Lecture Notes on General Relativity
Columbia University

January 16, 2013

Contents

Introduction

General Relativity is the classical theory that describes the evolution of systems under the effect of gravity. Its history goes back to 1915 when Einstein postulated that the laws of gravity can be expressed as a system of equations, the so-called Einstein equations. In order to formulate his theory, Einstein had to reinterpret fundamental concepts of our experience (such as time, space, future, simultaneity, etc.) in a purely geometrical framework. The goal of this course is to highlight the geometric character of General Relativity and unveil the fascinating properties of black holes, one of the most celebrated predictions of mathematical physics.

The course will start with a self-contained introduction to special relativity and then proceed to the more general setting of Lorentzian manifolds. Next the Lagrangian formulation of the Einstein equations will be presented. We will formally define the notion of black holes and prove the incompleteness theorem of Penrose (also known as singularity theorem). The topology of general black holes will also be investigated. Finally, we will present explicit spacetime solutions of the Einstein equations which contain black hole regions, such as the Schwarzschild, and more generally, the Kerr solution.

Chapter 1

Special Relativity

In both past and modern viewpoints, the universe is considered to be a continuum composed of events, where each event can be thought of as a point in space at an instant of time. We will refer to this continuum as the spacetime. The geometric properties, and in particular the causal structure of spacetimes in Newtonian physics and in the theory of relativity greatly differ from each other and lead to radically different perspectives for the physical world and its laws.

We begin by listing the key assumptions about spacetime in Newtonian physics and then proceed by replacing these assumptions with the postulates of special relativity.

1.1 Newtonian Physics

Main assumptions

The primary assumptions in Newtonian physics are the following

1. There is an absolute notion of time. This implies the notion of simultaneity is also absolute.

2. The speed of light is finite and observer dependent.

3. Observers can travel arbitrarily fast (in particular faster than c).

From the above one can immediately infer the existence of a time coordinate $t \in \mathbb{R}$ such that all the events of constant time t compose a 3-dimensional Euclidean space. The spacetime is topologically equivalent to $\mathbb{R}^4$ and admits a universal coordinate system (t, x_1, x^2, x_3).

Causal structure

Given an event $\mathbf{p}$ occurring at time $t_{\mathbf{p}}$, the spacetime can be decomposed into the following sets:

- Future of $\mathbf{p}$: Set of all events for which $t > t_{\mathbf{p}}$.

- Present of $\mathbf{p}$: Set of all events for which $t = t_{\mathbf{p}}$.

- Past of $\mathbf{p}$: Set of all events for which $t < t_{\mathbf{p}}$.

More generally, we can define the future (past) of a set S to be the union of the futures (pasts) of all points of S.

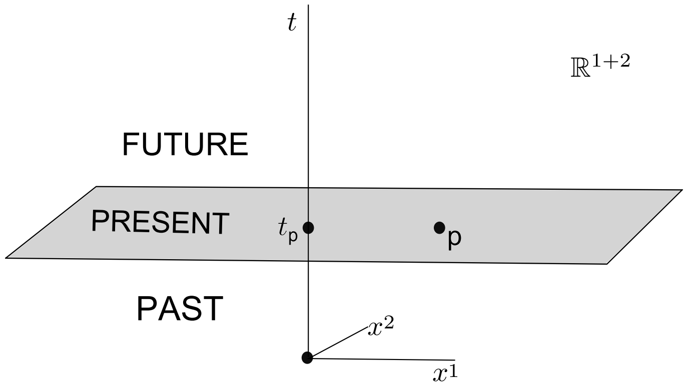


Figure 1.1: The Newtonian universe and its causal structure

From now on we consider geometric units with respect to which the light travels at speed $c = 1$ relative to observers at rest. If an observer at p emits a light beam in all directions of space, then the trajectory of this beam in spacetime will be a null cone with vertex at **p**. We can complete this cone by considering the trajectory of light beams that arrive at the event **p**.

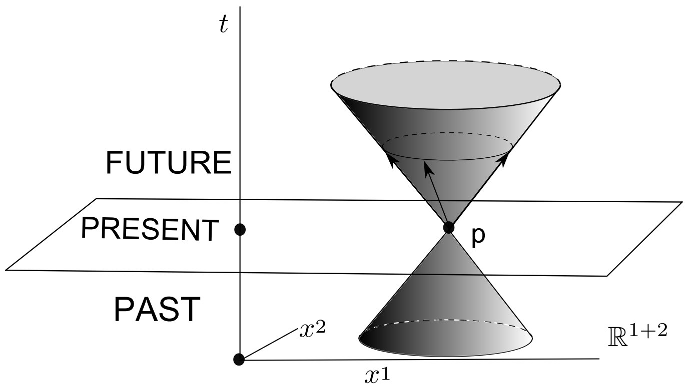


Figure 1.2: The Newtonian universe and light trajectories

It is important to emphasize that in Newtonian theory, in view of the existence of the absolute time t, one only works by projecting on the Euclidean space $\mathbb{R}^3$ and considering all quantities as functions of (the space and) time t.

Newton's theory gives a very accurate theory for objects moving at slow speeds in absence of strong gravitational fields. However, in several circumstances difficulties arise:

1. A more philosophical issue is that in Newtonian theory an observer is either at rest or in motion. But how could one determine if an observer O is (universally) at rest? Why can't a uniformly moving (relative to O) observer P be considered at rest since P is also not affected by any external influence?

2. For hundreds of years it has been known that in vacuum light propagates at a very high but constant speed, and no material has been observed to travel faster. If an observer P is moving at speed $c/5$ (relative to an observer O at rest) towards a light beam (which is moving at speed c relative to O), then the light would reach the observer P at speed $c + c/5$. However, astronomical observations of double stars should reveal such fast and slow light, but in fact the speeds turn out to be the same.

3. Light is the propagation of an electromagnetic disturbance and electromagnetic fields are governed by Maxwell's equations. However, these equations are not well-behaved in Newtonian theory; in particular, in this context, these laws are observer dependent and hence do not take the desired form of universal physical laws.

One of Einstein's contributions was his persistence that every physical law can be expressed independently of the choice of coordinates (we will return to this point later). It was this persistence along with his belief that Maxwell's equations are flawless that led to what is now known as special relativity.

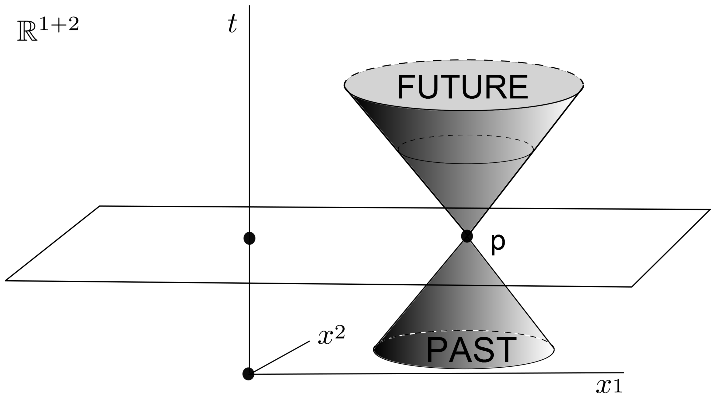
1.2 The Birth of Special Relativity

In 1905 Einstein published a paper titled "On the electrodynamics of moving bodies", where he described algebraic relations governing the motion of uniform observers so that Maxwell equations have the same form regardless of the observer's frame. In order to achieve his goal, Einstein had to assume the following

1. There is no absolute notion of time.

2. No observer or particle can travel faster than the speed of light c. The constant c should be considered as a physical law and hence does not depend on the observer who measures it.

The above immediately change the Newtonian perception for the spacetime, since under Einstein's assumptions the future (past) of an event $\mathbf{p}$ is confined to be the interior of the future (past) light cone with vertex at $\mathbf{p}$.



In 1908, Hermann Minkowski showed that Einstein's algebraic laws (and, in particular, the above picture) can be interpreted in a purely geometric way, by introducing a new kind of metric on $\mathbb{R}^4$, the so-called Minkowski metric.

1.3 The Minkowski Spacetime $\mathbb{R}^{3+1}$

Definition

A Minkowski metric g on the linear space $\mathbb{R}^4$ is a symmetric non-degenerate bilinear form with signature $(-,+,+,+)$. In other words, there is a basis $\{e_0, e_1, e_2, e_3\}$ such that

$$g(e_\alpha, e_\beta) = g_{\alpha\beta}, \;\; \alpha, \beta \in \{0, 1, 2, 3\},$$

where the matrix $g_{\alpha\beta}$ is given by

$$g = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Given such a frame (which for obvious reason will be called orthonormal), one can readily construct a coordinate system (t, x^1, x^2, x^3) of $\mathbb{R}^4$ such that at each point we have

$$e_0 = \partial_t, \quad e_i = \partial_{x^i}, \; i = 1, 2, 3.$$

Note that from now on in order to emphasize the signature of the metric we will denote the Minkowski spacetime by $\mathbb{R}^{3+1}$. With respect to the above coordinate system, the metric g can be expressed as a (0,2) tensor as follows:

$$g = -dt^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2. \tag{1.1}$$

Note that (for an arbitrary pseudo-Riemannian metric) one can still introduce a Levi–Civita connection and therefore define the notion the associated Christoffel symbols and geodesic curves and that of the Riemann, Ricci and scalar curvature. One can also define the volume form ϵ such that if $X_\alpha, \alpha = 0, 1, 2, 3$, is an orthonormal frame then $\epsilon(X_0, X_1, X_2, X_3) = \pm 1$. In the Minkowski case, the curvatures are all zero and the geodesics are lines with respect to the coordinate system (t, x^1, x^2, x^3).

1.3.1 Causality Theory

The fundamental new aspect of this metric is that it is not positive-definite. A vector $X \in \mathbb{R}^{3+1}$ is defined to be:

1. *spacelike*, if $g(X, X) > 0$,
2. *null*, if $g(X, X) = 0$,
3. *timelike*, if $g(X, X) < 0$.

If X is either timelike or null, then it is called *causal.* If $X = (t, x^1, x^2, x^3)$ is null vector at $\mathbf{p}$, then

$$t^2 = (x^1)^2 + (x^2)^2 + (x^3)^2$$

and hence X lies on cone with vertex at $\mathbf{p}$. In other words, all null vectors at $\mathbf{p}$ span a double cone, known as the *double null cone.* We denote by $\mathcal{S}_{\mathbf{p}}$,

$$\mathcal{S}_{\mathbf{p}} = \left\{ X \in T_{\mathbf{p}}\mathbb{R}^{3+1} \;:\; g(X, X) > 0 \right\},$$

the set of spacelike vectors at $\mathbf{p}$, by $\mathcal{I}_{\mathbf{p}}$,

$$\mathcal{I}_{\mathbf{p}} = \left\{ X \in T_{\mathbf{p}}\mathbb{R}^{3+1} \;:\; g(X,X) < 0 \right\},$$

the set of timelike vectors at $\mathbf{p}$, and by $\mathcal{N}_{\mathbf{p}}$,

$$\mathcal{N}_{\mathbf{p}} = \left\{ X \in T_{\mathbf{p}}\mathbb{R}^{3+1} \;:\; g(X,X) = 0 \right\},$$

the set of null vectors at $\mathbf{p}$. The set $\mathcal{S}_{\mathbf{p}}$ is open (and connected if $n > 1$). Note that the set $\mathcal{I}_{\mathbf{p}}$ is the interior of the solid double cone enclosed by $\mathcal{N}_{\mathbf{p}}$. Hence, $\mathcal{I}_{\mathbf{p}}$ is an open set consisting of two components which we may denote by $\mathcal{I}^+_{\mathbf{p}}$ and $\mathcal{I}^-_{\mathbf{p}}$. Then we can also decompose $\mathcal{N}_{\mathbf{p}} = \mathcal{N}^+_{\mathbf{p}} \cup \mathcal{N}^-_{\mathbf{p}}$, where

$$\mathcal{N}^+_{\mathbf{p}} = \partial\mathcal{I}^+_{\mathbf{p}}, \qquad \mathcal{N}^-_{\mathbf{p}} = \partial\mathcal{I}^-_{\mathbf{p}}.$$

The following questions arise: How is $\mathcal{N}^+_{\mathbf{p}}$ (or $\mathcal{I}^+_{\mathbf{p}}$) related to the future and past of p? How can one even discriminate between the future and past? To answer this we need to introduce the notion of time-orientability. A *time-orientation* of $(\mathbb{R}^{3+1}, g)$ is a continuous choice of a positive component $\mathcal{I}^+_{\mathbf{p}}$ at each $\mathbf{p} \in \mathbb{R}^{3+1}$. Then, we call $\mathcal{I}^+_{\mathbf{p}}$ (resp. $\mathcal{N}^+_{\mathbf{p}}$) the set of *future-directed* timelike (resp. null) vectors at p. Similarly we define the past-directed causal vectors. We also define:

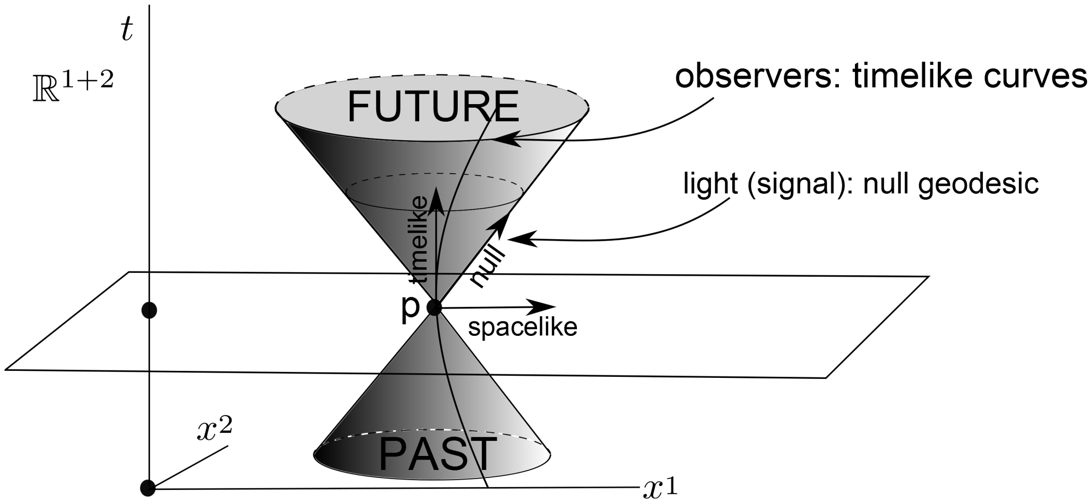
- The *causal future* $\mathcal{J}^+(\mathbf{p})$ of $\mathbf{p}$ by $\mathcal{J}^+(\mathbf{p}) = \mathcal{I}^+_{\mathbf{p}} \cup \mathcal{N}^+_{\mathbf{p}}$.
- The *chronological future* of $\mathbf{p}$ to simply be $\mathcal{I}^+_{\mathbf{p}}$.

We also define the causal future $\mathcal{J}^+(S)$ of a set S by

$$\mathcal{J}^+(S) = \bigcup_{\mathbf{p}\in S} \mathcal{J}^+_{\mathbf{p}}.$$

Causality, observers and particles, proper time

We can readily extend the previous causal characterizations for curves. In particular, a curve $\alpha : I \to \mathbb{R}^{3+1}$ is called future-directed timelike if $\dot{\alpha}(t)$ is a future-directed timelike vector at $\alpha(t)$ for all $t \in I$. Note that the worldline of an observer is represented by a timelike curve. If the observer is inertial, then he/she moves on a timelike geodesic. On the other hand, photons move on null geodesics (and hence information propagates along null geodesics).



We, therefore, see that the Minkowski metric provides a very elegant way to put the assumptions of Section 1.2 in a geometric context.

The *proper time* τ of an observer O is defined to be the parametrization of its worldline α such that $g(\dot{\alpha}(\tau), \dot{\alpha}(\tau)) = -1$. Note that no proper time can be defined for photons since for all parametrizations we have $g(\dot{\alpha}(\tau), \dot{\alpha}(\tau)) = 0$. However, in this case, it is usually helpful to consider affine parametrizations τ with respect to which $\nabla_{\dot{\alpha}}\dot{\alpha} = 0$.

Hypersurfaces

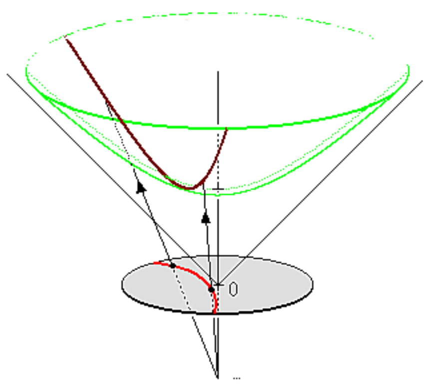
Finally, we have the following categories for hypersurfaces:

1. A hypesurface $\mathcal{H}$ is called spacelike, if the normal N_x at each point $x \in \mathcal{H}$ is timelike. In this case, $g\big|_{T_x\mathcal{H}}$ is positive-definite (i.e. $\mathcal{H}$ is a Riemannian manifold).
2. A hypesurface $\mathcal{H}$ is called null, if the normal N_x at each point $x \in \mathcal{H}$ is null. In this case, $g\big|_{T_x\mathcal{H}}$ is degenerate.
3. A hypesurface $\mathcal{H}$ is called timelike, if the normal N_x at each point $x \in \mathcal{H}$ is spacelike. In this case, $g\big|_{T_x\mathcal{H}}$ has signature $(-, +, +)$.

Examples of spacelike hypersurfaces

1. The hypersurfaces
$$H_\tau = \{t = \tau\}$$
are spacelike, since their unit normal is the timelike vector field ∂_0. Then, $\left(H_\tau, g\big|_{T_x\mathcal{H}_\tau}\right)$ is isometric to the 3-dimensional Euclidean space $\mathbb{E}^3$.

2. The hypersurface
$$H^3 = \{X : \; g(X,X) = -1 \text{ and } X \text{ future-directed } \}$$
is a spacelike hypersurface. Indeed, one can easily verify that X is the normal to H^3 at the endpoint of X. Then, $\left(H^3, g\big|_{T_x\mathcal{H}^3}\right)$ is isometric to the 3-dimensional hyperbolic space. The following figure depicts the radial projection which is an isometry from H^3 to the disk model.



Note that the hyperboloid consists of all points in spacetime where an observer located at $(0, 0, 0, 0)$ can be after proper time 1.

We remark that the Cauchy–Schwarz inequality is reversed for timelike vectors. If X, Y are two future-directed timelike vector fields then $g(X, Y) < 0$ and

$$-g(X, Y) \geq \sqrt{g(X, X) \cdot g(Y, Y)}.$$

Hence, there exists a real number ϕ such that

$$\cosh(\phi) = \frac{-g(X, Y)}{\sqrt{g(X, X) \cdot g(Y, Y)}}.$$

Then ϕ is called the hyperbolic angle of X and Y.

Examples of timelike hypersurfaces

1. The hypersurfaces
$$\mathcal{T}_\tau = \{x_1 = \tau\}$$
are timelike, since their normal is the spacelike vector field ∂_{x_1}. Then, $\left(\mathcal{T}_\tau, g\Big|_{T_x \mathcal{T}_\tau}\right)$ is isometric to the 3-dimensional Minkowski space $\mathbb{R}^{2+1}$.

2. The hypersurface
$$H^3_+ = \{X : \; g(X, X) = 1 \text{ and } X \text{ future-directed } \}$$
is a timelike hypersurface. Indeed, one can easily verify that X is the normal to H^3_+ at the endpoint of X.

Examples of null hypersurfaces

1. Let $n = (n_0, n_1, n_2, n_3)$ be a null vector. The planes given by the equation

$$P_n = \left\{(t, x^1, x^2, x^3) : \ n_0 t = n_1 x^1 + n_2 x^2 + n_3 x^3\right\}$$

are null hypersurfaces, since their normal is the null vector n.

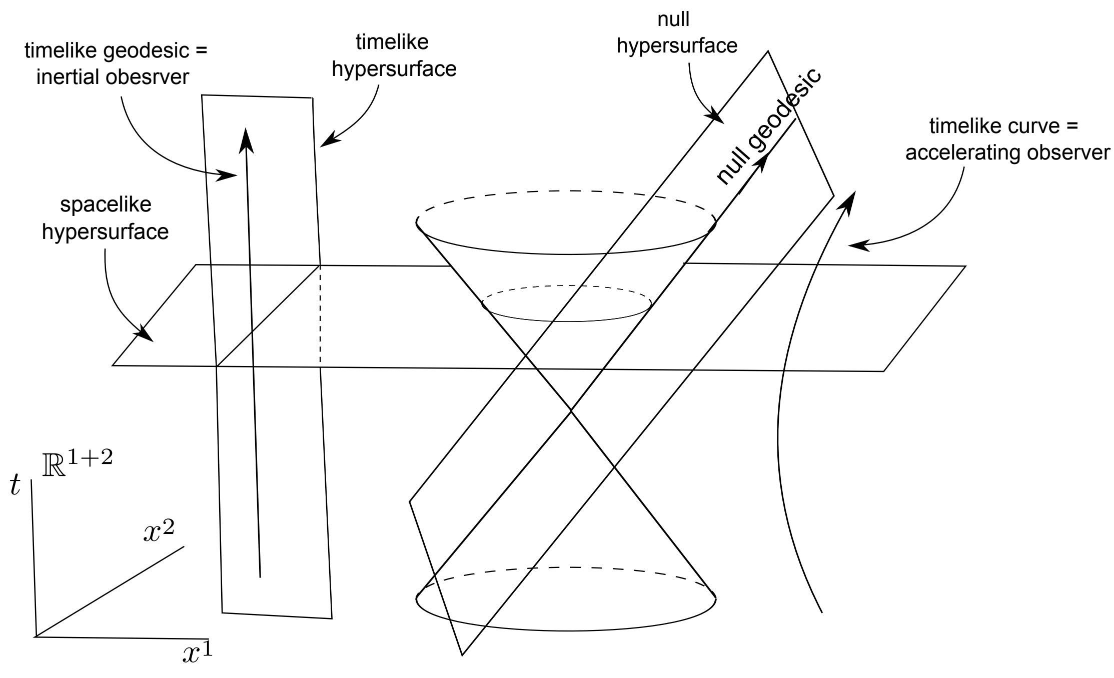
2. The (null) cone

$$C = \left\{(t, x^1, x^2, x^3) : \ t = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}\right\}$$

is also a null hypersurface. Its tangent plane at the endpoint of n is the plane P_n and hence its normal is the null vector n. Note that $C = \mathcal{N}^+(O)$, where O is the origin. Note also that $\mathcal{N}^-(O)$ is given by

$$\underline{C} = \left\{(t, x^1, x^2, x^3) : \ t = -\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}\right\}.$$

The above can be summarized in the following figure:



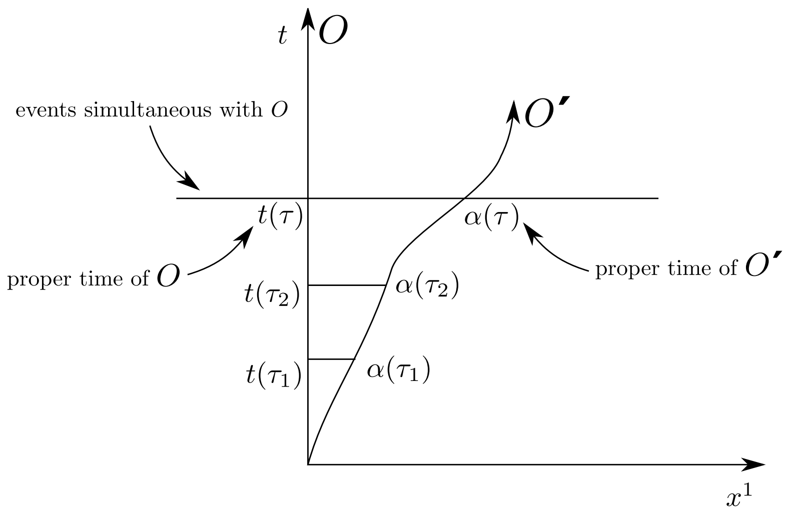
1.3.2 Inertial Observers, Frames of Reference and Isometies

Inertial Frames

Let O be an inertial observer moving on a timelike geodesic α. Let t be the proper time of O and (x^1, x^2, x^3) be a Euclidean coordinate system of the (spacelike) plane orthogonal to α at $\alpha(0)$. We will refer to the frame (t, x^1, x^2, x^3) as the frame associated to the inertial observer O.

Relativity of Time

Let O' be another (inertial or not) observer moving on the timelike curve α' which can be expressed as $\alpha'(\tau) = (t(\tau), x^1(\tau), x^2(\tau), x^3(\tau))$ with respect to the frame associated to O. Note that since α' is future-directed curve, the function $t \mapsto t(\tau)$ is invertible and hence we can write $\tau = \tau(t)$ giving us the following parametrization $\alpha'(t) = (t, x^1(\tau(t)), x^2(\tau(t)), x^3(\tau(t)))$. Note that the observer O at its proper time t sees that O' is located at the point $\alpha'_O(t) = (x^1(\tau(t)), x^2(\tau(t)), x^3(\tau(t)))$. Hence, the speed of O' relative to O is $\left|\frac{d}{dt}\alpha'_O(t)\right|$. As we shall see, we do not always have that $t = \tau$, in other words the time is relative to each observer.



Let O' be an inertial observer passing through the origin $(0,0,0,0)$ with respect to the frame of O and also moving in the x^1-direction at speed v with respect to O. Then, O' moves on the curve $\alpha'(t) = (t, vt, 0, 0)$; however, t is **not** the proper time of O' since $g(\dot\alpha(t), \dot\alpha(t)) = -1 + v^2$. However, if we consider the following parametrization $\alpha'(\tau) = (\gamma\tau, \gamma\tau v, 0, 0)$, where

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

then τ is the proper time for O' since $g(\dot\alpha'(\tau), \dot\alpha'(\tau)) = -1$. Note that $\gamma = cosh(\phi)$, where ϕ is the hyperbolic angle of $\alpha(t=1) = (1,0,0,0)$ and $\alpha'(\tau = 1)$.

Time dilation: If τ is the proper time of O', then from the above representation we obtain that $t = \gamma\tau$ along α'. Note that $\gamma > 1$ and so $t > \tau$, which implies that the proper time for the moving (with respect to O) observer O' runs slower compared to the time t that O measures for O'.

Isometries: Lorentz transformations

One of the most striking properties of Minkowski spacetime is that there exists an isometry F which maps O to O'. Relative to the frame associated to O this map takes the form:

$$F(t, x^1, x^2, x^3) = (\gamma \cdot (t + vx^1), \gamma \cdot (x^1 + vt), x^2, x^3).$$

Then

$$F[O(\lambda)] = F(\lambda, 0, 0, 0) = (\gamma\lambda, \gamma\lambda v, 0, 0) = \alpha'(\lambda) = O'(\lambda),$$

where $O(\lambda), O'(\lambda)$ denotes the position of the observers O, O', respectively, at proper time λ. Clearly, the map F, being an isometry, leaves the proper time of observers invariant.

Note that although there is no absolute notion of time (and space consisting of simultaneous events), null cones are absolute geometric constructions that do not depend on observers (in particular, the isometry F leaves the null cones invariant).

The isometry F is known as Lorentz transformation, or Lorentz boost in the x^1-direction or hyperbolic rotation. The latter name arises from the fact that such an isometry can be represented by a matrix whose form is similar to a Euclidean rotation but with trigonometric functions and angles replaced by hyperbolic trigonometric functions and angles. Note that the isometries $F = F_v$ (with $v^2 \leq 1$) correspond to the flow of the Killing field

$$H_{x^1} = t\partial_{x^1} + x^1\partial_t.$$

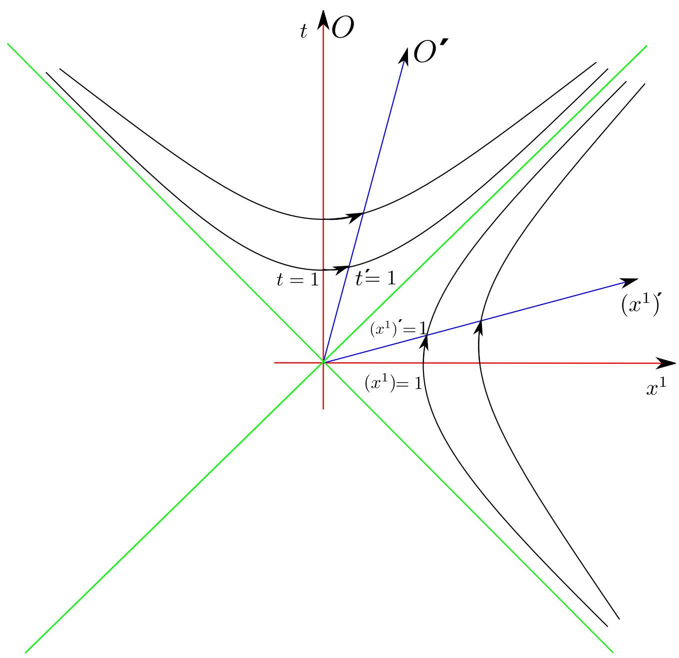
Clearly, one can define boosts in any direction.

Relativity of simultaneity

Having understood the relativity of time, we next proceed by investigating the relativity of space. The points simultaneous with observer O at the origin are all points such that $t = 0$. Then,

$$F[\{t=0\}] = F(0, x^1, x^2, x^3) = (\gamma x^1 v, \gamma x^1, x^2, x^3).$$

We extend t' such that $\{t' = 0\} = F[\{t = 0\}]$ and define the coordinates $\big((x^1)', (x^2)', (x^3)'\big)$ on the hypersurface $t' = 0$ such that if $\mathbf{p} \in \{t' = 0\}$, then $\mathbf{p} = \big((x^1)', (x^2)', (x^3)'\big)$ if $F^{-1}(\mathbf{p}) = (0, x^1, x^2, x^3)$. Note that the plane $t' = 0$ consists of all points simultaneous with O' at the origin and that one can define a global coordinate system $\big(t', (x^1)', (x^2)', (x^3)'\big)$. This is the reference frame associated to O'. In view of the fact that F is an isometry, the metric takes the form (1.1) with respect to the system $\big(t', (x^1)', (x^2)', (x^3)'\big)$.



Length Contraction: First note that distance between events is meaningful only for observers who consider the events to be simultaneous. At proper time $\tau = 1$ of O', observer O' measures that his distance from observer O is v. However, when $\tau = 1$, the measurements of O are such that O' is located at the point $(\gamma, \gamma v, 0, 0)$ with respect to the frame of O. Hence, the distance (after $\tau = 1$) that O measures between O and O' is $\gamma v > v$. In other words, O' measures shorter distances than O does. This phenomenon is called length contraction and is intimately connected to time dilation.

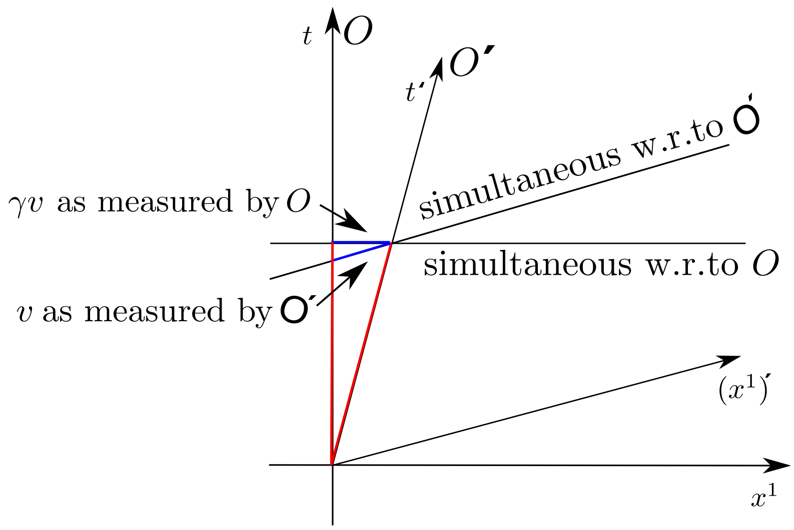


Figure 1.3: The difference of the hyperbolic lengths of the blue (resp. red) segments represents length contraction (resp. time dilation).

1.3.3 General and Special Covariance

The general covariance principle allows us to put physics in a geometric framework:

- **General covariance principle:** A*ll physical laws are independent of the choice of a particular coordinate system. In other words, the equations expressing physical laws must be written in terms of tensors.*

Since the only tensor that we have in special relativity is the metric g, then we can in fact assume the following:

- **Strong general covariance principle:** *All physical laws can be expressed in terms of the metric g and its tensorial expressions.*

Hence, physical laws change covariantly (i.e. tensorially) under change of coordinates (i.e. under diffeomorphisms). If we apply this principle for Minkowski spacetime and also restrict to its isometries, then we obtain the following:

- **Special covariance:** *The physical laws take exactly the same form when expressed in terms of the reference frames associated to inertial observers.*

1.3.4 Relativistic Mechanics

1. Time dilation

We will here generalize the result of the previous section on time dilation. Let O be an inertial observer and O' another (inertial or accelerating) observer. Let τ be the proper time of O' such that his/her trajectory is $\alpha'(\tau) = (t(\tau)), x^1(\tau), x^2, (\tau), x^3(\tau))$. As before, we can write $\tau = \tau(t)$ and hence $\alpha'(\tau(t)) = (t, x^1(\tau(t), x^2(\tau(t), x^3(\tau(t))$. If v is the speed of O' relative to O, the fact that $g(\dot{\alpha}'(\tau), \dot{\alpha}'(\tau)) = -1$ implies that

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1-v^2}} \geq 1 \Rightarrow t(\tau) \geq \tau.$$

2. Absoluteness of speed of light

If a particle moves of a null curve $\alpha'(\tau) = (t(\tau)), x^1(\tau), x^2, (\tau), x^3(\tau))$ then the speed of the particle with respect to an inertial observer O is

$$v = \frac{d\alpha'_O}{d\tau}\frac{d\tau}{dt} = 1,$$

where the last equation follows from $g(\dot{\alpha}'(\tau), \dot{\alpha}'(\tau)) = 0$.

3. Energy-momentum

Let $\alpha(\tau)$ represent the trajectory of a particle p with mass m. Then we have the following definitions:

- The *4-velocity* of p is the vector $U = \dot{\alpha}(\tau) = \frac{d\alpha}{d\tau}$.
- The *energy-momentum* of p is the vector $P = mU$.

Let's now see how an inertial observer O measures the above quantities. If the particle p moves at speed v relative to O, then by considering the frame associated to O, we obtain:

$$P = m\frac{d\alpha}{dt}\frac{dt}{d\tau} = \frac{m}{\sqrt{1-v^2}}\frac{d\alpha}{dt}.$$

1. The spatial component of the energy-momentum vector P is

$$P_O = \frac{m}{\sqrt{1-v^2}}\frac{d\alpha_O}{dt},$$

which is called the *momentum* of p as measured by O. (Compare this definition with the Newtonian definition in case $v \sim 0$.)

2. The temporal component of the energy-momentum vector P is

$$E_O = \frac{m}{\sqrt{1-v^2}} = m + \frac{1}{2}mv^2 + O(v^4),$$

which is called the *total energy* of p as measured by O. (Note, in particular, that E_O contains the kinetic energy $\frac{1}{2}mv^2$). Hence, the mass m is seen to be energy. If $v = 0$, then energy of p as measured by a co-observer is $E = m$, and by converting in conditional units, we obtain Einstein's famous equation

$$E = mc^2.$$

1.4 Conformal Structure

One is often interested in investigating the properties of isolated systems. In such situation one should only consider the local system and hence ignore the influence of matter at far distances. In other words, the asymptotic structure of the spacetime describing the geometry of an isolated system should like the asymptotical structure of Minkowski (recall that Minkowski spacetime represents the geometry a vacuum static highly symmetric topologically trivial universe). The goal of this section is to describe the global and asymptotic causal structure of Minkowski spacetime

One could start by considering the following foliation of Minkowski:

$$\mathbb{R}^{3+1} = \bigcup_{\tau\in\mathbb{R}} \mathcal{H}_\tau,$$

where the spacelike hypersurfaces $\mathcal{H}_\tau = \{t = \tau\}$ are as defined in Section 1.3. Note, however, this foliation does not capture the properties of null geodesics whose importance is manifest from the fact that signals travel along such curves. Indeed, an observer (like ourselves on earth) located far away from an isolated system under investigation must understand the asymptotic behavior of null geodesics in order to be able to measure radiation and other information sent from this system. For these reason, we will consider a foliation of Minkowski spacetime which captures the geometry of null geodesics emanating from points of a timelike geodesic. This is the so-called double null foliation.

As a final remark, note that since we want to understand the asymptotic structure, it will be convenient to apply conformal transformations on Minkowski spacetime in order to bring points at 'infinity' to finite distance. This procedure will allows us to reveal the structure of infinity. Note that conformal transformations preserve the causal structure, since they send timelike curves to timelike curves, spacelike curves to spacelike curves and null curves to null curves. In fact, conformal transformations send null geodesics to null geodesics (see Section 2.2).

1.4.1 The Double Null Foliation

Let us consider the timelike geodesic $\alpha(t) = (t, 0, 0, 0)$ where the coordinates are taken with respect to an inertial coordinate system (t, x^1, x^2, x^3). Recall that the future-directed null cone C_O with vertex at $O = \alpha(0)$ is given by

$$C_0 = \left\{ (t, x^1, x^2, x^3) : \; t = \sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2} \; \geq 0 \right\},$$

whereas the past-directed null cone $\underline{C}_O$ with vertex at $O = \alpha(0)$ is given by

$$\underline{C}_0 = \left\{ (t, x^1, x^2, x^3) : \; t = -\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2} \; \leq 0 \right\}.$$

In order to simplify the above expressions and capture the spherical symmetry of the null cones, it is convenient to introduce spherical coordinates (r, θ, ϕ) for the Euclidean hypersurfaces $\mathcal{H}_\tau$ such that $r = 0$ corresponds to the curve α. Then, in (t, r, θ, ϕ) coordinates, the Minkowski metric takes the form

$$g = -dt^2 + dr^2 + r^2 \cdot g_{\mathbb{S}^2(\theta,\phi)},$$

where $g_{\mathbb{S}^2}(\theta, \phi) = d\theta^2 + (\sin\theta)^2\, d\phi^2$ is the standard metric on the unit sphere. Then the future-directed null cone C_τ with vertex at $\alpha(\tau)$ is given by

$$C_\tau = \{(t, r, \theta, \phi) : \; t - r = \tau, \quad \tau \in \mathbb{R}\},$$

whereas the past-directed null cone $\underline{C}_\tau$ with vertex at $\alpha(\tau)$ is given by

$$\underline{C}_\tau = \{(t, r, \theta, \phi) : \; t + r = \tau, \quad \tau \in \mathbb{R}\}.$$

The above suggest that it is very convenient to convert to null coordinates (u, v, θ, ϕ) defined such that

$$\begin{aligned} u &= t - r, \\ v &= t + r. \end{aligned}$$

Note also that

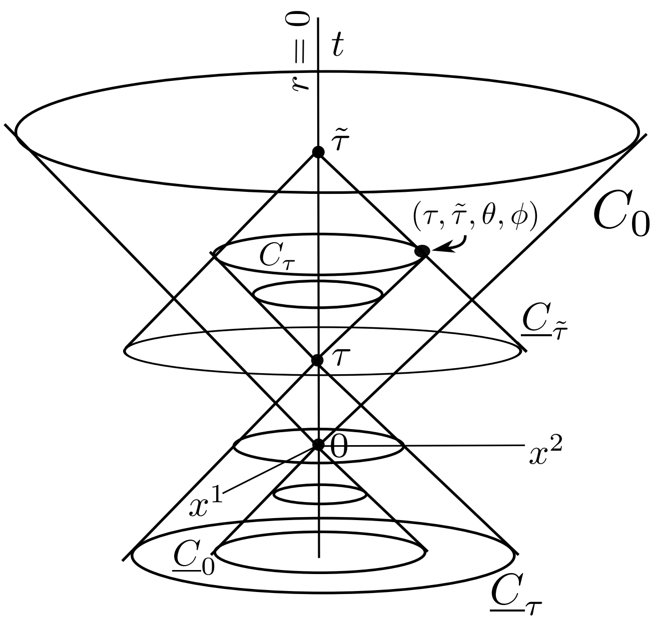
$$v \geq u \tag{1.2}$$

and $v = u$ if and only if $r = 0$. The metric with respect to null coordinates (u, v, θ, ϕ) takes the form

$$g = -dudv + \frac{1}{4}(u - v)^2 \cdot g_{\mathbb{S}^2(\theta,\phi)}$$

and the double null folation is given by the equations

$$\begin{aligned} C_\tau &= \{(u, v, \theta, \phi) : \; u = \tau, \quad \tau \in \mathbb{R}\}, \\ \underline{C}_\tau &= \{(u, v, \theta, \phi) : \; v = \tau, \quad \tau \in \mathbb{R}\}. \end{aligned}$$



Note that ∂_v (resp. ∂_u) is tangential to the null geodesics of the null cones C_τ (resp. $\underline{C}_\tau$). For a generalization of the double null foliation see Section 4.1.

1.4.2 The Penrose Diagram

The aim of the section is to describe the asymptotic structure of Minkowski space. In particular, we want to draw a "bounded" diagram whose boundary represents infinity and somehow respects the causal structure of Minkowski.

Clearly, $v \to +\infty$ along the null cones C_τ and similarly $u \to +\infty$ along the null cones $\underline{C}_\tau$. In order to bring the endpoint of null geodesics in finite distance, we consider the following change of coordinates:

$$
\begin{aligned}
\tan p &= v, \\
\tan q &= u,
\end{aligned}
\tag{1.3}
$$

with $p, q \in \left(\frac{\pi}{2}, \frac{\pi}{2}\right)$ and $p \geq q$. Then in (p, q, θ, ϕ) coordinates the metric takes the form

$$
g = \frac{1}{(\cos p)^2 \cdot (\cos q)^2} \left[-dp\, dq + \frac{1}{4} \sin^2(p - q) \cdot g_{\mathbb{S}^2(\theta, \phi)} \right].
$$

As expected, a consequence of the boundedness of the range of p, q is that the factor $\frac{1}{(\cos p)^2 \cdot (\cos q)^2}$ blows up as $p \to \pm\frac{\pi}{2}$ or $q \to \pm\frac{\pi}{2}$. In order to overcome this degeneracy, we consider the metric $\tilde{g}$ which in (p, q, θ, ϕ) takes the form

$$
\tilde{g} = -dp\, dq + \frac{1}{4} \sin^2(p - q) \cdot g_{\mathbb{S}^2(\theta, \phi)}. \tag{1.4}
$$

Clearly the metric $\tilde{g}$ is conformal to g. Note that $\nabla p = -\partial_q$, $\nabla q = -\partial_p$, where the ∇ is considered with respect to $\tilde{g}$, and therefore, the hypersurfaces

$$
\begin{aligned}
\widetilde{C}_\tau &= \{(p, q, \theta, \phi) : \ q = \tau, \quad \tau \in \mathbb{R}\}, \\
\widetilde{\underline{C}}_\tau &= \{(p, q, \theta, \phi) : \ p = \tau, \quad \tau \in \mathbb{R}\}.
\end{aligned}
$$

are null (with respect to $\tilde{g}$). Hence, if we suppress one angular direction, we can globally depict the manifold $(\widetilde{\mathcal{M}}, \tilde{g})$ covered by the coordinates (p, q, θ, ϕ) as follows:

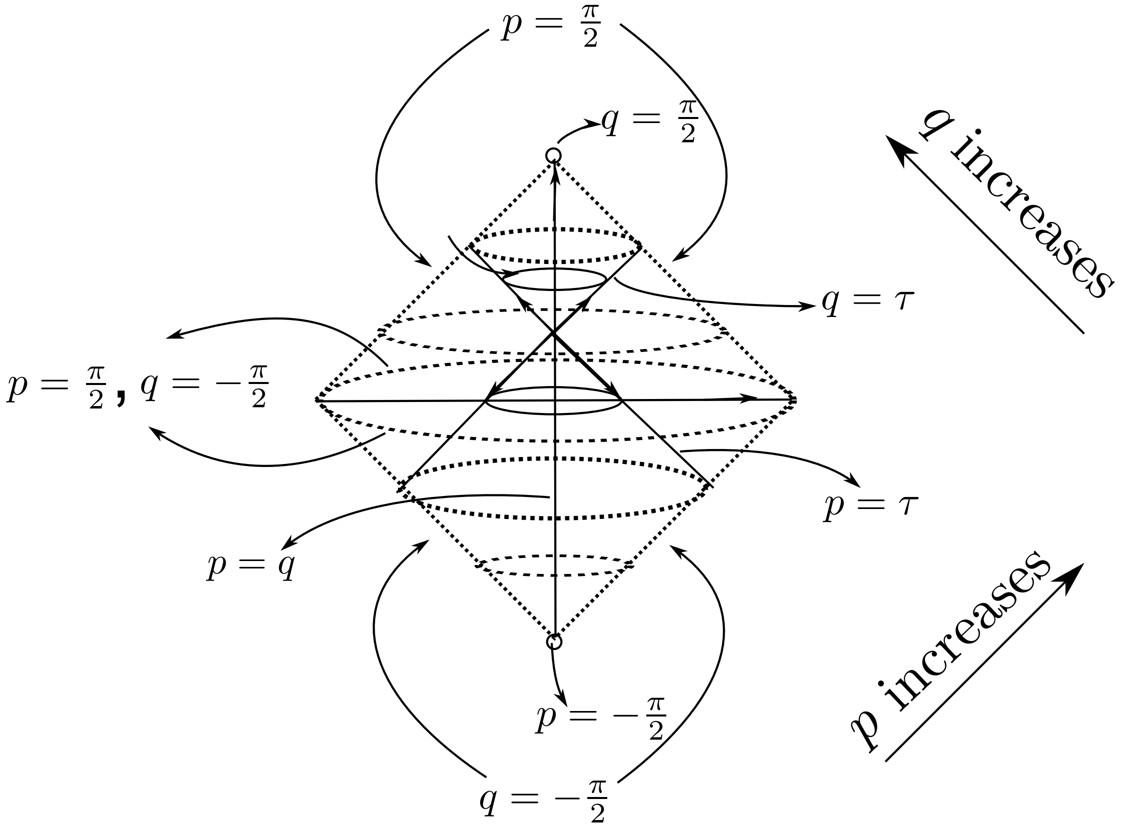


Figure 1.4: The manifold $(\widetilde{\mathcal{M}}, \tilde{g})$, which is conformal to Minkowski $\mathbb{R}^{3+1}$.

We define:

- *Future null infinity* $\mathcal{I}^+$ to be the endpoints of all future-directed null geodesics along which $r \to +\infty$.
- *Future timelike infinity* i^+ to be the endpoints of all future-directed timelike geodesics.
- *Spacelike infinity* i^0 to be the endpoint of all spacelike geodesics. This is in fact a point, and not a sphere, which can be thought of as the point at infinity of the one-point compactification of, say, the spacelike hypersurface $t = 0$.

Similarly, we define the past null infinity $\mathcal{I}^-$ and past timelike infinity i^-. Note that if we want to study an isolated system (located say at $r = 0$), then we should think of ourselves as moving along the future null infinity $\mathcal{I}^+$ and receiving the radiation/information sent from the system.

The boundary of the Figure 1.4 nicely depicts the above asymptotic structures of Minkowski space $\mathbb{R}^{3+1}$.

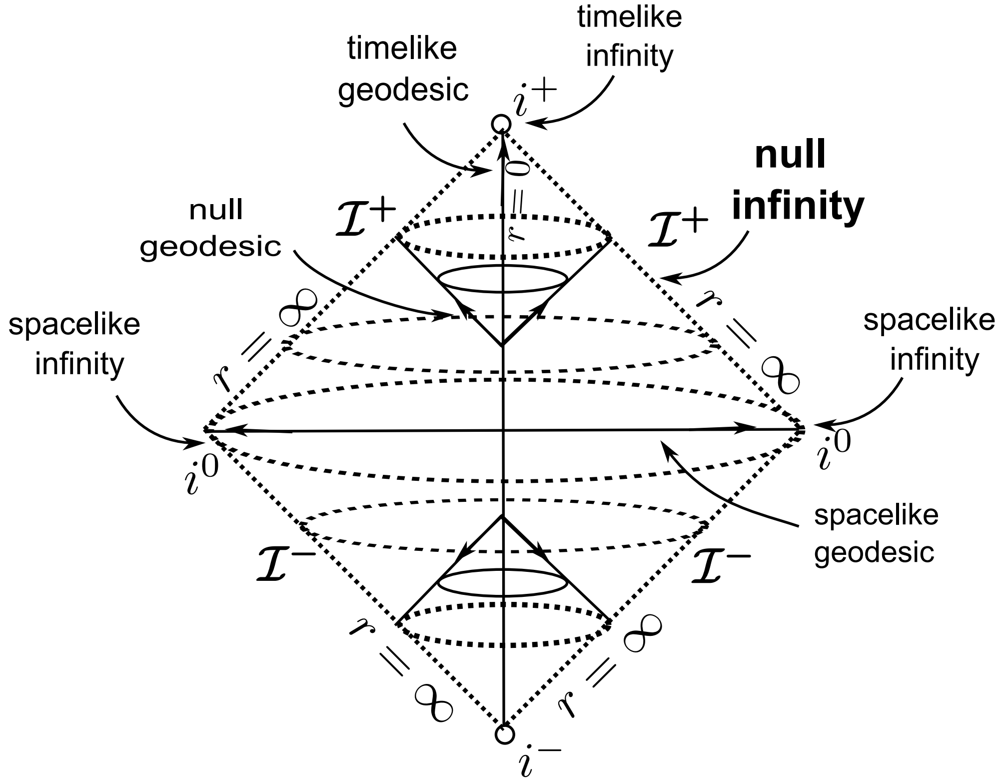


Figure 1.5: Identification of the boundary of the conformal diagram of Minkowski

Note that if a curve $\alpha(\tau) = (t(\tau), r(\tau), \theta(\tau), \phi(\tau))$ is such that as $\tau \to +\infty$

- $t \to +\infty$ and $r < \infty$, then α approaches i^+.
- $t \sim r \to +\infty$, then α approaches $\mathcal{I}^+$.
- $|t| < \infty$ and $r \to +\infty$, then α approaches i^0.
- $t \sim -r \to -\infty$, then α approaches $\mathcal{I}^-$.
- $t \to -\infty$ and $r < \infty$, then α approaches i^-.

