

Mathematical Relativity

José Natário

Lisbon, 2020

Contents

Preface	3
Chapter 1. Preliminaries	5
1. Special relativity	5
2. Differential geometry: Mathematicians vs physicists	9
3. General relativity	13
4. Exercises	14
Chapter 2. Exact solutions	17
1. Minkowski spacetime	17
2. Penrose diagrams	18
3. The Schwarzschild solution	21
4. Friedmann-Lemaître-Robertson-Walker models	27
5. Matching	37
6. Oppenheimer-Snyder collapse	39
7. Exercises	42
Chapter 3. Causality	55
1. Past and future	55
2. Causality conditions	59
3. Exercises	60
Chapter 4. Singularity theorems	63
1. Geodesic congruences	63
2. Energy conditions	66
3. Conjugate points	67
4. Existence of maximizing geodesics	74
5. Hawking's singularity theorem	78
6. Penrose's singularity theorem	80
7. Exercises	86
Chapter 5. Cauchy problem	89
1. Divergence theorem	89
2. Klein-Gordon equation	91
3. Maxwell's equations: constraints and gauge	95
4. Einstein's equations	98
5. Constraint equations	102
6. Einstein equations with matter	103

7. Exercises	104
Chapter 6. Positive mass theorem	107
1. Komar mass	107
2. Field theory	112
3. Einstein-Hilbert action	118
4. Gravitational waves	122
5. ADM mass	124
6. Positive mass theorem	127
7. Penrose inequality	128
8. Exercises	131
Chapter 7. Black holes	135
1. The Kerr solution	135
2. Killing horizons and the zeroth law	137
3. Smarr's formula and the first law	140
4. Second law	143
5. Hawking radiation and black hole thermodynamics	146
6. Exercises	147
Appendix: Mathematical concepts for physicists	151
Topology	151
Metric spaces	152
Hopf-Rinow theorem	152
Differential forms	152
Lie derivative	154
Cartan structure equations	156
Bibliography	159

Preface

These lecture notes were written for a one-semester course in mathematical relativity aimed at mathematics and physics students, which has been taught at Instituto Superior Técnico (Universidade de Lisboa) since 2010. They are not meant as an introduction to general relativity, but rather as a complementary, more advanced text, much like Part II of Wald's textbook [Wal84], on which they are loosely based. It is assumed that the reader is familiar at least with special relativity, and has taken a course either in Riemannian geometry (typically the mathematics students) or in general relativity (typically the physics students). In other words, the reader is expected to be proficient in (some version of) differential geometry and to be acquainted with the basic principles of relativity.

I thank the many colleagues and students who read this text, or parts of it, for their valuable comments and suggestions. Special thanks are due to my colleague and friend Pedro Girão.

CHAPTER 1

Preliminaries

In this initial chapter we give a very short introduction to special and general relativity for mathematicians. In particular, we relate the index-free differential geometry notation used in Mathematics (e.g. [O’N83, dC93, Boo03, GN14]) to the index notation used in Physics (e.g. [MTW73, Wal84, HE95]). As an exercise in index gymnastics, we derive the contracted Bianchi identities.

1. Special relativity

Consider an inertial frame S' moving with velocity v with respect to another inertial frame S along their common x -axis (Figure 1). According to classical mechanics, coordinate x' of a point P on the frame S' is related to its x coordinate on the frame S by

$$x' = x - vt.$$

Moreover, a clock in S' initially synchronized with a clock in S is assumed to keep the same time:

$$t' = t.$$

Thus the spacetime coordinates of events are related by a so-called **Galileo transformation**

$$\begin{cases} x' = x - vt \\ t' = t \end{cases}.$$

If the point P is moving, its velocity in S' is related to its velocity in S by

$$\frac{dx'}{dt'} = \frac{dx - vdt}{dt} = \frac{dx}{dt} - v.$$

This is in conflict with the experimental fact that the speed of light is the same in every inertial frame, indicating that classical mechanics is not correct. Einstein solved this problem in 1905 by replacing the Galileo transformation by the so-called **Lorentz transformation**:

$$\begin{cases} x' = \gamma(x - vt) \\ t' = \gamma(t - vx) \end{cases}.$$

Here

$$\gamma = \frac{1}{\sqrt{1 - v^2}},$$

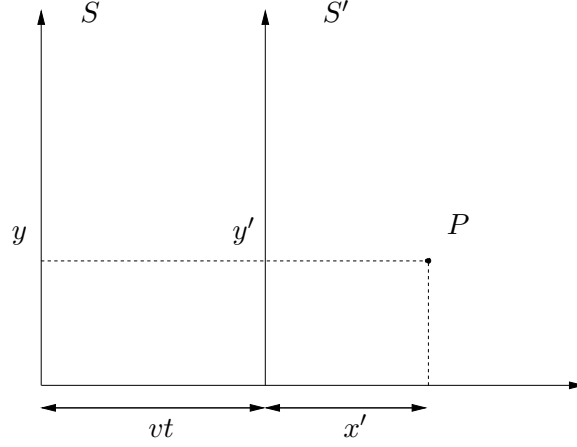


FIGURE 1. Galileo transformation.

and we are using units such that the speed of light is $c = 1$ (for example measuring time in years and distance in light-years). Note that if $|v|$ is much smaller than the speed of light, $|v| \ll 1$, then $\gamma \simeq 1$, and we retrieve the Galileo transformation (assuming $|v \frac{dx}{dt}| \ll 1$).

Under the Lorentz transformation velocities transform as

$$\frac{dx'}{dt'} = \frac{\gamma(dx - vdt)}{\gamma(dt - vdx)} = \frac{\frac{dx}{dt} - v}{1 - v\frac{dx}{dt}}.$$

In particular,

$$\frac{dx}{dt} = 1 \Rightarrow \frac{dx'}{dt'} = \frac{1 - v}{1 - v} = 1,$$

that is, the speed of light is the same in the two inertial frames.

In 1908, Minkowski noticed that

$$-(dt')^2 + (dx')^2 = -\gamma^2(dt - vdx)^2 + \gamma^2(dx - vdt)^2 = -dt^2 + dx^2,$$

that is, the Lorentz transformations could be seen as isometries of \mathbb{R}^4 with the indefinite metric

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 = -dt \otimes dt + dx \otimes dx + dy \otimes dy + dz \otimes dz.$$

DEFINITION 1.1. *The pseudo-Riemannian manifold $(\mathbb{R}^4, ds^2) \equiv (\mathbb{R}^4, \langle \cdot, \cdot \rangle)$ is called the **Minkowski spacetime**.*

Note that the set of vectors with zero square form a cone (the so-called **light cone**):

$$\langle v, v \rangle = 0 \Leftrightarrow -(v^0)^2 + (v^1)^2 + (v^2)^2 + (v^3)^2 = 0.$$

DEFINITION 1.2. *A vector $v \in \mathbb{R}^4$ is said to be:*

- (1) **timelike** if $\langle v, v \rangle < 0$;
- (2) **spacelike** if $\langle v, v \rangle > 0$;
- (3) **lightlike**, or **null**, if $\langle v, v \rangle = 0$.

- (4) **causal** if it is timelike or null;
 (5) **future-pointing** if it is causal and $\langle v, \frac{\partial}{\partial t} \rangle < 0$.

The same classification applies to (smooth) curves $c : [a, b] \rightarrow \mathbb{R}^4$ according to its tangent vector.

timelike future-pointing vector

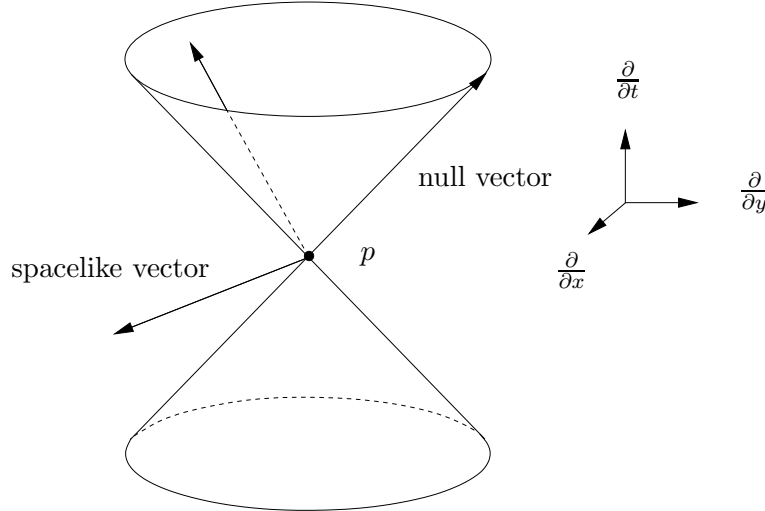


FIGURE 2. Minkowski geometry (traditionally represented with the t -axis pointing upwards).

The length $|\langle v, v \rangle|^{\frac{1}{2}}$ of a timelike (resp. spacelike) vector $v \in \mathbb{R}^4$ represents the time (resp. distance) measured between two events p and $p + v$ in the inertial frame where these events happen in the same location (resp. are simultaneous). If $c : [a, b] \rightarrow \mathbb{R}^4$ is a timelike curve then its length

$$\tau(c) = \int_a^b |\langle \dot{c}(s), \dot{c}(s) \rangle|^{\frac{1}{2}} ds$$

represents the **proper time** measured by the particle between events $c(a)$ and $c(b)$. We have:

PROPOSITION 1.3. (Twin paradox) *Of all timelike curves connecting two events $p, q \in \mathbb{R}^4$, the curve with **maximal** length is the line segment (representing inertial motion).*

PROOF. We may assume $p = (0, 0, 0, 0)$ and $q = (T, 0, 0, 0)$ on some inertial frame, and parameterize any timelike curve connecting p to q by the time coordinate:

$$c(t) = (t, x(t), y(t), z(t)).$$

Therefore

$$\tau(c) = \int_0^T |-1 + \dot{x}^2 + \dot{y}^2 + \dot{z}^2|^{\frac{1}{2}} dt = \int_0^T (1 - \dot{x}^2 - \dot{y}^2 - \dot{z}^2)^{\frac{1}{2}} dt \leq \int_0^T 1 dt = T.$$

□

Most problems in special relativity can be recast as questions about the geometry of the Minkowski spacetime.

PROPOSITION 1.4. (Doppler effect) *An observer moving with velocity v away from a source of light of period T measures the period to be*

$$T' = T \sqrt{\frac{1+v}{1-v}}.$$

PROOF. Figure 3 represents two light signals emitted by an observer at rest at $x = 0$ with a time difference T . These signals are detected by an observer moving with velocity v , who measures a time difference T' between them. Now, if the first signal is emitted at $t = t_0$, its history is the line $t = t_0 + x$. Consequently, the moving observer detects the signal at the event with coordinates

$$\begin{cases} t = t_0 + x \\ x = vt \end{cases} \Leftrightarrow \begin{cases} t = \frac{t_0}{1-v} \\ x = \frac{vt_0}{1-v} \end{cases}.$$

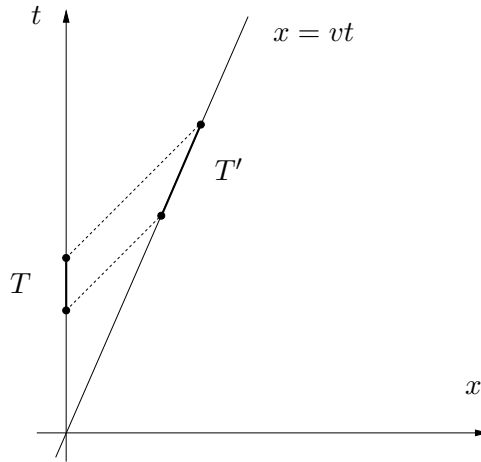


FIGURE 3. Doppler effect.

Similarly, the second light signal is emitted at $t = t_0 + T$, its history is the line $t = t_0 + T + x$, and it is detected by the moving observer at the event with coordinates

$$\begin{cases} t = \frac{t_0 + T}{1-v} \\ x = \frac{v(t_0 + T)}{1-v} \end{cases}.$$

Therefore the time difference between the signals as measured by the moving observer is

$$\begin{aligned} T' &= \sqrt{\left(\frac{t_0 + T}{1 - v} - \frac{t_0}{1 - v}\right)^2 - \left(\frac{v(t_0 + T)}{1 - v} - \frac{vt_0}{1 - v}\right)^2} \\ &= \sqrt{\frac{T^2}{(1 - v)^2} - \frac{v^2 T^2}{(1 - v)^2}} = T \sqrt{\frac{1 - v^2}{(1 - v)^2}} = T \sqrt{\frac{1 + v}{1 - v}}. \end{aligned}$$

□

In particular, two observers at rest in an inertial frame measure the same frequency for a light signal (Figure 4). However, because the gravitational field couples to all forms of energy (as $E = mc^2$), one expects that a photon climbing in a gravitational field to lose energy, hence frequency. In 1912, Einstein realized that this could be modelled by considering curved space-time geometries, so that equal line segments in a (flat) spacetime diagram do not necessarily correspond to the same length.

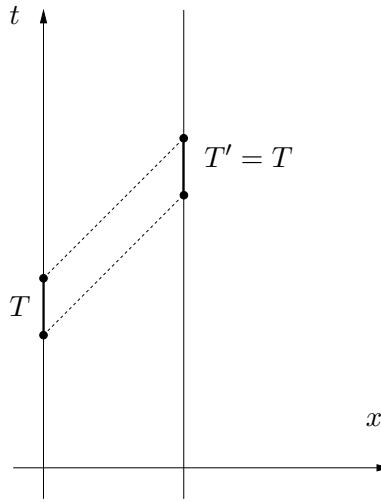


FIGURE 4. Minkowski geometry is incompatible with the gravitational redshift.

2. Differential geometry: Mathematicians vs physicists

Einstein's idea to incorporate gravitation into relativity was to replace the Minkowski spacetime $(\mathbb{R}^4, \langle \cdot, \cdot \rangle)$ by a curved four-dimensional **Lorentzian manifold** $(M, g) \equiv (M, \langle \cdot, \cdot \rangle)$. Here g is a **Lorentzian metric**, that is, a symmetric 2-tensor field such that at each tangent space $g = \text{diag}(-1, 1, 1, 1)$ in an appropriate basis. Just like in Riemannian geometry, g determines a

Levi-Civita connection, the unique connection ∇ which is symmetric and compatible with g :

$$\begin{aligned}\nabla_X Y - \nabla_Y X &= [X, Y]; \\ X \cdot \langle Y, Z \rangle &= \langle \nabla_X Y, Z \rangle + \langle Y, \nabla_X Z \rangle,\end{aligned}$$

for all vector fields X, Y, Z . The **curvature** of this connection is then given by the operator

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z.$$

The formulas above were written using the abstract notation usually employed by mathematicians. It is often very useful (especially when dealing with contractions) to use the more explicit notation usually adopted by physicists, which emphasizes the indices of the various objects when written in local coordinates:

Object	Mathematicians	Physicists
Vector field	X	X^μ
Tensor product	$X \otimes Y$	$X^\mu Y^\nu$
Metric	$g \equiv \langle \cdot, \cdot \rangle$	$g_{\mu\nu}$
Inner product	$g(X, Y) \equiv \langle X, Y \rangle$	$g_{\mu\nu} X^\mu Y^\nu$
Associated covector	$X^\sharp \equiv g(X, \cdot)$	$X_\nu \equiv g_{\mu\nu} X^\mu$
Covariant derivative	$\nabla_X Y$	$X^\mu \nabla_\mu Y^\nu$
Covariant derivative tensor	∇X	$\nabla_\mu X^\nu \equiv \partial_\mu X^\nu + \Gamma_{\mu\alpha}^\nu X^\alpha$

Here $\Gamma_{\mu\nu}^\alpha$ are the **Christoffel symbols** of the Levi-Civita connection; they can be computed from the components $g_{\mu\nu}$ of the metric tensor by the formula

$$\Gamma_{\mu\nu}^\alpha = \frac{1}{2} g^{\alpha\beta} (\partial_\mu g_{\nu\beta} + \partial_\nu g_{\mu\beta} - \partial_\beta g_{\mu\nu}),$$

and in turn be used to compute the components of the Riemann curvature tensor:

$$R_{\alpha\beta}{}^\mu{}_\nu = dx^\mu (R(\partial_\alpha, \partial_\beta)\partial_\nu) = \partial_\alpha \Gamma_{\beta\nu}^\mu - \partial_\beta \Gamma_{\alpha\nu}^\mu + \Gamma_{\alpha\gamma}^\mu \Gamma_{\beta\nu}^\gamma - \Gamma_{\beta\gamma}^\mu \Gamma_{\alpha\nu}^\gamma.$$

The covariant derivative tensor of a vector field X , not always emphasized in differential geometry courses for mathematicians, is simply the $(1, 1)$ -tensor field defined by

$$\nabla X(Y) = \nabla_Y X.$$

Also not always emphasized in differential geometry courses for mathematicians is the fact that any connection can be naturally extended to act on tensor fields (via the Leibnitz rule). For instance, if ω is a covector field (1-form) then one defines

$$(\nabla_X \omega)(Y) = X \cdot [\omega(Y)] - \omega(\nabla_X Y).$$

In local coordinates, this is

$$\begin{aligned} (X^\mu \nabla_\mu \omega_\nu) Y^\nu &= X^\mu \partial_\mu (\omega_\nu Y^\nu) - \omega_\nu (X^\mu \nabla_\mu Y^\nu) \\ &= X^\mu (\partial_\mu \omega_\nu) Y^\nu + X^\mu \omega_\nu \partial_\mu Y^\nu - \omega_\nu (X^\mu \partial_\mu Y^\nu + X^\mu \Gamma_{\mu\alpha}^\nu Y^\alpha) \\ &= (\partial_\mu \omega_\nu - \Gamma_{\mu\nu}^\alpha \omega_\alpha) X^\mu Y^\nu, \end{aligned}$$

that is,

$$\nabla_\mu \omega_\nu = \partial_\mu \omega_\nu - \Gamma_{\mu\nu}^\alpha \omega_\alpha.$$

The generalization for higher rank tensors is obvious: for instance, if T is a $(2, 1)$ -tensor then

$$\nabla_\alpha T_{\mu\nu}^\beta = \partial_\alpha T_{\mu\nu}^\beta + \Gamma_{\alpha\gamma}^\beta T_{\mu\nu}^\gamma - \Gamma_{\alpha\mu}^\gamma T_{\gamma\nu}^\beta - \Gamma_{\alpha\nu}^\gamma T_{\mu\gamma}^\beta.$$

Note that the condition of compatibility of the Levi-Civita connection with the metric is simply

$$\nabla g = 0.$$

In particular, the operations of raising and lowering indices commute with covariant differentiation.

As an exercise in index gymnastics, we will now derive a series of identities involving the Riemann curvature tensor. We start by rewriting its definition in the notation of the physicists:

$$\begin{aligned} R_{\alpha\beta}{}^\mu{}_\nu X^\alpha Y^\beta Z^\nu &= \\ &= X^\alpha \nabla_\alpha (Y^\beta \nabla_\beta Z^\mu) - Y^\alpha \nabla_\alpha (X^\beta \nabla_\beta Z^\mu) - (X^\alpha \nabla_\alpha Y^\beta - Y^\alpha \nabla_\alpha X^\beta) \nabla_\beta Z^\mu \\ &= (X^\alpha \nabla_\alpha Y^\beta) (\nabla_\beta Z^\mu) + X^\alpha Y^\beta \nabla_\alpha \nabla_\beta Z^\mu - (Y^\alpha \nabla_\alpha X^\beta) (\nabla_\beta Z^\mu) \\ &\quad - Y^\alpha X^\beta \nabla_\alpha \nabla_\beta Z^\mu - (X^\alpha \nabla_\alpha Y^\beta) \nabla_\beta Z^\mu + (Y^\alpha \nabla_\alpha X^\beta) \nabla_\beta Z^\mu \\ &= X^\alpha Y^\beta (\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha) Z^\mu. \end{aligned}$$

In other words,

$$R_{\alpha\beta}{}^\mu{}_\nu Z^\nu = (\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha) Z^\mu,$$

or, equivalently,

$$(1) \quad 2\nabla_{[\alpha} \nabla_{\beta]} Z_\mu = R_{\alpha\beta\mu\nu} Z^\nu,$$

where the square brackets indicate anti-symmetrization¹. This is readily generalized for arbitrary tensors: from

$$\begin{aligned} 2\nabla_{[\alpha} \nabla_{\beta]} (Z_\mu W_\nu) &= (2\nabla_{[\alpha} \nabla_{\beta]} Z_\mu) W_\nu + (2\nabla_{[\alpha} \nabla_{\beta]} W_\nu) Z_\mu \\ &= R_{\alpha\beta\mu\sigma} Z^\sigma W_\nu + R_{\alpha\beta\nu\sigma} W^\sigma Z_\mu \end{aligned}$$

one readily concludes that

$$2\nabla_{[\alpha} \nabla_{\beta]} T_{\mu\nu} = R_{\alpha\beta\mu\sigma} T^\sigma{}_\nu + R_{\alpha\beta\nu\sigma} T_\mu{}^\sigma.$$

Let us choose

$$Z_\mu = \nabla_\mu f \equiv \partial_\mu f$$

¹Thus $T_{[\alpha\beta]} = \frac{1}{2}(T_{\alpha\beta} - T_{\beta\alpha})$, $T_{[\alpha\beta\gamma]} = \frac{1}{6}(T_{\alpha\beta\gamma} + T_{\beta\gamma\alpha} + T_{\gamma\alpha\beta} - T_{\beta\alpha\gamma} - T_{\alpha\gamma\beta} - T_{\gamma\beta\alpha})$, etc.

in equation (1). We obtain

$$R_{[\alpha\beta\mu]\nu}Z^\nu = 2\nabla_{[[\alpha}\nabla_{\beta]}Z_{\mu]} = 2\nabla_{[\alpha}\nabla_{[\beta}Z_{\mu]}] = 0,$$

because

$$\nabla_{[\mu}Z_{\nu]} = \partial_{[\mu}Z_{\nu]} - \Gamma_{[\mu\nu]}^\alpha Z_\alpha = \partial_{[\mu}\partial_{\nu]}f = 0.$$

Since we can choose Z arbitrarily at a given point, it follows that

$$R_{[\alpha\beta\mu]\nu} = 0 \Leftrightarrow R_{\alpha\beta\mu\nu} + R_{\beta\mu\alpha\nu} + R_{\mu\alpha\beta\nu} = 0.$$

This is the so-called **first Bianchi identity**, and is key for obtaining the full set of symmetries of the Riemann curvature tensor:

$$R_{\alpha\beta\mu\nu} = -R_{\beta\alpha\mu\nu} = -R_{\alpha\beta\nu\mu} = R_{\mu\nu\alpha\beta}.$$

In the notation of the mathematicians, it is written as

$$R(X, Y)Z + R(Y, Z)X + R(Z, X)Y = 0$$

for all vector fields X, Y, Z .

Let us now take the covariant derivative of equation (1):

$$\nabla_\gamma R_{\alpha\beta\mu\nu}Z^\nu + R_{\alpha\beta\mu\nu}\nabla_\gamma Z^\nu = 2\nabla_\gamma\nabla_{[\alpha}\nabla_{\beta]}Z_\mu.$$

At any given point we can choose Z such that

$$\nabla_\gamma Z^\nu \equiv \partial_\gamma Z^\nu + \Gamma_{\gamma\delta}^\nu Z^\delta = 0.$$

Assuming this, we then obtain²

$$\begin{aligned} \nabla_{[\gamma}R_{\alpha\beta]\mu\nu}Z^\nu &= 2\nabla_{[\gamma}\nabla_{[\alpha}\nabla_{\beta]}]Z_\mu = 2\nabla_{[[\gamma}\nabla_{\alpha]}\nabla_{\beta]}Z_\mu \\ &= R_{[\gamma\alpha\beta]\delta}\nabla^\delta Z_\mu + R_{[\gamma\alpha|\mu\delta]}\nabla_{\beta]}Z^\delta = 0. \end{aligned}$$

Since we can choose Z arbitrarily at a given point, it follows that

$$(2) \quad \nabla_{[\alpha}R_{\beta\gamma]\mu\nu} = 0 \Leftrightarrow \nabla_\alpha R_{\beta\gamma\mu\nu} + \nabla_\beta R_{\gamma\alpha\mu\nu} + \nabla_\gamma R_{\alpha\beta\mu\nu} = 0$$

This is the so-called **second Bianchi identity**. In the notation of the mathematicians, it is written as

$$\nabla R(X, Y, Z, \cdot, \cdot) + \nabla R(Y, Z, X, \cdot, \cdot) + \nabla R(Z, X, Y, \cdot, \cdot) = 0$$

for all vector fields X, Y, Z .

Recall that the Riemann curvature tensor has only one independent contraction, called the **Ricci tensor**:

$$R_{\mu\nu} = R_{\alpha\mu}{}^\alpha{}_\nu.$$

The trace of the Ricci tensor, in turn, is known as the **scalar curvature**:

$$R = g^{\mu\nu}R_{\mu\nu}.$$

These quantities satisfy the so-called **contracted Bianchi identity**, which is obtained from (2) by contracting the pairs of indices (β, μ) and (γ, ν) :

$$\nabla_\alpha R - \nabla^\beta R_{\alpha\beta} - \nabla^\gamma R_{\alpha\gamma} = 0 \Leftrightarrow \nabla^\beta R_{\alpha\beta} - \frac{1}{2}\nabla_\alpha R = 0 \Leftrightarrow \nabla^\beta \left(R_{\alpha\beta} - \frac{1}{2}Rg_{\alpha\beta} \right) = 0.$$

²In the formula below the indices between vertical bars are not anti-symmetrized.

The contracted Bianchi identity is equivalent to the statement that the **Einstein tensor**

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$$

is divergenceless:

$$\nabla^\mu G_{\mu\nu} = 0.$$

3. General relativity

Newtonian gravity is described by a scalar function ϕ , called the **gravitational potential**. The equation of motion for a free-falling particle of mass m in Cartesian coordinates is

$$m\frac{d^2x^i}{dt^2} = -m\partial_i\phi \Leftrightarrow \frac{d^2x^i}{dt^2} = -\partial_i\phi.$$

Note that all free-falling particles describe the same trajectories (an observation dating back to Galileo). The gravitational potential is determined from the matter mass density ρ by the **Poisson equation**

$$\Delta\phi = 4\pi\rho$$

(using units such that Newton's gravitational constant is $G = 1$; this choice, together with $c = 1$, defines the so-called **geometrized units**, where lengths, time intervals and masses all have the same dimensions).

To implement his idea of describing gravity via a curved four-dimensional Lorentzian manifold (M, g) , Einstein had to specify (*i*) how free-falling particles would move on this manifold, and (*ii*) how to determine the curved metric g . Since free particles move along straight lines in the Minkowski spacetime, Einstein proposed that free falling particles should move along timelike geodesics. In other words, he suggested replacing the Newtonian equation of motion by the geodesic equation

$$\ddot{x}^\mu + \Gamma_{\alpha\beta}^\mu \dot{x}^\alpha \dot{x}^\beta = 0.$$

Moreover, Einstein knew that it is possible to define the **energy-momentum tensor** $T_{\mu\nu}$ of the matter content of the Minkowski spacetime, so that the conservation of energy and momentum is equivalent to the vanishing of its divergence:

$$\nabla^\mu T_{\mu\nu} = 0.$$

This inspired Einstein to propose that g should satisfy the so-called **Einstein field equations**:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}.$$

Here Λ is a constant, known as the **cosmological constant**. Note that the Einstein field equations imply, via the contracted Bianchi identity, that the energy-momentum tensor is divergenceless.

As a simple example, we consider a pressureless perfect fluid, known as **dust**. Its energy-momentum tensor is

$$T_{\mu\nu} = \rho U_\mu U_\nu,$$

where ρ is the dust rest density and U is a unit timelike vector field tangent to the histories of the dust particles. The equations of motion for the dust can be found from

$$\begin{aligned}\nabla^\mu T_{\mu\nu} = 0 &\Leftrightarrow [\nabla^\mu(\rho U_\mu)]U_\nu + \rho U_\mu \nabla^\mu U_\nu = 0 \\ &\Leftrightarrow \operatorname{div}(\rho U)U + \rho \nabla_U U = 0.\end{aligned}$$

Since U and $\nabla_U U$ are orthogonal (because $\langle U, U \rangle = -1$), we find

$$\begin{cases} \operatorname{div}(\rho U) = 0 \\ \nabla_U U = 0 \end{cases}$$

in the support of ρ . These are, respectively, the equation of conservation of mass and the geodesic equation. Thus the fact that free-falling particles move along geodesics can be seen as a consequence of the Einstein field equations (at least in this model).

4. Exercises

- (1) **Twin paradox:** Two twins, Alice and Bob, are separated on their 20th birthday. While Alice remains on Earth (which is an inertial frame to a very good approximation), Bob departs at 80% of the speed of light towards Planet X, 8 light-years away from Earth. Therefore Bob reaches his destination 10 years later (as measured on the Earth's frame). After a short stay, he returns to Earth, again at 80% of the speed of light. Consequently Alice is 40 years old when she sees Bob again.
- How old is Bob when they meet again?
 - How can the asymmetry in the twins' ages be explained? Notice that from Bob's point of view he is at rest in his spaceship and it is the Earth which moves away and then back again.
 - Imagine that each twin watches the other through a very powerful telescope. What do they see? In particular, how much time do they experience as they see one year elapse for their twin?
- (2) A particularly simple matter model is that of a smooth **massless scalar field** $\phi : M \rightarrow \mathbb{R}$, whose energy-momentum tensor is

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2}(\partial_\alpha \phi \partial^\alpha \phi)g_{\mu\nu}.$$

Show that if the Lorentzian manifold (M, g) satisfies the Einstein equations with this matter model then ϕ satisfies the **wave equation**

$$\square \phi = 0 \Leftrightarrow \nabla^\mu \partial_\mu \phi = 0.$$

- (3) The energy-momentum tensor for **perfect fluid** is

$$T_{\mu\nu} = (\rho + p)U_\mu U_\nu + pg_{\mu\nu},$$

where ρ is the fluid's rest density, p is the fluid's rest pressure, and U is a unit timelike vector field tangent to the histories of the fluid particles. Show that:

- (a) $(T_{\mu\nu}) = \text{diag}(\rho, p, p, p)$ in any orthonormal frame including U ;
 (b) the motion equations for the perfect fluid are

$$\begin{cases} \text{div}(\rho U) + p \text{div} U = 0 \\ (\rho + p)\nabla_U U = -(\text{grad } p)^\perp \end{cases} ,$$

where $^\perp$ represents the orthogonal projection on the spacelike hyperplane orthogonal to U .

CHAPTER 2

Exact solutions

In this chapter we present a number of exact solutions of the Einstein field equations, as well as their Penrose diagrams. These solutions will be used as examples or counter-examples to the theorems in the subsequent chapters. We also discuss the matching of two different solutions across a timelike hypersurface. A different perspective on Penrose diagrams can be found in [HE95].

1. Minkowski spacetime

The simplest solution of the Einstein field equations with zero cosmological constant in vacuum (i.e. with vanishing energy-momentum tensor) is the Minkowski spacetime, that is, \mathbb{R}^4 with the metric

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2.$$

Since this metric is flat, its curvature vanishes, and so do its Ricci and Einstein tensors. It represents a universe where there is no gravity whatsoever. Transforming the Cartesian coordinates (x, y, z) to spherical coordinates (r, θ, φ) yields

$$ds^2 = -dt^2 + dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2).$$

Performing the additional change of coordinates

$$\begin{cases} u = t - r & \text{(retarded time)} \\ v = t + r & \text{(advanced time)} \end{cases}$$

we obtain

$$ds^2 = -du dv + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2),$$

where

$$r(u, v) = \frac{1}{2}(v - u).$$

The coordinates (u, v) are called **null coordinates**: their level sets are null cones formed by outgoing/ingoing null geodesics emanating from the center. Note that they are subject to the constraint

$$r \geq 0 \Leftrightarrow v \geq u.$$

Finally, the coordinate change

$$(3) \quad \begin{cases} \tilde{u} = \tanh u \\ \tilde{v} = \tanh v \end{cases} \Leftrightarrow \begin{cases} u = \operatorname{arctanh} \tilde{u} \\ v = \operatorname{arctanh} \tilde{v} \end{cases}$$

brings the metric into the form

$$ds^2 = -\frac{1}{(1-\tilde{u}^2)(1-\tilde{v}^2)}d\tilde{u}d\tilde{v} + r^2(d\theta^2 + \sin^2\theta d\varphi^2),$$

where now

$$r(\tilde{u}, \tilde{v}) = \frac{1}{2}(\operatorname{arctanh}\tilde{v} - \operatorname{arctanh}\tilde{u})$$

and

$$(4) \quad -1 < \tilde{u} \leq \tilde{v} < 1.$$

Because (\tilde{u}, \tilde{v}) are also null coordinates, it is common to represent their axes tilted by 45° . The plane region defined by (4) is then represented in Figure 1.

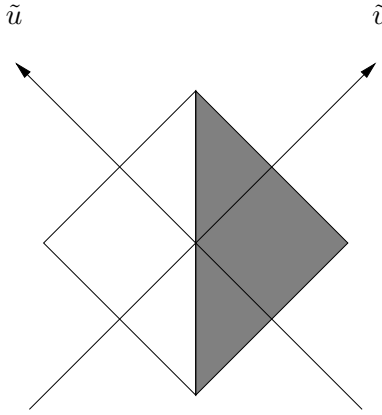


FIGURE 1. Range of the coordinates (\tilde{u}, \tilde{v}) .

This region is usually called the **Penrose diagram** for the Minkowski spacetime. If we take each point in the diagram to represent a sphere S^2 of radius $r(\tilde{u}, \tilde{v})$, the diagram itself represents the full spacetime manifold, in a way that makes causality relations apparent: any causal curve is represented in the diagram by a curve with tangent at most 45° from the vertical. In Figure 2 we represent some level hypersurfaces of t and r in the Penrose diagram. The former approach the point i^0 in the boundary of the diagram, called the **spacelike infinity**, whereas the later go from the boundary point i^- (**past timelike infinity**) to the boundary point i^+ (**future timelike infinity**). Finally, null geodesics start at the null boundary line \mathcal{I}^- (**past null infinity**) and end at the null boundary line \mathcal{I}^+ (**future null infinity**). These boundary points and lines represent ideal points at infinity, and do not correspond to actual points in the Minkowski spacetime.

2. Penrose diagrams

The concept of Penrose diagram can be easily generalized for any spherically symmetric space-time. Such spacetimes have metric

$$ds^2 = g_{AB}dx^A dx^B + r^2(d\theta^2 + \sin^2\theta d\varphi^2), \quad r = r(x^0, x^1),$$

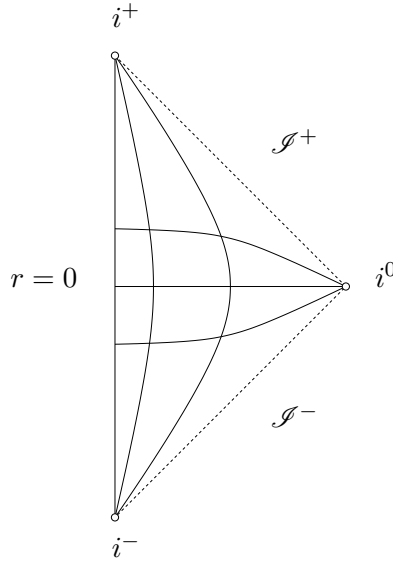


FIGURE 2. Penrose diagram for the Minkowski spacetime, with some level hypersurfaces of t and r represented.

where $g_{AB}dx^A dx^B$ is a Lorentzian metric on a 2-dimensional quotient manifold with boundary (which we assume to be diffeomorphic to a region of the plane). It turns out that any such metric is conformal to the Minkowski metric:

$$(5) \quad g_{AB}dx^A dx^B = -\Omega^2 du dv, \quad \Omega = \Omega(u, v).$$

This can be seen locally as follows: choose a spacelike line S , a coordinate u along it, and a coordinate w along a family of null geodesics emanating from S , so that S corresponds to $w = 0$ (Figure 3). Then near S we have

$$g_{uu} = \left\langle \frac{\partial}{\partial u}, \frac{\partial}{\partial u} \right\rangle > 0$$

and

$$g_{ww} = \left\langle \frac{\partial}{\partial w}, \frac{\partial}{\partial w} \right\rangle = 0.$$

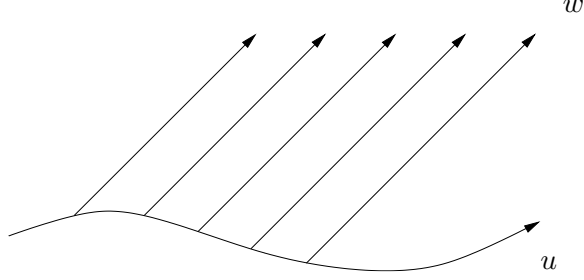
Therefore the 2-dimensional metric is written in these coordinates

$$g_{AB}dx^A dx^B = g_{uu}du^2 + 2g_{uw}dudw = g_{uu}du \left(du + \frac{2g_{uw}}{g_{uu}}dw \right).$$

As for any 1-form in a 2-dimensional manifold, we have

$$du + \frac{2g_{uw}}{g_{uu}}dw = f dv$$

for suitable functions f and v . Note that f cannot vanish, because (u, w) are local coordinates. Moreover, we can assume $f < 0$ by replacing v with $-v$ if necessary. Choosing $\Omega^2 = -fg_{uu}$ then yields (5).

FIGURE 3. Choice of the coordinates (u, w) .

We then see that any spherically symmetric metric can be written as

$$(6) \quad ds^2 = -\Omega^2 du dv + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)$$

with $\Omega = \Omega(u, v)$ and $r = r(u, v)$. By rescaling u and v if necessary, we can assume that the range of (u, v) is bounded, and hence obtain a Penrose diagram depicting the causal geometry. As we will see, this is extremely helpful in more complicated spherical symmetric solutions of the Einstein field equations.

REMARK 2.1. From

$$(g_{AB}) = \begin{pmatrix} 0 & -\frac{\Omega^2}{2} \\ -\frac{\Omega^2}{2} & 0 \end{pmatrix} \Rightarrow (g^{AB}) = \begin{pmatrix} 0 & -\frac{2}{\Omega^2} \\ -\frac{2}{\Omega^2} & 0 \end{pmatrix}$$

it is easily seen that

$$\partial_A \left(\sqrt{-\det(g_{CD})} g^{AB} \partial_B u \right) = 0 \Leftrightarrow \nabla_A \nabla^A u = 0.$$

and similarly for v . In other words, the null coordinates u and v are solutions of the wave equation in the 2-dimensional Lorentzian manifold:

$$\square u = \square v = 0.$$

This is the Lorentzian analogue of the so-called **isothermal coordinates** for Riemannian surfaces. The proof that they later exist locally is however slightly more complicated: given a point p on the surface, one chooses a local harmonic function with nonvanishing derivative,

$$\Delta u = 0, \quad (du)_p \neq 0,$$

and considers the equation

$$(7) \quad dv = \star du.$$

Here \star is the Hodge star, which for generic orientable n -dimensional pseudo-Riemannian manifolds is defined as follows: if $\{\omega^1, \dots, \omega^n\}$ is any positively oriented orthonormal coframe then

$$\star(\omega^1 \wedge \dots \wedge \omega^k) = \langle \omega^1, \omega^1 \rangle \dots \langle \omega^k, \omega^k \rangle \omega^{k+1} \wedge \dots \wedge \omega^n.$$

By the Poincaré Lemma, equation (7) can be locally solved, since

$$d \star du = \star \star d \star du = \star(\Delta u) = 0.$$

Moreover, v is itself harmonic, because

$$\Delta v = \star d \star dv = \star d \star \star du = \star d(-du) = 0.$$

Finally,

$$\|du\| = \|dv\| = \frac{1}{\Omega}$$

for some local function $\Omega > 0$, and so the metric is written in these coordinates as

$$ds^2 = \Omega^2 (du^2 + dv^2).$$

3. The Schwarzschild solution

If we try to solve the vacuum Einstein field equations with zero cosmological constant for a spherically symmetric Lorentzian metric, we obtain, after suitably rescaling the time coordinate, the **Schwarzschild metric**

$$ds^2 = - \left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)$$

(where $M \in \mathbb{R}$ is a constant). Note that for $M = 0$ we retrieve the Minkowski metric in spherical coordinates. Note also that if $M > 0$ then the metric is defined in two disconnected domains of coordinates, corresponding to $r \in (0, 2M)$ and $r \in (2M, +\infty)$.

The physical interpretation of the Schwarzschild solution can be found by considering the proper time of a timelike curve parameterized by the time coordinate:

$$\tau = \int_{t_0}^{t_1} \left[\left(1 - \frac{2M}{r}\right) - \left(1 - \frac{2M}{r}\right)^{-1} \dot{r}^2 - r^2 \dot{\theta}^2 - r^2 \sin^2 \theta \dot{\varphi}^2 \right]^{\frac{1}{2}} dt,$$

where $\dot{r} = \frac{dr}{dt}$, etc. The integrand L_S is the Lagrangian for geodesic motion in the Schwarzschild spacetime when parameterized by the time coordinate. Now for motions with speeds much smaller than the speed of light we have $\dot{r}^2 \ll 1$, etc. Assuming $\frac{M}{r} \ll 1$ as well we have

$$\begin{aligned} L_S &= \left[1 - \frac{2M}{r} - \left(1 - \frac{2M}{r}\right)^{-1} \dot{r}^2 - r^2 \dot{\theta}^2 - r^2 \sin^2 \theta \dot{\varphi}^2 \right]^{\frac{1}{2}} \\ &\simeq 1 - \frac{M}{r} - \frac{1}{2} \left(\dot{r}^2 + r^2 \dot{\theta}^2 + \sin^2 \theta \dot{\varphi}^2 \right) = 1 - L_N, \end{aligned}$$

where

$$L_N = \frac{1}{2} \left(\dot{r}^2 + r^2 \dot{\theta}^2 + r^2 \sin^2 \theta \dot{\varphi}^2 \right) + \frac{M}{r}$$

is precisely the Newtonian Lagrangian for the motion of a particle in the gravitational field of a point mass M . The Schwarzschild solution should therefore be considered the relativistic analogue of this field.

To write the Schwarzschild metric in the form (6) we note that the quotient metric is

$$\begin{aligned}
ds^2 &= - \left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 \\
&= - \left(1 - \frac{2M}{r}\right) \left[dt^2 - \left(1 - \frac{2M}{r}\right)^{-2} dr^2 \right] \\
&= - \left(1 - \frac{2M}{r}\right) \left[dt - \left(1 - \frac{2M}{r}\right)^{-1} dr \right] \left[dt + \left(1 - \frac{2M}{r}\right)^{-1} dr \right] \\
&= - \left(1 - \frac{2M}{r}\right) du dv,
\end{aligned}$$

where we define

$$u = t - \int \left(1 - \frac{2M}{r}\right)^{-1} dr = t - r - 2M \log |r - 2M|$$

and

$$v = t + \int \left(1 - \frac{2M}{r}\right)^{-1} dr = t + r + 2M \log |r - 2M|.$$

In the domain of coordinates $r > 2M$ we have $1 - \frac{2M}{r} > 0$, and so the quotient metric is already in the required form. Note however that, unlike what happened in the Minkowski spacetime, we now have

$$v - u = 2r + 4M \log |r - 2M| \in (-\infty, +\infty).$$

Consequently, by applying the coordinate rescaling (3) we obtain the full square, instead of a triangle (Figure 4). Besides the infinity points and null boundaries also present in the Penrose diagram for the Minkowski spacetime, there are two new null boundaries, \mathcal{H}^- (**past event horizon**) and \mathcal{H}^+ (**future event horizon**), where $r = 2M$.

It seems reasonable to expect that the metric can be extended across the horizons, since r does not tend to zero nor to infinity there; this expectation is confirmed by calculating the so-called **Kretschmann scalar**:

$$R_{\alpha\beta\mu\nu}R^{\alpha\beta\mu\nu} = \frac{48M^2}{r^6}.$$

This is perfectly well behaved as $r \rightarrow 2M$, and seems to indicate that the horizons are mere singularities of the coordinate system (t, r) . To show that this is indeed the case, note that in the (u, r) coordinate system the quotient metric is written

$$ds^2 = - \left(1 - \frac{2M}{r}\right) du^2 - 2du dr.$$

Since

$$\det \begin{pmatrix} -1 + \frac{2M}{r} & -1 \\ -1 & 0 \end{pmatrix} = -1,$$

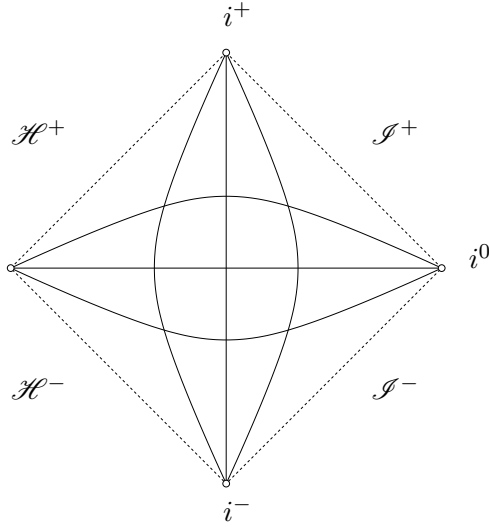


FIGURE 4. Penrose diagram for the region $r > 2M$ of the Schwarzschild spacetime, with some level hypersurfaces of t and r represented.

we see that the metric is well defined across $r = 2M$ in this coordinate system. Moreover, we know that it solves the Einstein equations in the coordinate domains $r < 2M$ and $r > 2M$; by continuity, it must solve it in the whole domain $r \in (0, +\infty)$. Note that the coordinate domains $r < 2M$ and $r > 2M$ are glued along $r = 2M$ so that the outgoing null geodesics $u = \text{constant}$ go from $r = 0$ to $r = +\infty$; in other words, the gluing is along the past event horizon \mathcal{H}^- .

To obtain the Penrose diagram for the coordinate domain $r < 2M$ we note that the quotient metric can be written as

$$ds^2 = - \left(1 - \frac{2M}{r}\right) du dv = - \left(\frac{2M}{r} - 1\right) du (-dv) = - \left(\frac{2M}{r} - 1\right) du dv',$$

where $v' = -v$. Since in this coordinate domain $\frac{2M}{r} - 1 > 0$, the quotient metric is in the required form. Note however that we now have

$$u + v' = -2r - 4M \log |r - 2M| \in (-4M \log(2M), +\infty),$$

and by setting

$$v'' = v' + 4M \log(2M)$$

we obtain

$$u + v'' > 0.$$

Consequently, by applying the coordinate rescaling (3) we obtain a triangle (Figure 5). There is now a spacelike boundary, where $r = 0$, and two null boundaries \mathcal{H}^- , where $r = 2M$. The Penrose diagram for the domain of the coordinates (u, r) can be obtained gluing the Penrose diagrams in Figures 4 and 5 along \mathcal{H}^- , so that the null geodesics $u = \text{constant}$ match (Figure 6).

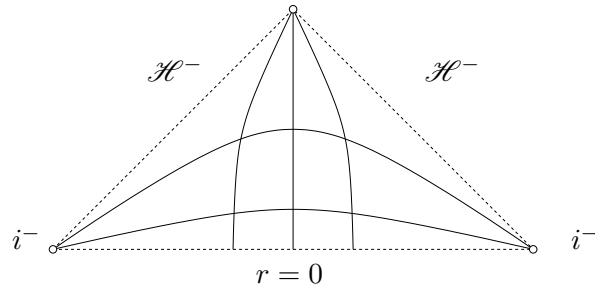


FIGURE 5. Penrose diagram for a region $r < 2M$ of the Schwarzschild spacetime, with some level hypersurfaces of t and r represented.

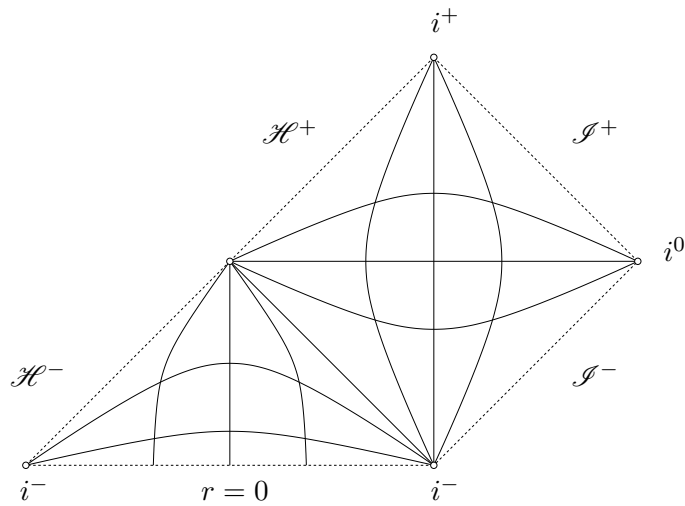


FIGURE 6. Penrose diagram for the domain of the coordinates (u, r) in the Schwarzschild spacetime, with some level hypersurfaces of t and r represented.

If instead we use the (v, r) coordinate system, the quotient metric is written

$$ds^2 = - \left(1 - \frac{2M}{r} \right) dv^2 + 2dv dr.$$

Again the metric is well defined across $r = 2M$ in this coordinate system, since

$$\det \begin{pmatrix} -1 + \frac{2M}{r} & 1 \\ 1 & 0 \end{pmatrix} = -1,$$

and solves the Einstein equations in the whole coordinate domain $r \in (0, +\infty)$. The coordinate domains $r < 2M$ and $r > 2M$ are now glued along $r = 2M$ so that the ingoing null geodesics $v = \text{constant}$ go from $r = +\infty$ to $r = 0$; in other words, the gluing is along the future event horizon \mathcal{H}^+ .

To obtain the Penrose diagram for the coordinate domain $r < 2M$ we note that the quotient metric can be written as

$$ds^2 = - \left(1 - \frac{2M}{r}\right) du dv = - \left(\frac{2M}{r} - 1\right) (-du) dv = - \left(\frac{2M}{r} - 1\right) du' dv,$$

where $u' = -u$. Since in this coordinate domain $\frac{2M}{r} - 1 > 0$, the quotient metric is in the required form. We have

$$u' + v = 2r + 4M \log |r - 2M| \in (-\infty, 4M \log(2M)),$$

and by setting

$$u'' = u' - 4M \log(2M)$$

we obtain

$$u'' + v < 0.$$

Consequently, by applying the coordinate rescaling (3) we obtain a triangle (Figure 7). Again there is a spacelike boundary, where $r = 0$, and two null boundaries \mathcal{H}^+ , where $r = 2M$. The Penrose diagram for the domain of the coordinates (v, r) can be obtained gluing the Penrose diagrams in Figures 4 and 7 along \mathcal{H}^+ , so that the null geodesics $v = \text{constant}$ match (Figure 8).

