mathcal{M}_-, \quad (\text{C.16a})$$

$$\mathcal{M}_+ := \mathbb{R} \times [0, +\infty) \times \mathbb{S}^2 \setminus \mathcal{R} \quad (\text{C.16b})$$

$$\mathcal{M}_- := \mathbb{R} \times (-\infty, 0] \times \mathbb{S}^2 \setminus \mathcal{R}. \quad (\text{C.16c})$$

In other words, \mathcal{M}_+ is the part $r \geq 0$ of \mathcal{M} , while \mathcal{M}_- is the part $r \leq 0$. Note that \mathcal{M}_+ and \mathcal{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, \mathcal{M}_+ contains \mathcal{M}_{I} , \mathcal{M}_{II} and a part of \mathcal{M}_{III} , while \mathcal{M}_- is entirely included in \mathcal{M}_{III} . The **Kerr-Schild coordinates** (\tilde{t}, x, y, z) on \mathcal{M}_+ are defined from the Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ by the following formulas:

$$\tilde{t} = \tilde{t} \quad (\text{C.17a})$$

$$x = (r \cos \tilde{\varphi} - a \sin \tilde{\varphi}) \sin \theta \quad (\text{C.17b})$$

$$y = (r \sin \tilde{\varphi} + a \cos \tilde{\varphi}) \sin \theta \quad (\text{C.17c})$$

$$z = r \cos \theta. \quad (\text{C.17d})$$

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 \mathcal{M}_+ and \mathcal{M}_- .

Remark 3: For $a = 0$, Eqs. (C.17b)-(C.17d) reduce to the standard relations between Cartesian and spherical coordinates in Euclidean space.

Remark 4: Equations (C.17b)-(C.17c) can be combined into a single relation:

$$x + iy = (r + ia)e^{i\tilde{\varphi}} \sin \theta. \quad (\text{C.18})$$

From Eqs. (C.17b)-(C.17c), we get

$$x^2 + y^2 = (r^2 + a^2) \sin^2 \theta. \quad (\text{C.19})$$

Combining with Eq. (C.17d) yields:

$$\boxed{\frac{x^2 + y^2}{r^2 + a^2} + \frac{z^2}{r^2} = 1}. \quad (\text{C.20})$$

This is a quadratic equation in r^2 . Solving it results in

$$r = \sqrt{\frac{1}{2} \left(x^2 + y^2 + z^2 - a^2 + \sqrt{(x^2 + y^2 + z^2 - a^2)^2 + 4a^2 z^2} \right)}. \quad (\text{C.21})$$

The components of \mathbf{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.17) (cf. the notebook D.5.4):

$$\mathbf{f} = -\mathbf{d}\tilde{t}^2 + \mathbf{d}x^2 + \mathbf{d}y^2 + \mathbf{d}z^2. \quad (\text{C.22})$$

This proves that (\tilde{t}, x, y, z) are Kerr-Schild coordinates, as announced.

The expression of the vector \mathbf{k} in terms of the Kerr-Schild coordinates is obtained similarly:

$$\mathbf{k} = \partial_{\tilde{t}} - \frac{rx + ay}{r^2 + a^2} \partial_x - \frac{ry - ax}{r^2 + a^2} \partial_y - \frac{z}{r} \partial_z. \quad (\text{C.23})$$

In this formula, r is to be considered as the function of (x, y, z) given by Eq. (C.21). For the associated 1-form, we get

$$\underline{\mathbf{k}} = -\mathbf{d}\tilde{t} - \frac{rx + ay}{r^2 + a^2} \mathbf{d}x - \frac{ry - ax}{r^2 + a^2} \mathbf{d}y - \frac{z}{r} \mathbf{d}z. \quad (\text{C.24})$$

The scalar factor H can be re-expressed from Eq. (C.13) in terms of z and r :

$$H = \frac{mr^3}{r^4 + a^2 z^2}. \quad (\text{C.25})$$

Remark 5: If $a = 0$ (Schwarzschild limit), we get

$$r = \sqrt{x^2 + y^2 + z^2}, \quad H = \frac{m}{r} \quad \text{and} \quad \underline{\mathbf{k}} = -\mathbf{d}\tilde{t} - \frac{x}{r} \mathbf{d}x - \frac{y}{r} \mathbf{d}y - \frac{z}{r} \mathbf{d}z. \quad (\text{C.26})$$

Remark 6: For $a \neq 0$, the relations (C.26) 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 $(g_{\alpha\beta})$ of the Kerr metric in Kerr-Schild coordinates can be read off by expanding the following expression of \mathbf{g} :

$$\boxed{\mathbf{g} = -\mathbf{d}\tilde{t}^2 + \mathbf{d}x^2 + \mathbf{d}y^2 + \mathbf{d}z^2 + \frac{2mr^3}{r^4 + a^2 z^2} \left(\mathbf{d}\tilde{t} + \frac{rx + ay}{r^2 + a^2} \mathbf{d}x + \frac{ry - ax}{r^2 + a^2} \mathbf{d}y + \frac{z}{r} \mathbf{d}z \right)^2}, \quad (\text{C.27})$$

which is obtained by combining Eqs (C.2), (C.22), (C.25) and (C.24).

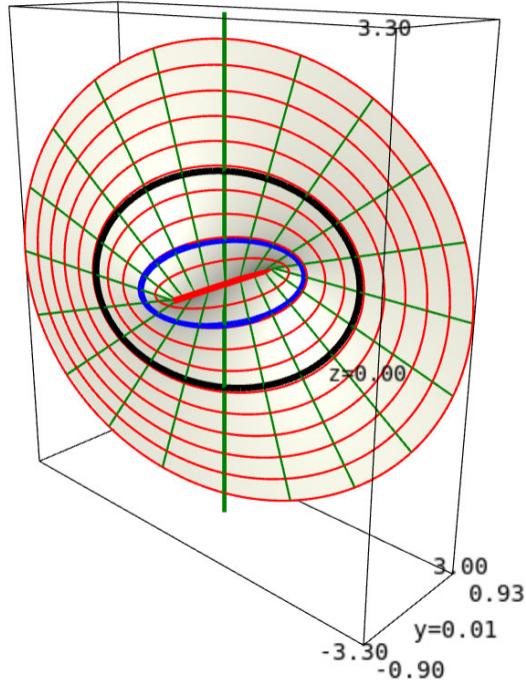


Figure C.1: Surface $\tilde{t} = \text{const}$, $\tilde{\varphi} = 0$ or π and $r \geq 0$ of the $a = 0.9\text{ m}$ Kerr spacetime depicted in terms of the Kerr-Schild coordinates (x, y, z) . The drawing is limited to $r \leq 3\text{ m}$. The vertical thick green line is the axis of rotation. On the 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.5.4]

Remark 7: It is clear on (C.27) that all metric components in Kerr-Schild coordinates are regular both at 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 $\eta = \partial_{\tilde{\varphi}}$ of Kerr spacetime [Eq. (10.39)] has the following expression in terms of Kerr-Schild coordinates:

$$\eta = -y \partial_x + x \partial_y. \quad (\text{C.28})$$

This is easily established from the chain rule and the partial derivative with respect to $\tilde{\varphi}$ of expressions (C.17). We notice on Eq. (C.28) that η is also a Killing vector for the flat metric f , namely the Killing vector generating spatial rotations about the z -axis.

The identity (C.20) shows that, in the Euclidean space spanned by the (x, y, z) coordinates, the surfaces of constant $r \neq 0$ are confocal¹ 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 (x, y, z) coordinates. Note that

¹In 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.

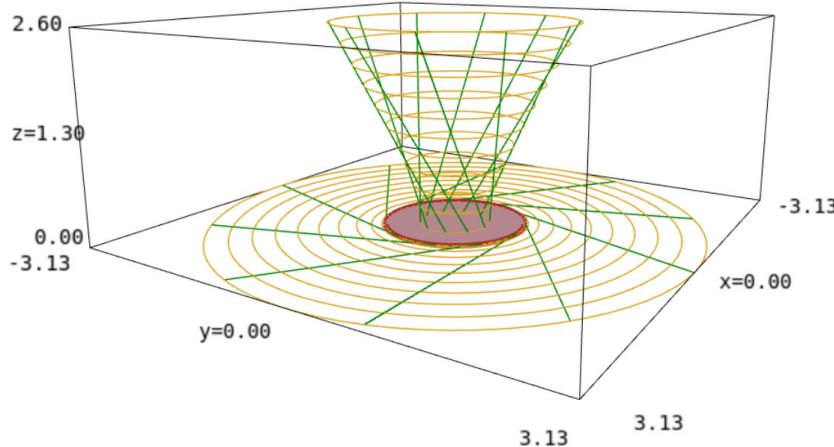


Figure C.2: Surfaces $(\tilde{t}, \theta) = \text{const}$ of the $a = 0.9 \text{ m}$ Kerr spacetime depicted in terms of the Kerr-Schild coordinates (x, y, z) . The disk-like surface in the plane $z = 0$ is for $\theta = \pi/2$, while the cone-like surface is for $\theta = \pi/6$. The pale brown lines 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.5.4]

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 $\mathcal{L}_{(v, \theta, \tilde{\varphi})}^{\text{in}}$ (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 η [cf. Eq. (C.28)].

C.2.3 The double-disk $r = 0$

For $r = 0$, the system (C.17) reduces to

$$\tilde{t} = \tilde{t} \quad (\text{C.29a})$$

$$x = -a \sin \theta \sin \tilde{\varphi} \quad (\text{C.29b})$$

$$y = a \sin \theta \cos \tilde{\varphi} \quad (\text{C.29c})$$

$$z = 0 \quad (\text{C.29d})$$

For $a \neq 0$ and a fixed value of \tilde{t} , the subset $r = 0$ of the Kerr spacetime \mathcal{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 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

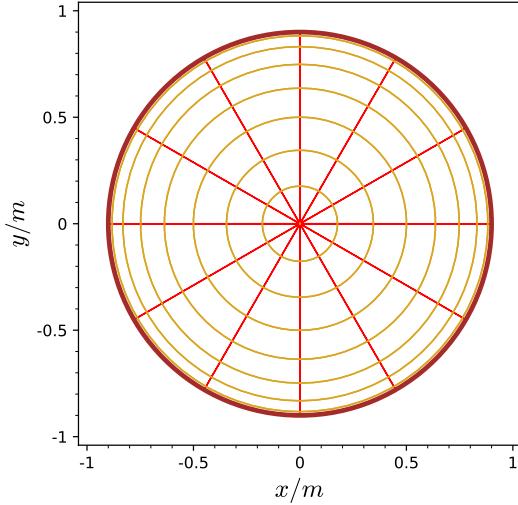


Figure C.3: Disk (actually double disk) $r = 0$ of the $a = 0.9\text{ m}$ Kerr spacetime depicted in terms of the Kerr-Schild coordinates (x, y) . The pale brown circles are the curves $\theta = \text{const}$, while the red segments are the curves $\tilde{\varphi} = \text{const}$. The disk boundary at $\sqrt{x^2 + y^2} = a$ is the curvature singularity of Kerr spacetime. [Figure produced with the notebook [D.5.4](#)]

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 \mathbf{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.29) maps both p and q to the same value $(\tilde{t}, -a \sin \theta \sin \tilde{\varphi}, a \sin \theta \cos \tilde{\varphi}, 0)$ 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.29), 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 \mathcal{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 \mathcal{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.17). A difference is when expressing the square root for r as in Eq. (C.21), one has to take the minus sign, so that

$$r := -\sqrt{\frac{1}{2} \left(x'^2 + y'^2 + z'^2 - a^2 + \sqrt{(x'^2 + y'^2 + z'^2 - a^2)^2 + 4a^2 z'^2} \right)}. \quad (\text{C.30})$$

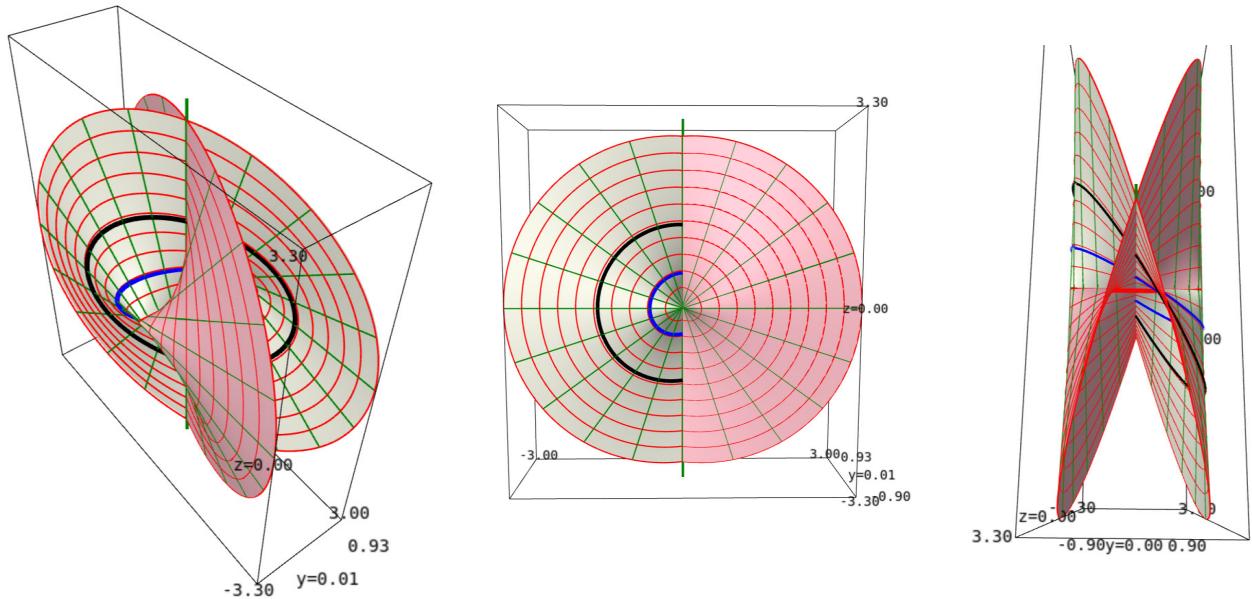


Figure C.4: Three views of the immersion of the full $\tilde{t} = \text{const}$ and $\tilde{\varphi} = 0$ or π surface of the $a = 0.9\text{ m}$ Kerr spacetime in the Euclidean space \mathbb{R}^3 , using the Kerr-Schild coordinates (x, y, z) for the $r \geq 0$ part (drawn in grey) and the Kerr-Schild 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.5.4; see this notebook for an interactive 3D view]

The two Kerr-Schild coordinate domains \mathcal{M}_+ and \mathcal{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 embedding² 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 \mathcal{M}_+ and \mathcal{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 $\mathcal{S}_{0,t}$.

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.17) with Eqs. (10.35).

Historical note : Kerr-Schild coordinates on Kerr spacetime have been introduced by Roy Kerr in the famous 1963 paper [311] announcing the discovery of the Kerr metric. They have been discussed further by Robert Boyer and Richard Lindquist in 1967 [71], Brandon Carter in 1968 [89] and Stephen Hawking

²See Sec. A.2.7 for the definitions of *immersion* and *embedding*.

and George Ellis in 1973 [265]. Generic Kerr-Schild metrics have been introduced and studied by Roy Kerr and Alfred Schild in 1965 [313].

Appendix D

SageMath computations

Contents

D.1	Introduction	771
D.2	Minkowski spacetime	772
D.3	Anti-de Sitter spacetime	772
D.4	Schwarzschild spacetime	772
D.5	Kerr spacetime	776
D.6	Evolution and thermodynamics	779

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 [461] and since then its development has been sustained by more than a hundred researchers (mostly mathematicians). Very good introductory textbooks about SageMath are [531, 305, 33].