Remark 1.4.1. *Although the above figures for $(\widetilde{\mathcal{M}}, \tilde{g})$ are very intuitive as been thought embedded in the Minkowski space $\mathbb{R}^{3+1}$, they are not rigorous since the metric $\tilde{g}$ is not isometric to the Minkowski metric. Indeed, the factor $\sin^2(p-q)$ highly distorts the geometry of the spheres $S_{\tau,\tau'} = \widetilde{C}_\tau \cap \underline{\widetilde{C}}_{\tau'}$ compared to the Minkowskian case. Not only this factor vanishes for the points where $p = q$ (which correspond to the points where $u = v$ and hence $r = 0$) but also degenerates as $p \to \frac{\pi}{2}$ and $q \to -\frac{\pi}{2}$, and hence the point denoted by i^0 above collapses to a point. In fact, the metric $\tilde{g}$ is locally isometric to the Einstein static metric which is the natural Lorentzian metric on $\mathbb{R} \times \mathbb{S}^3$. The latter can be thought of as the metric induced on the cylinder $\mathbb{R} \times \mathbb{S}^3$ embedded in the Minkowski space $\mathbb{R}^{4+1}$.*

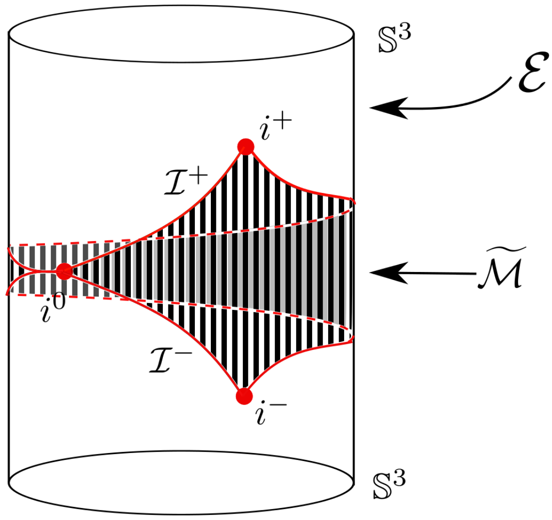


Figure 1.6: The red segments correspond to the boundary of the conformally compactified Minkowski spacetime in the Einstein static universe $\mathcal{E} = \mathbb{R} \times \mathbb{S}^3$. The points i^+, i^0, i^- are also points in $\mathcal{E}$. Every other point in $\widetilde{\mathcal{M}}$, apart from those that lie on the line connecting i^+ and i^-, represent a two-sphere in $\mathcal{E}$.

We can proceed further by suppressing all angular directions. Formally speaking, we consider the quotient $\widetilde{\mathcal{M}}/SO(3)$. Then, by (1.4), the metric on the quotient is simply $g_{\text{quot}} = -dp\,dq$, which coincides with the Minkowski spacetime $\mathbb{R}^{1+1}$. Hence, if we consider a planar section of Figure 1.5, then the resulting bounded 2-dimensional domain is embedded in the Minkowski spacetime $\mathbb{R}^{1+1}$:

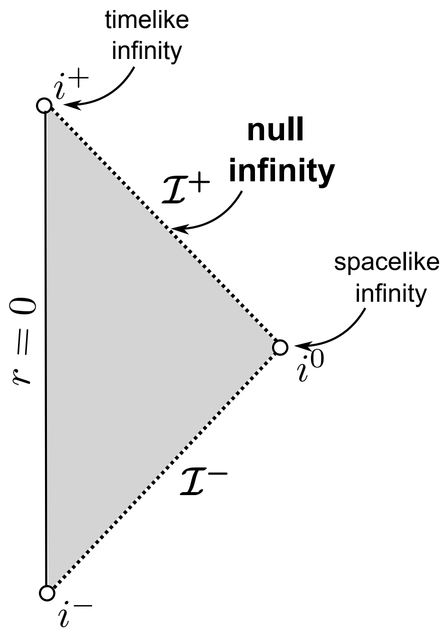


Figure 1.7: The Penrose diagram of Minkowski spacetime $\mathbb{R}^{3+1}$.

This diagram is called the Penrose diagram of Minkowski spacetime $\mathbb{R}^3$. All cones are collapsed to lines. Using the above diagram, one can read off the causal structure of the spacetime as is shown below:

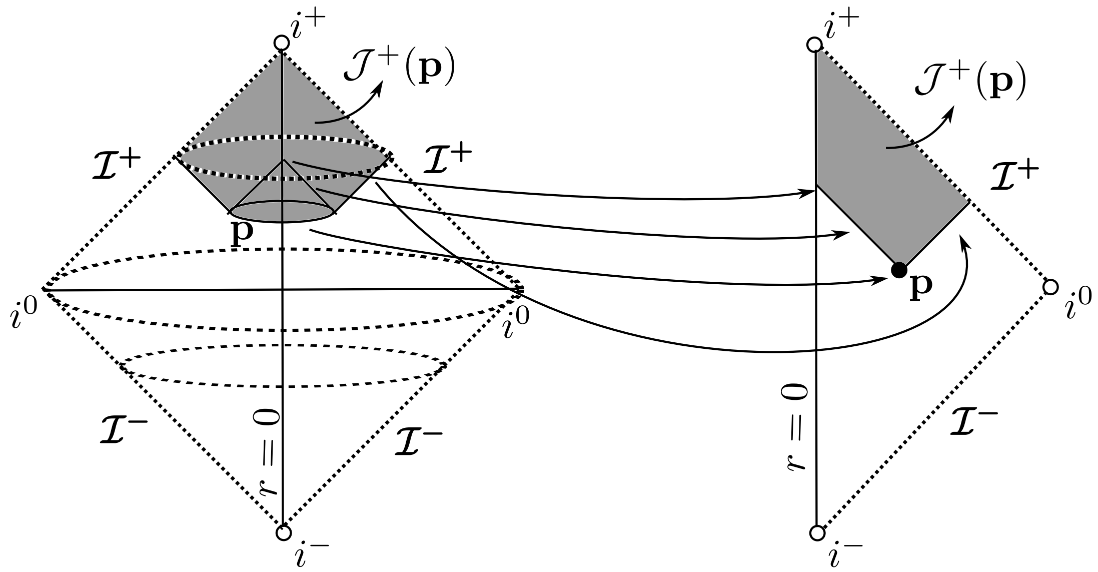


Figure 1.8: The Penrose diagram and causal structure.

More generally, one defines the Penrose diagram of a spherically symmetric spacetime to be the image of a bounded conformal transformation of the quotient spacetime in Minkowski spacetime $\mathbb{R}^{1+1}$. As already mentioned, the importance of such diagrams stems from fact that they allow one to read off the causal structure and recognize the asymptotic structure of a spacetime.

For example, we see that in Minkowski, the past of future null infinity $\mathcal{I}^+$ (where $r = \infty$) is the whole spacetime. That is to say, any point can send signals and thus communication with $\mathcal{I}^+$. However, we can construct spacetimes for which this is not the case. In other words, there are spacetimes which contain points which cannot communicate $\mathcal{I}^+$. The conformal diagram of such spacetime would be as follows:

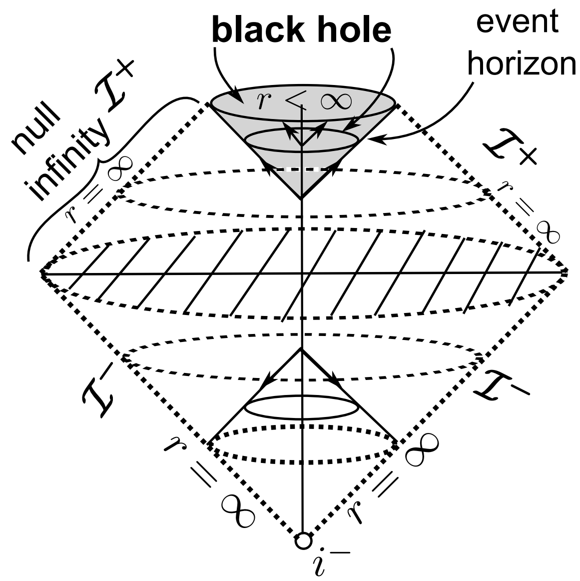


Figure 1.9: The shaded region is known as the black hole region.

The shaded region cannot send signal to $\mathcal{I}^+$ and for this reason is called *black hole*. Black holes are one of the most celebrated predictions of (general) relativity and, in fact, of mathematical physics. These notes will investigate geometric and analytic properties of such regions.

Remark 1.4.2. *The conformal symmetries of Minkowski spacetime are closely related to the global structure of the Penrose diagrams. For more details see Part II of these notes.*

1.5 Electromagnetism and Maxwell Equations

We conclude this section by an account of the relativistic version of Maxwell's equation governing the propagation of electromagnetic fields. After all, special relativity was discovered in order to incorporate Maxwell's theory.

The electric E and magnetic B fields are now incorporated in a new 2-form F on Minkowski space $\mathbb{R}^2$ and the Maxwell equations take the form

$$dF = 0, \qquad \mathrm{Div} F = -4\pi J,$$

where J is a 1-form representing the current density of electric charges. Note that since F is antisymmetric $\mathrm{Div}\mathrm{Div} F = 0$ and hence $\mathrm{Div} J = 0$. If ϵ denotes the volume form in $\mathbb{R}^{3+1}$, then E, B are recovered from F by the equations

$$F_{\mu\nu} = \begin{bmatrix} 0 & E_1 & E_2 & E_3 \\ -E_1 & 0 & -B_3 & B_2 \\ -E_2 & B_3 & 0 & -B_1 \\ -E_3 & -B_2 & B_1 & 0 \end{bmatrix}.$$

Since $dF = 0$, by Poincaré lemma we have that there exists a 1-form A, called the Maxwell potential, such that $F = dA$. If A is such a 1-form, then for any function f, the 1-form $A + df$

is also a Maxwell potential. Note that one can choose f such that the Maxwell potential satisfies $\mathrm{Div}A = 0$. This choice is called the Lorentz gauge. Under this gauge, the Maxwell equations take the form:

$$
\begin{aligned}
&\mathrm{Div}A = 0,\\
&\Box_g A = \mathrm{Div}dA = -4\pi J.
\end{aligned}
$$

Clearly, these equations are covariant under change of coordinates and invariant under isometries.

Chapter 2

Lorentzian Geometry

Introduction

As we have already seen, one needs to deal with curved spaces even in the framework of special relativity (recall that the set $\{X : \ g(X,X) = 1\}$ is a "curved" timelike hypersurface). In this chapter we provide the general framework for curved spaces and introduce the notions of Lorentzian geometry which are necessary for understanding the mathematical aspects of general relativity.

- A *Lorentzian manifold* $(\mathcal{M}, g)$ is a differentiable manifold of dimension $n+1$, endowed with a *Lorentzian metric* g, namely a differentiable assignment of a symmetric, non-degenerate bilinear form g_x with signature $(-,+,\cdots,+)$ in $T_x\mathcal{M}$ at each $x \in \mathcal{M}$.

We will mostly consider the case of four spacetime dimensions (and hence $n = 3$); though our results apply for the general case too.

- A *spacetime* manifold $(\mathcal{M}, g)$ is an orientable four-dimensional Lorentzian manifold.

In contrast to the Riemannian case, a given differentiable manifold might not admit a Lorentzian metric. The reason being that in a Lorentzian manifold there are "distinguished" directions which basically correspond to the $(-)$ of the signature. One can thus easily show the following

Proposition 2.0.1. *A smooth manifold $\mathcal{M}$ admits a Lorentzian metric if and only if it admits a non-vanishing vector field; in other words, $\mathcal{M}$ admits a Lorentzian metric if and only if it is either non-compact or compact with vanishing Euler characteristic.*

The formulas in Riemannian geometry for geodesics, parallel transport, curvatures etc. carry over to Lorentzian manifolds.

2.1 Causality I

Basic notions

The fundamental new aspect of Lorentzian metrics g is that g_x is not positive-definite in $T_x\mathcal{M}$. In fact, for each $x \in \mathcal{M}$, the linear space[1] $(T_x\mathcal{M}, g)$ is isometric to the Minkowski

[1] For simplicity, we will drop the index and thus by g we will also mean g_x.

spacetime $\mathbb{R}^{3+1}$ and therefore there exists a basis $(E_0, E_1, E_2, E_3))$ of $T_x\mathcal{M}$ such that

$$g(E_\alpha, E_\beta) = m_{\alpha\beta},$$

where $m_{\alpha\beta}$ is the Minkowski diagonal matrix $(-1, +1, +1, +1)$. Then, for any vector $X \in T_x\mathcal{M}$ we have $X = \sum_\alpha X^\alpha E_\alpha$ and thus

$$g(X, X) = -(X^0)^2 + (X^1)^2 + (X^2)^2 + (X^3)^2.$$

Note that unless $(\mathcal{M}, g)$ is locally isometric to Minkowski space the frame $(E_\alpha, \alpha = 0, 1, 2, 3)$ does not correspond to a coordinate frame. As in the flat Minkowski case, we say that a vector $X \in T_x\mathcal{M}$ is:

1. *spacelike*, if $g(X, X) > 0$,
2. *null*, if $g(X, X) = 0$,
3. *timelike*, if $g(X, X) < 0$.

If X is either timelike or null, then it is called *causal*. If X is null vector at $\mathbf{p}$, then

$$(X^0)^2 = (X^1)^2 + (X^2)^2 + (X^3)^2$$

and hence X lies on cone (which depends on the basis E_α) with vertex at x. In other words, all null vectors at x span a double cone, known as the *double null cone*. We denote by $\mathcal{N}_x$,

$$\mathcal{N}_x = \left\{ X \in T_x\mathbb{R}^{3+1} \;:\; g(X, X) = 0 \right\},$$

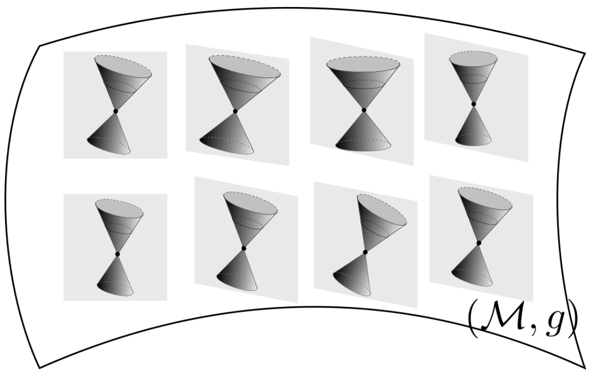
the set of null vectors in $T_x\mathcal{M}$, by $\mathcal{I}_x$,

$$\mathcal{I}_x = \left\{ X \in T_x\mathbb{R}^{3+1} \;:\; g(X, X) < 0 \right\},$$

the set of timelike vectors in $T_x\mathcal{M}$, and by $\mathcal{S}_x$,

$$\mathcal{S}_x = \left\{ X \in T_x\mathbb{R}^{3+1} \;:\; g(X, X) > 0 \right\},$$

the set of spacelike vectors in $T_x\mathcal{M}$. The set $\mathcal{S}_x$ is open (and connected if $n > 1$).



Note that the set $\mathcal{I}_x$ is the interior of the solid double cone enclosed by $\mathcal{N}_x$. Hence, $\mathcal{I}_x$ is an open set consisting of two components which we may denote by $\mathcal{I}_x^+$ and $\mathcal{I}_x^-$. Then we can also decompose $\mathcal{N}_x = \mathcal{N}_x^+ \cup \mathcal{N}_x^-$, where

$$\mathcal{N}_x^+ = \partial\mathcal{I}_x^+, \qquad \mathcal{N}_x^- = \partial\mathcal{I}_x^-.$$

A *time-orientation* of $(\mathcal{M}, g)$ is a continuous choice of a positive component $\mathcal{I}_x^+$ at each $x \in \mathcal{M}$. Then, we call $\mathcal{I}_x^+$ (resp. $\mathcal{N}_x^+$) the set of *future-directed* timelike (resp. null) vectors at x. Similarly we define the past-directed causal vectors.

Causality, observers and particles, proper time

We can readily extend the previous causal characterizations for curves. In particular, a curve $\alpha : I \to \mathcal{M}$ is called future-directed timelike if $\dot{\alpha}(t) \in T_{\alpha(t)}\mathcal{M}$ is a future-directed timelike vector at $\alpha(t)$ for all $t \in I$. Note that the worldline of an observer is represented by a timelike curve. As we shall see in the sequel, if an observer is inertial, then he/she moves on a timelike geodesic. On the other hand, photons move on null geodesics (and hence information propagates along null geodesics).

The *proper time* τ of an observer O is defined to be the parametrization of its worldline α such that $g(\dot{\alpha}(\tau), \dot{\alpha}(\tau)) = -1$. Note that no proper time can be defined for photons since for all parametrizations we have $g(\dot{\alpha}(\tau), \dot{\alpha}(\tau)) = 0$. However, in this case, it is usually helpful to consider affine parametrizations τ with respect to which $\nabla_{\dot{\alpha}}\dot{\alpha} = 0$.

Submanifolds

Let $\mathcal{N}$ be a submanifold of $\mathcal{M}$. Then $\mathcal{N}$ is called:

1. *spacelike* (or Riemannian), if the induced metric $g\Big|_{T_x\mathcal{N}}$ is positive-definite (and thus Riemannian).

2. *timelike* (or Lorentzian), if the induced metric $g\Big|_{T_x\mathcal{N}}$ has signature $(-, +, \cdots, +)$ and thus is Lorentzian.

3. *null*, if the induced metric $g\Big|_{T_x\mathcal{N}}$ is degenerate. Recall that a symmetric bilinear form d on a linear space V is called degenerate if there exists a vector $X \in V$ such that $d(X, Y) = 0$ for all $Y \in V$.

By Sylvester's law of inertia, these are all the possible case for the induced metric in a Lorentzian manifold. The particularly interesting case is when $\mathcal{N}$ is a hypersurface (of codimension one). In such a case, the hypersurface $\mathcal{N}$ can be characterized in terms of its normal (in $\mathcal{M}$) vector field N. Indeed:

1. A hypesurface $\mathcal{N}$ is called spacelike, if the normal N_x at each point $x \in \mathcal{N}$ is timelike.

2. A hypesurface $\mathcal{N}$ is called timelike, if the normal N_x at each point $x \in \mathcal{N}$ is spacelike.

3. A hypesurface $\mathcal{N}$ is called null, if the normal N_x at each point $x \in \mathcal{N}$ is null. In this case, N_x is tangential to $T_x\mathcal{N}$ and hence $g\Big|_{T_x\mathcal{N}}$ is degenerate.

The above are a consequence of the fact that the orthogonal complement of a timelike (resp. spacelike) vector in a Lorentzian space only consists of spacelike (resp. timelike) vectors. For more details about degenerate hyperplanes and hypersurfaces see Section 2.2.

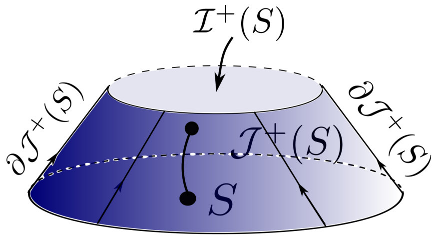
Causality theory

Clearly, a Lorentzian manifold $\mathcal{M}$ is generally not a linear space and hence does not coincide with the tangent planes at its points. For this reason, the definition of global causal structures require some additional care.

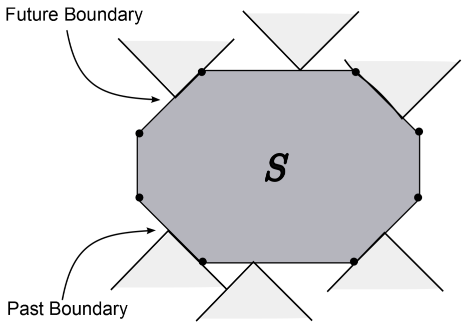
Definitions

We have the following definitions:

- The *causal future* $\mathcal{J}^+(S)$ of a set S is defined to be the set of all points which can be connected with a point of S through a future-directed causal curve.
- The *chronological future* $\mathcal{I}^+(S)$ of a set S is defined to be the set of all points which can be connected with a point of S through a future-directed timelike curve[2].



- A set S is called *achronal* if there is no timelike curve which intersects S twice. Hence, for all $p, q \in S$ we have $p \notin \mathcal{I}^+(q)$ and $q \notin \mathcal{I}^+(p)$.
- A set F is called *future* if for all $x \in F$ we have $\mathcal{I}^+(x) \subset F$.
- Let $\partial S = \overline{S} \cap \overline{\mathcal{M}/S}$ be the (topological) boundary of S in $\mathcal{M}$. An achronal piece B_1 of the boundary ∂S of S is called *future* if for all $x \in B_1$ we have $\mathcal{I}^+(x) \cap S = \emptyset$. Similarly, an achronal piece B_2 of the boundary ∂S of S is called *past* if for all $x \in B_2$ we have $\mathcal{I}^-(x) \cap S = \emptyset$.



Important properties

The definition of the exponential map exp_x at each point $x \in \mathcal{M}$ carries over in the case of Lorentzian manifolds. Let $U_x \subset \mathcal{M}$ be a normal neighborhood around x where exp_x is

[2]Note that $\mathcal{I}^+(x) \subset \mathcal{M}$ denotes the chronological future of x whereas $\mathcal{I}^+_x \subset T_x\mathcal{M}$ denotes the set of all timelike vectors in $T_x\mathcal{M}$.

a diffeomorphism: $exp_x : V_x \subset T_x\mathcal{M} \to U_x \subset \mathcal{M}$, where $V_x = \exp_x^{-1}(U_x)$. The following proposition determines locally the chronological future $\mathcal{I}^+(x)$ of a point x.

Proposition 2.1.1. *Let $(\mathcal{M}, g)$ be a Lorentzian manifold and $x \in \mathcal{M}$. Then,*

$$\mathcal{I}^+(x) \cap U_x = exp_x\big(\mathcal{I}_x^+ \cap V_x\big). \tag{2.1}$$

In particular, if $y \in \mathcal{I}^+(x) \cap U_x$, then y can be connected with x through a timelike geodesic. Furthermore, for any set S the chronological future (resp. past) $\mathcal{I}^+(S)$ (resp. $\mathcal{I}^-(S)$) is open in $\mathcal{M}$.

Proof. First note that

$$exp_x\big(\mathcal{I}_x^+ \cap V_x\big) \subseteq \mathcal{I}^+(x) \cap U_x.$$

Indeed, if $Y \in \mathcal{I}_x^+ \cap V_x$ then $exp_x(Y) = \alpha_Y(1)$, where α_Y is the unique affinely-parametrized geodesic such that $\alpha_Y(0) = x$ and $\dot{\alpha}_Y(0) = Y$. The geodesic α_Y is timelike since Y is timelike and the norm of the tangent to a geodesic is preserved along the geodesic:

$$\nabla_{\dot{\alpha}_Y}\Big(g(\dot{\alpha}_Y, \dot{\alpha}_Y)\Big) = 2g(\nabla_{\dot{\alpha}_Y}\dot{\alpha}_Y, \dot{\alpha}_Y) = 0.$$

We now prove the inverse inclusion. Suppose that it does not hold, that is to say suppose that there is a point $y \in \mathcal{I}^+(x) \cap U_x$ and a timelike curve γ such that $\gamma(0) = x$ and $\gamma(1) = y$ and such that $exp_x^{-1}(y) \notin \mathcal{I}_x^+ \cap V_x$. Then the curve $exp_x^{-1}(\alpha)$ initially enters the interior of the null cone at x and subsequently exits it.

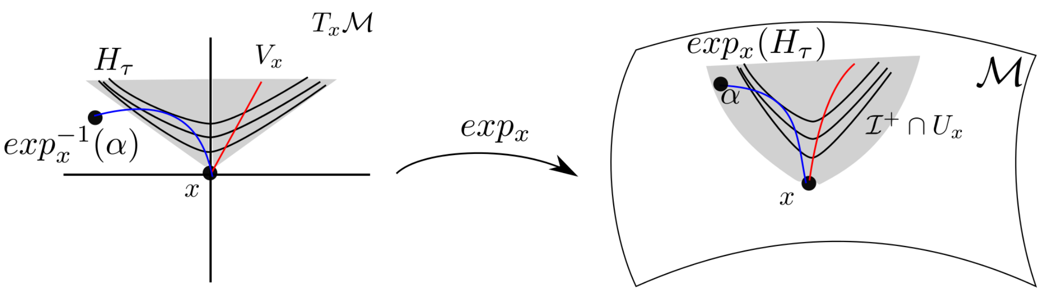
Recall the hyperboloids $H_\tau = \{X : \ g(X, X) = -\tau, \ \tau > 0\} \subset T_x\mathcal{M}$. These are spacelike hypersurfaces in $T_x\mathcal{M}$ and foliate $\mathcal{I}_x^+$. By a continuity argument we can easily deduce that there exists a point where the curve $exp_x^{-1}(\alpha)$ is tangent to a hyperboloid H_τ for some $\tau > 0$. Indeed, if $exp_x^{-1}(\alpha)$ were always transversal to H_τ then the function $f(s) = g(exp_x^{-1}(\alpha(s)), exp_x^{-1}(\alpha(s)))$ would be decreasing; however, this contradicts the fact that for some s_0 the point $\alpha(s_0)$ lies on the null cone and thus $f(s_0) = 0$ (note that $f(0) < 0$).

The proposition now follows from the following lemma which is the analogue of Gauss lemma in Riemannian geometry.

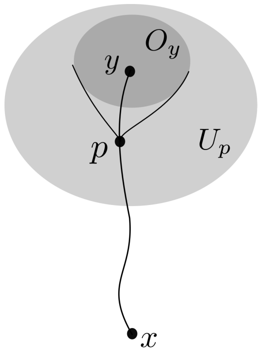
Lemma 2.1.1. ***(Gauss Lemma)*** *Let x be a point in a Lorentzian manifold $(\mathcal{M}, g)$ and $Y \in T_x\mathcal{M} \cap V_x$ be a timelike vector. Let also α_Y be the unique affinely-parametrized geodesic such that $\alpha_Y(0) = x$ and $\dot{\alpha}_Y(0) = Y$. Then, the timelike geodesic $\alpha_Y(t)$ is perpendicular to the hypersurfaces $exp_x(H_\tau)$ in $\mathcal{M}$. In particular, this implies that the hypersurfaces $exp_x(H_\tau)$ are spacelike.*

The proof of the lemma is identical to the Riemanian case and thus is omitted.

Returning to the proof of the proposition, we have that the curve $exp_x^{-1}(\alpha)$ cannot be tangent to H_τ for some τ, because if it were then the timelike curve α would have to be tangent to the spacelike hypersurface $exp_x(H_\tau)$, which, in view of Gauss' lemma, is a contradiction.



The above allows us the locally "compute" the chronological futures. We will use this result to show that the chronological future of sets is open. Since $\mathcal{I}^+(S) = \bigcup_{x\in S} \mathcal{I}^+(x)$ it suffices to show that $\mathcal{I}^+(x)$ is open for all points $x \in \mathcal{M}$. Let $y \in \mathcal{I}^+(x)$ and γ a curve such that $\gamma(0) = x$ and $\gamma(1) = y$. There exists a point p on γ close to y such that $y \in U_p$, where U_p is a normal neighborhood of p. Applying exp_p^{-1} we can pass to $T_p\mathcal{M}$ and obtain an neighborhood O_q of $exp_p^{-1}(q)$ lying in $\mathcal{I}_p^+ \subset T_p\mathcal{M}$. Then, by virtue of the previous results, the open set $exp_p(O_q)$ lies in $\mathcal{I}^+(p)$ and thus in $\mathcal{I}^+(x)$.



□

The following proposition describes the relation between the seemingly opposite notions of achronal and future sets.

Proposition 2.1.2. *Let F be a future set in a Lorentzian manifold $\mathcal{M}$. Then the topological boundary ∂F is a closed achronal three-dimensional locally Lipschitz submanifold of $\mathcal{M}$ such that $\partial_{mani}(\partial F) = \emptyset$, where ∂_{mani} denotes the boundary in the sense of manifolds.*

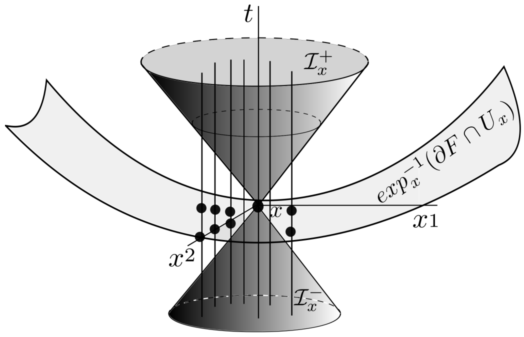
Proof. The topological boundary of a set is by definition closed.

Let now $x, y \in \partial F$ such that $y \in \mathcal{I}^+(x)$. Then there exists an open neighborhood O_x of x in $\mathcal{M}$ such that $O_x \subset \mathcal{I}^-(y)$. Since $x \in \partial F$ there exists a point $p \in O_x$ such that $p \in F$ and $y \in \mathcal{I}^+(p)$. There also exists an open set O_y of y such that $O_y \subset \mathcal{I}^+(p)$. However, since $y \in \partial F$, we have that $O_y \cap \mathcal{M}/F \neq \emptyset$, which contradicts the futureness of F.

The second part of the proposition deals with local notions and so it suffices to restrict our attention to a normal neighborhood U_x of a point $x \in \partial F$. Let $V_x \subset T_x\mathcal{M}$ be such that $exp_x(V_x) = U_x$ and consider inertial coordinates (t, x^1, x^2, x^3) on V_x.

Consider the lines $\gamma(t) = (t, c_1, c_2, c_3)$, i.e. the integral curves of ∂_t in V_x. These are all timelike. Consider also their image $exp_x(\gamma(t))$. In general, since exp_x is not an isometry in general, $exp_x(\gamma)(t)$ might fail to be timelike; however, since $dexp_x$ is the identity when resticted at the tangent plane of the origin of $T_x\mathcal{M}$, by the stability of timelike vectors, we can assume that U_x is sufficiently small so the curves $exp_x(\gamma(t))$ are timelike in U_x for all curves γ. Hence these curves can intersect ∂F at most once. We will show that the curves $exp_x(\gamma(t))$ intersect ∂F exactly once.

First note that $\mathcal{I}^+(x) \subset F$. Indeed, if $q \in \mathcal{I}^+(x)$ then there exists a $x' \in F$ close to x such that $q \in \mathcal{I}^+(x')$ and hence $q \in F$. Similarly, we can show that $\mathcal{I}^-(x) \subset \mathcal{M}/F$. Consider now the picture in $V_x \subset T_x\mathcal{M}$. The curves γ intersect both $\mathcal{I}_x^+ \cap V_x$ and $\mathcal{I}_x^- \cap V_x$. Passing to U_x in $\mathcal{M}$ by applying exp_x and in view of the previous results and Proposition 2.1.1 we have that the curves $exp_x(\gamma)$ intersect F and $\mathcal{M}/F$ and hence intersect ∂F at least (in fact exactly) once.



We will show that ∂F is a locally Lipschitz submanifold with no bounary (in the sense of manifolds). It suffices to prove that $exp_x^{-1}(\partial F \cap U_x)$ satisfies these properties. Since the integral curves γ of ∂_t intersect $exp_x^{-1}(\partial F \cap U_x)$ exactly once, we have that $exp_x^{-1}(\partial F \cap U_x)$ is the graph of the function $t = t(x^1, x^2, x^3)$ defined by

$$exp_x^{-1}(\partial F \cap U_x) = \left\{ \Big(t(x^1, x^2, x^3), x^1, x^2, x^3 \Big) : \ (x^1, x^2, x^3) \in \Big(\{t = 0\} \cap V_x \Big) \right\} \cong \mathbb{R}^3.$$

We will prove that the function $t = t(x^1, x^2, x^3)$ is Lipschitz (and thus continuous). Indeed, a point $q \in T_x\mathcal{M}$ cannot belong (under exp_x) to $\partial F \cap U_x$ if it belongs in a "modified" null cone with vertex at a point $p \in exp_x^{-1}(\partial F \cap U_x)$. By continuity around x we obtain that for a sufficiently small U_x (the size of which does not depend on F) we have

$$\frac{t(x^1, x^2, x^3) - t(\tilde{x}^1, \tilde{x}^2, \tilde{x}^3)}{\|(x^1, x^2, x^3) - (\tilde{x}^1, \tilde{x}^2, \tilde{x}^3)\|} \le 2.$$

and hence the function $t = t(x^1, x^2, x^3)$ is indeed Lipschitz.

The manifold ∂F has no boundary since for any point $x \in \partial F$ the neighborhood $\partial F \cap U_x$ of x in ∂F is homeomorphic to $\mathbb{R}^3$.

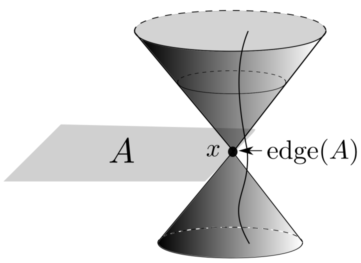
$\square$

Remarks

1. We will be particularly interested in the case where the future set F is such that $F = \mathcal{J}^+(S)$ for a set $S \subset \mathcal{M}$.

2. Proposition 2.1.2 holds in fact for more general achronal sets A and not just for the boundary of future sets. The only additional assumption we need to impose is that the achronal set A has no edge. This assumption implies that the curves $exp_x(\gamma(t))$ do intersect the set A. The precise definition of the edge is the following:

   - The *edge* of an achronal set A consists of all points $x \in \overline{A}$ such that every neighborhood U of x contains a timelike curve which intersects $\mathcal{I}^+(x) \cap U$ and $\mathcal{I}^-(x) \cap U$ but does not intersect A.



3. Note that the weak regularity of Proposition 2.1.2 cannot be improved. Note, for example, that the null cone (being the boundary of the future of a point in Minkowski) has a differentiable singularity at the vertex.

2.2 Null Geometry

Null geometry is a new kind of geometry that naturally lives on hypersurfaces of Lorentzian manifolds.

Definition and Applications

Let $(\mathcal{M}, g)$ be a $(3+1)-$dimensional Lorentzian manifold.

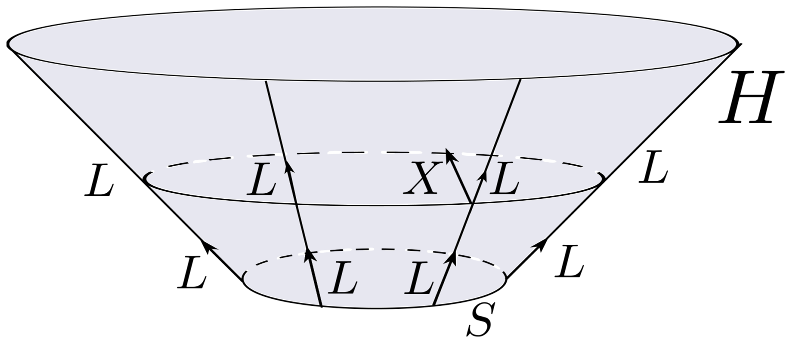
- A hypersurface H of $\mathcal{M}$ is called *null*, if at each point $x \in H$ the normal $L \in T_x\mathcal{M}$ to T_xH is a null vector, that is

$$g(L, L) = 0 \text{ and } g(L, X) = 0 : \forall X \in T_xH.$$

  Then, since $\dim(T_xH) = 3$ we have $T_xH = \langle L\rangle^\perp$ and since L is null we have that $L \in T_xH$ and hence for each $x \in \mathcal{M}$ the hyperplane T_xH is degenerate.

The structure of these hypersurfaces is of paramount importance in relativity. The following lemma has important applications for the geometry of a null hypersurface.

Lemma 2.2.1. *Let $x \in \mathcal{M}$ and H be a null hupersurface on $\mathcal{M}$ with $T_pH = \langle L\rangle^\perp$. Then, if $X \in T_xH$ then either X is null and $X \in \langle L\rangle$ or X is spacelike.*



Proof. If $X = (x_0, x_1, ..., x_n)$ and $L = (l_0, l_1, ...l_n)$ with respect to an orthonormal basis, then $g(L, X) = 0$ and $g(L, L) = 0$ can be written as $x_0l_0 = x_1l_1 + ... + x_nl_n$ and $l_0^2 = l_1^2 + ... + l_n^2$, respectively. By Cauchy–Schwarz inequality we obtain

$$(l_1^2 + ... + l_n^2)(x_1^2 + ... + x_n^2) \geq (l_1x_1 + ... + l_nx_n)^2 = (x_0l_0)^2,$$

which implies that $g(X, X) \geq 0$ and we have equality if and only if $X \in \langle L\rangle$.

$\square$

Null Generators

Therefore, we have a distinguished line bundle on H, which is the null lines spanned by the normal L to the hupersurface (recall that in this case the normal is also tangential to the hypersufrace). We next investigate the integral curves of this line bundle. We consider the covariant derivative $\nabla_L L$ and take: If $X \in TH$, then

$$g(\nabla_L L, X) = -g(L, \nabla_L X) = -g(L, \nabla_X L) - g(L, [L, X]) = -\frac{1}{2} X\Big(g(L, L)\Big) = 0, \qquad (2.2)$$

since $[L, X] \in TH$. Hence the integral curves are geodesics. We arrive at the following

Proposition 2.2.1. *A null hypersurface H is generated (ruled) by null geodesics whose tangent is the normal to H.*

We will refer to these null geodesics as the null generators of H. Note also that a natural choice for the normal L to H is such that L satisfies the geodesic equation $\nabla_L L = 0$.

Corollary 2.2.1. *Conformal transformations between Lorentzian manifolds map null hypersurfaces to null hypersurfaces and hence null geodesics to null geodesics.*

Optical Functions

Equation (2.2) holds only if X is tangential to H. We will next present a method to derive a similar formula for *any* vector X in $\mathcal{M}$.

A null hypersurface H can be considered to be a non-critical level set of a smooth function u, i.e. $H = \{u = c\}$, where c is a constant. In this case, the vector ∇u is normal to H and hence null, and hence u must satisfy the so-called *eikonal* equation

$$g(\nabla u, \nabla u) = g^{\mu\nu}(\partial_\mu u)(\partial_\nu u) = 0. \qquad (2.3)$$

For simplicity, let us denote $L = \nabla u$. If $x \in H$ then for any $X \in T_x\mathcal{M}$ we have:

$$\begin{aligned} g(\nabla_L L, X) &= \big(\nabla_L L\big)_\flat(X) = \big(\nabla_L L_\flat\big)(X) = \nabla_L(du)(X) \\ &= \nabla^2 u\,(L, X) = \nabla^2 u\,(X, L) = \nabla_X(du)(L) = g(\nabla_X L, L) = \frac{1}{2}\nabla_X\big(g(L, L)\big), \end{aligned}$$

which indeed generalizes (2.2). The musical isomorphism $\flat$ corresponds to the standard index-lowering. Hence, if $L = \nabla u$ then we obtain

$$\nabla_L L = \frac{1}{2}\nabla\Big(g(L, L)\Big). \qquad (2.4)$$

A particularly important application of formula (2.4) is when the function u is an optical function, namely:

- A differentiable function $u : \mathcal{M} \to \mathbb{R}$ is called *optical* if all level sets $\{u = c\}$ are null hypersurfaces. In this case, the eikonal equation is satisfied along all level sets of u.

If u is an optical function, and $L = \nabla u$, then $g(L, L) = 0$ everywhere and hence the gradient of this function also vanishes. Hence, the vector field L is a geodesic null normal to H since $\nabla_L L = 0$.

Remark: An example of a function u such that $u = 0$ is a null hypersurface but u is not an optical function is the following:

$$u(t, x^1, x^2, x^3) = -t^2 + (x^1)^2 + (x^2)^2 + (x^3)^2$$

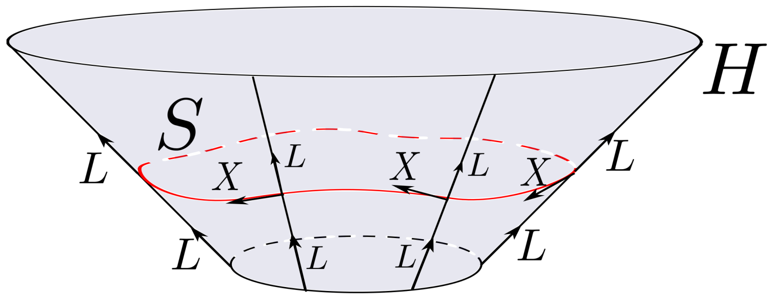
on Minkowski spacetime (with the origin deleted).

For more about optical functions see Section 4.1.

Sections of null hypersurfaces

- A *section* S of a null hypersurface H is a two-dimensional submanifold of H which intersects each null geonerator of H transversally.

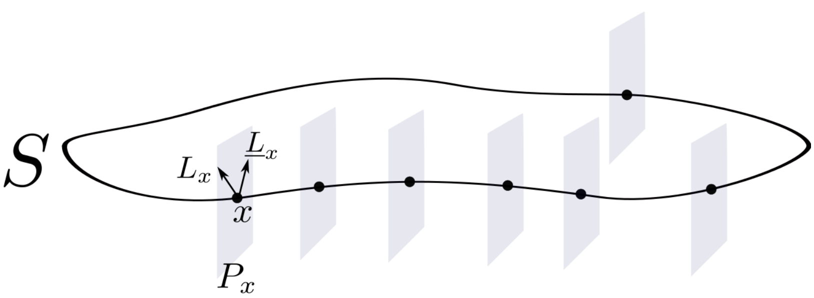
Clearly, by virtue of Lemma 2.2.1 and the above transversality assumption, any section S is a two-dimensional Riemannian manifold. Note that, since H is null, all null generators intersect S orthogonally.



We are mainly interested in the case where S is topologically homeomorphic to the two-dimensional sphere $\mathbb{S}^2$.

Spacelike surfaces and their associated null normal geodesic congruences

Let us now start with a surface S, namely a two-dimensional Riemannian manifold (homeomorphic to $\mathbb{S}^2$). For every point $x \in S$ we have $dim(T_xS) = 2$ and $g\big|_{T_xS}$ is positive definite. Hence, by Sylvester's law of inertia we have that the orthogonal complement $P_x = (T_xS)^\perp$ of T_xS in $T_x\mathcal{M}$ is a two-dimensional Lorentzian plane. In other words, $(P_x, g\big|_{P_x})$ is isometric to the two-dimensional Minkowski spacetime.



Since all vectors orthogonal to S at x belong in P_x the above discussion shows that there are exactly two null lines orthogonal[3] to S at x. Let L_x and $\underline{L}_x$ be two future-directed null vectors in P_x. Suppose that L_x projects to the exterior of S and $\underline{L}_x$ projects to the interior of S. We will call L_x the outer null normal to S at x and $\underline{L}_x$ the inner null normal to S at x. Although there is no natural normalization[4] for L_x (and $\underline{L}_x$), we require the choice of L_x (and $\underline{L}_x$) to depend differentiably on x so the resulting vectorfield L (and $\underline{L}$) along S is differentiable.

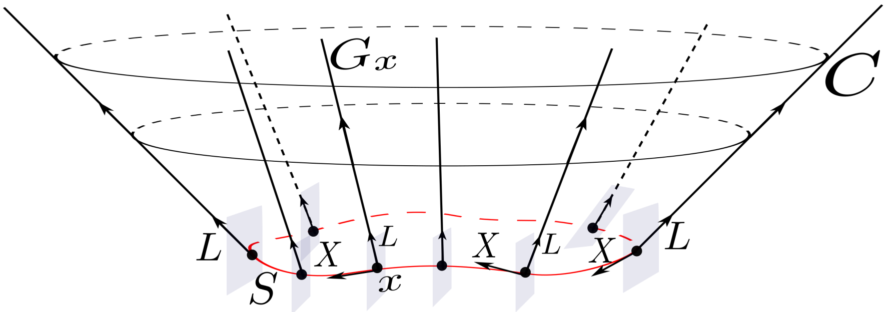
The vector bundle

$$\mathcal{P} = \bigcup_{x \in S} P_x$$

is called the normal bundle of S in $\mathcal{M}$. Note that $P_x = \langle L_x, \underline{L}_x \rangle$.

Let now $(L, \underline{L})$ be given future-directed null normal vector fields to S. For each $x \in S$ there is a unique affinely parametrized geodesic G_x with initial conditions (x, L_x), that is to say $G_x(0) = x$ and $\dot{G}_x(0) = L_x$. Similarly, there is a unique affinely parametrized geodesic $\underline{G}_x$ with initial conditions $(x, \underline{L}_x)$. We consider the sets in $\mathcal{M}$ formed by these geodesics:

$$C = \bigcup_{x \in S} G_x, \qquad \underline{C} = \bigcup_{x \in S} \underline{G}_x. \tag{2.5}$$



The null geodesics G_x (resp. $\underline{G}_x$) are called the null generators of C (resp. $\underline{C}$). We assume that $C, \underline{C}$ are smooth hypersurfaces (or that at least the part of $C, \underline{C}$ are smooth hypersurfaces). Note that in general $C, \underline{C}$ will not be globally smooth hypersurfaces. For example, the incoming null geodesic congruence of a standard sphere in Minkowski spacetime forms a cone which is singular at the vertex. For more about the regularity of $C, \underline{C}$ see Part II of these notes.

We have the following

Proposition 2.2.2. *The sets $C, \underline{C}$ defined above are null hypersurfaces.*

Before we prove Proposition 2.2.2 it is convenient to define affine foliations of $C, \underline{C}$. From now on we focus on the (smooth) hypersurface C.

The vector field L on S extends to a vector field L on C as the tangent field to the affinely parametrized null generators of C. Then L satisfies

$$\nabla_L L = 0$$

[3]Clearly, the null lines span a null cone at x, however there are only two null lines normal to S.

[4]For more about the normalization see the end of this section.

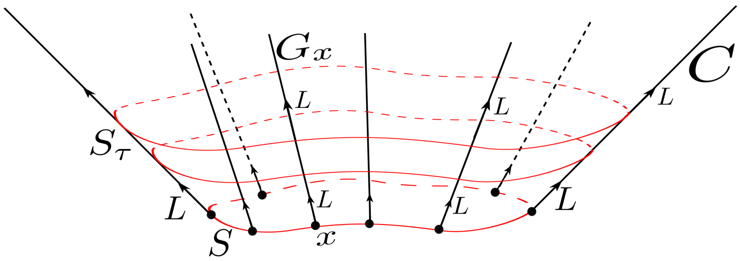
on C. Consider the flow $\mathcal{F}_\tau$ of L which is a one-parameter family of diffeomorphisms of C. We define the surfaces

$$S_\tau = \mathcal{F}_\tau(S)$$

which foliate C:

$$C = \bigcup_{\tau \geq 0} S_\tau. \tag{2.6}$$

Clearly $S_0 = S$ and S_τ are sections of C. Note that τ is the affine parameter of L, namely $L\tau = 1$ and $\tau = 0$ on S. For this reason we call the foliation defined by (2.6) an *affine foliation* of C. Understanding the geometric properties of the surfaces S_τ is of fundamental importance in Lorentzian geometry.



Let $y \in C$. Suppose that $y \in G_x$, and in fact $y = G_x(\tau)$. Then $y \in S_\tau$. Let (E_1, E_2) be a basis of T_xS. Then (L, E_1, E_2) is a basis for T_xC. We can propagate this basis along G_x according to

$$[L, E_i] = 0, \ \ i = 1, 2.$$

Then the vectors E_1, E_2 are tangential to the sections S_τ and in fact we have:

$$(E_i)_{G_x(\tau)} = d\mathcal{F}_\tau\big((E_i)_x\big), \ \ i = 1, 2.$$

The vector fields $E_i, i = 1, 2$ are called *normal*[5] *Jacobi fields* along the null geodesic G_x. They can be thought of as the infinitesimal displacement field of G_x through nearby null generators whose feet on S span a curve tangential to $E_i, i = 1, 2$. Indeed, if $\gamma(s) : (-\epsilon, \epsilon) \to S$ is a curve on S such that $\gamma(0) = x$ and $\dot{\gamma}(0) = E_i$, then we have a mapping

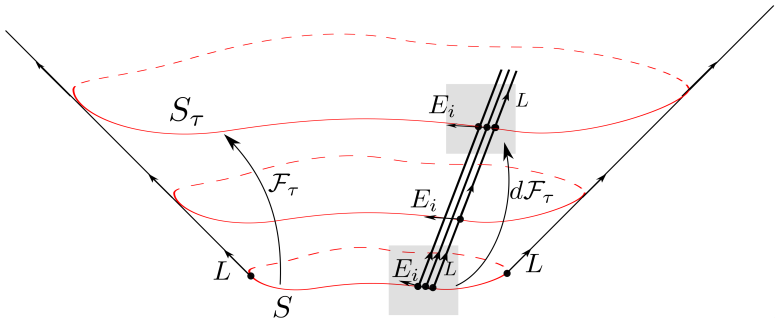
$$\Phi : (-\epsilon, \epsilon) \times [0, \infty) \to C$$

given by

$$\Phi(s, \tau) = G_{\gamma(s)}(\tau).$$

Then, if we fix τ, the curve $\alpha(s) = G_{\gamma(s)}(\tau)$ represents a displacement of $G_x(\tau)$ along nearby null generators. The infinitesimal displacement is given by $E_i = \frac{d}{ds} G_{\gamma(s)}(\tau)\big|_{s=0}$.

[5]Note that $g(L, E_i) = 0, i = 1, 2$.



The triple (L, E_1, E_2) is a basis for T_yC, the linear map $d\mathcal{F}_\tau : T_xC \to T_yC$ being an isomorphism (see Section 5.2 for a discussion of the case where this map is **not** an isomorphism).

Proof of Proposition 2.2.2. Let $y = G_x(\tau) \in C$ and (L, E_1, E_2) be a Lie propagated frame along the null generator G_x. We will show that T_yC is a null hyperplane. If $X \in \{L, E_1, E_2\}$, then

$$\nabla_L\big(g(L,X)\big) = g(\nabla_L L, X) + g(L, \nabla_L X) = g(L, \nabla_X L) = \frac{1}{2}\nabla_X\big(g(L,L)\big) = 0 \qquad (2.7)$$

along G_x, where we used the fact that $[L, X] = 0$ and that L is null on C. However, by assumption $L \perp \langle L, E_1, E_2\rangle$ at x and so $g(L, X) = 0$ for $\tau = 0$. In view of (2.7), we have that $g(L, X) = 0$ along G_x and hence the hyperplane T_yC is indeed null.