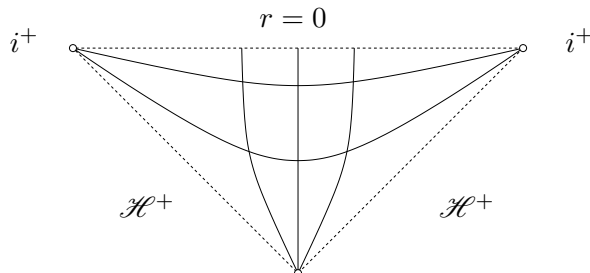


FIGURE 7. Penrose diagram for a region $r < 2M$ of the Schwarzschild spacetime, with some level hypersurfaces of t and r represented.

Both regions $r < 2M$ can of course be glued to the region $r > 2M$ simultaneously. Since they are invariant under reflections with respect to $t = 0$ (the vertical line through their common vertex), it is then clear that a mirror-reversed copy of the region $r > 2M$ can be glued to the surviving null boundaries \mathcal{H}^- and \mathcal{H}^+ (Figure 9). The resulting spacetime, known as the **maximal analytical extension** of the Schwarzschild solution, is a solution of the Einstein equations which cannot be extended any further, since $r \rightarrow 0$ or $r \rightarrow +\infty$ on the boundary of its Penrose diagram. Note that by continuity the Einstein equations hold at the point where the four Penrose diagrams intersect (known as the **bifurcate sphere**).

Let us now analyze in detail the Penrose diagram for the maximal analytic extension of the Schwarzschild spacetime. There are two **asymptotically flat** regions $r > 2M$, corresponding to two causally disconnected universes, joined by a **wormhole**. There are also two regions where $r < 2M$: a

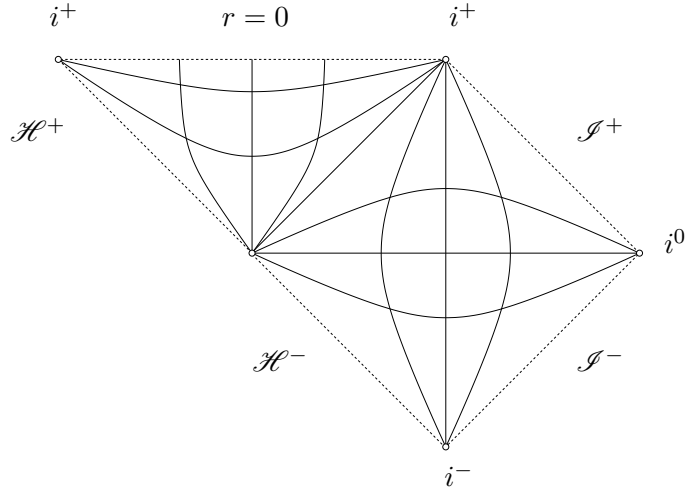


FIGURE 8. Penrose diagram for the domain of the coordinates (v, r) in the Schwarzschild spacetime, with some level hypersurfaces of t and r represented.

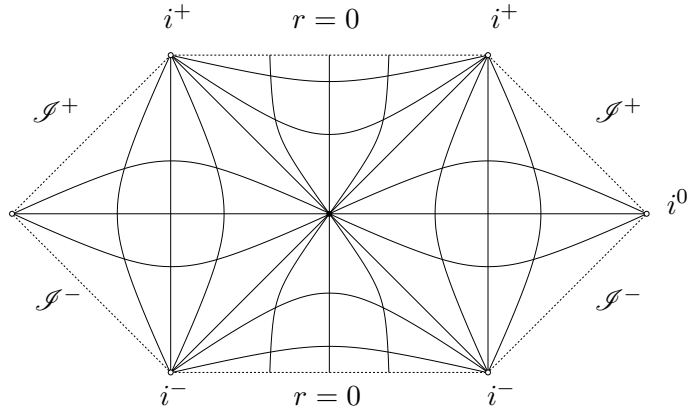


FIGURE 9. Penrose diagram for the maximal analytic extension of the Schwarzschild spacetime, with some level hypersurfaces of t and r represented.

black hole region, bounded by the future event horizons \mathcal{H}^+ , from which no causal curve can escape; and a **white hole region**, bounded by the past event horizons \mathcal{H}^- , from which every causal curve must escape. Note that the horizons themselves correspond to spheres which are propagating at the speed of light, but whose radius remains constant, $r = 2M$.

The black hole in the maximal analytic extension of the Schwarzschild spacetime is an **eternal black hole**, that is, a black hole which has always existed (as opposed to having formed by some physical process). We will see shortly how to use the Schwarzschild solution to model physically realistic black holes.

4. Friedmann-Lemaître-Robertson-Walker models

The simplest models of **cosmology**, the study of the Universe as a whole, are obtained from the assumption that space is homogeneous and isotropic (which is true on average at very large scales). It is well known that the only isotropic 3-dimensional Riemannian metrics are, up to scale, given by

$$dl^2 = \frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2),$$

where

$$k = \begin{cases} 1 & \text{for the standard metric on } S^3 \\ 0 & \text{for the standard metric on } \mathbb{R}^3 \\ -1 & \text{for the standard metric on } H^3 \end{cases}.$$

Allowing for a time-dependent scale factor $a(t)$ (also known as the “radius of the Universe”), we arrive at the **Friedmann-Lemaître-Robertson-Walker** (FLRW) family of Lorentzian metrics:

$$(8) \quad ds^2 = -dt^2 + a^2(t) \left[\frac{dr^2}{1 - kr^2} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right].$$

To interpret these metrics, we consider a general Lorentzian metric of the form

$$ds^2 = -e^{2\phi} dt^2 + h_{ij} dx^i dx^j = -e^{2\phi} dt^2 + dl^2.$$

The Riemannian metric dl^2 is readily interpreted as giving the distances measured between nearby observers with fixed space coordinates x^i in radar experiments: indeed, such observers measure proper time τ given by

$$d\tau^2 = e^{2\phi} dt^2.$$

The null geodesics representing a radar signal bounced by a given observer from a nearby observer (Figure 10) satisfy

$$ds^2 = 0 \Leftrightarrow e^{2\phi} dt^2 = dl^2 \Leftrightarrow d\tau^2 = dl^2 \Leftrightarrow d\tau = \pm dl.$$

Since the speed of light is $c = 1$, the distance traveled between the observers will be half the time between the emission and the reception of the signal:

$$\frac{(\tau + dl) - (\tau - dl)}{2} = dl.$$

Moreover, the unit timelike covector field tangent to the trajectories of the the observers with fixed space coordinates x^i is

$$U = -e^\phi dt \Leftrightarrow U_\mu = -e^\phi \nabla_\mu t.$$

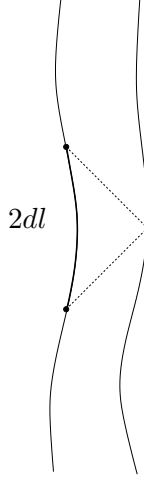


FIGURE 10. Distance between nearby observers.

Therefore

$$\begin{aligned}
 \nabla_U U^\mu &= -U^\nu \nabla_\nu (e^\phi \nabla^\mu t) = (U \cdot \phi) U^\mu - e^\phi U^\nu \nabla_\nu \nabla^\mu t \\
 &= (U \cdot \phi) U^\mu - e^\phi U^\nu \nabla^\mu \nabla_\nu t = (U \cdot \phi) U^\mu + e^\phi U^\nu \nabla^\mu (e^{-\phi} U_\nu) \\
 &= (U \cdot \phi) U^\mu - U^\nu U_\nu \nabla^\mu \phi + U^\nu \nabla^\mu U_\nu \\
 &= \nabla^\mu \phi + (U^\nu \nabla_\nu \phi) U^\mu + \frac{1}{2} \nabla^\mu (U^\nu U_\nu) = \nabla^\mu \phi + (U^\nu \nabla_\nu \phi) U^\mu,
 \end{aligned}$$

since $U^\nu U_\nu = -1$. In other words,

$$\nabla_U U = (\text{grad } \phi)^\perp,$$

where $^\perp$ represents the orthogonal projection on the spacelike hyperplane orthogonal to U .

Therefore the observers with fixed space coordinates in the FLRW models have zero acceleration, that is, they are free-falling (by opposition to the corresponding observers in the Schwarzschild spacetime, who must accelerate to remain at fixed $r > 2M$). Moreover, the distance between two such observers varies as

$$d(t) = a(t) \frac{d_0}{a_0} \Rightarrow \dot{d} = \dot{a} \frac{d_0}{a_0} = \frac{\dot{a}}{a} d.$$

This relation, known as the **Hubble law**, is often written as

$$v = Hd,$$

where v is the relative velocity and

$$H = \frac{\dot{a}}{a}$$

is the so-called **Hubble constant** (for historical reasons, since it actually varies in time).

We will model the matter content of the universe as an uniform dust of galaxies placed at fixed space coordinates (hence free-falling):

$$(9) \quad T = \rho(t)dt \otimes dt.$$

Plugging the metric (8) and the energy-momentum tensor (9) into the Einstein equations, and integrating once, results in the so-called **Friedmann equations**

$$\begin{cases} \frac{1}{2}\dot{a}^2 - \frac{\alpha}{a} - \frac{\Lambda}{6}a^2 = -\frac{k}{2} \\ \frac{4\pi}{3}\rho a^3 = \alpha \end{cases}$$

(where α is an integration constant). The first Friedmann equation is a first order ODE for $a(t)$; it can be seen as the equation of conservation of energy for a particle moving in the 1-dimensional effective potential

$$V(a) = -\frac{\alpha}{a} - \frac{\Lambda}{6}a^2$$

with energy $-\frac{k}{2}$. Once this equation has been solved, the second Friedmann equation yields $\rho(t)$ from $a(t)$. We now examine in detail the FLRW models arising from the solutions of these equations.

4.1. Milne universe. If we set $\alpha = \Lambda = 0$ then the first Friedmann equation becomes

$$\dot{a}^2 = -k.$$

Therefore either $k = 0$ and $\dot{a} = 0$, which corresponds to the Minkowski spacetime, or $k = -1$ and $\dot{a}^2 = 1$, that is

$$ds^2 = -dt^2 + t^2 dl_{H^3}^2,$$

where $dl_{H^3}^2$ represents the metric of the unit hyperbolic 3-space; this is the so-called **Milne universe**. It turns out that the Milne universe is isometric to an open region of the Minkowski spacetime, namely the region limited by the future (or past) light cone of the origin. This region is foliated by hyperboloids S_t of the form

$$T^2 - X^2 - Y^2 - Z^2 = t^2,$$

whose induced metric is that of a hyperbolic space of radius t (Figure 11). Note that the light cone corresponds to $a(t) = t = 0$, that is, the **Big Bang** of the Milne universe.

4.2. de Sitter universe. If $\alpha = 0$ and $\Lambda > 0$ we can choose units such that $\Lambda = 3$. The first Friedmann equation then becomes

$$\dot{a}^2 - a^2 = -k.$$

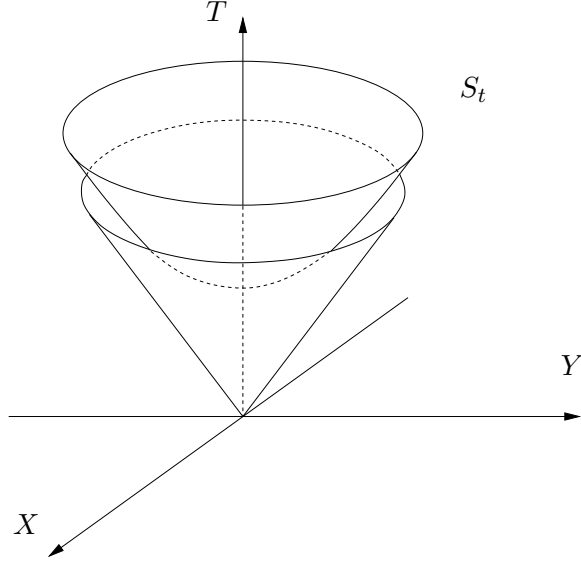


FIGURE 11. Milne universe.

In this case all three values $k = 1$, $k = 0$ and $k = -1$ are possible; the corresponding metrics are, respectively,

$$ds^2 = -dt^2 + \cosh^2 t dl_{S^3}^2;$$

$$ds^2 = -dt^2 + e^{2t} dl_{\mathbb{R}^3}^2;$$

$$ds^2 = -dt^2 + \sinh^2 t dl_{H^3}^2,$$

where $dl_{S^3}^2$, $dl_{\mathbb{R}^3}^2$ and $dl_{H^3}^2$ represent the metric of the unit 3-sphere, the Euclidean 3-space and the the unit hyperbolic 3-space.

It turns out that the last two models correspond to open regions of the first, which is then called the **de Sitter universe**. It represents a spherical universe which contracts to a minimum radius (1 in our units) and then re-expands. It is easily seen to be isometric to the unit hyperboloid

$$-T^2 + X^2 + Y^2 + Z^2 + W^2 = 1$$

in the Minkowski 5-dimensional spacetime (Figure 12).

To obtain the Penrose diagram for the de Sitter universe we write its metric as

$$\begin{aligned} ds^2 &= -dt^2 + \cosh^2 t [d\psi^2 + \sin^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2)] \\ &= \cosh^2 t [-d\tau^2 + d\psi^2 + \sin^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2)] \\ &= \cosh^2 t [-d\tau^2 + d\psi^2] + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \end{aligned}$$

where $\psi \in [0, \pi]$,

$$\tau = \int_{-\infty}^t \frac{dt}{\cosh t}$$

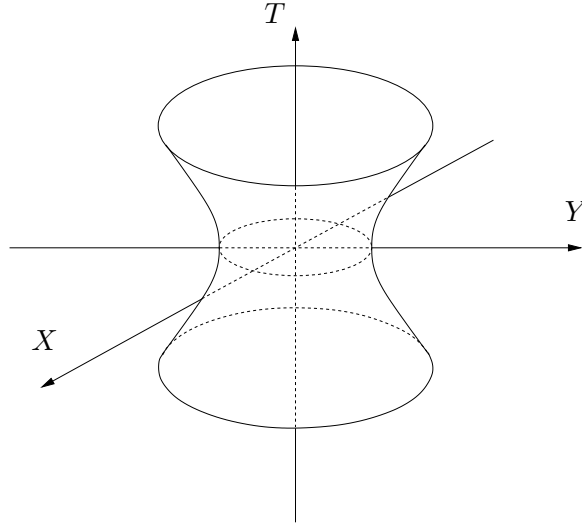


FIGURE 12. de Sitter universe.

and

$$r = \cosh t \sin \psi.$$

Since

$$\int_{-\infty}^{+\infty} \frac{dt}{\cosh t} = \pi,$$

we see that the quotient metric is conformal to the square $(0, \pi) \times [0, \pi]$ of the Minkowski 2-dimensional spacetime, and so the Penrose diagram is as depicted in Figure 12. Note that there are two lines where $r = 0$, corresponding to two antipodal points of the 3-sphere. A light ray emitted from one of these points at $t = -\infty$ has just enough time to reach the other point at $t = +\infty$ (dashed line in the diagram). Note also that in this case \mathcal{I}^- and \mathcal{I}^+ (defined as the past and future boundary points approached by null geodesics along which $r \rightarrow +\infty$) are spacelike boundaries.

4.3. Anti-de Sitter universe. If $\alpha = 0$ and $\Lambda < 0$ we can choose units such that $\Lambda = -3$. The the first Friedmann equation then becomes

$$\dot{a}^2 + a^2 = -k.$$

In this case only $k = -1$ is possible; the corresponding metric is

$$ds^2 = -dt^2 + \cos^2 t dl_{H^3}^2.$$

It turns out (see Exercise 5 in Chapter 5) that this model is an open region of the spacetime with metric

$$ds^2 = -\cosh^2 \psi dt^2 + d\psi^2 + \sinh^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2)$$

(where $\psi \in [0, +\infty)$), called the **anti-de Sitter universe**. It represents a static hyperbolic universe (with radius 1 in our units).

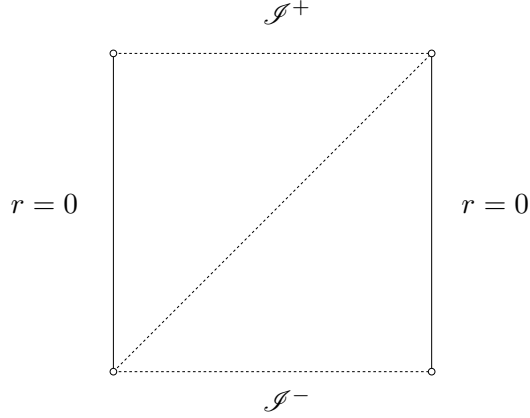


FIGURE 13. Penrose diagram for the de Sitter universe.

To obtain the Penrose diagram for the anti-de Sitter universe we write its metric as

$$\begin{aligned} ds^2 &= \cosh^2 \psi \left[-dt^2 + \frac{d\psi^2}{\cosh^2 \psi} \right] + \sinh^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2) \\ &= \cosh^2 \psi [-dt^2 + dx^2] + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \end{aligned}$$

where

$$x = \int_0^\psi \frac{d\psi}{\cosh \psi}$$

and

$$r = \sinh \psi.$$

Since

$$\int_0^{+\infty} \frac{d\psi}{\cosh \psi} = \frac{\pi}{2},$$

we see that the quotient metric is conformal to the strip $\mathbb{R} \times [0, \frac{\pi}{2})$ of the Minkowski 2-dimensional spacetime, and so the Penrose diagram is as depicted in Figure 14. The FLRW model above corresponds to the triangular region in the diagram. Note also that in this case $\mathcal{I}^- \equiv \mathcal{I}^+ \equiv \mathcal{I}$ is a timelike boundary.

4.4. Universes with matter and $\Lambda = 0$. If $\alpha > 0$ and $\Lambda = 0$, the first Friedmann equation becomes

$$\dot{a}^2 - \frac{2\alpha}{a} = -k.$$

In this case all three values $k = 1$, $k = 0$ and $k = -1$ are possible. Although it is possible to obtain explicit formulas for the solutions of these equations, it is simpler to analyze the graph of the effective potential $V(a)$ (Figure 15). Possibly by reversing and translating t , we can assume that all solutions are defined for $t > 0$, with $\lim_{t \rightarrow 0} a(t) = 0$, implying $\lim_{t \rightarrow 0} \rho(t) = +\infty$. Therefore all three models have a true singularity at $t = 0$, known as the

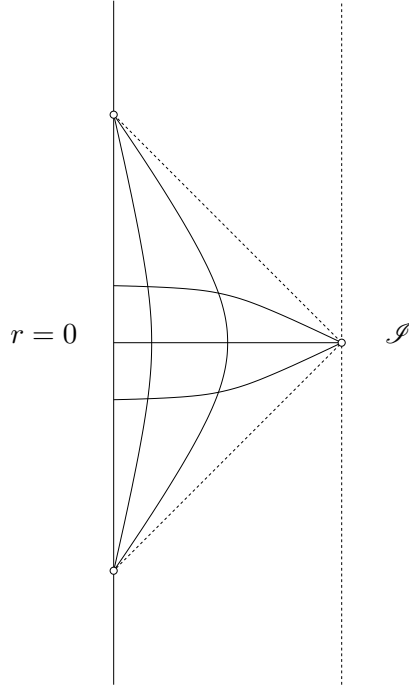


FIGURE 14. Penrose diagram for the anti-de Sitter universe.

Big Bang, where the scalar curvature $R = 8\pi\rho$ also blows up; this is not true for the Milne universe or the open region in the anti-de Sitter universe, which can be extended across the Big Bang. The spherical universe ($k = 1$) reaches a maximum radius 2α and re-collapses, forming a second singularity (the **Big Crunch**); the radius of the flat ($k = 0$) and hyperbolic ($k = -1$) universes increases monotonically.

To obtain the Penrose diagram for the spherical universe we write its metric as

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t) [d\psi^2 + \sin^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2)] \\ &= a^2(t) [-d\tau^2 + d\psi^2 + \sin^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2)] \\ &= a^2(t) [-d\tau^2 + d\psi^2] + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \end{aligned}$$

where $\psi \in [0, \pi]$,

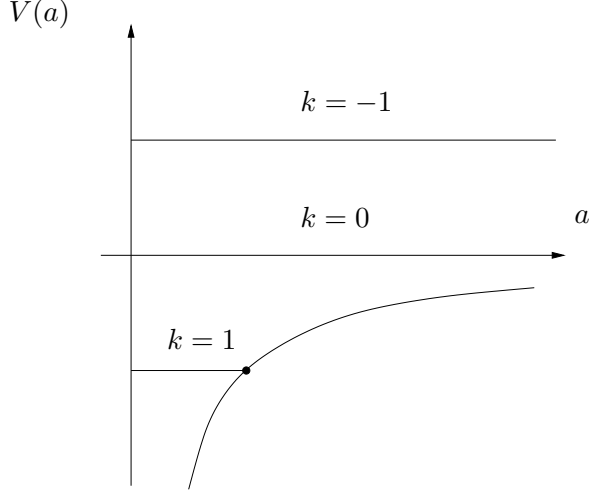
$$\tau = \int_0^t \frac{dt}{a(t)}$$

and

$$r = a(t) \sin \psi.$$

Since

$$\int_0^{t_{\max}} \frac{dt}{a(t)} = 2 \int_0^{a_{\max}} \frac{da}{a\dot{a}} = 2 \int_0^{2\alpha} \frac{da}{a\sqrt{\frac{2\alpha}{a} - 1}} = 2\pi,$$

FIGURE 15. Effective potential for FLRW models with $\Lambda = 0$.

we see that the quotient metric is conformal to the rectangle $(0, 2\pi) \times [0, \pi]$ of the Minkowski 2-dimensional spacetime, and so the Penrose diagram is as depicted in Figure 16. Note that there are two lines where $r = 0$, corresponding to two antipodal points of the 3-sphere. A light ray emitted from one of these points at $t = 0$ has just enough time to circle once around the universe and return at $t = t_{\max}$ (dashed line in the diagram). Note also that the Big Bang and the Big Crunch are spacelike boundaries.

To obtain the Penrose diagram for the flat universe we write its metric as

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t) [d\rho^2 + \rho^2 (d\theta^2 + \sin^2 \theta d\varphi^2)] \\ &= a^2(t) [-d\tau^2 + d\rho^2 + \rho^2 (d\theta^2 + \sin^2 \theta d\varphi^2)] \\ &= a^2(t) [-d\tau^2 + d\rho^2] + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \end{aligned}$$

where $\rho \in [0, +\infty)$ and $r = a(t)\rho$. Since

$$\int_0^{+\infty} \frac{dt}{a(t)} = \int_0^{+\infty} \frac{da}{a\dot{a}} = \int_0^{+\infty} \frac{da}{a\sqrt{\frac{2\alpha}{a}}} = +\infty,$$

we see that the quotient metric is conformal to the region $(0, +\infty) \times [0, +\infty)$ of the Minkowski 2-dimensional spacetime, and so the Penrose diagram is as depicted in Figure 17. Note that the Big Bang is a spacelike boundary.

The Penrose diagram for the hyperbolic universe turns out to be the same as for the flat universe. To see this we write its metric as

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t) [d\psi^2 + \sinh^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2)] \\ &= a^2(t) [-d\tau^2 + d\psi^2 + \sinh^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2)] \\ &= a^2(t) [-d\tau^2 + d\psi^2] + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2), \end{aligned}$$

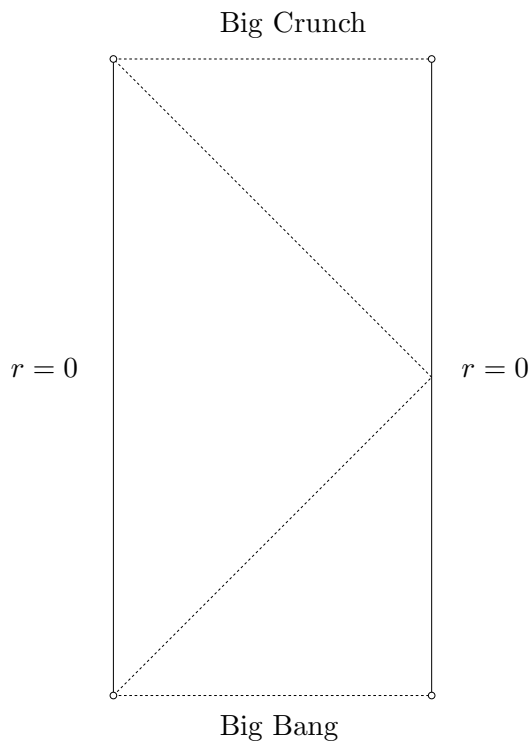


FIGURE 16. Penrose diagram for the spherical universe.

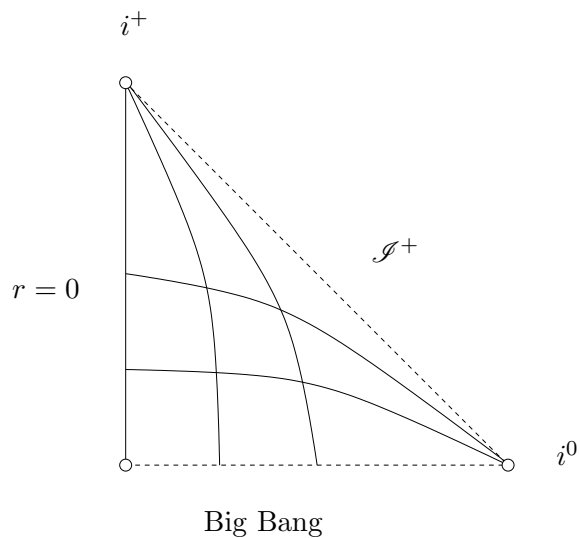


FIGURE 17. Penrose diagram for the flat and hyperbolic universes.

where $\psi \in [0, +\infty)$ and $r = a(t) \sinh \psi$, and note that

$$\int_0^{+\infty} \frac{dt}{a(t)} = \int_0^{+\infty} \frac{da}{a\dot{a}} = \int_0^{+\infty} \frac{da}{a\sqrt{\frac{2\alpha}{a} + 1}} = +\infty.$$

4.5. Universes with matter and $\Lambda > 0$. If $\alpha > 0$ and $\Lambda > 0$ we can choose units such that $\Lambda = 3$. The first Friedmann equation then becomes

$$\dot{a}^2 - \frac{2\alpha}{a} - a^2 = -k.$$

In this case all three values $k = 1$, $k = 0$ and $k = -1$ are possible. As before, we analyze the graph of the effective potential $V(a)$ (Figure 18). The hyperbolic and flat universes behave qualitatively like when $\Lambda = 0$, although $\dot{a}(t)$ is now unbounded as $t \rightarrow +\infty$, instead of approaching some constant. The spherical universe has a richer spectrum of possible behaviours, depending on α , represented in Figure 18 by drawing the line of constant energy $-k = -1$ at three different heights. The higher line (corresponding to $\alpha > \frac{\sqrt{3}}{9}$) yields a behaviour similar to that of the hyperbolic and flat universes. The intermediate line (corresponding to $\alpha = \frac{\sqrt{3}}{9}$) gives rise to an unstable equilibrium point $a = \frac{\sqrt{3}}{3}$, where the attraction force of the matter is balanced by the repulsion force of the cosmological constant; it corresponds to the so-called **Einstein universe**, the first cosmological model ever proposed. The intermediate line also yields two solutions asymptotic to the Einstein universe, one containing a Big Bang and the other endless expansion. Finally, the lower line (corresponding to $\alpha < \frac{\sqrt{3}}{9}$) yields two different types of behaviour (depending on the initial conditions): either similar to the spherical model with $\Lambda = 0$, or to the de Sitter universe.

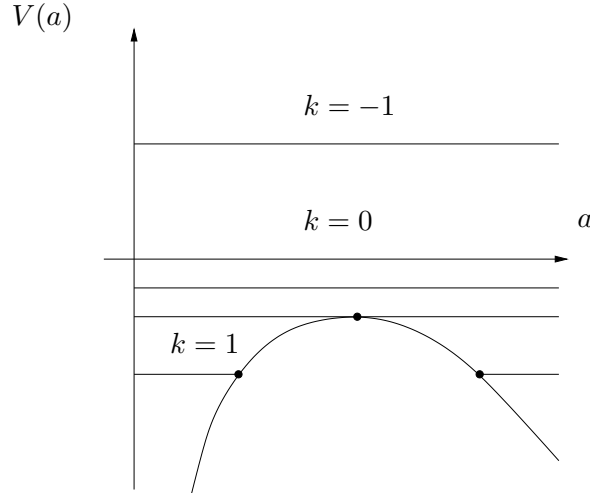


FIGURE 18. Effective potential for FLRW models with $\Lambda > 0$.

It is currently believed that the best model for our physical Universe is the flat universe with $\Lambda > 0$. If we write the first Friedmann equation as

$$H^2 \equiv \frac{\dot{a}^2}{a^2} = \frac{8\pi}{3}\rho + \frac{\Lambda}{3}$$

then the terms on the right-hand side are in the proportion 2 : 5 at the present time.

4.6. Universes with matter and $\Lambda < 0$. If $\alpha > 0$ and $\Lambda < 0$ we can choose units such that $\Lambda = -3$. The first Friedmann equation then becomes

$$\dot{a}^2 - \frac{2\alpha}{a} + a^2 = -k.$$

In this case all three values $k = 1$, $k = 0$ and $k = -1$ are possible. As before, we analyze the graph of the effective potential $V(a)$ (Figure 19). The qualitative behaviour of the hyperbolic, flat and spherical universes is the same as the spherical universe with $\Lambda = 0$, namely starting at a Big Bang and ending at a Big Crunch.

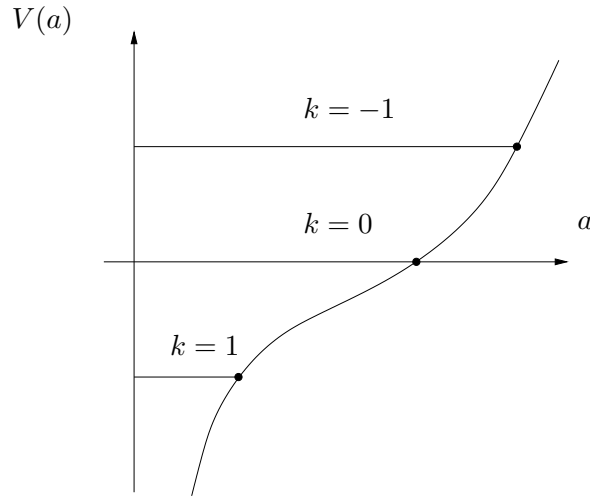


FIGURE 19. Effective potential for FLRW models with $\Lambda > 0$.

5. Matching

Let (M_1, g_1) and (M_2, g_2) be solutions of the Einstein field equations containing open sets U_1 and U_2 whose boundaries S_1 and S_2 are **timelike hypersurfaces**, that is, hypersurfaces whose induced metric is Lorentzian (or, equivalently, whose normal vector is spacelike). If S_1 is diffeomorphic to S_2 then we can identify them to obtain a new manifold M gluing U_1 to U_2 along $S_1 \cong S_2$ (Figure 20).

Let n be the unit normal vector to S_1 pointing out of U_1 , which we identify with the unit normal vector to S_2 pointing into U_2 . If (x^1, x^2, x^3) are local coordinates on $S \equiv S_1 \cong S_2$, we can construct a system of local coordinates (t, x^1, x^2, x^3) in a neighbourhood of S by moving a distance t along the geodesics with initial condition n . Note that U_1 , S and U_2

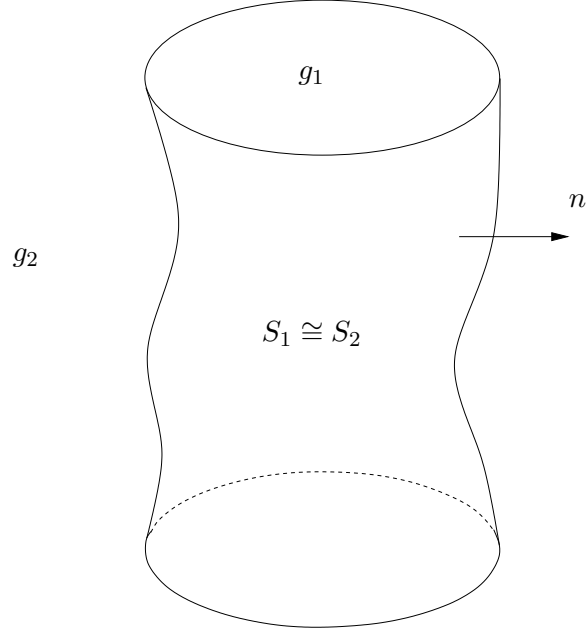


FIGURE 20. Matching two spacetimes.

correspond to $t < 0$, $t = 0$ and $t > 0$ in these coordinates. Since $\frac{\partial}{\partial t}$ is the unit tangent vector to the geodesics, we have

$$\begin{aligned} \frac{\partial}{\partial t} \left\langle \frac{\partial}{\partial t}, \frac{\partial}{\partial x^i} \right\rangle &= \left\langle \nabla_{\frac{\partial}{\partial t}} \frac{\partial}{\partial t}, \frac{\partial}{\partial x^i} \right\rangle + \left\langle \frac{\partial}{\partial t}, \nabla_{\frac{\partial}{\partial t}} \frac{\partial}{\partial x^i} \right\rangle \\ &= \left\langle \frac{\partial}{\partial t}, \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial t} \right\rangle = \frac{\partial}{\partial x^i} \left(\frac{1}{2} \left\langle \frac{\partial}{\partial t}, \frac{\partial}{\partial t} \right\rangle \right) = 0 \end{aligned}$$

($i = 1, 2, 3$), where we used

$$\nabla_{\frac{\partial}{\partial t}} \frac{\partial}{\partial x^i} - \nabla_{\frac{\partial}{\partial x^i}} \frac{\partial}{\partial t} = \left[\frac{\partial}{\partial t}, \frac{\partial}{\partial x^i} \right] = 0.$$

Since for $t = 0$ we have

$$\left\langle \frac{\partial}{\partial t}, \frac{\partial}{\partial x^i} \right\rangle = \left\langle n, \frac{\partial}{\partial x^i} \right\rangle = 0,$$

we see that $\frac{\partial}{\partial t}$ remains orthogonal to the surfaces of constant t . This result will be used repeatedly.

LEMMA 5.1. (Gauss Lemma I) *Let (M, g) be a Riemannian or a Lorentzian manifold, and $S \subset M$ a hypersurface whose normal vector field n satisfies $g(n, n) \neq 0$. The hypersurfaces S_t obtained from S by moving a distance t along the geodesics orthogonal to S remain orthogonal to the geodesics.*

The same ideas can be used to prove a closely related result.

LEMMA 5.2. (**Gauss Lemma II**) *Let (M, g) be a Riemannian or a Lorentzian manifold and $p \in M$. The hypersurfaces S_t obtained from p by moving a distance t along the geodesics through p remain orthogonal to the geodesics.*

In this coordinate system the metrics g_1 and g_2 are given on $t \leq 0$ and $t \geq 0$, respectively, by

$$g_A = dt^2 + h_{ij}^A(t, x) dx^i dx^j$$

($A = 1, 2$). Therefore we can define a continuous metric g on M if

$$h_{ij}^1(0, x) dx^i dx^j = h_{ij}^2(0, x) dx^i dx^j,$$

that is, if

$$g_1|_{TS} = g_2|_{TS}.$$

This also guarantees continuity of all tangential derivatives of the metric, but not of the normal derivatives. In order to have a C^1 metric we must have

$$\frac{\partial h_{ij}^1}{\partial t}(0, x) dx^i dx^j = \frac{\partial h_{ij}^2}{\partial t}(0, x) dx^i dx^j,$$

that is,

$$\mathcal{L}_n g_1|_{TS} = \mathcal{L}_n g_2|_{TS}.$$

Note that in this case the curvature tensor (hence the energy-momentum tensor) is at most discontinuous across S . More importantly, as shown in Exercise 8, the components T_{tt} and T_{ti} of the energy-momentum tensor are continuous across S (that is, the flow $T_{\mu\nu} n^\mu$ of energy and momentum across S is equal on both sides), implying that the energy-momentum tensor satisfies the integral version of the conservation equation $\nabla^\mu T_{\mu\nu} = 0$. Therefore we can consider (M, g) a solution of the Einstein equations.

Recall that

$$K_A = \frac{1}{2} \mathcal{L}_n g_A|_{TS_A}$$

is known as the **extrinsic curvature**, or **second fundamental form**, of S_A . We can summarize the discussion above in the following statement.

PROPOSITION 5.3. *Two solutions (M_1, g_1) and (M_2, g_2) of the Einstein field equations can be matched along diffeomorphic timelike boundaries S_1 and S_2 if and only if the induced metrics and second fundamental forms coincide:*

$$g_1 = g_2 \quad \text{and} \quad K_1 = K_2.$$

6. Oppenheimer-Snyder collapse

We can use the matching technique to construct a solution of the Einstein field equations which describes a spherical cloud of dust collapsing to a black hole. This is a physically plausible model for a black hole, as opposed to the eternal black hole.

Let us take (M_1, g_1) to be a flat collapsing FLRW universe:

$$g_1 = -d\tau^2 + a^2(\tau) [d\sigma^2 + \sigma^2 (d\theta^2 + \sin^2 \theta d\varphi^2)].$$

We choose S_1 to be the hypersurface $\sigma = \sigma_0$, with normal vector

$$n = \frac{1}{a} \frac{\partial}{\partial \sigma}.$$

The induced metric then is

$$g_{1|_{TS_1}} = -d\tau^2 + a^2(\tau)\sigma_0^2 (d\theta^2 + \sin^2 \theta d\varphi^2),$$

and the second fundamental form

$$K_1 = a(\tau)\sigma_0 (d\theta^2 + \sin^2 \theta d\varphi^2).$$

Here we used **Cartan's magic formula**: if ω is a differential form and X is a vector field then

$$\mathcal{L}_X \omega = X \lrcorner d\omega + d(X \lrcorner \omega).$$

Thus, for example,

$$\mathcal{L}_n d\tau = n \lrcorner d^2 \tau + d(n \lrcorner d\tau) = 0.$$

Note that the function $a(\tau)$ is constrained by the first Friedmann equation:

$$(10) \quad \dot{a}^2 = \frac{2\alpha}{a}$$

(we assume $\Lambda = 0$).

We now take (M_2, g_2) to be the Schwarzschild solution,

$$g_2 = -V dt^2 + V^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2),$$

where

$$V = 1 - \frac{2M}{r},$$

and choose S_2 to be a spherically symmetric timelike hypersurface given by the parameterization

$$\begin{cases} t = t(\tau) \\ r = r(\tau) \end{cases}.$$

The exact form of the functions $t(\tau)$ and $r(\tau)$ will be fixed by the matching conditions; for the time being they are constrained only by the condition that τ is the proper time along S_2 :

$$-V \dot{t}^2 + V^{-1} \dot{r}^2 = -1.$$

The induced metric is then

$$g_{2|_{TS_2}} = -d\tau^2 + r^2(\tau) (d\theta^2 + \sin^2 \theta d\varphi^2),$$

and so the two induced metrics coincide if and only if

$$(11) \quad r(\tau) = a(\tau)\sigma_0.$$

To simplify the calculation of the second fundamental form, we note that since S_1 is ruled by timelike geodesics, so is S_2 , because the induced metrics and extrinsic curvatures are the same (therefore so are the Christoffel symbols). Therefore $t(\tau)$ and $r(\tau)$ must be a solution of the radial geodesic equations, which are equivalent to

$$(12) \quad \begin{cases} V\dot{t} = E \\ -V\dot{t}^2 + V^{-1}\dot{r}^2 = -1 \end{cases} \Leftrightarrow \begin{cases} V\dot{t} = E \\ \dot{r}^2 = E^2 - 1 + \frac{2M}{r} \end{cases}$$

(where $E > 0$ is a constant). Equations (10), (11) and (12) are compatible if and only if

$$\begin{cases} E = 1 \\ M = \alpha\sigma_0^3 \end{cases},$$

that is, if and only if S_2 represents a spherical shell dropped from infinity with zero velocity and the mass parameter of the Schwarzschild spacetime is related to the density $\rho(\tau)$ of the collapsing dust by

$$M = \frac{4\pi}{3}r^3(\tau)\rho(\tau).$$

To compute the second fundamental form of S_2 we can then consider a family of free-falling spherical shells which includes S_2 . If s is the parameter indexing the shells (with S_2 corresponding to, say, $s = s_0$) then by the Gauss Lemma we can write the Schwarzschild metric in the form

$$g_2 = -d\tau^2 + A^2(\tau, s)ds^2 + r^2(\tau, s)(d\theta^2 + \sin^2\theta d\varphi^2)$$

(consider, for instance, the change of coordinates determined by the solution $t = t(\tau, s)$, $r = r(\tau, s)$ of the radial geodesic equations with initial conditions determined by $t(0, s) = 0$, $r(0, s) = s$, $\dot{r}(0, s) = 0$). The unit normal vector field n to the hypersurfaces of constant s is then

$$n = \frac{1}{A} \frac{\partial}{\partial s},$$

and so we have

$$K_2 = \frac{1}{2}\mathcal{L}_n g_2|_{TS_2} = r(n \cdot r)(d\theta^2 + \sin^2\theta d\varphi^2).$$

On the other hand, in Schwarzschild coordinates

$$n = V^{-1}\dot{r}\frac{\partial}{\partial t} + V\dot{t}\frac{\partial}{\partial r},$$

since n must be unit and orthogonal to

$$\dot{t}\frac{\partial}{\partial t} + \dot{r}\frac{\partial}{\partial r}.$$

Therefore we have

$$K_2 = rV\dot{t}(d\theta^2 + \sin^2\theta d\varphi^2) = rE(d\theta^2 + \sin^2\theta d\varphi^2),$$

or, using $E = 1$,

$$K_2 = r(\tau) (d\theta^2 + \sin^2 \theta d\varphi^2).$$

In other words, $K_1 = K_2$ follows from the previous conditions, and so we indeed have a solution of the Einstein equations.

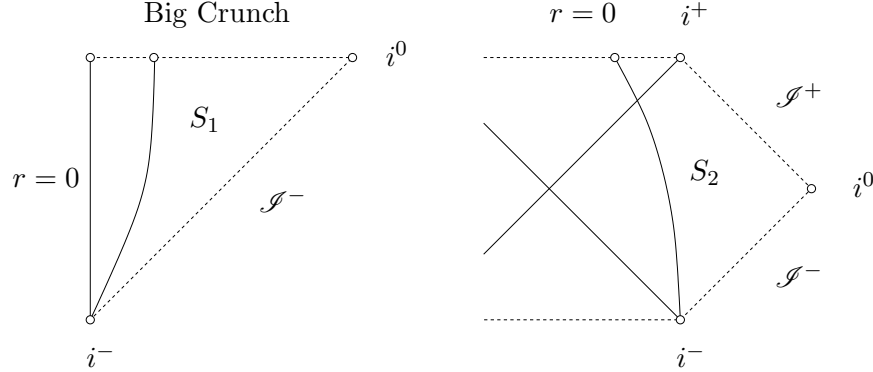


FIGURE 21. Matching hypersurfaces in the collapsing flat FLRW universe and the Schwarzschild solution.

To construct the Penrose diagram for this solution we represent S_1 and S_2 in the Penrose diagrams of the collapsing flat FLRW universe (obtained by reversing the time direction in the expanding flat FLRW universe) and the Schwarzschild solution (Figure 21). Identifying these hypersurfaces results in the Penrose diagram depicted in Figure 22.

7. Exercises

- (1) In this exercise we will solve the vacuum Einstein equations (without cosmological constant) for the spherically symmetric Lorentzian metric given by

$$ds^2 = -(A(t, r))^2 dt^2 + (B(t, r))^2 dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2),$$

where A and B are positive smooth functions.

- (a) Use Cartan's first structure equations,

$$\begin{cases} \omega_{\mu\nu} = -\omega_{\nu\mu} \\ d\omega^\mu + \omega^\mu{}_\nu \wedge \omega^\nu = 0 \end{cases},$$

to show that the nonvanishing connection forms for the orthonormal frame dual to

$$\omega^0 = A dt, \quad \omega^r = B dr, \quad \omega^\theta = r d\theta, \quad \omega^\varphi = r \sin \theta d\varphi$$

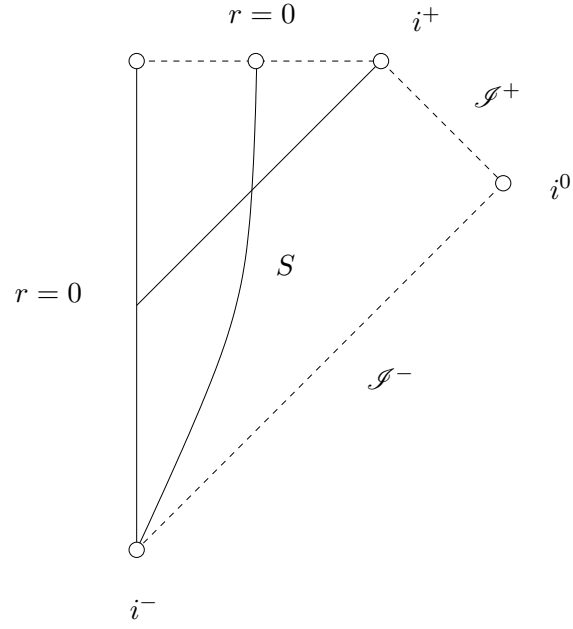


FIGURE 22. Penrose diagram for the Oppenheimer-Snyder collapse.

are (using the notation $\dot{} = \frac{\partial}{\partial t}$ and $\prime = \frac{\partial}{\partial r}$)

$$\omega^0_r = \omega^r_0 = \frac{A'}{B} dt + \frac{\dot{B}}{A} dr;$$

$$\omega^\theta_r = -\omega^r_\theta = \frac{1}{B} d\theta;$$

$$\omega^\varphi_r = -\omega^r_\varphi = \frac{\sin \theta}{B} d\varphi;$$

$$\omega^\varphi_\theta = -\omega^\theta_\varphi = \cos \theta d\varphi.$$

(b) Use Cartan's second structure equations,

$$\Omega^\mu_\nu = d\omega^\mu_\nu + \omega^\mu_\alpha \wedge \omega^\alpha_\nu,$$

to show that the curvature forms on this frame are

$$\begin{aligned}\Omega^0_r = \Omega^r_0 &= \left(\frac{A''B - A'B'}{AB^3} + \frac{\dot{A}\dot{B} - A\ddot{B}}{A^3B} \right) \omega^r \wedge \omega^0; \\ \Omega^0_\theta = \Omega^\theta_0 &= \frac{A'}{rAB^2} \omega^\theta \wedge \omega^0 + \frac{\dot{B}}{rAB^2} \omega^\theta \wedge \omega^r; \\ \Omega^0_\varphi = \Omega^\varphi_0 &= \frac{A'}{rAB^2} \omega^\varphi \wedge \omega^0 + \frac{\dot{B}}{rAB^2} \omega^\varphi \wedge \omega^r; \\ \Omega^\theta_r = -\Omega^r_\theta &= \frac{B'}{rB^3} \omega^\theta \wedge \omega^r + \frac{\dot{B}}{rAB^2} \omega^\theta \wedge \omega^0; \\ \Omega^\varphi_r = -\Omega^r_\varphi &= \frac{B'}{rB^3} \omega^\varphi \wedge \omega^r + \frac{\dot{B}}{rAB^2} \omega^\varphi \wedge \omega^0; \\ \Omega^\varphi_\theta = -\Omega^\theta_\varphi &= \frac{B^2 - 1}{r^2B^2} \omega^\varphi \wedge \omega^\theta.\end{aligned}$$

(c) Using

$$\Omega^\mu_\nu = \sum_{\alpha < \beta} R_{\alpha\beta}{}^\mu{}_\nu \omega^\alpha \wedge \omega^\beta,$$

determine the components $R_{\alpha\beta}{}^\mu{}_\nu$ of the curvature tensor in this orthonormal frame, and show that the nonvanishing components of the Ricci tensor in this frame are

$$\begin{aligned}R_{00} &= \frac{A''B - A'B'}{AB^3} + \frac{\dot{A}\dot{B} - A\ddot{B}}{A^3B} + \frac{2A'}{rAB^2}; \\ R_{0r} = R_{r0} &= \frac{2\dot{B}}{rAB^2}; \\ R_{rr} &= \frac{A'B' - A''B}{AB^3} + \frac{A\ddot{B} - \dot{A}\dot{B}}{A^3B} + \frac{2B'}{rB^3}; \\ R_{\theta\theta} = R_{\varphi\varphi} &= -\frac{A'}{rAB^2} + \frac{B'}{rB^3} + \frac{B^2 - 1}{r^2B^2}.\end{aligned}$$

Conclude that the nonvanishing components of the Einstein tensor in this frame are

$$\begin{aligned}G_{00} &= \frac{2B'}{rB^3} + \frac{B^2 - 1}{r^2B^2}; \\ G_{0r} = G_{r0} &= \frac{2\dot{B}}{rAB^2}; \\ G_{rr} &= \frac{2A'}{rAB^2} - \frac{B^2 - 1}{r^2B^2}; \\ G_{\theta\theta} = G_{\varphi\varphi} &= \frac{A''B - A'B'}{AB^3} + \frac{\dot{A}\dot{B} - A\ddot{B}}{A^3B} + \frac{A'}{rAB^2} - \frac{B'}{rB^3}.\end{aligned}$$

(d) Show that if we write

$$B(t, r) = \left(1 - \frac{2m(t, r)}{r}\right)^{-\frac{1}{2}}$$

for some smooth function m then

$$G_{00} = \frac{2m'}{r^2}.$$

Conclude that the Einstein equations $G_{00} = G_{0r} = 0$ are equivalent to

$$B = \left(1 - \frac{2M}{r}\right)^{-\frac{1}{2}},$$

where $M \in \mathbb{R}$ is an integration constant.

- (e) Show that the Einstein equation $G_{00} + G_{rr} = 0$ is equivalent to $A = \frac{\alpha(t)}{B}$ for some positive smooth function α .
- (f) Check that if A and B are as above then the remaining Einstein equations $G_{\theta\theta} = G_{\varphi\varphi} = 0$ are automatically satisfied.
- (g) Argue that it is always possible to rescale the time coordinate t so that the metric is written

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)$$

(the statement that any spherically symmetric solution of the vacuum Einstein equations without cosmological constant is of this form is known as **Birkhoff's theorem**).

- (2) Show that the Riemannian manifold obtained by gluing the hypersurfaces $t = 0$ of the two exterior regions in the maximally extended Schwarzschild solution along the horizon $r = 2M$ is isometric to the **Flamm paraboloid**, that is, the hypersurface in \mathbb{R}^4 with equation

$$\sqrt{x^2 + y^2 + z^2} = 2M + \frac{w^2}{8M}$$

(Figure 23).