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://relativite.obspm.fr/blackholes/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 Anti-de Sitter spacetime

D.3.1 Poincaré horizon in AdS_4 as a degenerate Killing horizon

This notebook accompanies Example 18 of Chap. 3.

https://nbviewer.jupyter.org/github/sagemanifolds/SageManifolds/blob/master/Notebooks/SM_anti_de_Sitter_Poincare_hor.ipynb

D.3.2 Conformal completion of AdS_4

This notebook introduces AdS_4 in various coordinate systems and provides the figure of Example 1 in Chap. 4.

https://nbviewer.jupyter.org/github/sagemanifolds/SageManifolds/blob/master/Notebooks/SM_anti_de_Sitter.ipynb

D.4 Schwarzschild spacetime

D.4.1 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.4.2 Solving the 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.4.3 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. (6.44).

https://nbviewer.jupyter.org/github/sagemanifolds/SageManifolds/blob/master/Notebooks/SM_basic_Schwarzschild.ipynb

D.4.4 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.4.5 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.4.6 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.4.7 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.4.8 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.4.9 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.4.10 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.4.11 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.4.12 Null geodesics in Schwarzschild spacetime with $b < b_c$

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.4.13 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.4.14 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.4.15 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.4.16 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.4.17 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.4.18 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.4.19 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.5 Kerr spacetime

D.5.1 Kerr metric as a solution of the 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.5.2 Kerr spacetime in Kerr coordinates

These two notebooks accompany Chap. 10: the Kerr metric is expressed in advanced Kerr coordinates $(v, r, \theta, \tilde{\varphi})$ (notebook 1) and in Kerr coordinates $(\tilde{t}, r, \theta, \tilde{\varphi})$ (notebook 2), the vacuum Einstein equation is checked, the outgoing and ingoing principal null geodesics are considered and the black hole surface gravity is computed.

1. https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_in_advanced_Kerr_coord.ipynb
2. https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_in_Kerr_coord.ipynb

D.5.3 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.5.4 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.5.5 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.5.6 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.5.7 Walker-Penrose Killing tensor on Kerr spacetime

This notebook accompanies Chap. 11; it shows that the tensor 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.5.8 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.5.9 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.5.10 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.5.11 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.5.12 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.5.13 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.5.14 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.5.15 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.5.16 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.5.17 Extremal Kerr spacetime

This notebook accompanies Chap. 13.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_extremal.ipynb

D.5.18 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.5.19 Isometric embeddings of (t, θ) sections of the extremal Kerr spacetime

This notebook accompanies Chap. 13.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Kerr_extremal_throat_emb.ipynb

D.5.20 Near-horizon extremal Kerr geometry

This notebook accompanies Chap. 13.

https://nbviewer.jupyter.org/github/sagemanifolds/SageManifolds/blob/master/Notebooks/SM_extremal_Kerr_near_horizon.ipynb

D.5.21 NHEK spacetime

This notebook accompanies Chap. 13.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/NHEK_spacetime.ipynb

D.6 Evolution and thermodynamics

D.6.1 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.6.2 Oppenheimer-Snyder collapse: spacetime diagrams

This notebook generates some figures for Chap. 14.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Oppenheimer_Snyder.ipynb

D.6.3 Oppenheimer-Snyder collapse: curvature

This notebook accompanies Chap. 14: it evaluates the curvature tensor and some associated scalars.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Oppenheimer_Snyder_curvature.ipynb

D.6.4 Vaidya spacetime

This notebook accompanies Chaps. 15 and 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>

D.6.5 Solving the ODE for outgoing radial null geodesics in Vaidya spacetime

This notebook accompanies Chap. 15.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Vaidya_solve_ode_out.ipynb

D.6.6 Naked singularity in Vaidya collapse

This notebook accompanies Chap. 15; it corresponds to the case where a naked singularity appears in the collapse of a radaition shell.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/Vaidya_nk_sing.ipynb

D.6.7 False trapped surface in Minkowski spacetime

This notebook generates Fig. 18.1.

https://nbviewer.jupyter.org/github/egourgoulhon/BHLectures/blob/master/sage/loc_cone_intersect.ipynb

Appendix E

Gyoto computations

Contents

E.1	Introduction	781
E.2	Image computations	781

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 [492]. 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/>
- Movie of Oppenheimer-Snyder collapse (Hirotaka Yoshino):
<http://ysnhrtk.web.fc2.com/animations.html>

Bibliography

- [1] R. Abraham and J.E. Marsden: *Foundations of Mechanics*, second edition, Addison-Wesley, Redwood City (1978).
- [2] M.A. Abramowicz and P.C. Fragile: *Foundations of Black Hole Accretion Disk Theory*, *Living Rev. Relativ.* **16**, 1 (2013).
- [3] 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 at <https://dlmf.nist.gov/>.
- [4] 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); arXiv: 1807.09409.
- [5] 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); arXiv: 2004.07187.
- [6] 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); arXiv: 1906.11238.
- [7] 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); arXiv: 1906.11242.
- [8] 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); arXiv: 1906.11243.
- [9] K. Akiyama et al. (the Event Horizon Telescope Collaboration): *The persistent shadow of the supermassive black hole of M 87. I. Observations, calibration, imaging, and analysis*, *Astron. Astrophys.* **681**, A79 (2024).
- [10] M. Alcubierre: *Introduction to 3+1 Numerical Relativity*, Oxford Univ. Press, Oxford (2008) DOI: 10.1093/acprof:oso/9780199205677.001.0001
- [11] 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); arXiv: 1304.0487.
- [12] W.L. Ames and K. Thorne: *The Optical Appearance of a Star that is Collapsing Through its Gravitational Radius*, *Astrophys. J.* **151**, 659 (1968).

- [13] L. Andersson, T. Bäckdahl, and P. Blue: *Geometry of Black Hole Spacetimes*, in *Asymptotic Analysis in General Relativity*, edited by T. Daudé, D. Häfner and J.-P. Nicolas, Cambridge Univ. Press, Cambridge (2018), p. 9; DOI: [10.1017/9781108186612.002](https://doi.org/10.1017/9781108186612.002); arXiv: [1610.03540](https://arxiv.org/abs/1610.03540).
- [14] N. Andersson: *Gravitational-Wave Astronomy: Exploring the Dark Side of the Universe*, Oxford Univ. Press, Oxford (2020); DOI: [10.1093/oso/9780198568032.001.0001](https://doi.org/10.1093/oso/9780198568032.001.0001).
- [15] S. Ansoldi: *Spherical black holes with regular center: a review of existing models including a recent realization with Gaussian sources*, in Proceedings of the second edition of the international workshop *Dynamics and Thermodynamics of Blackholes and Naked Singularities*, edited by L. Fatibene, M. Francaviglia, R. Giambò and G. Magli, published as arXiv:1205.2839 (2008); arXiv: [0802.0330](https://arxiv.org/abs/0802.0330).
- [16] T. Anson, E. Babichev, C. Charmousis and M. Hassaine: *Disforming the Kerr metric*, *J. High Energ. Phys.* **01(2021)**, 18 (2021); arXiv: [2006.06461](https://arxiv.org/abs/2006.06461).
- [17] S. Aretakis: *Nonlinear instability of scalar fields on extremal black holes*, *Phys. Rev. D* **87**, 084052 (2013); arXiv: [1304.4616](https://arxiv.org/abs/1304.4616).
- [18] S. Aretakis: *Horizon instability of extremal black holes*, *Adv. Theor. Math. Phys.* **19**, 507 (2015); arXiv: [1206.6598](https://arxiv.org/abs/1206.6598).
- [19] S. Aretakis: *Dynamics of Extremal Black Holes*, Springer, Cham (2018); DOI: [10.1007/978-3-319-95183-6](https://doi.org/10.1007/978-3-319-95183-6).
- [20] R. Arnowitt, S. Deser and C.W. Misner: *The dynamics of general relativity*, in *Gravitation: an introduction to current research*, edited by L. Witten, Wiley, New York (1962), p. 227; reprinted in *Gen. Relat. Grav.* **40**, 1997 (2008); arXiv: [gr-qc/0405109](https://arxiv.org/abs/gr-qc/0405109).
- [21] A. Ashtekar, J. Baez, A. Corichi, and K. Krasnov: *Quantum Geometry and Black Hole Entropy*, *Phys. Rev. Lett.* **80**, 904 (1998); arXiv: [gr-qc/9710007](https://arxiv.org/abs/gr-qc/9710007).
- [22] A. Ashtekar, C. Beetle, and J. Lewandowski: *Mechanics of rotating isolated horizons*, *Phys. Rev. D* **64**, 044016 (2001); arXiv: [gr-qc/0103026](https://arxiv.org/abs/gr-qc/0103026).
- [23] A. Ashtekar, C. Beetle, and J. Lewandowski: *Geometry of generic isolated horizons*, *Class. Quantum Grav.* **19**, 1195 (2002); arXiv: [gr-qc/0111067](https://arxiv.org/abs/gr-qc/0111067).
- [24] A. Ashtekar, L. Bombelli, and O. Reula: *The Covariant Phase Space of Asymptotically Flat Gravitational Fields*, in *Mechanics, Analysis and Geometry: 200 Years After Lagrange*, edited by M. Francaviglia, North Holland, Amsterdam (1991), p. 417; DOI: [10.1016/B978-0-444-88958-4.50021-5](https://doi.org/10.1016/B978-0-444-88958-4.50021-5).
- [25] A. Ashtekar, S. Fairhurst, and B. Krishnan: *Isolated horizons: Hamiltonian evolution and the first law*, *Phys. Rev. D* **62**, 104025 (2000); arXiv: [gr-qc/0005083](https://arxiv.org/abs/gr-qc/0005083).
- [26] A. Ashtekar and B. Krishnan: *Isolated and dynamical horizons and their applications*, *Living Rev. Relativ.* **7**, 10 (2004).
- [27] A. Ashtekar and A. Magnon-Ashtekar: *On conserved quantities in general relativity*, *J. Math. Phys.* **20**, 793 (1979).
- [28] E. Ayón-Beato and A. García: *The Bardeen model as a nonlinear magnetic monopole*, *Phys. Lett. B* **493**, 149 (2000); arXiv: [gr-qc/0009077](https://arxiv.org/abs/gr-qc/0009077).
- [29] C. Bambi: *Black Holes: A Laboratory for Testing Strong Gravity*, Springer, Singapore (2017); DOI: [10.1007/978-981-10-4524-0](https://doi.org/10.1007/978-981-10-4524-0).

- [30] C. Bambi (Ed.) : *Regular Black Holes: Towards a New Paradigm of Gravitational Collapse*, Springer, Singapore (2023); DOI: [10.1007/978-981-99-1596-5](https://doi.org/10.1007/978-981-99-1596-5).
- [31] R. Banerjee, B.R. Majhi, S.K. Modak, and S. Samanta: *Killing symmetries and Smarr formula for black holes in arbitrary dimensions*, *Phys. Rev. D* **82**, 124002 (2010); arXiv: [1007.5204](https://arxiv.org/abs/1007.5204).
- [32] J.F. Barbero G. and D. Pranzetti: *Black Hole Entropy in Loop Quantum Gravity*, in *Handbook of Quantum Gravity*, edited by C. Bambi, L. Modesto, and I. Shapiro, I., Springer, Singapore (2023); DOI: [10.1007/978-981-19-3079-9_104-1](https://doi.org/10.1007/978-981-19-3079-9_104-1); arXiv: [2212.13469](https://arxiv.org/abs/2212.13469).
- [33] G.V. Bard: *Sage for Undergraduates*, Americ. Math. Soc. (2015); preprint freely available from <http://www.gregory-bard.com/Sage.html>.
- [34] J.M. Bardeen: *Non-singular general relativistic gravitational collapse*, in *Abstracts of the 5th international conference on gravitation and the theory of relativity*, edited by V.A. Fock et al., Tbilisi University Press (1968).
- [35] J.M. Bardeen: *Kerr Metric Black Holes*, *Nature* **226**, 64 (1970).
- [36] J.M. Bardeen: *A Variational Principle for Rotating Stars in General Relativity*, *Astrophys. J.* **162**, 71 (1970).
- [37] 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.
- [38] J.M. Bardeen: *Rapidly Rotating Stars, Disks, and Black Holes*, in *Black Holes – Les astres occlus*, edited by C. DeWitt and B. DeWitt, Gordon and Breach, New York (1973), p. 241.
- [39] J.M. Bardeen, B. Carter, and S.W. Hawking: *The Four Laws of Black Hole Mechanics*, *Commun. Math. Phys.*, **31**, 161 (1973).
- [40] J. Bardeen and G.T. Horowitz: *Extreme Kerr throat geometry: A vacuum analog of $AdS_2 \times S^2$* , *Phys. Rev. D* **60**, 104030 (1999); arXiv: [hep-th/9905099](https://arxiv.org/abs/hep-th/9905099).
- [41] J.M. 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).
- [42] H. Barzegar, P.T. Chruściel, and M. Hörzinger: *Energy in higher-dimensional spacetimes*, *Phys. Rev. D* **96**, 124002 (2017); arXiv: [1708.03122](https://arxiv.org/abs/1708.03122).
- [43] 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](https://doi.org/10.1017/CBO9781139193344)
- [44] C.A.B. Bayona and N.R.F. Braga: *Anti-de Sitter boundary in Poincaré coordinates*, *Gen. Relativ. Gravit.* **39**, 1367 (2007); arXiv: [hep-th/0512182](https://arxiv.org/abs/hep-th/0512182).
- [45] D.L. Beckedorff: *Terminal Configurations of Stellar Evolution*, undergraduate senior thesis, Princeton University (1962); URL: <http://arks.princeton.edu/ark:/88435/dsp019593tx34r>.
- [46] K. Beckwith and C. Done: *Extreme gravitational lensing near rotating black holes*, *Month. Not. Roy. Astron. Soc.* **359**, 1217 (2005); arXiv: [astro-ph/0411339](https://arxiv.org/abs/astro-ph/0411339).
- [47] J.K. Beem, P.E. Ehrlich and K.L. Easley: *Global Lorentzian Geometry*, 2nd edition, Marcel Dekker, New York (1996); DOI: [10.1201/9780203753125](https://doi.org/10.1201/9780203753125).
- [48] R. Beig: *Arnowitt-Deser-Misner energy and g_{00}* , *Phys. Lett. A* **69**, 153 (1978).