□

Note that had we chosen an arbitrary null vector L_x (not necessarily normal to S) then we would not have obtained a null hypersurface. We will see several examples of timelike hypersurfaces spanned by null geodesics in the sequel.

We have the following definition:

- Given a surface S in $\mathcal{M}$, the outgoing null hypersurface C is called the*outer null geodesic congruence* normal to S. Similarly, the incoming null hypersurface $\underline{C}$ is called the*inner null geodesic congruence* normal to S.

The importance of the null geodesic congruences will become apparent in Section 2.4.

Remarks:

1. As we mentioned above, there is no natural normalization for the null normal to S vector fields $L, \underline{L}$. However, we can impose on the pair $(L, \underline{L})$ the normalization condition:

$$g(L, \underline{L}) = -1.$$

If L transforms according to

$$L \mapsto aL,$$

where a is an arbitrary positive differentiable function on S, then $\underline{L}$ transforms according to

$$\underline{L} \mapsto a^{-1}\underline{L}.$$

Given a normalized null pair $(L, \underline{L})$, we define the pair:

$$T = \frac{1}{\sqrt{2}}(L + \underline{L}), \qquad N = \frac{1}{\sqrt{2}}(L - \underline{L}).$$

One can easily see that for each $x \in S$, the pair (T_x, N_x) forms a positively-oriented orthonormal basis of P_x. If $(L', \underline{L}')$ is another normalized pair related to $(L, \underline{L})$ via the positive differentiable function α, then the pair (T', N'), obtained as above, gives another orthonormal basis of P_x for each $x \in S$. Note that the group $SO(1,1)$ of Lorentz boosts in two spacetime dimensions (described in Section 1.3.2) connects such positively-oriented orthonormal bases.

If S is contained in a spacelike hypersurface Σ such that S is the boundary of a compact region K in Σ, then a natural choice is to take T to be the unit future-directed timelike normal to Σ at S and N the unit outward normal to S in Σ.

2. Similar results apply for the (future) null geodesic cone with vertex a point $O \in \mathcal{M}$. This is defined as the set of all (future) null geodesic which emanate at the point O.

2.3 Global Hyperbolicity

The causal structure of Lorentzian manifolds might in some cases exhibit unphysical behavior. For example, there are Lorentzian manifolds with closed timelike curves. This kind of behavior is very pathological and so we want to impose conditions on the spacetimes in order to exclude it.

- An achronal hypersurface H is a *Cauchy hypersurface* if every intextendible causal curve intersects Σ exactly once.
- A spacetime $(\mathcal{M}, g)$ which possesses a Cauchy hypersurface is called *globally hyperbolic*.

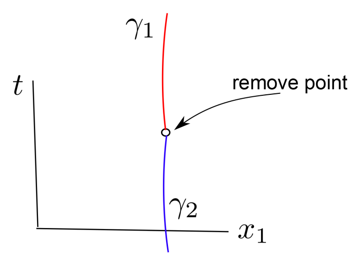
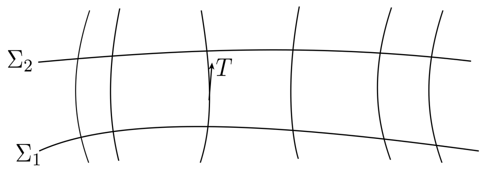


Figure 2.1: The spacetime $\mathbb{R}^{1+1}$ minus a point is not globally hyperbolic since any achronal hypersurface does not intersect one of the causal inextendible curves γ_1 or γ_2.

The existence of a Cauchy hypersurface is a global causal property of the spacetime. According to the following proposition, if a Cauchy hypersurface exists, then its topology is "unique".

Proposition 2.3.1. *If a time-orientable spacetime $(\mathcal{M}, g)$ admits two Cauchy hypersurfaces Σ_1, Σ_2 then Σ_1 is homeomorphic to Σ_2.*

Proof. Since $(\mathcal{M}, g)$ is time-orientable, there exists a non-vanishing timelike vector field T. The integral curves of T are timelike and hence intersect Σ_1 and Σ_2 exactly once.



The projection of Σ_1 onto Σ_2 is a continuous bijective map whose inverse is the projection of Σ_2 onto Σ_1. $\square$

Furthermore, the topology of a globally hyperbolic spacetime is completely determined by that of a Cauchy hypersurface:

Proposition 2.3.2. *If Σ is a Cauchy hypersurface for $(\mathcal{M}, g)$, then $\mathcal{M}$ is homeomorphic to $\Sigma \times \mathbb{R}$. In particular, this implies that there exists a global 'time' function $t : \mathcal{M} \to \mathbb{R}$ such that each level set $\Sigma_\tau = \{t = \tau\}$ is a (spacelike) Cauchy hypersurface and thus the vector field ∇t is everywhere timelike. Furthermore, the hypersurfaces Σ_τ foliate $\mathcal{M}$.*

Global hyperbolicity plays a role similar to that of completeness of Riemannian manifolds.

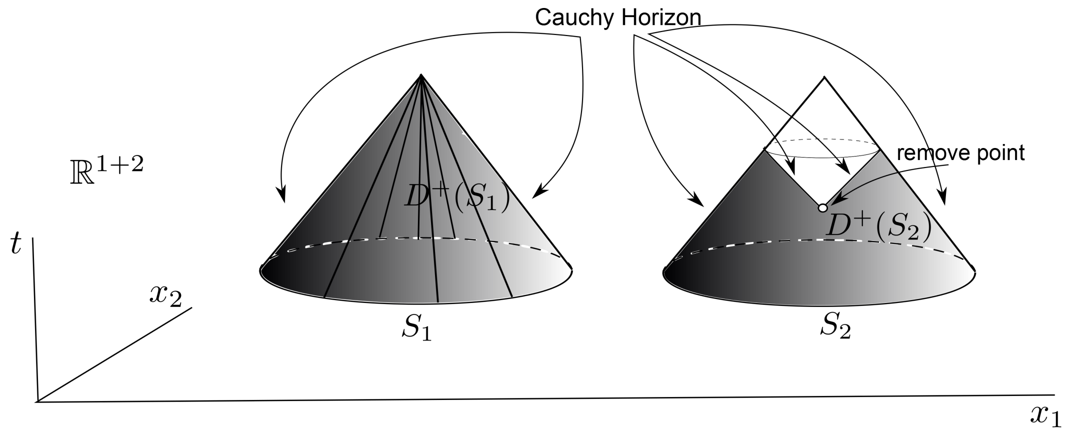
Proposition 2.3.3. *Let $(\mathcal{M}, g)$ be a globally hyperbolic spacetime. Let also $x, y \in \mathcal{M}$ with $y \in \mathcal{I}^+(x)$. Then there exists a timelike geodesic γ which connects x, y and maximizes the length function defined by the following formula*

$$L(\gamma) = \int_0^s \Big(- g(\dot{\gamma}(t), \dot{\gamma}(t)) \Big)^{\frac{1}{2}} dt.$$

A spacetime may not be globally hyperbolic; however there may exist subsets of such spacetimes which are indeed globally hyperbolic.

- Let A be an achronal three-dimensional hypersurface of a spacetime $(\mathcal{M}, g)$. Then the *Cauchy development* $D(A)$ of A is the biggest globally hyperbolic subset of $\mathcal{M}$ which admits A as a Cauchy hypersurface. In other words, $D(A)$ consists of all points $x \in \mathcal{M}$ such that every inextendible causal curve through x intersects A. The *future development* is defined to be $D^+(A) = \mathcal{J}^+(A) \cap D(A)$, and similary we can define the past development.

- The *future Cauchy horizon* of the development $D^+(A)$ in $\mathcal{M}$ is defined to be the future boundary of $D^+(A)$ in $\mathcal{M}$.



By virtue of Proposition 2.1.2, the Cauchy horizon is a three-dimensional Lipschitz submanifold.

2.4 Causality II

The following two propositions are extremely important for our analysis. They provide additional information regarding causal structures of globally hyperbolic spacetimes.

Recall that the topological boundary ∂S of set S in $\mathcal{M}$ is defined to be $\partial S = \overline{S} \cap \overline{\mathcal{M}/S}$.

Proposition 2.4.1. *Let K be a compact subset of a globally hyperbolic spacetime $(\mathcal{M}, g)$. Then $\mathcal{J}^+(K)$ is closed in $\mathcal{M}$ and hence $\partial\mathcal{J}^+(K) \subset \mathcal{J}^+(K)$.*

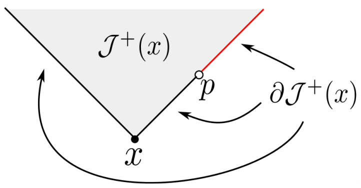
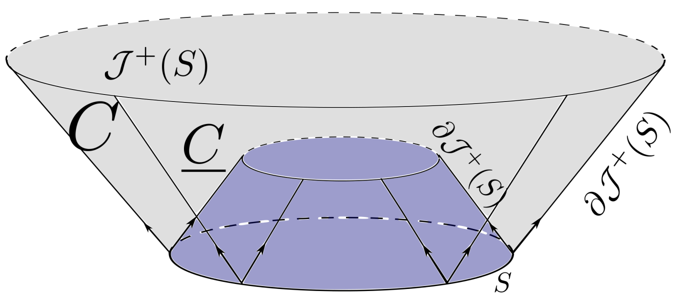


Figure 2.2: If we remove the point p from $\mathbb{R}^{1+1}$ then the set $\mathcal{J}^+(x)$ is not closed anymore.

Proposition 2.4.2. *Let S be a compact orientable two-dimensional surface in a globally hyperbolic spacetime $(\mathcal{M}, g)$ (for example let S be a surface homeomorphic to the sphere $\mathbb{S}^2$). Let C and $\underline{C}$ denote the (future) outgoing and incoming null geodesic congruence normal to S, respectively. Then,*

$$\partial\mathcal{J}^+(S) \subset C \bigcup \underline{C}.$$



Sketch of proof of Proposition 2.4.2: In view of Proposition 2.4.1, if $x \in \partial\mathcal{J}^+(S)$ then there is a caucal curve γ which connects x and a point of S. Basically, this is the only point in the proof where the global hyperbolicity is needed; contrast this with Figure 2.2. Clearly γ cannot be timelike, since otherwise x would not lie in the boundary of the future of S but in the interior instead. If γ is causal and not everywhere differentiable, then by a variation of γ we can find a strictly timelike curve which connects x with some point in S. Hence γ must be a null curve, and in fact by a variation argument again, it must be a null geodesic. One more variation of γ shows that γ must be orthogonal to S.

We will further characterize the boundary $\partial\mathcal{J}^+(S)$ in Section 5.3.

Chapter 3

Introduction to General Relativity

General relativity is a successful geometric theory of gravitation. In chapter 1 we saw why special relativity provides the natural framework for electromagnetism and Maxwell's equations. In this chapter, we describe the main geometric features of spacetimes modeling gravitational systems.

3.1 Equivalence Principle

It is not a priori obvious why the setting of Lorentzian geometry allows one to extend special relativity and thus incorporate gravity as well. However, in 1907 Einstein suggested the following principle (also known as Einstein's happiest thought):

Equivalence principle

- One cannot discriminate (at least locally) between accelerating systems and systems fixed in a gravitational field.

In other words, an observer who is fixed somewhere about the surface of the earth is accelerating. However, the observer cannot decide whether his acceleration is due to gravity or another force that forces him to accelerate. This also implies that the gravitational mass is equal to the inertial mass.

Since there is no 'magic' with gravity we have no way to distinguish gravity from other forces. For this reason, in general relativity, gravity is not described as a force field; rather we are forced to view gravity as an aspect of spacetime structure.

The equivalence principle may also be thought of as the following statement, known as *the geodesic hypothesis*:

- Inertial freely falling observers in a gravitational field move on timelike geodesics and photons move on null geodesics.

Note that freely falling observers are not accelerating. If one freely falls on earth then he will feel strange because his state is rapidly changing from the previous accelerating state of motion. Indeed, if a freely falling observer tries to lift his hand during the free fall then he will feel no weight and thus no external forces on him. This means that his motion is inertial and hence his worldline is a geodesic.

3.2 The Einstein Equations

The natural framework to develop general relativity is that of Lorentzian geometry. Indeed, in Minkowski spacetime the geodesics are just lines (with respect to inertial systems) and thus freely falling observers do not experience non-trivial dynamics. The only way to have a spacetime with non-trivial geodesics is by considering non-flat metrics. Hence, we seek for equations/laws that will determine the metric of the spacetime in the presence of matter fields.

According to Newtonian physics, laws of physics should depend on the second order derivative. Applied to our setting, this means that we need to impose conditions on the curvature of the spacetime $(\mathcal{M}, g)$. These conditions will depend on the matter model present $\mathcal{M}$.

Einstein equations:

$$Ric(g) - \frac{1}{2}R_{sc}(g)\, g = 8\pi\mathbf{T}, \tag{3.1}$$

where $Ric, R_{sc}(g)$ denote the Ricci and scalar curvature, respective, and $\mathbf{T}$ denotes the energy-momentum tensor of the present matter. Note that $\mathbf{T}$ is a $(0,2)$ symmetric divergence free tensor field.

We will be mostly interesting in the case where no matter is present. Then $\mathbf{T} = 0$ and 3.1 imply that the vacuum equations are:

Einstein-vacuum equations:

$$Ric(g) = 0. \tag{3.2}$$

Minkowski spacetime is the trivial solution of the vacuum equations. As we shall see in the sequel, most of the qualitative properties of (3.1) are already present in (3.2).

3.3 The Cauchy Problem

As with any physical theory, it is very important to confirm that general relativity is a locally well posed theory. In particular, we are interested in showing that given appropriate smooth initial data then there exists (at least locally in time) a unique smooth solution to the Einstein equations.

First we need to determine the "correct" notion for the initial data. We expect to have well-posedness only as long as $(\mathcal{M}, g)$ is globally hyperbolic. In this case, there exists a time functions t, such that the level sets $\mathcal{H}_\tau$ are spacelike Cauchy hypersurfaces and $\mathcal{M} = \cup_t \mathcal{H}_t$. Then, the vector field ∇t is normal to $\mathcal{H}_\tau$. We define the lapse function Φ by

$$\Phi = \frac{1}{\sqrt{-g(\nabla t, \nabla t)}}.$$

Then,

$$T = -\Phi^2 \cdot \nabla t \Rightarrow T \perp \mathcal{H}_\tau \text{ and } Tt = 1$$

and

$$N = \frac{1}{\Phi} \cdot T \Rightarrow N \perp \mathcal{H}_\tau \text{ and } g(N, N) = -1.$$

We can now consider a coordinate system (x^1, x^2, x^3) on $\mathcal{H}_0$. This can be propagated to any $\mathcal{H}_\tau$ by the flow of T, constructing a coordinate system (t, x^1, x^2, x^3) of $\mathcal{M}$. Then, the metric with respect to this system takes the form

$$g = -\Phi^2 dt^2 + \bar{g},$$

where $\bar{g} = \bar{g}(t)$ is the induced metric on $\mathcal{H}_t$.

We look for initial data to prescribe on $\mathcal{H}_0$. Since the equations $Ric(g) = 0$ are of second order in g, we must prescribe the metric $\bar{g}$ and also somehow the transversal derivatives $\partial_t \bar{g}_{ij}$ of $\bar{g}$ on $\mathcal{H}_0$. In the context, however, of the Cauchy problem, we want to a priori describe the initial data set only in terms of the geometry of $\mathcal{H}_0$, and hence we want to avoid prescribing the t-derivatives. This can be done by introducing the *second fundamental form k of $\mathcal{H}_0$ in $\mathcal{M}$*. More generally, the second fundamental form of $\mathcal{H}_t$ is defined to be the following $(0, 2)$ tensor field on $\mathcal{H}_t$:

$$k(X, Y) = g(\nabla_X N, Y),$$

where $X, Y \in T_p\mathcal{H}_t$. If $X, Y \in \{\partial_1, \partial_2, \partial_3\}$, where $\partial_i = \partial_{x^i}$, then we obtain[1]:

$$k_{ij} = \frac{1}{2\Phi} \frac{\partial \bar{g}_{ij}}{\partial t},$$

which is known as the *first variational formula.* This formula shows that the initial data set consists of the triplet $(\mathcal{H}_0, \bar{g}, k)$. Note that the first variational formula implies $\frac{\partial}{\partial t}(\sqrt{\bar{g}}) = \Phi \cdot trk \cdot \sqrt{\bar{g}}$.

The trace of the Codazzi equation and the double trace of the Gauss equation of the embedding of $\mathcal{H}_t$ in $(\mathcal{M}, g)$ give us respectively:

$$\begin{aligned} \overline{div} k - \overline{d} k &= Ric(N, \cdot), \\ \overline{R_{sc}} + (trk)^2 - |k|^2 &= R_{sc} + 2Ric(N, N), \end{aligned}$$

where the operators and quantities with the bar above refer to the induced geometry of $(\mathcal{H}_t, \bar{g})$.

In view of $Ric(g) = 0$, the right hand side of the Codazzi and Gauss equations vanishes. Hence, $(\bar{g}, k)$ must be prescribed so they satisfy these *constraint equations.* This completes the formulation of the initial value problem for the Einstein equations.

Let us now show that given such an initial data set, there exists a unique smooth local solution $(\mathcal{M}, g)$. Since the differential structure of $\mathcal{M}$ is known, it suffices to show the existence of g with respect to the coordinate system (t, x^1, x^2, x^3) defined above. Without loss of generality, we can take $\Phi = 1$ on $\mathcal{H}_0$. Then,

$$g|_{\mathcal{H}_0} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & \bar{g}_{ij} & \\ 0 & & & \end{pmatrix}, \partial_t g|_{\mathcal{H}_0} = \begin{pmatrix} \partial_t g_{tt} & \partial_t g_{tx^1} & \partial_t g_{tx^2} & \partial_t g_{tx^3} \\ \partial_t g_{tx^1} & & & \\ \partial_t g_{tx^2} & & 2k_{ij} & \\ \partial_t g_{tx^3} & & & \end{pmatrix}. \tag{3.3}$$

[1] We will not prove equations here. However, we prove all the analogous null structure equations in detail in Chapter 4

Note that $\partial_t g_{tx^\alpha}$ can be freely chosen initially for $\alpha = 0, 1, 2, 3$.

We next briefly describe Choque–Bruhat's argument for the existence of solutions to $Ric(g) = 0$ under the above initial conditions. Let us suppose that $t = x^0$. We define

$$\Gamma^a = g^{\mu\nu}\Gamma^a_{\mu\nu},$$

where $\Gamma^a_{\mu\nu}$ are the Christoffel symbols of g with respect to the coordinate system $x^a, a = 0, 1, 2, 3$. We also define

$$\Gamma_a = g_{ab}\Gamma^b, \qquad H_{\mu\nu} = R_{\mu\nu} - S_{\mu\nu}.$$

The crucial observation is that

$$P.P.(H_{\mu\nu}) = g^{\alpha\beta}\partial_\alpha\partial_\beta g_{\mu\nu} = P.P.(\Box_g g_{\mu\nu}),$$

where $P.P.$ =Principal Part. Given initial data as in (3.3), by standard theory of quasilinear equations, there exists a unique smooth local solution g to

$$H_{\mu\nu}(g) = 0.$$

See Chapter 9 for an introduction to the energy method which can be used to address the local well posedness of such equations.

However, we want g to satisfy $R_{\mu\nu}(g) = 0$ rather than $H_{\mu\nu} = 0$. In other words, we want to make sure that the solution g to $H_{\mu\nu}(g) = 0$ somehow also satisfies $S_{\mu\nu}(g) = 0$ or, in fact, $\Gamma_a = 0$.

If $H_{\mu\nu} = 0$ then

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = S_{\mu\nu} - \frac{1}{2}Sg_{\mu\nu} \;\Rightarrow\; \mathrm{Div}(S_{\mu\nu} - \frac{1}{2}Sg_{\mu\nu}) = 0,$$

since for any metric g we have $\mathrm{Div}(G_{\mu\nu}) = 0$, where $G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu}$. The second crucial observation is that

$$P.P.\Big(\mathrm{Div}(S_{\mu\nu} - \frac{1}{2}Sg_{\mu\nu})\Big) = g^{\alpha\beta}\partial_\alpha\partial_\beta\Gamma_\mu = P.P.(\Box_g\Gamma_\mu).$$

Hence, for a metric g solving $H_{\mu\nu}(g) = 0$, Γ_μ can be shown to be identically zero if Γ_μ and $\partial_t\Gamma_\mu$ are zero on $\mathcal{H}_0$. It can be easily shown that

$$\partial_t g_{tx^i} = \overline{\Gamma}_i,\; \partial_t g_{tt} = 2trk \;\Rightarrow\; \Gamma_\mu|_{\mathcal{H}_0} = 0.$$

However, we have now completely exhausted the freely chosen initial data and hence we would need a miracle so that $\partial_t\Gamma_\mu|_{\mathcal{H}_0} = 0$. The third crucial observation is that if $H_{\mu\nu} = 0$ then

$$\partial_t\;\Gamma_i|_{\mathcal{H}_0} = 2G_{tx^i}, \quad \partial_t\;\Gamma_t|_{\mathcal{H}_0} = 2G_{tt}.$$

However, the constraint equations (satisfied by the initial data $\overline{g}, k$) are precisely the equations $G_{tt} = G_{tx^i} = 0$. Hence, Γ_μ do in fact vanish to first order on $\mathcal{H}_0$ and hence, the previously constructed g is a solution to $Ric(g) = 0$.

Remarks

1. Note that the condition $\Gamma^\alpha = 0$ is equivalent to $\Box_g x^\alpha = 0$, i.e. the coordinate functions satisfy the wave equation. For this reason, these coordinates are called *wave coordinates* and this choice of coordinates is called *wave gauge*. In the Riemannian case, this is also known as *harmonic gauge*.

2. It is important to emphasize the role of global hyperbolicity to the uniqueness of the Einstein equations, and in fact, of general hyperbolic equations. This is also known as *finite propagation speed*. See Chapter 9 for a detailed discussion about this phenomenon.

3. The local well posednedss being understood (at least in the smooth class), one is interested in understanding global properties of the solutions to the Einstein equation. In general, this is extremely difficult. A priori, however, Choquet–Bruhat and Geroch have shown that, given an initial data set, there exists a uniqne maximal global hyperbolic spacetime (known as the *maximal development*) which solves the Einstein equations. By maximality here we mean the fact that any other globally hyperbolic solution can be isometrically embedded in the maximal development (in fact, for this reason, one should think of the maximal development as the "maximum" development).

3.4 Gravitational Redshift and Time Dilation

Based on qualitative properties of the postulates of relativity, (i.e. without precise knowledge of the spacetimes $(\mathcal{M}, g)$, which satisfy the Einstein equations) we can easily deduce that the proper time of observers depends on the gravitational potential of the region they live in. In particular, the closer an observer is at the gravitational source the slower his proper time passes.

Consider first the following thought experiment. Imagine two observers A, B in an accelerating spaceship such that their acceleration is congruent to the vector $\vec{AB}$. Suppose that A emits a photon towards B. Since photons travel with finite speed it will take some time for this photon to reach B. Since the spaceship is accelerating, at the time of reception of the photon by B, the speed of B is faster than the speed of A at the time of emission of the photon (note, however, that the distance of A, B is constant at all times). Invoking the Doppler effect we infer that B records the photon's frequency on arrival as lower than it was on departure from A. Hence, the frequency recorded by B is redshifted compared to the frequency of the photon at the time of emission from A. Since frequency is a basic measure of time, this implies that the proper time of B runs faster than the time of A as measured by B. In other words, 1sec for A is measured as 0.9secs for B (and hence B finds the time of A to be slower than his time), and conversely, 1sec for B is measured (by A) as 1.1secs for A. The latter is due to the blueshift of photons emitted from B as received by A.

Consider now two observers A, B who are constantly located at distance, say, 1km and 100km, respectively, away from the surface of the earth. In view of the equivalence principle, these two (non-freely falling) observers can be thought of as accelerating observers[2], both being accelerated in the outward direction $\vec{AB}$. Hence, the same principles apply as above (there also exists a relativistic version of the Doppler effect). When light travels from A to B it gets redshifted (as measured by B at the time of reception compared to the measurements of A at the time of emission) and blueshifted when it travels from B to A. This phenomenon is known as *gravitational redshift*. According to the above discussion, the time of B as measured by A is faster than the proper time of A and the time of A as measured by B is slower than the proper time of A. This phenomenon is known as *gravitational time*

[2]The acceleration is needed in order to counteract gravity.

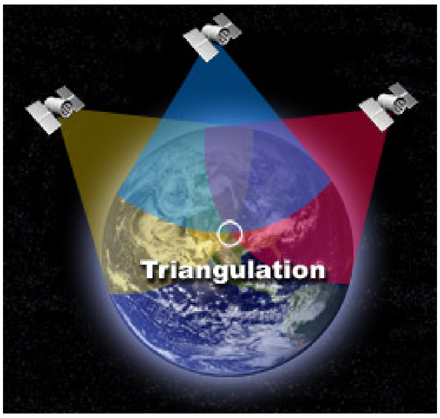
dilation. Confirmation of this effect has been achieved by comparison of atomic clocks flown in airplanes with clocks on the ground.

3.5 Applications

We next describe some predictions of general relativity without involving equations.

1. Global Positioning System (GPS)

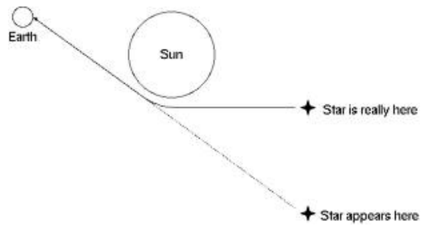
The Global Positioning System uses the effect of relativity in order to make correct measurements. The GPS is a network of over 30 satellites orbiting at distance 20,000km above the surface of the earth and moving at speed 14,000km/h, arranged such that at least four satellites are visible from any point on earth. Each satellite constantly transmits signals containing information on where the satellite is and what time the signal was transmitted. A GPS device on earth (constantly) receives this signal and calculates the distance of the device from the point where the signal was sent using the formula $d = ct$, where d is the distance, c the speed of light and t the time that it took for the signal to reach the device. Performing the same calculation using the signals sent from at least four visible satellites, the device is able to locate its distance from four known points in the space, and hence it can calculate its location. Note that although three satellites suffice, four or more satellites are needed for more accurate results.



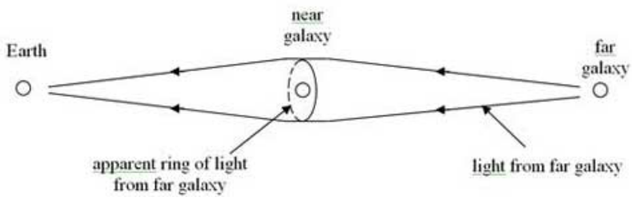
It is of paramount importance that the GPS devices calculate the time that signals had to travel from the satellites to reach the device. However, in view of the gravitational time dilation, the time on the satellites runs 45 microseconds/day faster compared to the time on the surface of the earth. Furthermore, in view of the motion of the satellites compared to the device, the effect special relativistic time dilation gives us that the time on satellites runs 7 microseconds/day slower. When special and general relativistic effects are combined, we deduce that satellite clocks run 38 microseconds/day faster. For this reason, the atomic clocks are adjusted when constructed on earth so they run 38 microseconds/day slower (and hence when on space they agree with the clocks on earth). If uncorrected, the total measurements would be off by 11km/day!

2. Light bending

Null geodesics in a general Lorentzian manifold are curves with non trivial curvature. In particular, according to relativity, a light ray that glazes the surface of the sun deflects from its original path. This bending of starlight passing near our sun has been observed during solar eclipses beginning with the 1919 expedition led by Eddington. This made relativity a very popular theory of gravitation overnight.



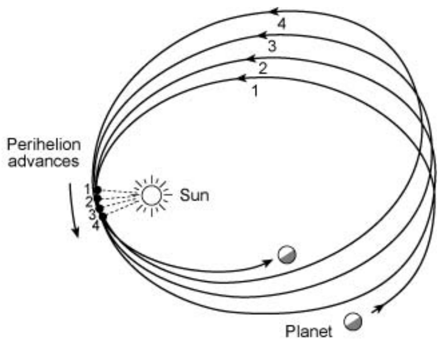
Furthermore, the position of the astronomical objects may be such that a galaxy appears to be in many positions at the same time:



In fact, in the case of black holes, relativity predicts the existence of lights which orbit the black hole. The region where orbiting photons live is called *photon sphere*.

3. Mercury's Perihelion Precession

Recall that the nearest point of the (elliptic) trajectory of a planet to the sun is called perihelion. A number of effects in our solar system cause the perihelia of planets to precess (rotate) around the sun. Mercury deviates from the precession predicted from these Newtonian effects by 43 arcseconds (=43/3600 of a degree) per century. On the other hand, relativity agrees closely with the observed amount of perihelion shift. Note that this effect is small; it requires a little over twelve million orbits for a full excess turn.



Chapter 4

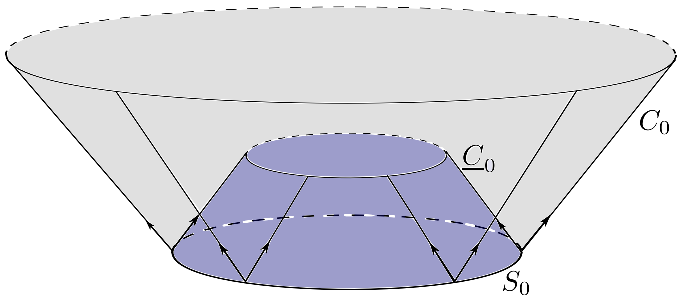
Null Structure Equations

4.1 The Double Null Foliation

Null Foliations and Optical Functions

Consider an open neighborhood of the spacetime $\mathcal{M}$. This region can be foliated by outgoing and incoming null hypersurfaces. This section is concerned with the formal definition of the double null foliation.

Let S_0 be an embedded 2-surface and $C_0, \underline{C}_0$ be the null hypersurfaces spanned by outgoing and incoming null geodesics normal to S_0.



Let $\Omega : S_0 \to \mathbb{R}$ be a smooth function on S_0 and L' be a null vector field normal to S_0 (and tangential to C_0). Let now $\underline{L}'$ be the null vector field normal to S_0 and tangent to the null geodesics of $\underline{C}_0$ such that

$$g(L', \underline{L}') = -\Omega^{-2}.$$

We extend $L', \underline{L}'$ on $C_0, \underline{C}_0$, respectively, such that

$$\nabla_{L'} L' = 0, \qquad \nabla_{\underline{L}'} \underline{L}' = 0.$$

We extend Ω to a function on the hypersurfaces C_0 and $\underline{C}_0$ and consider the vector fields

$$L = \Omega^2 L', \quad \underline{L} = \Omega^2 \underline{L}' \tag{4.1}$$

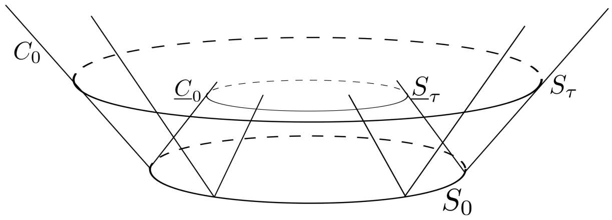
Define the function $\underline{u}$ on C_0 by

$$L\underline{u} = 1, \quad \text{with } \underline{u} = 0 \text{ on } S_0$$

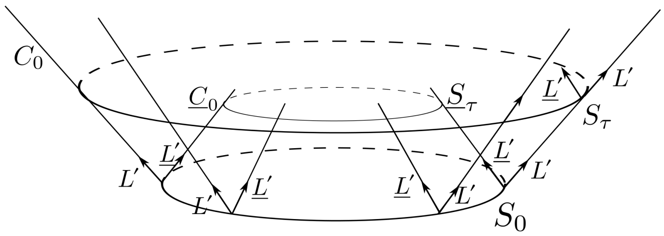
and similarly define the function u on $\underline{C}_0$ by

$$\underline{L}u = 1, \quad \text{with } u = 0 \text{ on } S_0.$$

Let S_τ be the embedded 2-surface on C_0 such that $\underline{u} = \tau$, and similarly, let $\underline{S}_\tau$ be the embedded 2-surface on $\underline{C}_0$ such that $u = \tau$.



We also define $\underline{L}'$ on C_0 such that $\underline{L}'$ is null and normal to S_τ and $g(L', \underline{L}') = -\Omega^{-2}$.



Consider the affinely parametrized null geodesics which emanate from the points on S_τ with initial tangent vector $\underline{L}'$. These geodesics, whose tangent we denote by $\underline{L}'$, span (in view of Proposition 2.2.2) null hypersurfaces which we will denote by $\underline{C}_\tau$. Hence, $C_0 \cap \underline{C}_\tau = S_\tau$. Similarly, we define L' globally (and the hypersurfaces C_τ such that their normal is L'). Extend the vector field $L, \underline{L}$ to global vector fields such that

$$L = \Omega^2 L', \quad \underline{L} = \Omega^2 \underline{L}'.$$

We will refer to Ω as the null lapse function. We also extend the functions $u, \underline{u}$ to global functions such that

$$Lu = 0, \qquad \underline{L}\,\underline{u} = 0.$$

Therefore,

$$C_\tau = \{u = \tau\}, \quad \underline{C}_\tau = \{\underline{u} = \tau\},$$

and hence $u, \underline{u}$ are optical functions. The importance of the renormalized vector fields $L, \underline{L}$ is manifest from the following

Proposition 4.1.1. *The optical functions $u, \underline{u}$ satisfy the following relations*

$$\nabla \underline{u} = -\underline{L}', \qquad \nabla u = -L'$$

and

$$L\underline{u} = 1, \qquad \underline{L}u = 1.$$

Proof. Since $u, \underline{u}$ are optical functions, by the results of Section 2.2 we have that $\nabla u, \nabla \underline{u}$ satisfy the geodesic equation. Since $L', \underline{L}'$ satisfy the geodesic equation as well, it suffices to show, for instance, that $\nabla \underline{u} = -\underline{L}'$ on C_0. Expressing $\nabla \underline{u}$ in terms of the null frame $(L', \underline{L}', e_1, e_2)$, where e_1, e_2 is a local frame on the spheres S_τ, we obtain

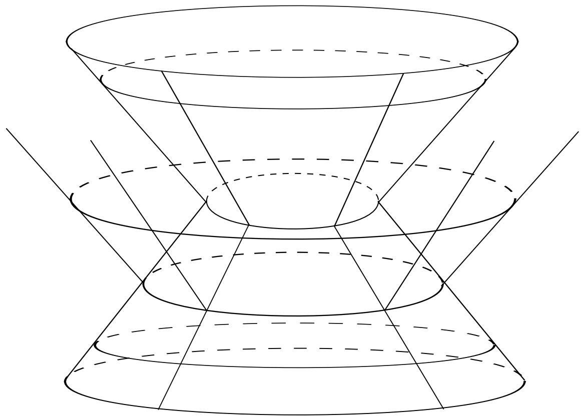
$$(\nabla \underline{u})^{\underline{L}'} = g^{\underline{L}'L'} \cdot (L'\underline{u}) + g^{\underline{L}'\underline{L}'} \cdot (\underline{L}'\underline{u}) = -\Omega^2 \cdot \Omega^{-2} = -1,$$

on C_0, and similarly, we obtain that the remaining components with respect to the above frame are zero. This proves the first equation. For the second, it suffices to notice that

$$L\underline{u} = g(L, \nabla \underline{u}) = g(L, -\underline{L}') = -\Omega^2 \cdot g(L', \underline{L}') = 1.$$

□

In this way, we foliate the spacetime with null hypersurfaces (being the level sets of $u, \underline{u}$) as depicted below



Gauge Freedom

The above analysis implies that the vector field L' on S_0 and the function Ω on $C_0 \cup \underline{C}_0$ can be freely chosen. This freedom reflects the freedom for the functions $u' = u'(u), \underline{u}' = \underline{u}'(\underline{u})$. Fixing Ω and L' determines up to additive constants the (optical) functions $u, \underline{u}$.

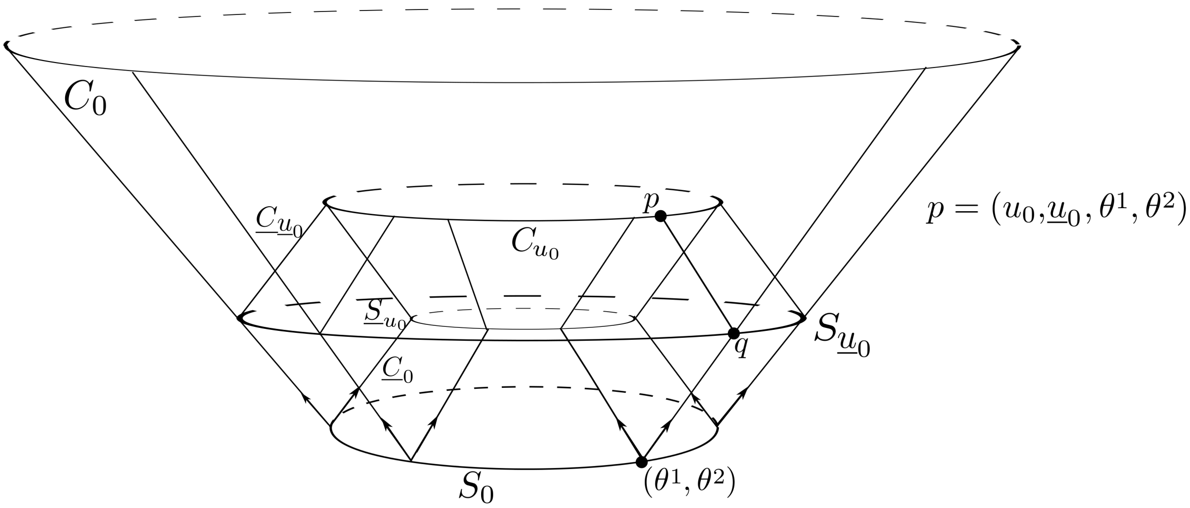
The Canonical Coordinate System

We saw above that we can construct at each point a *null frame* $(L, \underline{L}, e_1, e_2)$ adapted to the double null foliation (recall that e_1, e_2 is a local frame for the spheres $S_{u,\underline{u}}$). However, these vector fields do not arise from a coordinate system.

Using the optical functions $u, \underline{u}$ we will introduce a coordinate system suitably adapted to the corresponding double null foliation of the spacetime.

If $p \in \mathcal{M}$ then $p \in C_{u_0} \cap \underline{C}_{\underline{u}_0}$ and hence $u(p) = u_0, \underline{u}(p) = \underline{u}_0$. We next prescribe angular coordinates for the point p on the 2-surface $C_{u_0} \cap \underline{C}_{\underline{u}_0}$.

Let (θ^1, θ^2) denote a coordinate system on a domain of S_0. Suppose that the null generator of $\underline{C}_{\underline{u}_0}$ through p intersects C_0 at the point q and that the null generator of C_0 through the point q intersects the sphere S_0 at the point with coordinates (θ^1, θ^2). Then we assign to p the angular coordinates (θ^1, θ^2). Hence, the point p corresponds to the spacetime coordinates $(u_0, \underline{u}_0, \theta^1, \theta^2)$.



By construction we have everywhere:

$$\frac{\partial}{\partial u} = \underline{L}$$

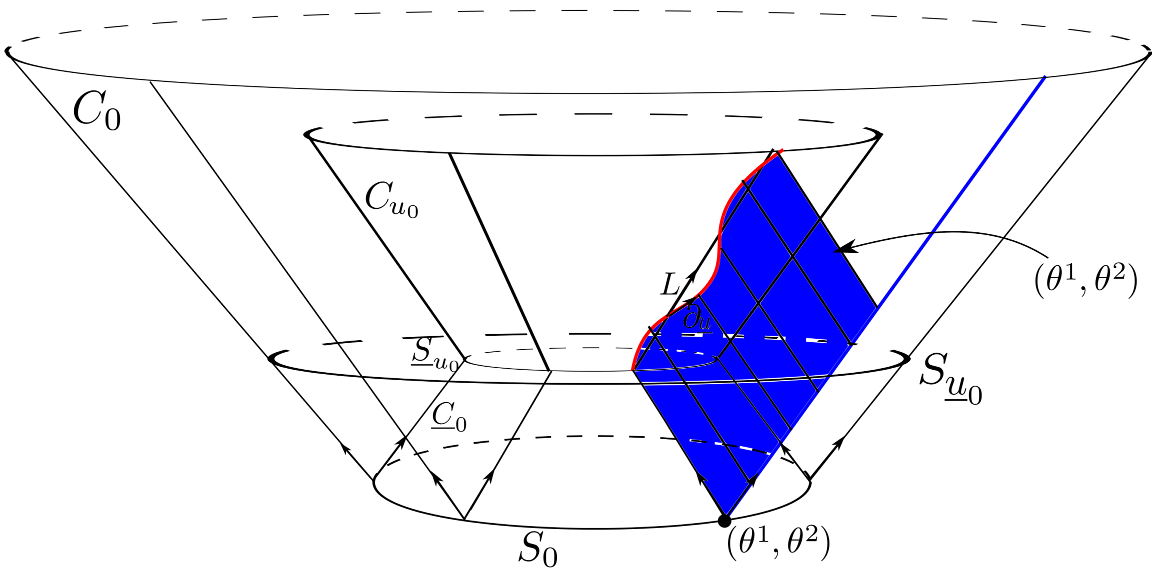
and

$$\frac{\partial}{\partial \theta^1}, \frac{\partial}{\partial \theta^2} \in TS_{u,\underline{u}},$$

whereas

$$\frac{\partial}{\partial \underline{u}} = L : \text{ on } C_0.$$

Note that the latter equation will not in general hold everywhere, as it is easily seen from the picture below:



From now on, for simplicity we denote $\partial_{\underline{u}} = \frac{\partial}{\partial \underline{u}}$, and so on. In general we have:

$$\partial_{\underline{u}} = L + b^i \partial_{\theta^i}.$$

By virtue of the equations $\underline{L} = \partial_u$ and $[\partial_u, \partial_{\underline{u}}] = [\partial_u, \partial_{\theta^i}] = 0$ we obtain:

$$[L, \underline{L}] = -\frac{\partial b^i}{\partial u}\partial_{\theta^i} \in TS_{u,\underline{u}},$$

and therefore,

$$\frac{\partial b^i}{\partial u} = -d\theta^i([L, \underline{L}]), \text{ and } b^i = 0 \text{ on } C_0 = \{u = 0\}\,. \tag{4.2}$$

Hence, the S-tangent vector field $b = b^i\partial_{\theta^i}$ is the obstruction to the integrability of $\langle L, \underline{L}\rangle = \big(TS_{u,\underline{u}}\big)^{\perp}$. In order to compute b it suffices to compute $[L, \underline{L}]$. Since $[L, \underline{L}] = \nabla_L \underline{L} - \nabla_{\underline{L}} L$, it suffices to compute the connection coefficients; we will do this in the next section.

The metric g with respect to the canonical coordinates is given by

$$g = -2\Omega^2 du d\underline{u} + (b^i b^j \not{g}_{ij}) dv d\underline{u} - 2(b^i \not{g}_{ij}) d\theta^j d\underline{u} + \not{g}_{ij}\, d\theta^i d\theta^j, \tag{4.3}$$

where $\not{g}$ denotes the induced metric on the 2-surfaces $S_{u,\underline{u}} = C_u \cap \underline{C}_{\underline{u}}$. We immediately obtain

$$\det(g) = -\Omega^2 \cdot \det(\not{g}\,). \tag{4.4}$$

Null Frames

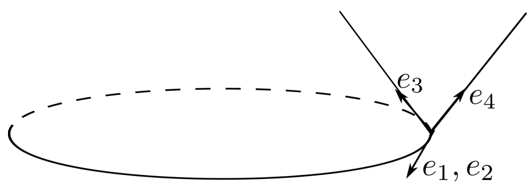
From now on we denote $S_{u,\underline{u}} = C_u \cap \underline{C}_{\underline{u}}$. If $\{e_1, e_2\} = \big(e_a\big)_{a=1,2}$ is an arbitrary frame on the spheres $S_{(u,\underline{u})}$ then we, in fact, have the following null frames:

- **Geodesic frame:** $(e_1, e_2, L', \underline{L}')$,
- **Equivariant frame:** $(e_1, e_2, L, \underline{L})$,
- **Normalized frame:** (e_1, e_2, e_3, e_4),

where

$$e_3 = \Omega \underline{L}', \qquad e_4 = \Omega L'.$$

We symmetrically rescaled $L', \underline{L}'$ above so that the equations take a symmetric (dual) form in the 3 and 4 directions.



Note that e_3, e_4 satisfy the normalization property

$$g(e_3, e_4) = -1. \tag{4.5}$$

4.2 Connection Coefficients

We consider the **normalized frame** (e_1, e_2, e_3, e_4) defined above. We define the connection coefficients with respect to this frame to be the smooth functions $\Gamma^{\lambda}_{\mu\nu}$ such that

$$\nabla_{e_\mu} e_\nu = \Gamma^{\lambda}_{\mu\nu} e_\lambda, \quad \lambda, \mu, \nu \in \{1, 2, 3, 4\}$$

Here ∇ denotes the connection of the spacetime metric g. We are mainly interested in the case where at least one of the indices λ, μ, ν is either 3 or 4 (otherwise, we obtain the Christoffel symbols with respect to the induced metric $\not{g}$). These coefficients are completely determined by the following components:

The components $\chi, \underline{\chi}, \eta, \underline{\eta}, \omega, \underline{\omega}, \zeta$:

$$\begin{aligned} \chi_{AB} &= g(\nabla_A e_4, e_B), \qquad & \underline{\chi}_{AB} &= g(\nabla_A e_3, e_B), \\ \eta_A &= g(\nabla_3 e_4, e_A), & \underline{\eta}_A &= g(\nabla_4 e_3, e_A), \\ \omega &= -g(\nabla_4 e_4, e_3), & \underline{\omega} &= -g(\nabla_3 e_3, e_4), \\ & \zeta_A = g(\nabla_A e_4, e_3) \end{aligned} \tag{4.6}$$

where $A, B \in \{1, 2\}$ and $\nabla_\mu = \nabla_{e_\mu}$. Note that $\underline{\zeta} = -\zeta$. The covariant tensor fields $\chi, \underline{\chi}, \eta, \underline{\eta}, \zeta$ are only defined on $T_x S_{u,v}$. **We can naturally extend these to tensor fields to be defined on $T_x\mathcal{M}$ by simply letting their value to be zero if they act on e_3 or e_4. Such tensor fields will in general be called <u>S-tensor fields</u>. Note that a vector field is an S-vector field if it is tangent to the spheres $S_{u,\underline{u}}$.** The importance the S-tensors originates from the fact that one is interested in understanding the embedding of $S_{u,\underline{u}}$ in C_u and $\underline{C}_{\underline{u}}$.

The connection coefficients Γ can be recovered by the following relations:

$$\begin{aligned} & \nabla_A e_B = \not\nabla_A e_B + \chi_{AB} e_3 + \underline{\chi}_{AB} e_4, \\ \nabla_3 e_A &= \not\nabla_3 e_A + \eta_A e_3, \quad & \nabla_4 e_A &= \not\nabla_4 e_A + \underline{\eta}_A e_4, \\ \nabla_A e_3 &= \underline{\chi}_A{}^{\sharp B} e_B + \zeta_A e_3, & \nabla_A e_4 &= \chi_A{}^{\sharp B} e_B - \zeta_A e_4, \\ \nabla_3 e_4 &= \eta^{\sharp A} e_A - \underline{\omega} e_4, & \nabla_4 e_3 &= \underline{\eta}^{\sharp A} e_A - \omega e_3, \\ \nabla_3 e_3 &= \underline{\omega} e_3, & \nabla_4 e_4 &= \omega e_4, \end{aligned} \tag{4.7}$$

Remarks:

1. Recall, that the second fundamental form of a manifold S embedded in a manifold $\mathcal{M}$ is defined to be the symmetric $(0, 2)$ tensor field II such that for each $x \in S$ we have

$$II_x : T_x S \times T_x S \to (T_x S)^\perp,$$

where the $\perp$ is defined via the decomposition $T_x\mathcal{M} = T_x S \oplus (T_x S)^\perp$. Specifically, if $X, Y \in T_x S$ then

$$II_x(X, Y) = \big(\nabla_X Y\big)^\perp$$

and hence

$$\nabla_X Y = \not\nabla_X Y + II(X, Y).$$

Here $\nabla\!\!\!/$ denotes the induced connection on S (which is taken by projecting the spacetime connection ∇ on T_xS).

The S-tensor fields $\chi, \underline{\chi}$ give us the projections of II_{AB} on e_3 and e_4, respectively. Indeed

$$II(X,Y) = \chi(X,Y)e_3 + \underline{\chi}(X,Y)e_4.$$

For this reason we will refer to $\chi, \underline{\chi}$ as the *null second fundamental forms* of $S_{u,\underline{u}}$ with respect to the null hypersurfaces $C_u, \underline{C}_{\underline{u}}$, respectively. One can easily verify that χ and $\underline{\chi}$ are symmetric (0,2) S-tensor fields. Indeed, a simple calculation shows that if X, Y are S-tangent vector fields then

$$\chi(X,Y) - \chi(Y,X) = g(e_4, [X,Y]),$$
$$\underline{\chi}(X,Y) - \underline{\chi}(Y,X) = g(e_3, [X,Y]).$$

Hence, $\chi, \underline{\chi}$ are symmetric if and only if $[X,Y] \perp e_3$ and $[X,Y] \perp e_4$ and thus if and only if $\langle e_3, e_4\rangle^{\perp} \ni [X,Y] \in TS_{u,\underline{u}}$. The symmetry of $\chi, \underline{\chi}$ is thus equivalent to the integrability of the orthogonal complement $\langle e_3, e_4\rangle^{\perp}$.

Furthermore, we can decompose χ and $\underline{\chi}$ into their trace and traceless parts by

$$\chi = \hat{\chi} + \frac{1}{2}(tr\chi)g\!\!\!/\,, \qquad \underline{\chi} = \hat{\underline{\chi}} + \frac{1}{2}(tr\underline{\chi})g\!\!\!/\,. \tag{4.8}$$

The trace of the S-tensor fields $\chi, \underline{\chi}$ (and more general S-tensor fields) is taken with respect to the induced metric $g\!\!\!/$. The trace $tr\chi$ is known as the *expansion* and the component $\hat{\chi}$ is called the *shear* of $S_{u,\underline{u}}$ with respect to C_u.