- (3) Recall that the nonvanishing components of the Einstein tensor of the spherically symmetric Lorentzian metric

$$ds^2 = -(A(t, r))^2 dt^2 + (B(t, r))^2 dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)$$

in the orthonormal frame dual to

$$\omega^0 = A dt, \quad \omega^r = B dr, \quad \omega^\theta = r d\theta, \quad \omega^\varphi = r \sin \theta d\varphi,$$

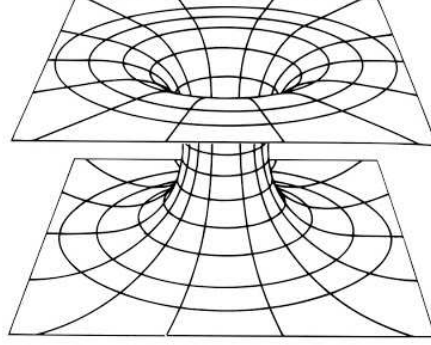


FIGURE 23. Two-dimensional analogue of the Flamm paraboloid.

are given by (using the notation $\dot{} = \frac{\partial}{\partial t}$ and $\prime = \frac{\partial}{\partial r}$)

$$G_{00} = \frac{2B'}{rB^3} + \frac{B^2 - 1}{r^2B^2} = \frac{2m'}{r^2};$$

$$G_{0r} = G_{r0} = \frac{2\dot{B}}{rAB^2};$$

$$G_{rr} = \frac{2A'}{rAB^2} - \frac{B^2 - 1}{r^2B^2};$$

$$G_{\theta\theta} = G_{\varphi\varphi} = \frac{A''B - A'B'}{AB^3} + \frac{\dot{A}\dot{B} - A\ddot{B}}{A^3B} + \frac{A'}{rAB^2} - \frac{B'}{rB^3},$$

where

$$B(t, r) = \left(1 - \frac{2m(t, r)}{r}\right)^{-\frac{1}{2}}.$$

(a) Assuming

- $G_{0r} = 0$ (so that B , and hence m , do not depend on t);
- $G_{00} + G_{rr} = 0$ (so that $A = \frac{\alpha(t)}{B}$ for some positive smooth function $\alpha(t)$);
- $\alpha(t) = 1$ (which can always be achieved by rescaling t),

show that

$$G_{\theta\theta} = G_{\varphi\varphi} = \frac{1}{2} (A^2)'' + \frac{1}{r} (A^2)'.$$

(b) Prove that the general spherically symmetric solution of the vacuum Einstein field equations with a cosmological constant Λ is the **Kottler metric**

$$g = - \left(1 - \frac{2M}{r} - \frac{\Lambda}{3}r^2\right) dt^2 + \left(1 - \frac{2M}{r} - \frac{\Lambda}{3}r^2\right)^{-1} dr^2 \\ + r^2 (d\theta^2 + \sin^2\theta d\varphi^2).$$

- (c) Obtain the Penrose diagram for the maximal extension of the Kottler solution with $\Lambda > 0$ and $0 < M < \frac{1}{3\sqrt{\Lambda}}$.
- (d) Consider now the spherically symmetric electromagnetic field

$$F = E(t, r) \omega^r \wedge \omega^0.$$

Show that this field satisfies the vacuum Maxwell equations

$$dF = d \star F = 0$$

(where \star is the Hodge star) if and only if

$$E(t, r) = \frac{e}{r^2}$$

for some constant $e \in \mathbb{R}$ (the electric charge in units for which $4\pi\epsilon_0 = 1$).

- (e) As we shall see in Chapter 6, this electromagnetic field corresponds to the energy-momentum tensor

$$T = \frac{E^2}{8\pi} \left(\omega^0 \otimes \omega^0 - \omega^r \otimes \omega^r + \omega^\theta \otimes \omega^\theta + \omega^\varphi \otimes \omega^\varphi \right).$$

Prove that the general spherically symmetric solution of the Einstein field equations with an electromagnetic field of this kind is the **Reissner-Nordström metric**

$$g = - \left(1 - \frac{2M}{r} + \frac{e^2}{r^2} \right) dt^2 + \left(1 - \frac{2M}{r} + \frac{e^2}{r^2} \right)^{-1} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2).$$

- (f) Obtain the Penrose diagram for the maximal extension of the Reissner-Nordström solution with $M > 0$ and $0 < e^2 < M^2$.
- (4) Consider the spherically symmetric Lorentzian metric

$$ds^2 = -dt^2 + a^2(t) \left(\frac{1}{1 - kr^2} dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2) \right),$$

where a is a positive smooth function.

- (a) Use Cartan's first structure equations,

$$\begin{cases} \omega_{\mu\nu} = -\omega_{\nu\mu} \\ d\omega^\mu + \omega^\mu{}_\nu \wedge \omega^\nu = 0 \end{cases},$$

to show that the nonvanishing connection forms for the orthonormal frame dual to

$$\begin{aligned} \omega^0 &= dt, & \omega^r &= a(t) (1 - kr^2)^{-\frac{1}{2}} dr, \\ \omega^\theta &= a(t) r d\theta, & \omega^\varphi &= a(t) r \sin \theta d\varphi \end{aligned}$$

are

$$\begin{aligned}
\omega^0_r &= \omega^r_0 = \dot{a} (1 - kr^2)^{-\frac{1}{2}} dr; \\
\omega^0_\theta &= \omega^\theta_0 = \dot{a} r d\theta; \\
\omega^0_\varphi &= \omega^\varphi_0 = \dot{a} r \sin \theta d\varphi; \\
\omega^\theta_r &= -\omega^r_\theta = (1 - kr^2)^{\frac{1}{2}} d\theta; \\
\omega^\varphi_r &= -\omega^r_\varphi = (1 - kr^2)^{\frac{1}{2}} \sin \theta d\varphi; \\
\omega^\varphi_\theta &= -\omega^\theta_\varphi = \cos \theta d\varphi.
\end{aligned}$$

(b) Use Cartan's second structure equations,

$$\Omega^\mu_\nu = d\omega^\mu_\nu + \omega^\mu_\alpha \wedge \omega^\alpha_\nu,$$

to show that the curvature forms on this frame are

$$\begin{aligned}
\Omega^0_r &= \Omega^r_0 = \frac{\ddot{a}}{a} \omega^0 \wedge \omega^r; \\
\Omega^0_\theta &= \Omega^\theta_0 = \frac{\ddot{a}}{a} \omega^0 \wedge \omega^\theta; \\
\Omega^0_\varphi &= \Omega^\varphi_0 = \frac{\ddot{a}}{a} \omega^0 \wedge \omega^\varphi; \\
\Omega^\theta_r &= -\Omega^r_\theta = \left(\frac{k}{a^2} + \frac{\dot{a}^2}{a^2} \right) \omega^\theta \wedge \omega^r; \\
\Omega^\varphi_r &= -\Omega^r_\varphi = \left(\frac{k}{a^2} + \frac{\dot{a}^2}{a^2} \right) \omega^\varphi \wedge \omega^r; \\
\Omega^\varphi_\theta &= -\Omega^\theta_\varphi = \left(\frac{k}{a^2} + \frac{\dot{a}^2}{a^2} \right) \omega^\varphi \wedge \omega^\theta.
\end{aligned}$$

(c) Using

$$\Omega^\mu_\nu = \sum_{\alpha < \beta} R_{\alpha\beta}{}^\mu{}_\nu \omega^\alpha \wedge \omega^\beta,$$

determine the components $R_{\alpha\beta}{}^\mu{}_\nu$ of the curvature tensor on this orthonormal frame, and show that the nonvanishing components of the Ricci tensor on this frame are

$$\begin{aligned}
R_{00} &= -\frac{3\ddot{a}}{a}; \\
R_{rr} &= R_{\theta\theta} = R_{\varphi\varphi} = \frac{\ddot{a}}{a} + \frac{2\dot{a}^2}{a^2} + \frac{2k}{a^2}.
\end{aligned}$$

Conclude that the nonvanishing components of the Einstein tensor on this frame are

$$\begin{aligned}
G_{00} &= \frac{3\dot{a}^2}{a^2} + \frac{3k}{a^2}; \\
G_{rr} &= G_{\theta\theta} = G_{\varphi\varphi} = -\frac{2\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - \frac{k}{a^2}.
\end{aligned}$$

- (d) Show that the Einstein equations with a cosmological constant Λ for a comoving pressureless perfect fluid of nonnegative density ρ , $G + \Lambda g = 8\pi\rho dt^2$, are equivalent to the system

$$\begin{cases} \frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = \frac{8\pi\rho}{3} + \frac{\Lambda}{3} \\ \frac{2\ddot{a}}{a} + \frac{\dot{a}^2}{a^2} + \frac{k}{a^2} = \Lambda \end{cases} .$$

Show that this system can be integrated to

$$\begin{cases} \frac{4\pi\rho}{3}a^3 = \alpha \\ \frac{1}{2}\dot{a}^2 - \frac{\alpha}{a} - \frac{\Lambda}{6}a^2 = -\frac{k}{2} \end{cases} ,$$

where α is a nonnegative integration constant.

- (e) Draw the Penrose diagram of the solutions with $\alpha > 0$, $\Lambda > 0$ and $k = 0$ (currently believed to model our physical Universe).
- (5) Compute the metrics of the following manifolds in the local coordinates indicated, and sketch the corresponding Penrose diagrams:
- (a) The region $T > \sqrt{X^2 + Y^2 + Z^2}$ of the 4-dimensional Minkowski spacetime using the parameterization

$$\begin{cases} T = t \cosh \psi \\ X = t \sinh \psi \sin \theta \cos \varphi \\ Y = t \sinh \psi \sin \theta \sin \varphi \\ Z = t \sinh \psi \cos \theta \end{cases} .$$

- (b) The hyperboloid $X^2 + Y^2 + Z^2 + W^2 = 1 + T^2$ in the 5-dimensional Minkowski spacetime using the parameterization

$$\begin{cases} T = \sinh t \\ X = \cosh t \sin \psi \sin \theta \cos \varphi \\ Y = \cosh t \sin \psi \sin \theta \sin \varphi \\ Z = \cosh t \sin \psi \cos \theta \\ W = \cosh t \cos \psi \end{cases} .$$

- (c) The region $W > 1$ of the same hyperboloid using the parameterization

$$\begin{cases} T = \sinh t \cosh \psi \\ X = \sinh t \sinh \psi \sin \theta \cos \varphi \\ Y = \sinh t \sinh \psi \sin \theta \sin \varphi \\ Z = \sinh t \sinh \psi \cos \theta \\ W = \cosh t \end{cases} .$$

- (d) The region $T > W$ of the same hyperboloid using the parameterization defined implicitly by

$$\begin{cases} T = W + e^t \\ X = e^t x \\ Y = e^t y \\ Z = e^t z \end{cases} .$$

- (e) The hyperboloid $T^2 + U^2 = 1 + X^2 + Y^2 + Z^2$ in \mathbb{R}^5 with the pseudo-Riemannian metric

$$ds^2 = -dT^2 - dU^2 + dX^2 + dY^2 + dZ^2$$

using the parameterization

$$\begin{cases} T = \cosh \psi \cos t \\ U = \cosh \psi \sin t \\ X = \sinh \psi \sin \theta \cos \varphi \\ Y = \sinh \psi \sin \theta \sin \varphi \\ Z = \sinh \psi \cos \theta \end{cases} .$$

- (6) Show that the anti-de Sitter metric

$$ds^2 = -\cosh^2 \psi dt^2 + d\psi^2 + \sinh^2 \psi (d\theta^2 + \sin^2 \theta d\varphi^2)$$

is a solution of the vacuum Einstein field equations with cosmological constant $\Lambda = -3$.

- (7) The 3-dimensional anti-de Sitter space can be obtained by “unwrapping” the hyperboloid $T^2 + U^2 = 1 + X^2 + Y^2$ in \mathbb{R}^4 with the pseudo-Riemannian metric

$$ds^2 = -dT^2 - dU^2 + dX^2 + dY^2.$$

In this exercise we identify \mathbb{R}^4 with the space of 2×2 matrices by the map

$$(T, U, X, Y) \mapsto \begin{pmatrix} T + X & U + Y \\ -U + Y & T - X \end{pmatrix}.$$

- (a) Show that the hyperboloid corresponds to the Lie group $SL(2, \mathbb{R})$ of 2×2 matrices with unit determinant.
- (b) Check that the squared norm of a vector $v \in \mathbb{R}^4$ in the metric above is $\langle v, v \rangle = -\det V$, where V is the 2×2 matrix associated to v . Conclude that the metric induced on the hyperboloid is bi-invariant (that is, invariant under left and right multiplication).
- (c) Use the Penrose diagram in Figure 24 and the fact that the one-parameter subgroups of a Lie group with a bi-invariant metric are geodesics of that metric to conclude that the exponential map $\exp : \mathfrak{sl}(2, \mathbb{R}) \rightarrow SL(2, \mathbb{R})$ is not surjective.

- (d) Write explicitly the matrices in $SL(2, \mathbb{R})$ which are not in the image of $\exp : \mathfrak{sl}(2, \mathbb{R}) \rightarrow SL(2, \mathbb{R})$ by using the parameterization

$$\begin{cases} T = \cosh \psi \cos t \\ U = \cosh \psi \sin t \\ X = \sinh \psi \cos \varphi \\ Y = \sinh \psi \sin \varphi \end{cases} .$$

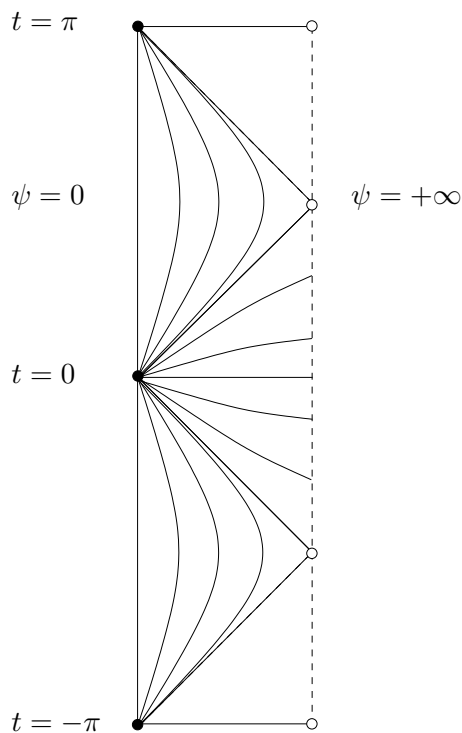


FIGURE 24. Exponential map on $SL(2, \mathbb{R})$.

- (8) Consider a Riemannian or Lorentzian metric given in the Gauss Lemma form

$$g = dt^2 + h_{ij}(t, x)dx^i dx^j,$$

so that the level sets of t are Riemannian or Lorentzian manifolds with induced metric $h(t) = h_{ij}dx^i dx^j$ and second fundamental form

$$K(t) = \frac{1}{2} \frac{\partial h_{ij}}{\partial t} dx^i dx^j.$$

Show that in these coordinates:

- (a) The Christoffel symbols are

$$\Gamma_{ij}^0 = -K_{ij}; \quad \Gamma_{jk}^i = \bar{\Gamma}_{jk}^i; \quad \Gamma_{0j}^i = K^i_j,$$

where $\bar{\Gamma}_{jk}^i$ are the Christoffel symbols of h .

(b) The components of the Riemann tensor are

$$\begin{aligned} R_{0i0}{}^j &= -\frac{\partial}{\partial t} K^j{}_i - K_{il} K^{lj}; \\ R_{ij0}{}^l &= -\bar{\nabla}_i K^l{}_j + \bar{\nabla}_j K^l{}_i; \\ R_{ijl}{}^m &= \bar{R}_{ijl}{}^m - K_{il} K^m{}_j + K_{jl} K^m{}_i, \end{aligned}$$

where $\bar{\nabla}$ is the Levi-Civita connection of h and $\bar{R}_{ijl}{}^m$ are the components of the Riemann tensor of h .

(c) The time derivative of the inverse metric is given by the formula

$$\frac{\partial h^{ij}}{\partial t} = -2K^{ij}.$$

(d) The components of the Ricci tensor are

$$\begin{aligned} R_{00} &= -\frac{\partial}{\partial t} K^i{}_i - K_{ij} K^{ij}; \\ R_{0i} &= -\bar{\nabla}_i K^j{}_j + \bar{\nabla}_j K^j{}_i; \\ R_{ij} &= \bar{R}_{ij} - \frac{\partial}{\partial t} K_{ij} + 2K_{il} K^l{}_j - K^l{}_i K_{lj}, \end{aligned}$$

where \bar{R}_{ij} are the components of the Ricci tensor of h .

(e) The scalar curvature is

$$R = \bar{R} - 2\frac{\partial}{\partial t} K^i{}_i - (K^i{}_i)^2 - K_{ij} K^{ij},$$

where \bar{R} is the scalar curvature of h .

(f) The component G_{00} of the Einstein tensor is

$$G_{00} = \frac{1}{2} \left(-\bar{R} + (K^i{}_i)^2 - K_{ij} K^{ij} \right).$$

This shows that the matching conditions guarantee the continuity of G_{00} and $G_{0i} = R_{0i}$.

(9) Recall that the nonvanishing components of the Einstein tensor of the static, spherically symmetric Lorentzian metric

$$ds^2 = -(A(r))^2 dt^2 + (B(r))^2 dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2)$$

in the orthonormal frame dual to

$$\omega^0 = A dt, \quad \omega^r = B dr, \quad \omega^\theta = r d\theta, \quad \omega^\varphi = r \sin \theta d\varphi,$$

are given by

$$\begin{aligned} G_{00} &= \frac{2B'}{rB^3} + \frac{B^2 - 1}{r^2B^2} = \frac{2m'}{r^2}; \\ G_{rr} &= \frac{2A'}{rAB^2} - \frac{B^2 - 1}{r^2B^2}; \\ G_{\theta\theta} = G_{\varphi\varphi} &= \frac{A''B - A'B'}{AB^3} + \frac{A'}{rAB^2} - \frac{B'}{rB^3}, \end{aligned}$$

where

$$B(r) = \left(1 - \frac{2m(r)}{r}\right)^{-\frac{1}{2}}.$$

In this exercise we will solve the Einstein equations (without cosmological constant) for a static perfect fluid of constant rest density ρ and rest pressure p , and match it to a Schwarzschild exterior.

(a) Show that

$$B(r) = (1 - kr^2)^{-\frac{1}{2}},$$

where $k = \frac{8\pi}{3}\rho$. Conclude that the spatial metric is that of a sphere S^3 with radius $\frac{1}{\sqrt{k}}$.

(b) Solve the ordinary differential equation $G_{rr} = G_{\theta\theta}$ to obtain

$$A(r) = C(1 - kr^2)^{\frac{1}{2}} + D,$$

where $C, D \in \mathbb{R}$ are integration constants.

(c) Show that the matching conditions to a Schwarzschild exterior of mass $M > 0$ across a surface $r = R$ are

$$\begin{cases} A(R) = \left(1 - \frac{2M}{R}\right)^{\frac{1}{2}} \\ A'(R) = \frac{M}{R^2} \left(1 - \frac{2M}{R}\right)^{-\frac{1}{2}} \\ B(R) = \left(1 - \frac{2M}{R}\right)^{-\frac{1}{2}} \end{cases}.$$

(d) Conclude that

$$A(r) = \frac{3}{2} \left(1 - \frac{2M}{R}\right)^{\frac{1}{2}} - \frac{1}{2} \left(1 - \frac{2Mr^2}{R^3}\right)^{\frac{1}{2}}.$$

(e) Show that

$$p(r) = \frac{k(1 - kr^2)^{\frac{1}{2}}}{\frac{3}{2}(1 - kR^2)^{\frac{1}{2}} - \frac{1}{2}(1 - kr^2)^{\frac{1}{2}}} - k.$$

What is the value of $p(R)$?

(f) Show that M and R must satisfy **Buchdahl's limit**:

$$\frac{2M}{R} < \frac{8}{9}.$$

What happens to $p(0)$ as $\frac{2M}{R} \rightarrow \frac{8}{9}$?

CHAPTER 3

Causality

In this chapter we briefly discuss the causality theory of a Lorentzian manifold, following [GN14]. We take a minimal approach; more details can be found in [O’N83, Wal84, Pen87, Nab88, HE95, Rin09].

1. Past and future

A spacetime (M, g) is said to be **time-orientable** if there exists a timelike vector field, that is, a vector field X satisfying $g(X, X) < 0$. In this case, we can define a time orientation on each tangent space $T_p M$ by declaring causal vectors $v \in T_p M$ to be **future-pointing** if $g(v, X_p) \leq 0$. It can be shown that any non-time-orientable spacetime admits a **time-orientable double covering** (just like any non-orientable manifold admits an orientable double covering).

Assume that (M, g) is **time-oriented** (i.e. time-orientable with a definite choice of time orientation). A timelike or causal curve $c : I \subset \mathbb{R} \rightarrow M$ is said to be **future-directed** if \dot{c} is future-pointing. The **chronological future** of $p \in M$ is the set $I^+(p)$ of all points to which p can be connected by a future-directed timelike curve. The **causal future** of $p \in M$ is the set $J^+(p)$ of all points to which p can be connected by a future-directed causal curve. Notice that $I^+(p)$ is simply the set of all events which are accessible to a particle with nonzero mass at p , whereas $J^+(p)$ is the set of events which can be causally influenced by p (as this causal influence cannot propagate faster than the speed of light). Analogously, the **chronological past** of $p \in M$ is the set $I^-(p)$ of all points which can be connected to p by a future-directed timelike curve, and the **causal past** of $p \in M$ is the set $J^-(p)$ of all points which can be connected to p by a future-directed causal curve.

In general, the chronological and causal pasts and futures can be quite complicated sets, because of global features of the spacetime. Locally, however, causal properties are similar to those of Minkowski spacetime. More precisely, we have the following statement:

PROPOSITION 1.1. *Let (M, g) be a time-oriented spacetime. Then each point $p_0 \in M$ has an open neighborhood $V \subset M$ such that the spacetime (V, g) obtained by restricting g to V satisfies:*

- (1) V is **geodesically convex**, that is, V is a **normal neighborhood** of each of its points such that given $p, q \in V$ there exists a unique geodesic (up to reparameterization) connecting p to q ;
- (2) $q \in I^+(p)$ if and only if there exists a future-directed timelike geodesic connecting p to q ;
- (3) $J^+(p) = \overline{I^+(p)}$;
- (4) $q \in J^+(p) \setminus I^+(p)$ if and only if there exists a future-directed null geodesic connecting p to q .

PROOF. Recall that the **exponential map** $\exp_p : U \subset T_p M \rightarrow M$ is the map given by

$$\exp_p(v) = c_v(1),$$

where c_v is the geodesic with initial conditions $c_v(0) = p$, $\dot{c}_v(0) = v$; equivalently, $\exp_p(tv) = c_v(t)$ (since $c_{tv}(1) = c_v(t)$). Recall also that V is a **normal neighborhood** of p if $\exp_p : U \rightarrow V$ is a diffeomorphism. The existence of geodesically convex neighborhoods is true for any affine connection and is proved for instance in [KN96].

To prove assertion (2), we start by noticing that if there exists a future-directed timelike geodesic connecting p to q then it is obvious that $q \in I^+(p)$. Suppose now that $q \in I^+(p)$; then there exists a future-directed timelike curve $c : [0, 1] \rightarrow V$ such that $c(0) = p$ and $c(1) = q$. Choose **normal coordinates** (x^0, x^1, x^2, x^3) , given by the parameterization

$$\varphi(x^0, x^1, x^2, x^3) = \exp_p(x^0 E_0 + x^1 E_1 + x^2 E_2 + x^3 E_3),$$

where $\{E_0, E_1, E_2, E_3\}$ is an orthonormal basis of $T_p M$ with E_0 timelike and future-pointing. These are global coordinates in V , since $\exp_p : U \rightarrow V$ is a diffeomorphism. Defining

$$\begin{aligned} W_p(q) &= -(x^0(q))^2 + (x^1(q))^2 + (x^2(q))^2 + (x^3(q))^2 \\ &= \sum_{\mu, \nu=0}^3 \eta_{\mu\nu} x^\mu(q) x^\nu(q), \end{aligned}$$

with $(\eta_{\mu\nu}) = \text{diag}(-1, 1, 1, 1)$, we have to show that $W_p(q) < 0$. Let $W_p(t) = W_p(c(t))$. Since $x^\mu(p) = 0$, we have $W_p(0) = 0$. Setting $x^\mu(t) = x^\mu(c(t))$, we obtain

$$\begin{aligned} \dot{W}_p(t) &= 2 \sum_{\mu, \nu=0}^3 \eta_{\mu\nu} x^\mu(t) \dot{x}^\nu(t); \\ \ddot{W}_p(t) &= 2 \sum_{\mu, \nu=0}^3 \eta_{\mu\nu} x^\mu(t) \ddot{x}^\nu(t) + 2 \sum_{\mu, \nu=0}^3 \eta_{\mu\nu} \dot{x}^\mu(t) \dot{x}^\nu(t), \end{aligned}$$

and consequently

$$\begin{aligned} \dot{W}_p(0) &= 0; \\ \ddot{W}_p(0) &= 2 \langle \dot{c}(0), \dot{c}(0) \rangle < 0. \end{aligned}$$

Therefore there exists $\varepsilon > 0$ such that $W_p(t) < 0$ for $t \in (0, \varepsilon)$.

By the Gauss Lemma, the level surfaces of W_p are orthogonal to the geodesics through p . Therefore, if $c_v(t) = \exp_p(tv)$ is the geodesic with initial condition $v \in T_pM$, we have

$$(\text{grad } W_p)_{c_v(1)} = a(v)\dot{c}_v(1).$$

Now

$$\begin{aligned} \langle (\text{grad } W_p)_{c_v(t)}, \dot{c}_v(t) \rangle &= \frac{d}{dt} W_p(c_v(t)) = \frac{d}{dt} \langle tv, tv \rangle \\ &= \frac{d}{dt} (t^2 \langle v, v \rangle) = 2t \langle v, v \rangle, \end{aligned}$$

and hence

$$\langle (\text{grad } W_p)_{c_v(1)}, \dot{c}_v(1) \rangle = 2 \langle v, v \rangle.$$

On the other hand,

$$\langle (\text{grad } W_p)_{c_v(1)}, \dot{c}_v(1) \rangle = \langle a(v)\dot{c}_v(1), \dot{c}_v(1) \rangle = a(v) \langle v, v \rangle.$$

We conclude that $a(v) = 2$, and therefore

$$(\text{grad } W_p)_{c_v(1)} = 2\dot{c}_v(1).$$

Consequently, $\text{grad } W_p$ is tangent to geodesics through p , being future-pointing on future-directed geodesics.

Suppose that $W_p(t) < 0$. Then $(\text{grad } W_p)_{c(t)}$ is timelike future-pointing, and so

$$\dot{W}(t) = \langle (\text{grad } W_p)_{c(t)}, \dot{c}(t) \rangle < 0,$$

as $\dot{c}(t)$ is also timelike future-pointing. We conclude that we must have $W_p(t) < 0$ for all $t \in [0, 1]$. In particular, $W_p(q) = W_p(1) < 0$, and hence there exists a future-directed timelike geodesic connecting p to q .

To prove assertion (3), let us see first that $\overline{I^+(p)} \subset J^+(p)$. If $q \in \overline{I^+(p)}$, then q is the limit of a sequence of points $q_n \in I^+(p)$. By (2), $q_n = \exp_p(v_n)$ with $v_n \in T_pM$ timelike future-pointing. Since \exp_p is a diffeomorphism, v_n converges to a causal future-pointing vector $v \in T_pM$, and so $q = \exp_p(v)$ can be reached from p by a future-directed causal geodesic. The converse inclusion $J^+(p) \subset \overline{I^+(p)}$ holds in general (cf. Proposition 1.2).

Finally, (4) is obvious from (3) and the fact that \exp_p is a diffeomorphism onto V . \square

This local behavior can be used to prove the following global result.

PROPOSITION 1.2. *Let (M, g) be a time oriented spacetime and $p \in M$. Then:*

- (1) $I^+(p)$ is open;
- (2) $J^+(p) \subset \overline{I^+(p)}$;
- (3) $I^+(p) = \text{int } J^+(p)$
- (4) if $r \in J^+(p)$ and $q \in I^+(r)$ then $q \in I^+(p)$;
- (5) if $r \in I^+(p)$ and $q \in J^+(r)$ then $q \in I^+(p)$.

PROOF. Exercise. \square

The twin paradox also holds locally for general spacetimes. More precisely, we have the following statement:

PROPOSITION 1.3. *Let (M, g) be a time-oriented spacetime, $p_0 \in M$ and $V \subset M$ a geodesically convex open neighborhood of p_0 . The spacetime (V, g) obtained by restricting g to V satisfies the following property: if $p, q \in V$ with $q \in I^+(p)$, c is the timelike geodesic connecting p to q and γ is any timelike curve connecting p to q , then $\tau(\gamma) \leq \tau(c)$, with equality if and only if γ is a reparameterization of c .*

PROOF. Any timelike curve $\gamma : [0, 1] \rightarrow V$ satisfying $\gamma(0) = p$, $\gamma(1) = q$ can be written as

$$\gamma(t) = \exp_p(r(t)n(t)),$$

for $t \in [0, 1]$, where $r(t) \geq 0$ and $\langle n(t), n(t) \rangle = -1$. We have

$$\dot{\gamma}(t) = (\exp_p)_* (\dot{r}(t)n(t) + r(t)\dot{n}(t)).$$

Since $\langle n(t), n(t) \rangle = -1$, we have $\langle \dot{n}(t), n(t) \rangle = 0$, and consequently $\dot{n}(t)$ is tangent to the level surfaces of the function $v \mapsto \langle v, v \rangle$. We conclude that

$$\dot{\gamma}(t) = \dot{r}(t)X_{\gamma(t)} + Y(t),$$

where X is the unit tangent vector field to timelike geodesics through p and $Y(t) = r(t)(\exp_p)_*\dot{n}(t)$ is tangent to the level surfaces of W_p (hence orthogonal to $X_{\gamma(t)}$). Consequently,

$$\begin{aligned} \tau(\gamma) &= \int_0^1 |\langle \dot{r}(t)X_{\gamma(t)} + Y(t), \dot{r}(t)X_{\gamma(t)} + Y(t) \rangle|^{\frac{1}{2}} dt \\ &= \int_0^1 (\dot{r}(t)^2 - |Y(t)|^2)^{\frac{1}{2}} dt \\ &\leq \int_0^1 \dot{r}(t) dt = r(1) = \tau(c), \end{aligned}$$

where we have used the facts that γ is timelike, $\dot{r}(t) > 0$ for all $t \in [0, 1]$ (as $\dot{\gamma}$ is future-pointing) and $\tau(c) = r(1)$ (as $q = \exp_p(r(1)n(1))$). It should be clear that $\tau(\gamma) = \tau(c)$ if and only if $|Y(t)| \equiv 0 \Leftrightarrow Y(t) \equiv 0$ ($Y(t)$ is spacelike or zero) for all $t \in [0, 1]$, implying that n is constant. In this case, $\gamma(t) = \exp_p(r(t)n)$ is, up to reparameterization, the geodesic through p with initial condition $n \in T_p M$. \square

There is also a local property characterizing null geodesics.

PROPOSITION 1.4. *Let (M, g) be a time-oriented spacetime, $p_0 \in M$ and $V \subset M$ a geodesically convex open neighborhood of p_0 . The spacetime (V, g) obtained by restricting g to V satisfies the following property: if for $p, q \in V$ there exists a future-directed null geodesic c connecting p to q and γ is a causal curve connecting p to q then γ is a reparameterization of c .*

PROOF. Since p and q are connected by a null geodesic, we conclude from Proposition 1.1 that $q \in J^+(p) \setminus I^+(p)$. Let $\gamma : [0, 1] \rightarrow V$ be a causal curve connecting p to q . Then we must have $\gamma(t) \in J^+(p) \setminus I^+(p)$ for all $t \in [0, 1]$, since $\gamma(t_0) \in I^+(p)$ implies $\gamma(t) \in I^+(p)$ for all $t > t_0$ (see Proposition 1.2). Consequently, we have

$$W_p(\gamma(t)) = 0 \Rightarrow \left\langle (\text{grad } W_p)_{\gamma(t)}, \dot{\gamma}(t) \right\rangle = 0,$$

where W_p was defined in the proof of Proposition 1.1. The formula

$$(\text{grad } W_p)_{c_v(1)} = 2\dot{c}_v(1),$$

which was proved for timelike geodesics c_v with initial condition $v \in T_pM$, must also hold for null geodesics (by continuity). Hence $\text{grad } W_p$ is tangent to the null geodesics ruling $J^+(p) \setminus I^+(p)$ and future-pointing. Since $\dot{\gamma}(t)$ is also future-pointing, we conclude that $\dot{\gamma}$ is proportional to $\text{grad } W_p$, and therefore γ is a reparameterization of a null geodesic, which must be c . \square

COROLLARY 1.5. *Let (M, g) be a time-oriented spacetime and $p \in M$. If $q \in J^+(p) \setminus I^+(p)$ then any future-directed causal curve connecting p to q must be a reparameterized null geodesic.*

2. Causality conditions

For physical applications, it is important to require that the spacetime satisfies reasonable causality conditions. The simplest of these conditions excludes time travel, i.e. the possibility of a particle returning to an event in its past history.

DEFINITION 2.1. *A spacetime (M, g) is said to satisfy the **chronology condition** if it does not contain closed timelike curves.*

This condition is violated by compact spacetimes:

PROPOSITION 2.2. *Any compact spacetime (M, g) contains closed timelike curves.*

PROOF. Taking if necessary the time-orientable double covering, we can assume that (M, g) is time-oriented. Since $I^+(p)$ is an open set for any $p \in M$, it is clear that $\{I^+(p)\}_{p \in M}$ is an open cover of M . If M is compact, we can obtain a finite subcover $\{I^+(p_1), \dots, I^+(p_N)\}$. Now if $p_1 \in I^+(p_i)$ for $i \neq 1$ then $I^+(p_1) \subset I^+(p_i)$, and we can exclude $I^+(p_1)$ from the subcover. Therefore, we can assume without loss of generality that $p_1 \in I^+(p_1)$, and hence there exists a closed timelike curve starting and ending at p_1 . \square

A stronger restriction on the causal behavior of the spacetime is the following:

DEFINITION 2.3. *A spacetime (M, g) is said to be **stably causal** if there exists a **global time function**, i.e. a smooth function $t : M \rightarrow \mathbb{R}$ such that $\text{grad}(t)$ is timelike.*

In particular, a stably causal spacetime is time-orientable. We choose the time orientation defined by $-\text{grad}(t)$, so that t increases along future-directed timelike curves. Notice that this implies that no closed timelike curves can exist, i.e. any stably causal spacetime satisfies the chronology condition. In fact, any small perturbation of a stably causal spacetime still satisfies the chronology condition (Exercise 4).

Let (M, g) be a time-oriented spacetime. A smooth future-directed causal curve $c : (a, b) \rightarrow M$ (with possibly $a = -\infty$ or $b = +\infty$) is said to be **future-inextendible** if $\lim_{t \rightarrow b} c(t)$ does not exist. The definition of a **past-inextendible** causal curve is analogous. The **future domain of dependence** of $S \subset M$ is the set $D^+(S)$ of all events $p \in M$ such that any past-inextendible causal curve starting at p intersects S . Therefore any causal influence on an event $p \in D^+(S)$ had to register somewhere in S , and one can expect that what happens at p can be predicted from data on S . Similarly, the **past domain of dependence** of S is the set $D^-(S)$ of all events $p \in M$ such that any future-inextendible causal curve starting at p intersects S . Therefore any causal influence of an event $p \in D^-(S)$ will register somewhere in S , and one can expect that what happened at p can be retrodicted from data on S . The **domain of dependence** of S is simply the set $D(S) = D^+(S) \cup D^-(S)$.

Let (M, g) be a stably causal spacetime with time function $t : M \rightarrow \mathbb{R}$. The level sets $S_a = t^{-1}(a)$ are said to be **Cauchy hypersurfaces** if $D(S_a) = M$. Spacetimes for which this happens have particularly good causal properties.

DEFINITION 2.4. *A stably causal spacetime possessing a time function whose level sets are Cauchy hypersurfaces is said to be **globally hyperbolic**.*

Notice that the future and past domains of dependence of the Cauchy hypersurfaces S_a are $D^+(S_a) = t^{-1}([a, +\infty))$ and $D^-(S_a) = t^{-1}((-\infty, a])$.

3. Exercises

- (1) Let (M, g) be the quotient of the 2-dimensional Minkowski spacetime by the discrete group of isometries generated by the map $f(t, x) = (-t, x + 1)$. Show that (M, g) is not time orientable.
- (2) Let (M, g) be a time oriented spacetime and $p \in M$. Show that:
 - (a) $I^+(p)$ is open;
 - (b) $J^+(p)$ is not necessarily closed;
 - (c) $J^+(p) \subset \overline{I^+(p)}$;
 - (d) $I^+(p) = \text{int } J^+(p)$
 - (e) if $r \in J^+(p)$ and $q \in I^+(r)$ then $q \in I^+(p)$;
 - (f) if $r \in I^+(p)$ and $q \in J^+(r)$ then $q \in I^+(p)$.
- (3) Consider the 3-dimensional Minkowski spacetime (\mathbb{R}^3, g) , where

$$g = -dt^2 + dx^2 + dy^2.$$

Let $c : \mathbb{R} \rightarrow \mathbb{R}^3$ be the curve $c(t) = (t, \cos t, \sin t)$. Show that although $\dot{c}(t)$ is null for all $t \in \mathbb{R}$ we have $c(t) \in I^+(c(0))$ for all $t > 0$. What kind of motion does this curve represent?

- (4) Let (M, g) be a stably causal spacetime and h an arbitrary symmetric $(2, 0)$ -tensor field with compact support. Show that for sufficiently small $|\varepsilon|$ the tensor field $g_\varepsilon = g + \varepsilon h$ is still a Lorentzian metric on M , and (M, g_ε) satisfies the chronology condition.
- (5) Let (M, g) be the quotient of the 2-dimensional Minkowski spacetime by the discrete group of isometries generated by the map $f(t, x) = (t + 1, x + 1)$. Show that (M, g) satisfies the chronology condition, but there exist arbitrarily small perturbations of (M, g) (in the sense of Exercise 4) which do not.
- (6) Let (M, g) be a time oriented spacetime and $S \subset M$. Show that:
 - (a) $S \subset D^+(S)$;
 - (b) $D^+(S)$ is not necessarily open;
 - (c) $D^+(S)$ is not necessarily closed.
- (7) Let (M, g) be the 2-dimensional spacetime obtained by removing the positive x -semi-axis of Minkowski 2-dimensional spacetime (cf. Figure 1). Show that:
 - (a) (M, g) is stably causal but not globally hyperbolic;
 - (b) there exist points $p, q \in M$ such that $J^+(p) \cap J^-(q)$ is not compact;
 - (c) there exist points $p, q \in M$ with $q \in I^+(p)$ such that the supremum of the lengths of timelike curves connecting p to q is not attained by any timelike curve.

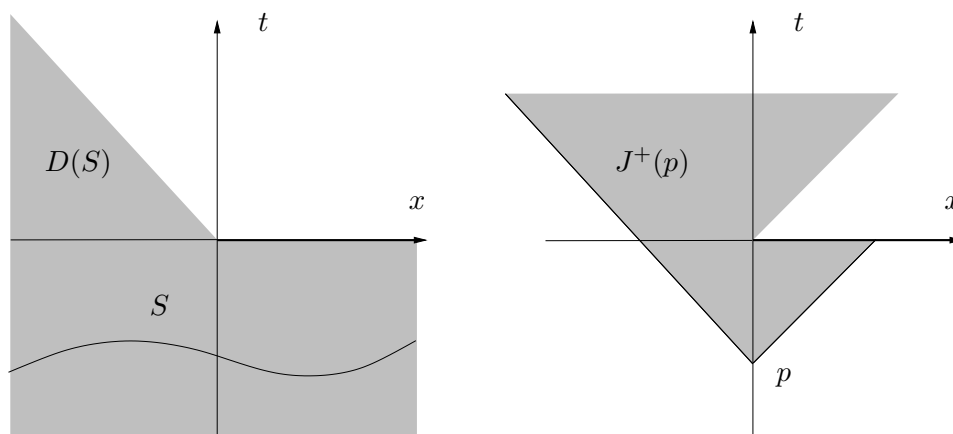


FIGURE 1. Stably causal but not globally hyperbolic spacetime.

- (8) Let (Σ, h) be a 3-dimensional Riemannian manifold. Show that the spacetime $(M, g) = (\mathbb{R} \times \Sigma, -dt \otimes dt + h)$ is globally hyperbolic if and only if (Σ, h) is complete.
- (9) Show that the following spacetimes are globally hyperbolic:

- (a) the Minkowski spacetime;
 - (b) the FLRW spacetimes;
 - (c) the region $\{r > 2m\}$ of the Schwarzschild spacetime;
 - (d) the region $\{r < 2m\}$ of the Schwarzschild spacetime;
 - (e) the maximal analytical extension of the Schwarzschild spacetime.
- (10) Let (M, g) be a global hyperbolic spacetime with Cauchy hypersurface S . Show that M is diffeomorphic to $\mathbb{R} \times S$.

CHAPTER 4

Singularity theorems

As we have seen in Chapter 2, both the Schwarzschild solution and the FLRW cosmological models display singularities, beyond which timelike and null geodesics cannot be continued. It was once thought that these solutions were singular due to their high degree of symmetry, and that more realistic spacetimes would be non-singular. In this chapter we show that this is not the case: any sufficiently small perturbation of these solutions will still be singular. We follow [Wal84] when discussing conjugate points and [GN14] for the details of the proofs. See also [O’N83, Pen87, Nab88, HE95].

1. Geodesic congruences

Let (M, g) be a Lorentzian manifold. A **congruence** of curves in an open set $U \subset M$ is the family of integral curves of a nonvanishing vector field X in U . We will assume that X is unit timelike and geodesic, that is,

$$\langle X, X \rangle = -1 \quad \text{and} \quad \nabla_X X = 0.$$

The properties of the congruence determined by X are best analyzed by considering its **second fundamental form**

$$B_{\mu\nu} = \nabla_\nu X_\mu.$$

This tensor is **purely spatial**, that is,

$$B_{\mu\nu} X^\mu = B_{\mu\nu} X^\nu = 0.$$

Indeed, since X is unit,

$$B_{\mu\nu} X^\mu = X^\mu \nabla_\nu X_\mu = \frac{1}{2} \nabla_\nu (X_\mu X^\mu) = 0.$$

On the other hand, because X is geodesic,

$$B_{\mu\nu} X^\nu = X^\nu \nabla_\nu X_\mu = \nabla_X X_\mu = 0.$$

PROPOSITION 1.1. *The second fundamental form B satisfies*

$$\nabla_X B_{\mu\nu} = -B_{\mu\alpha} B^\alpha{}_\nu + R_{\alpha\nu\mu\beta} X^\alpha X^\beta.$$

PROOF. We have

$$\begin{aligned} X^\alpha \nabla_\alpha B_{\mu\nu} &= X^\alpha \nabla_\alpha \nabla_\nu X_\mu = X^\alpha \nabla_\nu \nabla_\alpha X_\mu + X^\alpha R_{\alpha\nu\mu\beta} X^\beta \\ &= \nabla_\nu (X^\alpha \nabla_\alpha X_\mu) - (\nabla_\nu X^\alpha) (\nabla_\alpha X_\mu) + R_{\alpha\nu\mu\beta} X^\alpha X^\beta \\ &= -B_{\mu\alpha} B^\alpha{}_\nu + R_{\alpha\nu\mu\beta} X^\alpha X^\beta. \end{aligned}$$

□

Let $c(t, s)$ be a one-parameter family of geodesics of the congruence, parameterized such that

$$\frac{\partial c}{\partial t} = X$$

(Figure 1). The **geodesic deviation vector** associated to c is

$$Y = \frac{\partial c}{\partial s}.$$

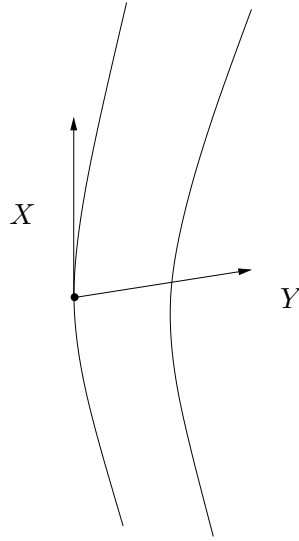


FIGURE 1. Geodesic deviation.

PROPOSITION 1.2. *The geodesic deviation vector satisfies*

$$\nabla_X Y^\mu = B^\mu{}_\nu Y^\nu.$$

PROOF. The definition of Y implies that

$$[X, Y] = 0 \Leftrightarrow \nabla_X Y - \nabla_Y X = 0.$$

Consequently, we have

$$\nabla_X Y^\mu = \nabla_Y X^\mu = Y^\nu \nabla_\nu X^\mu = B^\mu{}_\nu Y^\nu.$$

□

The equation for $\nabla_X B_{\mu\nu}$ then yields the following famous result.

PROPOSITION 1.3. *The geodesic deviation vector satisfies the **Jacobi equation***

$$\nabla_X \nabla_X Y = R(X, Y)X.$$

PROOF. We have

$$\begin{aligned}\nabla_X \nabla_X Y^\alpha &= \nabla_X (B^\alpha_\beta Y^\beta) = (\nabla_X B^\alpha_\beta) Y^\beta + B^\alpha_\beta \nabla_X Y^\beta \\ &= -B^{\alpha\mu} B_{\mu\beta} Y^\beta + R_{\mu\beta}{}^\alpha{}_\nu X^\mu X^\nu Y^\beta + B^\alpha_\beta B^\beta_\mu Y^\mu \\ &= R^\alpha{}_{\beta\mu\nu} X^\beta X^\mu Y^\nu.\end{aligned}$$

Alternatively, we can simply notice that

$$\nabla_X \nabla_X Y = \nabla_X \nabla_Y X = \nabla_Y \nabla_X X + R(X, Y)X = R(X, Y)X.$$

□

We now define the kinematic quantities associated to the congruence.

DEFINITION 1.4. *The **spatial metric** associated to the congruence is*

$$h_{\mu\nu} = g_{\mu\nu} + X_\mu X_\nu.$$

*The **expansion, shear and vorticity** are defined as¹*

$$\theta = h^{\mu\nu} B_{\mu\nu} = g^{\mu\nu} B_{\mu\nu},$$

$$\sigma_{\mu\nu} = B_{(\mu\nu)} - \frac{1}{3}\theta h_{\mu\nu},$$

$$\omega_{\mu\nu} = B_{[\mu\nu]},$$

so that we have the decomposition

$$B_{\mu\nu} = \frac{1}{3}\theta h_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu}.$$

Note that all the tensors above are purely spatial:

$$h_{\mu\nu} X^\nu = \sigma_{\mu\nu} X^\nu = \omega_{\mu\nu} X^\nu = 0,$$

Moreover, the trace of h is

$$h^{\mu\nu} h_{\mu\nu} = g^{\mu\nu} h_{\mu\nu} = g^{\mu\nu} (g_{\mu\nu} + X_\mu X_\nu) = 4 - 1 = 3,$$

and so the shear is traceless:

$$h^{\mu\nu} \sigma_{\mu\nu} = g^{\mu\nu} \sigma_{\mu\nu} = 0.$$

Fix a geodesic c , and let Y be a geodesic deviation vector along c . If Y is initially orthogonal to c then it will remain orthogonal:

$$X \cdot (X_\mu Y^\mu) = (\nabla_X X_\mu) Y^\mu + X_\mu \nabla_X Y^\mu = X_\mu B^{\mu\nu} Y_\nu = 0.$$

In an orthonormal frame $\{X, E_1, E_2, E_3\}$ parallel along c we then have

$$\dot{Y}^i = B_{ij} Y^j = \left(\frac{1}{3}\theta \delta_{ij} + \sigma_{ij} + \omega_{ij} \right) Y^j = \frac{1}{3}\theta Y^i + \sigma_{ij} Y^j + \omega_{ij} Y^j$$

($i = 1, 2, 3$). If we consider a small spacelike sphere in the hypersurface orthogonal to c and let it be carried by the geodesics of the congruence, we see that θ measures the rate at which the sphere's volume grows, σ describes the sphere's volume-preserving shape deformations, and ω gives the sphere's angular velocity.

¹Curved brackets indicate symmetrization: $B_{(\mu\nu)} = \frac{1}{2}(B_{\mu\nu} + B_{\nu\mu})$.

PROPOSITION 1.5. *The expansion of the congruence satisfies the **Raychaudhuri equation***

$$X \cdot \theta = -\frac{1}{3}\theta^2 - \sigma_{\mu\nu}\sigma^{\mu\nu} + \omega_{\mu\nu}\omega^{\mu\nu} - R_{\mu\nu}X^\mu X^\nu.$$

PROOF. Taking the trace of the equation for $\nabla_X B_{\mu\nu}$ (that is, contracting with $g^{\mu\nu}$) yields

$$\begin{aligned} X \cdot \theta &= -B_{\mu\nu}B^{\nu\mu} - R_{\alpha\beta}X^\alpha X^\beta \\ &= -\left(\frac{1}{3}\theta h_{\mu\nu} + \sigma_{\mu\nu} + \omega_{\mu\nu}\right) \left(\frac{1}{3}\theta h^{\mu\nu} + \sigma^{\mu\nu} - \omega^{\mu\nu}\right) - R_{\mu\nu}X^\mu X^\nu \\ &= -\frac{1}{3}\theta^2 - \sigma_{\mu\nu}\sigma^{\mu\nu} + \omega_{\mu\nu}\omega^{\mu\nu} - R_{\mu\nu}X^\mu X^\nu. \end{aligned}$$

□

2. Energy conditions

DEFINITION 2.1. *A given energy-momentum tensor $T_{\mu\nu}$, with trace $T = g^{\mu\nu}T_{\mu\nu}$, is said to satisfy:*

- (1) *the **strong energy condition (SEC)** if $T_{\mu\nu}X^\mu X^\nu + \frac{1}{2}T \geq 0$ for all unit timelike vectors X ;*
- (2) *the **weak energy condition (WEC)** if $T_{\mu\nu}X^\mu X^\nu \geq 0$ for all timelike vectors X ;*
- (3) *the **null energy condition (NEC)** if $T_{\mu\nu}X^\mu X^\nu \geq 0$ for all null vectors X ;*
- (4) *the **dominant energy condition (DEC)** if $-T^{\mu\nu}X_\nu$ is causal and future-pointing for all causal future-pointing vectors X .*

The weak energy condition is the reasonable requirement that any observer should measure a non-negative energy density, and the null energy condition can be thought of as the same requirement for observers moving at the speed of light. The dominant energy condition, on the other hand, demands that any observer should measure the flow of energy and momentum to be causal. To understand the strong energy condition, we write the Einstein equations as

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu}$$

(possibly including the cosmological constant in the energy-momentum tensor). Note that the trace of this equation yields

$$-R = 8\pi T,$$

and so the Einstein equations can also be written as

$$R_{\mu\nu} = 8\pi \left(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu} \right).$$

Therefore the strong energy condition simply requires that the Ricci tensor satisfies $R_{\mu\nu}X^\mu X^\nu \geq 0$ for all timelike vectors X (given that the Einstein equations are written as above).