- [49] J.D. Bekenstein: *Black holes and the second law*, *Lett. Nuovo Cimento* **4**, 737 (1972).
- [50] J.D. Bekenstein: *Extraction of Energy and Charge from a Black Hole*, *Phys. Rev. D* **7**, 949 (1973).
- [51] J.D. Bekenstein: *Black Holes and Entropy*, *Phys. Rev. D* **7**, 2333 (1973).
- [52] J.D. Bekenstein: *Generalized second law of thermodynamics in black-hole physics*, *Phys. Rev. D* **9**, 3292 (1974).
- [53] J. Ben Achour, H. Liu, H. Motohashi, S. Mukohyama, and K. Noui: *On rotating black holes in DHOST theories*, *J. Cosmo. Astropart. Phys.* **11**(2020), 001 (2020); arXiv: [2006.07245](#).
- [54] 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); arXiv: [astro-ph/0509839](#).
- [55] I. Bengtsson, E. Jakobsson, and J.M.M. Senovilla: *Trapped surfaces in Oppenheimer-Snyder black holes* *Phys. Rev. D* **88**, 064012 (2013)
DOI: [10.1103/PhysRevD.88.064012](#) arXiv: [1306.6486](#)
- [56] I. Bengtsson and P. Sandin: *Anti-de Sitter space, squashed and stretched*, *Class. Quantum Grav.* **23**, 971 (2006); arXiv: [gr-qc/0509076](#).
- [57] I. Bengtsson and J.M.M. Senovilla: *Note on trapped surfaces in the Vaidya solution*, *Phys. Rev. D* **79**, 024027 (2009); arXiv: [0809.2213](#).
- [58] M. Berger: *A Panoramic View of Riemannian Geometry*, Springer, Berlin (2003).
- [59] 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).
- [60] 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); arXiv: [1703.09525](#).
- [61] G.D. Birkhoff: *Relativity and Modern Physics*, Harvard University Press, Cambridge (1923).
- [62] R.L. Bishop and S.I. Goldberg: *Tensor Analysis on Manifolds*, Macmillan, New York (1968); reprinted by Dover, New York (1980).
- [63] 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); arXiv: [0803.2468](#).
- [64] L. Blanchet: *Post-Newtonian theory for gravitational waves*, *Living Rev. Relativ.* **27**, 4 (2024).
- [65] H. Bondi: *Spherically Symmetrical Models in General Relativity*, Month. Not. Roy. Astron. Soc. **107**, 410 (1947); reprinted in *Gen. Relativ. Gravit.* **31**, 1783 (1999).
- [66] I. Booth, L. Brits, J.A Gonzalez, and C. Van Den Broeck: *Marginally trapped tubes and dynamical horizons*, *Class. Quantum Grav.* **23**, 413 (2005); arXiv: [0506119](#).
- [67] I. Booth and S. Fairhurst: *Isolated, slowly evolving, and dynamical trapping horizons: Geometry and mechanics from surface deformations*, *Phys. Rev. D* **75**, 084019 (2007); arXiv: [gr-qc/0610032](#).
- [68] A. Börde: *Open and closed universes, initial singularities, and inflation*, *Phys. Rev. D* **50**, 3692 (1994)
DOI: [10.1103/PhysRevD.50.3692](#) arXiv: [gr-qc/9403049](#)

- [69] A. Börde: *Regular black holes and topology change*, *Phys. Rev. D* **55**, 7615 (1997); arXiv: gr-qc/9612057
- [70] R.H. Boyer: *Geodesic Killing orbits and bifurcate Killing horizons*, *Proc. Roy. Soc. Lond. A* **311**, 245 (1969).
- [71] R.H. Boyer and R.W. Lindquist: *Maximal Analytic Extension of the Kerr Metric*, *J. Math. Phys.* **8**, 265 (1967).
- [72] R.H. Boyer and T.G. Price: *An interpretation of the Kerr metric in general relativity*, *Math. Proc. Cambridge Phil. Soc.* **61**, 531 (1965).
- [73] D. Brill: *History of a black hole horizon*, *Grav. Cosmo.* **20**, 165 (2014).
- [74] R. Brout, S. Massar, R. Parentani, and P. Spindel: *A Primer for Black Hole Quantum Physics*, *Phys. Rep.* **260**, 329 (1995); arXiv: 0710.4345.
- [75] G.L. Bunting: *Proof of the uniqueness conjecture for black holes*, Ph.D. thesis, University of New England, Armidale, Australia (1983).
- [76] G.L. Bunting and A.K.M. Masood-ul-Alam: *Nonexistence of multiple black holes in asymptotically Euclidean static vacuum space-time*, *Gen. Relat. Grav.* **19**, 147 (1987).
- [77] L.M. Burko, G. Khanna, and A. Zenginoğlu: *Cauchy-horizon singularity inside perturbed Kerr black holes*, *Phys. Rev. D* **93**, 041501(R) (2016); arXiv: 1601.05120.
- [78] 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
- [79] A. Čadež and U. Kostić: *Optics in the Schwarzschild spacetime*, *Phys. Rev. D* **72**, 104024 (2005); arXiv: gr-qc/0405037.
- [80] L.-M. Cao: *Deformation of codimension-2 surfaces and horizon thermodynamics*, *J. High Energ. Phys.* **2011**, 112 (2011); arXiv: 1009.4540.
- [81] L. Capuano, L. Santoni, and E. Barausse: *Black hole hairs in scalar-tensor gravity and the lack thereof*, *Phys. Rev. D* **108**, 064058 (2023); arXiv: 2304.12750.
- [82] V. Cardoso: *The hole picture*, *Nature Review Phys.* **1**, 701 (2019); arXiv: 1910.04173.
- [83] V. Cardoso, F. Duque, and A. Foschi: *The light ring and the appearance of matter accreted by black holes*, *Phys. Rev. D* **103**, 104044 (2021); arXiv: 2102.07784.
- [84] S. Carlip: *Black hole thermodynamics*, *Int. J. Mod. Phys. D* **23**, 1430023 (2014); arXiv: 1410.1486.
- [85] B. Carr, K. Kohri, Y. Sendouda, and J. Yokoyama: *Constraints on primordial black holes*, *Rep. Prog. Phys.* **84**, 116902 (2021); arXiv: 2002.12778.
- [86] S.M. Carroll: *Spacetime and Geometry: An Introduction to General Relativity*, Addison Wesley (Pearson Education), San Francisco (2004); <https://www.preposterousuniverse.com/spacetimeandgeometry/>.
- [87] B. Carter: *Complete Analytic Extension of the Symmetry Axis of Kerr's Solution of Einstein's Equations*, *Phys. Rev.* **141**, 1242 (1966).
- [88] B. Carter: *Stationary Axisymmetric Systems in General Relativity*, Ph. D. Thesis, Cambridge University (1967); available at <https://luth.obspm.fr/~luthier/carter/Thesis/>.
- [89] B. Carter: *Global Structure of the Kerr Family of Gravitational Fields*, *Phys. Rev.* **174**, 1559 (1968).

- [90] B. Carter: *Hamilton-Jacobi and Schrodinger Separable Solutions of Einstein's Equations*, *Commun. Math. Phys.* **10**, 280 (1968).
- [91] B. Carter: *Killing horizons and orthogonally transitive groups in space-time*, *J. Math. Phys.* **10**, 70 (1969).
- [92] B. Carter: *Axisymmetric Black Hole Has Only Two Degrees of Freedom*, *Phys. Rev. Lett.* **26**, 331 (1971).
- [93] B. Carter: *Rigidity of a Black Hole*, *Nature Physical Science* **238**, 71 (1972).
- [94] 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).
- [95] 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).
- [96] B. Carter: *Perfect fluid and magnetic field conservation laws in the theory of black hole accretion rings*, in *Active Galactic Nuclei*, edited by C. Hazard and S. Mitton, Cambridge Univ. Press, Cambridge (1979), p. 273; available at <https://luth.obspm.fr/~luthier/carter/trav/Carter79.pdf>.
- [97] B. Carter: *Generalised mass variation formula for a stationary axisymmetric star or black hole with surrounding accretion disc*, in *Journées Relativistes 1979*, edited by I. Moret-Bailly and C. Latrémolière, Faculté des Sciences, Angers (1979), p. 166; arXiv: [astro-ph/0206113](https://arxiv.org/abs/astro-ph/0206113).
- [98] B. Carter: *The general theory of the mechanical, electromagnetic and thermodynamic properties of black holes*, in *General Relativity: An Einstein Centenary Survey*, edited by S.W. Hawking and W. Israel, Cambridge Univ. Press, Cambridge (1979), p. 294.
- [99] B. Carter: *Bunting identity and Mazur identity for non-linear elliptic systems including the black hole equilibrium problem*, *Commun. Math. Phys.* **99**, 563 (1985).
- [100] B. Carter: *Mathematical foundations of the theory of relativistic stellar and black hole configurations*, in *Gravitation in Astrophysics*, edited by B. Carter and J.B. Hartle, Plenum Press, New York (1987), p. 63.
- [101] B. Carter: *Extended tensorial curvature analysis for embeddings and foliations*, *Contemp. Math.* **203**, 207 (1997).
- [102] B. Carter: *Has the black hole equilibrium problem been solved?*, in *The Eighth Marcel Grossmann Meeting*, edited by T. Piran, World Scientific, Singapore (1999), p. 136; arXiv: [gr-qc/9712038](https://arxiv.org/abs/gr-qc/9712038).
- [103] B. Carter: *Energy dominance and the Hawking Ellis vacuum conservation theorem*, in *The Future of Theoretical Physics and Cosmology — Celebrating Stephen Hawking's Contributions to Physics*, edited by G.W. Gibbons, E.P.S. Shellard, and S.J. Rankin, Cambridge Univ. Press, Cambridge (2003), p. 177; arXiv: [gr-qc/0205010](https://arxiv.org/abs/gr-qc/0205010).
- [104] B. Carter: *Half Century of Black-Hole Theory: From Physicists' Purgatory to Mathematicians' Paradise*, in *A Century of Relativity Physics: ERE 2005; XXVIII Spanish Relativity*

- Meeting*, edited by L. Mornas and J. Diaz Alonso, *AIP Conference Proceedings* **841**, 29 (2006); arXiv: gr-qc/0604064.
- [105] M. Casals, S.E. Gralla, and P. Zimmerman: *Horizon instability of extremal Kerr black holes: Nonaxisymmetric modes and enhanced growth rate*, *Phys. Rev. D* **94**, 064003 (2016); arXiv: 1606.08505.
- [106] S. Chandrasekhar: *The Mathematical Theory of Black Holes*, Oxford Univ. Press, New York (1983); DOI: 10.1093/oso/9780198503705.001.0001.
- [107] Y. Choquet-Bruhat: *General Relativity and Einstein's Equations*, Oxford Univ. Press, New York (2009); DOI: 10.1093/acprof:oso/9780199230723.001.0001.
- [108] Y. Choquet-Bruhat: *Introduction to General Relativity, Black Holes, and Cosmology*, Oxford Univ. Press, New York (2015); DOI: 10.1093/oso/9780199666454.001.0001.
- [109] Y. Choquet-Bruhat, C. De Witt-Moretten, and M. Dillard-Bleick: *Analysis, Manifolds and Physics*, North-Holland, Amsterdam (1977).
- [110] Y. Choquet-Bruhat and R. Geroch: *Global aspects of the Cauchy problem in general relativity*, *Commun. Math. Phys.* **14**, 329 (1969).
- [111] D. Christodoulou: *Reversible and Irreversible Transformations in Black-Hole Physics*, *Phys. Rev. Lett.* **25**, 1596 (1970).
- [112] D. Christodoulou: *The Formation of Black Holes in General Relativity*, European Mathematical Society, Zürich (2009); DOI: 10.4171/068; arXiv: 0805.3880.
- [113] D. Christodoulou and R. Ruffini: *Reversible Transformations of a Charged Black Hole*, *Phys. Rev. D* **4**, 3552 (1971).
- [114] P.T. Chruściel: “No Hair” Theorems – Folklore, Conjectures, Results, in *Differential Geometry and Mathematical Physics, Contemporary Mathematics* **170**; edited by J.K. Beem and K.L. Duggal, American Mathematical Society, Providence (1994), p. 23; arXiv: gr-qc/9402032.
- [115] P.T. Chruściel: *On Rigidity of Analytic Black Holes*, *Commun. Math. Phys.* **189**, 1 (1997); arXiv: gr-qc/9610011.
- [116] 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); arXiv: gr-qc/0201053.
- [117] 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.
- [118] P.T. Chruściel: *Geometry of Black Holes*, Oxford Univ. Press, Oxford (2020); <https://global.oup.com/academic/product/geometry-of-black-holes-9780198855415>.
- [119] P.T. Chruściel, E. Delay, G. Galloway, and R. Howard: *Regularity of horizons and the area theorem*, *Annales Henri Poincaré* **2**, 109 (2001); arXiv: gr-qc/0001003.
- [120] P.T. Chruściel and G.J. Galloway: *Uniqueness of static black holes without analyticity*, *Class. Quantum Grav.* **27**, 152001 (2010); arXiv: 1004.0513.
- [121] P.T. Chruściel and J. Lopes Costa: *On uniqueness of stationary vacuum black holes*, *Astérisque* **321**, 195 (2008); arXiv: 0806.0016.
- [122] P.T. Chruściel, J. Lopes Costa, and M. Heusler: *Stationary Black Holes: Uniqueness and Beyond*, *Living Rev. Relativ.* **15**, 7 (2012).

- [123] P.T. Chruściel, M. Maliborski, and N. Yunes: *Structure of the singular ring in Kerr-like metrics*, *Phys. Rev. D* **101**, 104048 (2020); arXiv: [1912.06020](#).
- [124] P.T. Chruściel and L. Nguyen: *A Uniqueness Theorem for Degenerate Kerr–Newman Black Holes*, *Ann. Henri Poincaré* **11**, 585 (2010); arXiv: [1002.1737](#).
- [125] P.T. Chruściel, C.R. Ölz, and S.J. Szybka: *Space-time diagrammatics*, *Phys. Rev. D* **86**, 124041 (2012); arXiv: [1211.1718](#).
- [126] P.T. Chruściel, H.S. Reall and P. Tod: *On non-existence of static vacuum black holes with degenerate components of the event horizon*, *Class. Quantum Grav.* **23**, 549 (2006); arXiv: [gr-qc/0512041](#).
- [127] P.T. Chruściel and P. Tod: *The Classification of Static Electro–Vacuum Space–Times Containing an Asymptotically Flat Spacelike Hypersurface with Compact Interior*, *Commun. Math. Phys.* **271**, 577 (2007); arXiv: [gr-qc/0512043](#).
- [128] P.T. Chruściel and R.M. Wald: *Maximal hypersurfaces in stationary asymptotically flat spacetimes*, *Commun. Math. Phys.* **163**, 561 (1994); arXiv: [gr-qc/9304009](#).
- [129] P.T. Chruściel and R.M. Wald: *On the topology of stationary black holes*, *Class. Quantum Grav.* **11**, L147 (1994); arXiv: [gr-qc/9410004](#).
- [130] T. Chu, H.P. Pfeiffer, and M.I. Cohen: *Horizon dynamics of distorted rotating black holes*, *Phys. Rev. D* **83**, 104018 (2011); arXiv: [1011.2601](#).
- [131] M.I. Cohen, J.D. Kaplan, and M.A. Scheel: *Toroidal horizons in binary black hole inspirals*, *Phys. Rev. D* **85**, 024031 (2012); arXiv: [1110.1668](#).
- [132] M.I. Cohen, H.P. Pfeiffer, and M.A. Scheel: *Revisiting event horizon finders*, *Class. Quantum Grav.* **26**, 035005 (2009); arXiv: [0809.2628](#).
- [133] P. Collas and D. Klein: *Embeddings and time evolution of the Schwarzschild wormhole*, *Amer. J. Phys.* **80**, 203 (2012); arXiv: [1107.4871](#).
- [134] S. Collin-Zahn: *Des quasars aux trous noirs*, EDP Sciences, Les Ulis (2009).
- [135] G. Compère: *An introduction to the mechanics of black holes*, Lecture notes, Second Modave Summer School in Mathematical Physics (2006); arXiv: [gr-qc/0611129](#).
- [136] G. Compère: *The Kerr/CFT correspondence and its extensions*, *Living Rev. Relativ.* **20**, 1 (2017).
- [137] G. Compère: *Advanced Lectures on General Relativity*, Springer, Cham (2019); DOI: [10.1007/978-3-030-04260-8](#); arXiv: [1801.07064](#).
- [138] G. Compère and A. Druart: *Near-horizon geodesics of high spin black holes*, *Phys. Rev. D* **101**, 084042 (2020); erratum in *Phys. Rev. D* **102**, 029901 (2020); arXiv: [2001.03478](#).
- [139] G. Compère, K. Fransen, T. Hertog, and J. Long: *Gravitational waves from plunges into Gargantua*, *Class. Quantum Grav.* **35**, 104002 (2018); erratum in *Class. Quantum Grav.* **36**, 179501 (2019); arXiv: [1712.07130](#).
- [140] G. Contopoulos: *Orbits through the ergosphere of a Kerr black hole*, *Gen. Relat. Grav.* **16**, 43 (1984).
- [141] 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).