2. Note also that $\omega = \nabla_4(\log\Omega) =$ and $\underline{\omega} = \nabla_3(\log\Omega)$.

3. Let X be a vector tangential to a given sphere $S_{u,\underline{u}}$ at a point x. Then, if we extend X along the null generator γ of C_u passing through x according to the Jacobi equation $[L, X] = [\Omega e_4, X] = 0$, then we obtain an S-tangent vector field along γ. Note that in this case we obtain

$$\nabla_4 X = \nabla_X e_4 + \big(\nabla_X \log\Omega\big)e_4.$$

On the other hand, if we simply extend X such that $[e_4, X] = 0$ then, although X will be tangential to C_u, X will not be tangential to the sections $S_{u,\underline{u}}$ of C_u. This is because the sections the optical functions $u, \underline{u}$ are the affine parameters of the vector fields $\underline{L}, L$, respectively.

4. The S 1-form ζ is known as the *torsion*. If $d\!\!\!/$ denotes the exterior derivative on $S_{u,\underline{u}}$ then the S 1-forms $\eta, \underline{\eta}$ are related to ζ via

$$\eta = \zeta + d\!\!\!/(\log\Omega), \qquad \underline{\eta} = -\zeta + d\!\!\!/\log\Omega.$$

Proof. Let X be S-tangential and extend along the null generator of C_u according to the Jacobi equation, then

$$\underline{\eta}(x) = g(\nabla_4 e_3, X) = -g(e_3, \nabla_4 X) = -g(e_3, \nabla_X e_4) + \nabla_X(\log\Omega) = -\zeta(X) + \nabla_X \log\Omega. \tag{4.9}$$

We similarly show the analogous relation for η. $\square$

The above imply also that

$$\zeta = \frac{1}{2}(\eta - \underline{\eta}), \quad \nabla\!\!\!/ \log\Omega = \frac{1}{2}(\eta + \underline{\eta}).$$

The 1-forms $\eta, \underline{\eta}$ can be thought of as the torsion of the null hypersurfaces with respect to the geodesic vector fields. Indeed, the previous relations imply

$$\eta_A = \Omega^2 g(\nabla_A L', \underline{L}').$$

5. We have

$$\begin{aligned}
[L, \underline{L}] &= \nabla_L \underline{L} - \nabla_{\underline{L}} L = \Omega \cdot \Big(\nabla_4(\Omega e_3) - \nabla_3(\Omega e_4)\Big) \\
&= \Omega^2 \cdot \Big(\nabla_4 e_3 - \nabla_3 e_4 + (\nabla_4 \log\Omega) e_3 - (\nabla_3 \log\Omega) e_4\Big) \\
&= \Omega^2 \cdot \Big((\underline{\eta}^{\sharp A} - \eta^{\sharp A}) e_A - \omega e_3 + \underline{\omega} e_4 + (\nabla_4 \log\Omega) e_3 - (\nabla_3 \log\Omega) e_4\Big) \\
&= \Omega^2 \cdot \Big(\underline{\eta}^\sharp - \eta^\sharp\Big) \\
&= -2\Omega^2 \zeta^\sharp.
\end{aligned}$$

The above completely determines the vector field b defined in (4.2). In particular, it shows that $\frac{\partial b^i}{\partial u} = 2\Omega^2 (\zeta^\sharp)^i$ and hence the torsion ζ is the obstruction to the integrability of the timelike planes $\langle e_3, e_4 \rangle$ orthogonal to the spheres $S_{u,\underline{u}}$.

6. It is easy to show that for the standard spheres of radius r in the Minkowski spacetime we have

$$\hat{\chi} = \hat{\underline{\chi}} = 0, \quad tr\chi = \frac{2}{r}, \quad tr\underline{\chi} = -\frac{2}{r}. \tag{4.10}$$

Indeed, using the double null coordinate system we easily obtain, for example,

$$\chi(\partial_\theta, \partial_\theta) = \chi(\partial_\phi, \partial_\phi) = \frac{1}{r}, \quad \chi(\partial_\theta, \partial_\phi) = 0.$$

Note that $tr\underline{\chi} \to -\infty$ as $r \to 0$ (that is to say, the incoming expansion diverges we approach the vertex). For a generalization of this fact, see Section 5.2.

4.3 Curvature Components

We next decompose the Riemann curvature R in terms of the normalized null frame. First, we define the following components, which contain at most two S-tangential components (and hence at least 2 null components):

$$\begin{aligned}
\alpha_{AB} = R_{A4B4}, &\quad \underline{\alpha}_{AB} = R_{A3B3}, \\
\beta_A = R_{A434}, &\quad \underline{\beta}_A = R_{A334}, \\
\rho = R_{3434}, &\quad \sigma = \frac{1}{2}\epsilon\!\!/^{AB} R_{AB34}.
\end{aligned}$$

Note that $R(\cdot, \cdot, e_3, e_4)$, when restricted on $T_x S_{u,\underline{u}}$, is an antisymmetric form and hence collinear to the volume form $\epsilon\!\!/$ on $S_{u,\underline{u}}$. Furthermore, if

$$(*R)_{3434} = *\rho,$$

then $*\rho = 2\sigma$. Here, the dual $*R$ of the Riemann curvature is defined to be the (0,4) tensor:

$$(*R)_{\alpha\beta\gamma\delta} = \epsilon_{\mu\nu\alpha\beta}\, R^{\mu\nu}{}_{\gamma\delta}.$$

Clearly, the (0,2) S-tensor fields $\alpha, \underline{\alpha}$ are symmetric.

Proposition 4.3.1. *Suppose that $Ric = 0$. Then,*

$$tra = 0, tr\underline{a} = 0,$$

where the trace (of the S-tensor fields) is taken with respect to $\not{g}$, and the remaining components (with at least three spherical components or $R(\cdot, e_3, \cdot, e_4)$) are given by the following expressions:

$$\begin{aligned}
R_{A3BC} &= \not{g}_{AB}\underline{\beta}_C - \not{g}_{AC}\underline{\beta}_B,\\
R_{A4BC} &= -\not{g}_{AB}\beta_C + \not{g}_{AC}\beta_B,\\
R_{A3B4} &= \frac{1}{2}\sigma\not{\epsilon}_{AB} - \frac{1}{2}\rho\not{g}_{AB},\\
R_{ABCD} &= 2\rho(\not{g}_{AD}\not{g}_{BC} - \not{g}_{AC}\not{g}_{BD}).
\end{aligned}$$

Remark 4.3.1. *It is important to recall the basic properties of the volume form. If ϵ is the (spacetime) volume form with respect to g, and $\not{\epsilon}$ is the volume form on the spheres $S_{u,\underline{u}}$ with respect to the induced metric $\not{g}$, then*

$$\epsilon_{1234} = \epsilon(e_1, e_2, e_3, e_4) = 1 \text{ and } \epsilon_{AB34} = \not{\epsilon}_{AB}.$$

We note that, when working with $\not{\epsilon}$ it is very convenient to consider a (local) orthonormal S-frame e_1, e_2 on the spheres $S_{u,\underline{u}}$.

Hodge dual relative to $S_{u,\underline{u}}$

- *The (left) dual $*\xi$ of an S 1-form ξ is defined to be the following S 1-form*

  $$*\xi = \not{\epsilon}(\cdot, \xi^\sharp),$$

  or, using the frame:

  $$(*\xi)_A = \not{\epsilon}_{AB}\, \xi^{\sharp B}.$$

  Note also that

  $$**\xi = -\xi.$$

- *The dual $*\theta$ of an (0,2) S-tensor field θ is defined to be the following (0,2) $S-$tensor field*

  $$*\theta = \not{\epsilon}(\cdot, \theta^\sharp(\cdot)),$$

  or, using the frame:

  $$(*\theta)_{AB} = \not{\epsilon}_{AC}\, \theta^{\sharp C}{}_B.$$

  Note that $*\theta$ is symmetric if θ is traceless and $*\theta$ is traceless if θ is symmetric.

Proof of Proposition 4.3.1. We have $Ric_{\mu\nu} = (g^{-1})^{\kappa\lambda} R_{\kappa\mu\lambda\nu}$. The inverse metric g^{-1} can be written in terms of the normalized frame as follows:

$$g^{-1} = -e_3 \otimes e_4 - e_4 \otimes e_3 + \not{g}^{-1}.$$

Hence,

$$Ric_{\mu\nu} = -R_{3\mu4\nu} - R_{4\mu3\nu} + (\not{g}^{-1})^{CD} R_{C\mu D\nu}.$$

1. The equations $Ric_{33} = Ric_{44} = 0$

Note that $R_{44} = (g^{-1})^{CD} R_{C4D4} = tr\alpha$ and similarly, $R_{33} = tr\underline{\alpha}$.

2. The equations $Ric_{3A} = Ric_{4A} = 0$

We have $R_{4B} = -R_{344B} + (\not{g}^{-1})^{CA} R_{C4AB}$ and therefore,

$$(\not{g}^{-1})^{CA} R_{C4AB} = -\beta_B.$$

Observe now that $R(e_C, e_4, \cdot, \cdot)$ is an (0,2) antisymmetric S-form on $T_x S_{u,\underline{u}}$ and hence it is proportional to the induced volume form $\not{\epsilon}$, hence:

$$R_{C4AB} = \xi_C \not{\epsilon}_{AB}$$

where ξ is an S 1-form. Then,

$$(\not{g}^{-1})^{CA} R_{C4AB} = \xi_C \not{\epsilon}^{\sharp C}{}_{B} = -(*\xi)_B.$$

Therefore,

$$*\xi = \beta \Longrightarrow \xi = -(*\beta)$$

and

$$R_{C4AB} = -(*\beta)_C \not{\epsilon}_{AB} = -\not{g}_{CA}\beta_B + \not{g}_{CB}\beta_A.$$

3. The equations $Ric_{AB} = Ric_{34} = 0$

A reasoning similar as above implies that there exists a function f such that

$$R_{ABCD} = f \not{\epsilon}_{AB} \not{\epsilon}_{CD}.$$

Since $R_{AB} = 0$ we obtain

$$R_{3A4B} + R_{4A3B} = (\not{g}^{-1})^{CD} R_{CADB} = f \not{g}_{AB} \tag{4.11}$$

and therefore, taking the trace (and using that $tr\not{g} = \not{g}^{AB}\not{g}_{AB} = 2$) we obtain

$$f = \frac{1}{2}(\not{g}^{-1})^{AB}(R_{3A4B} + R_{4A3B}) = (\not{g}^{-1})^{AB} R_{A3B4}.$$

However, equation $Ric_{34} = 0$ gives us

$$(\not{g}^{-1})^{AB} R_{A3B4} = R_{4334} = -\rho \tag{4.12}$$

and hence $f = -\rho$, which in turn determines R_{ABCD}. Finally, we compute the (0,2) S-tensor field R_{A3B4}. For we consider its symmetric and antisymmetric parts:

$$R_{A3B4} + R_{B3A4} = -\rho \not{g}_{AB},$$
$$R_{A3B4} - R_{B3A4} = R_{AB34} = \sigma \not{\epsilon}_{AB},$$

where, in the last line above, we used the first Bianchi identity.

□

4.4 The Algebra Calculus of S-Tensor Fields

Consider a spacetime $(\mathcal{M}, g)$ and assume that it admits a double null foliation of spheres $S_{u,\underline{u}}$. Let $\not g$ be the induced metric on $S_{u,\underline{u}}$. Let $\not\epsilon$ denote the induced volume form on $S_{u,\underline{u}}$.

Definition of S-tensor fields:

A 1-form ξ of $\mathcal{M}$ is an S 1-form if $\xi(e_3) = \xi(e_4) = 0$. In other words, xi is specified by a smooth assignment of 1-form on $S_{u,\underline{u}}$ for each $S_{u,\underline{u}}$. Similarly, a p-covariant tensor θ on $\mathcal{M}$ is an S-tensor if θ gives the value zero whenever it acts on either e_3 or e_4. A vector field X is an S-tangent vector field (or S-vector field) if it is tangent to $S_{u,\underline{u}}$.

Musical isomorphisms

The (positive definite) metric $\not g$ induces canonical, known as musical, isomorphisms between the tangent bundle $TS_{u,\underline{u}}$ and the cotangent bundle $T^*S_{u,\underline{u}}$. Specifically, given a S-vector field X we define the S 1-form $X_\flat$ such that $X_\flat(Y) = \not g(X, Y)$, for all S-vector fields Y. Similarly, given an S 1-form ω we define the S-vector field $\omega^\sharp$ such that $\not g(\omega^\sharp, Y) = \omega(Y)$ for all S-vector fields Y.

The isomorphisms $\flat, \sharp$ can be extended to more general S-tensor fields. Let, for example, $T : TS_{u,\underline{u}} \to TS_{u,\underline{u}}$ be a $(1,1)$ tensor field. Then $T_\flat$ is a $(0,2)$ tensor field, given by

$$T_\flat(X, Y) = (T(X))_\flat(Y) = g(T(X), Y).$$

More generally, if $T \in TS_{u,\underline{u}} \otimes^n T^*S_{u,\underline{u}}$ is a tensor field of type $(1,n)$ then $T_\flat \in \otimes^{n+1}T^*S_{u,\underline{u}}$ is a tensor field of type $(0, n+1)$ given by

$$T_\flat(x_1, x_2, ..., X_n, Y) = (T(X_1, X_2, ..., X_n))_\flat(Y) = g(T(X_1, X_2, ..., X_n), Y).$$

On the other hand, if $T \in \otimes^2 T^*S_{u,\underline{u}}$ is of type $(0,2)$, then $T^\sharp \in TS_{u,\underline{u}} \otimes T^*S_{u,\underline{u}} = \mathrm{End}(TS_{u,\underline{u}})$ is of type $(1,1)$ such that if $X \in TS_{u,\underline{u}}$ then

$$T^\sharp(X) = (T(X, \cdot))^\sharp \in TS_{u,\underline{u}}.$$

Note that one could also define

$$T^\sharp(X) = (T(\cdot, X))^\sharp \in TS_{u,\underline{u}},$$

and these two definitions coincide if and only if T is symmetric. Otherwise one needs to prescribe explicitly how $\sharp$ acts, as above.

Trace

The trace of a (1,1) S-tensor field T is defined to be the contraction of the matrix of T in an arbitrary basis. If E_i is an S orthonormal frame then

$$trT = \sum_{i=1}^{n} \not g(T(E_i), E_i).$$

Given now a $(0,2)$ tensor field T, we define

$$trT = trT^\sharp.$$

In view of the fact that $T^\sharp$ depends on the metric $\not{g}$, the above trace is also called *metric trace*. For $(0,2)$ S-tensors one does not need to prescribe how $\sharp$ acts since

$$tr\Big(X \mapsto \big(T(X,\cdot)\big)^\sharp\Big) = tr\Big(X \mapsto \big(T(\cdot,X)\big)^\sharp\Big).$$

This can also be seen as the full contraction of the tensor product of the inverse metric $\not{g}^{-1}$ and T, i.e. $trT = (\not{g}^{-1})^{AB}T_{AB}$.

For higher order S-tensor fields, the trace is not uniquely defined and so one needs to prescribe how $\sharp$ acts (and hence which indices to raise and contract). We usually consider tensor fields $T(X,Y,Z,W)$ of type at most $(0,4)$. In that case, the trace in, say, X,Z variables is

$$\big(tr_{X,Z}T\big)(Y,W) = \sum_{i=1}^{n} T(E_i,Y,E_i,W).$$

Operations

Let θ,ϕ be (0,2) symmetric S-tensor fields. We define the (0,2) S-tensor field $\theta\times\phi$ such as

$$\theta\times\phi = \not{g}(\theta^\sharp(\cdot),\phi^\sharp(\cdot)), \qquad \text{i.e.} \qquad (\theta\times\phi)_{AB} = \not{g}_{CD}\theta_A^{\sharp C}\phi_B^{\sharp D} = \theta_{AC}\phi_B^{\sharp C}.$$

Note that $(\theta\times\phi)(X,Y) = (\phi\times\theta)(Y,X) := \widetilde{(\phi\times\theta)}(X,Y)$, where $\tilde{}$ denotes transposition. For such S-tensors we also define the functions (θ,ϕ) and $\theta\wedge\phi$ to be

$$(\theta,\phi) = tr(\theta\times\phi) = \theta_{AB}\phi^{\sharp\sharp AB}$$

$$\theta\wedge\phi = \frac{1}{2}\not{\epsilon}^{AB}(\theta\times\phi - \phi\times\theta)_{AB}.$$

For example, $\not{g}\times\theta = \theta$ and hence $(\not{g},\not{g}) = tr(\not{g}\times\not{g}) = tr\not{g} = 2$.

Differential operators

The exterior derivative on $S_{u,\underline{u}}$ is denoted by $\not{d}$. If ξ is an S 1-form then we define the curlξ to be the following function

$$\mathrm{curl}\,\xi = *\not{d}\,\xi,$$

where $*$ denotes the Hodge dual.

If T is an arbitrary S-tensor field, then we denote the induced gradient of T on $S_{u,\underline{u}}$ by $\not{\nabla}T$ and we extend $\not{\nabla}T$ to be an S-tensor field. Here $\not{\nabla}$ should be viewed as the connection on $S_{u,\underline{u}}$ associated to $\not{g}$.

The divergence of an S-vector field T is given by

$$\not{\mathrm{div}}T = tr(\not{\nabla}T) = tr\Big(X \mapsto \not{\nabla}_X T\Big),$$

whereas the divergence of an S 1-form ξ is given by

$$\not{\mathrm{div}}\xi = tr(\not{\nabla}\xi) = tr\Big(X \mapsto \big(\not{\nabla}_X\xi\big)^\sharp\Big),$$

More generally, if T is of type $(0,n)$ then $\mathrm{d\!/\!iv}T$ is of type $(0,n-1)$:

$$\begin{aligned}\mathrm{d\!/\!iv}T(Y_1,...,Y_{n-1}) &= tr\Big(X \mapsto \Big((\nabla\!\!\!\!/_X T)(\cdot,Y_1,...,Y_{n-1})\Big)^\sharp\Big)\\ &= \sum_{i=1}^n \Big(\nabla\!\!\!\!/_{E_i} T\Big)(E_i,Y_1,...Y_n),\end{aligned}$$

where $E_i, i=1,...,n$ form an orthonormal frame.

Let T be p-covariant S-tensor field. We define the projected covariant derivative $\nabla\!\!\!\!/_3 T$ to be the S-tensor field with the property

$$(\nabla\!\!\!\!/_3 T)(X_1,X_2,\ldots,X_p) = (\nabla_3 T)(X_1,X_2,\ldots,X_p),$$

for all S-vector fields X. Similarly we define $\nabla\!\!\!\!/_4 T$. Note that $\nabla\!\!\!\!/_3 g\!\!\!/ = \nabla\!\!\!\!/_4 g\!\!\!/ = 0$ and hence

$$\nabla\!\!\!\!/_3(trT) = tr(\nabla\!\!\!\!/_3 T), \qquad \nabla\!\!\!\!/_4(trT) = tr(\nabla\!\!\!\!/_4 T).$$

If ξ is an S 1-form and X is an S-vector field then

$$\nabla\!\!\!\!/_3(\xi^\sharp) = \Big(\nabla_3\xi\Big)^\sharp, \quad \nabla\!\!\!\!/_3(X_\flat) = \Big(\nabla\!\!\!\!/_3 X\Big)_\flat.$$

Let T be p-covariant S-tensor field. We define the projected lie derivative $\mathcal{L}\!\!\!/_3 T$ to be the S-tensor field with the property

$$(\mathcal{L}\!\!\!/_3 T)(X_1,X_2,\ldots,X_p) = (\mathcal{L}_3 T)(X_1,X_2,\ldots,X_p),$$

for all S-vector fields X. Similarly we define $\mathcal{L}\!\!\!/_4 T$. The importance of $\mathcal{L}\!\!\!/_4$ lies on the fact that, unlike $\nabla\!\!\!\!/_4$, it depends only on the differential structure of C_u.

In view of (4.7), and since $\mathcal{L}_X Y = \nabla_X Y - \nabla^Y X$, we obtain:

$$\begin{aligned}\mathcal{L}\!\!\!/_4\xi &= \nabla\!\!\!\!/_4\xi + \chi^\sharp\cdot\xi, \text{ if } \xi \text{ is an } S \text{ 1-form},\\ \mathcal{L}\!\!\!/_4\theta &= \nabla\!\!\!\!/_4\theta + \chi\times\theta + \theta\times\chi, \text{ if } \theta \text{ is an } S \text{ (0,2) tensor field},\end{aligned}$$

Finally, if ξ is an S 1-form we define the symmetrized traceless covariant derivative $\nabla\!\!\!\!/\hat{\otimes}\xi$ by the following

$$\nabla\!\!\!\!/\hat{\otimes}\xi = \nabla\!\!\!\!/\xi + \widetilde{\nabla\!\!\!\!/\xi} - (\mathrm{d\!/\!iv}\xi)g\!\!\!/ = \widehat{\mathcal{L}\!\!\!/_{\xi^\sharp} g\!\!\!/},$$

where $\widehat{\mathcal{L}\!\!\!/_{\xi^\sharp} g\!\!\!/}$ denotes the traceless part of $\mathcal{L}\!\!\!/_{\xi^\sharp} g\!\!\!/ := \mathcal{L}_{\xi^\sharp} g\!\!\!/$ on $S_{u,\underline{u}}$. It is easy to verify that $-\frac{1}{2}\nabla\!\!\!\!/\hat{\otimes}$ is the adjoint of the restriction of $\mathrm{d\!/\!iv}$ on symmetric and traceless (0,2) S-tensor fields.

4.5 Null Structure Equations

The ultimate goal is to understand the geometry $g\!\!\!/$ of the spheres $S_{u,\underline{u}}$ and also their embedding in C_u, $\underline{C}_v$, and more generally, in $\mathcal{M}$. For this reason, in this section, we will derive equations for the connection coefficients Γ defined in Section 4.2. Specifically, we will derive propagation equations along the null hypersurfaces $C_u, \underline{C_{\underline{u}}}$ and also elliptic equations on $S_{u,\underline{u}}$.

The propagation equations are of the following general form

$$\nabla\!\!\!/_4\Gamma = R + \Gamma\cdot\Gamma + D\!\!\!\!/\,\Gamma,$$
$$\nabla\!\!\!/_3\Gamma = R + \Gamma\cdot\Gamma + D\!\!\!\!/\,\Gamma,$$

where $D\!\!\!\!/ \in \{d\!\!\!/, \nabla\!\!\!/, \mathrm{d}\!\!\!/\mathrm{iv}\}$.

We will work with the normalized null frame (e_1, e_2, e_3, e_4).

The null structure equations

The first variational formulas

$$\begin{aligned} \mathcal{L}\!\!\!/_4 g\!\!\!/ &= 2\chi, \\ \mathcal{L}\!\!\!/_3 g\!\!\!/ &= 2\underline{\chi}. \end{aligned} \tag{4.13}$$

The second variational formulas

$$\begin{aligned} \nabla\!\!\!/_4\chi &= -\chi\times\chi - \alpha + \omega\chi, \\ \nabla\!\!\!/_3\underline{\chi} &= -\underline{\chi}\times\underline{\chi} - \alpha + \omega\underline{\chi}. \end{aligned} \tag{4.14}$$

The transversal propagation equations for $\chi, \underline{\chi}$

$$\begin{aligned} \nabla\!\!\!/_4\underline{\chi} &= \frac{1}{2}\nabla\!\!\!/\underline{\eta} + \frac{1}{2}\widetilde{\nabla\!\!\!/\underline{\eta}} + \underline{\eta}\otimes\underline{\eta} - \frac{1}{2}(\chi\times\underline{\chi} + \underline{\chi}\times\chi) - \omega\underline{\chi} + \rho g\!\!\!/\,, \\ \nabla\!\!\!/_3\chi &= \frac{1}{2}\nabla\!\!\!/\eta + \frac{1}{2}\widetilde{\nabla\!\!\!/\eta} + \eta\otimes\eta - \frac{1}{2}(\underline{\chi}\times\chi + \chi\times\underline{\chi}) - \underline{\omega}\chi + \rho g\!\!\!/\,. \end{aligned} \tag{4.15}$$

The equations for the torsions $\eta, \underline{\eta}$

$$\nabla\!\!\!/_4\eta = -\chi^\sharp\cdot(\eta - \underline{\eta}) - \beta, \quad \nabla\!\!\!/_3\underline{\eta} = -\underline{\chi}^\sharp\cdot(\underline{\eta} - \eta) + \underline{\beta}, \tag{4.16}$$

$$\nabla\!\!\!/_3\eta = -2\underline{\chi}\cdot\eta^\sharp - \underline{\beta} + 2d\!\!\!/\,\underline{\omega} + (\eta + \underline{\eta})\underline{\omega}. \tag{4.17}$$

$$\mathrm{curl}\,\underline{\eta} = -\,\mathrm{curl}\,\zeta = -\,\mathrm{curl}\,\eta = \chi\wedge\underline{\chi} - \sigma. \tag{4.18}$$

The propagation equations for $\omega, \underline{\omega}$

$$\begin{aligned} \nabla\!\!\!/_4\underline{\omega} &= -\,\omega\underline{\omega} + \frac{1}{2}\Omega^2\cdot\Big[2(\eta,\underline{\eta}) - |\eta|^2 - \rho\Big], \\ \nabla\!\!\!/_3\omega &= -\,\omega\underline{\omega} + \frac{1}{2}\Omega^2\cdot\Big[2(\eta,\underline{\eta}) - |\underline{\eta}|^2 - \rho\Big]. \end{aligned} \tag{4.19}$$

The Gauss equation

$$K = (\chi, \underline{\chi}) - tr\chi tr\underline{\chi} - \rho \tag{4.20}$$

The Codazzi equations

$$\begin{aligned} &\mathrm{d}\!\!\!/\mathrm{iv}\chi - d\!\!\!/\,tr\chi + \chi^\sharp\cdot\zeta - (tr\chi)\cdot\zeta = -\beta, \\ &\mathrm{d}\!\!\!/\mathrm{iv}\underline{\chi} - d\!\!\!/\,tr\underline{\chi} - \underline{\chi}^\sharp\cdot\zeta + (tr\underline{\chi})\cdot\zeta = \underline{\beta}. \end{aligned} \tag{4.21}$$

Remarks:

In view of the decomposition $\chi = \frac{1}{2}(tr\chi)\not{g} + \hat{\chi}$, the second variational formulas imply:

$$\begin{aligned}\not{\nabla}_4 tr\chi =& - |\chi|^2 + \omega tr\chi,\\ \not{\nabla}_3 tr\underline{\chi} =& - |\underline{\chi}|^2 + \omega tr\underline{\chi}.\end{aligned} \tag{4.22}$$

and

$$\begin{aligned}\not{\nabla}_4 \hat{\chi} =& \omega\hat{\chi} - (tr\chi)\hat{\chi} - \alpha,\\ \not{\nabla}_3 \hat{\underline{\chi}} =& \underline{\omega}\hat{\underline{\chi}} - (tr\underline{\chi})\hat{\underline{\chi}} - \underline{\alpha},\end{aligned} \tag{4.23}$$

where for the former equations we used that $[\not{\nabla}_3, tr] = [\not{\nabla}_4, tr] = 0$ and for latter ones we used that

$$|\chi|^2 = \frac{1}{2}(tr\chi)^2\not{g} + |\hat{\chi}|^2 \qquad \text{and} \qquad \hat{\chi}\times\hat{\chi} = \frac{1}{2}|\hat{\chi}|^2\not{g}, \tag{4.24}$$

since $\hat{\chi}$ is symmetric and traceless (0,2) S-tensor field. The equations (4.22) are also known as Raychaudhuri equations.

Note also that

$$(\chi, \underline{\chi}) = \frac{1}{2}tr\chi tr\underline{\chi} + (\hat{\chi}, \hat{\underline{\chi}}).$$

Therefore, if

$$\check{\rho} = -\rho + (\hat{\chi}, \hat{\underline{\chi}}). \tag{4.25}$$

then the Gauss equation can be written as

$$K = \check{\rho} - \frac{1}{2}tr\chi tr\underline{\chi}.$$

Note that the product $tr\chi tr\underline{\chi}$ does not depend on the normalization of e_3, e_4, i.e. it is invariant under the transformations $e_3 \mapsto ae_3, e_4 \mapsto \frac{1}{a}e_4$, with $a > 0$. The *mass aspect functions* $\mu, \underline{\mu}$ are defined by

$$\begin{aligned}\mu =& K + \frac{1}{2}tr\chi tr\underline{\chi} - \not{\mathrm{d}}\mathrm{iv}\zeta,\\ \underline{\mu} =& K + \frac{1}{2}tr\chi tr\underline{\chi} + \not{\mathrm{d}}\mathrm{iv}\zeta.\end{aligned} \tag{4.26}$$

If $A_{u,\underline{u}} = \int_{S_{u,\underline{u}}} d\not{g}$ is the area of $S_{u,\underline{u}}$ and $r_{u,\underline{u}} = \sqrt{\frac{A_{u,\underline{u}}}{4\pi}}$ is the 'radius' of $S_{u,\underline{u}}$, then we define the *Hawking mass* m of $S_{u,\underline{u}}$ by

$$m = \frac{r}{2}\Big(1 + \frac{1}{8\pi}\int_S tr\chi tr\underline{\chi}\Big). \tag{4.27}$$

The Gauss–Bonnet theorem on $S_{u,\underline{u}}$ implies

$$\frac{2m}{r} = \frac{1}{4\pi}\int_S \mu = \frac{1}{4\pi}\int_S \underline{\mu}.$$

Geometric structure equations

We first derive differential equations which rely only on the geometry of the double null foliation, without assuming the Einstein equations.

We will show these relations only for the 4-direction. The proof is identical for the 3-direction.

First variation formula:

Let $X, Y \in T_p S_{u,\underline{u}}$ and extend them along the null generator of C_u passing through p so they satisfy the Jacobi equation $[L, X] = [L, Y] = 0$. Then, $[e_4, X] = (\nabla_X \log \Omega) e_4$ and $[e_4, Y] = (\nabla_Y \log \Omega) e_4$. Then,

$$
\begin{aligned}
(\not{\mathcal{L}}_4 \not{g})(X, Y) = (\mathcal{L}_4 \not{g})(X, Y) &= \nabla_4\big(\not{g}(X, Y)\big) - \not{g}([e_4, X], Y) - \not{g}(X, [e_4, Y]) \\
&= \not{g}(\nabla_4 X, Y) + \not{g}(X, \nabla_4 X) = \not{g}(\nabla_X e_4, Y) + \not{g}(X, \nabla_Y e_4) = 2\chi(X, Y).
\end{aligned}
$$

Since this is true for all $X, Y \in T_p S_{u,\underline{u}}$ we obtain the result.

Second variation formula:

Let $X, Y \in T_p S_{u,\underline{u}}$ and extend them along the null generator of C_u passing through p so they satisfy the Jacobi equation $[L, X] = [L, Y] = 0$. Then, the commutators $[e_4, X] = (\nabla_X \log \Omega) e_4$ and $[e_4, Y] = (\nabla_Y \log \Omega) e_4$ are proportional to e_4. Recall also that $\nabla_X e_4 = \chi^\sharp(X) - \zeta(X) e_4$. Then,

$$
\begin{aligned}
(\not\nabla_4 \chi)(X, Y) &= \nabla_4\big(\chi(X, Y)\big) - \chi(\nabla_4 X, Y) - \chi(X, \nabla_4 Y) \\
&= \nabla_4\big(g(\nabla_X e_4, Y)\big) - \chi(\nabla_X e_4, Y) - \chi(X, \nabla_Y e_4) \\
&= g(\nabla_4 \nabla_X e_4, Y) + g(\nabla_X e_4, \nabla_4 Y) - \chi(\chi^\sharp(X), Y) - \chi(X, \chi^\sharp(Y)) \\
&= g\big(R(e_4, X) e_4 + \nabla_X \nabla_4 e_4 + \nabla_{[e_4, X]} e_4, Y\big) - (\chi \times \chi)(X, Y) \\
&= R(Y, e_4, e_4, X) + \omega g(\nabla_X e_4, Y) - (\chi \times \chi)(X, Y) \\
&= \big(-\alpha - \chi \times \chi + \omega \chi\big)(X, Y).
\end{aligned}
$$

The torsion equation:

Let $X \in T_p S_{u,\underline{u}}$. Then

$$
\begin{aligned}
(\not\nabla_4 \eta)(X) = (\nabla_4 \eta)(X) = \nabla_4(\eta(X)) - \eta(\nabla_4 X) &= \nabla_4\big(g(\nabla_3 e_4, X)\big) - g(\nabla_3 e_4, \nabla_4 X) \\
&= g(\nabla_4 \nabla_3 e_4, X) + g(\nabla_3 e_4, \nabla_4 X) - g(\nabla_3 e_4, \nabla_4 X) \\
&= g\big(R(e_4, e_3) e_4 + \nabla_3 \nabla_4 e_4 + \nabla_{[e_4, e_3]} e_4, X\big) \\
&= R(X, e_4, e_4, e_3) + \omega g(\nabla_3 e_4, X) + g\left(\nabla_{\left(\underline{\eta}^\sharp - \eta^\sharp - \omega e_3 + \underline{\omega} e_4\right)} e_4, X\right) \\
&= -\beta(X) + \omega \eta(X) + g\left(\nabla_{\left(\underline{\eta}^\sharp - \eta^\sharp\right)} e_4, X\right) - \omega \eta(X) \\
&= \left(-\beta + \chi \cdot \left(\underline{\eta}^\sharp - \eta^\sharp\right)\right)(X).
\end{aligned}
$$

Furthermore, for all scalar functions f we have the following commutation formula:

$$
[\not\nabla_3, \not{d}] f = \frac{1}{2}(\eta + \underline{\eta}) \nabla_3 f - \underline{\chi} \cdot (\not\nabla f)
$$

Proof. For any scalar function f and $X \in T_p S_{u,\underline{u}}$ we have,

$$
\begin{aligned}
\not\nabla_3(\not{d} f)(X) = \nabla_3(\not{d} f)(X) &= \nabla_3(X f) - (\not\nabla_3 X) f \\
&= X \nabla_3 f + [\nabla_3, X] f - (\not\nabla_3 X) f = X \nabla_3 f + (\nabla_3 X - \nabla_X e_3) f - (\not\nabla_3 X) f
\end{aligned}
$$

The result follows from the formulas (4.7). □

Since $\eta = -\underline{\eta} + 2\not{d}\log\Omega$ we obtain

$$\begin{aligned}
\nabla\!\!\!/_3\eta &= -\nabla\!\!\!/_3\underline{\eta} + 2\nabla\!\!\!/_3(\not{d}\log\Omega)\\
&= \underline{\chi}\cdot(\underline{\eta}^\sharp - \eta^\sharp) - \underline{\beta} + 2\nabla\!\!\!/_3(\not{d}\log\Omega)\\
&= \underline{\chi}\cdot(\underline{\eta}^\sharp - \eta^\sharp - 2\nabla\!\!\!/\log\Omega) - \underline{\beta} + 2\not{d}\,\underline{\omega} + (\eta + \underline{\eta})\underline{\omega}\\
&= -2\underline{\chi}\cdot\eta^\sharp - \underline{\beta} + 2\not{d}\,\underline{\omega} + (\eta + \underline{\eta})\underline{\omega}.
\end{aligned}$$

The propagation equation for $\underline{\omega}$:

Recall the equivariant pair $(L, \underline{L}) = (\Omega e_4, \Omega e_3)$. The following relations can be easily proved:

$$\begin{aligned}
&\Omega\underline{\omega} = -\frac{1}{2}\Omega^{-2}g(\nabla_{\underline{L}}\underline{L}, L),\\
&g\left(\nabla_{\underline{L}}\underline{L}, \nabla_L L\right) = -4\Omega^4\omega\underline{\omega},\\
&g\left(\nabla_{[L,\underline{L}]}\underline{L}, L\right) - 2\Omega^2 g\left(\nabla_{\zeta^\sharp}\underline{L}, L\right) = \Omega^4\cdot\left(\eta - \underline{\eta}, \eta\right),\\
&g(\nabla_L\underline{L}, L) = 0,\ \text{since } \nabla_L\underline{L} \text{ is } S\text{-tangent},\\
&g\left(\nabla_{\underline{L}}\nabla_L\underline{L}, L\right) = -\Omega^4\cdot(\eta, \underline{\eta}),\\
&g(R(L, \underline{L})\underline{L}, L) = R(L, \underline{L}, L, \underline{L}) = \Omega^4 R_{4343} = \Omega^4\rho.
\end{aligned}$$

Then,

$$\begin{aligned}
\nabla\!\!\!/_4(\Omega\underline{\omega}) &= \frac{1}{2}\frac{2}{\Omega^2}\frac{L\Omega}{\Omega}g\left(\nabla_{\underline{L}}, \underline{L}, L\right) - \frac{1}{2}\Omega^{-2}\cdot\left[g(\nabla_{\underline{L}}\underline{L}, \nabla_L L) + g\left(\nabla_L\nabla_{\underline{L}}\underline{L}, L\right)\right]\\
&= -2\Omega^2\omega\underline{\omega} + 2\Omega^2\omega\underline{\omega} - \frac{1}{2}\Omega^{-2}g\left(\nabla_{\underline{L}}\nabla_L\underline{L} + R(L, \underline{L})\underline{L} + \nabla_{[L,\underline{L}]}\underline{L}, L\right)\\
&= \frac{1}{2}\Omega^2\cdot\left[(\eta, \underline{\eta}) - \rho - (\eta - \underline{\eta}, \eta)\right]\\
&= \frac{1}{2}\Omega^2\cdot\left[2(\eta, \underline{\eta}) - |\eta|^2 - \rho\right].
\end{aligned}$$

The transversal propagation equation for χ and the curl equation for η can be obtained as follows:

Let $X, Y \in T_pS_{u,\underline{u}}$ and extend them along the null generator of C_u passing through p so they satisfy the Jacobi equation $[L, X] = [L, Y] = 0$. Then, $[e_4, X] = (\nabla_X\log\Omega)e_4$ and $[e_4, Y] = (\nabla_Y\log\Omega)e_4$. Then, using (4.9) we obtain

$$\nabla_X e_4 = \chi^\sharp\cdot X - \zeta(X)\cdot e_4, \qquad \nabla_4 X = \chi^\sharp\cdot X + \underline{\eta}(X)e_4$$

and similarly for Y. Furthermore, using also (4.7) one can easily verify the following:

$$\begin{aligned}
&g(\nabla_X e_3, \nabla_4 Y) = \Big[(\underline{\chi}\times\chi) + \frac{1}{2}(\underline{\eta}\otimes\underline{\eta}) - \frac{1}{2}(\eta\otimes\underline{\eta})\Big](X, Y),\\
&g(\nabla_X\nabla_4 e_3, Y) = (\nabla_X\underline{\eta})(Y) - \omega\underline{\chi}(X, Y),\\
&g\left(\nabla_{[\nabla_4, X]}e_3, Y\right) = \frac{1}{2}(\eta\otimes\underline{\eta})(X, Y) + \frac{1}{2}(\underline{\eta}\otimes\underline{\eta})(X, Y),\\
&g(\nabla_4\nabla_X e_3, Y) = g(\nabla_X\nabla_4 e_3 + \nabla_{[\nabla_4, X]}e_3, Y) + R(Y, e_3, e_4, X).
\end{aligned}$$

We have:

$$\begin{aligned}
\left(\nabla\!\!\!/_4\underline{\chi}\right)(X,Y) &= \left(\nabla_4\underline{\chi}\right)(X,Y) = \nabla_4\left(\underline{\chi}(X,Y)\right) - \underline{\chi}(\nabla_4 X, Y) - \chi(X,\nabla_4 Y)\\
&= \nabla_4\big(g(\nabla_X e_3, Y)\big) - \underline{\chi}(\chi^\sharp\cdot X, Y) - \underline{\chi}(X,\chi^\sharp\cdot Y)\\
&= g(\nabla_4\nabla_X e_3, Y) + g(\nabla_X e_3, \nabla_4 Y) - (\underline{\chi}\times\chi)(Y,X) - (\underline{\chi}\times\chi)(X,Y)\\
&= (\nabla_X\underline{\eta})(Y) + (\underline{\eta}\otimes\underline{\eta})(X,Y) - (\chi\times\underline{\chi})(X,Y) - R(X,e_4,Y,e_3) - \omega\underline{\chi}(X,Y).
\end{aligned} \tag{4.28}$$

If we consider the antisymmetric part of the above equation, then since the left hand side is symmetric and since $d\!\!\!/\,\underline{\eta} = \nabla\underline{\eta} - \widetilde{\nabla\underline{\eta}}$, we obtain

$$\begin{aligned}
d\!\!\!/\,\underline{\eta} &= \big(\chi\times\underline{\chi} - \underline{\chi}\times\chi\big)(X,Y) + R(X,e_4,Y,e_3) - R(Y,e_4,X,e_3)\\
&= \big(\chi\times\underline{\chi} - \underline{\chi}\times\chi\big)(X,Y) - R(X,Y,e_3,e_4),
\end{aligned}$$

where in the last equation we used the first Bianchi identity. We have thus obtained the equality of two 2-forms on $S_{u,\underline{u}}$. Hence, their Hodge duals are also equal:

$$\operatorname{curl}\underline{\eta} = *d\!\!\!/\,\underline{\eta} = \chi\wedge\underline{\chi} - \frac{1}{2}\not\epsilon^{AB}R_{AB34} = \chi\wedge\underline{\chi} - \sigma.$$

On the other hand, the symmetric part of (4.28) reads:

$$\nabla\!\!\!/_4\underline{\chi} = \frac{1}{2}\nabla\!\!\!/\underline{\eta} + \frac{1}{2}\widetilde{\nabla\!\!\!/\underline{\eta}} + \underline{\eta}\otimes\underline{\eta} - \frac{1}{2}(\chi\times\underline{\chi} + \underline{\chi}\times\chi) - \omega\underline{\chi} - \operatorname{symm}\Big(R(\cdot,e_4,\cdot,e_3)\Big). \tag{4.29}$$

The remaining two equations relate the geometry of $(S_{u,\underline{u}}, g\!\!\!/\,)$ and the second fundamental forms $\chi,\underline{\chi}$ with the geometry of the spacetime manifold $(\mathcal{M},g)$.

<u>The Gauss equation:</u>

Recall that if $\nabla\!\!\!/$ denotes the induced connection on $S_{u,\underline{u}}$ and X,Y are S vector fields then

$$\nabla_X Y = \nabla\!\!\!/_X Y + II(X,Y),$$

where $II(X,Y)\in (T_xS)^\perp$. If X,Y,Z,W are S vector fields and $R\!\!\!/$ denotes the induced Riemann curvature of $S_{u,\underline{u}}$ then the Gauss equation reads

$$R\!\!\!/(W,Z,X,Y) = R(W,Z,X,Y) + g\Big(II(Z,X), II(W,Y)\Big) - g\Big(II(W,X), II(Z,Y)\Big).$$

Proof. We have

$$\begin{aligned}
R(W,Z,X,Y) =& g\Big(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z, W\Big)\\
=& g\Big(\nabla_X\nabla\!\!\!/_Y Z + \nabla_X II(Y,Z) - \nabla_Y\nabla\!\!\!/_X Z - \nabla_Y II(X,Z) - \nabla\!\!\!/_{[X,Y]}Z, W\Big)\\
=& g\Big(\nabla\!\!\!/_X\nabla\!\!\!/_Y Z - \nabla\!\!\!/_Y\nabla\!\!\!/_X Z - \nabla\!\!\!/_{[X,Y]}Z, W\Big) - g\Big(II(Y,Z),\nabla_X W\Big) + g\Big(II(X,Z),\nabla_Y W\Big)\\
=& R\!\!\!/(W,Z,X,Y) - g\Big(II(W,X), II(Z,Y)\Big) + g\Big(II(Z,X), II(W,Y)\Big).
\end{aligned}$$

□

Let $\{e_A : A = 1, 2\}$ be a local frame field for $S_{u,\underline{u}}$. Setting $W = e_A, Z = e_B, X = e_C, Y = e_D$ and using that (see (4.7))

$$II(e_A, e_B) = \chi_{AB} e_3 + \underline{\chi}_{AB} e_4. \tag{4.30}$$

we can rewrite the Gauss equation as

$$R_{ABCD} = \not{R}_{ABCD} + \chi_{AC}\underline{\chi}_{BD} + \chi_{BD}\underline{\chi}_{AC} - \chi_{AD}\underline{\chi}_{BC} - \chi_{BC}\underline{\chi}_{AD}. \tag{4.31}$$

Since $S_{u,\underline{u}}$ is a two-dimensional (sub-)manifold, we have

$$\not{R}_{ABCD} = K \not{g}_{AB} \not{g}_{CD},$$

where K is the Gaussian curvature of $S_{u,\underline{u}}$. Therefore, we should think of (4.31) as an equation between functions. In fact, we can explicitly do this by taking the double trace of (4.31) and using $\not{g}_{AB}\not{g}^{AB} = 2$:

$$\frac{1}{2} R_{ABCD} \not{g}^{AC} \not{g}^{BD} = K + tr\chi tr\underline{\chi} - (\chi, \underline{\chi}). \tag{4.32}$$

The Codazzi equation:

The vector fields X, Y, Z are S-tangential. The Gauss equation expresses the S-tangential components of $R(X, Y)Z$ in terms of the intrinsic geometry of S and the second fundamental form. The Codazzi equation expresses the normal to S components of $R(X, Y)Z$ in terms of the derivative (i.e. normal connection) of the second fundamental form II. Let us recall the definition of the normal connection. Suppose that X is an S vector field and V is a normal to S vector field, i.e. $V \in (TS)^{\perp}$. Then, the normal connection ${}^n\nabla$ is defined such that

$${}^n\nabla_X V = \mathrm{nor}\big(\nabla_X V\big),$$

where $\mathrm{nor}\big(\nabla_X V\big)$ denotes the normal projection of $\nabla_X V$ on $(TS)^{\perp}$. We can also extend the normal connection to more general (normal) tensor fields on S. In particular, we have

$$\big({}^n\nabla_X II\big)(Y, Z) = {}^n\nabla_X\big(II(Y, Z)\big) - II(\not\nabla_X Y, Z) - II(Y, \not\nabla_X Z). \tag{4.33}$$

The Codazzi equation then reads

$$\mathrm{nor} R(X, Y)Z = \big({}^n\nabla_X II\big)(Y, Z) - \big(\nabla^n_Y II\big)(X, Z). \tag{4.34}$$

Proof.

$$\begin{aligned}
\mathrm{nor}R(X,Y)Z =& \mathrm{nor}\nabla_X\nabla_Y Z - \mathrm{nor}\nabla_Y\nabla_X Z - \mathrm{nor}\nabla_{[X,Y]}Z \\
=& II\big(X, \not\nabla_Y Z\big) + \nabla^n_X\big(II(Y,Z)\big) - II\big(Y, \nabla_X Z\big) - \nabla^n_Y\big(II(X,Z)\big) - II([X,Y],Z) \\
=& II\big(X, \not\nabla_Y Z\big) + II\big(\not\nabla_Y X, Z\big) - \nabla^n_Y\big(II(X,Z)\big) \\
& - II\big(Y, \not\nabla_X Z\big) - II(\not\nabla_X Y, Z) + \nabla^n_X\big(II(Y,Z)\big) \\
=& \big(\nabla^n_X II\big)(Y,Z) - \big(\nabla^n_Y II\big)(X,Z).
\end{aligned}$$

□

Observe now that (4.7), (4.30) and (4.33) imply

$$\big({}^{n}\nabla_X II\big)(Y,Z) = \Big[\chi(Y,Z)\zeta(X) + (\nabla\!\!\!/_X\chi)(Y,Z)\Big]\cdot e_3 + \Big[\underline{\chi}(Y,Z)\underline{\zeta}(X) + (\nabla\!\!\!/_X\underline{\chi})(Y,Z)\Big]\cdot e_4, \tag{4.35}$$

where $\underline{\zeta} = -\zeta$. Therefore, in our setting, we can rewrite the Codazzi equation (4.34) as follows

$$\begin{aligned}\mathrm{nor}R(X,Y)Z =& \Big[\chi(Y,Z)\zeta(X) + (\nabla\!\!\!/_X\chi)(Y,Z) - \chi(X,Z)\zeta(Y) - (\nabla\!\!\!/_Y\chi)(X,Z)\Big]\cdot e_3\\ &+ \Big[\underline{\chi}(Y,Z)\underline{\zeta}(X) + (\nabla\!\!\!/_X\underline{\chi})(Y,Z) - \underline{\chi}(X,Z)\underline{\zeta}(Y) - (\nabla\!\!\!/_Y\underline{\chi})(X,Z)\Big]\cdot e_4.\end{aligned} \tag{4.36}$$

The above finally gives us

$$\begin{aligned}R(e_3,Z,X,Y) &= -\underline{\chi}(Y,Z)\underline{\zeta}(X) - (\nabla\!\!\!/_X\underline{\chi})(Y,Z) + \underline{\chi}(X,Z)\underline{\zeta}(Y) + (\nabla\!\!\!/_Y\underline{\chi})(X,Z),\\ R(e_4,Z,X,Y) &= -\chi(Y,Z)\zeta(X) - (\nabla\!\!\!/_X\chi)(Y,Z) + \chi(X,Z)\zeta(Y) + (\nabla\!\!\!/_Y\chi)(X,Z).\end{aligned} \tag{4.37}$$

In view of the fact that

$$R(e_i,Z,X,Y) = \xi_i(Z)\cdot \epsilon\!\!/\,(X,Y), i = 3,4,$$

each of the above equations should be considered as an equation of one-forms acting on Z. By setting $X = e_A, Y = e_B, Z = e_C$ and taking the trace in the Z,X-entries we obtain:

$$(g\!\!\!/^{-1})^{AC}R_{4CAB} = -\chi_B^{\sharp A}\zeta_A - (\mathrm{d}\!\!/\mathrm{iv}\chi)_B + (tr\chi)\zeta_B + tr\nabla\!\!\!/_B\chi. \tag{4.38}$$

Since $[\nabla\!\!\!/_4, tr] = 0$ (since $\nabla\!\!\!/_4 g\!\!\!/ = 0$) we finally obtain:

$$(g\!\!\!/^{-1})^{AC}R_{4CAB} = \Big(-\chi^\sharp\cdot\zeta - \mathrm{d}\!\!/\mathrm{iv}\chi + (tr\chi)\cdot\zeta + d\!\!\!/\, tr\chi\Big)(e_B). \tag{4.39}$$

Proof of null structure equations

Let us now impose the Einstein equations (see Proposition 4.3.1).