Generically, the energy-momentum tensor is diagonalizable, that is, there exists an orthonormal frame $\{E_0, E_1, E_2, E_3\}$ in which the energy-momentum tensor is diagonal,

$$(T_{\mu\nu}) = \text{diag}(\rho, p_1, p_2, p_3).$$

The timelike eigenvalue ρ and is called the **rest energy density**, and the spacelike eigenvalues p_1, p_2, p_3 are known as the **principal pressures**. In terms of these eigenvalues, we have:

- (1) SEC $\Leftrightarrow \rho + \sum_{i=1}^3 p_i \geq 0$ and $\rho + p_i \geq 0$ ($i = 1, 2, 3$).
- (2) WEC $\Leftrightarrow \rho \geq 0$ and $\rho + p_i \geq 0$ ($i = 1, 2, 3$).
- (3) NEC $\Leftrightarrow \rho + p_i \geq 0$ ($i = 1, 2, 3$).
- (4) DEC $\Leftrightarrow \rho \geq |p_i|$ ($i = 1, 2, 3$).

Using this characterization, it is easy to see that the NEC is the weakest energy condition, that is, it is implied by any of the other conditions. The remaining three energy conditions are largely independent, except that the DEC implies the WEC. Notice that in particular the SEC does not imply the WEC.

3. Conjugate points

DEFINITION 3.1. *Let (M, g) be a Lorentzian manifold. A point $q \in M$ is said to be **conjugate** to $p \in M$ along a timelike geodesic c if there exists a nonvanishing solution Y of the Jacobi equation $\nabla_X \nabla_X Y = R(X, Y)X$ such that $Y_p = Y_q = 0$.*

Informally, two points p and q are conjugate along c if there exists a nearby timelike geodesic intersecting c at both p and q (Figure 2).

Chose an orthonormal frame $\{X, E_1, E_2, E_3\}$ parallel along c , where X is the unit tangent vector. Let Y be a geodesic deviation vector Y which vanishes at $p = c(0)$. If we write

$$Y = Y^0 X + Y^i E_i$$

then the Jacobi equation becomes

$$\begin{cases} \ddot{Y}^0 = 0 \\ \ddot{Y}^i = R^i{}_{00j} Y^j \end{cases} .$$

Since Y^0 is an affine function of the proper time τ , it must vanish identically if Y vanishes again along c . We will therefore assume that $Y^0 = 0$, that is, that Y is orthogonal to c . Since the remaining components of Y satisfy a linear ODE, we know that

$$Y^i(\tau) = A_{ij}(\tau) \dot{Y}^j(0),$$

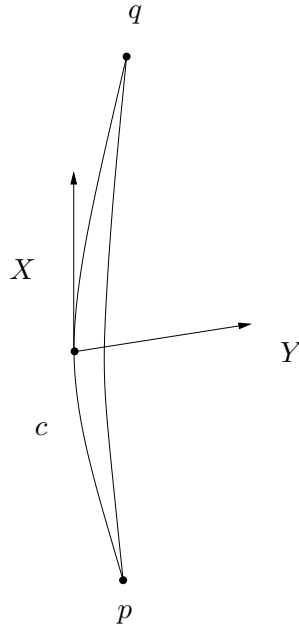


FIGURE 2. Geodesic deviation.

where $A(\tau)$ is the fundamental matrix solution vanishing at $\tau = 0$:

$$\begin{cases} A(0) = 0 \\ \dot{A}(0) = I \\ \ddot{A}_{ij} = R^i{}_{00k} A_{kj} \end{cases} .$$

Although $A(0) = 0$ is singular, $A(\tau)$ is not singular for $\tau > 0$ sufficiently small, because $\dot{A}(0) = I$. If $A(\tau)$ becomes singular for some $\tau_* > 0$ then $q = c(\tau_*)$ is conjugate to p (we just have to choose $\dot{Y}(0)$ to be a nonvanishing column vector in the kernel of $A(\tau_*)$).

Consider the congruence of timelike geodesics through p . Since on the one hand

$$\dot{Y}^i = B_{ij} Y^j$$

and on the other

$$\dot{Y} = \dot{A} \dot{Y}(0) = \dot{A} A^{-1} A \dot{Y}(0) = \dot{A} A^{-1} Y,$$

we conclude that

$$B = \dot{A} A^{-1},$$

and so

$$\theta = \text{tr } B = \text{tr}(\dot{A} A^{-1}) = \frac{d}{d\tau} \log(\det A).$$

Therefore the expansion of the congruence blows up if and only if the geodesic approaches the first conjugate point q .

THEOREM 3.2. *Let (M, g) be a 4-dimensional Lorentzian manifold satisfying the SEC, c a timelike geodesic and $p = c(0)$. Suppose that the expansion θ of the congruence of timelike geodesics through p takes a negative value $\theta_0 < 0$ at some point $r = c(\tau_0)$, with $\tau_0 > 0$. Then there exists a point q conjugate to p along c at a distance at most $\frac{3}{|\theta_0|}$ from r .*

PROOF. By the Gauss Lemma, we can find local coordinates (t, x^1, x^2, x^3) such that $X^\sharp = dt$, and so

$$dX^\sharp = 0 \Leftrightarrow \nabla_{[\mu} X_{\nu]} = 0.$$

In other words, the congruence has no vorticity, and the Raychaudhuri equation becomes

$$\frac{d\theta}{d\tau} = -\frac{1}{3}\theta^2 - \sigma_{\mu\nu}\sigma^{\mu\nu} - R_{\mu\nu}X^\mu X^\nu.$$

Because σ is purely spatial and (M, g) satisfies the SEC, we have

$$\frac{d\theta}{d\tau} \leq -\frac{1}{3}\theta^2 \Leftrightarrow -\frac{1}{\theta^2} \frac{d\theta}{d\tau} \geq \frac{1}{3} \Rightarrow \frac{1}{\theta} \geq \frac{1}{\theta_0} + \frac{1}{3}(\tau - \tau_0).$$

We conclude that $\frac{1}{\theta}$ vanishes, and so θ blows up, at proper time at most $\tau_0 + \frac{3}{|\theta_0|}$. \square

Let $c(t, s)$ be a one-parameter family of timelike curves connecting two points p and q :

$$c(t_0, s) = p \quad \text{and} \quad c(t_1, s) = q$$

for all s . Then the connecting vector

$$Y = \frac{\partial c}{\partial s},$$

which in general is not a Jacobi field, satisfies

$$Y_p = Y_q = 0.$$

We assume that $c(t, s)$ has been parameterized in such a way that Y does not vanish identically.

The tangent vector

$$X = \frac{\partial c}{\partial t}$$

is timelike, and if we define

$$f(t, s) = (-\langle X, X \rangle)^{\frac{1}{2}}$$

then the length of each curve is

$$\tau(s) = \int_{t_0}^{t_1} f(t, s) dt.$$

We have

$$\begin{aligned} \frac{d\tau}{ds} &= \int_{t_0}^{t_1} \frac{\partial f}{\partial s} dt = - \int_{t_0}^{t_1} \frac{1}{f} \langle X, \nabla_Y X \rangle dt = - \int_{t_0}^{t_1} \frac{1}{f} \langle X, \nabla_X Y \rangle dt \\ &= - \int_{t_0}^{t_1} X \cdot \left(\frac{1}{f} \langle X, Y \rangle \right) dt + \int_{t_0}^{t_1} \left\langle \nabla_X \left(\frac{X}{f} \right), Y \right\rangle dt \\ &= \int_{t_0}^{t_1} \left\langle \nabla_X \left(\frac{X}{f} \right), Y \right\rangle dt, \end{aligned}$$

where we used the Fundamental Theorem of Calculus and the fact that $Y_p = Y_q = 0$. This shows that the timelike curve c defined as $c(t) = c(t, 0)$ has extremal length among all timelike curves in such one-parameter families if and only if

$$\nabla_X \left(\frac{X}{f} \right) = 0,$$

that is, if and only if it is a timelike geodesic. Assume this to be the case. Then

$$\frac{d^2\tau}{ds^2}(0) = \int_{t_0}^{t_1} Y \cdot \left\langle \nabla_X \left(\frac{X}{f} \right), Y \right\rangle dt = \int_{t_0}^{t_1} \left\langle \nabla_Y \nabla_X \left(\frac{X}{f} \right), Y \right\rangle dt.$$

Assuming that $f(t, 0) = 1$ (that is, c is parameterized by its proper time) and $\langle X, Y \rangle = 0$ for $s = 0$ (which is always possible by reparameterizing $c(t, s)$) leads to

$$\frac{d^2\tau}{ds^2}(0) = \int_{t_0}^{t_1} \langle \nabla_Y \nabla_X X, Y \rangle dt.$$

Finally, using

$$R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X, Y]} Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z$$

we obtain

$$\begin{aligned} \frac{d^2\tau}{ds^2}(0) &= \int_{t_0}^{t_1} \langle \nabla_X \nabla_Y X - R(X, Y)X, Y \rangle dt \\ &= \int_{t_0}^{t_1} \langle \nabla_X \nabla_X Y - R(X, Y)X, Y \rangle dt. \end{aligned}$$

THEOREM 3.3. *A timelike curve c connecting the points $p, q \in M$ locally maximizes the proper time (along any one-parameter family of timelike curves connecting the same points) if and only if it is a timelike geodesic without conjugate points to p between p and q .*

PROOF. It is clear from what was done above that c being a geodesic is a necessary condition.

Let us assume that c has no conjugate points (to p , say) between p and q . In an orthonormal frame parallel along c we have

$$\frac{d^2\tau}{ds^2}(0) = \int_{t_0}^{t_1} Y^i \left(\ddot{Y}^i - R^i{}_{00j} Y^j \right) dt.$$

Because there are no conjugate points, the fundamental matrix solution $A(t)$ is nondegenerate for $t \in (t_0, t_1)$, and we can set

$$Y^i = A_{ij}Z^j.$$

We have

$$\ddot{Y}^i = A_{ij}\ddot{Z}^j + 2\dot{A}_{ij}\dot{Z}^j + \ddot{A}_{ij}Z^j = A_{ij}\ddot{Z}^j + 2\dot{A}_{ij}\dot{Z}^j + R^i{}_{00k}A_{kj}Z^j,$$

and so

$$\begin{aligned} \frac{d^2\tau}{ds^2}(0) &= \int_{t_0}^{t_1} A_{ij}Z^j \left(A_{ik}\ddot{Z}^k + 2\dot{A}_{ik}\dot{Z}^k \right) dt \\ &= \int_{t_0}^{t_1} Z^t A^t \left(A\ddot{Z} + 2\dot{A}\dot{Z} \right) dt \\ &= \int_{t_0}^{t_1} \left[\frac{d}{dt} \left(Z^t A^t A\dot{Z} \right) - \dot{Z}^t A^t A\dot{Z} - Z^t \dot{A}^t A\dot{Z} + Z^t A^t \dot{A}\dot{Z} \right] dt \\ &= - \int_{t_0}^{t_1} (A\dot{Z})^t A\dot{Z} dt + \int_{t_0}^{t_1} Z^t \left(A^t \dot{A} - \dot{A}^t A \right) \dot{Z} dt. \end{aligned}$$

Above we used the Fundamental Theorem of Calculus and the fact that $(AZ)^t A\dot{Z} = Y^t(\dot{Y} - \dot{A}Z) = Y^t(\dot{Y} - BY)$ vanishes at t_0 and t_1 (although B blows up as $(t - t_0)^{-1}$, Y vanishes as $t - t_0$ or faster by Taylor's formula). From $\dot{A} = BA$ we have

$$A^t \dot{A} - \dot{A}^t A = A^t BA - A^t B^t A = A^t (B - B^t) A = 2A^t \omega A = 0,$$

because the vorticity matrix ω vanishes for the congruence of timelike geodesics through p . Therefore

$$\frac{d^2\tau}{ds^2}(0) = - \int_{t_0}^{t_1} (A\dot{Z})^t A\dot{Z} dt \leq 0,$$

with equality if and only if

$$\dot{Z} \equiv 0 \Rightarrow Z \equiv 0 \Rightarrow Y \equiv 0$$

(note that if Z is constant then it must be zero because $0 = Y_q = A(t_1)Z$ and $A(t_1)$ is nonsingular). We conclude that c is indeed a maximum of the proper time along any one-parameter family of timelike curves connecting p and q .

On the other hand, if there exists a conjugate point along c between p and q , say $r = c(t^*)$, then let \hat{Y} be a nonvanishing Jacobi field such that $\hat{Y}(t_0) = \hat{Y}(t^*) = 0$ (in particular \hat{Y} is orthogonal to c), and let Y be the vector field along c that coincides with \hat{Y} between p and r and is zero between r and q . Similarly, let \hat{Z} be the (necessarily spacelike) vector field parallel along c such that $\hat{Z}(t^*) = -\nabla_X \hat{Y}(t^*)$, and let $Z(t) = \theta(t)\hat{Z}(t)$, where θ is a smooth function satisfying $\theta(t_0) = \theta(t_1) = 0$ and $\theta(t^*) = 1$. Finally, let Y_ε be the vector field along c defined by $Y_\varepsilon = Y + \varepsilon Z$, and consider a one-parameter family of curves $c_\varepsilon(t, s)$ such that $c_\varepsilon(t, 0) = c(t)$ and $Y_\varepsilon = \frac{\partial c_\varepsilon}{\partial s}$.

Since Y_ε is not C^1 , we must write the formula for the second derivative of the length as

$$\frac{d^2\tau}{ds^2}(0) = - \int_{t_0}^{t_1} \left(\langle \nabla_X Y_\varepsilon, \nabla_X Y_\varepsilon \rangle + \langle R(X, Y_\varepsilon)X, Y_\varepsilon \rangle \right) dt = I(Y_\varepsilon, Y_\varepsilon),$$

where the bilinear form I is clearly symmetric. Therefore

$$\frac{d^2\tau}{ds^2}(0) = I(Y, Y) + 2\varepsilon I(Y, Z) + \varepsilon^2 I(Z, Z).$$

Since Y is a Jacobi field between p and r , and zero between r and q , we have $I(Y, Y) = 0$. On the other hand,

$$\begin{aligned} I(Y, Z) &= - \int_{t_0}^{t^*} \left(\langle \nabla_X Y, \nabla_X Z \rangle + \langle R(X, Y)X, Z \rangle \right) dt \\ &= - \left[\langle \nabla_X Y, Z \rangle \right]_{t_0}^{t^*} + \int_{t_0}^{t^*} \left(\langle \nabla_X \nabla_X Y, Z \rangle - \langle R(X, Y)X, Z \rangle \right) dt \\ &= \langle \nabla_X \hat{Y}(t^*), \nabla_X \hat{Y}(t^*) \rangle > 0. \end{aligned}$$

Therefore for $\varepsilon > 0$ sufficiently small the one-parameter family $c_\varepsilon(t, s)$ contains curves whose length is greater than the length of c .

Figure 3 illustrates the geometric idea behind the proof above: c_s represents a generic curve of a one-parameter family corresponding to Y , and has the same length as c ; adding εZ changes c_s between points u and v , say, making it longer by the twin paradox. □

The results above can be generalized for timelike geodesics orthogonal to a spacelike hypersurface S . If one considers the congruence of such geodesics then at S

$$\begin{aligned} \nabla_X Y_\mu &= \nabla_Y X_\mu = Y^\nu (\nabla_\nu X_\mu) = Y^\nu (\nabla_{(\nu} X_{\mu)}) \\ &= \nabla_{(\mu} X_{\nu)} Y^\nu = \frac{1}{2} \mathcal{L}_X g_{\mu\nu} Y^\nu = K_{\mu\nu} Y^\nu. \end{aligned}$$

DEFINITION 3.4. *Let (M, g) be a Lorentzian manifold and let $S \subset M$ be a spacelike hypersurface with second fundamental form K . A point $q \in M$ is said to be **conjugate** to S along a timelike geodesic c orthogonal to S at some point $p \in S$ if there exists a nonvanishing solution Y of the Jacobi equation $\nabla_X \nabla_X Y = R(X, Y)X$ such that $Y_p \in T_p S$, $(\nabla_X Y_\mu)_p = (K_{\mu\nu} Y^\nu)_p$ and $Y_q = 0$.*

In an orthonormal frame $\{X, E_1, E_2, E_3\}$ parallel along c , again we can assume that $Y^0 = 0$, and have for the remaining components

$$Y^i(t) = A_{ij}(t)Y^j(0),$$

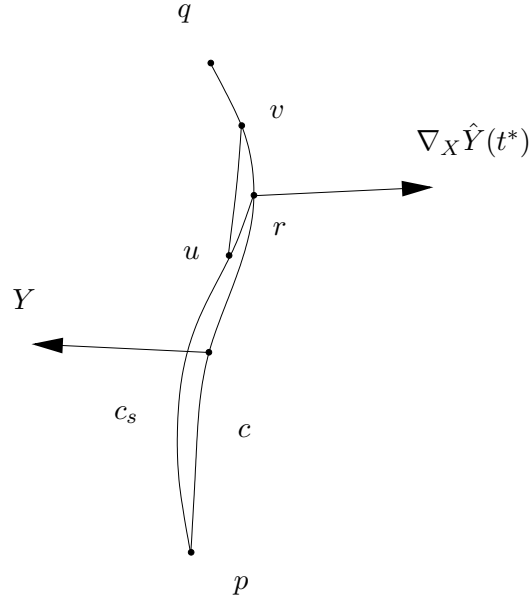


FIGURE 3. Proof of Theorem 3.3.

where $A(t)$ is the fundamental matrix solution:

$$\begin{cases} A(0) = I \\ \dot{A}(0) = K \\ \ddot{A}_{ij} = R^i{}_{00k} A_{kj} \end{cases} .$$

Arguing as above, we have the following result.

THEOREM 3.5. *Let (M, g) be a 4-dimensional Lorentzian manifold satisfying the SEC, $S \subset M$ a spacelike hypersurface and c a timelike geodesic orthogonal to S at some point $p \in S$. Suppose that the expansion θ of the congruence of timelike geodesics orthogonal to S takes a negative value $\theta_0 < 0$ at p . Then there exists a point q conjugate to S along c at a distance at most $\frac{3}{|\theta_0|}$ from S .*

THEOREM 3.6. *A timelike curve c connecting the spacelike hypersurface $S \subset M$ to the point $q \in M$ locally maximizes the proper time (along any one-parameter family of timelike curves connecting S to q) if and only if it is a timelike geodesic orthogonal to S without conjugate points to S between S and q .*

PROOF. The proof is basically the same as for curves connecting two points. The main differences are the in formula

$$\begin{aligned} \frac{d\tau}{ds} &= - \int_{t_0}^{t_1} X \cdot \left(\frac{1}{f} \langle X, Y \rangle \right) dt + \int_{t_0}^{t_1} \left\langle \nabla_X \left(\frac{X}{f} \right), Y \right\rangle dt \\ &= \frac{1}{f(t_0, s)} \langle X_p, Y_p \rangle + \int_{t_0}^{t_1} \left\langle \nabla_X \left(\frac{X}{f} \right), Y \right\rangle dt, \end{aligned}$$

which requires c to be orthogonal to S at $p = c(t_0)$; the fact that $A(t_0)$ does not vanish, but instead

$$\begin{aligned} (AZ)^t A\dot{Z} &= Y^t (\dot{Y} - \dot{A}Z) = Y^t (\dot{Y} - BAZ) \\ &= Y^t (\dot{Y} - BY) = Y^t (\dot{Y} - KY) = 0 \end{aligned}$$

at t_0 ; and the integrated formula

$$\begin{aligned} \frac{d^2\tau}{ds^2}(0) &= (K_{\mu\nu} Y^\mu Y^\nu)(p) - \int_{t_0}^{t_1} \left(\langle \nabla_X Y, \nabla_X Y \rangle + \langle R(X, Y)X, Y \rangle \right) dt \\ &= (K_{\mu\nu} Y^\mu Y^\nu)(p) + I(Y, Y), \end{aligned}$$

which vanishes when Y is a Jacobi field. \square

4. Existence of maximizing geodesics

PROPOSITION 4.1. *Let (M, g) be a globally hyperbolic spacetime, S a Cauchy hypersurface and $p \in D^+(S)$. Then $D^+(S) \cap J^-(p)$ is compact.*

PROOF. Let us define a **simple neighborhood** $U \subset M$ to be a geodesically convex open set diffeomorphic to an open ball whose boundary is a compact submanifold of a larger geodesically convex open set (therefore ∂U is diffeomorphic to S^3 and \bar{U} is compact). It is clear that simple neighborhoods form a basis for the topology of M . Also, it is easy to show that any open cover $\{V_\alpha\}_{\alpha \in A}$ has a countable, locally finite refinement $\{U_n\}_{n \in \mathbb{N}}$ by simple neighborhoods.

If $A = D^+(S) \cap J^-(p)$ were not compact, there would exist a countable, locally finite open cover $\{U_n\}_{n \in \mathbb{N}}$ of A by simple neighborhoods not admitting any finite subcover. Take $q_n \in A \cap U_n$ such that $q_m \neq q_n$ for $m \neq n$. The sequence $\{q_n\}_{n \in \mathbb{N}}$ cannot have accumulation points, since any point in M has a neighborhood intersecting only finite simple neighborhoods U_n . In particular, each simple neighborhood U_n contains only a finite number of points in the sequence (as \bar{U}_n is compact).

Set $p_1 = p$. Since $p_1 \in A$, we have $p_1 \in U_{n_1}$ for some $n_1 \in \mathbb{N}$. Let $q_n \notin U_{n_1}$. Since $q_n \in J^-(p_1)$, there exists a future-directed causal curve c_n connecting q_n to p_1 . This curve will necessarily intersect ∂U_{n_1} . Let $r_{1,n}$ be an intersection point. Since U_{n_1} contains only a finite number of points in the sequence $\{q_n\}_{n \in \mathbb{N}}$, there will exist infinite intersection points $r_{1,n}$. As ∂U_{n_1} is compact, these will accumulate to some point $p_2 \in \partial U_{n_1}$ (cf. Figure 4).

Because \bar{U}_{n_1} is contained in a geodesically convex open set V , which can be chosen so that $v \mapsto (\pi(v), \exp(v))$ is a diffeomorphism onto $V \times V$, we

have $p_2 \in J^-(p_1)$: if $\gamma_{1,n}$ is the unique causal geodesic connecting p_1 to $r_{1,n}$, parameterized by the global time function $t : M \rightarrow \mathbb{R}$, then the subsequence of $\{\gamma_{1,n}\}$ corresponding to a convergent subsequence of $\{r_{1,n}\}$ will converge to a causal geodesic γ_1 connecting p_1 to p_2 . If $S = t^{-1}(0)$ then we have $t(r_{1,n}) \geq 0$, implying that $t(p_2) \geq 0$ and hence $p_2 \in A$. Since $p_2 \notin U_{n_1}$, there must exist $n_2 \in \mathbb{N}$ such that $p_2 \in U_{n_2}$.

Since U_{n_2} contains only a finite number of points in the sequence $\{q_n\}_{n \in \mathbb{N}}$, an infinite number of curves c_n must intersect ∂U_{n_2} to the past of $r_{1,n}$. Let $r_{2,n}$ be the intersection points. As ∂U_{n_2} is compact, $\{r_{2,n}\}$ must accumulate to some point $p_3 \in \partial U_{n_2}$. Because \bar{U}_{n_2} is contained in a geodesically convex open set, $p_3 \in J^-(p_2)$: if $\gamma_{2,n}$ is the unique causal geodesic connecting $r_{1,n}$ to $r_{2,n}$, parameterized by the global time function, then the subsequence of $\{\gamma_{2,n}\}$ corresponding to convergent subsequences of both $\{r_{1,n}\}$ and $\{r_{2,n}\}$ will converge to a causal geodesic connecting p_2 to p_3 . Since $J^-(p_2) \subset J^-(p_1)$ and $t(r_{2,n}) \geq 0 \Rightarrow t(p_3) \geq 0$, we have $p_3 \in A$.

Iterating the procedure above, we can construct a sequence $\{p_i\}_{i \in \mathbb{N}}$ of points in A satisfying $p_i \in U_{n_i}$ with $n_i \neq n_j$ if $i \neq j$, such that p_i is connected to p_{i+1} by a causal geodesic γ_i . It is clear that γ_i cannot intersect S , for $t(p_{i+1}) > t(p_{i+2}) \geq 0$. On the other hand, the piecewise smooth causal curve obtained by joining the curves γ_i can easily be smoothed into a past-directed causal curve starting at p_1 which does not intersect S . Finally, such curve is inextendible: it cannot converge to any point, as $\{p_i\}_{i \in \mathbb{N}}$ cannot accumulate. But since $p_1 \in D^+(S)$, this curve would have to intersect S . Therefore A must be compact. \square

COROLLARY 4.2. *Let (M, g) be a globally hyperbolic spacetime and $p, q \in M$. Then*

- (i) $J^+(p)$ is closed;
- (ii) $J^+(p) \cap J^-(q)$ is compact.

PROOF. Exercise. \square

Proposition 4.1 is a key ingredient in establishing the following fundamental result.

THEOREM 4.3. *Let (M, g) be a globally hyperbolic spacetime with Cauchy hypersurface S , and $p \in D^+(S)$. Then, among all timelike curves connecting p to S , there exists a timelike curve with maximal length. This curve is a timelike geodesic, orthogonal to S .*

PROOF. Consider the set $T(S, p)$ of all timelike curves connecting S to p . Since we can always use the global time function $t : M \rightarrow \mathbb{R}$ as a parameter, these curves are determined by their images, which are compact subsets of the compact set $A = D^+(S) \cap J^-(p)$. As it is well known (see for instance [Mun00]), the set $C(A)$ of all compact subsets of A is a compact metric space for the **Hausdorff metric** d_H , defined as follows: if $d : M \times M \rightarrow \mathbb{R}$

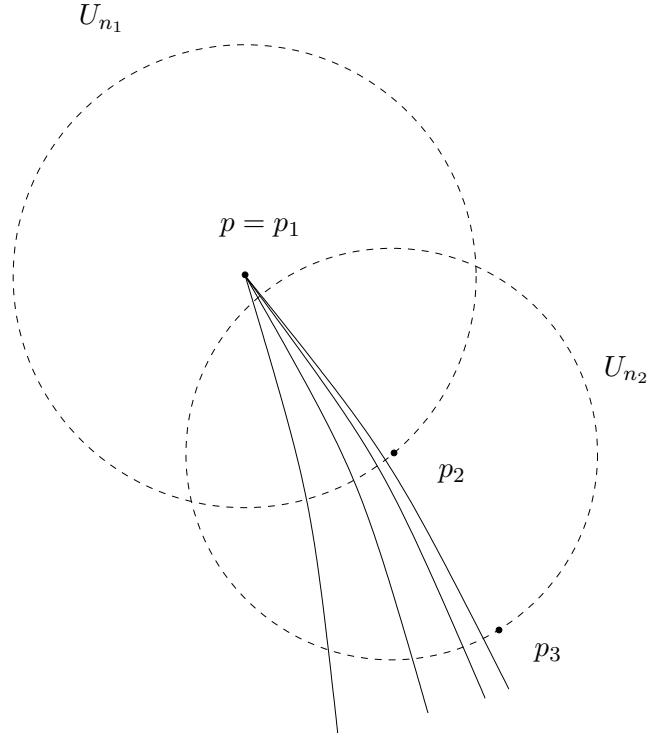


FIGURE 4. Proof of Proposition 4.1.

is a metric yielding the topology of M ,

$$d_H(K, L) = \inf\{\varepsilon > 0 \mid K \subset U_\varepsilon(L) \text{ and } L \subset U_\varepsilon(K)\},$$

where $U_\varepsilon(K)$ is a ε -neighborhood of K for the metric d . Therefore, the closure $C(S, p) = \overline{T(S, p)}$ is a compact subset of $C(A)$. It is not difficult to show that $C(S, p)$ can be identified with the set of **continuous causal curves** connecting S to p (a continuous curve $c : [0, t(p)] \rightarrow M$ is said to be **causal** if $c(t_2) \in J^+(c(t_1))$ whenever $t_2 > t_1$).

The length function $\tau : T(S, p) \rightarrow \mathbb{R}$ is defined by

$$\tau(c) = \int_0^{t(p)} |\dot{c}(t)| dt.$$

This function is **upper semicontinuous**, i.e. continuous for the topology

$$\mathcal{O} = \{(-\infty, a) \mid -\infty \leq a \leq +\infty\}$$

in \mathbb{R} . Indeed, let $c \in T(S, p)$ be parameterized by its arclength u . For a sufficiently small $\varepsilon > 0$, the function u can be extended to the ε -neighborhood $U_\varepsilon(c)$ in such a way that its level hypersurfaces are spacelike and orthogonal to c , that is, $-\text{grad } u$ is timelike and coincides with \dot{c} on c (cf. Figure 5). If $\gamma \in T(S, p)$ is in the open ball $B_\varepsilon(c) \subset C(A)$ for the Hausdorff metric d_H

then we can use u as a parameter, thus obtaining

$$du(\dot{\gamma}) = 1 \Leftrightarrow \langle \dot{\gamma}, \text{grad } u \rangle = 1.$$

Therefore $\dot{\gamma}$ can be decomposed as

$$\dot{\gamma} = \frac{1}{\langle \text{grad } u, \text{grad } u \rangle} \text{grad } u + X,$$

where X is spacelike and orthogonal to $\text{grad } u$, and so

$$|\dot{\gamma}| = \left| \frac{1}{\langle \text{grad } u, \text{grad } u \rangle} + \langle X, X \rangle \right|^{\frac{1}{2}}.$$

Given $\delta > 0$, we can choose $\varepsilon > 0$ sufficiently small so that

$$-\frac{1}{\langle \text{grad } u, \text{grad } u \rangle} < \left(1 + \frac{\delta}{2\tau(c)}\right)^2$$

on the ε -neighborhood $U_\varepsilon(c)$ (as $\langle \text{grad } u, \text{grad } u \rangle = -1$ along c). We have

$$\tau(\gamma) = \int_0^{t(p)} \left| \frac{d\gamma}{dt} \right| dt = \int_0^{t(p)} |\dot{\gamma}| \frac{du}{dt} dt = \int_{u(\gamma \cap S)}^{\tau(c)} |\dot{\gamma}| du,$$

where we have to allow for the fact that c is not necessarily orthogonal to S , and so the initial point of γ is not necessarily at $u = 0$ (cf. Figure 5). Consequently,

$$\begin{aligned} \tau(\gamma) &= \int_{u(\gamma \cap S)}^{\tau(c)} \left| -\frac{1}{\langle \text{grad } u, \text{grad } u \rangle} - \langle X, X \rangle \right|^{\frac{1}{2}} du \\ &< \int_{u(\gamma \cap S)}^{\tau(c)} \left(1 + \frac{\delta}{2\tau(c)}\right) du = \left(1 + \frac{\delta}{2\tau(c)}\right) (\tau(c) - u(\gamma \cap S)). \end{aligned}$$

Choosing ε sufficiently small so that

$$|u| < \left(\frac{1}{\tau(c)} + \frac{2}{\delta}\right)^{-1}$$

on $S \cap U_\varepsilon(c)$, we obtain $\tau(\gamma) < \tau(c) + \delta$, proving upper semicontinuity in $T(S, p)$. As a consequence, the length function can be extended to $C(S, p)$ through

$$\tau(c) = \lim_{\varepsilon \rightarrow 0} \sup \{ \tau(\gamma) \mid \gamma \in B_\varepsilon(c) \cap T(S, p) \}$$

(as for $\varepsilon > 0$ sufficiently small the supremum will be finite). Also, it is clear that if $c \in T(S, p)$ then the upper semicontinuity of the length forces the two definitions of $\tau(c)$ to coincide. The extension of the length function to $C(S, p)$ is trivially upper semicontinuous: given $c \in C(S, p)$ and $\delta > 0$, let $\varepsilon > 0$ be such that $\tau(\gamma) < \tau(c) + \frac{\delta}{2}$ for any $\gamma \in B_{2\varepsilon}(c) \cap T(S, p)$. Then it is clear that $\tau(c') \leq \tau(c) + \frac{\delta}{2} < \tau(c) + \delta$ for any $c' \in B_\varepsilon(c) \cap C(S, p)$.

Finally, we notice that the compact sets of \mathbb{R} for the topology \mathcal{O} are the sets with a maximum. Therefore, the length function attains a maximum at some point $c \in C(S, p)$. All that remains to be seen is that the maximum is also attained at a smooth timelike curve γ . To do so, cover c with finitely

many geodesically convex neighborhoods and choose points p_1, \dots, p_k in c such that $p_1 \in S$, $p_k = p$ and the portion of c between p_{i-1} and p_i is contained in a geodesically convex neighborhood for all $i = 2, \dots, k$. It is clear that there exists a sequence $c_n \in T(S, p)$ such that $c_n \rightarrow c$ and $\tau(c_n) \rightarrow \tau(c)$. Let $t_i = t(p_i)$ and $p_{i,n}$ be the intersection of c_n with $t^{-1}(t_i)$. Replace c_n by the sectionally geodesic curve γ_n obtained by joining $p_{i-1,n}$ to $p_{i,n}$ in the corresponding geodesically convex neighborhood. Then $\tau(\gamma_n) \geq \tau(c_n)$, and therefore $\tau(\gamma_n) \rightarrow \tau(c)$. Since each sequence $p_{i,n}$ converges to p_i , γ_n converges to the sectionally geodesic curve γ obtained by joining p_{i-1} to p_i ($i = 2, \dots, k$), and it is clear that $\tau(\gamma_n) \rightarrow \tau(\gamma) = \tau(c)$. Therefore γ is a point of maximum for the length. Finally, we notice that γ must be smooth at the points p_i , for otherwise we could increase its length by using the twin paradox. Therefore γ must be a timelike geodesic. It is also clear that γ must be orthogonal to S , for otherwise it would be possible to increase its length by small deformations. \square

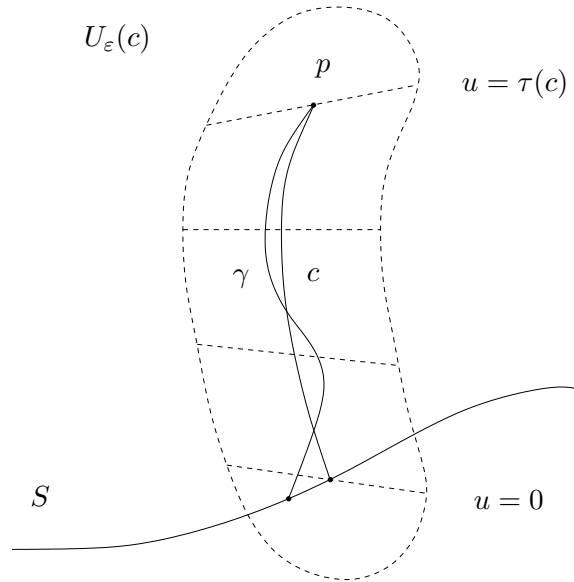


FIGURE 5. Proof of Theorem 4.3.

5. Hawking's singularity theorem

We have now all the necessary ingredients to prove the Hawking singularity theorem.

DEFINITION 5.1. *A spacetime (M, g) is said to be **singular** if it is not geodesically complete.*

THEOREM 5.2. (Hawking [Haw66]) *Let (M, g) be a globally hyperbolic spacetime satisfying the strong energy condition, and suppose that the expansion of the congruence of future-pointing timelike geodesics orthogonal to S satisfies $\theta \leq \theta_0 < 0$ on a Cauchy hypersurface S . Then (M, g) is singular.*

PROOF. We will show that no future-directed timelike geodesic orthogonal to S can be extended to proper time greater than $\tau_0 = -\frac{3}{\theta_0}$ to the future of S . Suppose that this was not so. Then there would exist a future-directed timelike geodesic c orthogonal to S , parameterized by proper time, defined in an interval $[0, \tau_0 + \varepsilon]$ for some $\varepsilon > 0$. Let $p = c(\tau_0 + \varepsilon)$. According to Theorem 4.3, there would exist a timelike geodesic γ with maximal length connecting S to p , orthogonal to S . Because $\tau(c) = \tau_0 + \varepsilon$, we would necessarily have $\tau(\gamma) \geq \tau_0 + \varepsilon$. Theorem 3.5 guarantees that γ would develop a conjugate point at a distance of at most τ_0 to the future of S , and Theorem 3.6 states that γ would cease to be maximizing beyond this point. Therefore we arrive at a contradiction. \square

REMARK 5.3. It should be clear that (M, g) is singular if the condition $\theta \leq \theta_0 < 0$ on a Cauchy hypersurface S is replaced by the condition $\theta \geq \theta_0 > 0$ on S . In this case, no **past-directed** timelike geodesic orthogonal to S can be extended to proper time greater than $\tau_0 = \frac{3}{\theta_0}$ to the **past** of S .

EXAMPLE 5.4.

- (1) The FLRW models with $\alpha > 0$ and $\Lambda = 0$ are globally hyperbolic, and satisfy the strong energy condition (as $\rho > 0$). Moreover, the expansion of the congruence tangent to $\frac{\partial}{\partial t}$ is $\theta = \frac{3\dot{a}}{a}$. Assume that the model is expanding at time t_0 . Then $\theta = \theta_0 = \frac{3\dot{a}(t_0)}{a(t_0)} > 0$ on the Cauchy hypersurface $S = \{t = t_0\}$, and hence Theorem 5.2 guarantees that this model is singular to the past of S (i.e. there exists a big bang). Moreover, Theorem 5.2 implies that this singularity is generic: any sufficiently small perturbation of an expanding FLRW model satisfying the strong energy condition will also be singular. Loosely speaking, any expanding universe must have begun at a big bang.
- (2) The region $\{r < 2m\}$ of the Schwarzschild solution is globally hyperbolic, and satisfies the strong energy condition (as $R_{\mu\nu} = 0$). The metric can be written in this region as

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

where

$$\tau = \int_r^{2m} \left(\frac{2m}{u} - 1\right)^{-\frac{1}{2}} du.$$

Therefore the inside of the black hole can be pictured as a cylinder $\mathbb{R} \times S^2$ whose shape is evolving in time. As $r \rightarrow 0$, the S^2 contracts

to a singularity, with the t -direction expanding. Since

$$K = \frac{dr}{d\tau} \left(-\frac{m}{r^2} dt^2 + r d\theta^2 + r \sin^2 \theta d\varphi^2 \right),$$

we have

$$\theta = \left(\frac{2m}{r} - 1 \right)^{-\frac{1}{2}} \left(\frac{2}{r} - \frac{3m}{r^2} \right).$$

Therefore we have $\theta = \theta_0 < 0$ on any Cauchy hypersurface $S = \{r = r_0\}$ with $r_0 < \frac{3m}{2}$, and hence Theorem 5.2 guarantees that the Schwarzschild solution is singular to the future of S . Moreover, Theorem 5.2 implies that this singularity is generic: any sufficiently small perturbation of the Schwarzschild solution satisfying the strong energy condition will also be singular. Loosely speaking, once the collapse has advanced long enough, nothing can prevent the formation of a singularity.

- (3) It should be noted that Theorem 5.2 proves geodesic incompleteness, not the existence of curvature singularities. For instance, it applies to the Milne universe, or a globally hyperbolic region of the anti-de Sitter universe, whose curvature is bounded (they are simply globally hyperbolic regions in larger, inextendible Lorentzian manifolds).

6. Penrose's singularity theorem

Let (M, g) be a globally hyperbolic spacetime, S a Cauchy hypersurface with future-pointing unit normal vector field N , and $\Sigma \subset S$ a compact 2-dimensional submanifold with unit normal vector field n in S . Let c_p be the null geodesic with initial condition $N_p + n_p$ for each point $p \in \Sigma$. We define a smooth map $\exp : (-\varepsilon, \varepsilon) \times \Sigma \rightarrow M$ for some $\varepsilon > 0$ as $\exp(r, p) = c_p(r)$.

DEFINITION 6.1. *The critical values of \exp are said to be **conjugate points** to Σ .*

Loosely speaking, conjugate points are points where geodesics starting orthogonally at nearby points of Σ intersect.

Let $q = \exp(r_0, p)$ be a point not conjugate to Σ . If φ is a local parameterization of Σ around p , then we can construct a system of local coordinates (u, r, x^2, x^3) on some open set $V \ni q$ by using the map

$$(u, r, x^2, x^3) \mapsto \exp(r, \psi_u(\varphi(x^2, x^3))),$$

where ψ_u is the flow along the timelike geodesics orthogonal to S and the map $\exp : (-\varepsilon, \varepsilon) \times \psi_u(\Sigma) \rightarrow M$ is defined as above.

Since $\frac{\partial}{\partial r}$ is tangent to null geodesics, we have $g_{rr} = \langle \frac{\partial}{\partial r}, \frac{\partial}{\partial r} \rangle = 0$. On the other hand, we have

$$\begin{aligned} \frac{\partial g_{r\mu}}{\partial r} &= \frac{\partial}{\partial r} \left\langle \frac{\partial}{\partial r}, \frac{\partial}{\partial x^\mu} \right\rangle = \left\langle \frac{\partial}{\partial r}, \nabla_{\frac{\partial}{\partial r}} \frac{\partial}{\partial x^\mu} \right\rangle \\ &= \left\langle \frac{\partial}{\partial r}, \nabla_{\frac{\partial}{\partial x^\mu}} \frac{\partial}{\partial r} \right\rangle = \frac{1}{2} \frac{\partial}{\partial x^\mu} \left\langle \frac{\partial}{\partial r}, \frac{\partial}{\partial r} \right\rangle = 0, \end{aligned}$$

for $\mu = 0, 1, 2, 3$. Since $g_{ru} = -1$ and $g_{r2} = g_{r3} = 0$ on $\psi_u(\Sigma)$, we have $g_{ru} = -1$ and $g_{r2} = g_{r3} = 0$ on V . Therefore the metric is written in this coordinate system as

$$g = \alpha du^2 - 2dudr + 2\beta_A dudx^A + \gamma_{AB} dx^A dx^B.$$

Since

$$\det \begin{pmatrix} \alpha & -1 & \beta_2 & \beta_3 \\ -1 & 0 & 0 & 0 \\ \beta_2 & 0 & \gamma_{22} & \gamma_{23} \\ \beta_3 & 0 & \gamma_{32} & \gamma_{33} \end{pmatrix} = -\det \begin{pmatrix} \gamma_{22} & \gamma_{23} \\ \gamma_{32} & \gamma_{33} \end{pmatrix},$$

we see that the functions

$$\gamma_{AB} = \left\langle \frac{\partial}{\partial x^A}, \frac{\partial}{\partial x^B} \right\rangle$$

form a positive definite matrix, and so g induces a Riemannian metric on the 2-dimensional surfaces $\exp(r, \psi_u(\Sigma))$, which are then spacelike. Since the vector fields $\frac{\partial}{\partial x^A}$ can always be defined along c_p , the matrix (γ_{AB}) is also well defined along c_p , even at points where the coordinate system breaks down, i.e. at points which are conjugate to Σ . These are the points for which $\gamma = \det(\gamma_{AB})$ vanishes, since only then will $\{\frac{\partial}{\partial u}, \frac{\partial}{\partial r}, \frac{\partial}{\partial x^2}, \frac{\partial}{\partial x^3}\}$ fail to be linearly independent. In fact the vector fields $\frac{\partial}{\partial x^A}$ are Jacobi fields along c_p .

It is easy to see that

$$\Gamma_{ur}^u = \Gamma_{rr}^u = \Gamma_{rA}^u = \Gamma_{rr}^r = \Gamma_{rr}^A = 0 \quad \text{and} \quad \Gamma_{rB}^A = \gamma^{AC} \beta_{CB},$$

where $(\gamma^{AB}) = (\gamma_{AB})^{-1}$ and $\beta_{AB} = \frac{1}{2} \frac{\partial \gamma_{AB}}{\partial r}$. Consequently,

$$\begin{aligned} R_{rr} &= R_{urr}{}^u + R_{Arr}{}^A = \left(-\frac{\partial \Gamma_{Ar}^A}{\partial r} - \Gamma_{Ar}^B \Gamma_{rB}^A \right) \\ &= -\frac{\partial}{\partial r} (\gamma^{AB} \beta_{AB}) - \gamma^{BC} \gamma^{AD} \beta_{CA} \beta_{DB}. \end{aligned}$$

The quantity

$$\theta = \gamma^{AB} \beta_{AB}$$

appearing in this expression is called the **expansion** of the null geodesics, and has an important geometric meaning:

$$\theta = \frac{1}{2} \operatorname{tr} \left((\gamma_{AB})^{-1} \frac{\partial}{\partial r} (\gamma_{AB}) \right) = \frac{1}{2} \frac{\partial}{\partial r} \log (\det (\gamma_{AB})) = \frac{\partial}{\partial r} \log (\det (\gamma_{AB}))^{\frac{1}{2}}.$$

Therefore the expansion yields the variation of the area element of the spacelike 2-dimensional surfaces $\exp(r, \psi_u(\Sigma))$. More importantly for our purposes, we see that a singularity of the expansion indicates a zero of $\det(\gamma_{AB})$, i.e. a conjugate point to $\psi_u(\Sigma)$.

PROPOSITION 6.2. *Let (M, g) be a globally hyperbolic spacetime satisfying the null energy condition, $S \subset M$ a Cauchy hypersurface with future-pointing unit normal vector field N , $\Sigma \subset S$ a compact 2-dimensional submanifold with unit normal vector field n in S and $p \in \Sigma$ a point where $\theta = \theta_0 < 0$. Then the null geodesic c_p with initial condition $N_p + n_p$ contains at least a point conjugate to Σ , at an affine parameter distance of at most $-\frac{2}{\theta_0}$ to the future of Σ (assuming that it can be extended that far).*

PROOF. Since (M, g) satisfies the null energy condition, we have $R_{rr} = R_{\mu\nu} \left(\frac{\partial}{\partial r}\right)^\mu \left(\frac{\partial}{\partial r}\right)^\nu \geq 0$. Consequently,

$$\frac{\partial \theta}{\partial r} + \gamma^{BC} \gamma^{AD} \beta_{CA} \beta_{DB} \leq 0.$$

Choosing an orthonormal basis (where $\gamma^{AB} = \delta_{AB}$), and using the inequality

$$(\text{tr } A)^2 \leq n \text{tr}(A^t A)$$

for square $n \times n$ matrices, it is easy to show that

$$\gamma^{BC} \gamma^{AD} \beta_{CA} \beta_{DB} = \beta_{BA} \beta_{AB} = \text{tr}((\beta_{AB})(\beta_{AB})^t) \geq \frac{1}{2} \theta^2.$$

Consequently θ must satisfy

$$\frac{\partial \theta}{\partial r} + \frac{1}{2} \theta^2 \leq 0.$$

Integrating this inequality yields

$$\frac{1}{\theta} \geq \frac{1}{\theta_0} + \frac{r}{2},$$

and hence θ must blow up at a value of r no greater than $-\frac{2}{\theta_0}$. \square

We define the **chronological future** and the **causal future** of the compact surface Σ as

$$I^+(\Sigma) = \bigcup_{p \in \Sigma} I^+(p) \quad \text{and} \quad J^+(\Sigma) = \bigcup_{p \in \Sigma} J^+(p)$$

(with similar definitions for the **chronological past** and the **causal past** of Σ). It is clear that $I^+(\Sigma)$, being the union of open sets, is itself open, and also that $J^+(\Sigma) \subset \overline{I^+(\Sigma)}$ and $I^+(\Sigma) = \text{int } J^+(\Sigma)$. On the other hand, it is easy to generalize Proposition 4.1 (and consequently Corollary 4.2) to the corresponding statements with compact surfaces replacing points. In particular, $J^+(\Sigma)$ is closed. Therefore

$$\partial J^+(\Sigma) = \partial I^+(\Sigma) = J^+(\Sigma) \setminus I^+(\Sigma),$$

and so, by a straightforward generalization of Corollary 1.5 in Chapter 3, every point in this boundary can be reached from a point in Σ by a future-directed null geodesic. Moreover, this geodesic must be orthogonal to Σ . Indeed, at Σ we have

$$\frac{\partial}{\partial u} = N \quad \text{and} \quad \frac{\partial}{\partial r} = N + n,$$

and so the metric takes the form

$$g = -du^2 - 2dudr + \gamma_{AB}dx^A dx^B.$$

If $c : I \subset \mathbb{R} \rightarrow M$ is a future-directed null geodesic with $c(0) \in \Sigma$, its initial tangent vector

$$\dot{c}(0) = \dot{u} \frac{\partial}{\partial u} + \dot{r} \frac{\partial}{\partial r} + \dot{x}^A \frac{\partial}{\partial x^A} = (\dot{u} + \dot{r})N + \dot{r}n + \dot{x}^A \frac{\partial}{\partial x^A}$$

must satisfy

$$\dot{u}(\dot{u} + 2\dot{r}) = \gamma_{AB}\dot{x}^A \dot{x}^B.$$

Since c is future-directed we must have $\dot{u} + \dot{r} > 0$. On the other hand, by choosing the unit normal to Σ on S to be either n or $-n$, we can assume $\dot{r} \geq 0$. If c is not orthogonal to Σ we then have

$$\gamma_{AB}\dot{x}^A \dot{x}^B > 0 \Rightarrow \dot{u}(\dot{u} + 2\dot{r}) > 0 \Rightarrow \dot{u} > 0.$$

Now the region where $u > 0$ and $r \geq 0$ is clearly a subset of $I^+(\Sigma)$, since its points can be reached from Σ by a sectionally smooth curve composed of an arc of timelike geodesic and an arc of null geodesic. Therefore, we see that if c is not orthogonal to Σ then $c(t) \in I^+(\Sigma)$ for all $t > 0$.

Even future-directed null geodesics orthogonal to Σ may eventually enter $I^+(\Sigma)$. A sufficient condition for this to happen is given in the following result.

PROPOSITION 6.3. *Let (M, g) be a globally hyperbolic spacetime, S a Cauchy hypersurface with future-pointing unit normal vector field N , $\Sigma \subset S$ a compact 2-dimensional submanifold with unit normal vector field n in S , $p \in \Sigma$, c_p the null geodesic through p with initial condition $N_p + n_p$ and $q = c_p(r)$ for some $r > 0$. If c_p has a conjugate point between p and q then $q \in I^+(\Sigma)$.*

PROOF. We will offer only a sketch of the proof. Let s be the first conjugate point along c_p between p and q . Since q is conjugate to p , there exists another null geodesic γ starting at Σ which (approximately) intersects c_p at s . The piecewise smooth null curve obtained by following γ between Σ and s , and c_p between s and q is a causal curve but not a null geodesic. This curve can be easily smoothed while remaining causal and nongeodesic, and so by the generalization of Corollary 1.5 in Chapter 3 we have $q \in I^+(\Sigma)$. \square

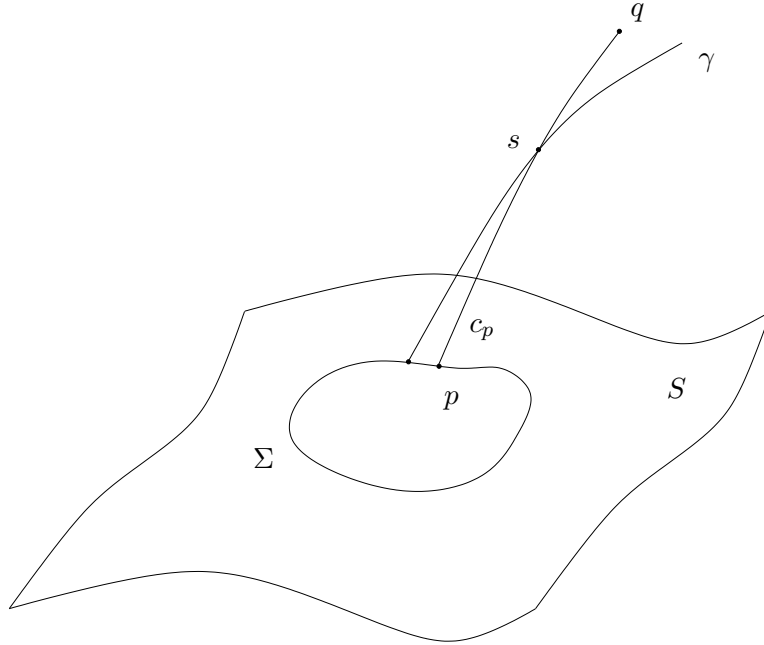


FIGURE 6. Proof of Proposition 6.3.