- [142] A. Coudray and J.-P. Nicolas: *Geometry of Vaidya spacetimes*, *Gen. Relat. Grav.* **53**, 73 (2021); arXiv: 2101.06544.
- [143] P.V.P. Cunha and C.A.R. Herdeiro: *Shadows and strong gravitational lensing: a brief review*, *Gen. Relat. Grav.* **50**, 42 (2018); arXiv: 1801.00860.
- [144] 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); arXiv: 1705.05461.
- [145] C.T. Cunningham and J.M. Bardeen: *The Optical Appearance of a Star Orbiting an Extreme Kerr Black Hole* *Astrophys. J.* **183**, 237 (1973).
- [146] E. Curiel: *A Potted Early History of “black hole”*, appendix to K. Landsman’s article [333] (2021).
- [147] H.D. Curtis: *Descriptions of 762 Nebulae and Clusters Photographed with the Crossley Reflector*, *Publications of Lick Observatory* **13**, 9 (1918).
- [148] 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), p. 97; available at <https://people.maths.ox.ac.uk/cmi/library/proceedings/cmip017c.pdf>; arXiv: 0811.0354.
- [149] T. Damour: *Black-hole eddy currents*, *Phys. Rev. D* **18**, 3598 (1978).
- [150] T. Damour: *Quelques propriétés mécaniques, électromagnétiques, thermodynamiques et quantiques des trous noirs*, Thèse de doctorat d’État, Université Paris 6 (1979); available at <https://pagesperso.ihes.fr/~damour/Articles/>.
- [151] T. Damour: *Surface effects in black hole physics*, in *Proceedings of the Second Marcel Grossmann Meeting on General Relativity*, edited by R. Ruffini, North Holland, Amsterdam (1982), p. 587; available at <https://pagesperso.ihes.fr/~damour/Articles/>.
- [152] 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, Birkhäuser, Basel (2004), p. 227; URL: <https://link.springer.com/book/9783764371067>; arXiv: hep-th/0401160.
- [153] G. Darmois: *Les équations de la gravitation einsteinienne*, Mémorial des sciences mathématiques, no. 25 (1927); available at https://www.numdam.org/item/?id=MSM_1927_25_1_0.
- [154] C. G. Darwin: *The gravity field of a particle*, *Proc. Roy. Soc. Lond. A* **249**, 180 (1959).
- [155] C. G. Darwin: *The gravity field of a particle. II*, *Proc. Roy. Soc. Lond. A* **263**, 39 (1961).
- [156] 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).
- [157] I. Davies and H.S. Reall: *Nonperturbative Second Law of Black Hole Mechanics in Effective Field Theory*, *Phys. Rev. Lett.* **132**, 171402 (2024); arXiv: 2312.07659.
- [158] F. de Felice: *Equatorial Geodesic Motion in the Gravitational Field of a Rotating Source*, *Nuovo Cimento B* **57**, 351 (1968).
- [159] 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).

- [160] 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>.
- [161] N. Deruelle and J.-P. Uzan: *Théories de la Relativité*, Belin, Paris (2014).
- [162] N. Deruelle and J.-P. Uzan: *Relativity in Modern Physics*, Oxford Univ. Press, Oxford (2018); DOI: [10.1093/oso/9780198786399.001.0001](https://doi.org/10.1093/oso/9780198786399.001.0001).
- [163] C. DeWitt and B. DeWitt (eds.): *Black Holes – Les astres occlus*, Gordon and Breach, New York (1973).
- [164] J. Dexter and E. Agol: *A Fast New Public Code for Computing Photon Orbits in a Kerr Spacetime*, *Astrophys. J.* **696**, 1616 (2009); arXiv: [0903.0620](https://arxiv.org/abs/0903.0620).
- [165] V.I. Dokuchaev: *Is there life inside black holes?*, *Class. Quantum Grav.* **28**, 235015 (2011); arXiv: [1103.6140](https://arxiv.org/abs/1103.6140).
- [166] 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); arXiv: [1812.06787](https://arxiv.org/abs/1812.06787).
- [167] V.I. Dokuchaev and N.O. Nazarova: *Event Horizon Image within Black Hole Shadow*, *J. Exp. Theor. Phys.* **128**, 578 (2019); arXiv: [1804.08030](https://arxiv.org/abs/1804.08030).
- [168] V.I. Dokuchaev and N.O. Nazarova: *Silhouettes of invisible black holes*, *Physics-Uspekhi* **63**, 583 (2020); arXiv: [1911.07695](https://arxiv.org/abs/1911.07695).
- [169] V.I. Dokuchaev and N.O. Nazarova: *Visible shapes of black holes M87* and SgrA**, *Universe* **6**, 154 (2020); arXiv: [2007.14121](https://arxiv.org/abs/2007.14121).
- [170] X. Dong: *Holographic entanglement entropy for general higher derivative gravity*, *J. High Energ. Phys.* **01(2014)**, 44 (2014); arXiv: [1310.5713](https://arxiv.org/abs/1310.5713).
- [171] A.G. Doroshkevich, Ya.B. Zeldovich and I.D. Novikov: *Gravitational Collapse of Non-symmetric and Rotating Masses* (in Russian), Zhurnal Éksperimental'noi i Teoreticheskoi Fiziki **49**, 170 (1965); English translation in *Soviet Physics JETP* **22**, 122 (1966).
- [172] S. Drasco and S.A. Hughes: *Rotating black hole orbit functionals in the frequency domain*, *Phys. Rev. D* **69**, 044015 (2004); arXiv: [astro-ph/0308479](https://arxiv.org/abs/astro-ph/0308479).
- [173] S. Drasco and S.A. Hughes: *Gravitational wave snapshots of generic extreme mass ratio inspirals*, *Phys. Rev. D* **73**, 024027 (2006); arXiv: [gr-qc/0509101](https://arxiv.org/abs/gr-qc/0509101).
- [174] 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).
- [175] D.M. Eardley and L. Smarr: *Time functions in numerical relativity: Marginally bound dust collapse*, *Phys. Rev. D* **19**, 2239 (1979).
- [176] A.S. Eddington: *A comparison of Whitehead's and Einstein's Formulæ*, *Nature* **113**, 192 (1924).
- [177] C.A. Egan and C.H. Lineweaver: *A Larger Estimate of the Entropy of the Universe*, *Astrophys. J.* **710**, 1825 (2010); arXiv: [0909.3983](https://arxiv.org/abs/0909.3983).
- [178] A. Einstein and N. Rosen: *The Particle Problem in the General Theory of Relativity*, *Phys. Rev.* **48**, 73 (1935).

- [179] J. Eisenstaedt: *Histoire et Singularités de la Solution de Schwarzschild (1915-1923)*, *Archive for History of Exact Sciences* **27**, 157 (1982).
- [180] J. Eisenstaedt: *Trajectoires et Impasses de la Solution de Schwarzschild*, *Archive for History of Exact Sciences* **37**, 275 (1987).
- [181] 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.
- [182] G.F.R. Ellis: *Relativistic cosmology*, in *Proceedings of the International School of Physics "Enrico Fermi"*, Course 47: *General relativity and cosmology*, edited by B. K. Sachs, Academic Press, New York and London (1971), p. 104; reprinted in *Gen. Relat. Grav.* **41**, 581 (2009).
- [183] H. Elvang and P. Figueras: *Black saturn*, *J. High Energ. Phys.* **05(2007)**, 050 (2007); arXiv: [hep-th/0701035](#).
- [184] R. Emparan, S. Ohashi and T. Shiromizu: *No-dipole-hair theorem for higher-dimensional static black holes*, *Phys. Rev. D* **82**, 084032 (2010); arXiv: [1007.3847](#).
- [185] R. Emparan and H.S. Reall: *A Rotating Black Ring Solution in Five Dimensions*, *Phys. Rev. Lett.* **88**, 101101 (2002); arXiv: [hep-th/0110260](#).
- [186] R. Emparan and H.S. Reall: *Black Holes in Higher Dimensions*, *Living Rev. Relativ.* **11**, 6 (2008).
- [187] J. Engle, K. Noui, and A. Perez: *Black Hole Entropy and SU(2) Chern-Simons Theory*, *Phys. Rev. Lett.* **105**, 031302 (2010); arXiv: [0905.3168](#).
- [188] H. Eschrig: *Topology and Geometry for Physics*, Springer, Berlin (2011); DOI: [10.1007/978-3-642-14700-5](#).
- [189] A. Fabbri and J. Navarro-Salas: *Modeling Black Hole Evaporation*, Imperial College Press, London, and World Scientific, Singapore (2005); DOI: [10.1142/p378](#).
- [190] H. Falcke: *Imaging black holes: past, present and future*, *J. Phys. Conf. Ser.* **942**, 012001 (2017); arXiv: [1801.03298](#).
- [191] H. Falcke, F. Melia, and E. Agol: *Viewing the Shadow of the Black Hole at the Galactic Center*, *Astrophys. J.* **528**, L13 (2000); arXiv: [astro-ph/9912263](#).
- [192] 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); arXiv: [2007.06732](#).
- [193] F. Fayos and R. Torres: *A class of interiors for Vaidya's radiating metric: singularity-free gravitational collapse*, *Class. Quantum Grav.* **25**, 175009 (2008).
- [194] D. Finkelstein: *Past-Future Asymmetry of the Gravitational Field of a Point Particle*, *Phys. Rev.* **110**, 965 (1958).
- [195] A. Fischer and J. Marsden: *The initial value problem and the dynamical formulation of general relativity*, in *General Relativity: an Einstein Centenary Survey*, edited by S.W. Hawking and W. Israel, Cambridge Univ. Press, Cambridge (1979), p. 138.
- [196] L. Flamm: *Beiträge zur Einsteinschen Gravitationstheorie*, *Physikalische Zeitschrift* **17**, 448 (1916); English translation in *Gen. Relativ. Gravit.* **47**, 72 (2015).
- [197] J. Frauendiener: *Conformal Infinity*, *Living Rev. Relativ.* **7**, 1 (2004).

- [198] J.L. Friedman, K. Schleich, and D.M. Witt: *Topological censorship*, *Phys. Rev. Lett.* **71**, 1486 (1993); erratum in *Phys. Rev. Lett.* **75**, 1872 (1995); arXiv: gr-qc/9305017.
- [199] H. Friedrich, I. Rácz, and R.M. Wald: *On the Rigidity Theorem for Spacetimes with a Stationary Event Horizon or a Compact Cauchy Horizon*, *Commun. Math. Phys.* **204**, 691 (1999); arXiv: gr-qc/9811021.
- [200] V.P. Frolov, P. Krtouš, and D. Kubiznák: *Black holes, hidden symmetries, and complete integrability*, *Living Rev. Relativ.* **20**, 6 (2017).
- [201] 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.
- [202] V.P. Frolov and K.S. Thorne: *Renormalized stress-energy tensor near the horizon of a slowly evolving, rotating black hole* *Phys. Rev. D* **39**, 2125 (1989).
- [203] V.P. Frolov and A. Zelnikov: *Introduction to Black Hole Physics*, Oxford Univ. Press, New York (2011); DOI: 10.1093/acprof:oso/9780199692293.001.0001.
- [204] C. Fronsdal: *Completion and Embedding of the Schwarzschild Solution*, *Phys. Rev.* **116**, 778 (1959).
- [205] R. Fujita & W. Hikida: *Analytical solutions of bound timelike geodesic orbits in Kerr spacetime*, *Class. Quantum Grav.* **26**, 135002 (2009); arXiv: 0906.1420.
- [206] D. Gajic: *Azimuthal instabilities on extremal Kerr*, submitted (2023); arXiv: 2302.06636.
- [207] G.J. Galloway: *Least area tori, black holes and topological censorship*, in *Differential Geometry and Mathematical Physics, Contemporary Mathematics* **170**; edited by J.K. Beem and K.L. Duggal, American Mathematical Society, Providence (1994), p. 113.
- [208] G.J. Galloway: *Null Geometry and the Einstein Equations*, in *The Einstein Equations and the Large Scale Behavior of Gravitational Fields*, edited by P.T. Chruściel and H. Friedrich, Birkhäuser, Basel (2004), p. 379; DOI: 10.1007/978-3-0348-7953-8_11; available at <https://www.math.miami.edu/~galloway/papers/cargproc.pdf>.
- [209] G.J. Galloway and R. Schoen: *A Generalization of Hawking's Black Hole Topology Theorem to Higher Dimensions*, *Commun. Math. Phys.* **266**, 571 (2006); arXiv: gr-qc/0509107.
- [210] G.J. Galloway and J.M.M. Senovilla: *Singularity theorems based on trapped submanifolds of arbitrary co-dimension*, *Class. Quantum Grav.* **27**, 152002 (2010); arXiv: 1005.1249.
- [211] S. Gao and R.M. Wald: “Physical process version” of the first law and the generalized second law for charged and rotating black holes, *Phys. Rev. D* **64**, 084020 (2001); arXiv: gr-qc/0106071.
- [212] D. Garfinkle and E.N. Glass: *Bertotti-Robinson and Melvin spacetimes*, *Class. Quantum Grav.* **28**, 215012 (2011); arXiv: 1109.1535.
- [213] K. Gebhardt, J. Adams, D. Richstone, T.R. Lauer, S.M. Faber, K. Gultekin, J. Murphy, and S. Tremaine: *The Black Hole Mass in M87 from Gemini/NIFS Adaptive Optics Observations*, *Astrophys. J.* **729**, 119 (2011); arXiv: 1101.1954.
- [214] R. Genzel, F. Eisenhauer and S. Gillessen: *The Galactic Center massive black hole and nuclear star cluster*, *Rev. Mod. Phys.* **82**, 3121 (2010); arXiv: 1006.0064.
- [215] R. Geroch: *What is a singularity in general relativity?* *Ann. Phys.* **48**, 526 (1968).

- [216] R. Geroch and J.B. Hartle: *Distorted black holes*, *J. Math. Phys.* **23**, 680 (1982).
- [217] G.W. Gibbons, D. Ida and T. Shiromizu: *Uniqueness of (dilatonic) charged black holes and black p-branes in higher dimensions*, *Phys. Rev. D* **66**, 044010 (2002); arXiv: [hep-th/0206136](#).
- [218] G.W. Gibbons, D. Ida and T. Shiromizu: *Uniqueness and Non-Uniqueness of Static Vacuum Black Holes in Higher Dimensions*, *Prog. Theor. Phys. Suppl.* **148**, 284 (2002); arXiv: [gr-qc/0203004](#).
- [219] G.W. Gibbons and M. Vyska: *The application of Weierstrass elliptic functions to Schwarzschild null geodesics*, *Class. Quantum Grav.* **29**, 065016 (2012); arXiv: [1110.6508](#).
- [220] 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); arXiv: [gr-qc/0611006](#).
- [221] V.L. Ginzburg: *The Magnetic Fields of Collapsing Masses and the Nature of Superstars* (in Russian), *Doklady Akademii Nauk SSSR* **156**, 43 (1964); English translation in *Soviet Physics Doklady* **9**, 329 (1964).
- [222] 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); arXiv: [gr-qc/0203086](#).
- [223] K. Glampedakis and G. Pappas: *Can supermassive black hole shadows test the Kerr metric?*, *Phys. Rev. D* **104**, L081503 (2021); arXiv: [2102.13573](#).
- [224] 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); arXiv: [1308.4760](#).
- [225] E. Gourgoulhon : *An introduction to relativistic hydrodynamics*, in *Stellar Fluid Dynamics and Numerical Simulations: From the Sun to Neutron Stars*, edited by M. Rieutord and B. Dubrulle, *EAS Publications Series* **21**, 43-79 (2006); arXiv: [gr-qc/0603009](#).
- [226] E. Gourgoulhon: *3+1 Formalism in General Relativity; Bases of Numerical Relativity*, Springer, Berlin (2012); URL: <https://relativite.obspm.fr/3p1>.
- [227] E. Gourgoulhon: *Special Relativity in General Frames: From Particles to Astrophysics*, Springer, Berlin (2013); URL: <https://relativite.obspm.fr/sperel>.
- [228] E. Gourgoulhon: *Relativité générale*, Master course lecture notes; URL: <https://luth.obspm.fr/~luthier/gourgoulhon/fr/master/relat.html>.
- [229] E. Gourgoulhon and J.L. Jaramillo: *A 3+1 perspective on null hypersurfaces and isolated horizons*, *Phys. Rep.* **423**, 159 (2006); arXiv: [gr-qc/0503113](#).
- [230] 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 Trzebiezowice, Poland (7-13 October 2007), edited by M. Abramowicz and O. Straub, *New Astronomy Reviews* **51**, 791 (2008); arXiv: [0803.2944](#).
- [231] 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).
- [232] S.E. Gralla: *Can the EHT M87 results be used to test general relativity?*, *Phys. Rev. D* **103**, 024023 (2021); arXiv: [2010.08557](#).