1. The equations $R_{44} = R_{33} = 0$:

The above equations are equivalent to $tr\alpha = tr\underline{\alpha} = 0$ which in turn are equivalent to (4.22).

2. The equations $R_{3A} = R_{4A} = 0$:

The equations $R_{4A} = 0$ are equivalent to $(g\!\!\!/^{-1})^{AC}R_{4CAB} = \beta_B$ which in turn, in view of (4.39), are equivalent to 4.21.

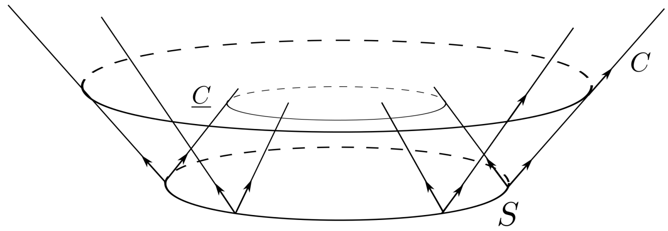
3. The equations $R_{34} = R_{AB} = 0$:

First note that (4.11), (4.12) and (4.29), (4.32) imply the null structure equations (4.15) and (4.20). Conversely, given (4.15) we can infer $R_{34} = 0$ from (4.29). Finally, given also (4.20) we have $R_{AB} = 0$.

4.6 The Characteristic Initial Value Problem

In Section 3.3 we discussed about the Cauchy problem for the Einstein equations. In particular, we saw that the initial data set consists of the triplet $(\mathcal{H}_0, \overline{g}, k)$, where $\mathcal{H}_0$ is a three-dimensional Riemannian manifold, $\overline{g}$ is the metric on $\mathcal{H}_0$ and k is a symmetric (0,2) tensor field on $\mathcal{H}_0$ and such that $\overline{g}, k$ satisfy the constraint equations. Recall that $\overline{g}, k$ are to be the first and second fundamental forms of $\mathcal{H}_0$ in $\mathcal{M}$, respectively.

In this section, we will discuss in detail the formulation of the characteristic initial value problem, i.e. the case where the initial Riemannian (spacelike) Cauchy hypesurface $\mathcal{H}_0$ is replaced by two degenerate (null) hypersurfaces $C \cup \underline{C}$ intersecting at a two-dimensional surface S.



Motivation

Let us first motivate the formulation of the characteristic initial value problem. Let us assume that $\not{g}$ is a given degenerate metric on $C \cup \underline{C}$ and let $\mathcal{M}$ be the arising spacetime manifold and g the Lorentzian metric which satisfies the Einstein equations extending $\not{g}$ on $C \cup \underline{C}$. Let us consider the double null foliation of $(\mathcal{M}, g)$ such that $\Omega = 1$ on $C \cup \underline{C}$. Let L be the geodesic vector field on C, which coincides with the normalized and equivariant vector field, and let $\underline{u}$ be its affine parameter such that $\underline{u} = 0$ on S. Then, we obtain a foliation of C which consists of the (spacelike) surfaces $S_\tau = \{\underline{u} = \tau\}$. The crucial observation is that the null second fundamental form χ on C, which recall that is defined to be the following (0,2) tensor field on C

$$\chi(X, Y) = g(\nabla_X e_4, Y),$$

where $X, Y \in T_pC$, is in fact, an tensor field which depends only on the intrinsic geometry of C (although $\nabla_X L$ depends on the spacetime metric g). Indeed, the first variational formula gives us

$$\chi = \frac{1}{2}\not{\mathcal{L}}_4 \not{g},$$

and since the Lie derivative $\not{\mathcal{L}}_L$ is intrinsic to the hypersurface C, we deduce that $\not{g}$ completely determines χ on C. On the other hand, by the Raychaudhuri equation we have

$$e_4(tr\chi) = -|\chi|^2 - tr\alpha,$$

and since χ and $tr\chi$ (and $\omega = L(\log\Omega) = 0$) are determined from $\not{g}$, we deduce that $tr\alpha$ is also determined. However, in view of the Einstein equations (see Section 4.3) we have

$$tr\alpha = Ric(e_4, e_4) = 0.$$

This shows that one cannot arbitrarily prescribe a degenerate metric $\not{g}$ on $C \cup \underline{C}$, since otherwise $tr\alpha$ would in general be non-zero.

A naive approach would of course be to mimic the formulation of the Cauchy problem in which case the characteristic initial data would consist of a degenerate metric $\not{g}$, and possibly other tensor fields on $C \cup \underline{C}$, subject to constraint equations. However, it turns out that there is another very elegant approach according to which the initial data do not contain the full knowledge of the degenerate metric $\not{g}$ on $C \cup \underline{C}$. On the other hand, if we want to consider the characteristic initial value problem, we need to make sure that $C \cup \underline{C}$ are to be null hypersurfaces in the to be constructed spacetime $(\mathcal{M}, g)$. The crucial step here is to recall Corollary 2.2.1 according to which conformal transformation respect null hypersurfaces and their null generators. In other words, instead of prescribing the full metric $\not{g}$ on $C \cup \underline{C}$, we could just prescribe the conformal class of the metric, namely

$$\mathrm{Conf}(\not{g}) = \{A\not{g}\,, A \in C^{\infty}(C \cup \underline{C}), A > 0\}\,.$$

In order to see if this approach is fruitful, we need to understand the conformal properties of the double null foliation.

Conformal properties of the double null foliation

Before, we proceed with an exposition of the conformal properties of the double null foliation, it is important to fix a foliation of $C \cup \underline{C}$ relative to which our quantities are constructed.

Fixing a foliation of $C \cup \underline{C}$

Let $L, \underline{L}$ be null vector fields tangential to S. We consider the (unique, given the vector fields $L, \underline{L}$ on S) foliation of $C \cup \underline{C}$ such that $\Omega = 1$ with respect to the actual (but unknown for the time being) metric $\not{g}$. This choice uniquely determines the vector fields $L, \underline{L}$ on $C, \underline{C}$, respectively. Note that this choice is analogous to the choice $\Phi = 1$ in Section 3.3. The affine parameters $\underline{u}, u$ of these vector fields define the sections of the foliation of the null hypersurfaces. In fact, the above choice determines the optical functions $u, \underline{u}$ globally. Hence, the equivariant vector fields for any metric $\tilde{g} \in \mathrm{Conf}(\not{g})$ coincide with $L, \underline{L}$.

For all metrics in $\mathrm{Conf}(\not{g})$, we assume that all the associated tensor fields (i.e. the connection coefficients and curvature components) are defined with respect to the above foliation. For example, if $\tilde{g} \in \mathrm{Conf}(\not{g})$ then the second fundamental form $\tilde{\chi}$ is defined by

$$\tilde{\chi}(X, Y) = \tilde{g}(\tilde{\nabla}_X L, Y),$$

where $\tilde{\nabla}$ is the connection of $\tilde{g}$. Note that L coincides with the geodesic, the normalized and the equivariant vector field with respect to the metric $\not{g}$. However, this is not the case for the metric $\tilde{g} = A\not{g}$, since in this case the corresponding null lapse $\tilde{\Omega} = \sqrt{A} \neq 1$ (this relation follows immediately from the fact that $(L, \underline{L})$ is an equivariant pair for both $\not{g}$ and $\tilde{g}$). This is why it is important to fix the vector field L and $\underline{L}$ with respect to which we calculate the connection coefficients and curvature components. Note that in view of the first variational formula we have

$$\chi = \frac{1}{2}\not{\mathcal{L}}_L \not{g}\,, \qquad \tilde{\chi} = \frac{1}{2}\not{\mathcal{L}}_L \tilde{g}.$$

Since $\tilde{g} = A\not{g}$, we have

$$\tilde{\chi} = \frac{1}{2}\not{\mathcal{L}}_L \tilde{g} = \frac{1}{2}\not{\mathcal{L}}_L (A\not{g}) = \frac{1}{2}(LA)\not{g} + A\chi = \frac{1}{2}(LA)\not{g} + A\chi.$$

Therefore, for the traceless parts $\hat{\tilde{\chi}}, \hat{\chi}$ we have

$$\hat{\tilde{\chi}} = \tilde{\chi} - \frac{1}{2}(tr\tilde{\chi})\tilde{g} = \frac{1}{2}(LA)\not{g} + A\chi - \frac{1}{2}A^{-1}(LA + Atr\chi)A\not{g} = A\hat{\chi},$$

where we used that $\tilde{g}^{AB} = A^{-1}\not{g}^{AB}$. Therefore,

$$|\hat{\tilde{\chi}}|^2_{\tilde{g}} = \tilde{g}^{AB}\tilde{g}^{CD}\hat{\tilde{\chi}}_{AC}\hat{\tilde{\chi}}_{BD} = A^{-2}\not{g}^{AB}\not{g}^{CD}A^2\hat{\chi}_{AB}\hat{\chi}_{CD} = |\hat{\chi}|^2_{\not{g}}. \tag{4.40}$$

Hence, the size $|\hat{\tilde{\chi}}|^2_{\tilde{g}}$ of the shear is conformally invariant! We denote $e = |\hat{\tilde{\chi}}|^2_{\tilde{g}}$.

Hence, so far we have fixed the optical functions $u, \underline{u}$, the conformal geometry of $C \cup \underline{C}$ is known as well as e. It is convenient from now on to work with a distinguished representative of Conf($\not{g}$). At this point, we need to assume that the metric $\not{g}$ is completely known on S. Then, using canonical coordinates $(\underline{u}, \theta)$ along C (and similarly for $\underline{C}$) we obtain consider the unique metric $\tilde{g}$ $\in$Conf($\not{g}$) such that

$$\sqrt{det\tilde{g}}\Big|_{(\underline{u},\theta)} = \sqrt{det\not{g}}\Big|_{(0,\theta)}.$$

Note that

$$\not{g} = \phi^2\tilde{g}$$

for some ϕ which is uniquely determined by the above condition. Clearly, we have $\phi|_S = 1$. Therefore, specifying of the conformal class Conf($\not{g}$) is equivalent to specifying the metric $\tilde{g}$.

The free data vs the full data

By the Raychaudhuri equation (for $\not{g}$ and hence $\Omega = 1, \omega = 0$) we obtain

$$\partial_{\underline{u}} tr\chi = \nabla_L tr\chi = -|\chi|^2_{\not{g}} = -\frac{1}{2}(tr\chi)^2 - |\hat{\chi}|^2_{\not{g}} = -\frac{1}{2}(tr\chi)^2 - e.$$

Recall now that

$$\phi^2(\underline{u}, \theta) = b(\theta) \cdot \sqrt{det\not{g}}\Big|_{(\underline{u},\theta)}, \quad b(\theta) = \frac{1}{\sqrt{detg}\big|_{(0,\theta)}}. \tag{4.41}$$

Then, if χ is the null second fundamental form associated with $\not{g}$, the first variational formula gives us

$$\chi_{AB} = \frac{1}{2}\partial_{\underline{u}}\not{g}_{AB} \;\Rightarrow\; \partial_{\underline{u}}(\sqrt{det\not{g}}) = (tr\chi) \cdot \sqrt{det\not{g}}. \tag{4.42}$$

Equations (4.41) and (4.42) imply:

$$tr\chi = \frac{\partial_{\underline{u}}\phi^2}{\phi^2} = \frac{2}{\phi} \cdot \partial_{\underline{u}}\phi. \tag{4.43}$$

Using now (4.41) and (4.43) we obtain that ϕ satisfies the following linear(!) equation:

$$\partial_{\underline{u}}\partial_{\underline{u}}\phi = -\frac{1}{2}e\phi.$$

Note that $\phi|_S = 1$ and, by (4.43), $\partial_{\underline{u}}\phi\big|_S = tr\chi$. Therefore, if we also know $tr\chi$ and $tr\underline{\chi}$ on S then we can determine ϕ on $C \cup \underline{C}$. This in turn determines $\not{g}$, and by first variational formula, the second fundamental forms χ, $\underline{\chi}$.

Let us try to determine the rest of the connection coefficients. By the choice of our gauge $\Omega = 1$ we have $\omega = \underline{\omega} = 0$ on $C \cup \underline{C}$. Hence, it remains to determine the torsion η (note that since $\Omega = 1$ we have $\eta = \zeta = -\underline{\eta}$. Note that we can eliminate the curvature component β from the Codazzi equations (4.21) and the torsion equations (4.16) and hence we obtain a linear propagation equation for the torsion (recall that $\chi, \underline{\chi}$ were just determined) and hence can be explicitly solved provided we know the value of the torsion η at S.

The curvature components can now be very easily computed using the null structure equations.

Formulation of characteristic initial value problem

Our previous analysis suggests the following:

The characteristic initial data set for the Einstein equations consists of a pair of three-dimensional hypersurfaces intersecting at a two-dimensional surface along with the (free) specification of the conformal class Conf($\not{g}$) of the degenerate metric $\not{g}$ on $C \cup \underline{C}$ as well as the full metric $\not{g}$, the expansions trχ, tr$\underline{\chi}$ and the torsion η on S.

Local well-posedness

Rendall has shown that for smooth characteristic initial data there exists a unique solution to the Einstein equations in a neighborhood of the surface S.

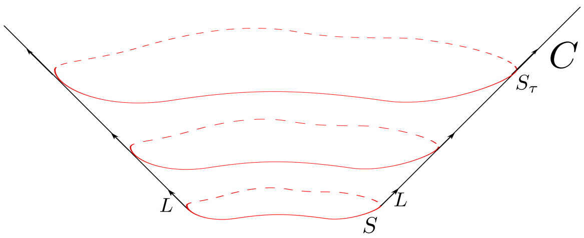
Luk has recently extended the above result to appropriate neighborhoods of the initial hypersurfaces $C, \underline{C}$.

Chapter 5

Applications to Null Hypersurfaces

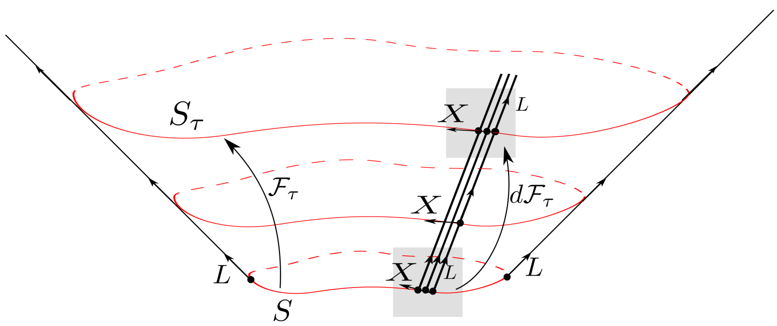
We restrict to one hypersurface C in a Lorentzian manifold $(\mathcal{M}, g)$, and hence for convenience, we fix the gauge such that $\Omega = 1$ on C. Then, the geodesic, the normalized and the equivariant null frames $(e_1, e_2, \underline{L}, L)$ all coincide on C (and L is tangential to C). Let S_τ be the level sets of the affine parameter of L on C. We will refer to $\cup_\tau S_\tau$ as the affine foliation of C. Since $\Omega = 1$ we have, in particular,

$$\nabla_L L = 0, \qquad \eta = \zeta = -\underline{\eta}.$$



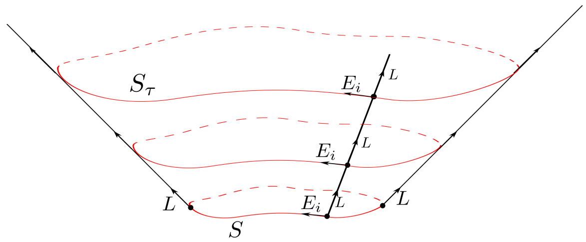
5.1 Jacobi Fields and Tidal Forces

A vector field X on C is called normal Jacobi vector field if $\mathcal{L}_L X = 0$. It is clear that X is tangent to the sections S_τ of the affine foliation of C. The reason X is called normal Jacobi is because X represents the infinitesimal displacement of the null generators around a fixed null generator.



Fermi frame

It is possible that the null generators approach another one as is the case for the incoming null geodesic congruence of a standard sphere in Minkowski. In this case, all null geodesics intersect at the vertex of the cone. In order to understand quantitatively the behavior of the null generators we need to express the normal Jacobi fields along a given null generator G_x in terms of an orthonormal basis $E_i(\tau), i = 1, 2$ which is tangent to the sections S_τ.



Let us first see how one can consider such basis. Starting with an orthonormal frame $E_i(0), i = 1, 2$ of $T_x S_0$, one ideally wants to propagate this along the null generator G_x such that at each point $G_x(\tau)$ one obtains an orthonormal frame of $T_{G_x(\tau)} S_\tau$. One way to do this, is by considering the so-called *Fermi frame* constructed by the following propagation equation:

$$\nabla_L E_i = -\zeta_i L$$

where $\zeta_i = \zeta(E_i)$. The above relation immediately implies that

$$L\big(g(E_i, K)\big) = 0,$$

where $K \in \{E_j, L, \underline{L}\}, j = 1, 2$, confirming the fact that if $E_i(0), i = 1, 2$, is an orthonormal frame of $T_x S_0$ then $E_i(\tau), i = 1, 2$, is an orthonormal frame of $T_{G_x(\tau)} S_\tau$.

Propagation equation of Jacobi fields

If now X is a normal Jacobi field along G_x, where $x \in S_0$, then it is very convenient to express X in terms of the Fermi frame E_i. Recall that X represents the infinitesimal displacement of null generators nearby G_x and E_i, on the other hand, remains orthonormal. We have:

$$X = X^i(\tau) \cdot E_i(\tau), \tag{5.1}$$

where as usual we sum over the repeated index i. Then,

$$\nabla_L X = \frac{dX^i(\tau)}{d\tau} \cdot E_i - X^i \cdot \zeta_i L. \tag{5.2}$$

On the other hand,

$$g(\nabla_L X, E_j) = g(\nabla_X L, E_J) = \chi(X, E_j), \tag{5.3}$$

giving us

$$\frac{dX^i(\tau)}{d\tau} = \chi_{ij} X^j. \tag{5.4}$$

This shows that the components X^i satisfy a linear system of first order equations. However, the connection coefficients χ_{ij} depend on the regularity of null hypersurface and not on the regularity of the ambient metric g. In order to obtain propagation equations for X^i which only depend on the Riemann curvature of g, we need to look at the second order equations.

The vector field X satisfies the more general Jacobi equation:

$$J_L X = \nabla_L \nabla_L X - R(L, X)L = 0. \tag{5.5}$$

Indeed, we have

$$R(L, X)L = \nabla_L \nabla_X L - \nabla_X \nabla_L L - \nabla_{[L,X]} L = \nabla_L \nabla_L X,$$

since $[L, X] = 0$ and $\nabla_L L = 0$. The operator J_L is also known as the *tidal operator* and appears naturally in the second variation of the length of G_x. Now, equation (5.5) implies

$$g(\nabla_L \nabla_L X, E_j) = g\big(R(L, X)L, E_j\big) = \alpha_{ij} X^i,$$

where α is the curvature component introduced in Section 4.3. On the other hand, (5.2) implies

$$\Big(\nabla_L \nabla_L X\Big)^i = \frac{d^2 X^i(\tau)}{d\tau^2}.$$

Hence,

$$\frac{d^2 X^i(\tau)}{d\tau^2} = \alpha_{ij} X^i. \tag{5.6}$$

Therefore, as long as the spacetime metric g remains smooth (and hence the curvature component α is smooth), the components X^i of the normal Jacobi vector field X with respect to the Fermi frame satisfy the above linear second order system of equations. In view of (5.4), the initial data $X^i(0), dX^i(0)/d\tau$ depend nearly on $X^i(0)$ and so does the solution $X^i(\tau)$. Therefore, there exists a smooth curve of matrices $M(\tau)$ such that

$$X^i(\tau) = M^i_j(\tau) X^j(0). \tag{5.7}$$

Clearly

$$M(0) = I, \text{ and hence } detM(0) = 1. \tag{5.8}$$

The matrix M measures the change of the normal Jacobi field X along G_x with respect to its initial value $X(0) = X_x$. For this reason, M is called the *deformation matrix*.

Using equation (5.4) we can obtain a propagation equation for M. Indeed, we have

$$\frac{dX^i(\tau)}{d\tau} = \frac{dM^i_k(\tau)}{d\tau} X^k(0) = \chi_{ij} X^j(\tau) = \chi_{ij} M^j_k X^k(0).$$

Therefore,

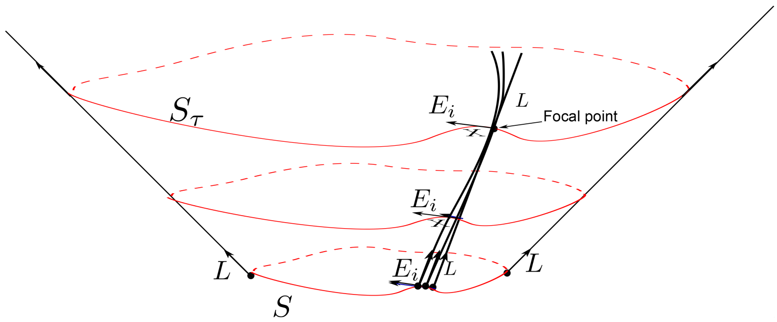
$$\frac{dM^i_k(\tau)}{d\tau} = \chi_{ij} M^j_k = (\chi \times M)_{ik},$$

or simply,

$$\frac{dM(\tau)}{d\tau} = \chi \times M. \tag{5.9}$$

5.2 Focal Points

Recall that as long as the Riemann curvature is smooth then so is the solution of (5.6). On the other hand, the deformation matrix M is invertible as long as the flow $\mathcal{F}_\tau$ of L induces isomorphisms between $T_x S_0$ and $T_{G_x(\tau)} S_\tau$. However, the null generators nearby G_x may converge to G_x in such a way such that $X = 0$ at some point $G_x(\tau_f)$ along G_x. In this case, the point $G_x(\tau_f)$ is called a *focal point* of C. For convenience, let τ_f denote the first time for which we have a focal point along G_x. The regularity of the null hypersurface breaks down at the focal points. We will see the exact meaning of this by expressing the vanishing of the normal Jacobi field X in terms of deformation matrix M.



Note that the matrix $M(\tau)$ is invertible for all $\tau \in [0, \tau_f)$ but not for $\tau = \tau_f$ and so

$$\chi = \frac{dM}{d\tau} \times M^{-1}. \tag{5.10}$$

Recalling the formula for the derivative of the determinant

$$\frac{ddetM}{d\tau} = (detM) \cdot tr \left(\frac{dM}{d\tau} \times M^{-1} \right)$$

we obtain

$$tr\chi = \frac{1}{detM} \frac{ddetM}{d\tau}. \tag{5.11}$$

Note that the trace of χ is taken with respect to the metric $\not{g}$ of S_τ which, with respect to the Fermi frame, is simply the identity matrix. Since

$$\frac{ddetM}{d\tau} \leq 0, \quad \text{and} \quad detM \to 0^+$$

as $\tau \to \tau_f$, and therefore,

$$tr\chi(\tau_f) = -\infty. \tag{5.12}$$

Hence the geometry of the null hypersurface C breaks down at the focal points. The quantity

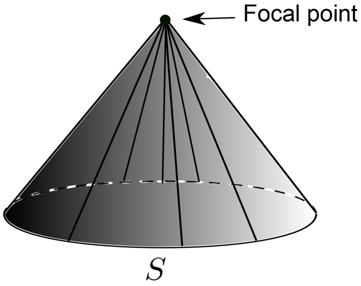
$$\min_{G_x, x \in S_0} \tau_f(G_x) \tag{5.13}$$

is called the radius of conjugacy of the null hypersurface C.

It is not always the case that nearby null generators intersect G_x at a focal point. If this happens then the point is called a *caustic point*.

Example

Recall that $tr\underline{\chi} = -\frac{2}{r}$ and hence note that indeed $tr\underline{\chi} \to -\infty$ as $r \to 0$ in Minkowski spacetime. Hence, the radius of conjugacy of the incoming null geodesic congruence normal to a standard sphere is equal to the radius of the sphere.



This behavior of the incoming null congruence of $\mathbb{S}^2(r)$ in Minkowski spacetime will be generalized in Section 5.4.

5.3 Causality III

Recall from Section 2.4 that if S is a closed two-dimensional surface in a globally hyperbolic time-orientable spacetime $(\mathcal{M}, g)$ and $C, \underline{C}$ are its (future) outgoing and incoming null geodesic congruence normal to S then

$$\partial \mathcal{J}^+(S) \subset C \bigcup \underline{C}.$$

Note that we always have

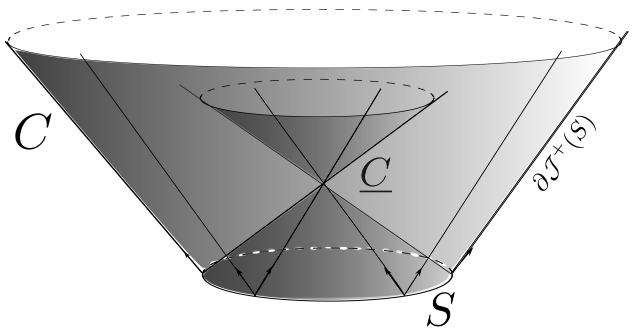
$$C \bigcup \underline{C} \subset \mathcal{J}^+(S),$$

however, it is *not* always the case that

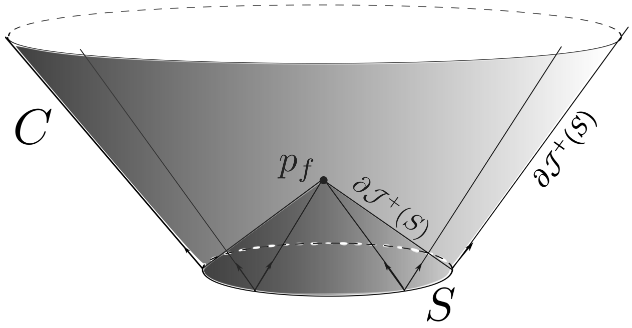
$$C \bigcup \underline{C} \subset \partial \mathcal{J}^+(S).$$

Although the boundary $\mathcal{J}^+(S)$ of the future of S is always part of $C \cup \underline{C}$, it may be the case that part of $C \cup \underline{C}$ lies in the interior of the future $\mathcal{J}^+(S)$. In this case, this part can be connected to S with timelike curves (and hence so does a neighborhood which thus belongs in the interior of the future of S).

For example, let us consider again the case a standard sphere S in Minkowski spacetime. In this case the null congruences $C \cup \underline{C}$ are depicted below



whereas the boundary of the future of S is the following



Hence, we can see that the part of $\underline{C}$ after the focal point p_f is not any more in the boundary of the future of S. This is in fact true for any globally hyperbolic Lorentzian manifold. Indeed, we have the following

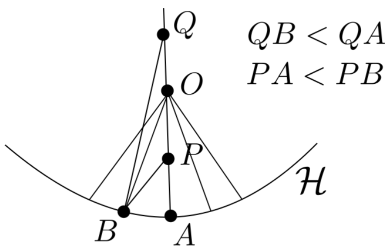
Proposition 5.3.1. *Let S be a closed two-dimensional surface in a globally hyperbolic time-orientable spacetime $(\mathcal{M}, g)$. Let C and $\underline{C}$ denote the (future) outgoing and incoming null geodesic congruence normal to S. Let C^* and $\underline{C}^*$ denote the part of C and $\underline{C}$, respectively, which do not contain any focal points. Then,*

$$\partial\mathcal{J}^+(S) \subset C^* \bigcup \underline{C}^*.$$

The proof of this proposition relies on a second variation of the part of the null generator G_x between the foot x in S and the focal point. The variation is appropriately taken to be in the direction of the normal Jacobi field which vanishes at the focal point. Then, this variation gives rise to infinitesimally close to G_x timelike curves connecting point of S and the focal point. Hence, the focal point is in the interior of $\mathcal{J}^+(S)$, from which it easily follows that the rest of the null generator G_x also lies in the interior of $\mathcal{J}^+(S)$.

Remark

A similar property of focal points is known in Riemannian, in fact even in Euclidean, geometry. For example, consider the following piece $\mathcal{H}$ of a circle in two-dimensional flat plane:



Note that if P is between A and the center O then the length of PA is the least distance of P from $\mathcal{H}$. On the other hand, the focal point O (where the normals to $\mathcal{H}$ intersect) is equidistant from all points of $\mathcal{H}$. Furthermore, if Q does line between O and A, then the length QA is **not** the least distance of Q from $\mathcal{H}$.

The above can be generalized to any complete Riemannian manifold. Note, finally, the importance of completeness for Riemannian geometry and global hyperbolicity for Lorentzian geometry.

5.4 Trapped Surfaces

Let S be a closed surface and $C \cup \underline{C}$ the null geodesic congruences normal to S and assume again $\Omega = 1$ on $C \cup \underline{C}$. We consider the geodesic vector fields $L, \underline{L}$ of $C, \underline{C}$, respectively, and let $\tau, \underline{\tau}$ be their affine parameters such that $S = \{\tau = 0\} = \{\underline{\tau} = 0\} \equiv S_0$.

The area and the second fundamental forms $\chi, \underline{\chi}$

Consider canonical coordinates $(\tau, \theta^1, \theta^2)$ on C (and similarly for $\underline{C}$), where $(\theta^1, \theta^2) \in \mathcal{U} \subset \mathbb{R}^2$ (see also Section 4.1). Let $\not{g}$ denote the induced metric on the sections S_τ of C. In view of the first variationa formula, and the formula for the derivative of the determinant of a matrix, we have

$$\nabla_L \left(\sqrt{det \not{g}} \right) = (tr\chi) \sqrt{det \not{g}}\,,$$

where $det \not{g}\,(\tau)$ denotes the determinant of the induced metric $\not{g}\,(\tau)$ on S_τ with respect to the coordinates $(\theta^1, \theta^2) \in \mathcal{U} \subset \mathbb{R}^2$. Note that

$$\text{Area}(S_\tau) = \int_{\mathcal{U}} \sqrt{det \not{g}\,(\tau)}\, d\theta^1 d\theta^2$$

and hence

$$\nabla_L \Big(\text{Area}(S_\tau) \Big) = \int_{\mathcal{U}} tr\chi \, d\mu_{\not{g}}\,, \tag{5.14}$$

and more generally,

$$\nabla_{fL} \Big(\text{Area}(S_\tau) \Big) = \int_{\mathcal{U}} f \cdot tr\chi \, d\mu_{\not{g}}\,, \tag{5.15}$$

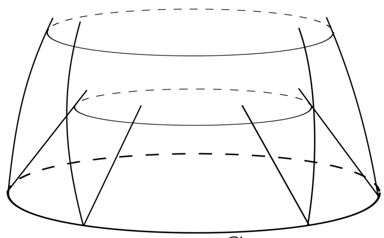
for any smooth non-negative function f on S_τ. *The equation* (5.15) *expresses the relation of the second fundamental form χ and the rate of change of the area of S_τ under infinitesimal displacements along the null generators of C. This is the reason $tr\chi$ is called the expansion of S_τ.*

Of course, the analogous equation is true for the hypersurface $\underline{C}$.

Definition of trapped surfaces

A trapped surface is, by definition, a closed two-dimensional surface S in $(\mathcal{M}, g)$ for which the area decreases under arbitrary (infinitesimal) displacements along the null generators of both null geodesic congruences $C \cup \underline{C}$ normal to S.

If $(\mathcal{M}, g)$ is globally hyperbolic, then since $C \cup \underline{C}$ bounds the future of S, we obtain that a trapped surface cannot expand in its future (hence the term trapped).



S

In view of formula (5.15), we have that for a trapped surface S, the integrals

$$\int_U f \cdot \mathrm{tr}\chi \, d\mu_{\not g} < 0, \qquad \int_U f \cdot \mathrm{tr}\underline{\chi} \, d\mu_{\not g} < 0$$

for all smooth non-negative functions f on S. Hence, an equivalent definition of the trapped surface is the following:

A trapped surface is a closed two-dimensional surface S in a Lorentzian manifold $(\mathcal{M}, g)$ such that the expansions

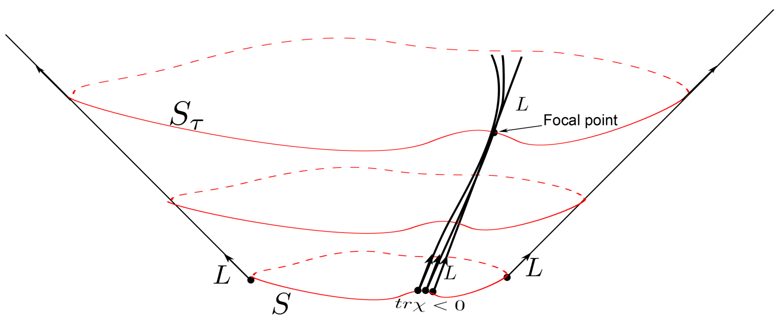
$$tr\chi < 0, \qquad tr\underline{\chi} < 0 \tag{5.16}$$

everywhere on S.

Trapped surfaces and focal points

The null generators of the incoming null hypersurface $\underline{C}$ contain focal points. Note also that the null expansion of a section of $\underline{C}$ is negative. The crucial observation is the following that these two properties are not unrelated:

Proposition 5.4.1. *Suppose that S is a closed two-dimensional surface (not necessarily trapped) in a Lorentzian manifold $(\mathcal{M}, g)$ which satisfies the Einstein equations $Ric(g) = 0$. If $tr\chi < 0$ at a point $x \in S$ then there exists a focal point on the null generator G_x of C emanating from the point x. A similar result holds for $\underline{C}$.*



Proof. In view of the Raychaudhuri equation (4.22), and (4.24), noting that $\omega = 0$ since $\Omega = 1$ on $C \cup \underline{C}$ we have

$$\nabla_L \big(tr\chi \big) = -\frac{1}{2}(tr\chi)^2 - |\hat{\chi}|^2 \leq 0. \tag{5.17}$$

Hence, if $T = tr\chi_x = tr\chi(0) < 0$ then $tr\chi(\tau) < 0$ for all $\tau \geq 0$. Therefore, by dividing with $(tr\chi)^2$ and forgetting about the shear term we obtain

$$\nabla_L \Big(-\frac{1}{tr\chi} \Big) \leq -\frac{1}{2} \tag{5.18}$$

and therefore

$$-\frac{1}{tr\chi} \leq -\frac{1}{T} - \frac{\tau}{2}. \tag{5.19}$$

Therefore, for

$$\tau_* = \frac{2}{-T} = \frac{2}{-tr\chi_x}, \tag{5.20}$$

we obtain

$$tr\chi(\tau_*) = -\infty \tag{5.21}$$

and hence the point $G_x(\tau_*)$ is the first focal point on G_x. □

Remark

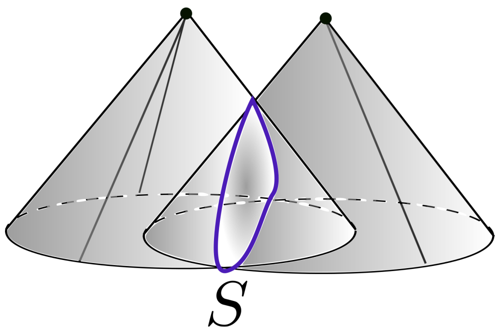
The equality part of (5.17) is true in view of the fact that $Ric(L, L) = tr\alpha = 0$. On the other hand for the inequality part of (5.17) we only need $tr\alpha = Ric(L, L) \geq 0$. The latter condition is known as the *positive null energy condition* and is weaker than the Einstein equations. Hence, Proposition 5.4.1 holds if $(\mathcal{M}, g)$ satisfies the positive null energy condition.

Trapped surfaces in Minkowski spacetime?

We saw earlier that the incoming null geodesic congruence $\underline{C}$ normal to a standard sphere S in Minkowski spacetime has everywhere negative expansion ($tr\chi = -2/r$) and every null generator of $\underline{C}$ contains a focal point (which is the vertex of the cone). On the other hand, the outgoing expansion is always positive ($tr\chi = 2/r$). The following question arises:

- *Can we deform the standard sphere in Minkowski spacetime so we obtain a trapped surface? More generally, is there a trapped surface in Minkowski spacetime?*

We will answer this question in the negative in the next section. For the time being, we remark that there are surfaces in Minkowski for which there are arbitrarily large regions where the incoming and outgoing null expansions $tr\underline{\chi}, tr\chi$ are both negative. Indeed, the idea is to deform a sphere so that it remains a section of the incoming null congruence $\underline{C}$ for which the expansion $tr\underline{\chi}$ (of any section of $\underline{C}$) is everywhere negative. For example, we can consider a surface S a large part of which lies in the intersection of the past null cones of two points p, q:



Note, however, that these regions do not cover the whole surface S. As mentioned above, the obstruction to obtaining surfaces with globally negative (both) null expansions is the subject of the next section.

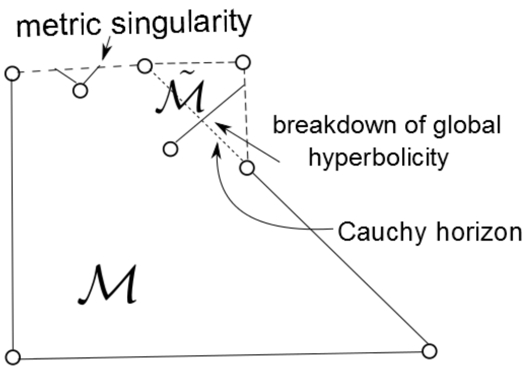
5.5 Penrose Incompleteness Theorem

Without further delay, let us state the theorem

Theorem 5.5.1 (Penrose Incompleness Theorem)**.** *Let $(\mathcal{M}, g)$ be a globally hyperbolic time-orientable (Hausdorff) spacetime with a non-compact Cauchy hypersurface $\mathcal{H}$ such that $\mathcal{M}$ contains a trapped surface S. If, in addition, $(\mathcal{M}, g)$ satisfies $Ric(L, L) \geq 0$ for all null vector fields L, then M is future null geodesically incomplete. In fact, at least one of the null generators of $C \cup \underline{C}$, the future null geodesic congruences normal to S, cannot be extended for all $\tau \geq 0$ in $\mathcal{M}$.*

Corollary 5.5.1. *In view of the geodesic completeness of Minkowski spacetime, there are no trapped surfaces.*

Remark 5.5.1. *The theorem, as shall be obvious by the proof, does not explain the reason why the null generator cannot be extended for infinite time in $\mathcal{M}$. It may be either that there is a metric (curvature) singularity somewhere along this null generator or that this null generator enters an extended manifold $\widetilde{\mathcal{M}}$ which is not globally hyperbolic. Both scenaria are depicted below:*



Quite surprisingly, the Penrose incompleteness theorem relies on a global topological argument. For this reason, we first summarize the basic results needed from differential topology:

1. A bijective continuous map from compact spaces to a Hausdorff space is a homeomorphism.
2. Let $\mathcal{N} \subset \mathcal{M}$ be an (injective) immersed topological submanifold with $\dim(N) = \dim(M) = n$ such that $\mathcal{N}$ is compact and $\mathcal{M}$ is a Hausdorff connected non-compact topological manifold. Then, $\partial_{mani}\mathcal{N} \neq \emptyset$, where $\partial_{mani}\mathcal{N}$ denotes the boundary of $\mathcal{N}$ in the sense of (topological) manifolds.

The second property can be easily proved by contradiction. Assume that $\partial_{mani}\mathcal{N} = \emptyset$. Then every point in $\mathcal{N}$ has a neighborhood homeomophic to $\mathcal{R}^n$ and hence, in view of the equal dimensions, it is also open in $\mathcal{M}$. By compactness, it follows that $\mathcal{N}$ is open in $\mathcal{M}$. However, $\mathcal{N}$ is compact and thus closed in $\mathcal{M}$, since $\mathcal{M}$ is Hausdorff. Hence, by connectedness, $\mathcal{M} = \mathcal{N}$, which contradicts the fact that $\mathcal{M}$ is non-compact.

Proof of Theorem 5.5.1. Recall that, from the results of Section 5.2, if $tr\chi_x = -k_x < 0$, $x \in S = \{\tau = 0\}$, then for the first focal point on the null generator $G_x \subset C$ appears at time $\tau = 2/k_x$. In view of the compactness of S we have

$$\sup_S tr\chi = k_C < 0, \qquad \sup_S tr\underline{\chi} = -k_{\underline{C}} < 0, \qquad \sup\left\{k_C, k_{\underline{C}}\right\} = k < 0.$$

We assume that $(\mathcal{M}, g)$ is future null geodesically complete. We consider the union $\mathcal{V}$ of all parts of the null generators of $C \cup \underline{C}$ for which $0 \leq \tau \leq 2/k$, that is

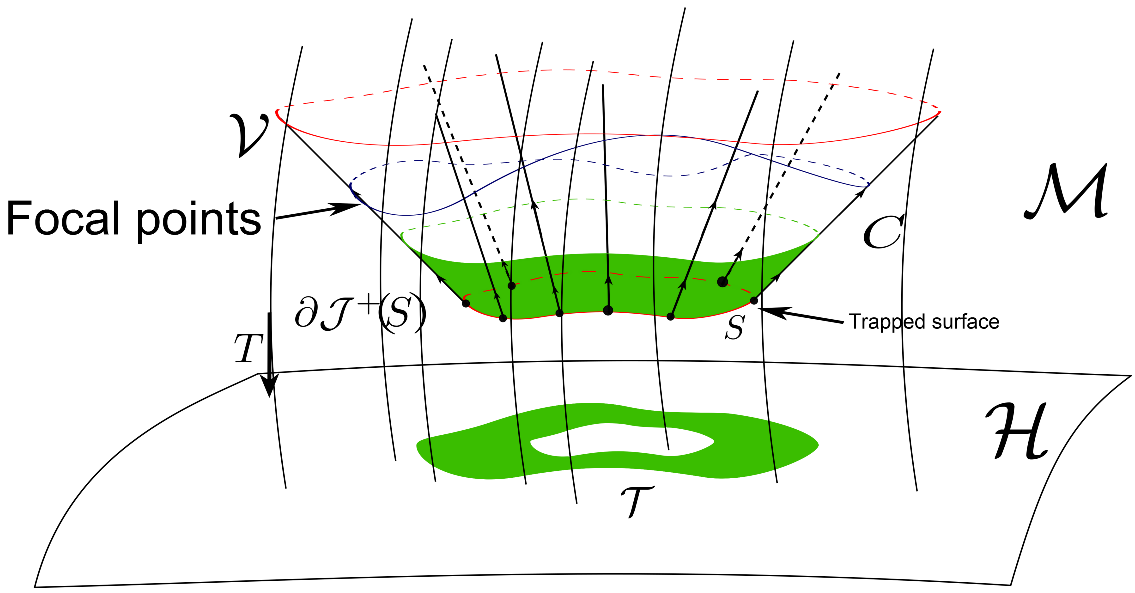
$$\mathcal{V} = \bigcup_{\substack{\tau \in [0,2/k] \\ x \in S}} \Big(G_x(\tau) \cup \underline{G}_x(\tau) \Big).$$

In view of our last assumption we have $\mathcal{V} \subset \mathcal{M}$. Furthermore, $\mathcal{V}$ is compact, since it is the union of two compact spaces, each being the image of the compact space $S \times [0, 2/k]$ under the continuous mapping which maps $(x, \tau) \in S \times [o, 2/k]$ to $G_x(\tau) \in \mathcal{V}$.

Clearly every null generator in $\mathcal{V}$ contains at least one focal point. Therefore, by the results of Section 5.3, we have

$$\partial \mathcal{J}^+(S) \subseteq \mathcal{V}.$$

Since V is compact and $\partial \mathcal{J}^+(S)$ (the topological boundary is closed by definition), we obtain that $\partial \mathcal{J}^+(S)$ is compact. We will next show that a global topological argument leads us to contradiction, and hence $(\mathcal{M}, g)$ cannot be future null geodesically complete.



In view of the time-orientability of $\mathcal{M}$, there is a global timelike vector field T whose integral curves are timelike foliate $\mathcal{M}$ and intersect the Cauchy hypersurface $\mathcal{H}$ exactly once. Furthermore, these integral curves intersect exactly once $\partial \mathcal{J}^+(S)$, since $\mathcal{J}^+(S)$ is a future set and the (topological) boundary of a future set is achronal three-dimensional Lipschitz submanifold without boundary in the sense of (topological) manifolds (see Proposition 2.1.2). The projection of $\partial \mathcal{J}^+(S)$ on $\mathcal{H}$ via the integral curves of T is thus an injective continuous mapping from $\partial \mathcal{J}^+(S)$ onto a subset $\mathcal{T}$ of $\mathcal{H}$. Therefore, by the property 1 above, we have that $\partial \mathcal{J}^+(S)$ is homeomorphic to $\mathcal{T}$ and thus $\mathcal{T}$ is a Lipschitz three-dimensional compact submanifold without boundary in the sense of (topological) manifolds ($\partial_{mani} \mathcal{T} = \emptyset$) in the non-compact three-dimensional manifold $\mathcal{H}$. Property 2 above shows that $\mathcal{T}$ must have non-empty boundary, which is contradiction.

$\square$

Remark 5.5.2. *Note that the general condition $Ric(L, L) \geq 0$ for all null vectors L guarantees that the volume of conical regions around null geodesics is smaller (or equal) than the Minkowskian regions and hence it can be considered as a focusing condition. Therefore, the Penrose theorem guarantees that an almost collapsed surface (i.e. a trapped surface) must lead to entirely collapsed geodesics (i.e. to focal points).*

5.6 Killing Horizons

We finish this chapter with a discussion about a special but very important class of null hypersurfaces. This special class of null hypersurface is characterized by the following property: Suppose that a null hypersurface C admits a normal ξ which can be extended in a spacetime neighborhood of $\mathcal{H}^+$ such that the extended ξ is Killing with respect to the metric g. Note that ξ does not have to be null outside the null hypersurface C.

Recall that a vector field ξ of $\mathcal{M}$ is called Killing if

$$\mathcal{L}_\xi g = 0.$$

This then implies that the map $T_\xi = \nabla\xi : X \mapsto \nabla_X \xi$ is antisymmetric since

$$g(\nabla_X \xi, Y) + g(\nabla_Y \xi, X) = 0, \tag{5.22}$$

for all vector fields X, Y of $\mathcal{M}$.

If C admits a Killing normal then it is called a *Killing horizon*. The Killing normal ξ does not have to coincide with the geodesic normal; if it does, i.e. if $\nabla_\xi \xi = 0$, then C is called *extremal horizon*. More generally, we will merely have

$$\nabla_\xi \xi = \kappa \xi, \tag{5.23}$$

for some function κ on C. The function κ is known as the *surface gravity* of C.

Remark

We could have defined a Killing horizon to be one such that $\mathcal{L}_\xi \not{g} = 0$, where $\not{g}$ is the induced metric on C and ξ null (and hence ξ is null Killing with respect to $\not{g}$). However, by the first variational formula, this is simply equivalent to $\chi = 0$ and does not depend on the normalization of ξ.

Properties of Killing horizons

1. Vanishing of the second fundamental form: $\chi = 0$

Proof. As before we assume that the null lapse function $\Omega = 1$ on C and L denote the geodesic vector field normal to C. We define the following $(0, 2)$ tensor field on C:

$$\chi_\xi(X, Y) = g(\nabla_X \xi, Y),$$

where $X, Y \in T_pC$, $p \in C$. Note that since X, Y are tangential to C, the values of χ_ξ depend only on the restriction of ξ on C. Since by assumption ξ is normal to C we have

$$\xi = f \cdot L,$$

where the function f on C is given by $f = -g(\xi, \underline{L})$. Then, since L is normal to C we obtain

$$\chi_\xi(X, Y) = g(\nabla_X fL, Y) = f \cdot \chi(X, Y). \tag{5.24}$$

Therefore, in view of the symmetry of χ, the tensor field χ_ξ is a symmetric (0,2) tensor field on C. On the other hand, in view of the Killing equation (5.22), the tensor field χ_ξ is also antisymmetric. Hence, $\chi_\xi = 0$ which, in view of (5.24), implies $\chi = 0$. □

2. <u>The relation of the surface gravity and the spacetime curvature: $\not\nabla\kappa = g(\xi, \underline{L}) \cdot \beta$.</u>

Proof. First note that κ is constant along the null generators of C; this can be easily shown by taking the pushforward of (5.23) under the flow of ξ (which consists of isometries). Hence, $L\kappa = 0$.

Since ξ is Killing it satisfies

$$\nabla^2_{X,Y}\xi := \nabla_X \nabla_Y \xi - \nabla_{\nabla_X Y}\xi = R(X, \xi)Y. \tag{5.25}$$

A proof of this relation is given at the end of this section. If X, Y are tangential to C, then we obtain

$$g(\nabla_X Y, L) = -g(Y, \nabla_X L) = 0$$

since $\chi = 0$. Hence $\nabla_X Y$ is also tangential to C and thus all the terms in (5.25) depend only on the restriction of ξ on C. Let us now assume that X is tangential to the sections S_τ of the affine foliation of C and $Y = \xi$. Then (5.25) becomes

$$\nabla_X \nabla_\xi \xi - \nabla_{\nabla_X \xi}\xi = R(X, \xi)\xi.$$

Note that $g(\nabla_X \xi, \xi) = \frac{1}{2} X\big(g(\xi, \xi)\big) = 0$ and that, if Z is any S vector field then $g(\nabla_X \xi, Z) = \chi_\xi(X, Z) = 0$. Therefore,

$$\nabla_X \xi = h \cdot \xi$$

for some function h on C (note that this function is related to the function f and the torsion ζ). Then, (5.25) becomes

$$R(X, \xi)\xi = \nabla_X(\kappa \cdot \xi) - h\nabla_\xi \xi = \big(\nabla_X \kappa\big) \cdot \xi + \kappa \nabla_X \xi - h \nabla_\xi \xi = \big(\nabla_X \kappa\big) \cdot \xi.$$

Taking the inner product with $\underline{L}$ we obtain

$$f^2 \cdot g\big(R(X, L)L, \underline{L}\big) = -f \nabla_X \kappa$$

and hence,

$$\nabla_X \kappa = -f \cdot R(\underline{L}, L, X, L) = g(\xi, \underline{L}) \cdot R(X, L, \underline{L}, L) = g(\xi, \underline{L}) \cdot \beta(X).$$

□

3. <u>The bifurcate sphere $\mathcal{S}$</u>

Recall that $\nabla_\xi \xi = \kappa\xi$, $\nabla_L L = 0$ and $\xi = f \cdot L$. Then,

$$\nabla_\xi \xi = f\nabla_L(fL) = f(Lf) \cdot L = (LF) \cdot \xi.$$

Therefore, $Lf = \kappa$, or in other words,

$$f(t) = \kappa \cdot t + c,$$

where t is the affine parameter of L and $c \in \mathbb{R}$. Then,

$$\xi = (\kappa \cdot t + c) \cdot L$$

and hence, if $\kappa \neq 0$, then for the section $\mathcal{S}$ of C which corresponds to $t = -\frac{c}{\kappa}$ is called the bifurcate sphere of C. Note that the bifurcate sphere $\mathcal{S}$ exists if the horizon is non-extremal. Furthermore, $\mathcal{S}$ coincides with a section of the affine foliation if and only if the surface gravity κ is globally constant on C. Clearly, an extremal horizon does not admit a bifurcate sphere.