DEFINITION 6.4. Let (M, g) be a globally hyperbolic spacetime and S a Cauchy hypersurface with future-pointing unit normal vector field N . A compact 2-dimensional submanifold $\Sigma \subset S$ with unit normal vector field n in S is said to be **trapped** if the expansions θ^+ and θ^- of the null geodesics with initial conditions $N + n$ and $N - n$ are both negative everywhere on Σ .

We have now all the necessary ingredients to prove the Penrose singularity theorem.

THEOREM 6.5. (**Penrose [Pen65]**) Let (M, g) be a connected globally hyperbolic spacetime with a noncompact Cauchy hypersurface S , satisfying the null energy condition. If S contains a trapped surface Σ then (M, g) is singular.

PROOF. Let $t : M \rightarrow \mathbb{R}$ be a global time function such that $S = t^{-1}(0)$. The integral curves of $\text{grad } t$, being timelike, intersect S exactly once, and $\partial I^+(\Sigma)$ at most once. This defines a continuous injective map $\pi : \partial I^+(\Sigma) \rightarrow S$, whose image is open. Indeed, if $q = \pi(p)$, then all points in some neighborhood of q are images of points in $\partial I^+(\Sigma)$, as otherwise there would be a sequence $q_k \in S$ with $q_k \rightarrow q$ such that the integral curves of $\text{grad } t$ through q_k would not intersect $\partial I^+(\Sigma)$. Letting r_k be the intersections of these curves with the Cauchy hypersurface $t^{-1}(t(r))$, for some point r to the future of p along the integral line of $\text{grad } t$, we would have $r_k \rightarrow r$, and so $r_k \in I^+(\Sigma)$ for sufficiently large k (as $I^+(\Sigma)$ is open), leading to a contradiction.

Since Σ is trapped (and compact), there exists $\theta_0 < 0$ such that the expansions θ^+ and θ^- of the null geodesics orthogonal to Σ both satisfy $\theta^+, \theta^- \leq \theta_0$. We will show that there exists a future-directed null geodesic orthogonal to Σ which cannot be extended to an affine parameter greater than $r_0 = -\frac{2}{\theta_0}$ to the future of Σ . Suppose that this was not so. Then, according to Proposition 6.2, any null geodesic orthogonal to Σ would have a conjugate point at an affine parameter distance of at most r_0 to the future of Σ , after which it would be in $I^+(\Sigma)$, by Proposition 6.3. Consequently, $\partial I^+(\Sigma)$ would be a (closed) subset of the compact set

$$\exp^+([0, r_0] \times \Sigma) \cup \exp^-([0, r_0] \times \Sigma)$$

(where \exp^+ and \exp^- refer to the exponential map constructed using the unit normals n and $-n$), hence compact. Therefore the image of π would also be compact, hence closed as well as open. Since M , and therefore S , are connected, the image of π would be S , which would then be homeomorphic to $\partial I^+(\Sigma)$. But S is noncompact by hypothesis, and we reach a contradiction. \square

REMARK 6.6. It should be clear that (M, g) is singular if the condition of existence of a trapped surface is replaced by the condition of existence of an **anti-trapped surface**, that is, a compact surface $\Sigma \subset S$ such that the expansions of null geodesics orthogonal to Σ are both positive. In this case, there exists a **past-directed** null geodesic orthogonal to Σ which cannot be extended to an affine parameter time greater than $r_0 = \frac{2}{\theta_0}$ to the **past** of Σ .

EXAMPLE 6.7.

- (1) The region $\{r < 2m\}$ of the Schwarzschild solution is globally hyperbolic, and satisfies the null energy condition (as $R_{\mu\nu} = 0$). Since r (or $-r$) is clearly a time function (depending on the choice of time orientation), it must increase (or decrease) along any future-pointing null geodesic, and therefore any sphere Σ of constant (t, r) is anti-trapped (or trapped). Since any Cauchy hypersurface is diffeomorphic to $\mathbb{R} \times S^2$, hence noncompact, we conclude from Theorem 6.5 that the Schwarzschild solution is singular to past (or future) of Σ . Moreover, Theorem 6.5 implies that this singularity is generic: any sufficiently small perturbation of the Schwarzschild solution satisfying the null energy condition will also be singular. Loosely speaking, once the collapse has advanced long enough, nothing can prevent the formation of a singularity.
- (2) The FLRW models with $\alpha > 0$ and $\Lambda = 0$ are globally hyperbolic, and satisfy the null energy condition. Moreover, radial null geodesics satisfy

$$\frac{dr}{dt} = \pm \frac{1}{a} \sqrt{1 - kr^2}.$$

Therefore, if we start with a sphere Σ of constant (t, r) and follow the orthogonal null geodesics along the direction of increasing or

decreasing r , we obtain spheres whose radii ar satisfy

$$\frac{d}{dt}(ar) = \dot{a}r + a\dot{r} = \dot{a}r \pm \sqrt{1 - kr^2}.$$

Assume that the model is expanding, with the big bang at $t = 0$, and spatially noncompact (in particular $k \neq 1$). Then, for sufficiently small $t > 0$, the sphere Σ is anti-trapped, and hence Theorem 6.5 guarantees that this model is singular to the past of Σ (i.e. there exists a big bang). Moreover, Theorem 6.5 implies that this singularity is generic: any sufficiently small perturbation of an expanding, spatially noncompact FLRW model satisfying the null energy condition will also be singular. Loosely speaking, any expanding universe must have begun at a big bang.

7. Exercises

- (1) Let g be a Lorentzian metric given in the Gauss Lemma form,

$$g = -dt^2 + h_{ij}dx^i dx^j,$$

and consider the geodesic congruence tangent to $\frac{\partial}{\partial t}$.

- (a) Show that

$$B_{ij} = \Gamma_{ij}^0 = \frac{1}{2} \frac{\partial h_{ij}}{\partial t},$$

that is,

$$B = \frac{1}{2} \mathcal{L}_{\frac{\partial}{\partial t}} g = K,$$

where K is the second fundamental form of the hypersurfaces of constant t .

- (b) Conclude that the expansion of the congruence of the galaxies in a FLRW model is

$$\theta = \frac{3\dot{a}}{a} = 3H.$$

- (2) Let (M, g) be a Lorentzian manifold.

- (a) Use the formula for the Lie derivative of a tensor,

$$(\mathcal{L}_X g)(Y, Z) = X \cdot g(Y, Z) - g(\mathcal{L}_X Y, Z) - g(Y, \mathcal{L}_X Z)$$

to show that

$$(\mathcal{L}_X g)(Y, Z) = g(\nabla_Y X, Z) + g(Y, \nabla_Z X).$$

- (b) Show that this formula can be written as

$$(\mathcal{L}_X g)_{\mu\nu} = \nabla_\mu X_\nu + \nabla_\nu X_\mu.$$

- (c) Suppose that X is a Killing vector field, i.e. $\mathcal{L}_X g = 0$. Use the Killing equation

$$\nabla_\mu X_\nu + \nabla_\nu X_\mu = 0$$

to show that X is a solution of the Jacobi equation. Give a geometric interpretation of this fact.

- (3) Let T be a diagonalizable energy-momentum tensor, that is, $(T_{\mu\nu}) = \text{diag}(\rho, p_1, p_2, p_3)$ on some orthonormal frame $\{E_0, E_1, E_2, E_3\}$. Show that:
- (a) T satisfies the SEC if and only if $\rho + \sum_{i=1}^3 p_i \geq 0$ and $\rho + p_i \geq 0$ ($i = 1, 2, 3$).
 - (b) T satisfies the WEC if and only if $\rho \geq 0$ and $\rho + p_i \geq 0$ ($i = 1, 2, 3$).
 - (c) T satisfies the DEC if and only if $\rho \geq |p_i|$ ($i = 1, 2, 3$).
 - (d) T satisfies the NEC if and only if $\rho + p_i \geq 0$ ($i = 1, 2, 3$).
 - (e) The first three conditions are independent except that the DEC implies the WEC.
 - (f) The first three conditions imply the NEC.
- (4) Let (M, g) be the globally hyperbolic Lorentzian manifold corresponding to the exterior region of the Schwarzschild solution, that is, $M = \mathbb{R} \times (\mathbb{R}^3 \setminus \overline{B_{2m}(0)})$ and

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

(with $m > 0$).

- (a) Show that for any $r_0 > 2m$ the curve

$$c(t) = \left(t, r_0, \frac{\pi}{2}, \sqrt{\frac{m}{r_0^3}} t\right)$$

is a timelike, null or spacelike geodesic, according to whether $r_0 > 3m$, $r_0 = 3m$ or $r_0 < 3m$.

- (b) Argue that the point $q = \left(\pi\sqrt{\frac{r_0^3}{m}}, r_0, \frac{\pi}{2}, \pi\right)$ is conjugate to the point $p = \left(0, r_0, \frac{\pi}{2}, 0\right)$ along c (note that this can be done without solving the Jacobi equation).
- (c) Show explicitly that if $r_0 > 3m$ then c stops being maximizing for $t > \pi\sqrt{\frac{r_0^3}{m}}$.

- (5) Let (M, g) be a globally hyperbolic spacetime and $p, q \in M$ with $q \in I^+(p)$. Show that among all timelike curves connecting p to q there exists a timelike curve with maximal length, which is a timelike geodesic.
- (6) Use ideas similar to those leading to the proof of Hawking's singularity theorem to prove **Myers's theorem**: if $(M, \langle \cdot, \cdot \rangle)$ is a complete Riemannian manifold whose Ricci curvature satisfies $R_{\mu\nu} X^\mu X^\nu \geq \varepsilon g_{\mu\nu} X^\mu X^\nu$ for some $\varepsilon > 0$ then M is compact. Can these ideas be used to prove a singularity theorem in Riemannian geometry?
- (7) Explain why Hawking's singularity theorem does not apply to each of the following geodesically complete Lorentzian manifolds:

- (a) Minkowski's spacetime;
 - (b) Einstein's universe;
 - (c) de Sitter's universe;
 - (d) Anti-de Sitter spacetime.
- (8) Consider the metric

$$ds^2 = \alpha du^2 - 2 du dr + 2\beta_A du dx^A + \gamma_{AB} dx^A dx^B.$$

- (a) Show that the Christoffel symbols satisfy

$$\Gamma_{ur}^u = \Gamma_{rr}^u = \Gamma_{rA}^u = \Gamma_{rr}^r = \Gamma_{rr}^A = 0 \quad \text{and} \quad \Gamma_{rB}^A = \gamma^{AC} \beta_{CB},$$

where $(\gamma^{AB}) = (\gamma_{AB})^{-1}$ and $\beta_{AB} = \frac{1}{2} \frac{\partial \gamma_{AB}}{\partial r}$.

- (b) Conclude that

$$R_{rr} = -\frac{\partial}{\partial r} (\gamma^{AB} \beta_{AB}) - \gamma^{AB} \gamma^{CD} \beta_{AC} \beta_{BD}.$$

- (9) Let (M, g) be a globally hyperbolic spacetime with Cauchy hypersurfaces S_0 and S_1 satisfying $S_1 \subset D^+(S_0)$, and $\Sigma \subset S_1$ a compact surface. Show that:
- (a) $D^+(S_0) \cap J^-(\Sigma)$ is compact;
 - (b) $J^-(\Sigma)$ is closed.
- (10) Explain why Penrose's singularity theorem does not apply to each of the following geodesically complete Lorentzian manifolds:
- (a) Minkowski's spacetime;
 - (b) Einstein's universe;
 - (c) de Sitter's universe;
 - (d) Anti-de Sitter spacetime.

CHAPTER 5

Cauchy problem

In this chapter we discuss the Cauchy problem for the Einstein field equations, following [Wal84]. We start by studying the Klein-Gordon equation, as a prototypical wave equation, and the the Maxwell equations, where the issues of constraints on the initial data and gauge freedom also arise. We then sketch the proof of the Choquet-Bruhat theorem and discuss the Lichnerowicz method for solving the constraint equations. See [Rin09] for a more complete discussion.

1. Divergence theorem

It is possible to define the divergence of a vector field on any orientable manifold where a volume form has been chosen.

DEFINITION 1.1. *Let M be an orientable n -dimensional manifold with volume form ϵ , and let X be a vector field on M . The **divergence** of X is the function $\operatorname{div} X$ such that*

$$d(X \lrcorner \epsilon) = (\operatorname{div} X)\epsilon.$$

The following result is a straightforward application of the Stokes theorem.

THEOREM 1.2. (Divergence theorem) *If M is a compact orientable n -dimensional manifold with boundary ∂M then*

$$\int_M (\operatorname{div} X)\epsilon = \int_{\partial M} X \lrcorner \epsilon.$$

PROPOSITION 1.3. *If (M, g) is a pseudo-Riemannian manifold with Levi-Civita connection ∇ then*

$$\operatorname{div} X = \nabla_\mu X^\mu.$$

PROOF. Take local coordinates such that $X = \partial_1$ around each point where X does not vanish. Since

$$\epsilon = \sqrt{|\det(g_{\mu\nu})|} dx^1 \wedge \dots \wedge dx^n,$$

we have

$$d(X \lrcorner \epsilon) = \mathcal{L}_X \epsilon - X \lrcorner d\epsilon = \partial_1 \log |\det(g_{\mu\nu})|^{\frac{1}{2}} \epsilon,$$

where we used $d\epsilon = 0$. Since

$$\partial_1 \log |\det A| = \operatorname{tr}(A^{-1} \partial_1 A)$$

for any matrix-valued function A in M , we have

$$\begin{aligned} d(X \lrcorner \epsilon) &= \frac{1}{2} (g^{\mu\nu} \partial_1 g_{\mu\nu}) \epsilon = \frac{1}{2} (g^{\mu\nu} (\mathcal{L}_X g)_{\mu\nu}) \epsilon \\ &= \frac{1}{2} g^{\mu\nu} (\nabla_\mu X_\nu + \nabla_\nu X_\mu) \epsilon = (\nabla_\mu X^\mu) \epsilon. \end{aligned}$$

This formula can be easily extended to the set of zeros of X : by continuity on the boundary, and trivially in the interior. \square

Assume now that (M, g) is a compact orientable n -dimensional Lorentzian manifold with boundary ∂M , and let $\{E_1, \dots, E_n\}$ be a positive orthonormal frame with $E_1 = N$ the outward unit normal vector on ∂M . The volume element is

$$\epsilon = -E_1^\sharp \wedge \dots \wedge E_n^\sharp,$$

and the volume element of ∂M with the induced orientation is

$$\sigma = \pm E_2^\sharp \wedge \dots \wedge E_n^\sharp$$

(according to whether N is timelike or spacelike). Therefore we have on ∂M

$$X \lrcorner \epsilon = \pm \langle X, N \rangle \sigma,$$

implying that the divergence theorem can be written as

$$\int_M (\operatorname{div} X) \epsilon = \int_{\partial M} \langle X, n \rangle \sigma,$$

where n is the outward unit normal vector in the points where it is spacelike, and is the inward unit normal vector in the points where it is timelike (Figure 1).

It may happen that ∂M has points where the normal is null, hence tangent to ∂M . In that case we choose the orthonormal frame such that

$$N = E_1 + E_2$$

is a null normal, with E_1 timelike and pointing outwards, and E_2 spacelike and (necessarily) pointing inwards. Then

$$\epsilon = -N^\sharp \wedge E_2^\sharp \wedge \dots \wedge E_n^\sharp$$

and so

$$X \lrcorner \epsilon = -\langle X, N \rangle \sigma,$$

where

$$\sigma = E_2^\sharp \wedge \dots \wedge E_n^\sharp$$

is a volume element on ∂M (compatible with the induced orientation, as E_1 points outwards). So we must choose in this case $n = -N$, that is, we must use the normal whose timelike component points inwards and whose spacelike component points outwards (Figure 1). Note that the magnitude (but not the sign) of the volume element σ will depend on the choice of n .

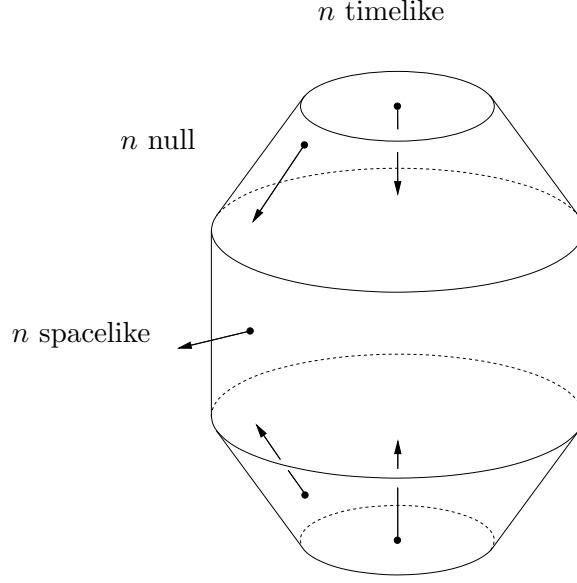


FIGURE 1. Normal vector for the divergence theorem on a Lorentzian manifold.

2. Klein-Gordon equation

Let (M, g) be a Lorentzian manifold. A smooth function $\phi : M \rightarrow \mathbb{R}$ is a solution of the **Klein-Gordon equation** if it satisfies

$$\square\phi - m^2\phi = 0 \Leftrightarrow \nabla^\mu \partial_\mu \phi - m^2\phi = 0.$$

For reasons that will become clear in Chapter 6, we define the **energy-momentum tensor** associated to this equation as

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2} g_{\mu\nu} (\partial_\alpha \phi \partial^\alpha \phi + m^2 \phi^2).$$

If ϕ is a solution of the Klein-Gordon equation then

$$\begin{aligned} \nabla^\mu T_{\mu\nu} &= \square\phi \partial_\nu \phi + \partial_\mu \phi \nabla^\mu \partial_\nu \phi - \partial_\alpha \phi \nabla_\nu \partial^\alpha \phi - m^2 \phi \partial_\nu \phi \\ &= (\square\phi - m^2\phi) \partial_\nu \phi = 0 \end{aligned}$$

Moreover, if (M, g) is time-oriented then it is possible to prove that T satisfies the dominant energy condition, that is, $T_{\mu\nu} X^\nu$ corresponds to a past-pointing causal vector whenever X is a future-pointing causal vector. Assume that X is a future-pointing timelike Killing vector field, and define $Y^\mu = T^{\mu\nu} X_\nu$. Then Y is a past-pointing causal vector field satisfying

$$\nabla_\mu Y^\mu = T^{\mu\nu} \nabla_\mu X_\nu = 0$$

(where we used the Killing equation $\nabla_{(\mu} X_{\nu)} = 0$ and the symmetry of T).

Let us now focus on the case when (M, g) is flat Minkowski spacetime and $X = \frac{\partial}{\partial t}$. Consider the Cauchy hypersurface $S_0 = \{t = t_0\}$, and let B_0

be the ball of radius R in that hypersurface:

$$B_0 = \{t = t_0, x^2 + y^2 + z^2 \leq R^2\}.$$

Let $S_1 = \{t = t_1\}$ be another Cauchy hypersurface, and consider the ball $B_1 = D(B_0) \cap S_1$ in that hypersurface (Figure 2). By the divergence theorem we have

$$\int_{B_0} \langle Y, X \rangle + \int_C \langle Y, n \rangle + \int_{B_1} \langle Y, -X \rangle = 0,$$

where C is the null portion of $\partial D(B_0)$ between S_0 and S_1 and n is a past-pointing normal. Since Y is a past-pointing causal vector we have $\langle Y, n \rangle \leq 0$, and so

$$(13) \quad \int_{B_1} \langle Y, X \rangle \leq \int_{B_0} \langle Y, X \rangle.$$

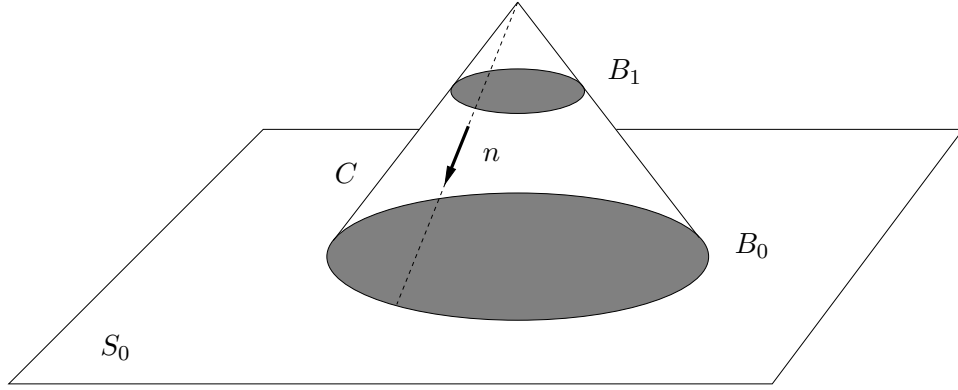


FIGURE 2. Proof of Proposition 2.1.

Note that

$$\begin{aligned} \langle Y, X \rangle &= T_{\mu\nu} X^\mu X^\nu = (X \cdot \phi)^2 + \frac{1}{2} (\partial_\alpha \phi \partial^\alpha \phi + m^2 \phi^2) \\ &= \frac{1}{2} [(\partial_0 \phi)^2 + (\partial_x \phi)^2 + (\partial_y \phi)^2 + (\partial_z \phi)^2 + m^2 \phi^2]. \end{aligned}$$

We conclude immediately that if ϕ is a solution of the Klein-Gordon equation and $\phi = \partial_0 \phi = 0$ in B_0 then $\phi = 0$ in B_1 , and indeed in $D(S_0)$ (since t_1 is arbitrary). Because the Klein-Gordon equation is linear we can then deduce the following result.

PROPOSITION 2.1. *Given two smooth functions $\phi_0, \psi_0 : S_0 \rightarrow \mathbb{R}$, there exists at most a solution ϕ of the Klein-Gordon equation satisfying $\phi = \phi_0$ and $\partial_0 \phi = \psi_0$ on S_0 .*

Given a smooth function $\phi : M \rightarrow \mathbb{R}$ and some set $A \subset \mathbb{R}^4$ we define the **Sobolev norm**

$$\|\phi\|_{H^1(A)}^2 = \int_A [\phi^2 + (\partial_0 \phi)^2 + (\partial_x \phi)^2 + (\partial_y \phi)^2 + (\partial_z \phi)^2].$$

In this definition A can either be an open set or a submanifold (in which case we use the induced volume form in the integral).

More generally, for each $k \in \mathbb{N}_0$ we define the Sobolev norms

$$\|\phi\|_{H^k(A)}^2 = \int_A \sum_{|\alpha| \leq k} (\partial^\alpha \phi)^2,$$

where $\alpha = (\alpha_0, \alpha_1, \alpha_2, \alpha_3) \in \mathbb{N}_0^4$, $|\alpha| = \alpha_0 + \alpha_1 + \alpha_2 + \alpha_3$, $\partial = (\partial_0, \partial_x, \partial_y, \partial_z)$ and $\partial^\alpha = \partial_0^{\alpha_0} \partial_x^{\alpha_1} \partial_y^{\alpha_2} \partial_z^{\alpha_3}$. Similarly, given a smooth function $\phi_0 : S_0 \rightarrow \mathbb{R}$ we define

$$\|\phi_0\|_{H^k(B_0)}^2 = \int_{B_0} \sum_{|\alpha| \leq k} (D^\alpha \phi_0)^2,$$

where now $D = (\partial_x, \partial_y, \partial_z)$. Inequality (13) can then be written as

$$\|\phi\|_{H^1(B_1)}^2 \leq C \|\phi\|_{H^1(B_0)}^2 \leq C \|\phi_0\|_{H^1(B_0)}^2 + C \|\psi_0\|_{H^0(B_0)}^2,$$

where $C > 0$ is a generic positive constant (not always the same). Integrating this inequality in t from $t_0 - R$ to $t_0 + R$ we obtain

$$\|\phi\|_{H^1(D(B_0))}^2 \leq C \|\phi_0\|_{H^1(B_0)}^2 + C \|\psi_0\|_{H^0(B_0)}^2.$$

On S_0 , all spatial partial derivatives of ϕ and $\partial_0 \phi$ are given by partial derivatives of ϕ_0 and ψ_0 . On the other hand, from the Klein-Gordon equation we have

$$\partial_0^2 \phi|_{S_0} = \partial_x^2 \phi_0 + \partial_y^2 \phi_0 + \partial_z^2 \phi_0 - m^2 \phi_0,$$

and so, by partial differentiation, we can obtain all spatial partial derivatives of $\partial_0^2 \phi$ on S_0 from partial derivatives of ϕ_0 . Differentiating the Klein-Gordon equation with respect to t yields

$$\partial_0^3 \phi|_{S_0} = \partial_x^2 \psi_0 + \partial_y^2 \psi_0 + \partial_z^2 \psi_0 - m^2 \psi_0,$$

and it should be clear that all partial derivatives of ϕ on S_0 are given by partial derivatives of ϕ_0 and ψ_0 . Note from the general partial derivative of the Klein-Gordon equation,

$$\partial_0^2 \partial^\alpha \phi = \partial_x^2 \partial^\alpha \phi + \partial_y^2 \partial^\alpha \phi + \partial_z^2 \partial^\alpha \phi - m^2 \partial^\alpha \phi,$$

that $\partial^\alpha \phi$ also satisfies the Klein-Gordon equation, and so

$$\begin{aligned} \|\partial^\alpha \phi\|_{H^1(B_1)}^2 &\leq C \|\partial^\alpha \phi\|_{H^1(B_0)}^2 \Rightarrow \\ \|\phi\|_{H^k(B_1)}^2 &\leq C \|\phi_0\|_{H^k(B_0)}^2 + C \|\psi_0\|_{H^{k-1}(B_0)}^2 \Rightarrow \\ \|\phi\|_{H^k(D(B_0))}^2 &\leq C \|\phi_0\|_{H^k(B_0)}^2 + C \|\psi_0\|_{H^{k-1}(B_0)}^2, \end{aligned}$$

whence

$$\|\phi\|_{H^k(D(B_0))} \leq C \|\phi_0\|_{H^k(B_0)} + C \|\psi_0\|_{H^{k-1}(B_0)}.$$

Another important norm is the **supremum norm**,

$$\|\phi\|_{C^0(A)} = \sup_{p \in A} |\phi(p)|.$$

More generally, we define the norms

$$\|\phi\|_{C^k(A)} = \sum_{|\alpha| \leq k} \sup_{p \in A} |\partial^\alpha \phi(p)|.$$

DEFINITION 2.2. *A set $A \subset \mathbb{R}^n$ is said to satisfy the **interior cone condition** if there exists a (closed) cone of height $h > 0$ and solid angle $\Omega > 0$ at the vertex such that for each point $p \in A$ it is possible to map the cone isometrically into A in such a way that the vertex is mapped to p .*

THEOREM 2.3. (Sobolev inequality) *If $A \subset \mathbb{R}^n$ satisfies the interior cone condition and $k > \frac{n}{2}$ then for any smooth function $f : A \rightarrow \mathbb{R}$*

$$\|f\|_{C^0(A)} \leq C \|f\|_{H^k(A)}.$$

PROOF. Exercise. □

For a solution of the Klein-Gordon equation we then have

$$\|\phi\|_{C^0(D(B_0))} \leq \|\phi\|_{H^3(D(B_0))} \leq C \|\phi_0\|_{H^3(B_0)} + C \|\psi_0\|_{H^2(B_0)}.$$

More generally,

$$\|\partial^\alpha \phi\|_{C^0(D(B_0))} \leq \|\partial^\alpha \phi\|_{H^3(D(B_0))} \leq C \|\phi_0\|_{H^{m+3}(B_0)} + C \|\psi_0\|_{H^{m+2}(B_0)},$$

where $m = |\alpha|$. We conclude that

$$\|\phi\|_{C^m(D(B_0))} \leq C \|\phi_0\|_{H^{m+3}(B_0)} + C \|\psi_0\|_{H^{m+2}(B_0)},$$

whence

$$\|\phi\|_{C^m(D(B_0))} \leq C \|\phi_0\|_{C^{m+3}(B_0)} + C \|\psi_0\|_{C^{m+2}(B_0)}.$$

THEOREM 2.4. *Given initial data $\phi_0, \psi_0 : S_0 \rightarrow \mathbb{R}$ for the Klein-Gordon equation, there exists a unique smooth solution ϕ satisfying $\phi = \phi_0$ and $\partial_0 \phi = \psi_0$ on S_0 . Moreover, if $B_0 \subset S_0$ is a ball then the solution in $D(B_0)$ depends only on the initial data in B_0 , and the map*

$$C^{m+3}(B_0) \times C^{m+2}(B_0) \ni (\phi_0, \psi_0) \mapsto \phi \in C^m(D(B_0))$$

is continuous.

PROOF. We just have to prove existence of solution. To do that, we note that if ϕ is a solution then we know all its partial derivatives on S_0 . Therefore, we can construct a power series for ϕ around each point of S_0 . If ϕ_0 and ψ_0 are analytic then the **Cauchy-Kowalewski theorem** guarantees that these series converge, and so there exists an analytic solution ϕ . If ϕ_0 and ψ_0 are smooth then there exist sequences $\phi_{0,n}$ and $\psi_{0,n}$ of analytic functions which converge to ϕ_0 and ψ_0 in all spaces $C^m(B_0)$. The corresponding analytic solutions ϕ_n of the Klein-Gordon equation thus form a Cauchy sequence in all spaces $C^m(D(B_0))$, and hence must converge in all these spaces to some function ϕ . This function is therefore smooth, and, passing to the limit, must satisfy the Klein-Gordon equation in all sets $(D(B_0))$. □

Using similar techniques, it is possible to prove a much stronger result.

THEOREM 2.5. *Let (M, g) be a globally hyperbolic spacetime with a Cauchy hypersurface S and future unit normal N . Then the **linear, diagonal second order hyperbolic system***

$$g^{\mu\nu}\nabla_\mu\partial_\nu\phi_i + A_{ij}^\mu\partial_\mu\phi_j + B_{ij}\phi_j + C_i = 0 \quad (i = 1, \dots, n),$$

where A_{ij}^μ , B_{ij} and C_i are smooth and ∇ is any connection, yields a well-posed Cauchy problem with initial data in S . More precisely, given smooth initial data $(\phi_1, \dots, \phi_n, N \cdot \phi_1, \dots, N \cdot \phi_n)$ on S there exists a unique smooth solution of the system, defined in M . Moreover, the solutions depend continuously on the initial data, and if two initial data sets coincide on some closed subset $B \subset S$ then the corresponding solutions coincide in $D(B)$.

To solve the Cauchy problem for the Einstein equations, we will need to solve more complicated systems of hyperbolic equations.

THEOREM 2.6. *Consider the **quasi-linear, diagonal second order hyperbolic system***

$$g^{\mu\nu}(x, \phi, \partial\phi)\nabla_\mu\partial_\nu\phi_i = F_i(x, \phi, \partial\phi) \quad (i = 1, \dots, n),$$

where $g^{\mu\nu}$ and F_i are smooth and ∇ is any connection on some manifold M . Let $(\phi_0)_1, \dots, (\phi_0)_n$ be a solution of this system, and define $(g_0)^{\mu\nu} = g^{\mu\nu}(x, \phi_0, \partial\phi_0)$. Assume that (M, g) is globally hyperbolic, and let S be a Cauchy hypersurface. Then the system above yields a well-posed Cauchy problem with initial data in S , in the following sense: given initial data in S sufficiently close to the initial data for $(\phi_0)_1, \dots, (\phi_0)_n$ there exists an open neighborhood V of S such that the system has a unique solution in V , and $(V, g(x, \phi, \partial\phi))$ is globally hyperbolic. Moreover, the solutions depend continuously on the initial data, and if two initial data sets coincide on some closed subset $B \subset S$ then the corresponding solutions coincide in $D(B)$.

PROOF. The idea of the proof is to start with the linear hyperbolic system

$$g^{\mu\nu}(x, \phi_0, \partial\phi_0)\nabla_\mu\partial_\nu\phi_i = F_i(x, \phi_0, \partial\phi_0) \quad (i = 1, \dots, n),$$

which by the previous theorem has a unique solution ϕ_1 close to ϕ_0 . Because of this, there exists a neighborhood V_1 of S such that $(V_1, g(x, \phi_1, \partial\phi_1))$ is globally hyperbolic with Cauchy hypersurface S , and so the system

$$g^{\mu\nu}(x, \phi_1, \partial\phi_1)\nabla_\mu\partial_\nu\phi_i = F_i(x, \phi_1, \partial\phi_1) \quad (i = 1, \dots, n),$$

again has a unique solution ϕ_2 close to ϕ_1 . Iterating this procedure we obtain a sequence ϕ_n which can then be shown to converge to the unique solution of the quasi-linear hyperbolic system. \square

3. Maxwell's equations: constraints and gauge

As a warm-up problem to solving the Einstein field equations we consider the considerably easier problem of solving the Maxwell equations without

sources in flat Minkowski spacetime. These equations can be split into what we shall call **constraint equations**,

$$\begin{cases} \operatorname{div} \mathbf{E} = 0 \\ \operatorname{div} \mathbf{B} = 0 \end{cases},$$

and **evolution equations**:

$$\begin{cases} \frac{\partial \mathbf{E}}{\partial t} = \operatorname{curl} \mathbf{B} \\ \frac{\partial \mathbf{B}}{\partial t} = -\operatorname{curl} \mathbf{E} \end{cases}.$$

One expects that the evolution equations completely determine $\mathbf{E}(t, \mathbf{x})$ and $\mathbf{B}(t, \mathbf{x})$ from initial data

$$\begin{cases} \mathbf{E}(0, \mathbf{x}) = \mathbf{E}_0(\mathbf{x}) \\ \mathbf{B}(0, \mathbf{x}) = \mathbf{B}_0(\mathbf{x}) \end{cases}.$$

This initial data, however, is not completely free: it must satisfy the constraint equations

$$\operatorname{div} \mathbf{E}_0 = \operatorname{div} \mathbf{B}_0 = 0.$$

This suffices to guarantee that the constraint equations are satisfied by $\mathbf{E}(t, \mathbf{x})$ and $\mathbf{B}(t, \mathbf{x})$, since they are preserved by the evolution: for instance,

$$\frac{\partial}{\partial t}(\operatorname{div} \mathbf{E}) = \operatorname{div} \left(\frac{\partial \mathbf{E}}{\partial t} \right) = \operatorname{div}(\operatorname{curl} \mathbf{B}) = 0.$$

As we will see, solving the Einstein field equations will also require splitting them into constraint equations and evolution equations. Another issue that will have to be dealt with, gauge freedom, also occurs when solving the Maxwell equations by using the electromagnetic **gauge potentials**. To do so, we note that two of the Maxwell equations (those which do not admit sources) are equivalent to the existence of a vector potential \mathbf{A} and a scalar potential ϕ in terms of which \mathbf{B} and \mathbf{E} can be written:

$$\begin{cases} \operatorname{div} \mathbf{B} = 0 \\ \operatorname{curl} \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t} \end{cases} \Leftrightarrow \begin{cases} \mathbf{B} = \operatorname{curl} \mathbf{A} \\ \mathbf{E} = -\operatorname{grad} \phi - \frac{\partial \mathbf{A}}{\partial t} \end{cases}.$$

These potentials, however, are nonunique: given any smooth function χ , the potentials

$$\begin{cases} \mathbf{A}' = \mathbf{A} + \operatorname{grad} \chi \\ \phi' = \phi - \frac{\partial \chi}{\partial t} \end{cases}$$

yield the same fields \mathbf{B} and \mathbf{E} (\mathbf{A}' and ϕ' are said to be related to \mathbf{A} and ϕ by a **gauge transformation**). The remaining Maxwell equations can now

be written as

$$\begin{cases} \operatorname{div} \mathbf{E} = 0 \\ \operatorname{curl} \mathbf{B} = \frac{\partial \mathbf{E}}{\partial t} \end{cases} \Leftrightarrow \begin{cases} \Delta \phi + \frac{\partial}{\partial t}(\operatorname{div} \mathbf{A}) = 0 \\ \operatorname{grad}(\operatorname{div} \mathbf{A}) - \Delta \mathbf{A} = -\operatorname{grad} \frac{\partial \phi}{\partial t} - \frac{\partial^2 \mathbf{A}}{\partial t^2} \end{cases} .$$

Therefore, if there exist gauge potentials satisfying

$$\operatorname{div} \mathbf{A} = -\frac{\partial \phi}{\partial t}$$

(the so-called **Lorentz gauge**) then these equations reduce to uncoupled wave equations:

$$\begin{cases} \square \phi = 0 \\ \square \mathbf{A} = 0 \end{cases}$$

To solve the Maxwell equations using the gauge potentials we then solve these wave equations with initial data ϕ_0 , $\left(\frac{\partial \phi}{\partial t}\right)_0$ and \mathbf{A}_0 , $\left(\frac{\partial \mathbf{A}}{\partial t}\right)_0$ satisfying:

- (1) $\operatorname{curl} \mathbf{A}_0 = \mathbf{B}_0$ (possible because $\operatorname{div} \mathbf{B}_0 = 0$);
- (2) $\phi_0 = 0$ (by choice);
- (3) $\left(\frac{\partial \mathbf{A}}{\partial t}\right)_0 = -\mathbf{E}_0$ (giving the correct initial electric field);
- (4) $\left(\frac{\partial \phi}{\partial t}\right)_0 = -\operatorname{div} \mathbf{A}_0$ (so that the Lorentz gauge condition holds).

These potentials will determine a solution of the Maxwell equations with the correct initial data **if the Lorentz gauge condition holds for all time**. Now

$$\square \left(\operatorname{div} \mathbf{A} + \frac{\partial \phi}{\partial t} \right) = \operatorname{div}(\square \mathbf{A}) + \frac{\partial}{\partial t}(\square \phi) = 0$$

and

$$\left(\operatorname{div} \mathbf{A} + \frac{\partial \phi}{\partial t} \right)_0 = 0.$$

Moreover,

$$\frac{\partial}{\partial t} \left(\operatorname{div} \mathbf{A} + \frac{\partial \phi}{\partial t} \right) = \operatorname{div} \left(\frac{\partial \mathbf{A}}{\partial t} \right) + \frac{\partial^2 \phi}{\partial t^2} = \operatorname{div} \left(\frac{\partial \mathbf{A}}{\partial t} \right) + \Delta \phi,$$

and so

$$\left(\frac{\partial}{\partial t} \left(\operatorname{div} \mathbf{A} + \frac{\partial \phi}{\partial t} \right) \right)_0 = -\operatorname{div} \mathbf{E}_0 = 0.$$

By uniqueness of solution of the wave equation, we conclude that the Lorentz gauge condition does hold for all time, and so the potentials obtained by solving the wave equation with the initial conditions above do determine the solution of the Maxwell equations with initial data \mathbf{B}_0 , \mathbf{E}_0 .

REMARK 3.1. The electromagnetic potentials can be seen as the components of the **electromagnetic potential one-form**

$$A = -\phi dt + A^1 dx + A^2 dy + A^3 dz.$$

Note that a gauge transformation can be written as

$$A' = A + d\chi.$$

The electric and magnetic fields can in turn be seen as the components of the **Faraday tensor**

$$\begin{aligned} F = dA &= E^1 dx \wedge dt + E^2 dy \wedge dt + E^3 dz \wedge dt \\ &+ B^1 dy \wedge dz + B^2 dz \wedge dx + B^3 dx \wedge dy. \end{aligned}$$

It should be obvious that F remains invariant under a gauge transformation. The Maxwell equations can be written as

$$dF = 0 \Leftrightarrow F = dA$$

and

$$d \star F = 0,$$

since

$$\begin{aligned} \star F &= E^1 dy \wedge dz + E^2 dz \wedge dx + E^3 dx \wedge dy \\ &- B^1 dx \wedge dt - B^2 dy \wedge dt - B^3 dz \wedge dt \end{aligned}$$

(that is, the Hodge star replaces \mathbf{B} with \mathbf{E} and \mathbf{E} with $-\mathbf{B}$).

4. Einstein's equations

Let (M, g) be a globally hyperbolic spacetime and $S \subset M$ a Cauchy hypersurface. Let us write g in the Gauss Lemma form near S ,

$$g = -dt^2 + h_{ij}(t, x)dx^i dx^j,$$

so that the level sets of t are Riemannian manifolds with induced metric $h(t) = h_{ij}dx^i dx^j$ and second fundamental form

$$K(t) = \frac{1}{2} \frac{\partial h_{ij}}{\partial t} dx^i dx^j.$$

For this choice of coordinates (gauge), finding the metric is equivalent to finding a time-dependent Riemannian metric $h(t)$ on S . The vacuum Einstein field equations $G_{\mu\nu} = 0$ can be split into **constraint equations**

$$\begin{cases} G_{00} = 0 \\ G_{0i} = 0 \end{cases} \Leftrightarrow \begin{cases} \bar{R} + (K^i_i)^2 - K_{ij}K^{ij} = 0 \\ \bar{\nabla}_i K^j_j - \bar{\nabla}_j K^j_i = 0 \end{cases}$$

and **evolution equations**

$$G_{ij} = 0 \Leftrightarrow \frac{\partial}{\partial t} K_{ij} = -\bar{R}_{ij} + 2K_{il}K^l_j - K^l_l K_{ij},$$

where $\bar{\nabla}$, \bar{R} and \bar{R}_{ij} are the Levi-Civita connection, the scalar curvature and the Ricci tensor of h . Note that the evolution equations allow us to evolve $h(t)$ and $K(t)$, whereas the constraint equations restrict their initial values $h(0)$ and $K(0)$. If the initial data satisfy the constraint equations, so

does the solution of the evolution equations. Indeed, the contracted Bianchi identities give us for free the equations

$$\begin{aligned}\nabla^\alpha G_{\alpha\beta} = 0 &\Leftrightarrow \nabla^0 G_{0\beta} + \nabla^i G_{i\beta} = 0 \Leftrightarrow -\nabla_0 G_{0\beta} + h^{ij} \nabla_j G_{i\beta} = 0 \\ &\Leftrightarrow \partial_0 G_{0\beta} - \Gamma_{00}^\alpha G_{\alpha\beta} - \Gamma_{0\beta}^\alpha G_{0\alpha} = h^{ij} (\partial_j G_{i\beta} - \Gamma_{ji}^\alpha G_{\alpha\beta} - \Gamma_{j\beta}^\alpha G_{i\alpha}).\end{aligned}$$

If the evolution equations hold, we have $G_{ij} = \partial_\alpha G_{ij} = 0$, and so the contracted Bianchi identities become

$$\begin{aligned}&\begin{cases} \partial_0 G_{00} = \Gamma_{00}^\alpha G_{\alpha 0} + \Gamma_{00}^\alpha G_{0\alpha} + h^{ij} (\partial_j G_{i0} - \Gamma_{ji}^\alpha G_{\alpha 0} - \Gamma_{j0}^\alpha G_{i\alpha}) \\ \partial_0 G_{0k} = \Gamma_{00}^0 G_{0k} + \Gamma_{0k}^\alpha G_{0\alpha} - h^{ij} (\Gamma_{ji}^0 G_{0k} + \Gamma_{jk}^0 G_{i0}) \end{cases} \\ &\Leftrightarrow \begin{cases} \partial_0 G_{00} = h^{ij} (\partial_j G_{i0} - K_{ji} G_{00} - \bar{\Gamma}_{ji}^k G_{k0}) \\ \partial_0 G_{0k} = K^i_k G_{0i} - h^{ij} (K_{ji} G_{0k} + K_{jk} G_{i0}) \end{cases}.\end{aligned}$$

This is a system of linear first order partial differential equations on $G_{0\beta}$; integrating the last three equations, and then the first, it is easy to see that, since the initial data vanishes at $t = 0$, the solution vanishes for all t .

REMARK 4.1. In general, any time function $t : M \rightarrow \mathbb{R}$ whose level sets S_t are Cauchy hypersurfaces can be completed into a system of local coordinates (t, x^1, x^2, x^3) . If

$$N = -\frac{\text{grad } t}{|\text{grad } t|}$$

is the future-pointing unit normal to S_t , we have the orthogonal decomposition

$$\frac{\partial}{\partial t} = \alpha N + \beta,$$

where the positive function α is known as the **lapse function** and the vector field β , tangent to S_t , is known as the **shift vector** (Figure 3). In these coordinates, the metric is written

$$\begin{aligned}g &= (-\alpha^2 + \beta_i \beta^i) dt^2 + 2\beta_i dt dx^i + h_{ij} dx^i dx^j \\ &= -\alpha^2 dt^2 + h_{ij} (dx^i - \beta^i dt)(dx^j - \beta^j dt).\end{aligned}$$

Note that the Riemannian metric of the Cauchy hypersurfaces S_t is still $h(t) = h_{ij} dx^i dx^j$; the lapse function and the shift vector merely specify how points with the same coordinates x^i in different Cauchy hypersurfaces S_t are related, which is a matter of choice, that is, a gauge freedom. The Gauss Lemma form of the metric, for instance, corresponds $\alpha = 1$ and $\beta = 0$, but this is not necessarily the best choice.

To prove the existence and uniqueness result for the vacuum Einstein field equations it is best to choose so-called **harmonic coordinates**, that is, coordinates x^μ satisfying the wave equation, $\square x^\mu = 0$. This equation can be written as

$$H^\mu \equiv \partial_\alpha g^{\alpha\mu} + \frac{1}{2} g^{\alpha\mu} g^{\rho\sigma} \partial_\alpha g_{\rho\sigma} = 0$$

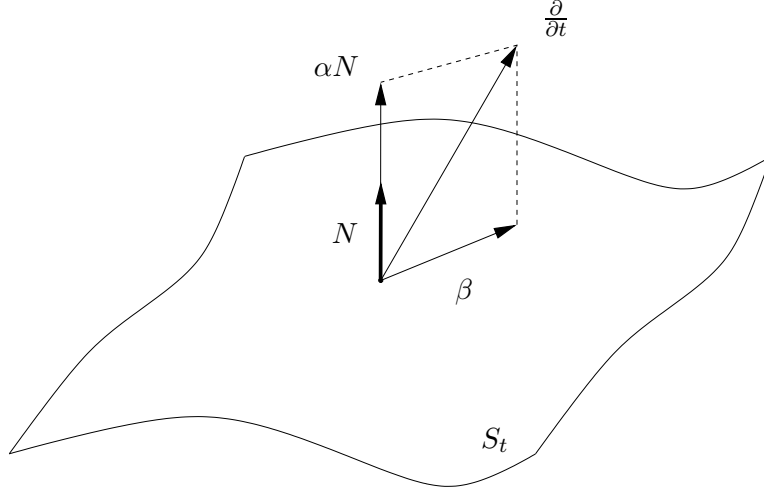


FIGURE 3. Lapse function and shift vector.

We define the **reduced Ricci tensor** to be

$$R_{\mu\nu}^H \equiv R_{\mu\nu} + g_{\alpha(\mu}\partial_{\nu)}H^\alpha = -\frac{1}{2}g^{\alpha\beta}\partial_\alpha\partial_\beta g_{\mu\nu} + F_{\mu\nu}(g, \partial g),$$

and the **reduced Einstein equations** to be

$$R_{\mu\nu}^H = 0.$$

Note that $R_{\mu\nu}^H = R_{\mu\nu}$ if the coordinates x^μ are harmonic; in this case, the reduced Einstein equations coincide with the Einstein equations. Moreover, the reduced Einstein equations are a quasi-linear, diagonal second order hyperbolic system for the components of the metric. Given initial data (h_{ij}, K_{ij}) satisfying the constraint equations, consider the following initial data for the reduced Einstein equations:

- (1) $g_{ij} = h_{ij}$ (forced);
- (2) $g_{i0} = 0$ (by choice);
- (3) $g_{00} = -1$ (by choice);
- (4) $\frac{\partial g_{ij}}{\partial t} = 2K_{ij}$ (forced);
- (5) $\frac{\partial g_{0\mu}}{\partial t}$ such that $H^\mu = 0$ in S .

If (h_{ij}, K_{ij}) is close to the trivial data $(\delta_{ij}, 0)$ for the Minkowski spacetime, Theorem 2.6 guarantees that we can solve the reduced Einstein equations in some open neighborhood V of S . From

$$G_{\mu\nu} = R_{\mu\nu}^H - \frac{1}{2}R^H g_{\mu\nu} - g_{\alpha(\mu}\partial_{\nu)}H^\alpha + \frac{1}{2}g_{\mu\nu}\partial_\alpha H^\alpha$$

it is easily seen that if $g_{\mu\nu}$ satisfies the reduced Einstein equations $R_{\mu\nu}^H = 0$ then the contracted Bianchi identities yield

$$\nabla^\mu G_{\mu\nu} = 0 \Rightarrow g^{\mu\nu}\partial_\mu\partial_\nu H^\alpha + A_\beta^{\alpha\mu}\partial_\mu H^\beta = 0.$$

Moreover, we have from the constraint equations

$$G_{\mu 0} = 0 \Rightarrow \partial_0 H^\mu = 0$$

on S . Therefore H^μ is the solution of a linear, diagonal second order hyperbolic system with vanishing initial conditions. We conclude that $H^\mu = 0$ in V , and therefore $g_{\mu\nu}$ solves the Einstein equations in V .

We can always assume that our initial data is close to the trivial data by rescaling: if x^μ are local coordinates such that $g_{\mu\nu} = \eta_{\mu\nu}$ at some point $p \in S$, where $\eta_{\mu\nu}$ is the Minkowski metric, we define new coordinates \bar{x}^μ by the formula

$$\bar{x}^\mu = \frac{1}{\lambda} x^\mu,$$

where $\lambda > 0$ is a constant. In these new coordinates the metric is

$$\bar{g}_{\mu\nu} = \frac{\partial x^\alpha}{\partial \bar{x}^\mu} \frac{\partial x^\beta}{\partial \bar{x}^\nu} g_{\alpha\beta} = \lambda^2 g_{\mu\nu}.$$

If this metric is a solution of the vacuum Einstein field equations, then so is

$$\tilde{g}_{\mu\nu} = \frac{1}{\lambda^2} \bar{g}_{\mu\nu} = g_{\mu\nu}.$$

Note that this metric satisfies

$$\frac{\partial \tilde{g}_{\mu\nu}}{\partial \bar{t}} = \frac{\partial g_{\mu\nu}}{\partial t} \frac{\partial t}{\partial \bar{t}} = \lambda \frac{\partial g_{\mu\nu}}{\partial t}.$$

Therefore for λ sufficiently small the initial data $(\tilde{g}_{\mu\nu}, \frac{\partial \tilde{g}_{\mu\nu}}{\partial \bar{t}})$ will be close to $(\eta_{\mu\nu}, 0)$.