- [233] S.E. Gralla, D.E. Holz, and R.M. Wald: *Black hole shadows, photon rings, and lensing rings*, *Phys. Rev. D* **100**, 024018 (2019); arXiv: [1906.00873](#).
- [234] S.E. Gralla and A. Lupsasca: *Lensing by Kerr black holes*, *Phys. Rev. D* **101**, 044031 (2020); arXiv: [1910.12873](#).
- [235] S.E. Gralla and A. Lupsasca: *Null geodesics of the Kerr exterior*, *Phys. Rev. D* **101**, 044032 (2020); arXiv: [1910.12881](#).
- [236] S.E. Gralla and A. Lupsasca: *Observable shape of black hole photon rings*, *Phys. Rev. D* **102**, 124003 (2020); arXiv: [2007.10336](#).
- [237] 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); arXiv: [1710.11112](#).
- [238] S.E. Gralla, A.P. Porfyriadis, and N. Warburton: *Particle on the innermost stable circular orbit of a rapidly spinning black hole*, *Phys. Rev. D* **92**, 064029 (2015); arXiv: [1506.08496](#).
- [239] S.E. Gralla and P. Zimmerman: *Critical exponents of extremal Kerr perturbations*, *Class. Quantum Grav.* **35**, 095002 (2018); arXiv: [1711.00855](#).
- [240] P. Grandclément: *Fully consistent rotating black holes in the cubic Galileon theory*, *Class. Quantum Grav.* **41**, 025012 (2024); arXiv: [2308.11245](#).
- [241] A.A. Grib and Y.V. Pavlov: *Black holes and particles with zero or negative energy*, *Theor. Math. Phys.* **190**, 268 (2017); arXiv: [1601.02592](#).
- [242] J.B. Griffiths and J. Podolský: *Exact Space-Times in Einstein's General Relativity*, Cambridge Univ. Press, Cambridge (2009); DOI: [10.1017/CBO9780511635397](#).
- [243] R. Grossman, J. Levin, and G. Perez-Giz: *Harmonic structure of generic Kerr orbits*, *Phys. Rev. D* **85**, 023012 (2012); arXiv: [1105.5811](#).
- [244] D. Grumiller, R. McNees and J. Salzer: *Black Holes and Thermodynamics: The First Half Century*, in *Quantum Aspects of Black Holes*, edited by X. Calmet, Springer, Cham (2015), p. 27; DOI: [10.1007/978-3-319-10852-0_2](#); arXiv: [1402.5127](#).
- [245] D. Grumiller and M.M. Sheikh-Jabbari: *Black Hole Physics: From Collapse to Evaporation*, Springer, Cham (2022); DOI: [10.1007/978-3-031-10343-8](#).
- [246] M. Guica, T. Hartman, W. Song, and A. Strominger: *The Kerr/CFT correspondence*, *Phys. Rev. D* **80**, 124008 (2009); arXiv: [0809.4266](#).
- [247] A. Gullstrand: *Allgemeine Lösung des statischen Einkörperproblems in der Einsteinschen Gravitationstheorie*, *Arkiv. Mat. Astron. Fys.* **16**, 1 (1922).
- [248] Y. Hagihara: *Theory of the Relativistic Trajectories in a Gravitational Field of Schwarzschild*, *Japanese Journal of Astronomy and Geophysics* **8**, 67 (1931).
- [249] K. Hajian and M.M. Sheikh-Jabbari: *Solution phase space and conserved charges: A general formulation for charges associated with exact symmetries*, *Phys. Rev. D* **93**, 044074 (2016); arXiv: [1512.05584](#).
- [250] P. Hájíček: *Exact models of charged black holes. I. Geometry of totally geodesic null hypersurface*, *Commun. Math. Phys.* **34**, 37 (1973).
- [251] P. Hájíček: *Exact models of charged black holes. II. Axisymmetric stationary horizons*, *Commun. Math. Phys.* **34**, 53 (1973).

- [252] P. Hájíček: *Can outside fields destroy black holes?*, *J. Math. Phys.* **15**, 1554 (1974).
- [253] J. Haláček and T. Ledvinka: *The analytic conformal compactification of the Schwarzschild spacetime*, *Class. Quantum Grav.* **31**, 015007 (2014); arXiv: 1401.1337.
- [254] D. Harlow and J. Wu: *Covariant phase space with boundaries*, *J. High Energ. Phys.* **10(2020)**, 146 (2020); arXiv: 1906.08616.
- [255] B.K. Harrison, K.S. Thorne, M. Wakano and J.A. Wheeler: *Gravitation Theory and Gravitational Collapse*, University of Chicago Press, Chicago (1965); URL: <https://archive.org/details/GravitationalTheoryandCollapseHarrison65-17293>.
- [256] J.B. Hartle: *Gravity: An Introduction to Einstein's General Relativity*, Addison Wesley (Pearson Education), San Fransisco (2003); URL: <https://archive.org/details/JBHartleGravityAnIntroductionToEinsteinsGeneralRelativityAddisonWesley20031>.
- [257] J.B. Hartle and S.W. Hawking: *Solutions of the Einstein-Maxwell equations with many black holes*, *Commun. Math. Phys.* **26**, 87 (1972); available at <https://projecteuclid.org/journalArticle/Download?urlId=cmp%2F1103858037>.
- [258] S.W. Hawking: *Gravitational radiation from colliding black holes*, *Phys. Rev. Lett.* **26**, 1344 (1971).
- [259] S.W. Hawking: *Black holes in general relativity*, *Commun. Math. Phys.* **25**, 152 (1972).
- [260] 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.
- [261] S.W. Hawking: *Black hole explosions?*, *Nature* **248**, 30 (1974).
- [262] S.W. Hawking: *Particle creation by black holes*, *Commun. Math. Phys.* **43**, 199 (1975); erratum in *Commun. Math. Phys.* **46**, 206 (1976).
- [263] S.W. Hawking: *A Brief History of Time: From the Big Bang to Black Holes*, Bantam Books, Toronto and New York (1988).
- [264] S.W. Hawking: *The Nature of Space and Time*, part of a series of six lectures with R. Penrose sponsored by Princeton University Press (1994); arXiv: hep-th/9409195.
- [265] S.W. Hawking and G.F.R. Ellis: *The large scale structure of space-time*, Cambridge Univ. Press, Cambridge (1973); DOI: 10.1017/CBO9780511524646.
- [266] S.W. Hawking and R. Penrose: *The singularities of gravitational collapse and cosmology*, *Proc. Roy. Soc. Lond. A* **314**, 529 (1970).
- [267] S.W. Hawking and R. Penrose: *The Nature of Space and Time*, Princeton Univ. Press, Princeton (2015).
- [268] S.A. Hayward: *General laws of black-hole dynamics*, *Phys. Rev. D* **49**, 6467 (1994); arXiv: gr-qc/9303006.
- [269] 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); arXiv: 1503.02172.
- [270] F. Hélein: *Multisymplectic formalism and the covariant phase space*, in *Variational Problems in Differential Geometry*, edited by R. Bielawski, K. Houston and M. Speight, Cambridge Univ. Press, Cambridge (2011), p. 94; DOI: 10.1017/CBO9780511863219.007; arXiv: 1106.2086.

- [271] C.A.R. Herdeiro and J.P.S. Lemos: *O buraco negro cinquenta anos depois: A gênese do nome*, *Gazeta de Fisica* **41**, 2 (2018); English translation in *The black hole fifty years after: Genesis of the name*, arXiv: [1811.06587](#).
- [272] W.C. Hernandez and C.W. Misner: *Observer Time as a Coordinate in Relativistic Spherical Hydrodynamics*, *Astrophys. J.* **143**, 452 (1966).
- [273] M. Heusler: *On the uniqueness of the Reissner-Nordström solution with electric and magnetic charge*, *Class. Quantum Grav.* **11**, L49 (1994).
- [274] M. Heusler: *Black Hole Uniqueness Theorems*, Cambridge Univ. Press, Cambridge (1996); DOI: [10.1017/CBO9780511661396](#).
- [275] M. Heusler: *On the uniqueness of the Papapetrou-Majumdar metric*, *Class. Quantum Grav.* **14**, L129 (1997); arXiv: [gr-qc/9607001](#).
- [276] D. Hilbert: *Die Grundlagen der Physik (Zweite Mitteilung)*, *Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-physikalische Klasse*, Jahr 1917, 53 (1917).
- [277] D. Hilbert: *Die Grundlagen der Physik II*, lectures given at Göttingen University in winter 1916-1917, lecture notes written by R. Bär; reprinted in [444], p. 162-307.
- [278] W.A. Hiscock, L.G. Williams, and D.M. Eardley: *Creation of particles by shell-focusing singularities*, *Phys. Rev. D* **26**, 751 (1982).
- [279] S. Hollands and A. Ishibashi: *Black hole uniqueness theorems in higher dimensional spacetimes*, *Class. Quantum Grav.* **29**, 163001 (2012); arXiv: [1206.1164](#).
- [280] S. Hollands, A. Ishibashi, and H.S. Reall: *A Stationary Black Hole Must be Axisymmetric in Effective Field Theory*, *Commun. Math. Phys.* **401**, 2757 (2023); arXiv: [2212.06554](#).
- [281] S. Hollands, A. Ishibashi, and R.M. Wald: *A Higher Dimensional Stationary Rotating Black Hole Must be Axisymmetric*, *Commun. Math. Phys.* **271**, 699 (2007); arXiv: [gr-qc/0605106](#).
- [282] S. Hollands and R.M. Wald: *Stability of Black Holes and Black Branes*, *Commun. Math. Phys.* **321**, 629 (2013); arXiv: [1201.0463](#).
- [283] S. Hollands, R.M. Wald and V.G. Zhang: *Entropy of dynamical black holes*, *Phys. Rev. D* **110**, 024070 (2024); arXiv: [2402.00818](#).
- [284] G.T. Horowitz: *Quantum States of Black Holes*, in *Black Holes and Relativistic Stars*, edited by R.M. Wald, Univ. Chicago Press, Chicago (1998), p. 241; arXiv: [gr-qc/9704072](#).
- [285] S. Husa and J. Winicour: *Asymmetric merger of black holes*, *Phys. Rev. D* **60**, 084019 (1999); arXiv: [gr-qc/9905039](#).
- [286] S. Hwang: *A Rigidity Theorem for Ricci Flat Metrics*, *Geometriae Dedicata* **71**, 5 (1998).
- [287] A.D. Ionescu and S. Klainerman: *Rigidity results in general relativity: A review*, *Surveys in Differential Geometry* **20**, 123 (2015); arXiv: [1501.01587](#).
- [288] W. Israel: *Event Horizons in Static Vacuum Space-Times*, *Phys. Rev.* **164**, 1776 (1967).
- [289] W. Israel: *Event horizons in static electrovac space-times*, *Commun. Math. Phys.* **8**, 245 (1968).
- [290] W. Israel: *Third Law of Black-Hole Dynamics: A Formulation and Proof*, *Phys. Rev. Lett.* **57**, 397 (1986).

- [291] V. Iyer and R.M. Wald: *Some properties of the Noether charge and a proposal for dynamical black hole entropy*, *Phys. Rev. D* **50**, 846 (1994); arXiv: [gr-qc/9403028](#).
- [292] T. Jacobson: *Introductory Lectures on Black Hole Thermodynamics*, lectures given at the University of Utrecht (1996); URL: <https://www.physics.umd.edu/grt/taj/776b/lectures.pdf>.
- [293] T. Jacobson, G. Kang, and R.C. Myers: *On black hole entropy*, *Phys. Rev. D* **49**, 6587 (1994); arXiv: [gr-qc/9312023](#).
- [294] T. Jacobson and Y.-A. Soong: *Slices of the Kerr ergosurface*, *Class. Quantum Grav.* **26**, 055014 (2009); arXiv: [0809.2369](#).
- [295] J. Jaffe: *Collapsing objects and the backward emission of light*, *Ann. Phys.* **55**, 374 (1969).
- [296] 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); DOI: [10.1142/S0218271811020366](#); arXiv: [1108.2408](#).
- [297] J.L. Jaramillo and E. Gourgoulhon: *Mass and Angular Momentum in General Relativity*, in *Mass and Motion in General Relativity*, edited by L. Blanchet, A. Spallicci and B. Whiting, *Fundamental Theories of Physics* **162**, 87, Springer, Dordrecht (2011); arXiv: [1001.5429](#).
- [298] M. Jaroszyński and A. Kurpiewski: *Optics near Kerr black holes: spectra of advection dominated accretion flows*, *Astron. Astrophys.* **326**, 419 (1997); arXiv: [astro-ph/9705044](#).
- [299] J.T. Jebsen: *Über die allgemeinen kugelsymmetrischen Lösungen der Einsteinschen Gravitationsgleichungen im Vakuum*, Arkiv för Matematik, Astronomi och Fysik **15**, Nr. 18, p. 1 (1921); English translation in *Gen. Relativ. Grav.* **37**, 2253 (2005).
- [300] T. Johannsen: *Sgr A* and general relativity*, *Class. Quantum Grav.* **33**, 113001 (2016); arXiv: [1512.03818](#).
- [301] 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); arXiv: [1005.1931](#).
- [302] N.V. Johansen and F. Ravndal: *On the discovery of Birkhoff's theorem*, *Gen. Relativ. Grav.* **38**, 537 (2006); arXiv: [physics/0508163](#).
- [303] 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); arXiv: [1907.04329](#).
- [304] M.D. Johnson et al: *The Black Hole Explorer: motivation and vision*, *Proc. of SPIE* **13092**, 130922D (2024); arXiv: [2406.12917](#).
- [305] D. Joyner and W. Stein: *Sage Tutorial*, CreateSpace (2014).
- [306] P. Kanti and E. Winstanley: *Hawking Radiation from Higher-Dimensional Black Holes*, in *Quantum Aspects of Black Holes*, edited by X. Calmet, *Fundamental Theories of Physics* **178**, 229, Springer, Cham (2015); arXiv: [1402.3952](#).
- [307] D. Kapec and A. Lupsasca: *Particle motion near high-spin black holes*, *Class. Quantum Grav.* **37**, 015006 (2020)
DOI: [10.1088/1361-6382/ab519e](#) arXiv: [1905.11406](#)

- [308] J. Katz: *A note on Komar's anomalous factor*, Class. Quantum Grav. **2**, 423 (1985)
DOI: [10.1088/0264-9381/2/3/018](https://doi.org/10.1088/0264-9381/2/3/018)
- [309] B.S. Kay and R.M. Wald: *Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on spacetimes with a bifurcate killing horizon*, Phys. Rep. **207**, 49 (1991)
DOI: [10.1016/0370-1573\(91\)90015-E](https://doi.org/10.1016/0370-1573(91)90015-E)
- [310] C. Kehle and R. Unger: *Gravitational collapse to extremal black holes and the third law of black hole thermodynamics*, J. Eur. Math. Soc., in press (2025); arXiv: [2211.15742](https://arxiv.org/abs/2211.15742).
- [311] 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](https://doi.org/10.1103/PhysRevLett.11.237)
- [312] 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](https://arxiv.org/abs/0706.1109)
- [313] R.P. Kerr and A. Schild: *A New Class of Vacuum Solutions of the Einstein Field Equations*, in *Atti del Convegno sulla Relatività 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](https://doi.org/10.1007/s10714-009-0857-z)
- [314] I. Khavkine: *Covariant phase space, constraints, gauge and the Peierls formula*, Int. J. Mod. Phys. A **29**, 1430009 (2014)
DOI: [10.1142/S0217751X14300099](https://doi.org/10.1142/S0217751X14300099) arXiv: [1402.1282](https://arxiv.org/abs/1402.1282)
- [315] A. King: *Supermassive Black Holes*, Cambridge Univ. Press, Cambridge (2023)
DOI: [10.1017/9781108768849](https://doi.org/10.1017/9781108768849)
- [316] O. Klein: *Einige Probleme der allgemeinen Relativitätstheorie*, in *Werner Heisenberg und die Physik unserer Zeit*, edited by F. Bopp, Vieweg+Teubner Verlag, Wiesbaden (1961), p. 58
DOI: [10.1007/978-3-663-05439-9_5](https://doi.org/10.1007/978-3-663-05439-9_5)
- [317] S. Kobayashi and K. Nomizu: *Foundations of Differential Geometry*, vol. 1, Interscience, New York (1963).
- [318] A. Komar: *Covariant Conservation Laws in General Relativity*, Phys. Rev. **113**, 934 (1959)
DOI: [10.1103/PhysRev.113.934](https://doi.org/10.1103/PhysRev.113.934)
- [319] A. Komar: *Asymptotic Covariant Conservation Laws for Gravitational Radiation*, Phys. Rev. **127**, 1411 (1962)
DOI: [10.1103/PhysRev.127.1411](https://doi.org/10.1103/PhysRev.127.1411)
- [320] E.-A. Kontou and K. Sanders: *Energy conditions in general relativity and quantum field theory*, Class. Quantum Grav. **37**, 193001 (2020)
DOI: [10.1088/1361-6382/ab8fcf](https://doi.org/10.1088/1361-6382/ab8fcf) arXiv: [2003.01815](https://arxiv.org/abs/2003.01815)
- [321] F. Kottler: *Über die physikalischen Grundlagen der Einsteinschen Gravitationstheorie*, Annalen der Physik **56**, 401 (1918)
DOI: [10.1002/andp.19183611402](https://doi.org/10.1002/andp.19183611402)
- [322] A. Krasiński: *Editor's Note: The Expanding Universe*, Gen. Relativ. Gravit. **29**, 637 (1997)
DOI: [10.1023/A:1018803604510](https://doi.org/10.1023/A:1018803604510)