4. <u>Conservation of the transversal second fundamental form: $\mathcal{L}_L \underline{\chi} = 0$ for extremal horizons.</u>

Proof. Let C be an extremal horizon, i.e. $\kappa = 0$. In this case, we can take the geodesic vector field $L = \xi$. Let X, Y be S tangential normal Jacobi vector fields, i.e. $[L, X] = [L, Y] = 0$.

In view of Corollary 5.6.2 we obtain

$$\mathcal{L}_L\big(\nabla_X Y\big) = \nabla_{\mathcal{L}_L X} Y + \nabla_X\big(\mathcal{L}_L Y\big) = 0. \tag{5.26}$$

On the other hand, in view of the first variational formula and $\chi = 0$, we have $\mathcal{L}_L \not{g} = 0$. Hence,

$$\mathcal{L}_L\big(\not{\nabla}_X Y\big) = 0. \tag{5.27}$$

Moreover,

$$\nabla_X Y = \not{\nabla}_X Y + \underline{\chi}(X, Y) \cdot L + \chi(X, Y) \cdot \underline{L} = \not{\nabla}_X Y + \underline{\chi}(X, Y) \cdot L.$$

Therefore, in view of (5.26), (5.27) we obtain

$$\mathcal{L}_L\big(\underline{\chi}(X, Y)\big) = 0, \tag{5.28}$$

which completes the proof (since $\mathcal{L}_L X = \mathcal{L}_L Y = 0$).

□

Properties of Killing fields

We finish this section by recalling several very important properties of Killing fields:

Lemma 5.6.1. *Let ξ be a Killing field of a pseudo-Riemannian manifold $(\mathcal{M}, g)$. Then for all vector fields X, Y the following holds*

$$\nabla^2_{(X,Y)}\xi = R(X, \xi)Y. \tag{5.29}$$

Proof. By taking the covariant derivative of the Killing equation (5.22) with respect to a vector field Z we obtain

$$g(\nabla_Z\nabla_X\xi, Y) + g(\nabla_X\xi, \nabla_Z Y) + g(\nabla_Z\nabla_Y\xi, X) + g\big(\nabla_Y\xi, \nabla_Z X\big) = 0,$$

or, using (5.22)

$$g(\nabla_Z\nabla_X\xi, Y) - g(\nabla_{\nabla_Z Y}\xi, \nabla_X) + g(\nabla_Z\nabla_Y\xi, Y) - g\big(\nabla_{\nabla_Z X}\xi, Y\big) = 0,$$

and thus

$$g\Big(\nabla^2_{(Z,X)}\xi, Y\Big) + g\Big(\nabla^2_{(Z,Y)}\xi, X\Big) = 0.$$

Similarly we obtain

$$g\Big(-\nabla^2_{(X,Y)}\xi, Z\Big) + g\Big(-\nabla^2_{(X,Z)}\xi, Y\Big) = 0,$$

$$g\Big(\nabla^2_{(Y,Z)}\xi, X\Big) + g\Big(\nabla^2_{(Y,X)}\xi, Z\Big) = 0.$$

By summing the previous three equations and the definition of the Riemann curvature we obtain

$$g\Big(\nabla^2_{(Z,Y)}\xi, X\Big) + g\Big(\nabla^2_{(Y,Z)}\xi, X\Big) + g\Big(R(Z,X)\xi, Y\Big) + g\Big(R(Y,X)\xi, Z\Big) = 0.$$

Furthermore,

$$2g\Big(\nabla^2_{(Z,Y)}\xi, X\Big) + g\Big(R(Y,Z)\xi, X\Big) + g\Big(R(Z,X)\xi, Y\Big) + g\Big(R(Y,X)\xi, Z\Big) = 0,$$

or

$$2g\Big(\nabla^2_{(Z,Y)}\xi, X\Big) + R(X,\xi,Y,Z) + R(Y,\xi,Z,X) + R(Z,\xi,Y,X) = 0.$$

Using $R(Y,\xi,Z,X) = R(X,Z,\xi,Y)$, $R(Z,\xi,Y,X) = -R(X,Y,Z,\xi)$ and the first Bianchi identity which reads

$$R(X,\xi,Y,Z) + R(X,Z,\xi,Y) = -R(X,Y,Z,\xi),$$

we obtain

$$2g\Big(\nabla^2_{(Z,Y)}\xi, X\Big) = 2R(X,Y,Z,\xi)$$

from which the required result follows by relabeling. □

Corollary 5.6.1. *A Killing field ξ is completely determined everywhere by the values of ξ and $\nabla\xi$ at a point $p \in \mathcal{M}$. Hence, the dimension of the linear space of Killing vector fields is at most $\frac{n(n+1)}{2}$, where $n = dim\mathcal{M}$.*

Proof. Consider another point q and a curve γ which connects p and q. Then (5.29) corresponds to a system of linear second order ODEs for the components of $\xi, \nabla\xi$. Moreover, ξ_p lies in an n-dimensional vector space, whereas $(\nabla\xi)_p$ lies in an $\frac{n(n-1)}{2}$-dimensional vector space, since it is axisymmetric.

□

Corollary 5.6.2. *Let ξ be a Killing vector field. Then*

$$\mathcal{L}_\xi\big(\nabla_X Y\big) = \nabla_{\mathcal{L}_\xi X} Y + \nabla_X\big(\mathcal{L}_\xi Y\big).$$

Proof. The proof follows immediately from Lemma 5.6.1 and the identity $\mathcal{L}_X Y = \nabla_X Y - \nabla_Y X$. □

Another corollary of Proposition 5.6.1 is the following:

Corollary 5.6.3. *Let ξ be a Killing field. Then we have*

$$\mathcal{L}_\xi R = \mathcal{L}_\xi Ric = \mathcal{L}_\xi R_{sc} = 0,$$

where R, Ric, R_{sc} is the Riemann, Ricci and scalar curvature, respectively.

Alternatively, one could use the definition of $\mathcal{L}_\xi$ using the flow of ξ, which consists of isometries.

Proof. From (5.6.1), for all vector fields X, Y we have

$$\mathcal{L}_\xi(\nabla_X Y) = \nabla_{\mathcal{L}_\xi X} Y + \nabla_X(\mathcal{L}_\xi Y),$$

where we have used that $\mathcal{L}_\xi X = \nabla_\xi X - \nabla_X \xi$. This applied several times implies

$$\mathcal{L}_\xi\Big(R(X,Y)Z\Big) = R(\mathcal{L}_\xi X, Y)Z + R(X, \mathcal{L}_\xi)Z + R(X,Y)\mathcal{L}_\xi Z,$$

from which it follows that $\mathcal{L}_\xi R = 0$. Since now Ric is the contraction of the tensor product of g^{-1} and R and the Lie derivative $\mathcal{L}_\xi$ commutes with contractions, satisfies the Leibniz rule and also $\mathcal{L}_\xi g^{-1} = 0$ we obtain $\mathcal{L}_\xi Ric = 0$. Similar argument applies for the scalar curvature. □

We can now prove the following

Proposition 5.6.1. *Let ξ be a Killing field. Then the following identities hold*

$$\begin{aligned}
&Div(\xi) = 0,\\
&Div\Big(R_{sc}\xi\Big) = 0,\\
&\Box_g \xi = Div(\nabla\xi) = Ric(\xi,\cdot),\\
&Div\Big(Ric(\xi,\cdot)\Big) = 0,
\end{aligned}$$

and hence the current

$$J = Div(\nabla\xi) \tag{5.30}$$

is conserved (i.e. divergence-free).

Proof. The first identity is trivial since for Killing ξ the $(1,1)$ tensor $\nabla\xi$ is antisymmetric and hence its trace vanishes. For the second identity observe that

$$Dic(R_{sc}\xi) = R_{sc}Div(\xi) + \nabla_\xi R_{sc} = 0.$$

Regarding the third identiy we have

$$\Box_g \xi = Div(\nabla\xi) = trace_{X,Y}\Big(\nabla^2_{(X,Y)}\xi\Big) = trace_{X,Y}\Big(R(X,\xi)Y\Big) = Ric(\xi,\cdot),$$

where we have used (5.29). For the last one, recall that the Einstein tensor

$$G = Ric - \frac{1}{2}R_{sc}\, g$$

satisfies $Div(G) = 0$. Moreover,

$$Div\Big(G(\xi,\cdot)\Big) = 0,$$

since

$$Div\Big(G(\xi,\cdot)\Big) = \Big(Div(G)\Big)(\xi) + \sum_{i=1}^{n} G(\nabla_{E_i}\xi, E_i),$$

and using the fact that G is symmetric and that $\nabla\xi$ is antisymmetric and noting that $\nabla_{E_i}\xi = \sum_{j=}^{n} g(\nabla_{E_i}\xi, E_j)E_j$. Furthermore,

$$Div\Big(G(\xi,\cdot)\Big) = Div\Big(Ric(\xi,\cdot)\Big) - \frac{1}{2}Div\Big(R_{sc}\, g(\xi,\cdot)\Big)$$

and since

$$Div\Big(R_{sc}\, g(\xi,\cdot)\Big) = Div\Big(R_{sc}(\xi_\flat)\Big) = Div\Big((R_{sc}\xi)_\flat\Big) = Div\Big(R_{sc}\xi\Big) = 0,$$

we obtain the required result.

Alternatively, note that $Div\Big(Div(\nabla\xi)\Big) = 0$, because $\nabla\xi$ is antisymmetric and the divergence of the divergence of an antisymmetric tensor vanishes. $\Box$

Corollary 5.6.4. *If $(\mathcal{M}, g)$ is Ricci flat (i.e. $Ric = 0$) and ξ is a Killing field, then ξ satisfies the Maxwell equations.*

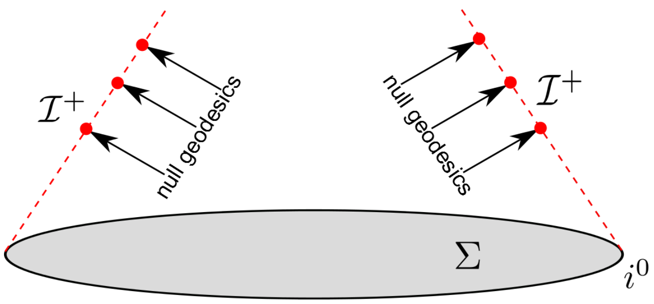
Chapter 6

Christodoulou's Memory Effect

In this chapter we describe the celebrated nonlinear memory effect of the gravitational field due to Christodoulou. The only way we can study isolated systems in the universe is by investigating the radiation/signals that reaches us from these systems. We therefore, need a well-defined notion for the region where radiation scatters. This is precisely the future null infinity $\mathcal{I}^+$.

6.1 The Null Infinity $\mathcal{I}^+$

Heuristically, the future null infinity $\mathcal{I}^+$ consists of all ideal limit points of null geodesics which reach arbitrarily large spatial distances.



Definition

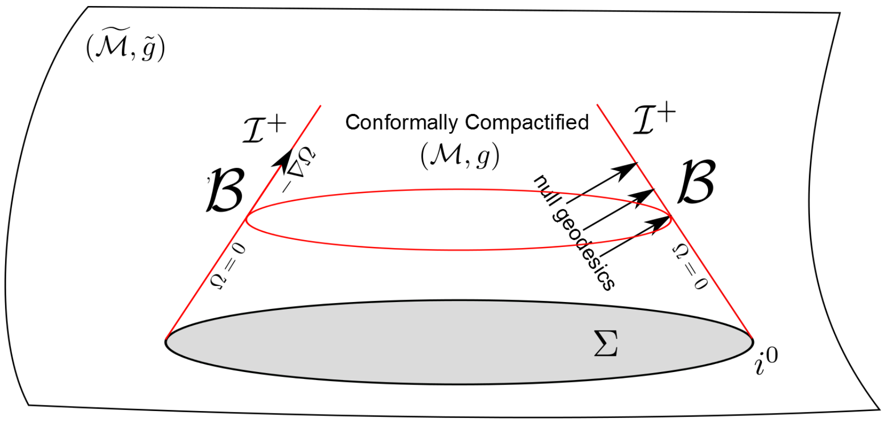
One can give a precise meaning to this construction by considering conformal transformations. Indeed, conformal transformations do not change the causal structure and, in fact, send null geodesics to null geodesics. Recall (see Section 1.4.2) that Minkowski spacetime can be conformally embedded in Einstein's static universe $\mathcal{E}$ such that the closure of the image is compact and hence has a boundary in $\mathcal{E}$. Recall also that this boundary consists of several components each of which is of fundamental importance in the study of the causal structure of the spacetime.

In this section we shall generalize these notions to more general Lorentzian manifolds. Let, therefore, $(\mathcal{M}, g)$ be a globally hyperbolic time-orientable Lorentzian manifold. Suppose that $(\mathcal{M}, g)$ can be conformally embedded in another Lorentzian manifold $(\tilde{\mathcal{M}}, g)$ such that the closure of the image (also known as the conformal compactification) has a boundary, which consists of two components; the future boundary and the past boundary, both intersecting at a point i^0. We shall refer to this point as the spacelike infinity.

Let us consider the future boundary $\mathcal{B}$ of the conformal compactification of $(\mathcal{M}, g)$. Let Ω^2 be the conformal factor which we assume that it satisfies the following properties

1. $\Omega > 0$ on $\mathcal{M}$,
2. Ω extends appropriately to $\mathcal{B}$ such that $\Omega = 0$ on $\mathcal{B}$,
3. Ω satisfies the eikonal equation $g(\nabla\Omega, \nabla\Omega) = 0$ on $\mathcal{B}$.

Then the future boundary $\mathcal{B}$ is a null hypersurface since it is a level set of Ω and hence its normal is the null vector field $\nabla\Omega$. We shall call this null hypersurface the future null infinity $\mathcal{I}^+$. Note that $\mathcal{I}^+$ is an incoming null hypersurface. For, $\mathcal{I}^+$ is the future boundary of a set.

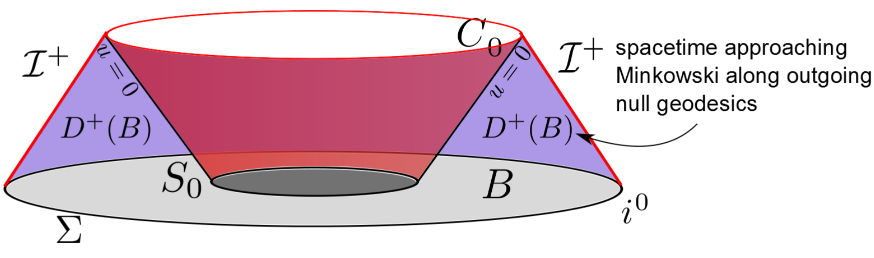


Asymptotical flatness and null infinity

Of course there are spacetime which do not admit such a null infinity. However, any asymptotically flat spacetime does so. Let us explain this further:

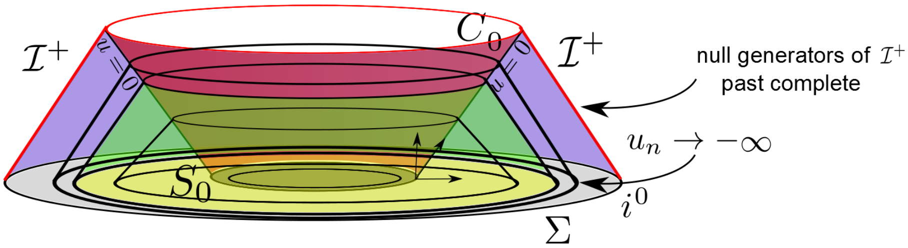
We are mostly interested in studying isolated systems in the universe. Hence, we can assume that arbitrarily far away from these systems the spacetime approaches Minkowski spacetime. Since general relativity is a dynamical theory we can only impose restrictions on a Cauchy hypersurface (say $\Sigma \sim \mathbb{R}^3$), since then the domain of dependence of Σ is uniquely determined. Hence, let us assume that the data on Σ are asymptotically flat, i.e. approach Minkowskean data at infinity. Then, there exists a sphere S_0 in Σ such that the data on its complement B in Σ is a small perturbation of Minkowski data.

Then by the stability of Minkowski spacetime proved by Christodoulou–Klainerman and Klainerman–Nicolo we can conclude that one can attach a piece of future null infinity at the domain of dependence $D^+(B)$ of B:



Note that the aforementioned works on the stability of Minkowski proved that the conformal factor Ω^2 extends to $\mathcal{I}^+$ as a function in $C^{1,\frac{1}{2}}$. In fact, Christodoulou has shown that for physically relevant spacetime the conformal factor Ω^2 cannot belong in C^2 (for more see also Part II of these notes).

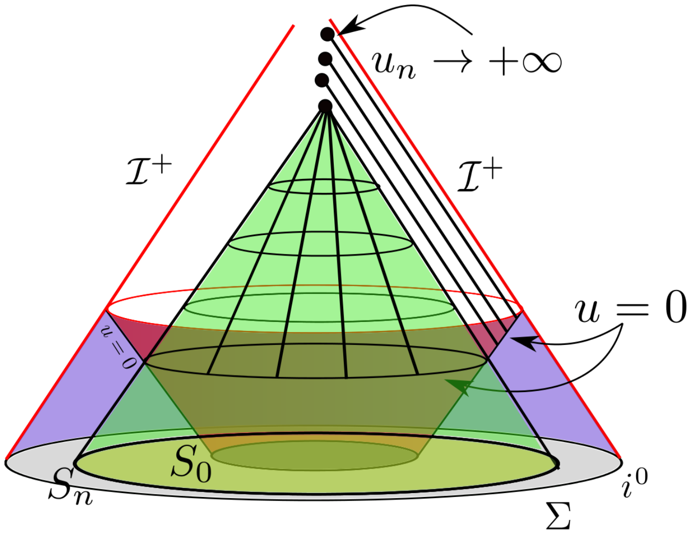
The works on the stability of Minkowski showed also that the null generators of this piece of null infinity are past complete in the following sense: Let C_0 be the outgoing null geodesic congruence normal to S_0. Let L be the outgoing null vector field normal to S_0 given by $L = \frac{1}{\sqrt{2}}(T + N)$, where T is the unit future directed vector field normal to Σ and N is the outward normal to S_0 in Σ. This choice then yields an affine foliation S_τ of C_0 (for which we can take $\tau \to +\infty$). For each section of this foliation we consider the affine foliation of the incoming null hypersurfaces as depicted below:



Then the affine time for which null generators of the incoming null hypersurfaces intersect Σ tends to infinity as $\tau \to +\infty$.

Completeness of null infinity

As we just explained, for asymptotically flat spacetimes the future null infinity is past complete. However, it is of fundamental significance to know if the future null infinity is also future complete, in a similar sense as above:



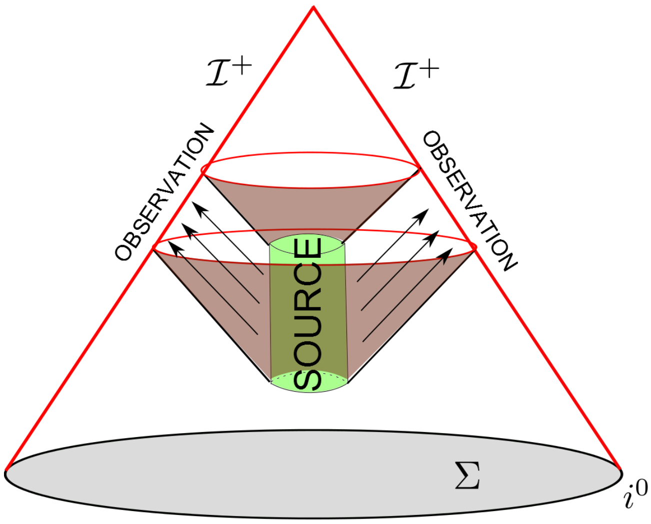
In this case, we measure the affine time u_τ for which the depicted null generators remain in the domain of dependence of the sections S_τ of C_0 and we require $u_\tau \to +\infty$ as $\tau \to +\infty$.

Weak cosmic censorship

Generic initial data for the Einstein equations lead to developments with complete future null infinity.

6.2 Tracing gravitational waves

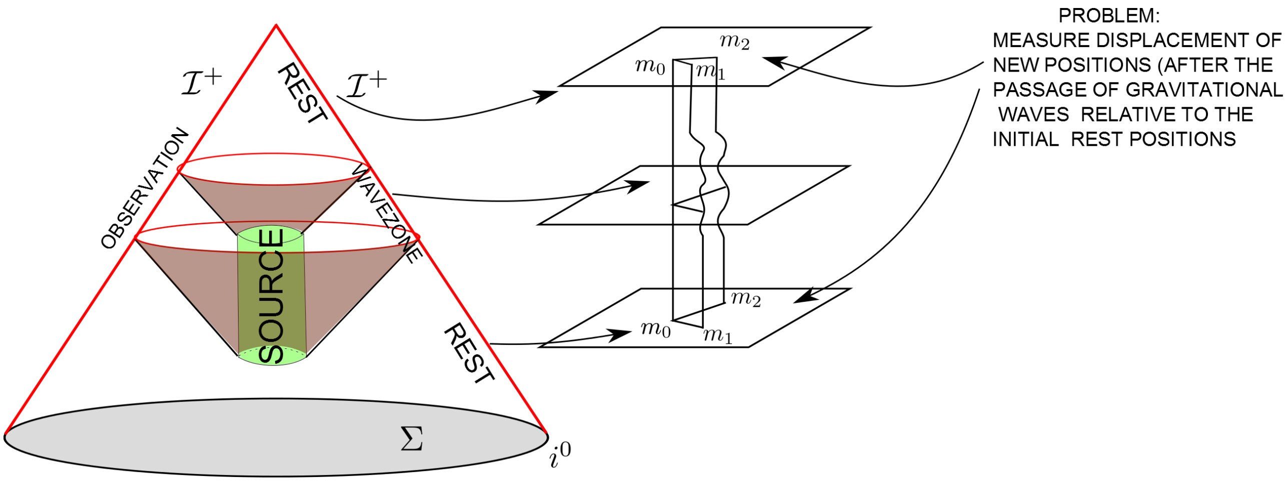
In this section we describe the basic setup of a gravitational wave experiment. One of the fundamental assumptions is that we, the observers, are located very far away from the source of the gravitational waves and hence we only rely on the gravitational radiation which reaches us. Hence we assume the experiment takes place along future null infinity.



The experiment is based on the following principle: "A gravitational wave forces masses to change their position". However, how can one measure this displacement? Considering just one mass would be hopeless, since masses move on timelike geodesics. Therefore, one needs to consider two or more masses and study their relative motion. If we find a difference in the relative distance of the masses then we could deduce that a gravitational wave has passed.

The most basic setup would be to consider two masses. However, in order to study completely the problem, we consider three (free) masses m_0, m_1, m_2 which lie initially (i.e. when the masses are at rest) on a plane $\Pi = \Pi(0)$. We assume that m_0, m_1, m_2 form initially a right isosceles triangle on $\Pi(0)$ and that the size of the equal sides is d.

We are then interested in investigating the relative motions of the pairs (m_0, m_1) and (m_0, m_2) and comparing these motions too.



Let T be tangent to the timelike geodesic γ of m_0 such that $\nabla_T T = 0$. Let also assume that initially

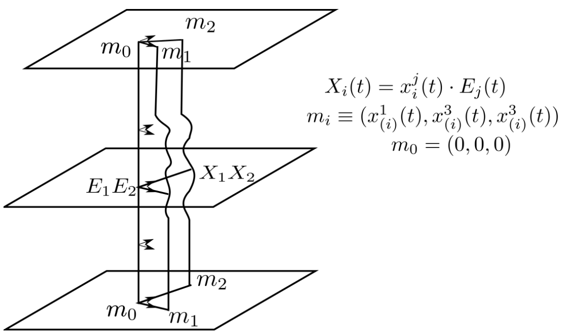
$$m_0 m_1 = X_1(0) = d \cdot E_1(0), \qquad m_0 m_2 = X_2(0) = d \cdot E_2(0),$$

where $(E_1(0), E_2(0))$ form an orthonormal frame on $\Pi(0)$. We then completely the orthonormal frame of the three-dimensional space: $(E_1(0), E_2(0), E_3(0))$. We parallely propagate this frame along γ according to $\nabla_T E_i = 0$, obtaining therefore an orthonormal frame $(E_1(t), E_2(t), E_3(t))$ known as the Fermi frame. Here t denotes the affine parameter of T.

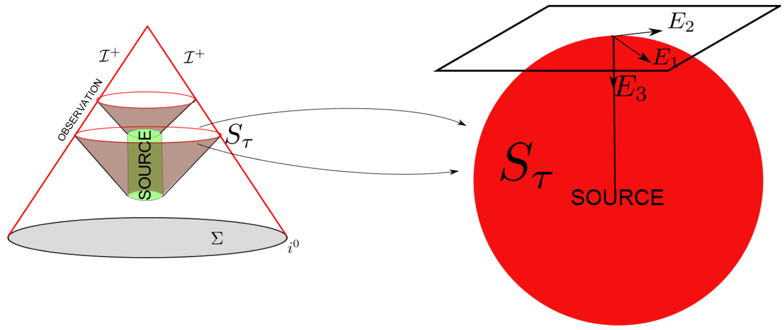
A priori the motion of the masses may not be constraint in the planes $\Pi(t) = \langle E_1(t), E_2(t) \rangle$. However, under very reasonable physical assumptions, the timelike geodesics of m_1, m_2 can be approximated by the associated Jacobi fields X_1, X_2 along γ given by following relations:

$$\begin{aligned} X_1 &= d \cdot E_1, \qquad \nabla_T X_1 = 0, \\ X_2 &= d \cdot E_2, \qquad \nabla_T X_2 = 0. \end{aligned}$$

Note that the relations $\nabla_T X_i = 0, i = 1, 2$ represent the fact that the masses are assumed to be at rest with respect to m_0 (hence no gravitational waves are present at the beginning of the experiment).

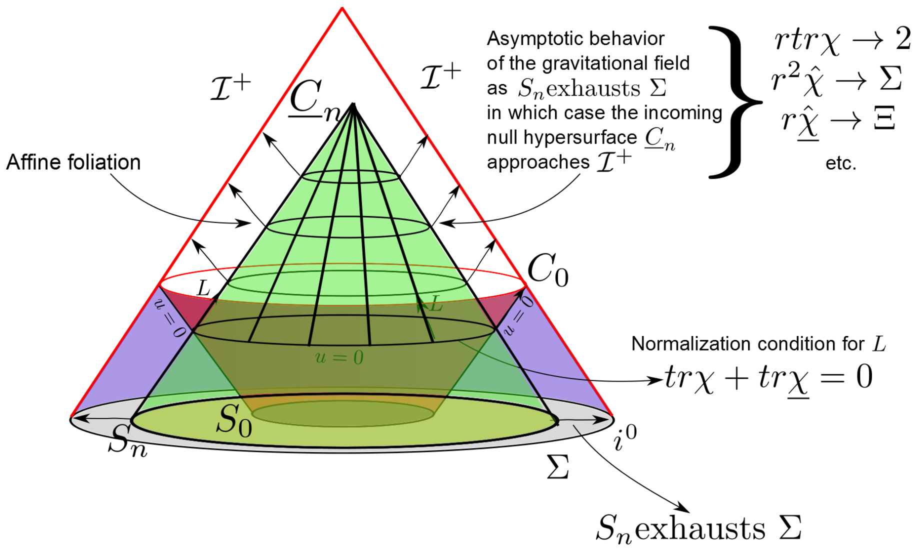


Recall that the whole configuration is supposed to propagate along a null generator of future null infinity. We can assume therefore, that the planes $\langle E_1(t), E_2(t) \rangle$ are tangential to the sections of null infinity along a fixed null generator. In this case, $E_3(t)$ points towards the direction of the source:



6.3 Peeling and Asymptotic Quantities

Consider the sections S_τ of the affine foliation of C_0 and the incoming null geodesic congruence $\underline{C}_\tau$ normal to S_τ. Renormalize $L, \underline{L}$ on $C_0 \cap \underline{C}_\tau$ such that $tr\chi + \mathrm{tr}\underline{\chi} = 0$. Consider then the associated affine foliation $\underline{S}_u$ of $\underline{C}_\tau$. Then, from the work of the stability of Minkowski spacetime (see Part II), we have precise limits for all appropriately rescaled geometric quantities (metric, connection coefficients, curvature components) of $\underline{S}_u$. The rescalings involve the radius coordinate r given by the area of the sections. Note that $r \to +\infty$ at $\mathcal{I}^+$.



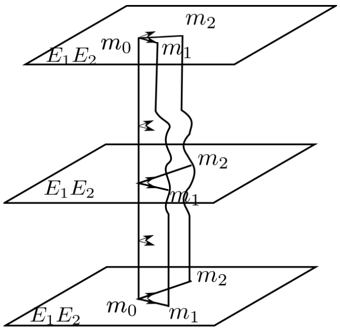
By renormalizing the null structure equations we obtain the asymptotic null structure equations for the asymptotic gravitational field along the null infinity $\mathcal{I}^+$.

6.4 The Memory Effect

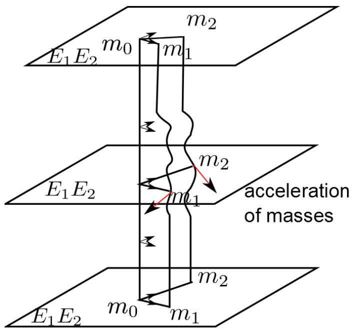
Assume the source is perpendicular to the horizontal plane spanned by E_1, E_2.

We will next describe the qualitative conclusions of the memory effect. For more precise quantitative results see Part II of these notes.

1. The motion of the masses is constrained in the horizontal plane $\langle E_1, E_2 \rangle$. Hence no movement in the vertical E_3 direction is to be observed.

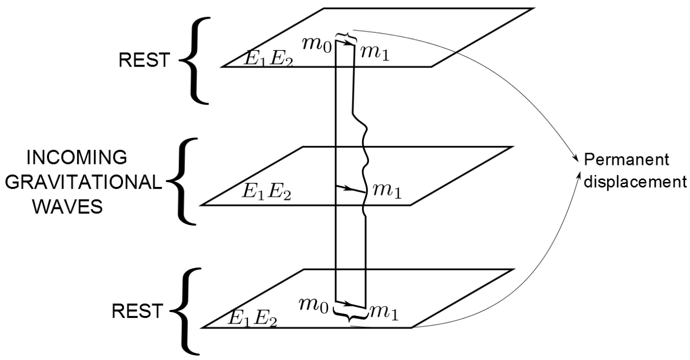


2. Acceleration of m_1 relative to m_0 is at each instance the rotation by 90^o degrees of the acceleration of m_2 relative to m_0.



3. The masses $m_{(i)}$ suffer a permanent displacement $\Delta x_{(i)}$ given by asymptotic quantities of the gravitational field (i.e. the asymptotics of the Riemann tensor).

This permanent displacement is known as the Christodoulou memory effect.



Chapter 7

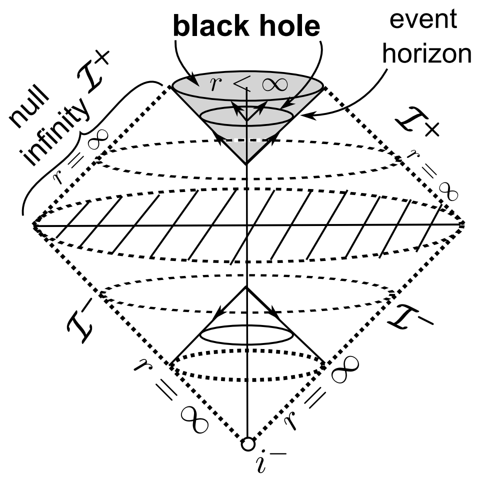
Black Holes

In this chapter we introduce the notion of a black hole region. We will discuss the intimate relation of black holes and trapped surfaces and then prove, using the machinery that we have developed so far general properties of such regions (known as black hole mechanics). We will next describe the main properties of the most basic black hole, namely the Schwarzschild black hole, and then proceed by presenting the Kerr metric.

7.1 Introduction

Recall that the future null infinity consists of the ideal endpoints of future-directed null geodesics which reach arbitrarily large distances. A region which cannot send signals to arbitrarily large distances is called *black hole*. Formally speaking, the black hole region in a spacetime $\mathcal{M}$ is the non-empty complement of the past of the future null infinity, i.e. $M/J^-(\mathcal{I}^+)$. The complementary region, i.e. the region which can communicate with $\mathcal{I}^+$ is called *domain of outer communications*. Their common boundary is called *future event horizon*.

Remark 7.1.1. *A usual additional condition in the definition of the black hole region is the* <u>*completeness*</u> *of future null infinity $\mathcal{I}^+$. This condition guarantees that ideal observes along $\mathcal{I}^+$ (such as ourselves) live* <u>*forever*</u> *yet never receive signals from the black hole region.*



7.2 Black Holes and Trapped Surfaces

The presence of a trapped surface in a vacuum spacetime implies the existence of a black hole region.

Proposition 7.2.1. *Let $(\mathcal{M}, g)$ be a vacuum spacetime. If S is a closed trapped surface in $\mathcal{M}$ then S cannot lie in the domain of outer communications $J^{-}(\mathcal{I}^{+})$. Hence $\mathcal{M}$ must have a black hole region and S must lie in it.*

Proof. Assume in the contrary that $S \subset J^{+}(\mathcal{I}^{+})$. Clearly there exists $p_1 \in \mathcal{I}^{+}$ such that $p_1 \notin J^{+}(S)$ and $p_2 \in \mathcal{I}^{+}$ such that $p_2 \in J^{+}(S)$. Hence, there must exist $p \in \mathcal{I}^{+}$ such that $p \in \partial J^{+}(S)$. However, in view of Proposition 2.4.2 we have $p \in C$, where C is the outgoing null geodesic congruence normal to S. Hence p lies on a null generator γ of C and since $p \in \mathcal{I}^{+}$ it follows that γ must be complete and always in $\partial J^{+}(S)$. However, as is shown in Section 5.4, since S is trapped, all null generators of $J^{+}(S)$ have finite affine length, contradiction.

□

Although the above proposition does not show that $\mathcal{I}^{+}$ must be complete, nonetheless it makes clear the need to study evolutionary criteria for the formation of trapped surfaces. For more on this see Part II of these notes.

7.3 Black Hole Mechanics

Let $(\mathcal{M}, g)$ contain a black hole region. The event horizon $\mathcal{H}^{+}$ must be a null hypersurface since it is the boundary of the past of a set. Of particular importance is the case where $\mathcal{H}^{+}$ is a Killing horizon (see Section 5.6). We then have the following

Proposition 7.3.1 (Zeroth law of black hole mechanics)**.** *If a Lorentzian manifold $(\mathcal{M}, g)$ satisfies the Einstein equations $Ric(g) = 0$ and contain a Killing horizon $\mathcal{H}^{+}$. Then the surface gravity κ is constant on $\mathcal{H}^{+}$.*

Proof. Recall that in Section 5.6 it was shown that

$$\nabla\!\!\!/ \kappa \sim \beta.$$

Hence in order to show that κ is constant on $\mathcal{H}^{+}$ is suffices to show that $\beta = 0$. However, this follows immediately from the null Codazzi equation 4.21 and the fact that $\chi = 0$ on the Killing horizon $\mathcal{H}^{+}$.

□

The remaining laws can be found in any standard textbook. We here mention that the first law is related to the conservation of energy and the second law is related to the non-decrease of the entropy. Finally, according to the third law, spacetimes which contain extremal black holes, i.e. such that $\kappa = 0$, should not form dynamically (see Chapter 10 for more a discussion of the wave propagation on such spacetimes).

7.4 Spherical Symmetry

7.4.1 General Setting

- A spacetime $(\mathcal{M}, g)$ is called *spherically symmetric* if $SO(3)$ acts by isometry and the orbits of this group are two-dimensional spheres. The area-radius coordinate $r = (A/4\pi)^{1/2}$ is very important (A is the area of the orbits). Then, in spherical coordinates (θ, ϕ) the induced metric on the symmetry spheres is $r^2(d\theta^2 + \sin^2\theta d\phi^2)$.

Birkhoff's theorem: The one-parameter Schwarzschild metrics g_M is the unique spherically symmetric one-parameter of metrics g which satisfy $Ric(g) = 0$.

It turns out that these metrics admit additional symmetries. We have the following definitions:

- A spacetime $(\mathcal{M}, g)$ is called *stationary*, if it admits a timelike Killing field, that is to say, if it admits a one-parameter family of isometries (the flow of the Killing field) whose orbits are timelike curves.
- A spacetime $(\mathcal{M}, g)$ is called *static*, if it is stationary and if, in addition, there exists a spacelike hypersurface Σ such which is orthogonal to the stationary Killing vector field.

Derivation of the metric

First, it is somewhat easier to assume that the spacetime is static and spherically symmetric (recall that Birkhoff's theorem does not require staticity; see below). Let us assume, in fact, that the spacetime admits a unique static Killing field T. Let Σ be a spacelike hypersurface orthogonal to T. One can easily see that Σ is foliated by spheres of symmetry. To show this it suffices to show that T is orthogonal to the spheres of symmetry. Indeed, since T is unique in the above sense, it must be invariant under the action of $SO(3)$; therefore, its projection on the spheres of symmetry must be an invariant vector field and hence it must vanish.

Coordinates on Σ: Let (θ, ϕ) be coordinates on a sphere of symmetry S embedded in Σ. If $p \in S$ then there exists a unique geodesic starting at p with initial velocity v such that $v \in T_p\Sigma$ and $v \perp T_pS$. Then we define (θ, ϕ) to be constant along such geodesics. We parametrize the spheres by the area coordinate r, obtaining the system (r, θ, ϕ) for Σ.

Coordinates on $\mathcal{M}$: We next use the staticity. Let $\Sigma = \Sigma_0$. Let t be the parameter of integral curves of T such that $t = 0$ on Σ_0. Then, $\mathcal{M} = \cup\Sigma_t$, where Σ_t is the image of Σ_0 under the flow of T. In this way, we can pushforward the system of Σ_0 onto $\mathcal{M}$ obtaining the system (t, r, θ, ϕ). It is easy to see that the metric with respect to this system must take the form

$$g = -f(r)dt^2 + h(r)dr^2 + r^2(d\theta^2 + \sin^2 d\phi^2),$$

where $f, h > 0$. Note that this system breaks down at the points where T and ∇r become collinear.

By the equations $Ric(g) = 0$, one then obtains that there exists a constant $M \in \mathbb{R}$ such that

$$h(r) = -\frac{1}{f(r)} = 1 - \frac{2M}{r}.$$

We next describe how to arrive to (7.4.1) without assuming staticity (completing the proof of Birkhoff's theorem). Consider a point $p \in S$ where S is a sphere of symmetry. There exists a 1-parameter subgroup I_p of $SO(3)$ which fixes p. Let $O(p)$ be the set of fixed points of I_p. Then $O(p)$ consists of two connected-components and is a orthogonal to S at q (and similarly it is orthogonal to any other sphere of symmetry at their common point). Hence $O(p)$ is a timelike two-dimensional plane. Let (θ, ϕ) be spherical coordinates on S. Extend (θ, ϕ) such that they are constant on each connected component of $O(p)$, for each $p \in S$. We can now introduce coordinates (t, r) coordinates on $O(p)$ and extend them globally such that t and r are constant on the spheres of symmetry. Then the metric with respect to (t, r, θ, ϕ) looks like (7.4.1), with the only difference being that $f = f(t, r)$ and $h = h(t, r)$. However, by solving the equations $Ric(g) = 0$ we again obtain the same solutions.

7.4.2 Schwarzschild Black Holes

The Schwarzschild metric in (t, r, θ, ϕ) is thus given by

$$g = -Ddt^2 + \frac{1}{D}dr^2 + r^2 g_{\mathbb{S}^2(\theta,\phi)},$$

where

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

and $M \in \mathbb{R}$ is a constant. We will be mainly interested in the case where $M > 0$. The metric is manifestly spherically symmetric and the vector field $T = \partial_t$ is Killing and timelike for $r > 2M$ and null at $r = 2M$, where however the metric component $g_{rr} = D^{-1}$ is singular (see below for more about this "singularity"). We note the topology of $\mathcal{M}$ is such that it is covered by the coordinate system (t, r, θ, ϕ) with respect to which the metric g makes sense.

Killing fields

Since $SO(3)$ acts by isometry, we have that there exists three linearly independent Killing fields $\Omega_1, \Omega_2, \Omega_3$ which arise from this action. As mentioned above, the vector field $T = \partial_t$ is also Killing. It turns out that all Killing fields are only those spanned by $T, \Omega_1, \Omega_2, \Omega_3$.

The geodesic flow

Let $\gamma(\tau) = (t(\tau), r(\tau), \theta(\tau), \phi(\tau))$ be a geodesic of $(\mathcal{M}, g)$. In view of the spherical symmetry we may assume that γ lies on the equatorial plane $\theta = \pi/2$. Suppose that γ is timelike or null. Without loss of generality we may assume that $\Omega_1 = \partial_\phi = \Phi$. Then we have the following integrals of motion:

1. $g(\dot{\gamma}, \dot{\gamma}) = -\kappa$, where $\kappa = 0$ or $\kappa = -1$.

2. $E = -g(T, \dot{\gamma})$, which we will refer to as the *energy* per unit mass of the particle.

3. $L = g(\Phi, \dot{\gamma})$, which we will refer to as the *angular momentum* per unit mass of the particle.

We can see that the Killing fields provide enough integrals of motion so the Hamilton–Jacobi equations for the geodesic flow are completely integrable. In particular, $r = r(\tau)$ satisfies the following (separated) ODE:

$$\left(\frac{dr}{d\tau}\right)^2 + V(r) = E^2,$$

where the potential V is given by

$$V(r) = \frac{1}{2}\kappa - \kappa\frac{M}{r} + \frac{L^2}{2r^2} - M\frac{L^2}{r^3}.$$

One can easily derive properties of causal geodesics depending on the values of E, L relative to M. In particular one can see that there exists stable circular (i.e of constant r) timelike geodesics for $r > 6M$, whereas the unstable circular timelike geodesics are restricted to $3M < r < 6M$. On the other hand, in the case of null geodesics, we only have unstable circular null geodesics at $r = 3M$.

The maximal extension

It turns out that the "singularity" at $r = 2M$ (for $M > 0$) is merely a coordinate singularity which can be overcome by a coordinate transformation. Consider the tortoise coordinate $r^* = r^*(r)$ such that

$$\frac{dr^*}{dr} = \frac{1}{D}.$$

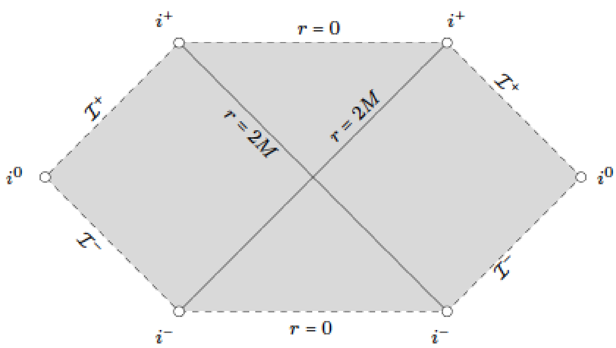
Then $r^* \to -\infty$ as $r \to 2M$ and $r^* \sim r$ for large r. If

$$v = t + r^*$$

then the metric in the coordinates[1] (v, r, θ, ϕ) takes the form

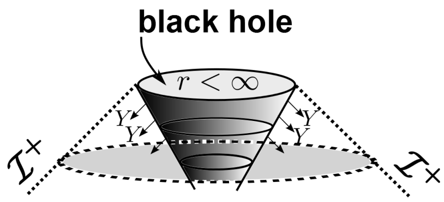
$$g = -Ddv^2 + 2dvdr + r^2 g_{\mathbb{S}^2(\theta,\phi)}, \tag{7.1}$$

where again $D = 1 - \frac{2M}{r}$. Clearly, this metric satisfies $Ric(g) = 0$ and can be extended beyond $r = 2M$ (although $r = 0$ is indeed a curvature singularity). In fact, the Kruskal maximal extension is depicted below:



[1]These coordinates are known as *ingoing Eddington–Finkelstein*.

We restrict to the domain covered by the (v, r, θ, ϕ) coordinate in the Schwarzschild spacetime. In this domain $T = \partial_v$ whereas $Y = \partial_r$ is tangential to incoming null geodesics (and is normalized so it differentiates with respect to r). In particular, Y is transversal to $\mathcal{H}^+$.



7.5 Kerr Black Holes

The Kerr metric with respect to the *Boyer-Lindquist coordinates* (t, r, θ, ϕ) is given by

$$g = g_{tt}dt^2 + g_{rr}dr^2 + g_{\phi\phi}d\phi^2 + g_{\theta\theta}d\theta^2 + 2g_{t\phi}dtd\phi,$$

where

$$g_{tt} = -\frac{\Delta - a^2\sin^2\theta}{\rho^2}, \quad g_{rr} = \frac{\rho^2}{\Delta}, \quad g_{t\phi} = -\frac{2Mar\sin^2\theta}{\rho^2},$$
$$g_{\phi\phi} = \frac{(r^2+a^2)^2 - a^2\Delta\sin^2\theta}{\rho^2}\sin^2\theta, \quad g_{\theta\theta} = \rho^2$$

with

$$\Delta = r^2 - 2Mr + a^2, \qquad \rho^2 = r^2 + a^2\cos^2\theta. \tag{7.2}$$

Schwarzschild corresponds to the case $a = 0$, subextreme Kerr to $|a| < M$ and extreme Kerr to $|a| = M$.

Note that the metric component g_{rr} is singular precisely at the points where $\Delta = 0$. To overcome this coordinate singularity we introduce the following functions $r^*(r), \phi^*(\phi, r)$ and $v(t, r^*)$ such that

$$r^* = \int \frac{r^2+a^2}{\Delta}, \quad \phi^* = \phi + \int \frac{a}{\Delta}, \quad v = t + r^*$$

In the *ingoing Eddington-Finkelstein coordinates* (v, r, θ, ϕ^*) the metric takes the form

$$g = g_{vv}dv^2 + g_{rr}dr^2 + g_{\phi^*\phi^*}(d\phi^*)^2 + g_{\theta\theta}d\theta^2 + 2g_{vr}dvdr + 2g_{v\phi^*}dvd\phi^* + 2g_{r\phi^*}drd\phi^*,$$

where

$$\begin{gathered} g_{vv} = -\left(1 - \frac{2Mr}{\rho^2}\right), \quad g_{rr} = 0, \quad g_{\phi^*\phi^*} = g_{\phi\phi}, \quad g_{\theta\theta} = \rho^2 \\ g_{vr} = 1, \quad g_{v\phi^*} = -\frac{2Mar\sin^2\theta}{\rho^2}, \quad g_{r\phi^*} = -a\sin^2\theta. \end{gathered} \tag{7.3}$$

For completeness, we include the computation for the inverse of the metric in (v, r, θ, ϕ^*) coordinates:

$$g^{vv} = \frac{a^2 \sin^2 \theta}{\rho^2}, \quad g^{rr} = \frac{\Delta}{\rho^2}, \quad g^{\phi^* \phi^*} = \frac{1}{\rho^2 \sin^2 \theta}, \quad g^{\theta\theta} = \frac{1}{\rho^2}$$

$$g^{vr} = \frac{r^2 + a^2}{\rho^2}, \quad g^{v\phi^*} = \frac{a}{\rho^2}, \quad g^{r\phi^*} = \frac{a}{\rho^2}.$$

Chapter 8

Lagrangian Theories and the Variational Principle

8.1 Matter Fields

Lagrangian mechanics, introduced by French mathematician Lagrange (1788), is a re-formulation of classical mechanics that combines conservation of momentum with conservation of energy. It was first introduced as an alternative way to calculate the trajectory of a particle in a dynamical system, but its applications are now far reaching.

The Lagrangian is a function associated to the system and contains all necessary information about its dynamics. Christodoulou defined the Lagrangian to be a n-form given by the composition of sections of appropriate vector bundles. In our case, the system is described by the maps $\phi : \mathcal{M} \to \mathbb{R}$ and its Lagrangian is a function

$$\mathcal{L} : \left(\mathbb{R}, T^*\mathcal{M}, g^{-1}\right) \to \mathbb{R} \tag{8.1}$$

where g^{-1} is the inverse of the pseudo-Riemannian metric on $\mathcal{M}$. In this way, the Lagrangian $\mathcal{L} = \mathcal{L}\left(\phi, d\phi, g^{-1}\right)$.

8.2 The Action Principle

According to the action principle the evolution of a dynamical system are is done in accordance with a quantity that takes an extremum value. For example, in Newtonian Mechanics, the evolution is such that $K - V$ takes the least value, where K is the **total** kinetic energy and V the **total** potential. Therefore, we only have to consider that $\mathcal{L} = K - V$ at each point and then take the integral of $\mathcal{L}$ over our manifold (in this case the real line-time) in order to obtain the total energies.

More generally, we need to integrate the Lagrangian over $\mathcal{M}$. For this reason, we use the natural volume form dg. Then, the number $S = \int_{\mathcal{M}} \mathcal{L} dg$ is called the action of the Lagragian.

Action Principle: "The actual evolution of the system (i.e. the actual ϕ) is such that S attains an extremum value."

Some remarks on this important principle:

- In case of Newtonian Mechanics, action principle is equivalent to the third law of Newton. However, Lagranges equations hold in any coordinate system, while Newtons are restricted to an inertial frame of reference.

- All the fundamental laws of physics can be written in terms of an action principle. This includes electromagnetism, general relativity, the standard model of particle physics, etc.

- This is an example of a variational principle and so the use of variational calculus is fundamental.