In this way we can obtain a local solution of the Einstein field equations in a neighborhood of each point $p \in S$. By uniqueness of solution of a quasi-linear, diagonal second order hyperbolic system we can glue these local solutions to obtain a global solution defined on an open neighborhood of S . In other words, given initial data (h, K) satisfying the constraint equations, there exists a globally hyperbolic spacetime (M, g) satisfying the Einstein field equations such that S is a Cauchy surface with induced metric h and second fundamental form K .

Finally, now that we have proved the existence of such solutions, it is possible to prove the existence of a **maximal** solution. The proof is as follows: if (M_1, g_1) and (M_2, g_2) are two solutions of the Einstein field equations containing S with the same initial data (h, K) , we say that $(M_1, g_1) \leq (M_2, g_2)$ if there is an isometric embedding $\psi : M_1 \rightarrow M_2$ preserving (S, h, K) . Note that it is possible that neither $(M_1, g_1) \leq (M_2, g_2)$ nor $(M_2, g_2) \leq (M_1, g_1)$. A set of solutions with the property that any two are related by \leq is called a **chain**; it is clear that every chain has an upper bound (the union up to isometric embeddings). Under these conditions, **Zorn's Lemma** guarantees that there is a maximal element (M, g) in the set of all solutions, that is, a solution which cannot be isometrically embedded into any other solution. It is possible to prove that this element is unique (if there were two such

maximal solutions it would be possible to patch them together to construct a larger solution). We then have the following fundamental result.

THEOREM 4.2. (**Choquet-Bruhat [FB55, CBG69]**) *Let (S, h) be a 3-dimensional Riemannian manifold and K a symmetric tensor field in S satisfying the constraint equations*

$$\begin{cases} \bar{R} + (K^i_i)^2 - K_{ij}K^{ij} = 0 \\ \bar{\nabla}_i K^j_j - \bar{\nabla}_j K^j_i = 0 \end{cases},$$

where $\bar{\nabla}$ and \bar{R} are the Levi-Civita connection and the scalar curvature of h . Then there exists a unique (up to isometry) 4-dimensional Lorentzian manifold (M, g) , called the **maximal Cauchy development** of (S, h, K) , satisfying:

- (i) (M, g) is a solution of the vacuum Einstein equations;
- (ii) (M, g) is globally hyperbolic with Cauchy surface S ;
- (iii) The induced metric and second fundamental forms of S are h and K ;
- (iv) Any 4-dimensional Lorentzian manifold satisfying (i) – (iii) can be isometrically embedded into (M, g) .

Moreover, if (S, h, K) and $(\bar{S}, \bar{h}, \bar{K})$ coincide on some closed subset $B \cong \bar{B}$ then $D(B)$ and $D(\bar{B})$ are isometric. Finally, g depends continuously on the initial data (h, K) (for appropriate topologies).

5. Constraint equations

To obtain initial data for the Einstein equations it is necessary to solve the nonlinear constraint equations

$$\begin{cases} \bar{R} + (K^i_i)^2 - K_{ij}K^{ij} = 0 \\ \bar{\nabla}_i K^j_j - \bar{\nabla}_j K^j_i = 0 \end{cases}.$$

The **Lichnerowicz method** for solving these equations is as follows: one starts by choosing an arbitrary Riemannian metric h and an arbitrary symmetric tensor K satisfying

$$\begin{cases} K^i_i = 0 \\ \bar{\nabla}_j K^j_i = 0 \end{cases}$$

(that is, K is traceless and divergenceless). These choices satisfy the second, but not the first, constraint equations. One then defines the conformally rescaled metric

$$\tilde{h} = u^4 h$$

and the rescaled symmetric tensor

$$\tilde{K} = u^{-2} K$$

Clearly \tilde{K} is still traceless, and it is easily seen that it is also divergenceless:

$$\tilde{\nabla}_j \tilde{K}^j_i = 0,$$

where $\tilde{\nabla}$ is the Levi-Civita connection of \tilde{h} . The first constraint equation for the metric \tilde{h} and the symmetric tensor \tilde{K} , on the other hand, becomes

$$\tilde{R} - \tilde{K}_{ij}\tilde{K}^{ij} = 0 \Leftrightarrow \bar{\Delta}u - \frac{1}{8}\bar{R}u + \frac{1}{8}u^{-7}K_{ij}K^{ij} = 0.$$

This is a nonlinear elliptic equation on one variable, much simpler than the original system. If one chooses the so-called time symmetric case $K = 0$ (the reason for this designation being that in the Gauss Lemma coordinates $t \rightarrow -t$ is clearly an isometry of the solution), this equation becomes linear:

$$\bar{\Delta}u - \frac{1}{8}\bar{R}u = 0.$$

As an example, choose h to be the Euclidean metric, $h_{ij} = \delta_{ij}$; then $\bar{R} = 0$ and the equation above is simply the Laplace equation

$$\Delta u = 0.$$

A simple solution, related to the gravitational field of a point mass, is

$$u = 1 + \frac{M}{2r}.$$

This solution leads exactly to the Schwarzschild solution, which in isotropic coordinates is written

$$ds^2 = - \left(\frac{1 - \frac{M}{2r}}{1 + \frac{M}{2r}} \right)^2 dt^2 + \left(1 + \frac{M}{2r} \right)^4 (dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2)).$$

Since the Laplace equation is linear, one can superimpose solutions. Thus an initial data set for a set of N black holes initially at rest can be obtained by choosing

$$u = \sum_{i=1}^N \left(1 + \frac{M_i}{2r_i} \right),$$

where $M_i > 0$ and r_i is the Euclidean distance to a fixed point $p_i \in \mathbb{R}^3$ ($i = 1, \dots, N$).

6. Einstein equations with matter

So far we have analyzed only the vacuum Einstein field equations. To include matter (and also a cosmological constant Λ) we must introduce matter fields, generically represented by ψ , and consider a system of the form

$$\begin{cases} G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu}(g, \psi) \\ \text{Field equations for } \psi \end{cases}.$$

If the equations for ψ form a hyperbolic system then Choquet-Bruhat's theorem still applies, with the constraint equations

$$\begin{cases} \bar{R} + (K^i_i)^2 - K_{ij}K^{ij} - 2\Lambda = 16\pi\rho \\ \bar{\nabla}_i K^j_j - \bar{\nabla}_j K^j_i = 8\pi J_i \end{cases}.$$

Here $\rho = T_{00}$ and $J_i = -T_{0i}$ are computed from h and the initial data for ψ . As an example, the Einstein-Klein-Gordon system is given by

$$\begin{cases} G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi (\partial_\mu \phi \partial_\nu \phi - \frac{1}{2} g_{\mu\nu} (\partial_\alpha \phi \partial^\alpha \phi + m^2 \phi^2)) \\ \nabla^\mu \partial_\mu \phi - m^2 \phi = 0 \end{cases},$$

and we have

$$\begin{cases} \rho = \frac{1}{2} (\psi_0^2 + h^{ij} \partial_i \phi_0 \partial_j \phi_0 + m^2 \phi_0^2) \\ J_i = -\psi_0 \partial_i \phi_0 \end{cases},$$

where ϕ_0 and ψ_0 are the initial data for ϕ .

7. Exercises

- (1) Let (M, g) be a time oriented Lorentzian manifold and X a future-pointing timelike vector field. Given a smooth function $\phi : M \rightarrow \mathbb{R}$ let

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2} g_{\mu\nu} (\partial_\alpha \phi \partial^\alpha \phi + m^2 \phi^2)$$

be the energy-momentum tensor associated to the Klein-Gordon equation for ϕ , and let Y be the vector field defined by

$$Y_\mu = T_{\mu\nu} X^\nu.$$

Show that:

- (a) $Y = (X \cdot \phi) \text{grad } \phi - \frac{1}{2} (\langle \text{grad } \phi, \text{grad } \phi \rangle + m^2 \phi^2) X$;
 (b) Y is causal.
 (c) Y is past-pointing.
- (2) (**Sobolev inequality**) Let Q be a closed solid cone in \mathbb{R}^n with height H , solid angle Ω and vertex at the origin. Let $\psi : \mathbb{R} \rightarrow \mathbb{R}$ be a smooth nonincreasing function with $\psi(r) = 1$ for $r < \frac{H}{3}$ and $\psi(r) = 0$ for $r > \frac{2H}{3}$. Show that:

- (a) For any smooth function $f : Q \rightarrow \mathbb{R}$ and any $k \in \mathbb{N}$ we have

$$f(0) = \frac{(-1)^k}{(k-1)!} \int_0^{R(\theta)} r^{k-1} \frac{\partial^k}{\partial r^k} (\psi(r) f(r, \theta)) dr,$$

where (r, θ) are the usual spherical coordinates in \mathbb{R}^n and $r = R(\theta)$ is the equation for the base of the cone (here $f(r, \theta)$ represents the function f written in spherical coordinates).

- (b) There exists a constant C , depending on k and Ω , such that

$$f(0) = C \int_Q r^{k-n} \frac{\partial^k}{\partial r^k} (\psi f).$$

- (c) For $k > \frac{n}{2}$ we have $|f(0)| \leq C' \|f\|_{H^k(Q)}$, where the constant C' depends on k , H and Ω only (you will need to use the Cauchy-Schwarz inequality for multiple integrals).

- (3) Consider a Lorentzian metric given in the Gauss Lemma form

$$g = -dt^2 + h_{ij}(t, x)dx^i dx^j,$$

so that the level sets of t are Riemannian manifolds with induced metric $h(t) = h_{ij}dx^i dx^j$ and second fundamental form

$$K(t) = \frac{1}{2} \frac{\partial h_{ij}}{\partial t} dx^i dx^j.$$

Show that in these coordinates:

- (a) The Christoffel symbols are

$$\Gamma_{ij}^0 = K_{ij}; \quad \Gamma_{jk}^i = \bar{\Gamma}_{jk}^i; \quad \Gamma_{0j}^i = K^i_j,$$

where $\bar{\Gamma}_{jk}^i$ are the Christoffel symbols of h .

- (b) The components of the Riemann tensor are

$$\begin{aligned} R_{0i0}^j &= -\frac{\partial}{\partial t} K^j_i - K_{il} K^{lj}; \\ R_{ij0}^l &= -\bar{\nabla}_i K^l_j + \bar{\nabla}_j K^l_i; \\ R_{ijl}^m &= \bar{R}_{ijl}^m + K_{il} K^m_j - K_{jl} K^m_i, \end{aligned}$$

where $\bar{\nabla}$ is the Levi-Civita connection of h and \bar{R}_{ijl}^m are the components of the Riemann tensor of h .

- (c) The time derivative of the inverse metric is given by the formula

$$\frac{\partial h^{ij}}{\partial t} = -2K^{ij}.$$

- (d) The components of the Ricci tensor are

$$\begin{aligned} R_{00} &= -\frac{\partial}{\partial t} K^i_i - K_{ij} K^{ij}; \\ R_{0i} &= -\bar{\nabla}_i K^j_j + \bar{\nabla}_j K^j_i; \\ R_{ij} &= \bar{R}_{ij} + \frac{\partial}{\partial t} K_{ij} - 2K_{il} K^l_j + K^l_l K_{ij}, \end{aligned}$$

where \bar{R}_{ij} are the components of the Ricci tensor of h .

- (e) The scalar curvature is

$$R = \bar{R} + 2\frac{\partial}{\partial t} K^i_i + (K^i_i)^2 + K_{ij} K^{ij},$$

where \bar{R} is the scalar curvature of h .

- (f) The component
- G_{00}
- of the Einstein tensor is

$$G_{00} = \frac{1}{2} \left(\bar{R} + (K^i_i)^2 - K_{ij} K^{ij} \right).$$

- (4) Let
- (M, g)
- be an
- $(n + 1)$
- dimensional Lorentzian manifold and
- (x^0, \dots, x^n)
- local coordinates on
- M
- . Show that:

(a) The condition for these coordinates to be harmonic is written

$$\nabla_\alpha \nabla^\alpha x^\mu = 0 \Leftrightarrow H^\mu \equiv \partial_\alpha g^{\alpha\mu} + \frac{1}{2} g^{\alpha\mu} g^{\rho\sigma} \partial_\alpha g_{\rho\sigma} = 0$$

($\mu = 0, \dots, n$).

(b) The reduced Ricci tensor is

$$R_{\mu\nu}^H \equiv R_{\mu\nu} + g_{\alpha(\mu} \partial_{\nu)} H^\alpha = -\frac{1}{2} g^{\alpha\beta} \partial_\alpha \partial_\beta g_{\mu\nu} + F_{\mu\nu}(g, \partial g).$$

(5) Denoting by h_0 the standard constant curvature metric on \mathbb{R}^3 , S^3 or H^3 , compute the cosmological constant and determine the maximal globally hyperbolic developments of the following sets of initial data for the vacuum Einstein equations (with cosmological constant):

- (a) $(\mathbb{R}^3, h_0, 0)$;
- (b) (\mathbb{R}^3, h_0, h_0) ;
- (c) $(S^3, h_0, 0)$;
- (d) $(H^3, h_0, 0)$;
- (e) (H^3, h_0, h_0) ;
- (f) $(\mathbb{R}^3, h_0, \text{diag}(p_1, p_2, p_3))$ with $p_1 + p_2 + p_3 = p_1^2 + p_2^2 + p_3^2 = 1$.

(6) Let (S, h, K) be an initial data set for the vacuum Einstein equations, with K traceless and divergenceless. Given a smooth positive function $u : S \rightarrow \mathbb{R}$, consider the conformally rescaled metric $\tilde{h} = u^4 h$ and the symmetric tensor $\tilde{K} = u^{-2} K$. By using normal coordinates when convenient, show that:

(a) The Christoffel symbols $\tilde{\Gamma}_{jk}^i$ of \tilde{h} are related to the Christoffel symbols $\bar{\Gamma}_{jk}^i$ of h by

$$\tilde{\Gamma}_{jk}^i = \bar{\Gamma}_{jk}^i + 2\partial_j(\log u)h^i{}_k + 2\partial_k(\log u)h^i{}_j - 2\partial^i(\log u)h_{jk}.$$

(b) \tilde{K} is divergenceless for the Levi-Civita connection of \tilde{h} .

(c) The Ricci tensor \tilde{R}_{ij} of \tilde{h} is related to the Ricci tensor \bar{R}_{ij} of h by

$$\begin{aligned} \tilde{R}_{ij} &= \bar{R}_{ij} - 2\bar{\nabla}_i \partial_j(\log u) - 2\bar{\Delta}(\log u)h_{ij} \\ &\quad + 4\partial_i(\log u)\partial_j(\log u) - 4|\text{grad}(\log u)|^2 h_{ij}. \end{aligned}$$

(d) The scalar curvature \tilde{R} of \tilde{h} is related to the scalar curvature \bar{R} of h by

$$\tilde{R} = u^{-4}\bar{R} - 8u^{-5}\bar{\Delta}u.$$

(7) Check that the metric

$$ds^2 = -\left(\frac{1 - \frac{M}{2r}}{1 + \frac{M}{2r}}\right)^2 dt^2 + \left(1 + \frac{M}{2r}\right)^4 (dr^2 + r^2(d\theta^2 + \sin^2\theta d\varphi^2))$$

is indeed the Schwarzschild metric by making the coordinate change

$$R = r \left(1 + \frac{M}{2r}\right)^2.$$

CHAPTER 6

Positive mass theorem

In this chapter we present the positive mass theorem. Following [Wal84], we start by defining the Komar mass for stationary spacetimes. We then discuss field theory and introduce the Einstein-Hilbert action as a means of motivating the definition of the ADM mass. Finally, we prove the (Riemannian) positive mass theorem and the (Riemannian) Penrose inequality for graphs, following [Lam10]. For more details see [Mar09, Bra11].

1. Komar mass

Recall that the Newtonian gravitational field satisfies

$$\operatorname{div} \mathbf{G} = -4\pi\rho,$$

where ρ is the mass density of the matter generating the field. Therefore the total mass of a given system is given by

$$M = \int_{\mathbb{R}^3} \rho = -\frac{1}{4\pi} \int_{\mathbb{R}^3} \operatorname{div} \mathbf{G} = -\frac{1}{4\pi} \int_{\Sigma} \langle \mathbf{G}, \mathbf{n} \rangle,$$

where Σ is any surface enclosing all the matter and \mathbf{n} is the outward unit normal. The fact that M does not depend on Σ is equivalent to the statement that $\operatorname{div} \mathbf{G} = 0$ in the region between any two such surfaces.

For a **static** Lorentzian metric, that is, a metric of the form

$$ds^2 = -e^{2\phi} dt^2 + h_{ij} dx^i dx^j$$

with ϕ and h not depending on t , the analogue of the gravitational field is minus the acceleration of the observers with constant space coordinates x^i , that is, $-\operatorname{grad} \phi$ (see Chapter 2). On the other hand, since the metric is static, we expect that the energy computed at a given surface should be multiplied by the redshift factor e^ϕ to obtain its reference value at infinity. The relativistic analogue of the formula above is then

$$M = \frac{1}{4\pi} \int_{\Sigma} e^\phi (\partial_\mu \phi) n^\mu = \frac{1}{4\pi} \int_{\Sigma} (\partial_\mu e^\phi) n^\mu.$$

We have

$$\partial_\mu e^\phi = \partial_\mu (-K^\nu K_\nu)^{\frac{1}{2}} = -e^{-\phi} K^\nu \nabla_\mu K_\nu = -N^\nu \nabla_\mu K_\nu,$$

where $K = \frac{\partial}{\partial t}$ is the timelike Killing vector field and $N = e^{-\phi} \frac{\partial}{\partial t}$ is the unit timelike vector with the same direction, that is, the future-pointing unit

normal to the hypersurfaces S_t of constant t (see Figure 1). Therefore, we can write

$$M = -\frac{1}{4\pi} \int_{\Sigma} (\nabla_{\mu} K_{\nu}) n^{\mu} N^{\nu} = \frac{1}{4\pi} \int_{\Sigma} (\nabla_{\mu} K_{\nu}) N^{\mu} n^{\nu}.$$

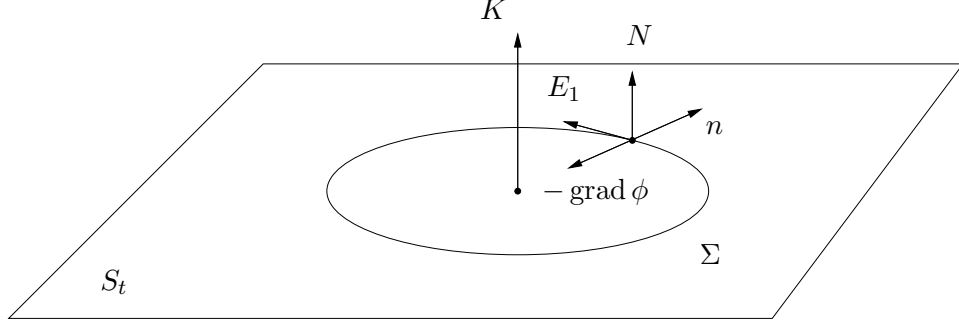


FIGURE 1. Computing the Komar mass on a static spacetime.

Because K is a Killing vector field, $\nabla_{\mu} K_{\nu}$ is a 2-form; more precisely,

$$(dK^{\sharp})_{\mu\nu} = \nabla_{\mu} K_{\nu} - \nabla_{\nu} K_{\mu} = 2\nabla_{[\mu} K_{\nu]}.$$

If E_1 and E_2 are two unit vector fields tangent to Σ such that $\{N, n, E_1, E_2\}$ is a positive orthonormal frame (see Figure 1), and so $\{-N^{\sharp}, n^{\sharp}, E_1^{\sharp}, E_2^{\sharp}\}$ is a positive orthonormal coframe, then we can expand

$$\nabla K^{\sharp} = -\nabla K^{\sharp}(N, n) N^{\sharp} \wedge n^{\sharp} + \dots,$$

and so

$$\begin{aligned} M &= \frac{1}{4\pi} \int_{\Sigma} \nabla K^{\sharp}(N, n) = \frac{1}{4\pi} \int_{\Sigma} \nabla K^{\sharp}(N, n) E_1^{\sharp} \wedge E_2^{\sharp} \\ &= -\frac{1}{4\pi} \int_{\Sigma} -\nabla K^{\sharp}(N, n) \star (N^{\sharp} \wedge n^{\sharp}) = -\frac{1}{4\pi} \int_{\Sigma} \star \nabla K^{\sharp}, \end{aligned}$$

that is

$$M = -\frac{1}{8\pi} \int_{\Sigma} \star dK^{\sharp}.$$

This expression is the so-called **Komar mass**. Although we arrived at this expression by considering a **static** space, it actually works for any **stationary** spacetime. In other words, the timelike Killing vector field K does not have to be hypersurface-orthogonal.

To show that the Komar mass is well defined, that is, that $\star dK^{\sharp}$ is a closed 2-form in vacuum, we start by noticing that if X is any vector field then

$$(\operatorname{div} X)\epsilon = d(X \lrcorner \epsilon) = d \star X^{\sharp},$$

whence

$$\operatorname{div} X = -\star d \star X^{\sharp}.$$

In particular, for any smooth function ϕ

$$\square\phi = \operatorname{div} \operatorname{grad} \phi = -\star d\star(\operatorname{grad} \phi)^\sharp = -\star d\star d\phi.$$

In local coordinates, we have

$$\begin{aligned} \star d\phi &= \epsilon(\operatorname{grad} \phi, \cdot, \cdot, \cdot) = \sqrt{|\det(g_{\mu\nu})|} dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 (\operatorname{grad} \phi, \cdot, \cdot, \cdot) \\ &= \sqrt{|\det(g_{\mu\nu})|} \partial^0 \phi dx^1 \wedge dx^2 \wedge dx^3 - \sqrt{|\det(g_{\mu\nu})|} \partial^1 \phi dx^0 \wedge dx^2 \wedge dx^3 + \dots \end{aligned}$$

Therefore

$$\begin{aligned} d\star d\phi &= \partial_0 \left(\sqrt{|\det(g_{\mu\nu})|} \partial^0 \phi \right) dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \\ &\quad + \partial_1 \left(\sqrt{|\det(g_{\mu\nu})|} \partial^1 \phi \right) dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 + \dots \end{aligned}$$

and we obtain the useful formula

$$\square\phi = \frac{1}{\sqrt{|\det(g_{\mu\nu})|}} \partial_\alpha \left(\sqrt{|\det(g_{\mu\nu})|} \partial^\alpha \phi \right).$$

It is natural to try and generalize this formula for arbitrary k -forms.

DEFINITION 1.1. *If ω is a k -form then its **Hodge d'Alembertian** is the k -form*

$$\square_H \omega = -\star d\star d\omega - d\star d\star \omega.$$

It turns out that in general the Hodge d'Alembertian does not coincide with the usual d'Alembertian

$$\square\omega = \nabla_\mu \nabla^\mu \omega$$

(sometimes called the **rough d'Alembertian**). The relation between these two operators for 1-forms is given by the following result.

THEOREM 1.2. (Weitzenbock formula) *If ω is a 1-form then*

$$\square_H \omega - \square\omega = -\operatorname{Ric}(\omega^\sharp, \cdot).$$

PROOF. We have

$$\begin{aligned} (\star d\star d\omega)_\delta &= \frac{3 \cdot 2}{3! 2!} \epsilon_{\gamma\alpha\beta\delta} \nabla^\gamma (\epsilon^{\mu\nu\alpha\beta} \nabla_\mu \omega_\nu) = \frac{1}{2!} \epsilon_{\alpha\beta\gamma\delta} \epsilon^{\mu\nu\alpha\beta} \nabla^\gamma \nabla_\mu \omega_\nu \\ &= (-g^\mu{}_\gamma g^\nu{}_\delta + g^\mu{}_\delta g^\nu{}_\gamma) \nabla^\gamma \nabla_\mu \omega_\nu = -\nabla^\mu \nabla_\mu \omega_\delta + \nabla^\nu \nabla_\delta \omega_\nu \end{aligned}$$

and

$$\begin{aligned} (d\star d\star \omega)_\delta &= \frac{4}{4!} \nabla_\delta (\epsilon_{\gamma\nu\alpha\beta} \nabla^\gamma (\epsilon^{\mu\nu\alpha\beta} \omega_\mu)) = \frac{1}{3!} \epsilon_{\gamma\nu\alpha\beta} \epsilon^{\mu\nu\alpha\beta} \nabla_\delta \nabla^\gamma \omega_\mu \\ &= -g^\mu{}_\gamma \nabla_\delta \nabla^\gamma \omega_\mu = -\nabla_\delta \nabla^\mu \omega_\mu. \end{aligned}$$

Therefore

$$(\star d\star d\omega + d\star d\star \omega)_\delta = -(\square\omega)_\delta + \nabla^\mu \nabla_\delta \omega_\mu - \nabla_\delta \nabla^\mu \omega_\mu.$$

The Weitzenbock formula now follows from

$$\nabla_\mu \nabla_\delta \omega^\mu - \nabla_\delta \nabla_\mu \omega^\mu = R_{\mu\delta}{}^\mu{}_\nu \omega^\nu = R_{\delta\nu} \omega^\nu.$$

□

Now let $K = \frac{\partial}{\partial t}$ be the timelike Killing vector field in a stationary spacetime. Then

$$\begin{aligned} d \star K^\sharp &= d \left(\frac{\partial}{\partial t} \lrcorner \left(\sqrt{|\det(g_{\mu\nu})|} dt \wedge dx^1 \wedge dx^2 \wedge dx^3 \right) \right) \\ &= d \left(\sqrt{|\det(g_{\mu\nu})|} dx^1 \wedge dx^2 \wedge dx^3 \right) \\ &= \frac{\partial}{\partial t} \sqrt{|\det(g_{\mu\nu})|} dx^1 \wedge dx^2 \wedge dx^3 = 0, \end{aligned}$$

since the metric coefficients do not depend on t . From the Weitzenböck formula we then have

$$\star d \star dK^\sharp = -\square K^\sharp + Ric(K, \cdot).$$

Now, adding cyclic permutations of the fundamental identity for the Riemann tensor,

$$\begin{aligned} \nabla_\mu \nabla_\nu K_\alpha - \nabla_\nu \nabla_\mu K_\alpha &= R_{\mu\nu\alpha}{}^\beta K_\beta, \\ \nabla_\nu \nabla_\alpha K_\mu - \nabla_\alpha \nabla_\nu K_\mu &= R_{\nu\alpha\mu}{}^\beta K_\beta, \\ \nabla_\alpha \nabla_\mu K_\nu - \nabla_\mu \nabla_\alpha K_\nu &= R_{\alpha\mu\nu}{}^\beta K_\beta, \end{aligned}$$

and using the Killing equation and the first Bianchi identity, we have

$$\nabla_\mu \nabla_\nu K_\alpha = -R_{\nu\alpha\mu}{}^\beta K_\beta,$$

whence

$$\square K_\alpha = -R_\alpha{}^\beta K_\beta.$$

We conclude that

$$\star d \star dK^\sharp = 2Ric(K, \cdot) \Leftrightarrow d \star dK^\sharp = 2 \star Ric(K, \cdot),$$

implying that $\star dK^\sharp$ is indeed closed in vacuum, that is, the Komar mass is well defined: any two homologous compact orientable surfaces Σ_1 and Σ_2 which enclose the matter content of the stationary spacetime can be used to compute it (Figure 2).

If Σ is the boundary of a spacelike 3-dimensional manifold B whose future-pointing unit normal is N then the Komar mass can be written as

$$\begin{aligned} M &= -\frac{1}{8\pi} \int_\Sigma \star dK^\sharp = -\frac{1}{8\pi} \int_B d \star dK^\sharp = -\frac{1}{4\pi} \int_B \star Ric(K, \cdot) \\ &= -\frac{1}{4\pi} \int_B \epsilon(Ric(K, \cdot)^\sharp, \cdot, \cdot, \cdot) = -\frac{1}{4\pi} \int_B -\langle (Ric(K, \cdot)^\sharp, N) \rangle \\ &= \frac{1}{4\pi} \int_B R_{\mu\nu} K^\mu N^\nu = 2 \int_B \left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right) K^\mu N^\nu \end{aligned}$$

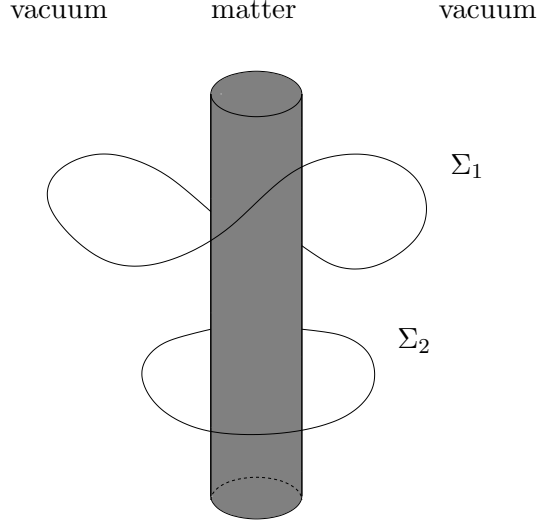


FIGURE 2. Computing the Komar mass with two homologous surfaces.

(we used $\epsilon = -N^\sharp \wedge \sigma$ in the second line, where σ is the volume element for B). Note that the Komar mass is not, as one might have guessed, the integral

$$M' = \int_B T_{\mu\nu} K^\mu N^\nu.$$

This integral is also well defined in a stationary spacetime, since

$$\nabla^\mu (T_{\mu\nu} K^\nu) = (\nabla^\mu T_{\mu\nu}) K^\nu + T_{\mu\nu} \nabla^\mu K^\nu = 0,$$

due to the contracted Bianchi identity and the Killing equation. However, we also have

$$\nabla^\mu (T K_\mu) = K^\mu \nabla_\mu T + T \nabla^\mu K_\mu = 0,$$

because T is constant along K , which has zero divergence (by contracting the Killing equation). Therefore we also have

$$\nabla^\mu \left(\left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right) K^\nu \right) = 0.$$

To have an idea of what exactly is measured by the Komar mass, consider a static spacetime, where it is possible to choose B such that $N = e^{-\phi} K$. In this case, for a perfect fluid (whose flow lines are necessarily the integral curves of K), we have

$$M = 2 \int_B \left(T(N, N) + \frac{1}{2} T \right) e^\phi = 2 \int_B \left(\rho + \frac{1}{2} (-\rho + 3p) \right) e^\phi = \int_B (\rho + 3p) e^\phi.$$

Thus we see that the Komar mass also includes the pressure; this is reminiscent of the Newtonian formula for the internal energy of a monoatomic

gas,

$$U = \frac{3}{2}pV.$$

2. Field theory

Let us consider the problem of how to define the energy of a field ψ in flat Minkowski space. The field equations for the field ψ are usually the equations for the critical points of an **action**

$$S = \int_{\mathbb{R}^4} \mathcal{L}(\psi, \partial\psi) dt dx^1 dx^2 dx^3,$$

obtained by integrating a **Lagrangian density** \mathcal{L} (where we assume that the field decays fast enough so that S is well defined). The field equations are then

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \right) - \frac{\partial \mathcal{L}}{\partial \psi} = 0.$$

Indeed, if $\psi(\lambda)$ is a one-parameter family of fields such that $\psi(0)$ is a critical point of the action, and $\delta \equiv \frac{d}{d\lambda}|_{\lambda=0}$, then

$$\begin{aligned} \delta S &= \int_{\mathbb{R}^4} \left(\frac{\partial \mathcal{L}}{\partial \psi} \delta \psi + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \delta(\partial_\mu \psi) \right) = \int_{\mathbb{R}^4} \left(\frac{\partial \mathcal{L}}{\partial \psi} \delta \psi + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \partial_\mu(\delta \psi) \right) \\ &= \int_{\mathbb{R}^4} \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \delta \psi \right) + \int_{\mathbb{R}^4} \left(\frac{\partial \mathcal{L}}{\partial \psi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \right) \right) \delta \psi. \end{aligned}$$

For field variations $\delta \psi$ with compact support we then have

$$\delta S = \int_{\mathbb{R}^4} \left(\frac{\partial \mathcal{L}}{\partial \psi} - \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \right) \right) \delta \psi.$$

The **canonical energy-momentum tensor** is defined as

$$T^\mu{}_\nu = \frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \partial_\nu \psi - \mathcal{L} \delta^\mu{}_\nu$$

and satisfies

$$\begin{aligned} \partial_\mu T^\mu{}_\nu &= \partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \right) \partial_\nu \psi + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \psi)} \partial_\mu \partial_\nu \psi \\ &\quad - \frac{\partial \mathcal{L}}{\partial \psi} (\partial_\mu \psi) \delta^\mu{}_\nu - \frac{\partial \mathcal{L}}{\partial(\partial_\alpha \psi)} (\partial_\mu \partial_\alpha \psi) \delta^\mu{}_\nu = 0. \end{aligned}$$

Defining the **Hamiltonian density** to be

$$\mathcal{H} = -T^0{}_0 = -\frac{\partial \mathcal{L}}{\partial(\partial_0 \psi)} \partial_0 \psi + \mathcal{L},$$

it is then clear from the divergence theorem that the **Hamiltonian**

$$H = \int_{S_t} \mathcal{H} dx^1 dx^2 dx^3$$

is independent of the hypersurface S_t of constant t chosen to compute the integral (assuming that the field decays fast enough so that H is well defined

and the boundary integral corresponding to the divergence term vanishes). The Hamiltonian can be identified with the total energy of the field ψ , which is therefore constant in time.

If we have N fields ψ_1, \dots, ψ_N instead of a single field ψ , then it is easily seen that the field equations are

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial (\partial_\mu \psi_i)} \right) - \frac{\partial \mathcal{L}}{\partial \psi_i} = 0 \quad (i = 1, \dots, N),$$

the canonical energy-momentum tensor is

$$T^\mu{}_\nu = \frac{\partial \mathcal{L}}{\partial (\partial_\mu \psi_i)} \partial_\nu \psi_i - \mathcal{L} \delta^\mu{}_\nu$$

(summed over i), and the Hamiltonian density is

$$\mathcal{H} = -T^0{}_0 = -\frac{\partial \mathcal{L}}{\partial (\partial_0 \psi_i)} \partial_0 \psi_i + \mathcal{L}.$$

Note carefully that up until this point the metric was not used to raise or lower indices, or even to apply the divergence theorem. This will be important in Section 5.

If we use Cartesian coordinates then the tensor

$$T^{\mu\nu} = \frac{\partial \mathcal{L}}{\partial (\partial_\mu \psi_i)} \partial^\nu \psi_i - \mathcal{L} g^{\mu\nu}$$

satisfies the conservation equation

$$\nabla_\mu T^{\mu\nu} = 0.$$

This provides a method for obtaining the energy-momentum tensor that is used in the Einstein field equations for the various matter models. We now list some simple examples.

2.1. Klein-Gordon field. The Lagrangian density for the Klein-Gordon field is

$$\mathcal{L} = \frac{1}{2} (g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi + m^2 \phi^2).$$

Consequently the canonical energy-momentum tensor for the Klein-Gordon field is

$$T^{\mu\nu} = \partial^\mu \phi \partial^\nu \phi - \frac{1}{2} g^{\mu\nu} (\partial_\alpha \phi \partial^\alpha \phi + m^2 \phi^2),$$

as claimed in Chapter 5.

2.2. Electromagnetic field. In this case we can take as our fields the electromagnetic potentials A_0, A_1, A_2, A_3 . The Lagrangian density (in units where $4\pi\epsilon_0 = 1$) is

$$\mathcal{L} = \frac{1}{16\pi} g^{\alpha\mu} g^{\beta\nu} F_{\alpha\beta} F_{\mu\nu},$$

where

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu.$$

Using

$$\frac{\partial F_{\alpha\beta}}{\partial(\partial_\mu A_\nu)} = \delta^\mu_\alpha \delta^\nu_\beta - \delta^\nu_\alpha \delta^\mu_\beta,$$

it is easily seen that

$$\partial_\mu \left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\nu)} \right) - \frac{\partial \mathcal{L}}{\partial A_\nu} = 0 \Leftrightarrow \partial_\mu F^{\mu\nu} = 0,$$

which are indeed the Maxwell equations $d \star F = 0$, as

$$\begin{aligned} \epsilon_{\mu\gamma\delta\nu} \nabla^\mu \epsilon^{\alpha\beta\gamma\delta} F_{\alpha\beta} &= \epsilon_{\mu\nu\gamma\delta} \epsilon^{\alpha\beta\gamma\delta} \nabla^\mu F_{\alpha\beta} \\ &= -2(\delta^\mu_\alpha \delta^\nu_\beta - \delta^\nu_\alpha \delta^\mu_\beta) \nabla^\mu F_{\alpha\beta} = -2\nabla^\mu F_{\mu\nu}. \end{aligned}$$

Note that the equations $dF = 0$ follow automatically from the definition $F = dA$. The canonical energy-momentum tensor for the electromagnetic field is

$$T_{\text{can}}^{\mu\nu} = \frac{\partial \mathcal{L}}{\partial(\partial_\mu A_\alpha)} \partial^\nu A_\alpha - \mathcal{L} g^{\mu\nu} = \frac{1}{4\pi} \left(F^{\mu\alpha} \partial^\nu A_\alpha - \frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} g^{\mu\nu} \right).$$

This tensor is neither symmetric nor gauge-invariant. However,

$$F^{\mu\alpha} \partial^\nu A_\alpha = F^{\mu\alpha} F^\nu{}_\alpha + F^{\mu\alpha} \partial_\alpha A^\nu,$$

where the first term is symmetric and gauge-invariant, and the second term is divergenceless:

$$\partial_\mu (F^{\mu\alpha} \partial_\alpha A^\nu) = F^{\mu\alpha} \partial_\mu \partial_\alpha A^\nu = 0$$

(because F is antisymmetric and the partial derivatives commute). Therefore the tensor

$$T^{\mu\nu} = \frac{1}{4\pi} \left(F^{\mu\alpha} F^\nu{}_\alpha - \frac{1}{4} F_{\alpha\beta} F^{\alpha\beta} g^{\mu\nu} \right)$$

is symmetric, gauge-invariant and divergenceless. This is the true energy-momentum tensor for the electromagnetic field.

2.3. Relativistic elasticity. A continuous medium can be described by a Riemannian 3-manifold (S, k) (the **relaxed configuration**) and projection map $\pi : \mathbb{R}^4 \rightarrow S$ whose level sets are timelike curves (the worldlines of the medium particles), as shown in Figure 3.

If we choose local coordinates $(\bar{x}^1, \bar{x}^2, \bar{x}^3)$ on S then we can think of π as a set of three scalar fields $\bar{x}^1, \bar{x}^2, \bar{x}^3 : \mathbb{R}^4 \rightarrow \mathbb{R}$. For a given worldline, we can complete this set of scalar fields into local coordinates $(\bar{t}, \bar{x}^1, \bar{x}^2, \bar{x}^3)$ for \mathbb{R}^4 such that \bar{t} is the proper time along that worldline and its level sets are orthogonal to it:

$$g = -d\bar{t}^2 + h_{ij} d\bar{x}^i d\bar{x}^j \quad (\text{on the worldline}).$$

Notice that the orthogonal metric

$$h = h_{ij} d\bar{x}^i d\bar{x}^j$$

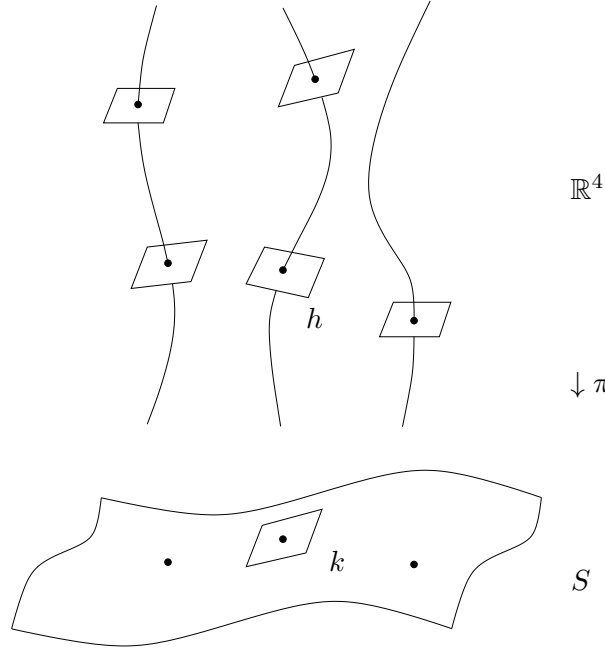


FIGURE 3. A continuous medium in Minkowski's spacetime.

can be thought of as a time-dependent Riemannian metric on S , describing the local deformations of the medium along each worldline, that is, the deviations from the natural metric

$$k = k_{ij} d\bar{x}^i d\bar{x}^j.$$

We can compute the (inverse) metric h from

$$h^{ij} = g^{\mu\nu} \frac{\partial \bar{x}^i}{\partial x^\mu} \frac{\partial \bar{x}^j}{\partial x^\nu},$$

which does not depend on the choice of \bar{t} . In other words, the metric h is a quadratic function of the partial derivatives of the fields $\bar{x}^1, \bar{x}^2, \bar{x}^3$. An elastic Lagrangian density \mathcal{L} for these fields is obtained by assuming that $\mathcal{L} = \mathcal{L}(\bar{x}^i, h^{ij})$. The canonical energy-momentum tensor is

$$T^{\mu\nu} = \frac{\partial \mathcal{L}}{\partial (\partial_\mu \bar{x}^i)} \partial^\nu \bar{x}^i - \mathcal{L} g^{\mu\nu},$$

and so in the coordinate system $(\bar{t}, \bar{x}^1, \bar{x}^2, \bar{x}^3)$ we have

$$T^{\bar{0}\bar{0}} = \mathcal{L},$$

that is, the elastic Lagrangian density is just the rest energy density $\rho = T^{\bar{0}\bar{0}}$ measured by each particle in the medium. The choice of $\rho = \rho(\bar{x}^i, h^{ij})$ is called the **elastic law** of the continuous medium.

We define **homogeneous and isotropic materials** to be those for which ρ depends only on the eigenvalues (s_1^2, s_2^2, s_3^2) of h_{ij} with respect to k_{ij} (that is, the eigenvalues of the matrix (h_{ij}) in a frame where $k_{ij} = \delta_{ij}$). Note that (s_1, s_2, s_3) are the stretch factors along the principal directions given by the eigenvectors of h_{ij} , that is, the distance between two nearby points along a principal direction in the current configuration, as measured by the metric h , divided by the distance between the same points in the relaxed configuration, as measured by the metric k .

Assume that $k_{ij} = \delta_{ij}$, that is, that (S, k) is the Euclidean space. We define the more convenient variables

$$\lambda_0 = \det(h^{ij}) = \frac{1}{(s_1 s_2 s_3)^2};$$

$$\lambda_1 = \text{tr}(h^{ij}) = \frac{1}{s_1^2} + \frac{1}{s_2^2} + \frac{1}{s_3^2};$$

$$\lambda_2 = \text{tr cof}(h^{ij}) = \frac{1}{(s_1 s_2)^2} + \frac{1}{(s_2 s_3)^2} + \frac{1}{(s_3 s_1)^2}.$$

Note that

$$s_1 s_2 s_3 = \left(\frac{1}{\lambda_0}\right)^{\frac{1}{2}}$$

is the volume occupied in the deformed state by a unit volume of material in the relaxed configuration. Equivalently,

$$n = (\lambda_0)^{\frac{1}{2}}$$

is the number density of particles of the medium in the deformed state, if we normalize the number density in the relaxed configuration to be 1 particle per unit volume. Elastic media whose elastic law depends only on n ,

$$\rho = \rho(\lambda_0),$$

are simply perfect fluids. To check this, we note that from the formula for the inverse of a matrix we have

$$h_{ij} = \frac{1}{\lambda_0} A^{ji} = \frac{1}{\lambda_0} A^{ij},$$

where A^{ij} is the (i, j) -cofactor of (h^{ij}) . On the other hand, from the Laplace expansion for determinants we have

$$\lambda_0 = \sum_{j=1}^3 h^{ij} A^{ij} \quad (\text{no sum over } i),$$

and so

$$\frac{\partial \lambda_0}{\partial h^{ij}} = A^{ij} = \lambda_0 h_{ij}.$$

Therefore

$$\begin{aligned}
T^{\mu\nu} &= \frac{d\rho}{d\lambda_0} \frac{\partial\lambda_0}{\partial h^{ij}} \frac{\partial h^{ij}}{\partial(\partial_\mu \bar{x}^k)} \partial^\nu \bar{x}^k - \rho g^{\mu\nu} \\
&= \frac{d\rho}{d\lambda_0} \lambda_0 h_{ij} (g^{\mu\alpha} \delta^i_k \partial_\alpha \bar{x}^j + g^{\mu\alpha} \partial_\alpha \bar{x}^i \delta^j_k) \partial^\nu \bar{x}^k - \rho g^{\mu\nu} \\
&= 2\lambda_0 \frac{d\rho}{d\lambda_0} h_{ij} \partial^\mu \bar{x}^i \partial^\nu \bar{x}^j - \rho g^{\mu\nu}.
\end{aligned}$$

Since

$$h_{\mu\nu} = h_{ij} \partial_\mu \bar{x}^i \partial_\nu \bar{x}^j$$

is simply the metric on the hyperplanes orthogonal to the worldlines, that is,

$$h_{\mu\nu} = g_{\mu\nu} + U_\mu U_\nu,$$

where U is the unit tangent vector to the worldlines, we obtain

$$\begin{aligned}
T^{\mu\nu} &= 2\lambda_0 \frac{d\rho}{d\lambda_0} U^\mu U^\nu + \left(2\lambda_0 \frac{d\rho}{d\lambda_0} - \rho \right) g^{\mu\nu} \\
&= (\rho + p) U^\mu U^\nu + p g^{\mu\nu}
\end{aligned}$$

with

$$p = 2\lambda_0 \frac{d\rho}{d\lambda_0} - \rho,$$

which is indeed the energy-momentum tensor of a perfect fluid. For example, dust corresponds to the elastic law $\rho = \rho_0 \sqrt{\lambda_0}$ (for some positive constant ρ_0), yielding $p = 0$, and a stiff fluid, with equation of state $p = \rho$, is given by the choice $\rho = \rho_0 \lambda_0$. The ‘‘hard phase’’ rigid fluid introduced by Christodoulou, with equation of state $p = \rho - \rho_0$, corresponds to $\rho = \frac{\rho_0}{2}(\lambda_0 + 1)$.

To obtain elastic materials that are not fluids we must choose elastic laws that also depend on λ_1 and λ_2 . For instance, an elastic law is said to be **quasi-Hookean** if it is of the form

$$\rho = \hat{p}(n) + \hat{\mu}(n)\sigma,$$

where σ is a **shear scalar**, that is, a non-negative function of the stretch factors such that $\sigma = 0$ if and only if $s_1 = s_2 = s_3$. The functions \hat{p} and $\hat{\mu}$ are called the **unsheared energy density** and the **rigidity modulus** of the elastic material. Examples of these are the **John quasi-Hookean material**, corresponding to the shear scalar

$$\sigma = \frac{s_1^2 + s_2^2 + s_3^2}{(s_1^2 s_2^2 s_3^2)^{\frac{1}{3}}} - 3,$$

and the the **Karlovini-Samuelsson quasi-Hookean material**, corresponding to the shear scalar

$$\sigma = \frac{1}{12} \left[\left(\frac{s_1}{s_2} - \frac{s_2}{s_1} \right)^2 + \left(\frac{s_1}{s_3} - \frac{s_3}{s_1} \right)^2 + \left(\frac{s_2}{s_3} - \frac{s_3}{s_2} \right)^2 \right].$$

It is easily seen that the first elastic law is of the form

$$\rho = f(\lambda_0) + g(\lambda_0)\lambda_2,$$

whereas the second is of the form

$$\rho = f(\lambda_0) + g(\lambda_0)\lambda_1\lambda_2.$$

Other examples are the **stiff ultra-rigid material** of Karlovini and Samuelson, given by

$$\rho = \frac{\rho_0}{4}(\lambda_2 + 1),$$

and the **Brotas rigid solid**, given by

$$\rho = \frac{\rho_0}{8}(\lambda_0 + \lambda_1 + \lambda_2 + 1)$$

(where ρ_0 is a positive constant).

3. Einstein-Hilbert action

The variational formulation of field theories is coordinate-free, and so it gives a simple method to write the field equations on an arbitrary coordinate system. We must be careful, however, to note that the action written in the new coordinate system must include the Jacobian of the coordinate transformation:

$$S = \int_{\mathbb{R}^4} \mathcal{L}(\psi, \partial\psi) \sqrt{|\det(g_{\mu\nu})|} dx^0 dx^1 dx^2 dx^3.$$

This suggests that to generalize the field equations to an arbitrary curved spacetime (M, g) one should consider actions of this form, where \mathcal{L} should be invariant under coordinate changes. In particular, one may wonder if there is a Lagrangian action for the metric itself which yields the Einstein field equations. The answer to this question is affirmative, and the corresponding action is known as the **Einstein-Hilbert action**:

$$S = \int_M R \sqrt{|\det(g_{\mu\nu})|} dx^0 dx^1 dx^2 dx^3 = \int_M R \epsilon,$$

where R and ϵ are the scalar curvature and the volume element of the metric g , and the integral is over an arbitrary (oriented) manifold M . Note that R depends on g and its first and second partial derivatives, unlike the Lagrangian densities that we encountered before. Instead of deriving the Euler-Lagrange equations for this case, we will proceed in a more geometric way. We note here, however, that this action is exceptional in that it leads to second-order equations for the metric, as opposed to the fourth order equations that are typical of Lagrangian densities depending on second partial derivatives.