- [323] A. Krasiński, E. Verdaguer and R.P. Kerr: *Editorial note to: R. P. Kerr and A. Schild, A new class of vacuum solutions of the Einstein field equations*, Gen. Relativ. Gravit. **41**, 2469 (2009)
DOI: [10.1007/s10714-009-0856-0](https://doi.org/10.1007/s10714-009-0856-0)
- [324] B. Krishnan: *Quasi-local Black Hole Horizons*, in *Springer Handbook of Spacetime*, edited by A. Ashtekar and V. Petkov, Springer, Berlin, Heidelberg (2014), p. 527
DOI: [10.1007/978-3-642-41992-8_25](https://doi.org/10.1007/978-3-642-41992-8_25) arXiv: [1303.4635](https://arxiv.org/abs/1303.4635)
- [325] M.D. Kruskal: *Maximal Extension of Schwarzschild Metric*, Phys. Rev. **119**, 1743 (1960)
DOI: [10.1103/PhysRev.119.1743](https://doi.org/10.1103/PhysRev.119.1743)
- [326] H.K. Kunduri and J. Lucietti: *Classification of Near-Horizon Geometries of Extremal Black Holes*, Living Rev. Relativ. **16**, 8 (2013)
DOI: [10.12942/lrr-2013-8](https://doi.org/10.12942/lrr-2013-8)
- [327] H.K. Kunduri and J. Lucietti: *No static bubbling spacetimes in higher dimensional Einstein-Maxwell theory*, Class. Quantum Grav. **35**, 054003 (2018)
DOI: [10.1088/1361-6382/aaa744](https://doi.org/10.1088/1361-6382/aaa744) arXiv: [1712.02668](https://arxiv.org/abs/1712.02668)
- [328] H.K. Kunduri, J. Lucietti and H.S. Reall: *Near-horizon symmetries of extremal black holes*, Class. Quantum Grav. **24**, 4169 (2007)
DOI: [10.1088/0264-9381/24/16/012](https://doi.org/10.1088/0264-9381/24/16/012) arXiv: [0705.4214](https://arxiv.org/abs/0705.4214)
- [329] Y. Kuroda: *Naked Singularities in the Vaidya Spacetime*, Prog. Theor. Phys. **72**, 63 (1984)
DOI: [10.1143/PTP.72.63](https://doi.org/10.1143/PTP.72.63)
- [330] J. Lafontaine: *An Introduction to Differential Manifolds*, Springer, Cham (2015)
DOI: [10.1007/978-3-319-20735-3](https://doi.org/10.1007/978-3-319-20735-3)
- [331] K. Lake: *Revisiting the Darmois and Lichnerowicz junction conditions*, Gen. Relativ. Gravit. **49**, 134 (2017)
DOI: [10.1007/s10714-017-2300-1](https://doi.org/10.1007/s10714-017-2300-1) arXiv: [1705.01090](https://arxiv.org/abs/1705.01090)
- [332] K. Lake and R.C. Roeder: *Note on the optical appearance of a star collapsing through its gravitational radius*, Astrophys. J. **232**, 277 (1979)
DOI: [10.1086/157286](https://doi.org/10.1086/157286)
- [333] K. Landsman: *Singularities, Black Holes, and Cosmic Censorship: A Tribute to Roger Penrose*, Foundations of Physics **51**, 42 (2021)
DOI: [10.1007/s10701-021-00432-1](https://doi.org/10.1007/s10701-021-00432-1) arXiv: [2101.02687](https://arxiv.org/abs/2101.02687)
- [334] K. Landsman: *Penrose's 1965 singularity theorem: from geodesic incompleteness to cosmic censorship*, Gen. Relat. Grav. **54**, 115 (2022)
DOI: [10.1007/s10714-022-02973-w](https://doi.org/10.1007/s10714-022-02973-w) arXiv: [2205.01680](https://arxiv.org/abs/2205.01680)
- [335] S. Lang: *Differential and Riemannian Manifolds*, Springer-Verlag, New York (1995).
- [336] D. Langlois: *Relativité générale*, Vuibert, Paris (2013).
- [337] D. Langlois: *Dark energy and modified gravity in degenerate higher-order scalar-tensor (DHOST) theories: A review*, Int. J. Mod. Phys. D **28**, 1942006 (2019)
DOI: [10.1142/S0218271819420069](https://doi.org/10.1142/S0218271819420069) arXiv: [1811.06271](https://arxiv.org/abs/1811.06271)

- [338] 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](https://doi.org/10.1103/PhysRevD.89.024041) arXiv: [1310.7499](https://arxiv.org/abs/1310.7499)
- [339] 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>
- [340] M. Le Bellac: *Quantum Physics*, Cambridge Univ. Press, Cambridge (2006)
DOI: [10.1017/CBO9780511616471](https://doi.org/10.1017/CBO9780511616471)
- [341] J. M. Lee: *Introduction to Topological Manifolds*, 2nd edition, Springer, New-York (2011)
DOI: [10.1007/978-1-4419-7940-7](https://doi.org/10.1007/978-1-4419-7940-7)
- [342] J. M. Lee: *Introduction to Smooth Manifolds*, 2nd edition, Springer, New-York (2013)
DOI: [10.1007/978-1-4419-9982-5](https://doi.org/10.1007/978-1-4419-9982-5)
- [343] J. M. Lee: *Introduction to Riemannian Manifolds*, 2nd edition, Springer, Cham (2018)
DOI: [10.1007/978-3-319-91755-9](https://doi.org/10.1007/978-3-319-91755-9)
- [344] J. Lee and R.M. Wald: *Local symmetries and constraints*, J. Math. Phys. **31**, 725 (1990)
DOI: [10.1063/1.528801](https://doi.org/10.1063/1.528801)
- [345] 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](https://doi.org/10.1023/A:1018855621348)
- [346] R. W. Lindquist, R. A. Schwartz, and C. W. Misner: *Vaidya's Radiating Schwarzschild Metric*, Phys. Rev. **137**, B1364 (1965)
DOI: [10.1103/PhysRev.137.B1364](https://doi.org/10.1103/PhysRev.137.B1364)
- [347] J.-P. Luminet: *Image of a spherical black hole with thin accretion disk*, Astron. Astrophys. **75**, 228 (1979).
- [348] J.-P. Luminet: *An Illustrated History of Black Hole Imaging: Personal Recollections (1972-2002)*, preprint (2019); arXiv: [1902.11196](https://arxiv.org/abs/1902.11196).
- [349] A. Lupsasca, A. Cárdenas-Avendaño, D.C.M. Palumbo, M.D. Johnson, S.E. Gralla, D.P. Marrone, P. Galison, P. Tiede, and L.Keeble: *The Black Hole Explorer: photon ring science, detection, and shape measurement*, Proc. of SPIE **13092**, 130926Q (2024); arXiv: [2406.09498](https://arxiv.org/abs/2406.09498)
- [350] D. Lynden-Bell: *Galactic Nuclei as Collapsed Old Quasars*, Nature **223**, 690 (1969)
DOI: [10.1038/223690a0](https://doi.org/10.1038/223690a0)
- [351] J.H. MacGibbon, B.J. Carr, and D.N. Page: *Do evaporating black holes form photospheres?*, Phys. Rev. D **78**, 064043 (2008)
DOI: [10.1103/PhysRevD.78.064043](https://doi.org/10.1103/PhysRevD.78.064043) arXiv: [0709.2380](https://arxiv.org/abs/0709.2380)
- [352] S.D. Majumdar: *A Class of Exact Solutions of Einstein's Field Equations*, Phys. Rev. **72**, 390 (1947)
DOI: [10.1103/PhysRev.72.390](https://doi.org/10.1103/PhysRev.72.390)

- [353] J. Maldacena and L. Susskind: *Cool horizons for entangled black holes*, Fortschritte der Physik **61**, 781 (2013)
DOI: [10.1002/prop.201300020](https://doi.org/10.1002/prop.201300020) arXiv: [1306.0533](https://arxiv.org/abs/1306.0533)
- [354] J.-A. Marck: *Journey to a black hole*, movie (1991)
<https://www.youtube.com/watch?v=50qop50ltrM>
see also <https://images.cnrs.fr/video/46>
- [355] J.-A. Marck: *Short-cut method of solution of geodesic equations for Schwarzschild black hole*, Class. Quantum Grav. **13**, 393 (1996)
DOI: [10.1088/0264-9381/13/3/007](https://doi.org/10.1088/0264-9381/13/3/007) arXiv: [gr-qc/9505010](https://arxiv.org/abs/gr-qc/9505010)
- [356] K. Martel and E. Poisson: *Regular coordinate systems for Schwarzschild and other spherical spacetimes*, Amer. J. Phys. **69**, 476 (2001)
DOI: [10.1119/1.1336836](https://doi.org/10.1119/1.1336836) arXiv: [gr-qc/0001069](https://arxiv.org/abs/gr-qc/0001069)
- [357] A.K.M. Masood-ul-Alam: *Uniqueness proof of static charged black holes revisited*, Class. Quantum Grav. **9**, L53 (1992)
DOI: [10.1088/0264-9381/9/5/001](https://doi.org/10.1088/0264-9381/9/5/001)
- [358] 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](https://doi.org/10.1126/science.270.5238.941)
- [359] N.E. Mavromatos and V.A. Mitsou: *Magnetic monopoles revisited: Models and searches at colliders and in the Cosmos*, Int. J. Mod. Phys. A **35**, 2030012 (2020)
DOI: [10.1142/S0217751X20300124](https://doi.org/10.1142/S0217751X20300124) arXiv: [2005.05100](https://arxiv.org/abs/2005.05100)
- [360] P.O. Mazur: *Proof of uniqueness of the Kerr-Newman black hole solution*, J. Phys. A: Math. Gen. **15**, 3173 (1982)
DOI: [10.1088/0305-4470/15/10/021](https://doi.org/10.1088/0305-4470/15/10/021)
- [361] P.O. Mazur: *Black hole uniqueness from a hidden symmetry of Einstein's gravity*, Gen. Relat. Grav. **16**, 211 (1984)
DOI: [10.1007/BF00762536](https://doi.org/10.1007/BF00762536)
- [362] P.O. Mazur: *A global identity for nonlinear σ -models*, Phys. Lett. A **100**, 341 (1984)
DOI: [10.1016/0375-9601\(84\)91084-3](https://doi.org/10.1016/0375-9601(84)91084-3)
- [363] P.O. Mazur: *Black Hole Uniqueness Theorems*, unpublished report (2001)
arXiv: [hep-th/0101012](https://arxiv.org/abs/hep-th/0101012)
- [364] 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](https://doi.org/10.1051/0004-6361/201628829) arXiv: [1608.05063](https://arxiv.org/abs/1608.05063)
- [365] J. W. Milnor: *On manifolds homeomorphic to the 7-sphere*, Ann. Math. **64**, 399 (1956).
- [366] H. Mineur: *La mécanique des masses variables. Le problème des deux corps*, Annales scientifiques de l'École Normale Supérieure, Ser. 3, **50**, 1 (1933)
DOI: [10.24033/asens.825](https://doi.org/10.24033/asens.825)
- [367] 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](https://doi.org/10.1103/PhysRevD.67.084027) arXiv: [gr-qc/0302075](https://arxiv.org/abs/gr-qc/0302075)

- [368] C.W. Misner: *Interpretation of Gravitational-Wave Observations*, Phys. Rev. Lett. **28**, 994 (1972)
DOI: [10.1103/PhysRevLett.28.994](https://doi.org/10.1103/PhysRevLett.28.994)
- [369] 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](https://doi.org/10.1103/PhysRev.136.B571)
- [370] C.W. Misner, K.S. Thorne, and J.A. Wheeler: *Gravitation*, Freeman, New York (1973); reprinted by Princeton University Press, Princeton (2017)
URL: <https://press.princeton.edu/books/ebook/9781400889099/gravitation>
- [371] 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](https://doi.org/10.1088/0264-9381/25/19/195015) arXiv: [0805.1451](https://arxiv.org/abs/0805.1451)
- [372] H. Müller zum Hagen, D.C. Robinson, and H.J. Seifert: *Black holes in static vacuum space-times*, Gen. Relat. Grav. **4**, 53 (1973)
DOI: [10.1007/BF00769760](https://doi.org/10.1007/BF00769760)
- [373] G. Muñoz: *Orbits of massless particles in the Schwarzschild metric: Exact solutions*, Amer. J. Phys. **82**, 564 (2014)
DOI: [10.1119/1.4866274](https://doi.org/10.1119/1.4866274)
- [374] R.C. Myers: *Higher-dimensional black holes in compactified space-times*, Phys. Rev. D **35**, 455 (1987)
DOI: [10.1103/PhysRevD.35.455](https://doi.org/10.1103/PhysRevD.35.455)
- [375] R.C. Myers and M.J. Perry: *Black holes in higher dimensional space-times*, Ann. Phys. **172**, 304 (1986)
DOI: [10.1016/0003-4916\(86\)90186-7](https://doi.org/10.1016/0003-4916(86)90186-7)
- [376] 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](https://doi.org/10.3847/2041-8213/ab518c) arXiv: [1910.02957](https://arxiv.org/abs/1910.02957)
- [377] V.V. Narlikar and P.C. Vaidya: *A Spherically Symmetrical Non-Static Electromagnetic Field*, Nature **159**, 642 (1947)
DOI: [10.1038/159642a0](https://doi.org/10.1038/159642a0)
- [378] J. Nash: *The Imbedding Problem for Riemannian Manifolds*, Ann. Math. **63**, 20 (1956)
DOI: [10.2307/1969989](https://doi.org/10.2307/1969989)
- [379] E.T. Newman, E. Couch, K. Chinnapared, A. Exton, A. Prakash, and R. Torrence: *Metric of a Rotating, Charged Mass*, J. Math. Phys. **6**, 918 (1965)
DOI: [10.1063/1.1704351](https://doi.org/10.1063/1.1704351)
- [380] R.P.A.C. Newman: *Topology and stability of marginal 2-surfaces*, Class. Quantum Grav. **4**, 277 (1987)
DOI: [10.1088/0264-9381/4/2/011](https://doi.org/10.1088/0264-9381/4/2/011)
- [381] J.-P. Nicolas: *The Conformal Approach to Asymptotic Analysis*, in *From Riemann to Differential Geometry and Relativity*, edited by L. Ji, A. Papadopoulos and S. Yamada, Springer, Cham (2017), p. 571
DOI: [10.1007/978-3-319-60039-0_19](https://doi.org/10.1007/978-3-319-60039-0_19) arXiv: [1508.02592](https://arxiv.org/abs/1508.02592)