If a system admits a Lagrangian, then by definition the action principle holds. This principle will give us the laws of evolution and this is what we study next.

Euler Lagrange Equations

The Euler Lagrange equations are the equations that the action principle yields. Consider the coordinate system $\mathbf{x} = (x^1, ..., x^n)$. We want to extremize

$$S(\phi) = \int_{\mathcal{M}} \mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg = \int_{\mathcal{M}} \mathcal{L}\left(\phi\left(x^1, \ldots, x^n\right), \frac{\partial \phi}{\partial x^1}, \ldots, \frac{\partial \phi}{\partial x^n}, g^{-1}\right) dg.$$

The *first variational principle* gives us

$$0 = \left.\frac{d}{d\epsilon}\right|_{\epsilon=0} S(\phi + \epsilon h),$$

for any $h : \mathcal{M} \to \mathbb{R}$ that vanishes on the boundary $\partial\mathcal{M}$ so that the "endpoints" of the system are the same in the variation. Chain rule yields

$$0 = \int_{\mathcal{M}} \left(h\frac{\partial \mathcal{L}}{\partial \phi} + \sum_{k=1}^{n} \frac{\partial \mathcal{L}}{\partial \phi_k} \frac{\partial h}{\partial x^k} \right) dg \tag{8.2}$$

where $\phi_k = \frac{\partial \phi}{\partial x^k}$. If $\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]^k$ is the vector field with components $\frac{\partial \mathcal{L}}{\partial \phi_k}$ (w.r.t. the system $\mathbf{x}$) then applying divergence theorem for $\mathbf{X} = h\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]$ we take

$$\begin{aligned} \int_{\mathcal{M}} \operatorname{Div}\left(h\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]\right) dg &= \int_{\mathcal{M}} \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]^k \frac{\partial h}{\partial x^k} + h \operatorname{Div}\left(\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]\right) \right) dg \\ &= \int_{\partial\mathcal{M}} \left(\left(\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right] h \right) \cdot \mathbf{n} \right) dg_{n-1}. \end{aligned} \tag{8.3}$$

However, the boundary condition for h makes the boundary integral vanish. Therefore

$$\int_{\mathcal{M}} \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]^k \frac{\partial h}{\partial x^k} \right) dg = \int_{\mathcal{M}} \left(-h \operatorname{Div}\left(\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]\right) \right) dg$$

and so equation (8.2) can be written as

$$0 = \int_{\mathcal{M}} \left(h\frac{\partial \mathcal{L}}{\partial \phi} - h \operatorname{Div}\left(\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]\right) \right) dg. \tag{8.4}$$

Since (8.4) is true for any such h, the fundamental lemma of calculus of variation gives

$$\frac{\partial \mathcal{L}}{\partial \phi} - \operatorname{Div}\left(\left[\frac{\partial \mathcal{L}}{\partial d\phi}\right]\right) = 0. \tag{8.5}$$

This is exactly the Euler Lagrange equation. In general, if $\phi : \mathcal{M} \to \mathcal{N}$ then we have such an equation for every component of ϕ.

8.3 Derivation of the Energy Momentum Tensor

We now have all the necessary machinery to compute the energy momentum tensor of Lagrangian matter fields. We remind that we are seeking for a $(0,2)$ tensor field T that is symmetric and divergence-free. The first condition is rather physically reasonable condition and the second expresses the infinitesimal conservation of energy. Let us suppose that the Lagrangian is $\mathcal{L}\left(\phi, d\phi, g^{-1}\right)$ and the action of the Lagrangian is $S(\phi) = \int_{\mathcal{M}} \mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg$. We know that the Lagrangian contains all physical information concerning the system and the forces acting on it and the action of the Lagrangian gives us the laws of evolution of the system. Therefore, if we want to have a version of the conservation of energy then we should study more closely (and geometrically!) the action S.

Suppose we want to compute the integral $S(\phi)$. For such a computation we definitely need a coordinate system $\mathbf{x}$ on $\mathcal{M}$. On the other hand, every diffeomorphism $f : \mathcal{M} \to \mathcal{M}$ induces a pullback coordinate system $\mathbf{y} = f^*\mathbf{x}$. Therefore, our problem now is to construct diffeomorphisms of $\mathcal{M}$. But what are the most natural diffeomorphisms we can consider on $\mathcal{M}$? Let us look at vector fields of $\mathcal{M}$. Therefore, let $\mathbf{X}$ be an arbitrary vector field. Since we seek for an object on $\mathcal{M}$ with **local** properties we may assumme that $\mathbf{X}$ is compactly supported, i.e. $\mathbf{X} = 0$ outside a compact region $\mathcal{U} \subset \mathcal{M}$. Now let $\mathcal{F}_t$ be the associated flow of $\mathbf{X}$. Each such diffeomorphism $\mathcal{F}_t$ defines a pullback coordinate system $\mathbf{y}_t = \mathcal{F}_t^*\mathbf{x}$ whose change of coordinates is $\mathcal{F}_t$. Then

$$\int_{\mathcal{M}} \mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg = \int_{\mathcal{F}_t^{-1}(\mathcal{M})} \mathcal{F}_t^*\left(\mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg\right). \tag{8.6}$$

Clearly $\mathcal{F}_t^{-1}(\mathcal{M}) = \mathcal{M}$ and so

$$\int_{\mathcal{M}} \mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg = \int_{\mathcal{M}} \mathcal{F}_t^*\left(\mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg\right). \tag{8.7}$$

Equation (8.7) is the key identity! Indeed, (8.7) gives us

$$0 = \int_{\mathcal{M}} \left(\mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg - \mathcal{F}_t^*\left(\mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg\right)\right)$$

or

$$0 = \int_{\mathcal{M}} \frac{\mathcal{F}_t^*\left(\mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg\right) - \mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg}{t}$$

or

$$0 = \int_{\mathcal{M}} \lim_{t\to 0} \left(\frac{\mathcal{F}_t^*\left(\mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg\right) - \mathcal{L}\left(\phi, d\phi, g^{-1}\right) dg}{t}\right)$$

and by the definition of the Lie derivative we get

$$0 = \int_{\mathcal{M}} \mathcal{L}_{\mathbf{X}} \left(\mathcal{L} \left(\phi, d\phi, g^{-1} \right) dg \right). \tag{8.8}$$

We were led to equation (8.8) by only using the geometric structure of $\mathcal{M}$. Applying the chain and Leibnitz rule we obtain

$$0 = \int_{\mathcal{M}} \left(\left(\frac{\partial \mathcal{L}}{\partial \phi} \mathcal{L}_{\mathbf{X}} \phi + \frac{\partial \mathcal{L}}{\partial d\phi} \mathcal{L}_{\mathbf{X}} d\phi + \frac{\partial \mathcal{L}}{\partial g^{-1}} \mathcal{L}_{\mathbf{X}} g^{-1} \right) dg + \mathcal{L} \cdot \mathcal{L}_{\mathbf{X}} dg \right). \tag{8.9}$$

We now have the following:

1. Cartan's identity gives us

$$\mathcal{L}_{\mathbf{X}} d\phi = d(\mathbf{X}\phi) = d \left(\mathcal{L}_{\mathbf{X}} \phi \right). \tag{8.10}$$

2. Consider the vector field $\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right]^k = \left[\frac{\partial \mathcal{L}}{\partial \phi_k} \right]$, where $\phi_k = \frac{\partial \phi}{\partial x^k}$. Then

$$\operatorname{Div} \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \mathcal{L}_{\mathbf{X}} \phi \right) = d \left(\mathcal{L}_{\mathbf{X}} \phi \right) \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \right) + \left(\mathcal{L}_{\mathbf{X}} \phi \right) \cdot \operatorname{Div} \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \right). \tag{8.11}$$

3. Applying divergence theorem for the vector field $\left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \mathcal{L}_{\mathbf{X}} \phi \right)$ we get

$$\int_{\mathcal{M}} \operatorname{Div} \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \mathcal{L}_{\mathbf{X}} \phi \right) dg = \int_{\partial \mathcal{M}} \left(\left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \mathcal{L}_{\mathbf{X}} \phi \right) \cdot \mathbf{n} \right) dg_{n-1} = 0, \tag{8.12}$$

since, the boundary condition $\mathbf{X} = 0$ implies that $\mathcal{L}_{\mathbf{X}} \phi = \mathbf{X} \left(\phi \right) = 0$ on $\partial \mathcal{M}$.

4. Equations (8.10), (8.11) and (8.12) imply that

$$\begin{aligned} \int_{\mathcal{M}} \left(\frac{\partial \mathcal{L}}{\partial d\phi} \mathcal{L}_{\mathbf{X}} d\phi \right) dg &= \int_{\mathcal{M}} d \left(\mathcal{L}_{\mathbf{X}} \phi \right) \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \right) dg \\ &= \int_{\mathcal{M}} - \left(\mathcal{L}_{\mathbf{X}} \phi \right) \cdot \operatorname{Div} \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \right) dg. \end{aligned} \tag{8.13}$$

Therefore, using equation (8.13) we can write (8.9) as

$$0 = \int_{\mathcal{M}} \left(\left(\frac{\partial \mathcal{L}}{\partial \phi} \mathcal{L}_{\mathbf{X}} \phi - \left(\mathcal{L}_{\mathbf{X}} \phi \right) \cdot \operatorname{Div} \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \right) + \frac{\partial \mathcal{L}}{\partial g^{-1}} \mathcal{L}_{\mathbf{X}} g^{-1} \right) dg + \mathcal{L} \cdot \mathcal{L}_{\mathbf{X}} dg \right)$$

or

$$0 = \int_{\mathcal{M}} \left(\left(\left(\frac{\partial \mathcal{L}}{\partial \phi} - \operatorname{Div} \left(\left[\frac{\partial \mathcal{L}}{\partial d\phi} \right] \right) \right) \mathcal{L}_{\mathbf{X}} \phi + \frac{\partial \mathcal{L}}{\partial g^{-1}} \mathcal{L}_{\mathbf{X}} g^{-1} \right) dg + \mathcal{L} \cdot \mathcal{L}_{\mathbf{X}} dg \right).$$

However, according to Euler Lagrange equation (8.5) we obtain

$$0 = \int_{\mathcal{M}} \left(\left(\frac{\partial \mathcal{L}}{\partial g^{-1}} \mathcal{L}_{\mathbf{X}} g^{-1} \right) dg + \mathcal{L} \cdot \mathcal{L}_{\mathbf{X}} dg \right),$$

which can be written as

$$0 = \int_{\mathcal{M}} \left(\frac{\partial \mathcal{L}}{\partial g^{-1}} \mathcal{L}_{\mathbf{X}} g^{-1} + \mathcal{L} \cdot \left(\frac{1}{2} g_{ij} \left(\mathcal{L}_{\mathbf{X}} g \right)^{ij} \right) \right) dg$$

or

$$0 = \int_{\mathcal{M}} \left(\frac{\partial \mathcal{L}}{\partial g^{ij}} \left(\mathcal{L}_{\mathbf{X}} g^{-1} \right)^{ij} + \mathcal{L} \cdot \left(\frac{1}{2} g_{ij} \left(\mathcal{L}_{\mathbf{X}} g \right)^{ij} \right) \right) dg$$

or,

$$0 = \int_{\mathcal{M}} \left(\left(-\frac{\partial \mathcal{L}}{\partial g^{ij}} + \frac{1}{2} g_{ij} \mathcal{L} \right) \left(\mathcal{L}_{\mathbf{X}} g \right)^{ij} \right) dg. \tag{8.14}$$

The left term in the integrad of (8.14) is independent of the vector field $\mathbf{X}$ and so let's define the tensor Q by

$$Q_{ij} = \left(\frac{\partial \mathcal{L}}{\partial g^{ij}} - \frac{1}{2} g_{ij} \mathcal{L} \right). \tag{8.15}$$

Clearly Q is symmetric. Also, since the integral of the product of Q and a derivative is zero then by integrating by parts we expect the integral of the product of the divergence of Q and a vector field to be zero. Since the latter is going to be true for any vector field then we expect that the divergence of T vanishes. This is exactly what we were looking for and so it is the most natural candidate for the energy momentrum tensor. First, let us prove in detail the following

Theorem 8.3.1. *The tensor Q defined by* (9.1) *is divergence- free.*

Proof. We have

$$\left(\mathcal{L}_{\mathbf{X}} g \right)^{ij} = \mathbf{X}^{i;j} + \mathbf{X}^{j;i}$$

therefore equation (8.14) can be written as

$$0 = \int_{\mathcal{M}} \left(Q_{ij} \left(\mathbf{X}^{i;j} + \mathbf{X}^{j;i} \right) \right) dg$$

or

$$0 = \int_{\mathcal{M}} \left(\left(Q_{ij} \mathbf{X}^{i;j} \right) + \left(Q_{ij} \mathbf{X}^{j;i} \right) \right) dg$$

and since Q is symmetric we equivalently have

$$0 = \int_{\mathcal{M}} \left(\left(Q_{ij} \mathbf{X}^{i;j} \right) + \left(Q_{ji} \mathbf{X}^{j;i} \right) \right) dg$$

or

$$0 = \int_{\mathcal{M}} \left(Q_{ij} \mathbf{X}^{j;i} \right) dg. \tag{8.16}$$

We have

$$\mathrm{Div}\left(Q\mathbf{X} \right) = \left(\mathrm{Div}\left(Q \right) \right) \left(\mathbf{X} \right) + Q \left(\nabla \mathbf{X} \right)$$

and so

$$\int_{\mathcal{M}} \mathrm{Div}\left(Q\mathbf{X} \right) dg = \int_{\mathcal{M}} \left(\mathrm{Div}\left(Q \right) \right) \left(\mathbf{X} \right) dg + \int_{\mathcal{M}} \left(Q_{ji} \mathbf{X}^{j;i} \right) dg. \tag{8.17}$$

Applying divergence theorem for the vector field $\mathbf{Y} = Q\mathbf{X} = Q\left(\mathbf{X} \right)$ we get

$$\int_{\mathcal{M}} \mathrm{Div}\left(Q\mathbf{X} \right) dg = \int_{\partial \mathcal{M}} \left(\left(Q\mathbf{X} \right) \cdot \mathbf{n} dg_{n-1} \right) = 0 \tag{8.18}$$

since we have assumed that $\mathbf{X}$ is zero on the boundary $\partial\mathcal{M}$. Therefore, equations (8.16),(8.17) and (8.18) give

$$0 = \int_{\mathcal{M}} (\mathrm{Div}\,(Q))\,(\mathbf{X})\,dg, \tag{8.19}$$

which is true for any compactly supported X. This implies that $\mathrm{Div}\,(Q) = 0$. Indeed, if there is a point $p \in \mathcal{M}$ where $\mathrm{Div}\,(Q) \neq 0$ then we can find a vector X_p at p such that $\mathrm{Div}\,(Q)\,(X_p) > 0$. By continuity we can extend X_p to a compactly supported vector field whose support contains p and is sufficiently "small" and such that $\mathrm{Div}\,(Q)\,(X_p) > 0$. Then integrating this we get a strictly positive number with contradicts (8.19). □

We now have the following definition:

Definition 8.3.1. *The tensor Q given by* (9.1) *is called the energy momentum tensor of the Lagrangian matter field ϕ.*

8.4 Application to Linear Waves

One of the most important examples of matter fields are those given by the following Lagrangian

$$\mathcal{L}\left(\phi, d\phi, g^{-1}\right) = g^{-1}\left(d\phi, d\phi\right) = g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi (= g(\nabla\phi, \nabla\phi)). \tag{8.20}$$

Then the Euler Lagrange equations (8.5) imply that ϕ satisfies

$$\triangle_g \phi = 0, \tag{8.21}$$

where $\triangle_g$ is the Laplacian on $(\mathcal{M}, g)$. If g is positive definite, then ϕ is a harmonic function on the Riemannian manifold $(\mathcal{M}, g)$. If g is a Lorentzian metric then ϕ is a linear wave. In this case, there is a special symbol for the Laplacian on Lorentzian manifolds:

$$\Box_g \phi = 0. \tag{8.22}$$

Equation (9.1) gives us the energy momentum tensor of scalar linear waves on Lorentzian manifolds

$$Q_{\mu\nu} = \partial_\mu\phi\partial_\nu\phi - \frac{1}{2}g^{-1}\left(d\phi, d\phi\right)g_{\mu\nu}. \tag{8.23}$$

This tensor plays a fundamental role in the analysis of the evolution of waves.

8.5 Noether's Theorem

We saw how exploiting the geometry of the space and the Lagrangian structure of the system to derive an infinitesimal conservation of energy. If one wants to derive stronger versions of these laws, then stronger conditions have to be imposed on $\mathcal{M}$ too. Noether's theorem expresses in the most general way these conditions.

Let us see this more closely. The geometry of the space gives us equation (8.7) which is the key for the derivation of energy momentum tensor. It says that the "average" of those Lagrangians over the manifold $\mathcal{M}$ is the same. Obviously, a stronger condition that may be imposed is the following

$$\mathcal{L}\left(\phi, d\phi, g^{-1}\right) = \mathcal{F}_t^*\left(\mathcal{L}\left(\phi, d\phi, g^{-1}\right)\right). \tag{8.24}$$

This expresses a one-parameter symmetries of the Lagrangian. Then Noether's theorem gives a conservation law for the Euler Lagrange equation. For example, if a physical system behaves the same regardless of how it is oriented in space, its Lagrangian is rotationally symmetric; from this symmetry, Noether's theorem shows the angular momentum of the system must be conserved.

In the case of linear waves, Noether's theorem states that every Killing fields gives rise to a conservation law.

Chapter 9

Hyperbolic Equations

In this chapter we will study the properties of hyperbolic equations and emphasize the importance of global hyperbolicity in connection to hyperbolic equations.

We shall first consider the case of the scalar wave equation.

9.1 The Energy Method

Let (M, g) be a Lorentzian manifold and $\psi : \mathcal{M} \to \mathbb{R}$ a scalar function. The energy momentum tensor associated to ψ is defined to be the following symmetric $(0, 2)$ tensor field

$$Q[\psi] = d\psi \otimes d\psi - \frac{1}{2} g^{-1}(d\psi, d\psi) g. \tag{9.1}$$

The following proposition reveals the connection of Q and the wave equation.

Proposition 9.1.1. *For any function ψ we have the following identity*

$$\operatorname{Div} Q[\psi] = (\Box_g \psi) \cdot d\psi \tag{9.2}$$

Proof. Let us first compute its divergence for arbitrary ψ. Recall that the covariant derivative ∇ satisfies

$$\nabla(A \otimes B) = (\nabla A) \otimes B + A \otimes (\nabla B)$$

and that $\operatorname{Div} A = (\nabla^{\mu} A)_{\mu}$. Therefore,

$$\begin{aligned}
(\nabla^a T)_{\mu\nu} &= ((\nabla^a d\psi) \otimes d\psi)_{\mu\nu} + (d\psi \otimes (\nabla^a d\psi))_{\mu\nu} - \frac{1}{2} g^{-1}(d\psi, d\psi)(\nabla^a g)_{\mu\nu} \\
&\quad - \frac{1}{2} \left(\nabla^a (g^{-1}(d\psi, d\psi))\right) g_{\mu\nu} \\
&= (\nabla^a d\psi)_{\mu} \cdot (d\psi)_{\nu} + (d\psi)_{\mu} \cdot (\nabla^a d\psi)_{\nu} - g^{-1}(\nabla^a d\psi, d\psi) \cdot g_{\mu\nu},
\end{aligned}$$

where we used that $\nabla^a g^{-1} = 0$. Let now $a = \mu$. Observe also that

$$g^{-1}(\nabla^{\mu} d\psi, d\psi) \cdot g_{\mu\nu} = (\nabla^{\mu} d\psi)^b \cdot (d\psi)_b \cdot g_{\mu\nu} = (\nabla^b d\psi)^{\mu} \cdot (d\psi)_b \cdot g_{\mu\nu} = (\nabla^b d\psi)_{\nu} \cdot (d\psi)_b,$$

where we used that the Hessian is symmetric (recall the Einstein's notation for summation). Therefore,

$$\operatorname{Div} Q = (\operatorname{Div} d\psi) \cdot d\psi = (\Box_g \psi) \cdot d\psi.$$

□

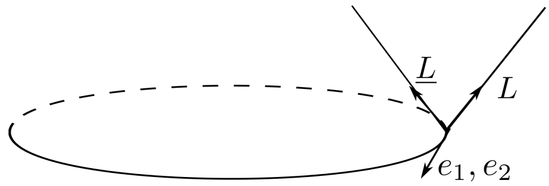
On top of equation (9.2), we also have the following

Proposition 9.1.2. *Let V_1, V_2 be two future directed timelike vector fields. Then the quantity $Q(V_1, V_2)$ is positive definite in $d\psi$, that is to say*

$$Q(V_1, V_2) \geq c \sum_{\alpha=0}^{3} (\partial_\alpha \psi)^2 .$$

By continuity, if one of V_1, V_2 is null then $Q(V_1, V_2)$ is non-negative definite in $d\psi$.

Proof. Let $x \in \mathcal{M}$ and $V_1, V_2 \in T_x\mathcal{M}$ be future-directed timelike vectors. Consider the plane $M = \langle V_1, V_2 \rangle$, which is manifestly timelike. The orthogonal complement $\Pi = \langle M \rangle^{\perp}$ is spacelike, and let e_1, e_2 be an orthonormal frame of it. Let also $L, \underline{L}$ be future-directed null vectors in M such that $g(L, \underline{L}) = -1$. Then $L, \underline{L}, e_1, e_2$ form a frame for $T_x\mathcal{M}$.



With respect to this frame, we have

$$|\nabla \psi|^2 = -2(L\psi)(\underline{L}\psi) + (e_1\psi)^2 + (e_2\psi)^2.$$

Therefore,

$$Q(L, L) = (L\psi)^2, \quad Q(\underline{L}, \underline{L}) = (\underline{L}\psi)^2, \quad Q(L, \underline{L}) = (e_1\psi)^2 + (e_2\psi)^2. \tag{9.3}$$

The result now follows from the fact that $V_i = a_i L + b_i \underline{L}$ for some $a_i, b_i > 0$, $i = 1, 2$.

□

See Appendix 8 for more about the tensor Q.

The Vector Field Method

For investigating the evolution of linear waves we will use the so-called vector field method. This is a geometric and robust method and involves mainly L^2 estimates. The main idea is to construct appropriate (0,1) currents and use the divergence identity in appropriate regions. We usually consider currents P_μ that depend on the geometry of $(\mathcal{M}, g)$ and are such that both P_μ and $\nabla^\mu P_\mu$ depend only on the first order derivatives of ψ. Such currents are in general called compatible currents. This can be achieved by using the wave equation to make all second order derivatives disappear and end up with something that highly depends on the geometry of the spacetime. There is a general method for producing such currents using the energy momentum tensor Q.

Energy Currents

Since Q is a $(0, 2)$ tensor we need to contract it with vector fields of $\mathcal{M}$. It is here where the geometry of $\mathcal{M}$ makes its appearance. We have the following definition

Definition 9.1.1. *Given a vector field V we define the J^V current by*

$$J^V_\mu[\psi] = Q_{\mu\nu}[\psi]V^\nu = Q[\psi](V, \cdot). \tag{9.4}$$

Note that we will usually drop ψ and thus simply write Q instead of $Q[\psi]$ etc. We say that we use the vector field V as a *multiplier*[1] if we apply the divergence identity for the current J^V_μ. The divergence of this current is

$$\mathrm{Div}(J) = \mathrm{Div}\,(QV) = \mathrm{Div}\,(Q)\,V + Q\,(\nabla V)\,, \tag{9.5}$$

where $(\nabla V)^{ij} = \left(g^{ki}\nabla_k V\right)^j = \left(\nabla^i V\right)^j$. If ψ is a solution of the wave equation then $\mathrm{Div}\,(Q) = 0$ and, therefore, $\nabla^\mu J^V_\mu$ is an expression of the 1-jet of ψ.

Definition 9.1.2. *Given a vector field V the scalar current K^V is defined by*

$$K^V[\psi] = Q[\psi]\,(\nabla V) = Q_{ij}[\psi]\left(\nabla^i V\right)^j = \frac{1}{2}Q_{\mu\nu}[\psi]\pi^{\mu\nu}_V, \tag{9.6}$$

where $\pi^{\mu\nu}_V = (\mathcal{L}_V g)^{\mu\nu} = (\nabla^\mu V)^\nu + (\nabla^\nu V)^\mu$ is the deformation tensor of V.

The first term of the right hand side of (9.5) does not vanish when we commute the wave equation with a vector field that is not Killing (or, more generally, when ψ does satisfy the wave equation). Therefore, we also have the following definition.

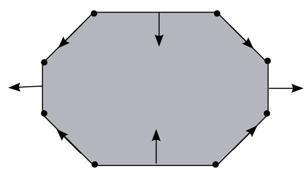
Definition 9.1.3. *Given a vector field V we define the scalar current $\mathcal{E}^V$ by*

$$\mathcal{E}^V[\psi] = (\mathrm{Div}\,Q)\,V = (\Box_g\psi)\,d\psi\,(V) = (\Box_g\psi)\,(V\psi)\,. \tag{9.7}$$

Note that Stokes' theorem gives us:

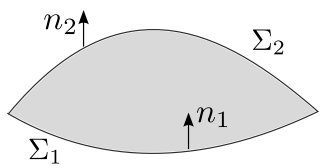
$$\int_{\mathcal{R}} (\Box_g\psi)\,(V\psi) + \int_{\mathcal{R}} \frac{1}{2}Q_{\mu\nu}[\psi]\pi^{\mu\nu}_V = \int_{\partial\mathcal{R}} Q(V, n_{\partial\mathcal{R}}),$$

where the direction of the normal to the boundary vector $n_{\partial\mathcal{R}}$ is determined by the following figure (thought of as embedded in $\mathbb{R}^{1+1}$)



9.2 A Priori Estimate

Consider two spacelike hypersurfaces Σ_0, Σ_τ which enclose the spacetime region $\mathcal{R}$ foliated by hypersurfaces Σ_t, where t is a time function.



[1]the name comes from the fact that the tensor Q is multiplied by V.

We suppose that ψ satisfy the inhomogeneous equation

$$\Box_g \psi = F.$$

Let also V be a timelike vector field and

$$f(t) = \int_{\Sigma_t} Q(V, n_{\Sigma_t}) \sim \int_{\Sigma_t} \sum_\alpha (\partial_\alpha \psi)^2.$$

Then,

$$f(\tau) + \int_{\mathcal{R}} F \cdot (V\psi) + \int_{\mathcal{R}} K^V[\psi] = f(0),$$

where $K^V[\psi]$ is given by (9.6). Since V and n_{Σ_t} are timelike we have

$$\int_{\Sigma_t} \left| K^V[\psi] \right| \leq C f(t).$$

Recall now the coarea formula

$$\int_0^\tau \left(\int_{\Sigma_t} f \right) dt = \int_{\mathcal{R}} f |\nabla t| \sim \int_{\mathcal{R}} f.$$

By the energy identity and coarea formula, there exists a uniform constant $C > 0$ such that for all solutions ψ to the inhomogeneous wave equation we have

$$f(\tau) \leq f(0) + \int_0^\tau F^2 dt + \int_0^\tau C f(t) dt. \tag{9.8}$$

The fundamental theorem of calculus gives us

$$\psi^2(\tau, \cdot) \leq \psi^2(0, \cdot) + \int_0^\tau C \Big(\psi^2(t, \cdot) + (\partial\psi)^2(t, \cdot) \Big) dt. \tag{9.9}$$

Recall also the Grönwall inequality: If f, A, B are non-negative functions and A is non-decreasing then

$$f(\tau) \leq A(\tau) + \int_0^\tau B(t) f(t) dt \quad \Longrightarrow \quad f(\tau) \leq e^{\int_0^\tau B(t)dt} A(\tau).$$

By adding (9.8) and (9.9) and using Grönwall inequality we then obtain the following estimate:

$$\|\psi\|^2_{H^1(\Sigma_\tau)} \leq e^{C\tau} \left(\|\psi\|^2_{H^1(\Sigma_0)} + \|F\|^2_{L^2(\mathcal{R})} \right). \tag{9.10}$$

Symmetries and Conservation Laws

If the spacetime admits a timelike Killing field V and apply it as a multiplier then, since its deformation tensor identically vanishes, if ψ satisfies the inhomogeneous equation, then the divergence of the current $J^V_\mu[\psi]$ is simply equal to $F \cdot (V\psi)$. Then,

$$\begin{aligned}
\sup_{t \in [0,\tau]} f(t) &\leq f(0) + \int_0^\tau \int_{\Sigma_t} |F \cdot V\psi| \leq f(0) + \int_0^\tau \|F\|_{L^2(\Sigma_t)} \cdot \|V\psi\|_{L^2(\Sigma_t)} \\
&\leq f(0) + \left(\sup_{t \in [0,\tau]} \|V\psi\|_{L^2(\Sigma_t)} \right) \cdot \int_0^\tau \|F\|_{L^2} (\Sigma_t) \\
&\leq f(0) + \frac{1}{2} \left(\sup_{t \in [0,\tau]} f(t) \right) + C \left(\int_0^\tau \|F\|_{L^2(\Sigma_t)} \right)^2
\end{aligned}$$

Hence,

$$\sup_{t\in[0,\tau]} f(t) \le f(0) + C\left(\int_0^\tau \|F\|_{L^2(\Sigma_t)}\right)^2. \tag{9.11}$$

Furthermore, we can easily obtain an improved estimate for ψ itself

$$\sup_{t\in[0,\tau]} \int_{\Sigma_t} \psi^2 \le \int_{\Sigma_0} \psi^2 + C\left(\int_0^\tau \left(f(t)\right)^{1/2}\right)^2. \tag{9.12}$$

Using (9.11) in (9.12) and added those two together we take

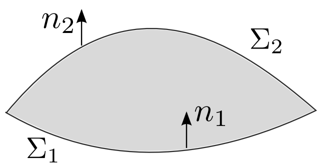
$$\sup_{t\in[0,\tau]} \|\psi\|^2_{H^1(\Sigma_t)} \le C\tau\left(\|\psi\|^2_{H^1(\Sigma_0)} + \left(\int_0^\tau \|F\|_{L^2(\Sigma_t)}\right)^2\right). \tag{9.13}$$

Note that (9.13) is stronger than (9.10).

Remark 9.2.1. *Estimate* (9.13) *holds for perturbations of Minkowski (even though they do not, in general, admit (timelike) Killing fields).*

If now V is Killing and ψ satisfies the homogeneous wave equation (i.e. $F = 0$) then

$$\int_{\Sigma_\tau} Q(V, n_{\Sigma_\tau}) = \int_{\Sigma_0} Q(V, n_{\Sigma_0}). \tag{9.14}$$



This statement corresponds to a conservation law. This is partly the content of a deep theorem of Noether[2]. This conservation law is particularly important if V is timelike since in this case the associated integrand quantities are positive definite in $d\psi$.

9.3 Well-posedness of the Wave Equation

Let Σ_0 be a spacelike hypersurface in a Lorentzian manifold $(\mathcal{M}, g)$. We next study the Cauchy problem to the wave equation

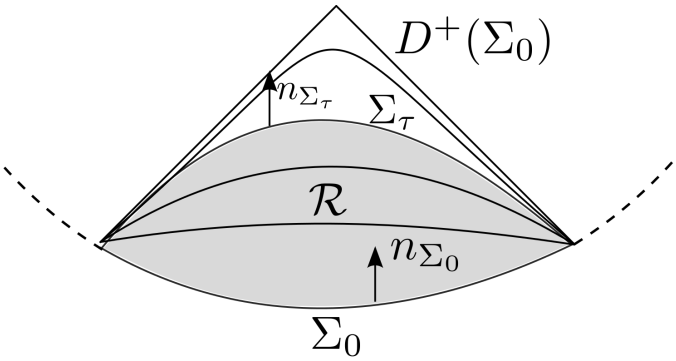
$$\Box_g \psi = F,$$
$$\psi|_{\Sigma_0} = f_0, \;\; n_{\Sigma_0}\psi|_{\Sigma_0} = f_1.$$

The regularity of F, f_0, f_1 is always such that the right hand sides of our estimates is finite.

Uniqueness: finite speed propagation

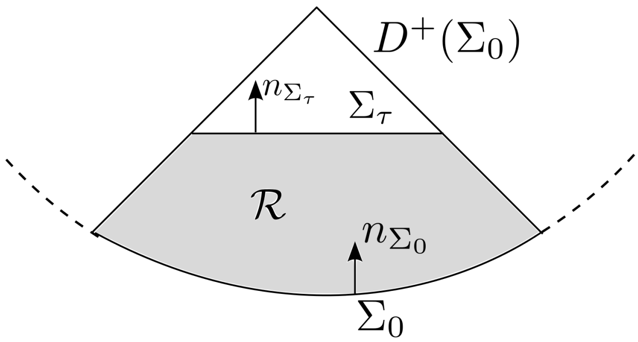
By virtue of the global hyperbolicity of $D^+(\Sigma_0)$ there exists a time function t and hence we can consider a spacelike foliation Σ_t of $D^+(\Sigma_0)$.

[2]According to this theorem, any continuous family of isometries gives rise to a conservation law.



Then if we suppose that ψ_1, ψ_2 satisfy the same inhomogeneous wave equation with same initial data then $\psi = \psi_1 - \psi_2$ satisfies the linear wave equation with trivial data. Then, (9.10) implies that $\|\psi\|_{H^1(\Sigma_t)} = 0$ for all t and hence $\psi = 0$ in $D^+(\Sigma_0)$ which shows the required uniqueness result.

In fact, we could have also considered the following foliation



and note that the null boundaries contribute non-negative definite terms in the energy identity. Estimate (9.10) again shows that if the data are trivial on Σ_0, then the solution is trivial in $D^+(\Sigma_0)$. In other words, non-trivial data on the dotted line outside the hypersurface Σ_0 take some time to affect the solutions to the wave equation (in particular, such data cannot affect ψ in $D^+(\Sigma_0)$). This phenomenon is a characteristic feature of hyperbolic equations and is called *finite speed propagation* (and also known as *weak Huygen's principle* and *domain of dependence property*).

Existence of solutions to the linear wave equation

We first prove the existence of solutions to the inhomogeneous wave equation with trivial initial data. Indeed, a solution to the general problem can be obtained as follows: Let $\tilde{\psi}$ be any function with the same initial data as the given one and let ψ be the solution $\Box\psi = F - \Box_g\tilde{\psi}$ with trivial data. Then, $\psi + \tilde{\psi}$ is the solution of the general problem.

Hence, we look for solutions to $\Box_g\psi = F$, with $f_0 = 0, f_1 = 0$.

Fix $\tau > 0$ and consider the spacetime region $\mathcal{R} = \cup_{t\in[0,\tau]}\Sigma_t$. We will show that a solution exists in $\mathcal{R}$, and since this will be true for any $\tau > 0$ the global existence in the interior of $D^+(\Sigma_0)$ follows.

Since our estimates hold for specific functions spaces, namely Sobolev spaces, it follows that the solutions will also have to exist in Sobolev spaces. For this reason, we first introduce the notion of a *weak solution.* Let

$$C_o(\mathcal{R}) = \left\{ \phi \in C^\infty(\mathcal{R}),\, \phi|_{\Sigma_\tau} = \partial\phi|_{\Sigma_\tau} = \phi|_{\Sigma_0} = \partial\phi|_{\Sigma_0} = 0 \right\}.$$

Then a weak solution to $\Box\psi = F$, with trivial initial data, is a distribution ψ such that

$$\langle \psi, \Box_g \phi \rangle = \int_{\mathcal{R}} F \cdot \phi \tag{9.15}$$

for all $\phi \in C_o(\mathcal{R})$. The motivation of this definition is of course the case where ψ is smooth and thus (9.15) holds in the classical sense, where $\langle \psi, \Box_g \phi \rangle = \int_{\mathcal{R}} \psi \cdot \Box_g \phi$. We will next show that if F is sufficiently regular, then ψ can in fact be represented by a function in an appropriate functional space such that (9.15) makes sense in the following sense:

$$\int_{\mathcal{R}} \psi \cdot \Box_g \phi = \int_{\mathcal{R}} F \cdot \phi,$$

where both integrals make sense provided $\phi \in C_o(\mathcal{R})$.

Remark 9.3.1. *We need to introduce the important notation for functional spaces we will be considering. If $u \in X([0,\tau]; Y(\Sigma_t))$, with $t \in [0,\tau]$, then $u|_{\Sigma_t}$ belongs in the functional space $Y(\Sigma_t)$ and the function $\tilde{u} : [0,\tau] \to Y$ lies in the space $X([0,\tau])$, where $\tilde{u}(t) = u_{\Sigma_t}$.*

For example, if $u \in L^1([0,\tau]; H^1)$, then $\|u\|_{L^1([0,\tau];H^1)} = \int_0^\tau \|u\|_{H^1(\Sigma_t)}\, dt$ is finite. Therefore, estimate (9.13) *can also be written as*

$$\|\psi\|_{L^\infty([0,\tau];H^1)} \leq C\tau \left(\|\psi\|_{H^1(\Sigma_0)} + \|F\|_{L^1([0,\tau];H^1)} \right).$$

Assume that $F \in L^2(\mathcal{R})$. Then

$$\begin{aligned} |\langle \psi, \Box_g \phi \rangle| = \left| \int_{\mathcal{R}} F \cdot \phi \right| &\leq \int_{\mathcal{R}} F^2 \int_{\mathcal{R}} \phi^2 \leq C \|F\|^2_{L^2(\mathcal{R})} \cdot \int_0^\tau \int_{\Sigma_t} \phi^2 \\ &\overset{(9.10)}{\leq} C\tau \|F\|^2_{L^2(\mathcal{R})} \int_{\mathcal{R}} (\Box_g \phi)^2 = C_{F,\tau} \int_{\mathcal{R}} (\Box_g \phi)^2, \end{aligned}$$

where we used the a priori estimate (9.10) for the smooth function ϕ (note also that by assumption the H^1 norm of ϕ on Σ_0 vanishes). It is important to emphasize that the a priori estimate allowed us to obtain the above estimate where the constant $C_{F,\tau}$, of course, does not depend on ϕ. Hence, if we consider the linear subspace

$$\mathcal{S}(\mathcal{R}) = \{ \Box_g \phi : \ \phi \in C_0(\mathcal{R}) \}$$

of the Banach space $L^2(\mathcal{R})$ then we have that the distribution ψ is in fact a bounded linear operator on $\mathcal{S}(\mathcal{R})$ (equipped with the $L^2(\mathcal{R})$-norm). Therefore, by Hahn-Banach we can extend ψ to bounded linear functional on $L^2(\mathcal{R})$. By Riesz representation theorem (or in other words, by duality) the $L^2(\mathcal{R})$ function ψ is in fact represented by a function $\psi \in L^2(\mathcal{R})$ so that

$$\langle \psi, \Box_g \phi \rangle = \int_{\mathcal{R}} \psi \cdot \Box_g \phi = \int_{\mathcal{R}} F \cdot \phi,$$

for all $\phi \in C_o(\mathcal{R})$.

Note that for the general problem with initial data, by the trace theorem, it suffices to take $f_0 \in H^{3/2}(\Sigma_0)$, $f_1 \in H^{1/2}(\Sigma_0)$. In fact, one can show that if $k \geq 0$ and

$$f_0 \in H^{k+1}(\Sigma_0), \quad f_1 \in H^k(\Sigma_0), \quad F \in H^k(\mathcal{R})$$

then there exists a unique solution

$$\psi \in H^{k+1}(\mathcal{R}) \text{ and } \psi \in H^{k+1}(\Sigma_\tau), \quad \psi \in H^k(\Sigma_\tau).$$

This can be shown by generalizing (9.10) for all $k \geq 0$ and using Cauchy-Kowalewsky theorem. Indeed, one can approximate (in the appropriate Sobolev norms) the differential operators and the initial data by analytic functions and solve in the analytic class the new equations. By the a priori estimates we have that the $H^{k+1}(\mathcal{R})$ norm of these analytic functions are uniformly bounded and hence the sequence converges weakly in $H^{k+1}(\mathcal{R})$. This limit is then the solution to the original problem.

Remark 9.3.2. *We can obtain slightly stronger results if $(\mathcal{M}, g)$ admits a timelike Killing field V and $F \in L^1([0,\tau]; L^2(\Sigma_t))$. Indeed, letting $\phi \in C_o(\mathcal{R})$ we obtain*

$$\begin{aligned}
|\langle \psi, \Box_g \phi \rangle| &= \left| \int_{\mathcal{R}} F \cdot \phi \right| \leq \int_0^\tau \int_{\Sigma_t} |F \cdot \phi| \leq \int_0^\tau \|F\|_{L^2(\Sigma_t)} \cdot \|\phi\|_{L^2(\Sigma_t)} \\
&\leq \sup_{t \in [0,\tau]} \|\phi\|_{L^2(\Sigma_t)} \cdot \int_0^\tau \|F\|_{L^2(\Sigma_t)} = C_{F,\tau} \sup_{t \in [0,\tau]} \|\phi\|_{L^2(\Sigma_t)} \leq C_{F,\tau} \int_0^T \|\Box_g \phi\|_{L^2(\Sigma_t)} \\
&= C_{F,\tau} \|\Box_g \phi\|_{L^1([0,\tau];L^2)} .
\end{aligned}$$

Clearly, ψ is a bounded functional on $C_o(\mathcal{R})$ with respect to the $L^1([0,\tau]; L^2)$-norm and hence by Hahn-Banach extends to a bounded functional of $L^1([0,\tau]; L^2)$. Hence, by duality ψ is in fact represented by a function $\psi \in L^\infty([0,\tau]; L^2)$ such that

$$\langle \psi, \Box_g \phi \rangle = \int_{\mathcal{R}} \psi \cdot \Box_g \phi = \int_{\mathcal{R}} F \cdot \phi.$$

In fact, we can improve the last inequality used above such that $\|\Box_g \psi\|_{L^2(\Sigma_t)}$ is replaced by $\|\Box_g \psi\|_{H^{-1}(\Sigma_t)}$. This then implies that $\psi \in L^\infty([0,\tau]; H^1)$. Assuming that F is smooth, one can show that

$$\psi \in C([0,\tau]; H^k) \cap C^1([0,\tau]; H^{k-1})$$

for all $k \in \mathbb{N}$. To remove the smoothness assumption on F, we choose a sequence of smooth functions F_m such that $F_m \to F$ in $L^1\big([0,\tau]; H^l\big)$ for a fixed l. For each of these F_m we obtain a solution ψ_m with trivial initial data, as above, which can be easily seen to form a Cauchy sequence in the Banach space $C([0,\tau]; H^l) \cap C^1([0,\tau]; H^{l-1})$. Hence the limit of this sequence solves the equation in $C([0,\tau]; H^l) \cap C^1([0,\tau]; H^{l-1})$.

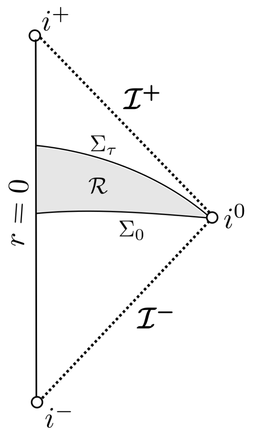
9.4 The Wave Equation on Minkowski spacetime

Pointwise and energy boundedness

Consider the Minkowski spacetime $\mathbb{R}^{3+1}$. Let ψ solve the homogeneous wave equation $\Box_g \psi = 0$ with initial data (f_0, f_1) prescribed on a Cauchy hypersurface Σ_0 (for example, we can take $\Sigma_0 = \{t = 0\}$). Clearly, the vector field

$$T = \partial_t$$

is globally timelike and Killing and hence its application as a multiplier gives us the following



By applying the divergence identity in the region $\mathcal{R} = \cup_{0 \le t \le \tau} \Sigma_t$ and using the fact that Σ_τ is spacelike for all τ (and hence the normal n_{Σ_τ} is timelike), we obtain a uniform constant $C > 0$ such that for all solutions ψ to the wave equation we have

$$\int_{\Sigma_\tau} \sum_a (\partial_\alpha \psi)^2 \le C \int_{\Sigma_0} \sum_a (\partial_\alpha \psi)^2. \tag{9.16}$$

Vector fields can be applied as multipliers but also as *commutators*. In other words, since ∂_t is Killing, we have that $\partial_t \psi$ also satisfies the wave equation. In fact, we can do the same for all translation vector fields $\partial_a, \alpha = 0, 1, 2, 3$. Repeating this, we can in fact bound the following higher order quantity:

$$\int_{\Sigma_\tau} \sum_{a,k} (\partial_\alpha^k \psi)^2 \le C \int_{\Sigma_0} \sum_{a,k} (\partial_\alpha^k \psi)^2, \tag{9.17}$$

with $k \ge 1$.

It remains to bound $\|\psi\|_{L^2(\Sigma_\tau)}$. Estimate (9.10) (or even (9.13)) is not at all satisfactory since it does not provide a uniform in τ bound. Consider spherical coordinates (t, r, θ, ϕ). Then for $\Sigma_\tau = \{t = \tau\}$ we obtain the 1-dimensional identity

$$\begin{aligned} \int_0^\infty \psi^2 dr &= [r\psi^2]_{r=0}^{r=\infty} - 2 \int_0^\infty r\psi(\partial_r \psi) dr \\ &\le \epsilon \int_0^\infty \psi^2 dr + \frac{1}{\epsilon} \int_0^\infty (\partial_r \psi)^2 r^2 dr, \end{aligned}$$

where we assumed that $r\psi^2 \to 0$ at infinity (recalling the finite speed propagation, this is true if, for example, the initial data are compactly supported). Hence, there exists a uniform constant $C > 0$ such that

$$\int_0^\infty \frac{1}{r^2} \psi^2 r^2 dr \le C \int_0^\infty (\partial_r \psi)^2 r^2 dr.$$

Integrating over $\mathbb{S}^2$ we obtain

$$\int_{\Sigma_\tau} \frac{1}{r^2} \psi^2 \le C \int_{\Sigma_\tau} (\partial_r \psi)^2 \le C \int_{\Sigma_0} \sum_a (\partial_\alpha \psi)^2. \tag{9.18}$$

Although we can only bound a weight L^2 norm of ψ, this bound suffices to apply Sobolev inequality and thus obtain uniform pointwise bounds.

Pointwise and energy decay

Let

$$f(t) = \int_{\Sigma_t} Q(V, n_{\Sigma_t}) \sim \int_{\Sigma_t} \sum_{\alpha} (\partial_\alpha \psi)^2.$$

In order to prove that $f(\tau)$ decays as $\tau \to +\infty$ one would ideally show that there exists a uniform constant $C > 0$ such that

$$\int_{\tau_1}^{\tau_2} f(t)dt \leq Cf(\tau_1).$$

In view of the boundedness of energy, this would then imply that $f(\tau)$ decays faster than any polynomial rate. It turns out, however, that such an estimate does not hold in Minkowski. Indeed, as we shall see, the spacetime integral of the energy can only be bounded by quantities with stronger weights (in other words, the left hand side above cannot just be $f(\tau_1)$).

For this reason, we divide the problem into two separate ones. The first one has to do with the region $r \leq R$, for some $R > 0$, and the second one with $r \leq R$.

Spacetime estimates in the spatially compact region are traditionally called Integrated Local Energy Decay (ILED). Such an estimate in Minkowski goes back to Morawetz 1968.

Since we want to obtain bounds for a spacetime integral, we want to apply vector field multipliers for which the current $K[\psi]$ is non-negative definite. So, clearly the Killing field T is not an option. If we use the vector field

$$X = f(r)\partial_r$$

(with respect to (t, r, θ, ϕ) coordinates) then

$$K^X[\psi] = \left(\frac{f'}{2} + \frac{f}{r}\right)(\partial_t\psi)^2 + \left(\frac{f'}{2} - \frac{f}{r}\right)(\partial_r\psi)^2 - \frac{f'}{2}\left|\nabla\!\!\!/\psi\right|^2,$$

where K^X is given by (9.6). Unfortunately, there is no f which makes all coefficients above positive. For this reason we need to modify the above current. For we consider the current

$$J_\mu^{X,h_1,h_2,w}[\psi] = J_\mu^X + h_1(r)\,\psi\nabla_\mu\psi + h_2(r)\,\psi^2(\nabla_\mu w),$$

where h_1, h_2, w are functions on $\mathcal{M}$. Then we have

$$\begin{aligned}\tilde{K}^X &= \nabla^\mu J_\mu^{X,h_1,h_2,w} = K^X + \nabla^\mu(h_1\psi\nabla_\mu\psi) + \nabla^\mu\left(h_2\psi^2(\nabla_\mu w)\right)\\ &= K^X + h_1(\nabla^a\psi\nabla_a\psi) + (\nabla^\mu h_1 + 2h_2\nabla^\mu w)\,\psi\nabla_\mu\psi + (\nabla_\mu h_2\nabla^\mu w + h_2(\Box_g w))\,\psi^2.\end{aligned}$$

By taking $h_2 = 1, h_1 = 2G, w = -G$ we make the coefficient of $\psi\nabla_\mu\psi$ vanish. Therefore, let us define

$$J_\mu^{X,1}[\psi] \doteq Q(X,\cdot) + 2G\psi(\nabla_\mu\psi) - (\nabla_\mu G)\,\psi^2 = J_\mu^X + 2G\psi(\nabla_\mu\psi) - (\nabla_\mu G)\,\psi^2. \tag{9.19}$$

Then

$$K^{X,1} \doteq \nabla^\mu J_\mu^{X,1} = K^X + 2G\left(-(\partial_t\psi)^2 + (\partial_r\psi)^2 + \left|\nabla\!\!\!/\psi\right|^2\right) - (\Box_g G)\,\psi^2.$$

Therefore, if we take G such that[3]

$$-2G = -\left(\frac{f'}{2} + \frac{f}{r}\right) \Rightarrow G = \frac{f'}{4} + \frac{f}{2r} \tag{9.20}$$

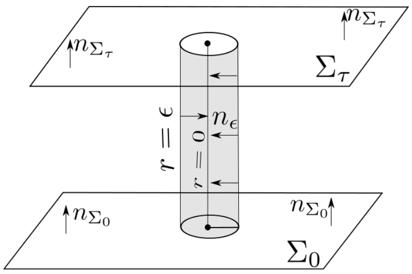
then

$$\begin{aligned} K^{X,1}[\psi] &= f'\,(\partial_r\psi)^2 + \frac{f}{r}\,|\nabla\!\!\!/\psi|^2 - (\Box_g G)\,\psi^2 \\ &= f'\,(\partial_r\psi)^2 + \frac{f}{r}\,|\nabla\!\!\!/\psi|^2 - \left(\frac{1}{4}f''' + \frac{1}{r}f''\right)\psi^2 \end{aligned} \tag{9.21}$$

If we take

$$\boxed{J^{X,1}_\mu\psi, \quad f=1, \quad G=\frac{1}{2r}}$$

then $K^{X,1}[\psi] = \frac{1}{r}|\nabla\!\!\!/\psi|^2$. We will apply the energy identity in the region $r \geq \epsilon$, and we will let $\epsilon \to 0$.