To obtain the Euler-Lagrange equations for the Einstein-Hilbert action we start by considering two affine connections ∇ and $\tilde{\nabla}$ on M . Because

$$(\tilde{\nabla}_X - \nabla_X)(fY) = f(\tilde{\nabla}_X - \nabla_X)Y,$$

there exists a tensor C such that

$$(\tilde{\nabla}_X - \nabla_X)Y = C(X, Y) \Leftrightarrow \tilde{\nabla}_\mu Y^\nu = \nabla_\mu Y^\nu + C_{\mu\alpha}^\nu Y^\alpha.$$

If both ∇ and $\tilde{\nabla}$ are symmetric then

$$0 = (\tilde{\nabla}_X Y - \tilde{\nabla}_Y X) - (\nabla_X Y - \nabla_Y X) = C(X, Y) - C(Y, X),$$

that is, C is symmetric:

$$C_{\mu\nu}^\alpha = C_{\nu\mu}^\alpha.$$

Using the Leibnitz rule, it is easy to determine the relation between the covariant derivatives of any tensor using the two connections: for example,

$$\tilde{\nabla}_\alpha T_{\mu\nu}^\beta = \nabla_\alpha T_{\mu\nu}^\beta + C_{\alpha\gamma}^\beta T_{\mu\nu}^\gamma - C_{\alpha\mu}^\gamma T_{\gamma\nu}^\beta - C_{\alpha\nu}^\gamma T_{\mu\gamma}^\beta.$$

Assume now that $\tilde{\nabla}$ is the Levi-Civita connection for the metric g . Then we have

$$0 = \tilde{\nabla}_\alpha g_{\mu\nu} = \nabla_\alpha g_{\mu\nu} - C_{\alpha\mu}^\beta g_{\beta\nu} - C_{\alpha\nu}^\beta g_{\mu\beta} = \nabla_\alpha g_{\mu\nu} - C_{\nu\alpha\mu} - C_{\mu\alpha\nu}.$$

By subtracting this identity from its cyclic permutations,

$$\begin{aligned} \nabla_\mu g_{\nu\alpha} &= C_{\alpha\mu\nu} + C_{\nu\mu\alpha}, \\ \nabla_\nu g_{\alpha\mu} &= C_{\mu\nu\alpha} + C_{\alpha\nu\mu}, \end{aligned}$$

we readily obtain

$$\begin{aligned} 2C_{\alpha\mu\nu} &= \nabla_\mu g_{\nu\alpha} + \nabla_\nu g_{\alpha\mu} - \nabla_\alpha g_{\mu\nu} \Leftrightarrow \\ (14) \quad C_{\mu\nu}^\alpha &= \frac{1}{2} g^{\alpha\beta} (\nabla_\mu g_{\nu\beta} + \nabla_\nu g_{\mu\beta} - \nabla_\beta g_{\mu\nu}). \end{aligned}$$

Moreover, we have

$$\begin{aligned} \tilde{\nabla}_\mu \tilde{\nabla}_\nu X^\alpha &= \tilde{\nabla}_\mu (\nabla_\nu X^\alpha + C_{\nu\beta}^\alpha X^\beta) = \nabla_\mu \nabla_\nu X^\alpha + C_{\mu\beta}^\alpha \nabla_\nu X^\beta - C_{\mu\nu}^\beta \nabla_\beta X^\alpha \\ &\quad + \nabla_\mu C_{\nu\beta}^\alpha X^\beta + C_{\nu\beta}^\alpha \nabla_\mu X^\beta + C_{\mu\gamma}^\alpha C_{\nu\beta}^\gamma X^\beta - C_{\mu\nu}^\gamma C_{\gamma\beta}^\alpha X^\beta, \end{aligned}$$

whence

$$\begin{aligned} \tilde{R}_{\mu\nu}{}^\alpha{}_\beta X^\beta &= (\tilde{\nabla}_\mu \tilde{\nabla}_\nu - \tilde{\nabla}_\nu \tilde{\nabla}_\mu) X^\alpha = R_{\mu\nu}{}^\alpha{}_\beta X^\beta \\ &\quad + \left(\nabla_\mu C_{\nu\beta}^\alpha - \nabla_\nu C_{\mu\beta}^\alpha + C_{\mu\gamma}^\alpha C_{\nu\beta}^\gamma - C_{\nu\gamma}^\alpha C_{\mu\beta}^\gamma \right) X^\beta, \end{aligned}$$

that is,

$$(15) \quad \tilde{R}_{\mu\nu}{}^\alpha{}_\beta = R_{\mu\nu}{}^\alpha{}_\beta + \nabla_\mu C_{\nu\beta}^\alpha - \nabla_\nu C_{\mu\beta}^\alpha + C_{\mu\gamma}^\alpha C_{\nu\beta}^\gamma - C_{\nu\gamma}^\alpha C_{\mu\beta}^\gamma.$$

Note that we retrieve the usual formulae for the Christoffel symbols and the Riemann tensor from equations (14) and (15) in the case when $\nabla_\mu = \partial_\mu$.

Let $g(\lambda)$ be a one-parameter family of Lorentzian metrics on M , and choose ∇ and $\tilde{\nabla}$ to be the Levi-Civita connections of $g(0)$ and $g(\lambda)$. The

difference between these connections is a tensor $C(\lambda)$ with $C(0) = 0$. Again setting $\delta \equiv \frac{d}{d\lambda}|_{\lambda=0}$, we have from (14)

$$\begin{aligned}\delta C_{\mu\nu}^{\alpha} &= \frac{1}{2}g^{\alpha\beta} (\nabla_{\mu}\delta g_{\nu\beta} + \nabla_{\nu}\delta g_{\mu\beta} - \nabla_{\beta}\delta g_{\mu\nu}) \\ &= \frac{1}{2} (\nabla_{\mu}\delta g_{\nu}^{\alpha} + \nabla_{\nu}\delta g_{\mu}^{\alpha} - \nabla^{\alpha}\delta g_{\mu\nu}),\end{aligned}$$

where $g_{\mu\nu}$ means $g_{\mu\nu}(0)$ and all indices are raised or lowered with this metric. Note carefully that δg means the tensor $\delta g_{\mu\nu}$, possible with some indices raised. Thus for instance

$$\delta(g^{\mu\nu}) = -g^{\mu\alpha}g^{\nu\beta}\delta g_{\alpha\beta} = -\delta g^{\mu\nu}.$$

From (15) we have

$$\delta R_{\mu\nu}^{\alpha}{}_{\beta} = \nabla_{\mu}\delta C_{\nu\beta}^{\alpha} - \nabla_{\nu}\delta C_{\mu\beta}^{\alpha},$$

whence

$$\begin{aligned}\delta R_{\mu\nu} &= \nabla_{\alpha}\delta C_{\mu\nu}^{\alpha} - \nabla_{\mu}\delta C_{\alpha\nu}^{\alpha} \\ &= \frac{1}{2}\nabla_{\alpha} (\nabla_{\mu}\delta g_{\nu}^{\alpha} + \nabla_{\nu}\delta g_{\mu}^{\alpha} - \nabla^{\alpha}\delta g_{\mu\nu}) - \frac{1}{2}\nabla_{\mu}\nabla_{\nu}\delta g_{\alpha}^{\alpha}.\end{aligned}$$

Note that

$$g^{\mu\nu}\delta R_{\mu\nu} = -\nabla_{\mu}\nabla^{\mu}\delta g_{\nu}^{\nu} + \nabla^{\mu}\nabla^{\nu}\delta g_{\mu\nu} = \nabla^{\mu}(-\nabla_{\mu}\delta g_{\nu}^{\nu} + \nabla^{\nu}\delta g_{\mu\nu})$$

is a divergence with respect to the metric $g(0)$, and will vanish when integrated for variations of the metric with compact support.

The variation of the Einstein-Hilbert action is

$$\delta S = \delta \int_M R\epsilon = \int_M \delta(R\epsilon) = \int_M (\delta R\epsilon + R\delta\epsilon).$$

We have

$$\delta R = \delta(g^{\mu\nu}R_{\mu\nu}) = -\delta g^{\mu\nu}R_{\mu\nu} + \nabla^{\mu}(-\nabla_{\mu}\delta g_{\nu}^{\nu} + \nabla^{\nu}\delta g_{\mu\nu})$$

and, using the identity

$$\delta \det A = (\det A) \operatorname{tr}(A^{-1}\delta A),$$

for any matrix-valued function A ,

$$\begin{aligned}\delta\epsilon &= \delta\sqrt{-\det(g_{\mu\nu})} dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \\ &= -\frac{1}{2}(\det(g_{\mu\nu})) g^{\mu\nu}\delta g_{\mu\nu} (-\det(g_{\mu\nu}))^{-\frac{1}{2}} dx^0 \wedge dx^1 \wedge dx^2 \wedge dx^3 \\ &= \frac{1}{2}g_{\mu\nu}\delta g^{\mu\nu}\epsilon.\end{aligned}$$

We conclude that

$$\begin{aligned}\delta S &= - \int_M \left(R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right) \delta g^{\mu\nu} \epsilon + \int_M \nabla^\mu (-\nabla_\mu \delta g_\nu{}^\nu + \nabla^\nu \delta g_{\mu\nu}) \epsilon \\ &= - \int_M G_{\mu\nu} \delta g^{\mu\nu} \epsilon.\end{aligned}$$

for variations of the metric with compact support. We conclude that the Euler-Lagrange equations for the Einstein-Hilbert action are the Einstein equations $G_{\mu\nu} = 0$.

REMARK 3.1. It is easy to see that the Einstein tensor of a compact surface (M, g) vanishes identically. Therefore we have

$$\delta \int_M R \epsilon = 0$$

automatically. If g_1 and g_2 are any two Riemannian metrics on M then $g(\lambda) = (1 - \lambda)g_1 + \lambda g_2$ is a Riemannian metric which interpolates between g_1 and g_2 . We then have

$$\frac{d}{d\lambda} \int_M R(\lambda) \epsilon(\lambda) = 0 \Rightarrow \int_M R_1 \epsilon_1 = \int_M R_2 \epsilon_2,$$

that is, the integral of the scalar curvature does not depend on the metric. This statement is known as the **Gauss-Bonnet theorem**.

To include matter fields ψ and a cosmological constant Λ in the Einstein equations we consider the action

$$S = \int_M \left(\mathcal{L}(g, \psi) - \frac{1}{16\pi} (R - 2\Lambda) \right) \epsilon.$$

It is clear that

$$\delta S = \int_M E(\psi) \delta \psi \epsilon + \frac{1}{16\pi} \int_M (G_{\mu\nu} + \Lambda g_{\mu\nu} - 8\pi T_{\mu\nu}) \delta g^{\mu\nu} \epsilon,$$

where we have defined

$$\delta \int_M \mathcal{L}(g, \psi) \epsilon = \int_M E(\psi) \delta \psi \epsilon - \frac{1}{2} \int_M T_{\mu\nu} \delta g^{\mu\nu} \epsilon.$$

The Euler-Lagrange equations are then the Einstein equations with sources plus the field equations for ψ :

$$\begin{cases} G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu} \\ E(\psi) = 0 \end{cases}.$$

The energy-momentum tensor is then

$$T_{\mu\nu} = -2 \frac{\delta \mathcal{L}}{\delta g^{\mu\nu}} - \mathcal{L} g_{\mu\nu},$$

where the second term comes from the variation of the volume element and we have set

$$\delta \mathcal{L} = \frac{\delta \mathcal{L}}{\delta g^{\mu\nu}} \delta g^{\mu\nu} + E(\psi) \delta \psi.$$

It is interesting to note that the energy-momentum tensor as defined here often agrees with what one would expect from the canonical energy-momentum tensor associated to the Lagrangian density \mathcal{L} in Minkowski's spacetime.

4. Gravitational waves

Since we have computed the variation of the Ricci tensor, we make a short digression to discuss the linearized Einstein vacuum equations, which describe the propagation of gravitational waves on a fixed solution (M, g) of the full nonlinear vacuum equations $R_{\mu\nu} = \Lambda g_{\mu\nu}$. The linearized equations are simply

$$\delta R_{\mu\nu} = \Lambda \delta g_{\mu\nu},$$

that is,

$$(16) \quad \nabla_\alpha (\nabla_\mu \delta g_\nu^\alpha + \nabla_\nu \delta g_\mu^\alpha - \nabla^\alpha \delta g_{\mu\nu}) - \nabla_\mu \nabla_\nu \delta g_\alpha^\alpha = 2\Lambda \delta g_{\mu\nu}.$$

Note that some variations of the metric are trivial, in that they arise from the diffeomorphism invariance of the Einstein equations: if ψ_λ is a one-parameter family of diffeomorphisms, then the variation

$$g(\lambda) = \psi_\lambda^* g$$

yields metrics which are isometric to $g = g(0)$, but expressed in different coordinates. In this case we have

$$\delta g = \mathcal{L}_V g,$$

that is

$$\delta g_{\mu\nu} = \nabla_\mu V_\nu + \nabla_\nu V_\mu,$$

where V is the vector field defined at each point $p \in M$ by

$$V_p = \frac{d}{d\lambda|_{\lambda=0}} \psi_\lambda(p).$$

Therefore, the linearized Einstein equations have gauge freedom: we can always add a Lie derivative of g to any given variation of the metric without altering its physical meaning. This corresponds to gauge transformations of the form

$$\delta g_{\mu\nu} \rightarrow \delta g_{\mu\nu} + \nabla_\mu V_\nu + \nabla_\nu V_\mu.$$

We will now construct a gauge where equations (16) look particularly simple. To do that, we consider the trace-reversed metric perturbation

$$\overline{\delta g}_{\mu\nu} = \delta g_{\mu\nu} - \frac{1}{2}(\delta g_\alpha^\alpha)g_{\mu\nu},$$

which transforms under a gauge transformation as

$$\overline{\delta g}_{\mu\nu} \rightarrow \overline{\delta g}_{\mu\nu} + \nabla_\mu V_\nu + \nabla_\nu V_\mu - (\nabla_\alpha V^\alpha)g_{\mu\nu},$$

so that its divergence transforms as

$$\begin{aligned}\nabla^\mu \overline{\delta g}_{\mu\nu} &\rightarrow \nabla^\mu \overline{\delta g}_{\mu\nu} + \square V_\nu + \nabla_\mu \nabla_\nu V^\mu - \nabla_\nu \nabla_\alpha V^\alpha \\ &= \nabla^\mu \overline{\delta g}_{\mu\nu} + \square V_\nu + R_{\mu\nu}{}^\mu{}_\alpha V^\alpha \\ &= \nabla^\mu \overline{\delta g}_{\mu\nu} + \square V_\nu + R_{\nu\alpha} V^\alpha.\end{aligned}$$

Assume that (M, g) is globally hyperbolic. By solving the wave equation

$$\square V_\nu + R_{\nu\alpha} V^\alpha + \nabla^\mu \overline{\delta g}_{\mu\nu} = 0,$$

we can then change to a gauge where

$$\nabla^\mu \overline{\delta g}_{\mu\nu} = 0,$$

that is,

$$\nabla^\mu \delta g_{\mu\nu} = \frac{1}{2} \nabla_\nu \delta g_\alpha{}^\alpha.$$

Taking the trace of equation (16) we then obtain

$$(17) \quad \square \delta g_\alpha{}^\alpha + 2\Lambda \delta g_\alpha{}^\alpha = 0.$$

Note that we still have a residual gauge freedom, corresponding to vector fields V such that

$$\square V_\nu + R_{\nu\alpha} V^\alpha = 0.$$

Choosing V and its derivatives on some Cauchy hypersurface S such that

$$\delta g_\alpha{}^\alpha + 2\nabla_\alpha V^\alpha = N \cdot (\delta g_\alpha{}^\alpha + 2\nabla_\alpha V^\alpha) = 0,$$

where N is the future-pointing unit normal to S , and solving the wave equation for V above, we guarantee the existence of a gauge where

$$\delta g_\alpha{}^\alpha = N \cdot (\delta g_\alpha{}^\alpha) = 0$$

on S . The wave equation (17) then guarantees that

$$\delta g_\alpha{}^\alpha = 0$$

on M , and so

$$\nabla^\mu \delta g_{\mu\nu} = 0.$$

This is the so-called **transverse traceless gauge**. In this gauge, the linearized Einstein equation (16) can be written as

$$\nabla_\alpha \nabla_\mu \delta g_\nu{}^\alpha + \nabla_\alpha \nabla_\nu \delta g_\mu{}^\alpha - \square \delta g_{\mu\nu} = 2\Lambda \delta g_{\mu\nu}.$$

Using

$$\begin{aligned}\nabla_\alpha \nabla_\mu \delta g_\nu{}^\alpha &= \nabla_\mu \nabla_\alpha \delta g_\nu{}^\alpha + R_{\alpha\mu\nu\beta} \delta g^{\beta\alpha} + R_{\alpha\mu}{}^\alpha{}_\beta \delta g_\nu{}^\beta \\ &= R_{\alpha\mu\nu\beta} \delta g^{\alpha\beta} + R_{\mu\beta} \delta g_\nu{}^\beta \\ &= R_{\alpha\mu\nu\beta} \delta g^{\alpha\beta} + \Lambda \delta g_{\mu\nu},\end{aligned}$$

we finally obtain

$$\square \delta g_{\mu\nu} - 2R_{\alpha\mu\nu\beta} \delta g^{\alpha\beta} = 0.$$

5. ADM mass

If we write the metric in the Gauss Lemma form,

$$g = -dt^2 + h_{ij}(t, x)dx^i dx^j$$

then from the exercises in Chapter 5 we have

$$R = \bar{R} + 2\frac{\partial}{\partial t}(K^i{}_i) + (K^i{}_i)^2 + K_{ij}K^{ij}.$$

Setting

$$X = K^i{}_i \frac{\partial}{\partial t},$$

we have

$$\operatorname{div} X = \nabla_0 X^0 + \nabla_i X^i = \partial_0 X^0 + \Gamma_{i0}^i X^0 = \partial_0 X^0 + K^i{}_i X^0,$$

that is,

$$\frac{\partial}{\partial t}(K^i{}_i) = \operatorname{div} X - (K^i{}_i)^2.$$

We conclude that

$$R = \bar{R} - (K^i{}_i)^2 + K_{ij}K^{ij} + 2\operatorname{div} X,$$

and so the Einstein-Hilbert action corresponds to the Lagrangian density

$$\mathcal{L} = \sqrt{\det(h_{ij})} \left(\bar{R} - (K^i{}_i)^2 + K_{ij}K^{ij} \right).$$

Therefore the Hamiltonian density is

$$\begin{aligned} \mathcal{H} &= -\frac{\partial \mathcal{L}}{\partial(\partial_0 h_{ij})} \partial_0 h_{ij} + \mathcal{L} = -\frac{\partial \mathcal{L}}{\partial(K_{ij})} K_{ij} + \mathcal{L} \\ &= -\sqrt{\det(h_{ij})} \left(2K^{ij} - 2K^l{}_l h^{ij} \right) K_{ij} + \mathcal{L} \\ &= \sqrt{\det(h_{ij})} \left(\bar{R} + (K^i{}_i)^2 - K_{ij}K^{ij} \right) = 2\sqrt{\det(h_{ij})} G_{00} = 0 \end{aligned}$$

for any solution of the vacuum Einstein field equations. That is, the total energy associated to the fields h_{ij} is simply zero.

To try to obtain a nonzero quantity we reconsider the boundary terms that were discarded when varying the Einstein-Hilbert action:

$$\int_M \nabla^\mu (-\nabla_\mu \delta g_\nu{}^\nu + \nabla^\nu \delta g_{\mu\nu}) \epsilon = \int_{\partial M} (-\nabla_\mu \delta g_\nu{}^\nu + \nabla^\nu \delta g_{\mu\nu}) n^\mu \omega,$$

where we take M to be a manifold with a timelike boundary ∂M at infinity, tangent to $\frac{\partial}{\partial t}$, with unit normal n and volume element ω (Figure 4). Note that it is not necessary to consider the flux of the vector field X which we discarded above as it is orthogonal to n . The boundary integral can be written as

$$\int_{\mathbb{R}} \int_{\Sigma} \left(-\partial_i (h^{jk} \delta h_{jk}) + h^{jk} \bar{\nabla}_k \delta h_{ij} \right) n^i \sigma dt,$$

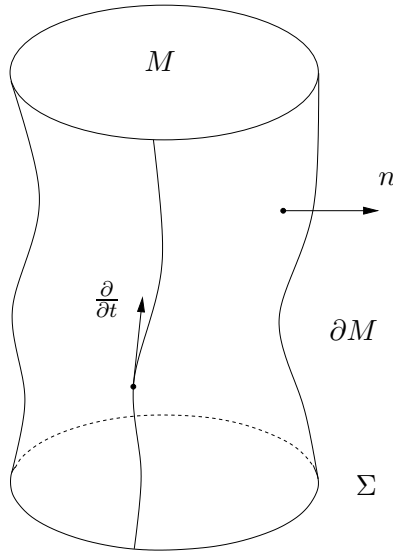


FIGURE 4. Computing the boundary terms for the Einstein-Hilbert action.

where we take ∂M to be the flow by $\frac{\partial}{\partial t}$ of a spacelike surface Σ and $\omega = dt \wedge \sigma$. Suppose that h approaches the Euclidean metric at infinity, so that in an appropriate coordinate system (x^1, x^2, x^3) we have

$$h_{ij} = \delta_{ij} + O(r^{-p}), \quad \delta h_{ij} = O(r^{-p}), \quad \partial_k h_{ij}, \partial_k \delta h_{ij}, \Gamma_{ij}^k = O(r^{-p-1})$$

as $r \rightarrow +\infty$, with $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$ and $p > \frac{1}{2}$. Then the boundary integral can be written as

$$\begin{aligned} & \lim_{r \rightarrow \infty} \int_{\mathbb{R}} \int_{S_r} (-\partial_i \delta h_{jj} + \partial_j \delta h_{ij}) \frac{x^i}{r} dt \\ &= \delta \lim_{r \rightarrow \infty} \int_{\mathbb{R}} \int_{S_r} (-\partial_i h_{jj} + \partial_j h_{ij}) \frac{x^i}{r} dt = \delta I, \end{aligned}$$

where S_r is a coordinate sphere of radius r . Note that δ can be brought outside the integral because we have replaced the metric-dependent terms $\bar{\nabla}$, n and σ .

If S is the Einstein-Hilbert action, we then have

$$\delta S = - \int_M G_{\mu\nu} \delta g^{\mu\nu} + \delta I,$$

and so the Einstein equations hold if and only if

$$\delta(S - I) = 0.$$

One can think of I as the integral of a singular Lagrangian density \mathcal{I} . The Einstein-Hilbert Lagrangian density \mathcal{L} should therefore be replaced by $\mathcal{L} - \mathcal{I}$, and so, since \mathcal{I} does not depend on K_{ij} , the corresponding Hamiltonian

density \mathcal{H} should be replaced by $\mathcal{H} - \mathcal{I} = -\mathcal{I}$. The Hamiltonian should then be

$$H = - \lim_{r \rightarrow \infty} \int_{S_r} (\partial_j h_{ij} - \partial_i h_{jj}) \frac{x^i}{r}.$$

This Hamiltonian suggests a definition of the total energy of the gravitational field at a given time slice, provided that h approaches the Euclidean metric at infinity.

DEFINITION 5.1. *A 3-dimensional Riemannian manifold (S, h) is said to be **asymptotically flat** if there exist:*

- (i) *A compact set $K \subset S$ such that $S \setminus K$ is diffeomorphic to $\mathbb{R}^3 \setminus \overline{B_1}(0)$;*
- (ii) *A chart (x^1, x^2, x^3) on $S \setminus K$ (called a **chart at infinity**) such that*

$$|h_{ij} - \delta_{ij}| + r |\partial_k h_{ij}| + r^2 |\partial_k \partial_l h_{ij}| = O(r^{-p}) \text{ and } \bar{R} = O(r^{-q})$$

for some $p > \frac{1}{2}$ and $q > 3$, where $r^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$ and \bar{R} is the scalar curvature of h .

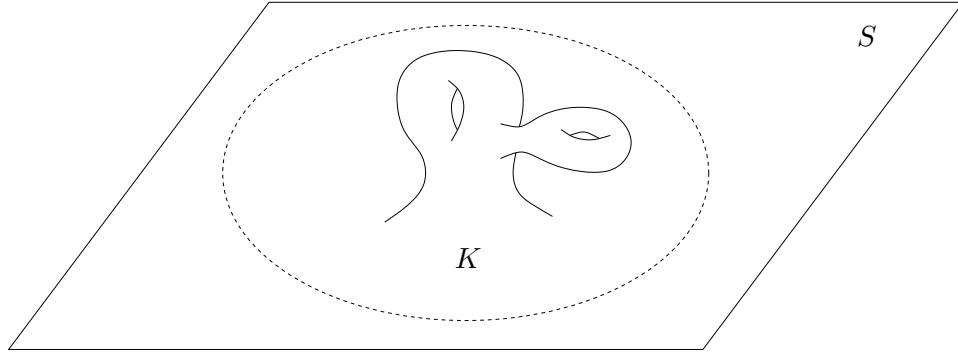


FIGURE 5. Asymptotically flat manifold.

DEFINITION 5.2. (**Arnowitt-Deser-Misner [ADM61]**) *The **ADM mass** of an asymptotically flat Riemannian manifold (S, h) is*

$$M = \lim_{r \rightarrow +\infty} \frac{1}{16\pi} \int_{S_r} (\partial_j h_{ij} - \partial_i h_{jj}) \frac{x^i}{r},$$

where S_r is a sphere of radius r in the chart at infinity (x^1, x^2, x^3) .

Note that in this definition the Hamiltonian has been multiplied by the factor $-\frac{1}{16\pi}$ that is used when coupling to matter fields.

THEOREM 5.3. (**Ashtekar [AMA79]**) *If (S, h) is asymptotically flat and the maximal Cauchy development of (S, h, K) is stationary then the Komar mass of the maximal Cauchy development coincides with the ADM mass of (S, h) .*

THEOREM 5.4. (**Bartnik [Bar86]**) *The ADM mass is well defined, that is, it does not depend on the choice of the chart at infinity.*

6. Positive mass theorem

Since the gravitational field is attractive, its energy is presumably negative, at least for bound states. On the other hand, we expect gravitational waves to carry positive energy. It is therefore an important question to decide whether there is a lower bound for the ADM mass (the inexistence of which would signal an instability). In the simplest case of time-symmetric initial data ($K = 0$) the restriction equations reduce to

$$\bar{R} = 16\pi\rho,$$

and we expect $\rho \geq 0 \Leftrightarrow \bar{R} \geq 0$ for reasonable matter fields. In this case, we have the following famous result.

THEOREM 6.1. (Schoen and Yau [SY81]) *Let (S, h) be a complete asymptotically flat Riemannian 3-manifold with scalar curvature $\bar{R} \geq 0$. Then:*

- (i) *Its ADM mass is nonnegative, $M \geq 0$.*
- (ii) *If $M = 0$ then $S = \mathbb{R}^3$ and h is the Euclidean metric.*

PROOF. Following [Lam10], we give a proof of (i) only for asymptotically flat Riemannian 3-manifolds S that are graphs of smooth functions $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ with the metric h induced by the Euclidean metric of \mathbb{R}^4 . Using the Cartesian coordinates (x^1, x^2, x^3) of \mathbb{R}^3 as global coordinates on the graph we have

$$h_{ij} = \delta_{ij} + \partial_i f \partial_j f.$$

From that expression one can show that the scalar curvature of the graph is the divergence of a vector field on \mathbb{R}^3 :

$$\bar{R} = \partial_i \left(\frac{1}{1 + |\text{grad } f|^2} (\partial_i f \partial_j \partial_j f - \partial_i \partial_j f \partial_j f) \right).$$

On the other hand,

$$\begin{aligned} M &= \lim_{r \rightarrow +\infty} \frac{1}{16\pi} \int_{S_r} (\partial_j h_{ij} - \partial_i h_{jj}) \frac{x^i}{r} \\ &= \lim_{r \rightarrow +\infty} \frac{1}{16\pi} \int_{S_r} (\partial_i \partial_j f \partial_j f + \partial_i f \partial_j \partial_j f - 2\partial_i \partial_j f \partial_j f) \frac{x^i}{r} \\ &= \lim_{r \rightarrow +\infty} \frac{1}{16\pi} \int_{S_r} (\partial_i f \partial_j \partial_j f - \partial_i \partial_j f \partial_j f) \frac{x^i}{r}. \end{aligned}$$

Since (S, h) is asymptotically flat, the derivatives of f approach zero with certain decays as $r \rightarrow +\infty$. One can then easily show that

$$\begin{aligned} M &= \lim_{r \rightarrow +\infty} \frac{1}{16\pi} \int_{S_r} \left(\frac{1}{1 + |\text{grad } f|^2} (\partial_i f \partial_j \partial_j f - \partial_i \partial_j f \partial_j f) \right) \frac{x^i}{r} \\ &= \frac{1}{16\pi} \int_{\mathbb{R}^3} \partial_i \left(\frac{1}{1 + |\text{grad } f|^2} (\partial_i f \partial_j \partial_j f - \partial_i \partial_j f \partial_j f) \right) \\ &= \frac{1}{16\pi} \int_{\mathbb{R}^3} \bar{R} \geq 0. \end{aligned}$$

We do not prove the rigidity statement (ii). It is interesting to note that this statement, together with the formula above for the ADM mass, implies that any graph with zero scalar curvature is flat. \square

The volume element of the graph is

$$\epsilon = \sqrt{1 + |\text{grad } f|^2} dx^1 \wedge dx^2 \wedge dx^3,$$

since the eigenvalues of the matrix (h_{ij}) are $1 + |\text{grad } f|^2$ for the eigenvector $\text{grad } f$ and 1 for the eigenvectors orthogonal to $\text{grad } f$. Consequently the ADM mass is

$$M = \frac{1}{16\pi} \int_S \frac{\bar{R}}{\sqrt{1 + |\text{grad } f|^2}} \epsilon = \int_S \frac{\rho}{\sqrt{1 + |\text{grad } f|^2}} \epsilon < \int_S \rho \epsilon.$$

The difference

$$M - \int_S \rho \epsilon < 0$$

can be thought of as the (negative) gravitational binding energy.

7. Penrose inequality

The positive mass theorem admits a refinement in the case when black holes are present, known as the Penrose inequality. The idea is that black hole horizons correspond to minimal surfaces Σ on the Riemannian manifold (S, h) , each contributing with a mass M at least as big as the mass of a Schwarzschild black hole with the same event horizon area A :

$$M \geq \sqrt{\frac{A}{16\pi}}.$$

To understand the motivation for this inequality, recall that in the proof of the Penrose singularity theorem we defined the outward null expansion of a 2-surface Σ on a Cauchy hypersurface S as

$$\theta = \frac{1}{2} \gamma^{AB} \frac{\partial \gamma_{AB}}{\partial r} = \frac{1}{2} \text{tr} \left(\mathcal{L}_{\frac{\partial}{\partial r}} g \right)_{|_{T\Sigma}}$$

If N is the future-pointing unit normal to S and n is the outward unit normal to Σ on S we then have

$$\theta = \frac{1}{2} \text{tr} (\mathcal{L}_{N+n} g)_{|_{T\Sigma}} = \text{tr} K_{|_{T\Sigma}} + \frac{1}{2} \text{tr} (\mathcal{L}_n g)_{|_{T\Sigma}},$$

where we used the fact that $\frac{\partial}{\partial r} = N + n$ on Σ , and also that if X, Y are tangent to Σ then

$$(\mathcal{L}_Z g)(X, Y) = \langle X, \nabla_Y Z \rangle + \langle Y, \nabla_X Z \rangle$$

depends only on Z along Σ . For time-symmetric initial data ($K = 0$) this becomes

$$\theta = \frac{1}{2} \operatorname{tr}(\mathcal{L}_n g)|_{T\Sigma} = \frac{1}{2} \operatorname{tr}(\mathcal{L}_n h)|_{T\Sigma} = \operatorname{tr} \kappa,$$

where h is the metric induced by g on S and κ is the second fundamental form of Σ on S . We conclude that Σ is **marginally trapped**, that is, has zero outward null expansion, if and only $\operatorname{tr} \kappa = 0$, which is precisely the condition for Σ to be a minimal surface. Now a marginally trapped surface anticipates the formation of trapped surfaces, which will lead to geodesic incompleteness of the resulting spacetime, presumably due to singularities. If one believes the **weak cosmic censorship conjecture**, these singularities should only occur inside black holes, and so there should be a black hole horizon enveloping any marginally trapped surface.

THEOREM 7.1. (**Huisken and Ilmanen [HI01], Bray [Bra01]**) *Let (S, h) be a complete asymptotically flat Riemannian manifold with scalar curvature $\bar{R} \geq 0$. Then:*

- (i) *Its ADM mass satisfies $M \geq \sqrt{\frac{A}{16\pi}}$, where A is the sum of the areas of the outer minimal surfaces.*
- (ii) *If $M = \sqrt{\frac{A}{16\pi}}$ then the restriction of $(S, h, 0)$ to the exterior of the outer minimal surfaces coincides with the initial data for the Schwarzschild solution of mass M outside the event horizon.*

PROOF. Following [**Lam10**], we give a proof of only for asymptotically flat Riemannian 3-manifolds S that are graphs of smooth functions $f : U \subset \mathbb{R}^3 \rightarrow \mathbb{R}$, with the metric h induced by the Euclidean metric of \mathbb{R}^4 . Here

$$U = \mathbb{R}^3 \setminus \bigcup_{a=1}^N K_a,$$

where K_1, \dots, K_N are disjoint convex compact sets with smooth boundaries $\partial K_1, \dots, \partial K_N$ to which f extends as a constant and where $|\operatorname{grad} f| \rightarrow +\infty$ (so that they are minimal surfaces of the graph). Applying the divergence theorem as in the proof of the positive mass theorem for graphs we now obtain

$$M = \frac{1}{16\pi} \int_U \bar{R} + \frac{1}{16\pi} \sum_{a=1}^N \int_{\partial K_a} \left(\frac{1}{1 + |\operatorname{grad} f|^2} (\partial_i f \partial_j \partial_j f - \partial_i \partial_j f \partial_j f) \right) n^i,$$

where n is the outward unit normal to ∂K_a . Now as is well known the Laplacian Δf of f in U is related to the Laplacian $\bar{\Delta} f$ of f in ∂K_a by

$$\Delta f = \bar{\Delta} f + H f(n, n) + (n \cdot f) \operatorname{tr} \kappa,$$

where Hf is the Hessian of f and κ is the second fundamental form of ∂K_a in \mathbb{R}^3 . Since f is constant on ∂K_a we have $\bar{\Delta}f = 0$ and so

$$\begin{aligned} (\partial_i f \partial_j \partial_j f - \partial_i \partial_j f \partial_j f) n^i &= (n \cdot f) \Delta f - Hf(n, \text{grad } f) \\ &= (n \cdot f) Hf(n, n) + (n \cdot f)^2 \text{tr } \kappa - \langle \text{grad } f, n \rangle Hf(n, n) \\ &= \langle \text{grad } f, n \rangle^2 \text{tr } \kappa = |\text{grad } f|^2 \text{tr } \kappa, \end{aligned}$$

where we used $\text{grad } f = \langle \text{grad } f, n \rangle n$. Therefore

$$\begin{aligned} M &= \frac{1}{16\pi} \int_U \bar{R} + \frac{1}{16\pi} \sum_{a=1}^N \int_{\partial K_a} \left(\frac{|\text{grad } f|^2}{1 + |\text{grad } f|^2} \right) \text{tr } \kappa \\ &= \frac{1}{16\pi} \int_U \bar{R} + \frac{1}{16\pi} \sum_{a=1}^N \int_{\partial K_a} \text{tr } \kappa. \end{aligned}$$

Now **Minkowski's inequality** for the smooth boundaries of compact convex sets states that

$$\int_{\partial K_a} \text{tr } \kappa \geq \sqrt{16\pi A_a},$$

where A_a is the area of ∂K_a . Since $\bar{R} \geq 0$ we conclude that

$$M \geq \sum_{a=1}^N \sqrt{\frac{A_a}{16\pi}} \geq \sqrt{\frac{1}{16\pi} \sum_{a=1}^N A_a},$$

where we used $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b > 0$.

We do not prove the rigidity statement (ii). It is interesting to note that this statement, together with the formula above for the ADM mass, implies that any graph of the type considered above with zero scalar curvature is a Flamm paraboloid. \square

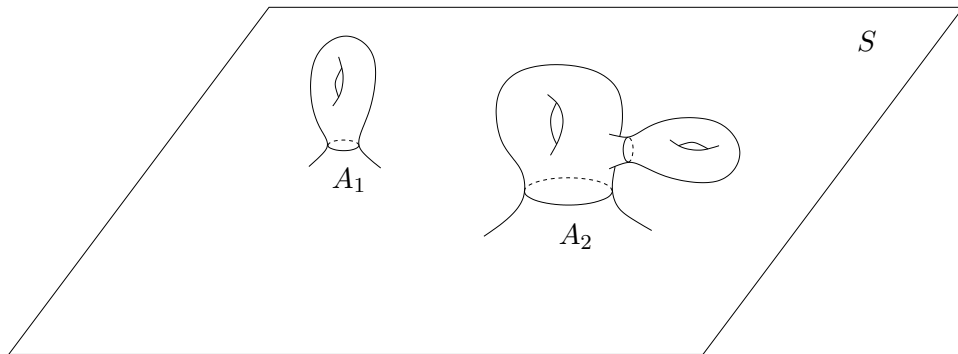


FIGURE 6. Penrose inequality; the sum of the areas of the outer minimal surfaces is $A = A_1 + A_2$.

8. Exercises

- (1) Let g be a static spherically symmetric Lorentzian metric on \mathbb{R}^4 whose matter fields have spatially compact support and satisfy the dominant energy condition. There exist smooth functions $\phi = \phi(r)$ and $m = m(r)$ such that

$$g = -e^{2\phi(r)} dt^2 + \frac{dr^2}{1 - \frac{2m(r)}{r}} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2).$$

Show that:

- (a) The Einstein equations imply

$$\begin{aligned} \frac{dm}{dr} &= 4\pi r^2 \rho; \\ \frac{d\phi}{dr} &= \frac{m + 4\pi r^3 p}{r(r - 2m)}, \end{aligned}$$

where ρ and p are the energy density and the radial pressure as measured by the static observers.

- (b) There exist constants $M \geq 0$ and $\Phi \in \mathbb{R}$ such that

$$\lim_{r \rightarrow +\infty} m(r) = M \quad \text{and} \quad \lim_{r \rightarrow +\infty} \phi(r) = \Phi.$$

- (c) If we choose the coordinate t such that $\Phi = 0$ then M is the Komar mass of g with respect to the timelike Killing vector field $\frac{\partial}{\partial t}$.

- (d) The constant M satisfies $M \leq E$, where

$$E = \int_{\{t=0\}} \rho,$$

with equality exactly when g is the Minkowski metric.

- (2) Starting with the appropriate Lagrangian density, show that the Klein-Gordon equation can be written as

$$\frac{1}{\sqrt{|\det(g_{\mu\nu})|}} \partial_\alpha \left(\sqrt{|\det(g_{\mu\nu})|} \partial^\alpha \phi \right) - m^2 \phi = 0.$$

- (3) Starting with the Einstein-Hilbert-Klein-Gordon action

$$S = \int_M \left[\frac{1}{2} (g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi + m^2 \phi^2) - \frac{1}{16\pi} R \right] \epsilon$$

obtain the energy-momentum tensor for ϕ :

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - \frac{1}{2} g_{\mu\nu} (\partial_\alpha \phi \partial^\alpha \phi + m^2 \phi^2).$$

Note that this agrees with what one would expect from the canonical energy-momentum tensor.

- (4) Show that the Einstein-Hilbert action is equivalent to the (first order, non-geometric) **Einstein action**

$$S = \int_M g^{\alpha\beta} \left(\Gamma_{\alpha\gamma}^{\delta} \Gamma_{\beta\delta}^{\gamma} - \Gamma_{\alpha\beta}^{\gamma} \Gamma_{\gamma\delta}^{\delta} \right) \sqrt{|\det(g_{\mu\nu})|} dx^0 dx^1 dx^2 dx^3,$$

by showing that they differ by a term of the form $\partial_\alpha Q^\alpha$, where

$$Q^\alpha = \left(g^{\beta\gamma} \Gamma_{\beta\gamma}^\alpha - g^{\alpha\beta} \Gamma_{\beta\gamma}^\gamma \right) \sqrt{|\det(g_{\mu\nu})|}.$$

The following formulae will be useful:

$$\partial_\alpha \sqrt{|\det(g_{\mu\nu})|} = -\frac{1}{2} g_{\beta\gamma} \partial_\alpha g^{\beta\gamma} \sqrt{|\det(g_{\mu\nu})|}$$

and

$$\nabla_\alpha g^{\beta\gamma} = 0 \Leftrightarrow \partial_\alpha g^{\beta\gamma} = -\Gamma_{\alpha\delta}^\beta g^{\delta\gamma} - \Gamma_{\alpha\delta}^\gamma g^{\beta\delta}.$$

- (5) Let $f : \mathbb{R}^3 \rightarrow \mathbb{R}$ be a smooth function and consider the metric h induced on its graph S by the Euclidean metric in \mathbb{R}^4 ,

$$h_{ij} = \delta_{ij} + \partial_i f \partial_j f.$$

Show that:

- (a) The inverse metric is

$$h^{ij} = \delta_{ij} - \frac{\partial_i f \partial_j f}{1 + |\text{grad } f|^2}.$$

- (b) The Christoffel symbols are

$$\bar{\Gamma}_{ij}^k = \frac{\partial_k f \partial_i \partial_j f}{1 + |\text{grad } f|^2}.$$

- (c) The Ricci tensor is

$$\begin{aligned} \bar{R}_{ij} &= \frac{\partial_i \partial_j f \partial_k \partial_k f - \partial_i \partial_k f \partial_j \partial_k f}{1 + |\text{grad } f|^2} \\ &\quad + \frac{\partial_k f \partial_l f \partial_i \partial_k f \partial_j \partial_l f - \partial_k f \partial_l f \partial_i \partial_j f \partial_k \partial_l f}{(1 + |\text{grad } f|^2)^2}. \end{aligned}$$

- (d) The scalar curvature is

$$\begin{aligned} \bar{R} &= \frac{\partial_i \partial_i f \partial_j \partial_j f - \partial_i \partial_j f \partial_i \partial_j f}{1 + |\text{grad } f|^2} \\ &\quad - \frac{2 \partial_j f \partial_k f (\partial_i \partial_i f \partial_j \partial_k f - \partial_i \partial_j f \partial_i \partial_k f)}{(1 + |\text{grad } f|^2)^2}. \end{aligned}$$

- (e) The scalar curvature can be written as

$$\bar{R} = \partial_i \left(\frac{1}{1 + |\text{grad } f|^2} (\partial_i f \partial_j \partial_j f - \partial_i \partial_j f \partial_j f) \right).$$

- (6) Let h be the spherically symmetric Riemannian metric defined in \mathbb{R}^3 by

$$h = \frac{dr^2}{1 - \frac{2m(r)}{r}} + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2),$$

where m is a smooth function whose derivative has compact support.

- (a) Check that in Cartesian coordinates we have

$$h_{ij} = \delta_{ij} + \frac{\frac{2m(r)}{r^3}}{1 - \frac{2m(r)}{r}} x^i x^j.$$

- (b) Show that if the limit

$$M = \lim_{r \rightarrow +\infty} m(r)$$

exists then h is asymptotically flat with ADM mass M (which in particular coincides with the Komar mass when appropriate).

- (c) Check that h has scalar curvature

$$\bar{R} = \frac{4}{r^2} \frac{dm}{dr},$$

and use this to prove the Riemannian positive mass theorem for h .

- (d) Show that $r = r_0$ is a minimal surface if and only if $m(r_0) = \frac{r_0}{2}$ (in which case r is a well-defined coordinate only for $r > r_0$), and use this to prove the Riemannian Penrose inequality for h .

- (7) Consider a Riemannian metric given in the Gauss Lemma form

$$g = dt^2 + h_{ij}(t, x) dx^i dx^j,$$

so that the hypersurface $t = 0$ is a Riemannian manifold with induced metric $h = h_{ij} dx^i dx^j$ and second fundamental form

$$K = \frac{1}{2} \frac{\partial h_{ij}}{\partial t} dx^i dx^j.$$

Show that:

- (a) The Laplacian operators Δ and $\bar{\Delta}$ of g and h are related by

$$\Delta f = \bar{\Delta} f + Hf(\partial_0, \partial_0) + (\partial_0 f) \operatorname{tr} K,$$

where Hf is the Hessian of f .

- (b) The metric induced on the hypersurface $t = \lambda f(x)$ is

$$h(\lambda) = [h_{ij}(\lambda f(x), x) + \lambda^2 \partial_i f \partial_j f] dx^i dx^j.$$

- (c) The first variation of this metric is

$$\delta h \equiv \frac{d}{d\lambda} \Big|_{\lambda=0} h(\lambda) = 2f K_{ij} dx^i dx^j.$$

(d) The first variation of the volume element is

$$\delta\sigma \equiv \frac{d}{d\lambda}|_{\lambda=0} \sigma(\lambda) = f \operatorname{tr} K \sigma.$$

(e) The second variation of the volume element is

$$\begin{aligned} \delta^2\sigma &\equiv \frac{d^2}{d\lambda^2}|_{\lambda=0} \sigma(\lambda) \\ &= \left[\frac{1}{2}f^2 \left(\bar{R} - R + (K^i_i)^2 - K_{ij}K^{ij} \right) + |\operatorname{grad} f|^2 \right] \sigma, \end{aligned}$$

where R and \bar{R} are the scalar curvatures of g and h .

(f) There is no metric on the 3-torus with positive scalar curvature. (You will need to use the fact that any metric in the 3-torus admits a minimizing 2-torus; this type of idea is used in the proof of the rigidity statement in the positive mass theorem.)

(8) Consider the surfaces Σ_t obtained from the boundary $\partial K \equiv \Sigma_0$ of a compact convex set $K \subset \mathbb{R}^3$ by flowing a distance t along the unit normal. Let $A(t)$ and $V(t)$ be the area of Σ_t and the volume bounded by Σ_t .

(a) Show that

$$\dot{A}(0) = \int_{\partial K} \operatorname{tr} \kappa,$$

where κ is the second fundamental form of Σ .

(b) Prove that $\ddot{A}(t) = 8\pi$, implying $A(t) = 4\pi t^2 + \dot{A}(0)t + A(0)$.

(c) Conclude that $V(t) = \frac{4\pi}{3}t^3 + \frac{\dot{A}(0)}{2}t^2 + A(0)t + V(0)$.

(d) Use the isoperimetric inequality $A(t)^3 \geq 36\pi V(t)^2$ to prove Minkowski's inequality:

$$\dot{A}(0) \geq \sqrt{16\pi A(0)}.$$

CHAPTER 7

Black holes

In this chapter we study black holes and the laws of black hole thermodynamics, following [Tow97] (see also [BCH73, Poi07]). An elementary discussion of quantum field theory in curved spacetime and the Hawking radiation can be found in [Car03].

1. The Kerr solution

General rotating black holes are described by the **Kerr metric**, given in the so-called **Boyer-Lindquist coordinates** by

$$ds^2 = - \left(1 - \frac{2Mr}{\rho^2} \right) dt^2 - \frac{4Mar \sin^2 \theta}{\rho^2} dt d\varphi + \frac{\rho^2}{\Delta} dr^2 + \rho^2 d\theta^2 + \left(r^2 + a^2 + \frac{2Ma^2 r \sin^2 \theta}{\rho^2} \right) \sin^2 \theta d\varphi^2,$$

where

$$\begin{aligned} \rho^2 &= r^2 + a^2 \cos^2 \theta, \\ \Delta &= r^2 - 2Mr + a^2, \end{aligned}$$

and $M, a \in \mathbb{R}$ are constants. Note that the Schwarzschild metric is a particular case, corresponding to $a = 0$. It is possible to prove that the Kerr metric solves the vacuum Einstein field equations (see for instance [O'N95]).

The Kerr metric is not spherically symmetric, but admits a two-dimensional group of isometries, generated by the Killing vector fields $X = \frac{\partial}{\partial t}$ and $Y = \frac{\partial}{\partial \varphi}$. The Komar mass associated to X is

$$M_{\text{Komar}} = -\frac{1}{8\pi} \int_{\Sigma} \star dX^{\sharp},$$

where Σ is a 2-surface of constant (t, r) , and can be computed to be

$$M_{\text{Komar}} = M.$$

The expression for the Komar mass in terms of the energy-momentum tensor, given in Chapter 6, suggests the definition of the **Komar angular momentum** as

$$J_{\text{Komar}} = \frac{1}{16\pi} \int_{\Sigma} \star dY^{\sharp}$$

(note the change in sign and absolute value of the constant, due to the fact that Y is now spacelike and essentially orthogonal to the timelike unit

normal N). The same exact argument as was done for the Komar mass shows that J_{Komar} does not depend on the choice of Σ . Performing the calculation for the Kerr metric yields

$$J_{\text{Komar}} = Ma,$$

and so the parameter a can be interpreted as the angular momentum per unit mass.

The Killing vector X becomes null on the hypersurface given by the equation

$$r = M + \sqrt{M^2 - a^2 \cos^2 \theta},$$

known as the **ergosphere**. However, it is easy to show that the metric induced on this hypersurface is Lorentzian, and so it cannot be the black hole event horizon (since it can be crossed both ways by timelike curves). The event horizon corresponds to the hypersurface $r = r_+$, where

$$r_+ = M + \sqrt{M^2 - a^2}.$$

Indeed, the function Δ changes sign on this hypersurface, and so $\text{grad } r$ becomes timelike (meaning that r must decrease along causal curves). Note that the ergosphere encloses the event horizon, touching it only at the poles, as shown in Figure 1. The region in between, where X is spacelike, is called the **ergoregion**, because matter fields satisfying the dominant energy condition can have negative energy there, and so extract energy from the black hole when absorbed (a mechanism known as the **Penrose process** in the case of particles or **superradiance** in the case of fields). Note also that the existence of an event horizon requires that $|a| \leq M$. If $|a| < M$ the black hole is said to be **subextremal**, and if $|a| = M$ it is called **extremal**.

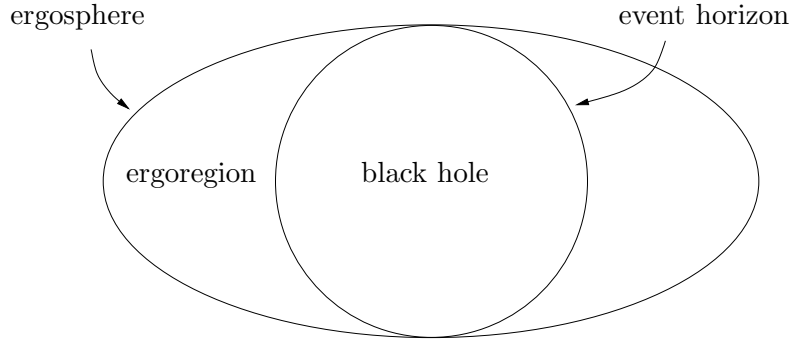


FIGURE 1. Spacelike cross-section of the Kerr solution.

A time-oriented spacetime (M, g) is said to be **asymptotically flat** if it contains an open set \mathcal{I} where the metric is well approximated (in a certain sense which we will not make precise) by the Minkowski metric in the region $\{x^2 + y^2 + z^2 > R^2\}$, for sufficiently large $R > 0$. For such spacetimes we can define the **black hole region** as $\mathcal{B} = M \setminus J^-(\mathcal{I})$, so that \mathcal{B} consists of

the events which cannot send signals to \mathcal{I} . If $\mathcal{B} \neq \emptyset$, the spacetime is called a **black hole spacetime**, and $\mathcal{H}^+ = \partial\mathcal{B}$ is called the **event horizon**. Finally, an asymptotically flat spacetime is called **stationary** if it admits a Killing vector field which is timelike on \mathcal{I} .