- [382] B.C. Nolan: *Sectors of spherical homothetic collapse*, Class. Quantum Grav. **18**, 1651 (2001)
DOI: [10.1088/0264-9381/18/9/304](https://doi.org/10.1088/0264-9381/18/9/304) arXiv: [gr-qc/0010032](https://arxiv.org/abs/gr-qc/0010032)
- [383] B.C. Nolan: *Odd-parity perturbations of self-similar Vaidya spacetime*, Class. Quantum Grav. **24**, 177 (2007)
DOI: [10.1088/0264-9381/24/1/010](https://doi.org/10.1088/0264-9381/24/1/010) arXiv: [gr-qc/0608040](https://arxiv.org/abs/gr-qc/0608040)
- [384] G. Nordström: *On the Energy of the Gravitational Field in Einstein's Theory*, Verhandl. Koninkl. Ned. Akad. Wetenschap., Afdel. Natuurk. **20**, 1238 (1918); available at <https://archive.org/details/proceedingsofsec202koni/page/1238>
- [385] 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/abstract>
- [386] I.D. Novikov and V.P. Frolov: *Physics of Black Holes*, Kluwer, Dordrecht (1989)
DOI: [10.1007/978-94-017-2651-1](https://doi.org/10.1007/978-94-017-2651-1)
- [387] 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.
- [388] M. Nozawa, T. Shiromizu, K. Izumi and S. Yamada: *Divergence equations and uniqueness theorem of static black holes*, Class. Quantum Grav. **35**, 175009 (2018)
DOI: [10.1088/1361-6382/aad206](https://doi.org/10.1088/1361-6382/aad206) arXiv: [1805.11385](https://arxiv.org/abs/1805.11385)
- [389] B. O'Neill: *Semi-Riemannian Geometry, with Applications to Relativity*, Academic Press, New York (1983).
- [390] B. O'Neill: *The geometry of Kerr black holes*, A.K. Peters, Wellesley (MA) (1995); reprinted by Dover, Mineola (NY) (2014).
- [391] J.R. Oppenheimer and H. Snyder: *On Continued Gravitational Contraction*, Phys. Rev. **56**, 455 (1939)
DOI: [10.1103/PhysRev.56.455](https://doi.org/10.1103/PhysRev.56.455)
- [392] 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](https://doi.org/10.1088/0004-637X/742/2/84) arXiv: [1106.3689](https://arxiv.org/abs/1106.3689)
- [393] D.N. Page: *Particle emission rates from a black hole: Massless particles from an uncharged, nonrotating hole*, Phys. Rev. D **13**, 198 (1976)
DOI: [10.1103/PhysRevD.13.198](https://doi.org/10.1103/PhysRevD.13.198)
- [394] D.N. Page: *Hawking radiation and black hole thermodynamics*, New J. Phys. **7**, 203 (2005)
DOI: [10.1088/1367-2630/7/1/203](https://doi.org/10.1088/1367-2630/7/1/203) arXiv: [hep-th/0409024](https://arxiv.org/abs/hep-th/0409024)
- [395] 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](https://doi.org/10.1086/152990)
- [396] P. Painlevé: *La mécanique classique et la théorie de la relativité*, Comptes Rendus Acad. Sci. (Paris) **173**, 677 (1921)
URL: <http://www.bibnum.education.fr/physique/relativite/la-mecanique-classique-et-la-theorie-de-la-relativite>
- [397] E. Papantonopoulos (ed.): *Physics of Black Holes*, Springer, Berlin (2009).

- [398] A. Papapetrou: *A Static Solution of the Gravitational Field for an Arbitrary Charge-Distribution*, Proc. R. Irish Acad. A, **51**, 191 (1947)
URL: <https://www.jstor.org/stable/20488481>
- [399] A. Papapetrou : *Formation of a Singularity and Causality*, in *A Random Walk in Relativity and Cosmology*, edited by N. Dadhich, J. Krishna Rao, J.V. Narlikar and C.V. Vishveshwara, Wiley Eastern, New Dehli (1985), p. 184.
- [400] R. Parentani and P. Spindel: *Hawking radiation*, Scholarpedia **6**, 6958 (2011)
DOI: [10.4249/scholarpedia.6958](https://doi.org/10.4249/scholarpedia.6958)
- [401] A.W. Peet: *TASI Lectures on Black Holes in String Theory*, in *Strings, Branes and Gravity*, edited by J.A. Harvey, S. Kachru, and E. Silverstein, World Scientific, Singapore (2001), p. 353
DOI: [10.1142/9789812799630_0003](https://doi.org/10.1142/9789812799630_0003) arXiv: [hep-th/0008241](https://arxiv.org/abs/hep-th/0008241)
- [402] R. Penrose: *Asymptotic Properties of Fields and Space-Times*, Phys. Rev. Lett. **10**, 66 (1963)
DOI: [10.1103/PhysRevLett.10.66](https://doi.org/10.1103/PhysRevLett.10.66)
- [403] 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](https://doi.org/10.1007/s10714-010-1110-5)
- [404] R. Penrose: *Gravitational collapse and space-time singularities*, Phys. Rev. Lett. **14**, 57 (1965)
DOI: [10.1103/PhysRevLett.14.57](https://doi.org/10.1103/PhysRevLett.14.57)
- [405] R. Penrose: *Zero Rest-Mass Fields Including Gravitation: Asymptotic Behaviour*, Proc. Roy. Soc. Lond. A **284**, 159 (1965)
[http://www.jstor.org/stable/2415306](https://www.jstor.org/stable/2415306)
- [406] 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.
- [407] R. Penrose: *Gravitational Collapse: The Role of General Relativity*, Rivista del Nuovo Cimento, Numero Speciale **1**, 252 (1969); reprinted in Gen. Relativ. Gravit. **34**, 1141 (2002)
DOI: [10.1023/A:1016578408204](https://doi.org/10.1023/A:1016578408204)
- [408] R. Penrose and G.R. Floyd: *Extraction of Rotational Energy from a Black Hole*, Nature Phys. Sci. **229**, 177 (1971)
DOI: [10.1038/physci229177a0](https://doi.org/10.1038/physci229177a0)
- [409] A. Perez: *Black holes in loop quantum gravity*, Rep. Prog. Phys. **80**, 126901 (2017)
DOI: [10.1088/1361-6633/aa7e14](https://doi.org/10.1088/1361-6633/aa7e14) arXiv: [1703.09149](https://arxiv.org/abs/1703.09149)
- [410] 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](https://doi.org/10.7916/D8P84JVB)
- [411] V. Perlick: *Gravitational Lensing from a Spacetime Perspective*, Living Rev. Relativ. **7**, 9 (2004)
DOI: [10.12942/lrr-2004-9](https://doi.org/10.12942/lrr-2004-9)
- [412] P. Peter and J.-P. Uzan: *Primordial Cosmology*, Oxford Univ. Press, Oxford (2009).

- [413] J. Plebański and A. Krasiński: *An Introduction to General Relativity and Cosmology*, Cambridge Univ. Press, Cambridge (2006)
DOI: [10.1017/CBO9780511617676](https://doi.org/10.1017/CBO9780511617676)
- [414] M.A. Podurets: *Asymptotic behavior of the optical luminosity of a star in gravitational collapse in terms of general relativity theory* (in Russian), Astronomicheskii Zhurnal **41**, 1090 (1964); English translation in Soviet Astronomy **8**, 868 (1965)
URL: <https://ui.adsabs.harvard.edu/abs/1965SvA.....8..868P/abstract>
- [415] E. Poisson: *A Relativist's Toolkit, The Mathematics of Black-Hole Mechanics*, Cambridge Univ. Press, Cambridge (2004)
DOI: [10.1017/CBO9780511606601](https://doi.org/10.1017/CBO9780511606601)
- [416] E. Poisson and W. Israel: *Internal structure of black holes*, Phys. Rev. D **41**, 1796 (1990)
DOI: [10.1103/PhysRevD.41.1796](https://doi.org/10.1103/PhysRevD.41.1796)
- [417] E. Poisson and C.M. Will: *Gravity: Newtonian, Post-Newtonian, Relativistic*, Cambridge Univ. Press, Cambridge (2014)
DOI: [10.1017/CBO9781139507486](https://doi.org/10.1017/CBO9781139507486)
- [418] A. Pound and B. Wardell: *Black Hole Perturbation Theory and Gravitational Self-Force*, in *Handbook of Gravitational Wave Astronomy*, edited by C. Bambi, S. Katsanevas, and K.D. Kokkotas, Springer, Singapore (2021)
DOI: [10.1007/978-981-15-4702-7_38-1](https://doi.org/10.1007/978-981-15-4702-7_38-1) arXiv: [2101.04592](https://arxiv.org/abs/2101.04592)
- [419] W.H. Press and S.A. Teukolsky: *Floating Orbits, Superradiant Scattering and the Black-hole Bomb*, Nature **238**, 211 (1972)
DOI: [10.1038/238211a0](https://doi.org/10.1038/238211a0)
- [420] I. Rácz: *A simple proof of the recent generalizations of Hawking's black hole topology theorem*, Class. Quantum Grav. **25**, 162001 (2008)
DOI: [10.1088/0264-9381/25/16/162001](https://doi.org/10.1088/0264-9381/25/16/162001) arXiv: [0806.4373](https://arxiv.org/abs/0806.4373)
- [421] I. Rácz and R.M. Wald: *Global extensions of spacetimes describing asymptotic final states of black holes*, Class. Quantum Grav. **13**, 539 (1996)
DOI: [10.1088/0264-9381/13/3/017](https://doi.org/10.1088/0264-9381/13/3/017) arXiv: [gr-qc/9507055](https://arxiv.org/abs/gr-qc/9507055)
- [422] A. Rajantie: *Introduction to magnetic monopoles*, Contemporary Physics **53**, 195 (2012)
DOI: [10.1080/00107514.2012.685693](https://doi.org/10.1080/00107514.2012.685693) arXiv: [1204.3077](https://arxiv.org/abs/1204.3077)
- [423] P. Ramond and A. Le Tiec: *Multipolar particles in helically symmetric spacetimes*, Class. Quantum Grav. **38**, 135022 (2021)
DOI: [10.1088/1361-6382/abebef](https://doi.org/10.1088/1361-6382/abebef) arXiv: [2005.00602](https://arxiv.org/abs/2005.00602)
- [424] H.S. Reall: *Higher-Dimensional Black Holes*, in *General Relativity, Cosmology and Astrophysics; Perspectives 100 years after Einstein's stay in Prague*, edited by J. Bičák and T. Ledvinka, Springer, Cham (2014), p. 245
DOI: [10.1007/978-3-319-06349-2_12](https://doi.org/10.1007/978-3-319-06349-2_12)
- [425] H.S. 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>
- [426] H.S. Reall: *Third law of black hole mechanics for supersymmetric black holes and a quasilocal mass-charge inequality*, Phys. Rev. D **110**, 124059 (2024); arXiv: [2410.11956](https://arxiv.org/abs/2410.11956).

- [427] 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](https://doi.org/10.1088/0004-637X/742/2/83) arXiv: [1106.3688](https://arxiv.org/abs/1106.3688)
- [428] H. Reissner: *Über die Eigengravitation des elektrischen Feldes nach der Einsteinschen Theorie*, *Annalen der Physik* **50**, 106 (1916)
DOI: [10.1002/andp.19163550905](https://doi.org/10.1002/andp.19163550905)
- [429] C.S. Reynolds: *Observational Constraints on Black Hole Spin*, *Ann. Rev. Astron. Astrophys.* **59**, 117 (2021)
DOI: [10.1146/annurev-astro-112420-035022](https://doi.org/10.1146/annurev-astro-112420-035022) arXiv: [2011.08948](https://arxiv.org/abs/2011.08948)
- [430] L. Rezzolla and O. Zanotti: *Relativistic Hydrodynamics*, Oxford Univ. Press, Oxford (2013)
DOI: [10.1093/acprof:oso/9780198528906.001.0001](https://doi.org/10.1093/acprof:oso/9780198528906.001.0001)
- [431] A. Riazuelo: *Trous noirs et trous de vers*, *Pour la Science* **372**, 36 (2008).
- [432] 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](https://doi.org/10.1142/S0218271819500421) arXiv: [1511.06025](https://arxiv.org/abs/1511.06025)
- [433] 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](https://doi.org/10.1142/S0218271820501096) arXiv: [2008.04384](https://arxiv.org/abs/2008.04384)
- [434] 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](https://doi.org/10.1023/A:1015347106729)
- [435] D.C. Robinson: *Classification of black holes with electromagnetic fields*, *Phys. Rev. D* **10**, 458 (1974)
DOI: [10.1103/PhysRevD.10.458](https://doi.org/10.1103/PhysRevD.10.458)
- [436] D.C. Robinson: *Uniqueness of the Kerr Black Hole*, *Phys. Rev. Lett.* **34**, 905 (1975)
DOI: [10.1103/PhysRevLett.34.905](https://doi.org/10.1103/PhysRevLett.34.905)
- [437] D.C. Robinson: *A simple proof of the generalization of Israel's theorem*, *Gen. Relat. Grav.* **8**, 695 (1977)
DOI: [10.1007/BF00756322](https://doi.org/10.1007/BF00756322)
- [438] D.C. Robinson: *Four decades of black hole uniqueness theorems*, in *The Kerr spacetime*, edited by D.L. Wiltshire, M. Visser, and S.M. Scott, Cambridge Univ. Press, Cambridge (2009), p. 115; available at
https://nms.kcl.ac.uk/david.robinson/web_page/blackholes.pdf
- [439] 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](https://doi.org/10.1098/rspa.1962.0036)
- [440] C. Rovelli: *Black Hole Entropy from Loop Quantum Gravity*, *Phys. Rev. Lett.* **77**, 3288 (1996)
DOI: [10.1103/PhysRevLett.77.3288](https://doi.org/10.1103/PhysRevLett.77.3288) arXiv: [gr-qc/9603063](https://arxiv.org/abs/gr-qc/9603063)

- [441] C. Rovelli: *Quantum Gravity*, Cambridge Univ. Press, Cambridge (2004); available at <https://www.cpt.univ-mrs.fr/~rovelli/book.pdf>
DOI: 10.1017/CBO9780511755804
- [442] R. Ruffini and J.A. Wheeler: *Introducing the black hole*, Physics Today **24**, 30 (1971)
DOI: 10.1063/1.3022513
- [443] J. Santiago and M. Visser: *Tolman temperature gradients in a gravitational field*, Eur. J. Phys. **40**, 025604 (2019)
DOI: 10.1088/1361-6404/aaff1c arXiv: 1803.04106
- [444] T. Sauer and U. Majer (eds.): *David Hilbert's Lectures on the Foundations of Physics, 1915-1927*, Springer, Berlin (2009)
DOI: 10.1007/b12915
- [445] 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 arXiv: 0810.1767
- [446] W. Schmidt: *Celestial mechanics in Kerr spacetime*, Class. Quantum Grav. **19**, 2743 (2002)
DOI: 10.1088/0264-9381/19/10/314 arXiv: gr-qc/0202090
- [447] P. Schneider, J. Ehlers, and E.E. Falco: *Gravitational Lenses*, Springer-Verlag, Berlin (1992).
- [448] 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
- [449] O. Semerák: *Stationary frames in the Kerr field*, Gen. Relat. Grav. **25**, 1041 (1993)
DOI: 10.1007/BF00763554
- [450] J.M.M. Senovilla: *Singularity Theorems and Their Consequences*, Gen. Relat. Grav. **30**, 701 (1998)
DOI: 10.1023/A:1018801101244 arXiv: 1801.04912
- [451] J.M.M. Senovilla: *A critical appraisal of the singularity theorems*, Phil. Trans. R. Soc. A. **380**, 20210174 (2022)
DOI: 10.1098/rsta.2021.0174 arXiv: 2108.07296
- [452] J.M.M. Senovilla: *The influence of Penrose's singularity theorem in general relativity*, Gen. Relat. Grav. **54**, 151 (2022)
DOI: 10.1007/s10714-022-03038-8 arXiv: 2206.13925
- [453] 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
- [454] S.L. Shapiro: *Thermal radiation from stellar collapse to a black hole*, Phys. Rev. D **40**, 1858 (1989)
DOI: 10.1103/PhysRevD.40.1858
- [455] M. Shibata: *Numerical Relativity*, World Scientific, Singapore (2016); DOI: 10.1142/9692.