Since $r = \epsilon$ is timelike and its unit normal is $n_\epsilon = -\partial_r$ we obtain

$$\int_{\mathcal{R}\cap\{r\geq\epsilon\}} K^{X,1}[\psi] = \int_{\Sigma_0\cup\Sigma_\tau} J^{X,1}_\mu[\psi]n^\mu + \int_{r=\epsilon} J^{X,1}_\mu[\psi]n^\mu_\epsilon.$$

The boundary integral over $\Sigma_0 \cup \Sigma_\tau$ can be estimated using (9.17) and (9.18). The integral over $r = \epsilon$ is equal to

$$\int_{r=\epsilon} J^{X,1}_\mu[\psi]n^\mu_\epsilon = \int_0^\tau \int_{\mathbb{S}^2(\theta,\phi)} \Big(-J^X_r[\psi] - \frac{1}{\epsilon}\psi(\nabla_r\psi) - \frac{1}{\epsilon^2}\psi^2\Big)\epsilon^2 dg_{\mathbb{S}^2}dt$$

Therefore, as $\epsilon \to 0$, the only term that remains is negative, and hence has the correct sign. Therefore, we obtain,

$$\int_{\mathcal{R}} \frac{1}{r}|\nabla\!\!\!/\psi|^2 + \int_0^\tau \psi^2(r=0)dt \leq C\int_{\Sigma_0} J^T_\mu[\psi]n^\mu_{\Sigma_0}. \tag{9.22}$$

Morawetz's original argument was that, although the above estimate bounds the angular derivatives only, one can obtain exactly the same estimate by simply translating the origin of the spheres. In order to bound all spatial derivatives, it suffices to add three estimates for which the origins of the spheres are not colinear. This way we can retrieve the $\partial_r\psi$ derivative in (9.22). This argument however heavily uses the symmetries of Minkowski in

[3] We could have also chosen $G = \frac{f}{2r}$, however this choice is not very suitable for more general spacetimes.

a very non-robust way. It turns out that we can derive similar estimates for $\partial_r\psi$ using the energy method. Indeed, if we consider the current

$$\boxed{J_\mu^{X,1}[\psi], \quad f = -\frac{1}{r+1}}$$

then,

$$K^{X,1}[\psi] = \frac{1}{(r+1)^2}(\partial_r\psi)^2 + \gamma(r)\psi^2 - \frac{1}{r(r+1)}|\nabla\!\!\!\!/\,\psi|^2,$$

where $\gamma(r) > 0$ and $\gamma(r) \sim \frac{1}{r^4}$ for large r. The angular derivatives have already been bounded. Since $G \sim \frac{1}{r}$ for small r, the boundary integrals can be bounded as previously noting again the factor r^2 in the volume form of the integral over Σ_τ (which is needed as $r \to 0^+$). Hence,

$$\int_{\mathcal{R}} \frac{1}{(r+1)^2}(\partial_r\psi)^2 \leq C \int_{\Sigma_0} J_\mu^T[\psi] n_{\Sigma_0}^\mu. \tag{9.23}$$

Finally, in order to bound the derivative $\partial_t\psi$ we use the current

$$\boxed{J_\mu^X[\psi], \quad f = \frac{1}{r+1}}$$

Then,

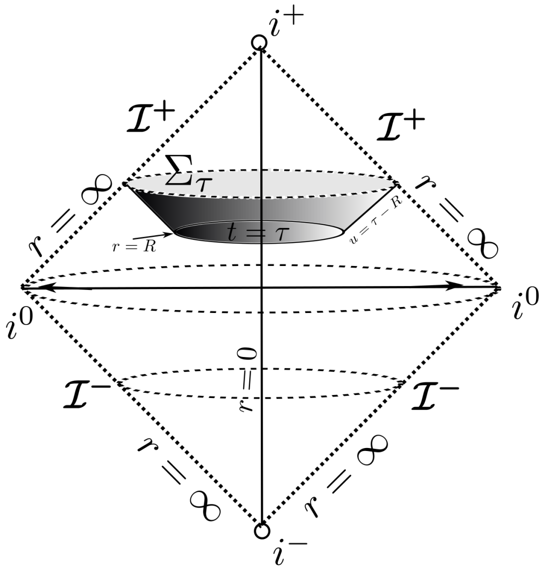
$$K^X[\psi] = \frac{(r+2)}{2r(r+1)^2}(\partial_t\psi)^2 + \alpha_1(r)(\partial_r\psi)^2 + \frac{1}{2(r+1)^2}|\nabla\!\!\!\!/\,\psi|^2$$

where $\alpha_1, \sim \frac{1}{r^2}$ for large r and $\alpha_1 \sim \frac{1}{r}$ as $r \to 0^+$. Hence, all the error terms and the the boundary integrals can be bounded using the previous estimates and noting the r^2 term in the volume form.

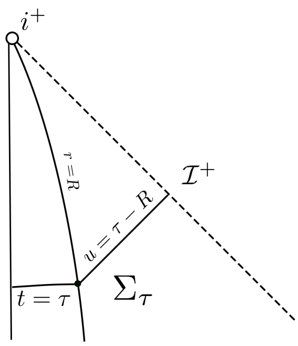
Therefore, given $R > 0$ there exists a constant C_R such that for all $r \leq R$ we have the following **integrated local energy decay estimate**

$$\int_{\tau_1}^{\tau_2} \int_{\Sigma_t \cap \{r \leq R\}} Q(T, n_{\Sigma_t}) \sim \int_{\mathcal{R} \cap \{r \leq R\}} \Big((\partial_t\psi)^2 + (\partial_r\psi)^2 + |\nabla\!\!\!\!/\,\psi|^2 \Big) \leq C_R \int_{\Sigma_{\tau_1}} Q(T, n_{\Sigma_{\tau_1}}). \tag{9.24}$$

We now consider the region $r \geq R$. Dafermos and Rodnianski have introduced a very elegant and robust method to derive non-degenerate estimates in this region. In particular, their method uses only r-weights and no t-weights. We assume that Σ_τ coincides with $t = \tau$ for $r \leq R$ and $u = \tau_R$ for $r \geq R$, as depicted below:

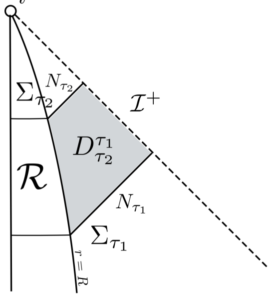


Suppressing all angular variables we obtain the following diagram



Consider now $\tilde{\Sigma}_\tau = \varphi_\tau(\tilde{\Sigma}_0)$, where φ_τ is the flow of T. For arbitrary $\tau_1 < \tau_2$ we define

$$
\begin{aligned}
&\tilde{\mathcal{R}}_{\tau_1}^{\tau_2} = \cup_{\tau \in [\tau_1, \tau_2]} \tilde{\Sigma}_\tau, \\
&\tilde{\mathcal{D}}_{\tau_1}^{\tau_2} = \tilde{\mathcal{R}}_{\tau_1}^{\tau_2} \cap \{r \geq R\}, \\
&\tilde{N}_\tau = \tilde{\Sigma}_\tau \cap \{r \geq R\}, \\
&\Delta_{\tau_1}^{\tau_2} = \tilde{\mathcal{R}}_{\tau_1}^{\tau_2} \cap \{r = R\}.
\end{aligned}
$$



Proposition 9.4.1. ***(Dafermos–Rodnianski hierarchy)*** *There exists a constant $C > 0$ such that for all solutions ψ to the wave equation we have*

$$\int_{\tau_1}^{\tau_2} \left(\int_{N_\tau} Q(T, n_{N_\tau}) \right) d\tau \, \leq C \int_{\Sigma_{\tau_1}} Q(T, n_{\Sigma_{\tau_1}}) + C \int_{N_{\tau_1}} r^{-1} (\partial_v \Psi)^2,$$

$$\int_{\tau_1}^{\tau_2} \left(\int_{N_\tau} r^{-1} (\partial_v \Psi)^2 \right) d\tau \, \leq C \int_{\Sigma_{\tau_1}} Q(T, n_{\Sigma_{\tau_1}}) + C \int_{N_{\tau_1}} (\partial_v \Psi)^2,$$

where $\Psi = r\psi$ and ∂_v is considered with respect to the null system (u, v).

This hierarchy is an application of the following more general r-weighted estimate of Dafermos and Rodnianski:

Proposition 9.4.2. *Suppose $p \in \mathbb{R}$, ψ is a solution to the wave equation and $\Psi = r\psi$. Then,*

$$\begin{aligned} &\int_{N_{\tau_2}} r^p \frac{(\partial_v \Psi)^2}{r^2} + \int_{\mathcal{D}_{\tau_1}^{\tau_2}} r^{p-1} p \frac{(\partial_v \Psi)^2}{r^2} + \int_{\mathcal{D}_{\tau_1}^{\tau_2}} \frac{r^{p-1}}{4} (2-p) \, |\nabla\!\!\!/ \psi|^2 \\ &\qquad \leq C \int_{\Sigma_{\tau_1}} Q(T, n_{\Sigma_{\tau_1}}) + \int_{N_{\tau_1}} r^p \frac{(\partial_v \Psi)^2}{r^2} \end{aligned} \tag{9.25}$$

Proof. We first consider the cut-off function $\zeta : [R, +\infty) \to [0, 1]$ such that

$$\begin{aligned} \zeta(r) = 0 \text{ for all } r \in [R, R + 1/2] \,, \\ \zeta(r) = 1 \text{ for all } r \in [R + 1, +\infty) \,. \end{aligned}$$

Let $q = p - 2$. We consider the vector field

$$V = r^q \partial_v$$

which we apply as multiplier acting on the function $\zeta\Psi$ in the region $\mathcal{D}_{\tau_1}^{\tau_2}$. Then

$$\int_{\mathcal{D}_{\tau_1}^{\tau_2}} K^V[\zeta\Psi] + \mathcal{E}^V[\zeta\Psi] = \int_{\partial\mathcal{D}_{\tau_1}^{\tau_2}} Q[\zeta\Phi](V, n^\mu).$$

Note that for $r \geq R + 1$ we have $K^V[\zeta\Psi] = K^V[\Psi]$ and $\mathcal{E}^V[\zeta\Psi] = \mathcal{E}^V[\Psi]$. Then,

$$\begin{aligned} K^V[\Psi] &= Q_{\mu\nu} \left(\nabla^\mu \left(r^q \partial_v \right) \right)^\nu = Q_{\mu\nu} \left(\left(\nabla^\mu r^q \right) \partial_v \right)^\nu + Q_{\mu\nu} r^q \left(\nabla^\mu \partial_v \right)^\nu \\ &= r^q K^V + Q_{\mu v} \left(\nabla^\mu r^q \right) \\ &= 2 r^{q-1} (\partial_u \Psi)(\partial_v \Psi) + q r^{q-1} (\partial_v \Psi)^2 - q \frac{r^{q-1}}{4} \, |\nabla\!\!\!/ \Psi|^2 \,. \end{aligned}$$

Note that since ψ solves the wave equation, Ψ satisfies

$$4 \partial_u \partial_v \Psi - \Delta\!\!\!/ \Psi = 0$$

and so

$$\begin{aligned} \Box_g \Psi &= -\frac{2}{r} (\partial_u \Psi - \partial_v \Psi) - 4 (\partial_u \partial_v \Psi) + \Delta\!\!\!/ \Psi \\ &= -\frac{2}{r} (\partial_u \Psi - \partial_v \Psi), \end{aligned}$$

which, as expected, depends only on the 1-jet of Ψ. Therefore,

$$\begin{aligned}\mathcal{E}^V(\Psi) &= r^q(\partial_v\Psi)(\Box_g\Psi)\\ &= -2r^{q-1}(\partial_u\Psi)(\partial_v\Psi) + 2r^{q-1}(\partial_v\Psi)^2.\end{aligned}$$

Thus

$$K^V[\Psi] + \mathcal{E}^V[\Psi] = (q+2)r^{q-1}(\partial_v\Psi)^2 - q\frac{r^{q-1}}{4}\left|\nabla\!\!\!/\Psi\right|^2.$$

In view of the cut-off function ζ all the integrals over Δ vanish. Clearly, all error terms that arise in the region[4] $\mathcal{W} = \overline{\text{supp}(\zeta - 1)} = \{R \leq r \leq R+1\}$ are quadratic forms of the 1-jet of ψ and, therefore, in view of the local integrated energy decay, these integrals are bounded by $\int_{\Sigma_{\tau_1}} Q(T, n_{\Sigma_{\tau_1}})$. The boundary integrals are:

$$\int_{\partial\mathcal{D}_{\tau_1}^{\tau_2}} Q[\zeta\Psi](V,n) = \int_{N_{\tau_1}} r^q\left(\partial_v\zeta\Psi\right)^2 - \int_{N_{\tau_2}} r^q\left(\partial_v\zeta\Psi\right)^2 - \int_{\mathcal{I}^+}\frac{1}{4}\left|\nabla\!\!\!/\Psi\right|^2.$$

The last two integrals on the right hand side appear with the right sign. The error terms produced by the cut-off ζ in the region $\mathcal{W}$ are controlled by the flux of T through Σ_{τ_1}.

$\square$

The second estimate of Proposition 9.4.1 follows trivially by taking $p = 2$. Regarding the first estimate, taking $p = 1$ implies

$$\int_{\mathcal{D}_{\tau_1}^{\tau_2}}\frac{1}{r^2}\left(\partial_v\Psi\right)^2 + \frac{1}{r^2}\left|\nabla\!\!\!/\Psi\right|^2 \leq C\int_{\Sigma_{\tau_1}} Q(T, n_{\Sigma_{\tau_1}}) + C\int_{N_{\tau_1}} r^{-1}\left(\partial_v\Psi\right)^2. \tag{9.26}$$

Note now that $\left|\nabla\!\!\!/\Psi\right|^2 = r^2\left|\nabla\!\!\!/\psi\right|^2$, (9.26) and

$$\int_{\mathcal{D}_{\tau_1}^{\tau_2}}\frac{1}{r^2}\left(\partial_v\zeta\Psi\right)^2 = \int_{\mathcal{D}_{\tau_1}^{\tau_2}}\frac{1}{r^2}\left(\partial_v\zeta\psi\right)^2 + \int_{\mathcal{D}_{\tau_1}^{\tau_2}}\frac{1}{2r^2}\partial_v(r\zeta\psi^2) = \int_{\mathcal{D}_{\tau_1}^{\tau_2}}\frac{1}{r^2}\left(\partial_v\zeta\psi\right)^2 + \int_{\mathcal{I}^+}\frac{1}{4r}(\zeta\psi)^2.$$

Again all the error terms can be bounded by the local integrated energy decay.

Remark 9.4.1. *The reason we introduced the function Ψ is because the weight r that it contains makes it non-degenerate ($\psi = 0$ on $\mathcal{I}^+$ but Ψ does not vanish there in general). The reason we have divided by r^2 in (9.25) is because we want to emphasize the weight that corresponds to ψ and not to Ψ.*

Therefore, if $\Psi = r\psi$ and

$$f(\tau) = \int_{\Sigma_\tau} Q(T, n_{\Sigma_\tau}), \qquad h_1(\tau) = \int_{N\tau}\frac{1}{r}(\partial_v\Psi)^2, \qquad h_2(\tau) = \int_{N\tau}(\partial_v\Psi)^2$$

and $\tau_1 < \tau_2$ then

$$\begin{gathered} f(\tau_1) \leq f(\tau_2),\\ \int_{\tau_1}^{\tau_2} f(t)dt \leq Cf(\tau_1) + Ch_1(\tau_1),\\ \int_{\tau_1}^{\tau_2} h_1(t)dt \leq Cf(\tau_1) + h_2(\tau_1).\end{gathered}$$

[4] The weights in r play no role in this region.

The first two inequalities imply that

$$f(\tau) \leq C\frac{D}{\tau},$$

where D depends on the initial data. The last inequality implies that there exists a dyadic sequence[5] τ_n such that

$$h_1(\tau_n) \leq C\frac{D}{\tau_n},$$

where D depends again on the initial data. The second estimate then gives us

$$(\tau_{n+1} - \tau_n)f(\tau_{n+1}) \leq \int_{\tau_n}^{\tau_{n+1}} f(t)dt \leq Cf(\tau_n) + Ch(\tau_n) \leq C\frac{D}{\tau},$$

which, by the properties of the dyadic sequence (and the first inequality), implies

$$f(\tau) \leq C\frac{D}{\tau^2}.$$

Pointwise bounds: First note that for $r_0 \geq R > 0$ we have

$$\psi^2(r_0, \omega) = \left(\int_{r_0}^{+\infty} (\partial_\rho \psi)\, d\rho\right)^2 \leq \left(\int_{r_0}^{+\infty} (\partial_\rho \psi)^2 \rho^2 d\rho\right)\left(\int_{r_0}^{+\infty} \frac{1}{\rho^2} d\rho\right) = \frac{1}{r_0}\left(\int_{r_0}^{+\infty} (\partial_\rho \psi)^2 \rho^2 d\rho\right),$$

where ∂_ρ is a derivative tangential to Σ_τ. Therefore,

$$\int_{\mathbb{S}^2} \psi^2(r_0, \omega)d\omega \leq \frac{1}{r_0}\int_{\mathbb{S}^2}\int_{r_0}^{+\infty} (\partial_\rho \psi)^2 \rho^2 d\rho d\omega \leq \frac{C}{r_0}\int_{\Sigma_\tau \cap \{r \geq r_0\}} Q(N, n_{\Sigma_\tau}), \tag{9.27}$$

For the interior region $r \leq R$ we can commute with T and apply elliptic estimates. Hence, commuting with either T and the angular operators and using Sobolev inequality gives us

$$|\psi| \leq C\frac{D}{r^{1/2} \cdot \tau}, \qquad |\psi| \leq C\frac{D}{r \cdot \tau^{1/2}}.$$

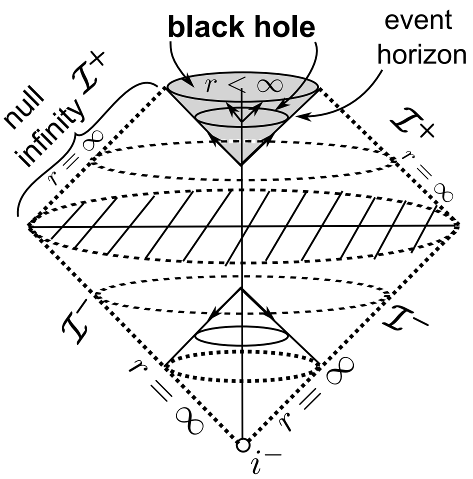
[5] A dyadic sequence τ_n has the property that $(\tau_{n+1} - \tau_n) \sim \tau_n \sim \tau_{n+1}$.

Chapter 10

Wave Propagation on Black Holes

10.1 Introduction

Recall that the conformal diagram of a black hole looks like the following



10.2 Pointwise and Energy Boundedness

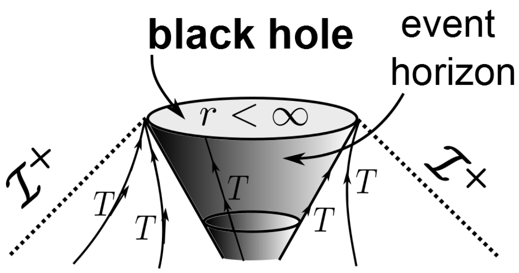
We will study the properties of the linear homogeneous wave equation on black hole backgrounds. Although we shall be particularly interested in the Schwarzschild case, our analysis will cover more general black holes for which the following hold:

1. The spacetime is asymptotically flat, that is to say the metric g approaches the Minkowski metric for large r, where r is an appropriate radius coordinate.

2. There exists a Killing field T which is timelike outside the event horizon $\mathcal{H}^+$ and null and normal on $\mathcal{H}^+$.

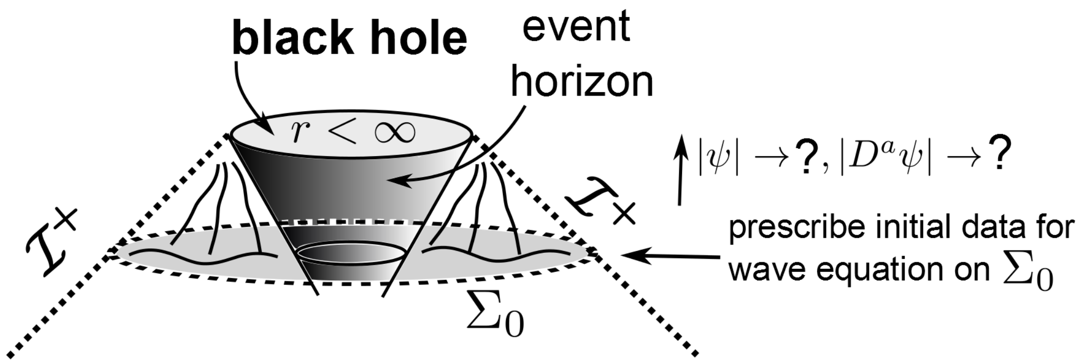
3. The coordinate r is such that it is constant along the integral curves of T and such that $\mathcal{H}^+ = \{r = r_{\mathcal{H}^+}\}$. The domain of outer communications then corresponds to $r > r_{\mathcal{H}^+}$.

4. The event horizon has positive surface gravity $\kappa > 0$. The surface gravity is defined by the following relation
$$\nabla_T T = \kappa T.$$



Let Σ_0 be a spacelike hypersurface which crosses the event horizon. We will study the following Cauchy problem

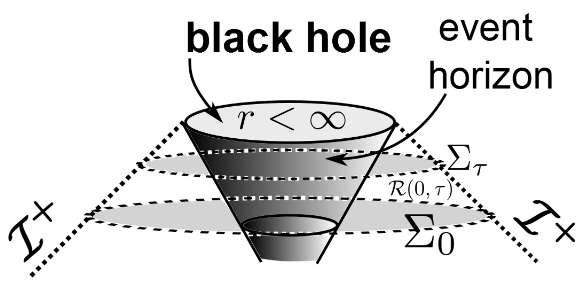
$$\Box_g \psi = 0,$$
$$\psi|_{\Sigma_0} = f_0, \quad T\psi|_{\Sigma_0} = f_1.$$



We assume that f_0, f_1 are sufficiently regular and we want to understand the evolution of ψ in $D^+(\Sigma_0)$ up to and including the horizon $\mathcal{H}^+$.

Basic a priori estimate

We define $\Sigma_\tau = \mathcal{F}_\tau(\Sigma_0)$, where $\mathcal{F}_\tau$ is the flow of T.



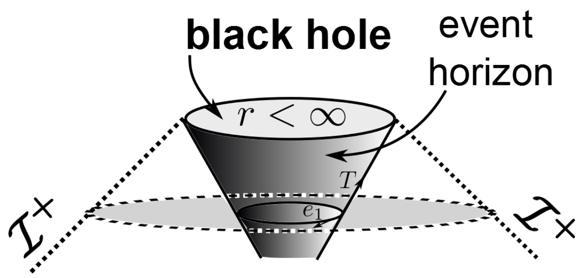
Applying the Killing field T as a multiplier in region $\mathcal{R}(0,\tau) = \cup_{0\leq t\leq\tau}\Sigma_t$ and noting that $Q(T, n_{\mathcal{H}^+}) \geq 0$, we obtain the following estimate

$$\int_{\Sigma_\tau} Q(T, n_{\Sigma_\tau}) \leq \int_{\Sigma_0} Q(T, n_{\Sigma_0}). \tag{10.1}$$

In view, however, of (9.3), we can see that

$$Q(T, n_{\Sigma_\tau}) \sim (T\psi)^2 + (e_1\psi)^2 + (e_2\psi)^2 \sim (T\psi)^2 + |\nabla\!\!\!/\psi|^2,$$

on the horizon $\mathcal{H}^+$ where T is null and tangential to it and n_{Σ_τ} timelike. Note that $\nabla\!\!\!/$ denotes the gradient on the sphere $S_{r_0} = \Sigma_\tau \cap \{r = r_0\}$.



This implies that (10.1) does not provide an estimate for all four derivatives of ψ up to and including the horizon, but in fact, the left hand side of (10.1) degenerates with respect to the transversal to $\mathcal{H}^+$ derivative (note that T, e_1, e_2 are all tangential to $\mathcal{H}^+$). For this reason we will refer to estimate (10.1) as the boundedness of the degenerate energy.

This however loss of derivatives is very bad since this degeneracy does not allow us to apply Sobolev inequalities and obtain pointwise estimates up to and including $\mathcal{H}^+$. One of the fundamental difficulties of the study of the wave equation on black holes is precisely to overcome this degeneracy.

The Schwarzschild case: Consider a spacelike hypersurface Σ_0 which crosses $\mathcal{H}^+$ and coincides with $t = 0$ for sufficiently large r. We define $\Sigma_\tau = \mathcal{F}_\tau(\Sigma_0)$, where $\mathcal{F}_\tau$ is the flow of T. Then,

$$Q(T, n_{\Sigma_\tau}) \sim (T\psi)^2 + \left(1 - \frac{2M}{r}\right)(Y\psi)^2 + |\nabla\!\!\!/\psi|^2.$$

On other hand, if N is a strictly timelike vector field then

$$Q(N, n_{\Sigma_\tau}) \sim (T\psi)^2 + (Y\psi)^2 + |\nabla\!\!\!/\psi|^2.$$

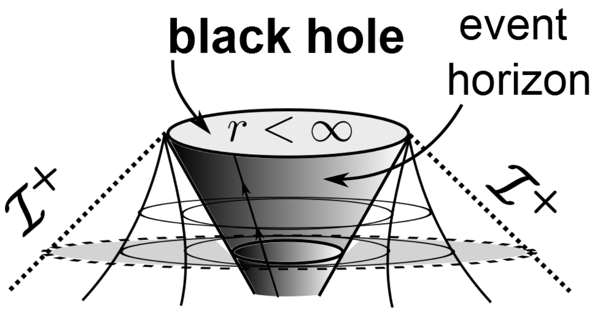
Boundedness of the non-degenerate energy

We will next present the argument of Dafermos and Rodnianski (2005) for the boundedness of the non-degenerate energy.

It is clear from the above discussion that in order to obtain non-degenerate bounds we need to apply a timelike vector field as multiplier; such a multiplier, however, will not have vanishing spacetime term $K^N[\psi]$. In this context, Dafermos and Rodnianski showed that under assumptions 1–4 above there exists $r_{\mathcal{H}^+} < r_0 < r_1$ and a timelike $\mathcal{F}_\tau$−invariant vector field N such that

1. $K^N[\psi] \sim Q(N, n_{\Sigma_\tau})$, for $r_{\mathcal{H}^+} \leq r \leq r_0$.
2. $\left|K^N[\psi]\right| \leq CQ(T, n_{\Sigma_\tau})$, for $r_0 \leq r \leq r_1$.
3. $N = T$ (and hence $K^N[\psi] = 0$), for $r \geq r_1$,

where K^N is given by (9.6).



Proof. For simplicity, we will prove this result only for metrics which take the form (7.1) for a general D. Consider the following ansatz

$$N = N^v(r)T + N^r(r)Y$$

Then

$$K^N[\psi] = F_{vv}\left(T\psi\right)^2 + F_{rr}\left(Y\psi\right)^2 + F_{\nabla\!\!\!/}\left|\nabla\!\!\!/\psi\right|^2 + F_{vr}\left(T\psi\right)\left(Y\psi\right),$$

where the coefficients are given by

$$F_{vv} = \left(\partial_r N^v\right), F_{rr} = D\left[\frac{\left(\partial_r N^r\right)}{2} - \frac{N^r}{r}\right] - \frac{N^r D'}{2}, F_{\nabla\!\!\!/} = -\frac{1}{2}\left(\partial_r N^r\right), F_{vr} = D\left(\partial_r N^v\right) - \frac{2N^r}{r},$$

where $D' = \frac{dD}{dr}$. Note that since $g\left(N,N\right) = -D\left(N^v\right)^2 + 2N^vN^r$, $g\left(N,T\right) = -DN^v + N^r$, and so $N^r\left(r = M\right)$ can not be zero (otherwise the vector field N would not be timelike).

The crucial observation is the following: The surface gravity is given by $\kappa = \frac{1}{2}D'$ and, therefore, by assumption we must have $D' > 0$.

We can thus take for $r = r_{\mathcal{H}^+}$ (which is a root of D) $N^v, \partial_r N^v$ to be very large, N^r negative and $\partial_r N^r$ negative. Then by Cauchy–Schwarz inequality we obtain the required positivity of $K^N[\psi]$ on $\mathcal{H}^+$ and thus, by stability considerations, on a neighborhood of $\mathcal{H}^+$ where $r_{\mathcal{H}^+} \leq r \leq r_0$. We can then smoothly extend N such that $N = T$ for $r \geq r_1 > r_0$. □

We now have all tools in place in order to prove the boundedness of the non-degenerate energy. Writing the energy identity for $J^N_\mu[\psi]$ in region $\mathcal{R}(\tau', \tau)$ we obtain

$$\int_{\Sigma_\tau} Q(N, n_{\Sigma_\tau}) + \int_{\mathcal{H}^+} Q(N, n_{\mathcal{H}^+}) + \int_{\mathcal{R}} K^N[\psi] = \int_{\Sigma_{\tau'}} Q(N, n_{\Sigma_{\tau'}}),$$

where $K^N[\psi] = \mathrm{Div}\left(Q(N, \cdot)\right)$. Since

$$\int_{\mathcal{R}} K^N[\psi] = \int_{r \leq r_0} K^N[\psi] + \int_{r_0 \leq r \leq r_1} K^N[\psi]$$

and since the flux on $\mathcal{H}^+$ is non-negative definite we obtain

$$\begin{aligned}\int_{\Sigma_\tau} Q(N, n_{\Sigma_\tau}) + \int_{r\leq r_0} K^N[\psi] &\leq \int_{r_0\leq r\leq r_1} -K^N[\psi] + \int_{\Sigma_{\tau'}} Q(N, n_{\Sigma_{\tau'}}) \\ &\leq C\int_{r_0\leq r\leq r_1} Q(T, n_\Sigma) + \int_{\Sigma_{\tau'}} Q(N, n_{\Sigma_{\tau'}}).\end{aligned}$$

We can now add a big multiple of $\int_{r\geq r_0} Q(T, n_\Sigma)$. If $f(\tau) = \int_{\Sigma_\tau} Q(N, n_{\Sigma_\tau})$, in view of the properties of N, the coarea formula and the boundedness of the degenerate energy we obtain

$$f(\tau) + b\int_{\tau'}^{\tau} f(t)dt \leq C(\tau - \tau')f(0) + f(\tau').$$

Rearranging terms, dividing by $(\tau - \tau')$ and taking the limit as $\tau' \to \tau$ tends

$$f'(\tau) + bf(\tau) \leq Cf(0).$$

This implies

$$\left(f(\tau)e^{b\tau} - \frac{C}{b}f(0)e^{b\tau}\right)' \leq 0$$

and so

$$f(\tau) \leq Bf(0)$$

for some uniform constant $B > 0$.

Remark 10.2.1. *The positivity of the surface gravity is intimately related to the so-called redshift effect that takes place along the horizon $\mathcal{H}^+$. For this reason, the vector field N is sometimes referred to as the redshift vector field.*

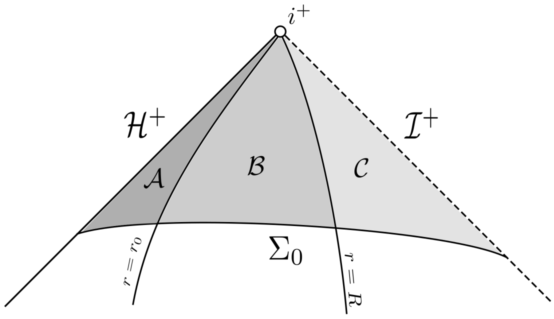
Remark 10.2.2. *As in the Minkowksi case, the vector field N can be applied as a multiplier as well as a commutator to give us higher order bounds.*

Remark 10.2.3. *For a special class of black holes, the surface gravity κ vanishes. Such black holes are called extremal. In these cases, the above construction for N does not work. Relevant results for such black holes have been obtained by Aretakis.*

Remark 10.2.4. *Although for the general subextremal family of axisymmetric Kerr black holes (see also Section 10.3) the surface gravity is positive and so the above construction for N applies, the Killing field T fails to be timelike (in fact causal) everywhere outside the event horizon $\mathcal{H}^+$. This means that in Kerr one does not have the trivial estimate* (10.1)*. Of course, this implies that the above proof for the boundedness of the non-degenerate energy fails. It turns out that for such backgrounds, in order to show any boundedness estimate one must in fact show something stronger, i.e. a decay estimate.*

10.3 Pointwise and Energy Decay

The discussion about energy decay on Minkowski in Section 9.4 suggests that one needs to prove a local integrated energy decay and also show an appropriate hierarchy of estimates in a neighborhood of infinity. It turns out, that the Dafermos–Rodnianski hierarchy still holds for Schwarzschild (and in fact more general) backgrounds. Hence, the problem is reduced in obtaining estimate in the spatially compact region $\{r_{\mathcal{H}^+} \leq r \leq R\}$, for (arbitrary but fixed) $R > r_{\mathcal{H}^+}$. In fact, in view of the redshift construction presented in Section 10.2, it suffices to show a spacetime estimate in region $\mathcal{B}$ depicted below



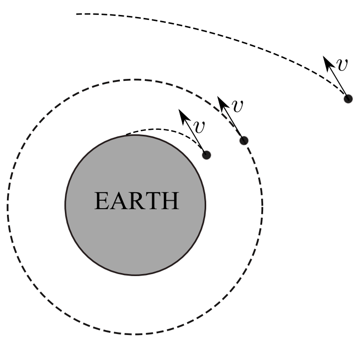
In order to obtain spacetime estimates in region $\mathcal{B}$ we need to know specific properties of the metric, and in particular, of the geodesic flow. Hence, the general assumptions 1–4 of Section 10.2 which lead to boundedness do not suffice to yield decay too (in fact, one does not expect to obtain decay using only assumptions). For this reason, we restrict our attention to Schwarzschild and Kerr (see below) spacetimes.

Schwarzschild

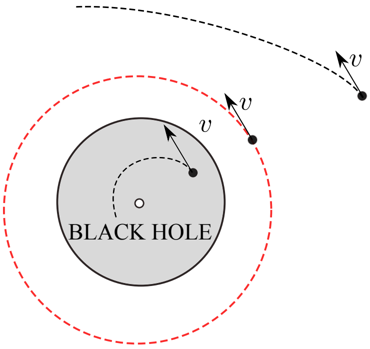
Let us first consider Schwarzschild backgrounds. As in Minkowski spacetime, one uses appropriately modified energy currents with multiplier ∂_r. In fact, it turns out that the multiplier

$$X = \partial_{r^*},$$

with respect to (t, r^*, θ, ϕ) is more suitable. The constructions of the multipliers (first presented by Dafermos and Rodnianski), although they are similar to the constructions for Minkowski, arise several technical difficulties. For example, one needs to decompose ψ in angular frequencies; that is to say, one needs to consider the Fourier series of ψ on the spheres of symmetry and derive uniform estimates for every projection separately. However, the most striking new feature of these constructions comes from the non-trivial dynamics of the geodesic flow. The most relevant property of the geodesic flow of Schwarzschild is that the three-dimensional timelike hypersurface $r = 3M$ is spanned by null geodesics, which orbit the black hole region (and hence are not orthogonal to the spheres of symmetry). Indeed, the (continuity) principle that applies to satellites orbiting the earth

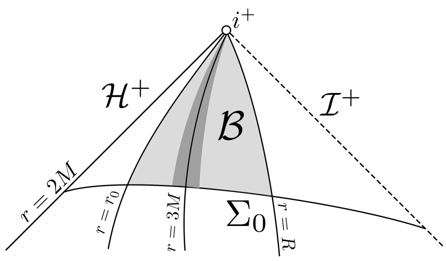


can be extended for Schwarzschild and thus obtain (trapped) null geodesics which neither scatter to null infinity (where $r = +\infty$) nor cross the event horizon $\mathcal{H}^+$.



The hypersurface $r = 3M$ is known as the *photon sphere*.

Returning now to the integrated local energy decay estimate, the presence of the photon sphere implies that the energy gets concentrated for long time around the photon sphere.



This implies that an estimate in region $\mathcal{B}$ must either degenerate on the photon sphere or lose derivatives. Specifically, one can only show the following

$$\int_{\mathcal{B}} (\partial_{r^*}\psi)^2 + \psi^2 + (r-3M)^2\Big((T\psi)^2 + |\nabla\!\!\!/\psi|^2\Big) \leq \int_{\Sigma_0} Q[\psi](T, n_{\Sigma_0}).$$

Note that it is in fact the tangential to the photon sphere derivatives that degenerate above. Hence, commuting with the Killing T (and thus replacing ψ above by $T\psi$) we obtain

$$\int_{\mathcal{B}} \sum_a (\partial_\alpha \psi)^2 + \psi^2 \leq C \int_{\Sigma_0} Q[\psi](T, n_{\Sigma_0}) + Q[T\psi](T, n_{\Sigma_0}).$$

This loss of derivatives is known as the *trapping effect*. One can now obtain the pointwise and energy decay results (which also lose one more derivative compared to Minkowski).

Remark 10.3.1. *The redshift effect, as already mentioned, degenerates for extremal black holes ($\kappa = 0$). For this reason, one needs to derive separate estimates in region $\mathcal{A}$. Work of Aretakis has shown that, in the extremal case, the trapping effect takes place along the horizon $\mathcal{H}^+$ too.*

Kerr

The Kerr metric admits less Killing fields and this makes the proof of estimates on such backgrounds more elaborate. Dafermos and Rodnianski (2010) have proved decay for

Kerr for all $|a| < M$. We next describe the main features that arise from passing from Schwarzschild ($M > 0, |a| = 0$) to Kerr ($M > 0, M > |a| \geq 0$) and also sketch the main points of the approach of Dafermos and Rodnianski.

Ergoregion and superradiance

If $|a| \neq 0$ then the Killing field $T = \partial_t = \partial_v$ is spacelike in a region $\mathcal{E}$ outside $\mathcal{H}^+$ known as *ergoregion*. In the context of $\Box_g \psi = 0$, this means that the flux $Q[\psi](T, , n_{\Sigma_\tau})$ fails to be non-negative definite. Hence, we can have

$$\int_{\mathcal{I}^+} Q(T, n_{\mathcal{I}^+}) > \int_{\Sigma_0} Q(T, n_{\Sigma_0}).$$

This phenomenon is called *superradiance*. Note that $\mathcal{H}^+ \subset \mathcal{E}$ and $\sup_{\mathcal{E}} r = 2M$. As we have already mentioned, the lack of the a priori degenerate (at $\mathcal{H}^+$) estimate for the T-flux implies that one needs to show decay directly. One can prove decay for Minkowski and Schwarzschild by constructing currents with positive definite divergence and using the conservation of the T-flux to bound the future boundary terms. Hence, in Kerr one needs to find an alternative approach in order to bound the boundary terms. This is in fact done using a bootstrap argument (see below).

Remark 10.3.2. *Superradiance is effectively absent if ψ is assumed to be axisymmetric. Indeed, if $\Phi\psi = 0$ then*

$$J^T_\mu[\psi] n^\mu_{\Sigma_\tau} \sim (T\psi)^2 + D(r)(Y\psi)^2 + |\nabla\!\!\!/ \psi|^2$$

for $r \leq R$ where the constants in $\sim$ depend on M and R and D degenerates only on $\mathcal{H}^+$.

Trapping

In Schwarzschild, all trapped null geodesics either live on $r = 3M$ or approach this hypersurface. In Kerr, on the other hand, there are trapped null geodesics for an open range of the radius coordinate r and therefore they can only be understood in phase space (and not just in physical space). This implies that the trapping effect, which must be taken into account for any dispersive estimate, is much more complicated than in the Schwarzschild case.

Remark 10.3.3. *If ψ is axisymmetric ($\Phi\psi = 0$), then it only "sees" the trapped null geodesics which are orthogonal to the vector field Φ. The behavior of the latter geodesics is much more favorable since they all approach a unique $r = r_{a,M}$ hypersurface.*

The bootstrap

The desired estimates close in a bootstrap setting which makes use of the smooth dependence of the Kerr metric on the angular momentum a. Specifically, given any $0 < a_0 < M$ and R sufficiently large, one considers the set $\mathcal{S} \subset [0, a_0]$ such that $|a| \in \mathcal{S}$ if there exists a constant $C = C(a)$ which depends on M, a such that for all solutions ψ to the wave equation we have

$$\int_{\Sigma_\tau} \sum_\alpha (\partial_\alpha \psi)^2 \leq C \int_{\Sigma_0} \sum_\alpha (\partial_\alpha \psi)^2,$$
$$\int_{\mathcal{R}\cap\{r\leq R\}} \sum_\alpha (\chi \partial_\alpha \psi)^2 + \psi^2 \leq C \int_{\Sigma_0} \sum_\alpha (\partial_\alpha \psi)^2.$$

for all $\tau \geq 0$, where $\chi = \chi(r) \geq 0$ and vanishes only in the physical space projection of the trapped set.

Since $[0, a_0]$ is a connected set, it suffices to show that $\mathcal{S}$ is non-empty, open and closed in $[0, a_0]$. The non-emptiness follows from the Schwarzschild result ($0 \in \mathcal{S}$).

In order to show that $\mathcal{S}$ is open, it suffices to show that if $a \in \mathcal{S}$ then $(a - \epsilon, a + \epsilon) \subset \mathcal{S}$ for some $\epsilon > 0$. Then, if $a \in \mathcal{S}$, the smooth dependence on the angular momentum of the metric allows us to use Cauchy stability considerations. We note that these considerations have to be localized away from the trapped set where the above estimate degenerates. In the openness part of the proof, we expect to allow $C(a + e) \to +\infty$ as $e \to \epsilon$.

In order to show that $\mathcal{S}$ is closed in $[0, a_0]$, it suffices to prove that if $(a - \epsilon, a + \epsilon) \subset \mathcal{S}$ then $a - \epsilon, a + \epsilon \in \mathcal{S}$. The way to show this is the following: Since $(a - \epsilon, a + \epsilon) \subset \mathcal{S}$ we have that $C(\tilde{a}) < \infty$ for $\tilde{a} \in (a - \epsilon, a + \epsilon)$. It, therefore, suffices to improve the boundedness of the constants $C(\tilde{a})$, that is to say, it suffices to prove that

$$C(\tilde{a}) < \infty \ \ \forall \tilde{a} \in (a - \epsilon, a + \epsilon) \ \Rightarrow \ C(\tilde{a}) < C(a_0) \ \ \forall \tilde{a} \in (a - \epsilon, a + \epsilon).$$

Appealing again to the smooth dependence of the metric on the angular momentum, we obtain that $C(a - \epsilon) < C(a_0)$ and $C(a + \epsilon) < C(a_0)$.

Frequency localization

As we mentioned above, one needs to understand the the properties of the geodesic flow, and in particular, of the trapped null geodesics. In Schwarzschild, on top of the stationary Killing field T, one has a complete set of spherical Killing fields traditionally denoted $\Omega_1, \Omega_2, \Omega_3$. These Killing fields provide enough conserved quantities to deduce that the Hamilton–Jacobi equations separate. In Kerr, although the only Killing vector fields are T and Φ, the Hamilton–Jacobi equations still separate in view of a third non-trivial conserved quantity discovered by Carter. Penrose and Walker showed that the complete integrability of the geodesic flow has its fundamental origin in the existence of an *irreducible Killing tensor*.

A Killing 2-tensor is a symmetric 2-tensor K which satisfies $\nabla_{(\alpha} K_{\beta\gamma)} = 0$. For example, the metric is always a Killing tensor. A Killing tensor will be called *irreducible* if it can not be constructed from the metric and other Killing vector fields.

A Killing tensor K yields a conserved quantity for geodesics. Indeed, if γ is a geodesic then $K_{\alpha\beta}\dot{\gamma}^\alpha \dot{\gamma}^\beta$ is a constant of the motion. It turns out that in Ricci flat spacetimes the separability of Hamilton–Jacobi, the separability of the wave equation and the existence of an irreducible Killing tensor are equivalent. Moverover, a Killing tensor gives rise to a symmetry operator $K = \nabla_\alpha \left(K^{\alpha\beta} \nabla_\beta\right)$ with the property $[K, \Box_g] = 0$. In the Kerr spacetime the symmetry operator K associated to Carter's irreducible Killing 2-tensor takes the form

$$K\psi = \Delta\!\!\!\!/_{\mathbb{S}^2} \psi - \Phi^2 \psi + \left(a^2 \sin^2 \theta\right) T^2 \psi. \tag{10.2}$$

One of the main insights of the approach of Dafermos and Rodnianski is the use of the separability of the wave equation (in (t,r,θ,ϕ) coordinates) as a way to frequency (micro-)localize the energy currents. This now allows us to construct multipliers for each frequency separately and thus capture trapping.

We have the following frequency decomposition of solutions ψ to the wave equation in Kerr:

$$\psi(t,r,\theta,\phi)=\frac{1}{\sqrt{2\pi}}\int_{\omega\in\mathbb{R}}\overbrace{\sum_{m,\ell}\underbrace{R_{\omega,m,\ell}(r)\cdot S^{(a\omega)}_{m\ell}(\theta,\phi)}_{\text{Oblate Spheroidal Expansion}}\cdot e^{-i\omega t}}^{\text{Fourier Expansion}}d\omega.$$

The functions $S^{(a\omega)}_{m\ell}(\theta,\phi)$, $m,\ell\in\mathbb{Z}, l\geq|m|$, are the eigenfunctions of the elliptic operator

$$P_{(a\omega)}=-\not\Delta_{\mathbb{S}^2}-(a\omega)^2\cos^2\theta.$$

If $\lambda^{(a\omega)}_{m\ell}$ are the corresponding eigenvalues, then we define the following angular frequency

$$\Lambda^{(a\omega)}_{m\ell}=\lambda^{(a\omega)}_{m\ell}+(a\omega)^2\geq 0.$$

If we also define $u^{(a\omega)}_{m\ell}(r)=\sqrt{(r^2+a^2)}\cdot R_{\omega,m,\ell}(r)$ then by suppressing the indices, the function $u=u(r^*)$ satisfies the Carter's equation

$$u''+(\omega^2-V(r))u=0,$$

where V is a potential function.

In view of Parseval's identity we have

$$\int_{-\infty}^{+\infty}\int_{\mathbb{S}^2}\psi^2r^2dg_{\mathbb{S}^2}dt\sim\int_{-\infty}^{+\infty}\sum_{m,l}\left|u^{(a\omega)}_{ml}(r)\right|^2d\omega,$$

$$\int_{r=c}(T\psi)^2\sim\int_{-\infty}^{+\infty}\sum_{m,l}\omega^2\left|u(r=c)\right|^2d\omega,$$

$$\int_{r=c}|\not\nabla\psi|^2\sim\int_{-\infty}^{+\infty}\sum_{m,l}\Lambda^{(a\omega)}_{m\ell}\left|u(r=c)\right|^2d\omega,$$

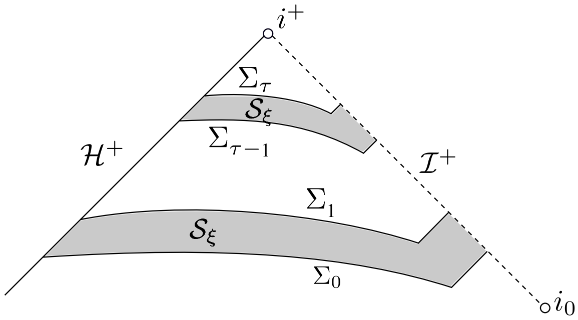
$$\int_{r=c}(\partial_{r^*}\psi)^2\sim\int_{-\infty}^{+\infty}\sum_{m,l}\left[\left|\frac{d}{dr^*}u(r=c)\right|^2+\left|u(r=c)\right|^2\right]d\omega.$$

Hence, in order to derive spacetime estimates, it suffices to construct microlocal currents $J^{(a\omega)}_{m\ell}[u]$ such that for $r_{\mathcal{H}^+}<r_0<R_0$ there exists a uniform in frequencies constant b such that

$$b\int_{r_0}^{R_0}\Big[|u|^2+|u'|^2+\left[\Lambda+\omega^2\right]|u|^2\Big]dr^*\leq J[u](r=R_0)-J[u](r=r_0)$$

for all frequencies (ω,Λ), where Λ is an angular frequency.

Major technical difficulty: This frequency decomposition of ψ requires taking the Fourier transform in time which in turn requires knowing some decay in time. Since this is not known a priori when proving the openness part of the bootstrap, we need to cut off in time as follows: Let ξ_τ be a cut-off function such that $\xi_\tau(\tilde{\tau})=0$ for $\tilde{\tau}\leq 0$ and $\tilde{\tau}\geq\tau$ and $\xi_\tau(\tilde{\tau})=1$ for $1\leq\tilde{\tau}\leq\tau-1$. Then the support of $\nabla\xi_\tau$ is the region $S_\xi=\{0\leq\tilde{\tau}\leq 1\}\cup\{\tau-1\leq\tilde{\tau}\leq\tau\}$:



Let $\psi_{✂} = \xi_\tau \psi$ be the compactly supported in time function which arises from the cut-off ξ_τ multiplied to the solution ψ of the wave equation. This cut-off version of ψ then satisfies the following inhomogeneous wave equation

$$\Box_g \psi_{✂} = F,$$

where

$$F = 2\nabla^\mu \xi_\tau \nabla_\mu \psi + (\Box \xi_\tau)\, \psi. \tag{10.3}$$

It turns out that one can still frequency localize $\psi_{✂}$ and thus try to apply the above framework. The only difference is that the cut-off creates error terms that need to be estimated. For closeness, since we have established estimates from the openness part of the proof, we can remove the future cut-off and thus try to prove quantitative estimates with precise control on the constants $C(\tilde{a})$. For details of the constructions see the work of Dafermos and Rodnianski.

Index