The importance of the Kerr solution stems from the following result.

THEOREM 1.1. (**Israel [Isr67], Carter [Car71], Hawking [Haw72], Robinson [Rob75], Chruściel and Costa [CC08]**) *The Kerr solution is the only real-analytic, stationary black hole spacetime satisfying the vacuum Einstein equations.*

2. Killing horizons and the zeroth law

DEFINITION 2.1. *A **Killing horizon** is a null surface which is orthogonal to a nonvanishing Killing vector field.*

Note that the Killing vector field is therefore null on the corresponding Killing horizon, and tangent to it. Killing horizons are important due to the following result.

THEOREM 2.2. (**Hawking [Haw72]**) *The event horizon of a stationary black hole spacetime whose matter fields satisfy hyperbolic equations and the weak energy condition is a Killing horizon.*

PROPOSITION 2.3. *The integral curves of a nonvanishing normal to a null hypersurface (e.g. the Killing vector field orthogonal to a Killing horizon) are reparameterized null geodesics.*

PROOF. Let Z be nonvanishing and normal to a null hypersurface S , and let $p \in S$. For any tangent vector $v \in T_p S$ it is easy to construct a local vector field V which is tangent to S , commutes with Z and satisfies $V_p = v$. We have

$$\langle V, \nabla_Z Z \rangle = -\langle \nabla_Z V, Z \rangle = -\langle \nabla_V Z, Z \rangle = -\frac{1}{2} V \cdot \langle Z, Z \rangle = 0.$$

Since p and v are arbitrary, we conclude that $\nabla_Z Z$ is orthogonal to S , i.e.

$$\nabla_Z Z = kZ,$$

for some function $k : S \rightarrow \mathbb{R}$. □

DEFINITION 2.4. *If \mathcal{H} is a Killing horizon associated to the Killing vector field Z then the function $k : \mathcal{H} \rightarrow \mathbb{R}$ such that*

$$\nabla_Z Z = kZ,$$

*on \mathcal{H} is called the **surface gravity** of \mathcal{H} relative to Z .*

THEOREM 2.5. (**Zeroth law of black hole thermodynamics**) *The surface gravity of a Killing horizon on a spacetime satisfying the dominant energy condition is a constant function.*

PROOF. Let Z be a Killing vector field associated to a Killing horizon \mathcal{H} . Since Z is orthogonal to \mathcal{H} , we have from the Frobenius theorem that

$$Z^\sharp \wedge dZ^\sharp = 0$$

on \mathcal{H} . Because Z^\sharp is nonvanishing, and can therefore be completed to a coframe, we necessarily have

$$dZ^\sharp = 2Z^\sharp \wedge U^\sharp \Leftrightarrow \nabla_\alpha Z_\beta = Z_\alpha U_\beta - U_\alpha Z_\beta$$

on \mathcal{H} , for some vector field U . If the vector fields X and Y are tangent to \mathcal{H} then

$$(18) \quad X^\alpha Y^\beta \nabla_\alpha Z_\beta = 0,$$

and consequently

$$(19) \quad Y^\alpha Z^\beta \nabla_\alpha X_\beta = -Y^\alpha X^\beta \nabla_\alpha Z_\beta = 0.$$

on \mathcal{H} . Taking the derivative of (18) along any vector field V tangent to \mathcal{H} yields

$$\begin{aligned} V^\mu X^\alpha Y^\beta \nabla_\mu \nabla_\alpha Z_\beta &= -V^\mu (\nabla_\mu X^\alpha) Y^\beta \nabla_\alpha Z_\beta - V^\mu X^\alpha (\nabla_\mu Y^\beta) \nabla_\alpha Z_\beta \\ &= -V^\mu (Y^\beta \nabla_\mu X^\alpha + X^\alpha \nabla_\mu Y^\beta) (Z_\alpha U_\beta - U_\alpha Z_\beta) = 0, \end{aligned}$$

in view of (19). On the other hand, we saw in Chapter 6 that any Killing vector field Z satisfies

$$(20) \quad \nabla_\mu \nabla_\alpha Z_\beta = -R_{\alpha\beta\mu}{}^\nu Z_\nu,$$

whence

$$(21) \quad R_{\alpha\beta\mu\nu} X^\alpha Y^\beta V^\mu Z^\nu = 0$$

for any three vector fields X, Y, V tangent to \mathcal{H} . If X and Y are orthonormal then they can always be completed to a local frame $\{X, Y, Z, W\}$ such W is null, orthogonal to X and Y , and normalized against Z ,

$$Z_\mu W^\mu = -1.$$

In this frame the metric is written

$$g_{\mu\nu} = X_\mu X_\nu + Y_\mu Y_\nu - Z_\mu W_\nu - W_\mu Z_\nu,$$

and so we obtain, for any vector field V tangent to \mathcal{H} ,

$$(22) \quad \begin{aligned} R_{\alpha\beta} V^\alpha Z^\beta &= R_{\alpha\mu\beta\nu} V^\alpha Z^\beta g^{\mu\nu} = -R_{\alpha\mu\beta\nu} V^\alpha Z^\beta Z^\mu W^\nu \\ &= V^\alpha Z^\beta W^\nu \nabla_\alpha \nabla_\beta Z_\nu, \end{aligned}$$

where we used (21) and (20).

We have

$$(23) \quad k = -\langle W, \nabla_Z Z \rangle = -W^\nu Z^\mu \nabla_\mu Z_\nu.$$

Differentiating (23) along a vector field V tangent \mathcal{H} yields

$$V \cdot k = -(V^\alpha \nabla_\alpha W^\nu) Z^\mu \nabla_\mu Z_\nu - W^\nu (V^\alpha \nabla_\alpha Z^\mu) \nabla_\mu Z_\nu - W^\nu Z^\mu V^\alpha \nabla_\alpha \nabla_\mu Z_\nu.$$

The first term in the right-hand side of this equation is

$$-(V^\alpha \nabla_\alpha W^\nu) k Z_\nu = k W^\nu V^\alpha \nabla_\alpha Z_\nu = k W^\nu V^\alpha (Z_\alpha U_\nu - U_\alpha Z_\nu) = k V^\alpha U_\alpha,$$

whereas the second term is

$$-W^\nu V^\alpha (Z_\alpha U^\mu - U_\alpha Z^\mu) \nabla_\mu Z_\nu = V^\alpha U_\alpha W^\nu Z^\mu \nabla_\mu Z_\nu = -k V^\alpha U_\alpha.$$

Therefore these terms cancel out, and, using (22), we obtain

$$V \cdot k = -R_{\alpha\beta} V^\alpha Z^\beta.$$

From (22) it is clear that

$$R_{\alpha\beta} Z^\alpha Z^\beta = 0,$$

implying that the vector field

$$I_\alpha = R_{\alpha\beta} Z^\beta$$

is tangent to \mathcal{H} . By Einstein's equation, we have

$$I_\alpha = \left(\frac{1}{2} R g_{\alpha\beta} - \Lambda g_{\alpha\beta} + 8\pi T_{\alpha\beta} \right) Z^\beta = \frac{1}{2} R Z_\alpha - \Lambda Z_\alpha + 8\pi T_{\alpha\beta} Z^\beta,$$

where R is the scalar curvature, Λ is the cosmological constant and $T_{\alpha\beta}$ is the energy-momentum tensor. Therefore the vector field

$$J_\alpha = T_{\alpha\beta} Z^\beta$$

is also tangent to \mathcal{H} . Since $T_{\alpha\beta}$ satisfies the dominant energy condition, the vector field J must be causal, and since it is tangent to \mathcal{H} it can only be null and parallel to Z . We conclude that the vector field I is also proportional to Z , and so

$$V \cdot k = -I_\alpha V^\alpha = 0,$$

that is, k is constant along \mathcal{H} . \square

Let λ be a local coordinate that is an affine parameter for the null geodesics along \mathcal{H} , that is,

$$\left\langle \frac{\partial}{\partial \lambda}, \frac{\partial}{\partial \lambda} \right\rangle = 0 \quad \text{and} \quad \nabla_{\frac{\partial}{\partial \lambda}} \frac{\partial}{\partial \lambda} = 0$$

on \mathcal{H} . Then it is easy to see that

$$Z = k(\lambda - \lambda_0) \frac{\partial}{\partial \lambda}$$

on \mathcal{H} , where λ_0 may depend on the null geodesic. In other words, Z vanishes on some cross-section of \mathcal{H} . If Z and $\frac{\partial}{\partial \lambda}$ are future-pointing then Z vanishes to the past when k is positive (that is, $\lambda > \lambda_0$), and to the future if k is negative (that is, $\lambda < \lambda_0$).

If t is a local coordinate in a neighborhood of \mathcal{H} such that

$$Z = \frac{\partial}{\partial t}$$

then we have on \mathcal{H}

$$Z = \frac{d\lambda}{dt} \frac{\partial}{\partial \lambda} = k(\lambda - \lambda_0) \frac{\partial}{\partial \lambda},$$

implying that

$$\lambda - \lambda_0 = Ce^{kt},$$

where C may depend on the null geodesic. By rescaling λ conveniently we may assume that

$$Z = e^{kt} \frac{\partial}{\partial \lambda}$$

on \mathcal{H} .

Now consider a timelike congruence crossing \mathcal{H} , with tangent unit timelike vector field U satisfying

$$[Z, U] = 0.$$

This can be accomplished by taking a timelike hypersurface transverse to Z , ruled by timelike curves, and moving each curve by the flow of Z . The quantity

$$E = - \left\langle \frac{\partial}{\partial \lambda}, U \right\rangle = - \left\langle e^{-kt} Z, U \right\rangle$$

represents the energy of a given null geodesic as measured by an observer of the congruence when crossing the Killing horizon, and is related to the frequency of the associated wave. Since Z is a Killing field, we have

$$\begin{aligned} Z \cdot E &= -\mathcal{L}_Z \left\langle e^{-kt} Z, U \right\rangle = - \left\langle \mathcal{L}_Z(e^{-kt} Z), U \right\rangle - \left\langle e^{-kt} Z, \mathcal{L}_Z U \right\rangle \\ &= - \left\langle [Z, e^{-kt} Z], U \right\rangle - \left\langle e^{-kt} Z, [Z, U] \right\rangle = \left\langle ke^{-kt} Z, U \right\rangle = -kE, \end{aligned}$$

implying that

$$E = E_0 e^{-kt}.$$

In other words, the energy of the null geodesic as measured by the observers of the congruence decreases exponentially if $k > 0$ (**redshift effect**), and increases exponentially if $k < 0$ (**blueshift effect**).

3. Smarr's formula and the first law

Unlike the case of the Schwarzschild black hole, where the event horizon is the Killing horizon corresponding to $X = \frac{\partial}{\partial t}$, the event horizon of the Kerr black hole is a Killing horizon for the Killing vector field

$$Z = X + \Omega Y,$$

where $Y = \frac{\partial}{\partial \varphi}$ and $\Omega \in \mathbb{R}$ is an appropriate constant.

DEFINITION 3.1. Ω is called the **angular velocity of the event horizon**.

To find Ω , we write the quadratic equation in ω for the vector $X + \omega Y$ to be null at some point with $r > r_+$ (so that we can use Boyer-Lindquist coordinates):

$$- \left(1 - \frac{2Mr}{\rho^2} \right) - \frac{4Mar \sin^2 \theta}{\rho^2} \omega + \left(r^2 + a^2 + \frac{2Ma^2 r \sin^2 \theta}{\rho^2} \right) \sin^2 \theta \omega^2 = 0.$$

The discriminant of this equation has the simple form $\Delta \sin^2 \theta$, and consequently vanishes when we take $r = r_+$, in which case the quadratic equation has the single solution

$$\Omega = \frac{a}{r_+^2 + a^2} = \frac{a}{2Mr_+}.$$

This solution is the limit as $r \rightarrow r_+$ of the values of $\omega = \omega(r, \theta)$ such that $X + \omega Y$ is null, and so it must coincide with the angular velocity of the event horizon.

From the expressions of the Komar mass and angular momentum, it is clear that

$$M - 2\Omega J = -\frac{1}{8\pi} \int_{\Sigma} \star dZ^\sharp$$

for any compact orientable 2-surface Σ enclosing the event horizon \mathcal{H} . Let us consider the case when Σ is a spacelike cross-section of \mathcal{H} (Figure 2). We can uniquely define a future-pointing unit timelike vector field N and an unit spacelike vector field n , both orthogonal to Σ , such that $Z = N + n$. Because Z is a Killing vector field, $\nabla_\mu Z_\nu$ is a 2-form; more precisely,

$$(dZ^\sharp)_{\mu\nu} = \nabla_\mu Z_\nu - \nabla_\nu Z_\mu = 2\nabla_\mu Z_\nu.$$

If E_1 and E_2 are two unit vector fields tangent to Σ such that $\{N, n, E_1, E_2\}$ is a positive orthonormal frame, and so $\{-N^\sharp, n^\sharp, E_1^\sharp, E_2^\sharp\}$ is a positive orthonormal coframe, then we can expand

$$\nabla Z^\sharp = -\nabla Z^\sharp(N, n)N^\sharp \wedge n^\sharp + \dots$$

Therefore,

$$M - 2\Omega J = -\frac{1}{4\pi} \int_{\Sigma} \star \nabla Z^\sharp = \frac{1}{4\pi} \int_{\Sigma} \nabla Z^\sharp(N, n)E_1^\sharp \wedge E_2^\sharp.$$

Since

$$\nabla Z^\sharp(N, n) = \nabla Z^\sharp(Z, n) = \langle \nabla_Z Z, n \rangle = \langle kZ, n \rangle = k,$$

we finally obtain the **Smarr formula**:

$$M = \frac{kA}{4\pi} + 2\Omega J,$$

where A is the area of the cross-section Σ (which in particular is the same for all cross-sections of \mathcal{H}).

A cross section of the event horizon can be obtained by taking the limit as $r \rightarrow r_+$ of a surface of constant (t, r) . The induced metric is then

$$ds^2 = (r_+^2 + a^2 \cos^2 \theta) d\theta^2 + \left(r_+^2 + a^2 + \frac{2Ma^2 r_+ \sin^2 \theta}{r_+^2 + a^2 \cos^2 \theta} \right) \sin^2 \theta d\varphi^2,$$

and its area is

$$A = 4\pi(r_+^2 + a^2) = 8\pi(M^2 + M\sqrt{M^2 - a^2}),$$

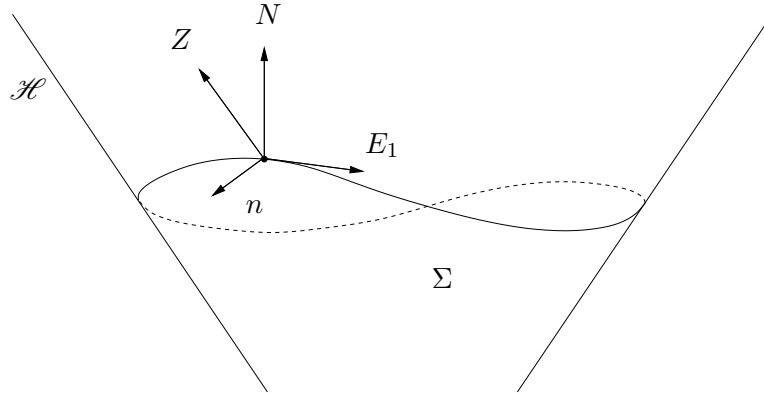


FIGURE 2. Spacelike cross-section of the event horizon.

whence

$$M = \sqrt{\frac{A}{16\pi} + \frac{4\pi J^2}{A}}.$$

Noting that $M = M(A, J)$ is homogeneous of degree $1/2$, we know from Euler's homogeneous function theorem that

$$\frac{1}{2}M = \frac{\partial M}{\partial A}A + \frac{\partial M}{\partial J}J.$$

On the other hand, it is easy to check that

$$\frac{\partial M}{\partial J} = \Omega.$$

Smarr's formula then implies the following result:

THEOREM 3.2. (First law of black hole thermodynamics) *The function $M = M(A, J)$ giving the mass of a Kerr black hole as a function of the area of (spacelike cross-sections of) its horizon and its angular momentum satisfies*

$$dM = \frac{k}{8\pi}dA + \Omega dJ.$$

This formula provides an easy way to compute the surface gravity of a Kerr black hole horizon:

$$k = \frac{1}{4M} - M\Omega^2.$$

In particular, for an extremal black hole ($r_+ = a = M$) we have

$$\Omega = \frac{1}{2M} \Rightarrow k = 0.$$

4. Second law

We consider arbitrary test fields propagating on a Kerr background. Apart from ignoring their gravitational backreaction, we make no further hypotheses on the fields: they could be any combination of scalar or electromagnetic fields, fluids, elastic media, or other types of matter. By the Einstein equation, their combined energy-momentum tensor T must satisfy

$$\nabla_\mu T^{\mu\nu} = 0.$$

Using the symmetry of T and the Killing equation,

$$\nabla_\mu X_\nu + \nabla_\nu X_\mu = 0,$$

we have

$$\nabla_\mu (T^{\mu\nu} X_\nu) = 0.$$

This conservation law suggests that the total field energy on a given spacelike hypersurface S extending from the black hole event horizon \mathcal{H}^+ to infinity (Figure 3) should be

$$E = \int_S T^{\mu\nu} X_\nu N_\mu,$$

where N is the future-pointing unit normal to S .

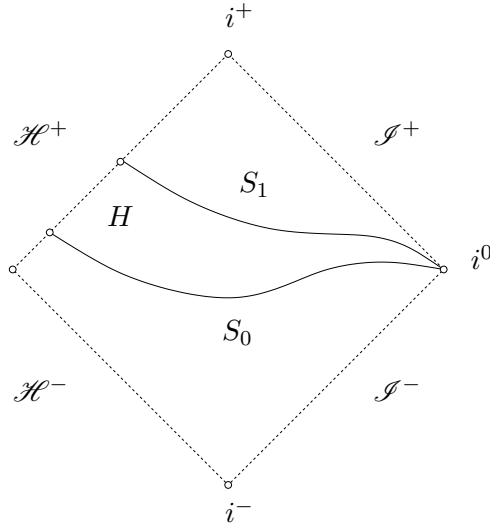


FIGURE 3. Penrose diagrams for the region of outer communication of the Kerr spacetime.

Analogously, the total field angular momentum on a spacelike hypersurface S extending from the event horizon to infinity is

$$(24) \quad L = - \int_S T^{\mu\nu} Y_\nu N_\mu,$$

where the minus sign accounts for the timelike unit normal.

Consider now two such spacelike hypersurfaces, S_0 and S_1 , with S_1 to the future of S_0 (Figure 3). The energy absorbed by the black hole across the subset H of \mathcal{H}^+ between S_0 and S_1 is then

$$\Delta M = \int_{S_0} T^{\mu\nu} X_\nu N_\mu - \int_{S_1} T^{\mu\nu} X_\nu N_\mu,$$

whereas the angular momentum absorbed by the black hole across H is

$$\Delta J = - \int_{S_0} T^{\mu\nu} Y_\nu N_\mu + \int_{S_1} T^{\mu\nu} Y_\nu N_\mu.$$

Therefore, we have

$$\Delta M - \Omega \Delta J = \int_{S_0} T^{\mu\nu} Z_\nu N_\mu - \int_{S_1} T^{\mu\nu} Z_\nu N_\mu.$$

Because Z is also a Killing vector field,

$$\nabla_\mu (T^{\mu\nu} Z_\nu) = 0,$$

and so the divergence theorem, applied to the region bounded by S_0 , S_1 and H , yields

$$\Delta M - \Omega \Delta J = \int_H T^{\mu\nu} Z_\nu Z_\mu$$

(we use $-Z$ as the null normal on H). Therefore we have the following result.

THEOREM 4.1. (Second law of black hole thermodynamics, test field version) *If the energy-momentum tensor T corresponding to any collection of test fields propagating on a Kerr background satisfies the null energy condition at the event horizon then the energy ΔM and the angular momentum ΔJ absorbed by the black hole satisfy*

$$\Delta M \geq \Omega \Delta J.$$

If we think of Kerr black holes as stationary states and imagine that the interaction of a Kerr black hole with test fields results in a new Kerr black hole, then, in view of the first law of black hole thermodynamics, we can rewrite the result above as

$$dA = \frac{8\pi}{k} (dM - \Omega dJ) \geq 0,$$

that is, the area of the event horizon of a Kerr black hole can only increase as a result of its interaction with test fields.

In fact, it is possible to prove a general result about the area of the event horizon of a black hole spacetime.

PROPOSITION 4.2. *The event horizon of a black hole spacetime is ruled by null geodesics.*

PROOF. Given the causal structure of Minkowski's spacetime, it is clear that $J^-(\mathcal{I})$ coincides with $I^-(\mathcal{I})$, an open set. Let $p \in \mathcal{H}$ be any point in the event horizon and let $U \ni p$ be a simple neighborhood (see Proposition 4.1 in Chapter 4). Given a sequence $\{p_n\} \subset U \cap J^-(\mathcal{I})$ converging to p , let c_n be a future-pointing causal curve connecting p_n to \mathcal{I} , let q_n be the first intersection of c_n with ∂U , and let γ_n be the future-pointing causal geodesic connecting p_n to q_n in U (see Figure 4). Since the exponential map centered on any point in U is a diffeomorphism, these geodesics converge to a future-pointing causal geodesic γ in U with initial point p . Note that γ cannot enter $J^-(\mathcal{I})$, because then we would have $p \in J^-(\mathcal{I}) \equiv \text{int } J^-(\mathcal{I})$, in contradiction with $p \in \mathcal{H} \equiv \partial J^-(\mathcal{I})$. Moreover, every point in γ is the limit of points in γ_n , hence points in $J^-(\mathcal{I})$. We conclude that γ is a curve on $\overline{J^-(\mathcal{I})} \setminus \text{int } J^-(\mathcal{I}) = \mathcal{H}$. Finally, γ cannot be a timelike geodesic, because then the sequence q_n would enter the open set $I^+(p)$, and we would again have $p \in J^-(\mathcal{I})$. We conclude that γ is a null geodesic. If we extend γ maximally towards the future we obtain a future-inextendible null geodesic; covering this curve with simple neighborhoods and applying similar arguments to the above, one can easily show that it never leaves \mathcal{H} . \square

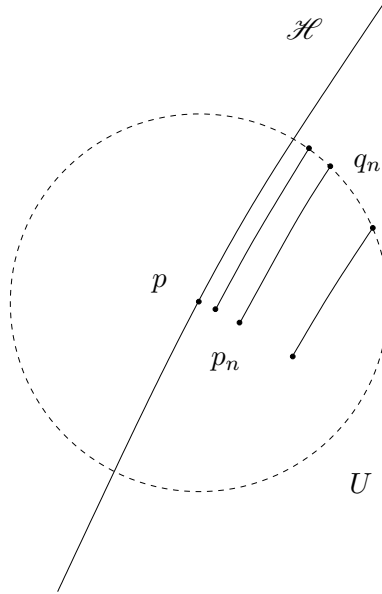


FIGURE 4. Proof of Proposition 4.2.

THEOREM 4.3. (Second law of black hole thermodynamics, Hawking's version [Haw72]) *If the energy-momentum tensor of a black hole spacetime satisfies the null energy condition at the event horizon and the null geodesics ruling the event horizon are complete towards the future then their expansion is never negative. In particular, the area of any spacelike cross-section of the event horizon cannot decrease towards the future.*

PROOF. Note that the null geodesics ruling the event horizon are orthogonal to any spacelike cross-section Σ , so that the discussion preceding the proof of Penrose's singularity theorem in Chapter 4 applies. Suppose that the expansion of some null geodesic were negative at some point $p \in \mathcal{H}$. Then, by the analogue of Proposition 6.2 in Chapter 4, that null geodesic would have a conjugate point to the future of p , after which, by the analogue of Proposition 6.3 in Chapter 4, it would leave \mathcal{H} . Since this would contradict Proposition 4.2, the expansion can never be negative. \square

5. Hawking radiation and black hole thermodynamics

Recall the three laws of thermodynamics for, say, a gas:

- **Zeroth law:** The temperature is constant throughout the gas when thermal equilibrium has been reached.
- **First law:** The internal energy U as a function of the gas entropy S and volume V satisfies

$$dU = TdS - pdV,$$

where T is the temperature and p is the pressure.

- **Second law:** The entropy of the gas cannot decrease towards the future.

These are remarkably similar to the three laws of black hole thermodynamics, if we identify the black hole mass M with the internal energy, (some multiple of) the horizon's surface gravity with the black hole's temperature and (some multiple of) the horizon's area with the black hole's entropy. Inspired by this analogy, Bekenstein [Bek72] proposed in 1972 that black holes are indeed thermodynamic systems. Hawking initially resisted this suggestion, since objects in thermal equilibrium at a given temperature must emit black body radiation, which black holes cannot (classically) do. In 1974, however, Hawking [Haw74] applied methods of quantum field theory in curved spacetime to show that black holes do indeed emit particles with a thermal spectrum corresponding to the temperature

$$T = \frac{k}{2\pi}.$$

In particular, this fixed the black hole entropy as

$$S = \frac{A}{4}.$$

6. Exercises

- (1) To compute the Komar mass and angular momentum of the Kerr solution we consider the region $r \gg M, a$.

- (a) Show that a positive orthonormal coframe is approximately given in this region by

$$\begin{aligned}\omega^0 &\sim dt, & \omega^r &\sim dr, & \omega^\theta &\sim r d\theta, \\ \omega^\varphi &\sim r \sin \theta d\varphi - \frac{2Ma \sin \theta}{r^2} dt.\end{aligned}$$

- (b) Establish the following asymptotic formulas:

$$\begin{aligned}X^\# &\sim -\left(1 - \frac{2M}{r}\right) dt - \frac{2Ma \sin^2 \theta}{r} d\varphi; \\ Y^\# &\sim -\frac{2Ma \sin^2 \theta}{r} dt + r^2 \sin^2 \theta d\varphi; \\ dX^\# &\sim \frac{2M}{r^2} \omega^0 \wedge \omega^r + \dots; \\ dY^\# &\sim -\frac{6Ma \sin^2 \theta}{r^2} \omega^0 \wedge \omega^r + \dots\end{aligned}$$

- (c) Prove that $M_{\text{Komar}} = M$ and $J_{\text{Komar}} = Ma$.

- (2) Show that the metric induced on the ergosphere is

$$\begin{aligned}ds^2 &= -2a \sin^2 \theta dt d\varphi + \frac{2M^3 r}{(r-M)^2} d\theta^2 \\ &\quad + (r^2 + a^2 + a^2 \sin^2 \theta) \sin^2 \theta d\varphi^2,\end{aligned}$$

where $r = M + \sqrt{M^2 - a^2 \cos^2 \theta}$. Prove that this metric is Lorentzian.

- (3) The symmetry semiaxis $\theta = 0$ is a totally geodesic submanifold of the Kerr solution, with metric

$$ds^2 = -\frac{\Delta}{r^2 + a^2} dt^2 + \frac{r^2 + a^2}{\Delta} dr^2.$$

Obtain the maximal analytical extension of this submanifold. Note that $r = 0$ is not a singularity, and so the metric can be continued for negative values of r .

- (4) Consider the static and spherically symmetric metric given in local coordinates by

$$ds^2 = -V(r) dt^2 + V(r) dr^2 + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2).$$

- (a) Show that this metric can be written in the form

$$ds^2 = -V(r) dv^2 + 2dv dr + r^2 (d\theta^2 + \sin^2 \theta d\varphi^2),$$

where

$$v = t + \int \frac{dr}{V(r)},$$

- (b) Assume that there V has an isolated zero at $r = r_H$, so that this hypersurface is a Killing horizon. Show that the corresponding surface gravity relative to $\frac{\partial}{\partial v}$ is

$$k = \frac{1}{2}V'(r_H).$$

- (5) Consider a one-parameter family of null geodesics $\gamma(\tau, \lambda)$ connecting two timelike curves $c_0(\tau) = \gamma(\tau, 0)$ and $c_1(\tau) = \gamma(\tau, 1)$. Show that

$$\frac{\partial}{\partial \lambda} \left\langle \frac{\partial \gamma}{\partial \tau}, \frac{\partial \gamma}{\partial \lambda} \right\rangle = 0,$$

and use this to prove that if τ is the proper time for the curve c_0 and

$$\tau'(\tau) = \int_{\tau_0}^{\tau} |\dot{c}_1(t)| dt$$

is the proper time for the curve c_1 then

$$\frac{d\tau'}{d\tau} = \frac{E_0}{E_1},$$

where

$$E_0 = - \left\langle \dot{c}_0, \frac{\partial \gamma}{\partial \lambda} \right\rangle \quad \text{and} \quad E_1 = - \left\langle \frac{\dot{c}_1}{|\dot{c}_1|}, \frac{\partial \gamma}{\partial \lambda} \right\rangle$$

are the energies of the null geodesic $\gamma(\tau, \lambda)$ as measured by the observers corresponding to c_0 and c_1 .

- (6) Compute the area, the angular velocity and the surface gravity of a Kerr black hole horizon (you may find the relation $r_+^2 + a^2 = 2Mr_+$ to be useful here).
- (7) Prove that test fields satisfying the null energy condition at the event horizon cannot destroy an extremal Kerr black hole. More precisely, prove that if an extremal black hole is characterized by the physical quantities (M, J) , and absorbs energy and angular momentum $(\Delta M, \Delta J)$ by interacting with the test fields, then the metric corresponding to the physical quantities $(M + \Delta M, J + \Delta J)$ represents, to first order in ΔM and ΔJ , either a subextremal or an extremal Kerr black hole.
- (8) Show that if a test field satisfying the null energy condition at the event horizon extracts energy from a Kerr black hole then $a = J/M$ always decreases. What fraction of the black hole's mass can be extracted?
- (9) Check that the second law holds for the black hole resulting from an Oppenheimer-Snyder collapse.
- (10) Use the second law of black hole thermodynamics to:
- (a) Prove that a Schwarzschild black hole cannot split into two Kerr black holes;

- (b) Give an upper bound to the energy released in the form of gravitational waves when two Kerr black holes coalesce to form a Schwarzschild black hole. Can the efficiency of this process ever exceed 50%?

Appendix: Mathematical concepts for physicists

In this appendix we list some mathematical concepts which will be used in the main text, for the benefit of readers whose background is in Physics.

Topology

DEFINITION 6.1. A **topological space** is a set M with a **topology**, that is, a list of the **open subsets** of M , satisfying:

- (1) Both \emptyset and M are open;
- (2) Any union of open sets is open;
- (3) Any finite intersection of open sets is open.

All the usual topological notions can now be defined. For instance, a **closed set** is a set whose complement is open. The **interior** $\text{int } A$ of a subset $A \subset M$ is the largest open set contained in A , its **closure** \bar{A} is the smallest closed set containing A , and its **boundary** is $\partial A = \bar{A} \setminus \text{int } A$.

The main object of topology is the study of limits and continuity.

DEFINITION 6.2. A sequence $\{p_n\}$ is said to **converge** to $p \in M$ if for any open set $U \ni p$ there exists $k \in \mathbb{N}$ such that $p_n \in U$ for all $n \geq k$.

DEFINITION 6.3. A map $f : M \rightarrow N$ between two topological spaces is said to be **continuous** if for each open set $U \subset N$ the preimage $f^{-1}(U)$ is an open subset of M . A bijection f is called a **homeomorphism** if both f and its inverse f^{-1} are continuous.

A system of local coordinates on a manifold is an example of a homeomorphism between the coordinate neighborhood and an open set of \mathbb{R}^n .

Two fundamental concepts in topology are compactness and connectedness.

DEFINITION 6.4. A subset $A \subset M$ is said to be **compact** if every cover of A by open sets admits a finite subcover. It is said to be **connected** if it is impossible to write $A = (A \cap U) \cup (A \cap V)$ with U, V disjoint open sets and $A \cap U, A \cap V \neq \emptyset$.

The following result generalizes the theorems of Weierstrass and Bolzano.

THEOREM 6.5. *Continuous maps carry compact sets to compact sets, and connected sets to connected sets.*

Metric spaces

DEFINITION 6.6. A **metric space** is a set M and a **distance function** $d : M \times M \rightarrow [0, +\infty)$ satisfying:

- (1) **Positivity:** $d(p, q) \geq 0$ and $d(p, q) = 0$ if and only if $p = q$;
- (2) **Symmetry:** $d(p, q) = d(q, p)$;
- (3) **Triangle inequality:** $d(p, r) \leq d(p, q) + d(q, r)$,

for all $p, q, r \in M$.

The **open ball** with center p and radius ε is the set

$$B_\varepsilon(p) = \{q \in M \mid d(p, q) < \varepsilon\}.$$

Any metric space has a natural topology, whose open sets are unions of open balls. In this topology $p_n \rightarrow p$ if and only if $d(p_n, p) \rightarrow 0$, $F \subset M$ is closed if and only if every convergent sequence in F has limit in F , and $K \subset M$ is compact if and only if every sequence in K has a sublimit in K .

A fundamental notion for metric spaces is completeness.

DEFINITION 6.7. A sequence $\{p_n\}$ in M is said to be a **Cauchy sequence** if for all $\varepsilon > 0$ there exists $N \in \mathbb{N}$ such that $d(p_n, p_m) < \varepsilon$ for all $m, n > N$. A metric space is said to be **complete** if all its Cauchy sequences converge.

In particular any compact metric space is complete.

Hopf-Rinow theorem

DEFINITION 6.8. A Riemannian manifold is said to be **geodesically complete** if any geodesic is defined for every value of its parameter.

DEFINITION 6.9. Let (M, g) be a connected Riemannian manifold and $p, q \in M$. The **distance** between p and q is defined as

$$d(p, q) = \inf\{l(\gamma) \mid \gamma \text{ is a smooth curve connecting } p \text{ to } q\}.$$

It is easily seen that (M, d) is a metric space. Remarkably, the completeness of this metric space is equivalent to geodesic completeness.

THEOREM 6.10. (Hopf-Rinow) A connected Riemannian manifold (M, g) is geodesically complete if and only if (M, d) is a complete metric space.

Differential forms

DEFINITION 6.11. A **differential-form** ω of degree k is simply a completely anti-symmetric k -tensor: $\omega_{\alpha_1 \dots \alpha_k} = \omega_{[\alpha_1 \dots \alpha_k]}$.

For instance, covector fields are differential forms of degree 1. Differential forms are useful because of their rich algebraic and differential structure.

DEFINITION 6.12. If ω is a k -form and η is an l -form then their **exterior product** is the $(k + l)$ -form

$$(\omega \wedge \eta)_{\alpha_1 \dots \alpha_k \beta_1 \dots \beta_l} = \frac{(k+l)!}{k!l!} \omega_{[\alpha_1 \dots \alpha_k} \eta_{\beta_1 \dots \beta_l]},$$

and the **exterior derivative** of ω is the $(k + 1)$ -form

$$(d\omega)_{\alpha_1 \dots \alpha_k} = (k+1) \nabla_{[\alpha} \omega_{\alpha_1 \dots \alpha_k]},$$

where ∇ is any symmetric connection.

It is easy to see that any k -form ω is given in local coordinates by

$$(25) \quad \omega = \sum_{\alpha_1 < \dots < \alpha_k} \omega_{\alpha_1 \dots \alpha_k}(x) dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_k},$$

and therefore has $\binom{n}{k}$ independent components on an n -dimensional manifold.

PROPOSITION 6.13. If ω , η and θ are differential forms then:

- (1) $\omega \wedge (\eta \wedge \theta) = (\omega \wedge \eta) \wedge \theta$;
- (2) $\omega \wedge \eta = (-1)^{(\deg \omega)(\deg \eta)} \eta \wedge \omega$;
- (3) $\omega \wedge (\eta + \theta) = \omega \wedge \eta + \omega \wedge \theta$;
- (4) $d(\omega + \eta) = d\omega + d\eta$;
- (5) $d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^{\deg \omega} \omega \wedge d\eta$;
- (6) $d^2\omega = 0$.

It is clear from these properties that if the k -form ω is given in local coordinates by (25) above then

$$d\omega = \sum_{\alpha_1 < \dots < \alpha_k} \sum_{\alpha} \partial_{\alpha} \omega_{\alpha_1 \dots \alpha_k} dx^{\alpha} \wedge dx^{\alpha_1} \wedge \dots \wedge dx^{\alpha_k}.$$

The last property in Proposition 6.13 has a converse, known as the **Poincaré Lemma**.

LEMMA 6.14. (Poincaré) If $d\omega = 0$ then locally $\omega = d\eta$.

A related result is the **Frobenius Theorem**. Here we present a particular case of this result.

THEOREM 6.15. (Frobenius) The nonvanishing 1-form ω is locally orthogonal to a family of hypersurfaces if and only if $\omega \wedge d\omega = 0$.

To prove the easy direction in this equivalence it suffices to note that ω is locally orthogonal to a family of hypersurfaces $\{f = \text{constant}\}$ if and only if $\omega = gdf$ for some nonvanishing function g . Note that in particular this is always true for 1-forms in 2-dimensional manifolds.

We will now assume that our n -dimensional manifold is **oriented**, that is, that an orientation can be, and has been, consistently chosen on every tangent space. Any n -form ω is written in local coordinates as

$$\omega = a(x) dx^1 \wedge \dots \wedge dx^n.$$

If the coordinate system is positive, that is, if the coordinate basis $\{\partial_1, \dots, \partial_n\}$ has positive orientation at all points, we define

$$\int_U \omega = \int_{x(U)} a(x) dx^1 \dots dx^n,$$

where U is the coordinate neighborhood. This formula does not depend on the choice of local coordinates because $dx^1 \wedge \dots \wedge dx^n$ transforms by the determinant of the change of variables.

THEOREM 6.16. (Stokes) *If M is an oriented n -dimensional manifold with boundary ∂M and ω is an $(n-1)$ -form then*

$$\int_M d\omega = \int_{\partial M} \omega.$$

In this theorem the orientations of M and ∂M are related as follows: if ∂M is a level set of x^1 in a positive coordinate system and ∂_1 points outwards then the coordinate system (x^2, \dots, x^n) on ∂M is positive.

If M has a metric g then its **volume element** is the n -form ϵ which is 1 when contracted with a positive orthonormal frame. In positive local coordinates we have

$$\epsilon = \sqrt{|\det(g_{\mu\nu})|} dx^1 \wedge \dots \wedge dx^n.$$

It is easily seen that $\nabla \epsilon = 0$, where ∇ is the Levi-Civita connection. If ω is a k -form then its **Hodge dual** is the $(n-k)$ -form $\star\omega$ given by

$$(\star\omega)_{\beta_1 \dots \beta_{n-k}} = \frac{1}{k!} \omega^{\alpha_1 \dots \alpha_k} \epsilon_{\alpha_1 \dots \alpha_k \beta_1 \dots \beta_{n-k}}.$$

The operator \star , called the Hodge star, can alternatively be defined as follows: if $\{\omega^1, \dots, \omega^n\}$ is any positively oriented orthonormal coframe (so that the volume element is $\epsilon = \omega^1 \wedge \dots \wedge \omega^n$) then

$$\begin{aligned} \star(\omega^1 \wedge \dots \wedge \omega^k) &= g(\omega^1, \omega^1) \dots g(\omega^k, \omega^k) \omega^{k+1} \wedge \dots \wedge \omega^n \\ &= \epsilon((\omega^1)^\sharp, \dots, (\omega^k)^\sharp, \dots). \end{aligned}$$

Lie derivative

A vector field X can be identified with the differential operator that corresponds to taking derivatives along X . In local coordinates, this operator is given by

$$X \cdot f = X^\mu \partial_\mu f.$$

It turns out that the commutator of two vector fields X and Y , regarded as differential operators, is also a vector field:

$$\begin{aligned} [X, Y] \cdot f &= X \cdot (Y^\mu \partial_\mu f) - Y \cdot (X^\mu \partial_\mu f) \\ &= (X \cdot Y^\mu) \partial_\mu f + Y^\mu X^\nu \partial_\nu \partial_\mu f - (Y \cdot X^\mu) \partial_\mu f - X^\mu Y^\nu \partial_\nu \partial_\mu f \\ &= (X \cdot Y^\mu - Y \cdot X^\mu) \partial_\mu f. \end{aligned}$$

DEFINITION 6.17. The **Lie bracket** of two vector fields X and Y is the vector field

$$[X, Y] = (X \cdot Y^\mu - Y \cdot X^\mu) \partial_\mu.$$

This operation is intimately related with the exterior derivative.

PROPOSITION 6.18. If ω is a 1-form then

$$d\omega(X, Y) = X \cdot \omega(Y) - Y \cdot \omega(X) - \omega([X, Y])$$

for all vector fields X and Y .

PROOF. In local coordinates we have

$$\begin{aligned} d\omega(X, Y) &= (\partial_\mu \omega_\nu - \partial_\nu \omega_\mu) X^\mu Y^\nu = Y^\nu X \cdot \omega_\nu - X^\nu Y \cdot \omega_\nu \\ &= X \cdot (\omega_\nu Y^\nu) - \omega_\nu X \cdot Y^\nu - Y \cdot (\omega_\nu X^\nu) + \omega_\nu Y \cdot X^\nu \\ &= X \cdot \omega(Y) - Y \cdot \omega(X) - \omega_\nu (X \cdot Y^\nu - Y \cdot X^\nu). \end{aligned}$$

□

If the vector field X is nonzero at some point p then there exists a coordinate system defined in a neighborhood of p such that $X = \partial_1$. In fact, we just have to fix local coordinates (x^2, \dots, x^n) on a hypersurface Σ transverse to X at p and let x^1 be the parameter for the flow of X starting at Σ . If T is any tensor, we define its **Lie derivative** along X as the tensor with components

$$(\mathcal{L}_X T)_{\beta_1 \dots \beta_l}^{\alpha_1 \dots \alpha_k} = \partial_1 T_{\beta_1 \dots \beta_l}^{\alpha_1 \dots \alpha_k}.$$

This can be extended to points where X vanishes by continuity. Although this definition seems to depend on the coordinate system, it is actually invariant. To check this, we just have to find an invariant expression for the Lie derivative of functions, vector fields and 1-forms and then notice that the Leibnitz rule applies.

PROPOSITION 6.19. If X is a vector field then:

- (1) $\mathcal{L}_X f = X \cdot f$ for functions f ;
- (2) $\mathcal{L}_X Y = [X, Y]$ for vector fields Y ;
- (3) $\mathcal{L}_X \omega = X \lrcorner d\omega + d(X \lrcorner \omega)$ for 1-forms ω

(where \lrcorner means contraction in the first index).

PROOF. The formula for functions is immediate. In the coordinate system where $X = \partial_1$,

$$\mathcal{L}_X Y = \partial_1 Y^\mu \partial_\mu = (X \cdot Y^\mu - Y \cdot X^\mu) \partial_\mu = [X, Y].$$

Finally, we have

$$\begin{aligned} (X \lrcorner d\omega + d(X \lrcorner \omega))(Y) &= d\omega(X, Y) + Y \cdot \omega(X) = X \cdot \omega(Y) - \omega([X, Y]) \\ &= \mathcal{L}_X(\omega(Y)) - \omega(\mathcal{L}_X Y) = (\mathcal{L}_X \omega)(Y), \end{aligned}$$

where we used the Leibnitz rule. This formula is sometimes called **Cartan's magic formula**. □

Cartan structure equations

Let $\{E_\mu\}$ be an orthonormal frame, and $\{\omega^\mu\}$ the corresponding orthonormal coframe, so that

$$\omega^\mu(E_\nu) = \delta^\mu{}_\nu.$$

Note that the metric can be written as

$$ds^2 = \eta_{\mu\nu} \omega^\mu \otimes \omega^\nu,$$

where $(\eta_{\mu\nu}) = \text{diag}(-1, 1, 1, 1)$ is the flat space metric (which we will use to raise and lower indices).

DEFINITION 6.20. *The **connection forms** associated to the orthonormal frame $\{E_\mu\}$ are the 1-forms $\omega^\mu{}_\nu$ such that*

$$\nabla_X E_\nu = \omega^\mu{}_\nu(X) E_\mu$$

*for all vector fields X . The **curvature forms** associated to this frame are the 2-forms $\Omega^\mu{}_\nu$ such that*

$$R(X, Y)E_\nu = \Omega^\mu{}_\nu(X, Y)E_\mu$$

for all vector fields X, Y .

Note that the components of the Riemann tensor in the orthonormal frame can be retrieved from the curvature forms by noticing that

$$\Omega^\mu{}_\nu = R_{\alpha\beta}{}^\mu{}_\nu \omega^\alpha \otimes \omega^\beta = \sum_{\alpha < \beta} R_{\alpha\beta}{}^\mu{}_\nu \omega^\alpha \wedge \omega^\beta.$$

These forms can be computed by using the so-called **Cartan structure equations**. This is by far the most efficient way to compute the curvature.

THEOREM 6.21. *The connection forms are the unique solution of **Cartan's first structure equations***

$$\begin{cases} \omega_{\mu\nu} = -\omega_{\nu\mu} \\ d\omega^\mu + \omega^\mu{}_\nu \wedge \omega^\nu = 0 \end{cases},$$

*and the curvature forms are given by **Cartan's second structure equations***

$$\Omega^\mu{}_\nu = d\omega^\mu{}_\nu + \omega^\mu{}_\alpha \wedge \omega^\alpha{}_\nu.$$

PROOF. The first condition is equivalent to

$$X \cdot \langle E_\mu, E_\nu \rangle = 0$$

for all vector fields X , which in turn is equivalent to the compatibility of the connection with the metric. Using

$$d\omega^\mu(X, Y) = X \cdot \omega^\mu(Y) - Y \cdot \omega^\mu(X) - \omega^\mu([X, Y])$$

for all vector fields X and Y , it is easy to see that the second condition is equivalent to

$$[E_\alpha, E_\beta] = \nabla_{E_\alpha} E_\beta - \nabla_{E_\beta} E_\alpha,$$

which in turn is equivalent to the symmetry of the connection. Since the Levi-Civita connection is the only connection which is symmetric and compatible with the metric, we conclude that Cartan's first structure equations have a unique solution.

Finally, the third condition can be derived by writing

$$\Omega^\mu{}_\nu(X, Y)E_\mu = R(X, Y)E_\nu = \nabla_X \nabla_Y E_\nu - \nabla_Y \nabla_X E_\nu - \nabla_{[X, Y]} E_\nu$$

in terms of the connection forms. \square

Bibliography

- [ADM61] R. Arnowitt, S. Deser, and C. Misner, *Coordinate invariance and energy expressions in general relativity*, Phys. Rev. **122** (1961), 997–1006.
- [AMA79] A. Ashtekar and A. Magnon-Ashtekar, *On conserved quantities in general relativity*, J. Math. Phys. **20** (1979), 793–800.
- [Bar86] R. Bartnik, *The mass of an asymptotically flat manifold*, Commun. Pure Appl. Math. **39** (1986), 661–693.
- [BCH73] J. Bardeen, B. Carter, and S. Hawking, *The four laws of black hole mechanics*, Commun. Math. Phys. **31** (1973), 161–170.
- [Bek72] J. Bekenstein, *Black holes and the second law*, Nuovo Cim. Lett. **4** (1972), 737–740.
- [Boo03] W. Boothby, *An introduction to differentiable manifolds and Riemannian geometry*, Academic Press, 2003.
- [Bra01] H. Bray, *Proof of the Riemannian Penrose inequality using the positive mass theorem*, J. Diff. Geom. **59** (2001), 177–267.
- [Bra11] ———, *On the positive mass, Penrose, and ZAS inequalities in general dimension*, arXiv:1101.2230 (2011).
- [Car71] B. Carter, *Axisymmetric black hole has only two degrees of freedom*, Phys. Rev. Lett. **26** (1971), 331–.
- [Car03] S. Carroll, *Spacetime and geometry: An introduction to general relativity*, Pearson, 2003.
- [CBG69] Y. Choquet-Bruhat and R. Geroch, *Global aspects of the Cauchy problem in general relativity*, Commun. Math. Phys. **14** (1969), 329–335.
- [CC08] P. Chruściel and J. Costa, *On uniqueness of stationary vacuum black holes*, Astérisque **321** (2008), 195–265.
- [dC93] M. do Carmo, *Riemannian geometry*, Birkhäuser, 1993.
- [FB55] Y. Fourès-Bruhat, *Théorème d’existence pour certains systèmes d’équations aux dérivées partielles non linéaires*, Acta Math. **88** (1955), 141–225.
- [GN14] L. Godinho and J. Natário, *An introduction to Riemannian geometry: With applications to mechanics and relativity*, Springer, 2014.
- [Haw66] S. Hawking, *The occurrence of singularities in cosmology*, Proc. Roy. Soc. Lon. A **294** (1966), 511–521.
- [Haw72] ———, *Black holes in general relativity*, Commun. Math. Phys. **25** (1972), 152–166.
- [Haw74] ———, *Black holes explosions?*, Nature **248** (1974), 30–31.
- [HE95] S. Hawking and G. Ellis, *The large scale structure of space-time*, Cambridge University Press, 1995.
- [HI01] G. Huisken and T. Ilmanen, *The inverse mean curvature flow and the Riemannian Penrose inequality*, J. Diff. Geom. **59** (2001), 353–437.
- [Isr67] W. Israel, *Event horizons in static vacuum space-times*, Phys. Rev. **164** (1967), 1776–1779.
- [KN96] S. Kobayashi and K. Nomizu, *Foundations of differential geometry*, vol. I and II, Wiley, 1996.

- [Lam10] M.-K. Lam, *The graphs cases of the Riemannian positive mass and Penrose inequalities in all dimensions*, [arXiv:1010.4256](#) (2010).
- [Mar09] M. Mars, *Present status of the Penrose inequality*, *Class. Quant. Grav.* **26** (2009), 193001.
- [MTW73] C. Misner, K. Thorne, and J. A. Wheeler, *Gravitation*, Freeman, 1973.
- [Mun00] J. Munkres, *Topology*, Prentice-Hall, 2000.
- [Nab88] G. Naber, *Spacetime and singularities – an introduction*, Cambridge University Press, 1988.
- [O’N83] B. O’Neill, *Semi-Riemannian geometry*, Academic Press, 1983.
- [O’N95] ———, *The geometry of Kerr black holes*, A K Peters, 1995.
- [Pen65] R. Penrose, *Gravitational collapse and space-time singularities*, *Phys. Rev. Lett.* **14** (1965), 57–59.
- [Pen87] ———, *Techniques of differential topology in relativity*, Society for Industrial and Applied Mathematics, 1987.
- [Poi07] E. Poisson, *A relativist’s toolkit: The mathematics of black-hole mechanics*, Cambridge University Press, 2007.
- [Rin09] H. Ringström, *The Cauchy problem in general relativity*, European Mathematical Society, 2009.
- [Rob75] D. Robinson, *Uniqueness of the Kerr black hole*, *Phys. Rev. Lett.* **34** (1975), 905–906.
- [SY81] R. Shoen and S.-T. Yau, *Proof of the positive mass theorem ii*, *Commun. Math. Phys.* **79** (1981), 231–260.
- [Tow97] P. Townsend, *Black holes*, [arXiv:gr-qc/9707012](#) (1997).
- [Wal84] R. Wald, *General relativity*, University of Chicago Press, 1984.