- [456] M. Siino: *Topological Appearance of Event Horizon: What Is the Topology of the Event Horizon That We Can See?*, Prog. Theor. Phys. **99**, 1 (1998)
DOI: [10.1143/PTP.99.1](https://doi.org/10.1143/PTP.99.1) arXiv: [gr-qc/9803099](https://arxiv.org/abs/gr-qc/9803099)
- [457] M. Siino: *Topology of event horizons* Phys. Rev. D **58**, 104016 (1998)
DOI: [10.1103/PhysRevD.58.104016](https://doi.org/10.1103/PhysRevD.58.104016) arXiv: [gr-qc/9701003](https://arxiv.org/abs/gr-qc/9701003)
- [458] L. Smarr: *Mass Formula for Kerr Black Holes*, Phys. Rev. Lett. **30**, 71 (1973); errata in Phys. Rev. Lett. **30**, 521 (1973)
DOI: [10.1103/PhysRevLett.30.71](https://doi.org/10.1103/PhysRevLett.30.71) and [10.1103/PhysRevLett.30.521](https://doi.org/10.1103/PhysRevLett.30.521) (errata)
- [459] A.A. Starobinsky: *Amplification of waves during reflection from a rotating “black hole”* (in Russian), Zhur. Eksp. Teor. Fiz. **64**, 48 (1973); English translation in Soviet Physics JETP **37**, 28 (1973)
URL: <http://jetp.ras.ru/cgi-bin/e/index/e/37/1/p28?a=list>
- [460] 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](https://doi.org/10.1103/PhysRevD.101.064007) arXiv: [1912.07609](https://arxiv.org/abs/1912.07609)
- [461] W. Stein and D. Joyner: *SAGE: System for Algebra and Geometry Experimentation*, Commun. Comput. Algebra **39**, 61 (2005).
- [462] B. Steinmüller, A.R. King, and J.-P. Lasota: *Radiating bodies and naked singularities*, Phys. Lett. A **51**, 191 (1975)
DOI: [10.1016/0375-9601\(75\)90525-3](https://doi.org/10.1016/0375-9601(75)90525-3)
- [463] N. Straumann: *General Relativity*, 2nd edition, Springer, Dordrecht (2013)
DOI: [10.1007/978-94-007-5410-2](https://doi.org/10.1007/978-94-007-5410-2)
- [464] A. Strominger and C. Vafa: *Microscopic origin of the Bekenstein-Hawking entropy*, Phys. Lett. B **379**, 99 (1996)
DOI: [10.1016/0370-2693\(96\)00345-0](https://doi.org/10.1016/0370-2693(96)00345-0) arXiv: [hep-th/9601029](https://arxiv.org/abs/hep-th/9601029)
- [465] 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>
- [466] 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](https://doi.org/10.1103/PhysRevD.46.1453)
- [467] D. Sudarsky and R.M. Wald: *Mass formulas for stationary Einstein-Yang-Mills black holes and a simple proof of two staticity theorems*, Phys. Rev. D **47**, R5209(R) (1993)
DOI: [10.1103/PhysRevD.47.R5209](https://doi.org/10.1103/PhysRevD.47.R5209) arXiv: [gr-qc/9305023](https://arxiv.org/abs/gr-qc/9305023)
- [468] J.L. Synge: *The Gravitational Field of a Particle*, Proc. Roy. Irish Acad. A **53**, 83 (1950)
<http://www.jstor.org/stable/20488511>
- [469] 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](https://doi.org/10.1093/mnras/131.3.463)
- [470] L.B. Szabados: *Quasi-Local Energy-Momentum and Angular Momentum in General Relativity*, Living Rev. Relativ. **12**, 4 (2009)
DOI: [10.12942/lrr-2009-4](https://doi.org/10.12942/lrr-2009-4)

- [471] 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](https://doi.org/10.1023/A:1020744914721)
- [472] F.R. Tangherlini: *Schwarzschild field in n dimensions and the dimensionality of space problem*, *Il Nuovo Cimento* **27**, 636 (1963)
DOI: [10.1007/BF02784569](https://doi.org/10.1007/BF02784569)
- [473] C. H. Taubes: *Gauge theory on asymptotically periodic 4-manifolds*, *J. Differential Geom.* **25**, 363 (1987).
- [474] E. Teo: *Spherical Photon Orbits Around a Kerr Black Hole*, *Gen. Relativ. Gravit.* **35**, 1909 (2003)
DOI: [10.1023/A:1026286607562](https://doi.org/10.1023/A:1026286607562)
- [475] S.A. Teukolsky: *The Kerr metric*, *Class. Quantum Grav.* **32** 124006 (2015)
DOI: [10.1088/0264-9381/32/12/124006](https://doi.org/10.1088/0264-9381/32/12/124006) arXiv: [1410.2130](https://arxiv.org/abs/1410.2130)
- [476] S.A. Teukolsky and W.H. Press: *Perturbations of a rotating black hole. III. Interaction of the hole with gravitational and electromagnetic radiation*, *Astrophys. J.* **193**, 443 (1974)
DOI: [10.1086/153180](https://doi.org/10.1086/153180)
- [477] K.S. Thorne: *Disk-Accretion onto a Black Hole. II. Evolution of the Hole*, *Astrophys. J.* **191**, 507 (1974)
DOI: [10.1086/152991](https://doi.org/10.1086/152991)
- [478] K.S. Thorne: *Black Holes and Time Warps*, Norton, New York (1994).
- [479] 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](https://doi.org/10.1093/mnras/198.2.339)
- [480] K.S. Thorne, W.H. Zurek, and R.H. Price: *The Thermal Atmosphere of a Black Hole*, in *Black holes: The Membrane Paradigm*, edited by K.S. Thorne, R.H. Price and D.A. MacDonald, Yale University Press, New Haven (1986), p. 280.
- [481] 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](https://doi.org/10.1023/A:1018891418565)
- [482] 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](https://arxiv.org/abs/gr-qc/9707012)
- [483] W.G. Unruh: *Notes on black-hole evaporation*, *Phys. Rev. D* **14**, 870 (1976)
DOI: [10.1103/PhysRevD.14.870](https://doi.org/10.1103/PhysRevD.14.870)
- [484] W.G. Unruh: *Universal coordinates for Schwarzschild black holes* (2014)
arXiv: [1401.3393](https://arxiv.org/abs/1401.3393)
- [485] P.C. Vaidya: *The External Field of a Radiating Star in General Relativity*, *Current Science* **12**, 183 (1943); reprinted in *Gen. Relativ. Gravit.* **31**, 119 (1999)
DOI: [10.1023/A:1018871522880](https://doi.org/10.1023/A:1018871522880)
- [486] P.C. Vaidya: *The Gravitational Field of a Radiating Star*, *Proc. Indian Acad. Sci. A* **33**, 264 (1951); reprinted in *Gen. Relativ. Gravit.* **31**, 121 (1999)
DOI: [10.1023/A:1018875606950](https://doi.org/10.1023/A:1018875606950)

- [487] P.C. Vaidya: '*Newtonian' Time in General Relativity*', *Nature* **171**, 260 (1953); reprinted in *Gen. Relativ. Gravit.* **31**, 137 (1999)
DOI: [10.1023/A:1018827723788](https://doi.org/10.1023/A:1018827723788)
- [488] K. Van Aelst: *Note on equatorial geodesics in circular spacetimes*, *Class. Quantum Grav.* **37**, 207001 (2020)
DOI: [10.1088/1361-6382/aba80c](https://doi.org/10.1088/1361-6382/aba80c)
- [489] K. Van Aelst, E. Gourgoulhon, P. Grandclément, and C. Charmousis: *Hairy rotating black holes in cubic Galileon theory*, *Class. Quantum Grav.* **37**, 035007 (2020)
DOI: [10.1088/1361-6382/ab6391](https://doi.org/10.1088/1361-6382/ab6391) arXiv: [1910.08451](https://arxiv.org/abs/1910.08451)
- [490] 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>
- [491] F.H. Vincent, E. Gourgoulhon, and J. Novak: *3+1 geodesic equation and images in numerical spacetimes*, *Class. Quantum Grav.* **29**, 245005 (2012)
DOI: [10.1088/0264-9381/29/24/245005](https://doi.org/10.1088/0264-9381/29/24/245005) arXiv: [1208.3927](https://arxiv.org/abs/1208.3927)
- [492] 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](https://doi.org/10.1088/0264-9381/28/22/225011) arXiv: [1109.4769](https://arxiv.org/abs/1109.4769)
- [493] 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](https://doi.org/10.1051/0004-6361/202037787) arXiv: [2002.09226](https://arxiv.org/abs/2002.09226)
- [494] C. V. Vishveshwara: *Generalization of the "Schwarzschild Surface" to Arbitrary Static and Stationary Metrics*, *J. Math. Phys.* **9**, 1319 (1968)
DOI: [10.1063/1.1664717](https://doi.org/10.1063/1.1664717)
- [495] M. Visser: *Hawking Radiation without Black Hole Entropy*, *Phys. Rev. Lett.* **80**, 3436 (1998)
DOI: [10.1103/PhysRevLett.80.3436](https://doi.org/10.1103/PhysRevLett.80.3436) arXiv: [gr-qc/9712016](https://arxiv.org/abs/gr-qc/9712016)
- [496] 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](https://arxiv.org/abs/0706.0622)
- [497] M. Visser and Z. Yan: *Properties of dynamical black hole entropy*, *J. High Energ. Phys.* **10(2024)**, 029 (2024)
DOI: [10.1007/JHEP10\(2024\)029](https://doi.org/10.1007/JHEP10(2024)029) arXiv: [2403.07140](https://arxiv.org/abs/2403.07140)
- [498] R.M. Wald: *General relativity*, University of Chicago Press, Chicago (1984).
- [499] R.M. Wald: *On identically closed forms locally constructed from a field*, *J. Math. Phys.* **31**, 2378 (1990).
- [500] R.M. Wald: *Black Holes and Thermodynamics*, in *Black Hole Physics*, edited by V. De Sabbata and Z. Zhang, Kluwer Academic Publisher, Dordrecht (1992), p. 55; DOI: [10.1007/978-94-011-2420-1_2](https://doi.org/10.1007/978-94-011-2420-1_2).
- [501] R.M. Wald: *Black hole entropy is the Noether charge*, *Phys. Rev. D* **48**, R3427 (1993); arXiv: [gr-qc/9307038](https://arxiv.org/abs/gr-qc/9307038).

- [502] R.M. Wald: *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*, University of Chicago Press, Chicago (1994)
URL: <https://press.uchicago.edu/ucp/books/book/chicago/Q/bo3684008.html>
- [503] R.M. Wald: “Nernst theorem” and black hole thermodynamics, *Phys. Rev. D* **56**, 6467 (1997); arXiv: [gr-qc/9704008](#).
- [504] R.M. Wald: The thermodynamics of black holes, *Living Rev. Relativ.* **4**, 6 (2001).
- [505] R.M. Wald: The Entropy of Black Holes, in *Black Holes Inside and Out 2024: visions for the future of black hole physics*, edited by L. Buoninfante, R. Carballo-Rubio, V. Cardoso, F. Di Filippo and A. Eichhorn (2024), p. 29; arXiv: [2410.14414](#).
- [506] R.M. Wald and A. Zoupas: General definition of “conserved quantities” in general relativity and other theories of gravity, *Phys. Rev. D* **61**, 084027 (2000); arXiv: [gr-qc/9911095](#).
- [507] M. Walker and R. Penrose: On Quadratic First Integrals of the Geodesic Equations for Type {22} Spacetimes, *Commun. Math. Phys.* **18**, 265 (1970).
- [508] 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); arXiv: [1802.06166](#).
- [509] A.C. Wall: Ten proofs of the generalized second law, *J. High Energ. Phys.* **06(2009)**, 021 (2009); arXiv: [0901.3865](#).
- [510] A.C. Wall: A Second Law for higher curvature gravity, *Int. J. Mod. Phys. D* **24**, 1544014 (2015); arXiv: [1504.08040](#).
- [511] D. Wallace: The case for black hole thermodynamics part I: Phenomenological thermodynamics, *Studies in History and Philosophy of Modern Physics* **64**, 52 (2018); arXiv: [1710.02724](#).
- [512] D. Wallace: The case for black hole thermodynamics part II: Statistical mechanics, *Studies in History and Philosophy of Modern Physics* **66**, 103 (2019); arXiv: [1710.02725](#).
- [513] 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); arXiv: [1304.7273](#).
- [514] N. Warburton, L. Barack, and N. Sago: Isofrequency pairing of geodesic orbits in Kerr geometry, *Phys. Rev. D* **87**, 084012 (2013); arXiv: [1301.3918](#).
- [515] B. Waugh and K. Lake: Double-null coordinates for the Vaidya metric, *Phys. Rev. D* **34**, 2978 (1986).
- [516] H. Weyl: Zur Gravitationstheorie, *Annalen der Physik* **54**, 117 (1917).
- [517] H. Weyl: Über die statischen kugelsymmetrischen Lösungen von Einsteins “kosmologischen” Gravitationsgleichungen, *Physikalische Zeitschrift* **20**, 31 (1919).
- [518] J.A. Wheeler: Feynman and Jacob Bekenstein, Web of Stories (1996); URL: <https://www.webofstories.com/play/john.wheeler/84>.
- [519] J.A. Wheeler: *Geons, Black Holes, and Quantum Foam: a Life in Physics*, Norton, New York (1998).
- [520] M. Wielgus et al.: Monitoring the Morphology of M87* in 2009-2017 with the Event Horizon Telescope, *Astrophys. J.* **901**, 67 (2020); arXiv: [2009.11842](#).

- [521] D.C. Wilkins: *Bound Geodesics in the Kerr Metric*, *Phys. Rev. D* **5**, 814 (1972).
- [522] J.W. York: *Kinematics and dynamics of general relativity*, in *Sources of gravitational radiation*, edited by L.L. Smarr, Cambridge University Press, Cambridge (1979), p. 83.
- [523] H. Yoshino, K. Takahashi, and K. Nakao: *How does a collapsing star look?*, *Phys. Rev. D* **100**, 084062 (2019); arXiv: [1908.04223](https://arxiv.org/abs/1908.04223).
- [524] A. Zaffaroni: *AdS black holes, holography and localization*, *Living Rev. Relativ.* **23**, 2 (2020); arXiv: [1902.07176](https://arxiv.org/abs/1902.07176).
- [525] Ya.B. Zeldovich: *Semiclosed Worlds in the General Theory of Relativity* (in Russian), *Zhur. Eksp. Teor. Fiz.* **43**, 1037 (1962); English translation in *Soviet Physics JETP* **16**, 732 (1963).
- [526] Ya.B. Zeldovich: *Generation of Waves by a Rotating Body* (in Russian), *Pisma Red. Zhur. Eksp. Teor. Fiz.* **14**, 270 (1971); English translation in *Soviet Physics JETP Letters* **14**, 180 (1971).
- [527] Ya.B. Zeldovich: *Amplification of Cylindrical Electromagnetic Waves Reflected from a Rotating Body* (in Russian), *Zhur. Eksp. Teor. Fiz.* **62**, 2076 (1972); English translation in *Soviet Physics JETP* **35**, 1085 (1972)
URL: <http://jetp.ras.ru/cgi-bin/e/index/e/35/6/p1085?a=list>
- [528] Ya.B. Zeldovich and I.D. Novikov: *Relativistic Astrophysics. I.* (in Russian), *Uspekhi Fizicheskikh Nauk* **84**, 377 (1964); English translation in *Soviet Physics Uspekhi* **7**, 763 (1965).
- [529] Ya.B. Zeldovich and I.D. Novikov: *Relativistic Astrophysics, vol. 1: Stars and Relativity*, University of Chicago Press, Chicago (1971).
- [530] Ya.B. Zeldovich and I.D. Novikov: *Relativistic Astrophysics, vol. 2: The Structure and Evolution of the Universe*, University of Chicago Press, Chicago (1983).
- [531] P. Zimmermann et al.: *Computational Mathematics with SageMath*, SIAM (2018); freely available from <https://www.sagemath.org/sagebook/>.
- [532] R.L. Znajek: *Black hole electrodynamics and the Carter tetrad*, *Month. Not. Roy. Astron. Soc.* **179**, 457 (1977).