Conformal Scattering Theory

Lionel MASON[1] **& Jean-Philippe NICOLAS**[2]

[1]The Mathematical Institute, University of Oxford,
AWB, ROQ, Woodstock Rd, Oxford OX2 6GG, United Kingdom
lmason@maths.ox.ac.uk

[2]LMBA, UMR CNRS 6205, 29200 Brest, France,
Jean-Philippe.Nicolas@univ-brest.fr

Contents

Chapter 1

Introduction

Scattering theory can be schematically described as follows. We study an evolution equation

$$\mathcal{L}(\phi) = 0\,, \tag{1.1}$$

on $\mathbb{R}_t \times \Sigma$ where Σ is an n-dimensional manifold with or without boundary. Our goal is to identify, for each solution ϕ in a certain class to be defined, some functions $\phi_\pm$ that characterise the behaviour of ϕ in the distant future/past. Moreover, the correspondance between the solution and its future (resp. past) asymptotic behaviour ϕ_+ (resp. ϕ_-) should be one-to-one, i.e. there are mappings, the so-called wave operators,

$$W^\pm \,:\, \phi \mapsto \phi^\pm$$

that are invertible and allow to define the scattering map

$$S := (W^-)^{-1} W^+$$

that summarises the whole evolution of the field. The functions $\phi^\pm$ may serve as asymptotic data (referred to as scattering data) to reconstruct the field via the inverse wave operators. These scattering data may or may not be functions on Σ as we shall see in various examples in this book.

One of us (JPN) wishes to express his deepest thanks to Alain Bachelot for his marvellous lecture course on scattering theory that he was fortunate to attend in the early 1990's. They were unfortunately never published, probably because Alain felt that they were not bringing any new material compared to the book by Lax and Phillips. However his presentation differed sensibly from the book it was based on and was extremely detailed and pedagogical. Chapter 5 owes much to the lecture notes of this course.

Chapter 2

Some notions of special and general relativity

2.1 Minkowski space-time

Minkowski space $\mathbb{M}$ is $\mathbb{R}^4$ endowed with the Minkowski metric, whose expression in Cartesian coordinates is given by (the speed of light being taken equal to 1)

$$\eta = \mathrm{d}t^2 - \mathrm{d}x^2 - \mathrm{d}y^2 - \mathrm{d}z^2 \,. \tag{2.1}$$

Another useful expression of the metric η is in terms of spherical coordinates. It is particularly useful in order to perform an explicit conformal compactification. Is it a straightforward calculation to show that

$$\eta = \mathrm{d}t^2 - \mathrm{d}r^2 - r^2\mathrm{d}\omega^2 \,,\ \mathrm{d}\omega^2 = \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\varphi^2 \,, \tag{2.2}$$

where the spherical coordinates (r, θ, φ) are related to (x, y, z) by

$$x = r\sin\theta\cos\varphi \,,\ y = r\sin\theta\sin\varphi \,,\ z = r\sin\theta \,.$$

The metric $\mathrm{d}\omega^2$ defined in (2.2) is the euclidian metric on the 2-sphere.

The Minkowski metric acts on vectors at a point or vector fields on $\mathbb{M}$ as follows

$$\begin{gathered}
V = V^0\frac{\partial}{\partial t} + V^1\frac{\partial}{\partial x} + V^2\frac{\partial}{\partial y} + V^3\frac{\partial}{\partial z} \,,\ W = W^0\frac{\partial}{\partial t} + W^1\frac{\partial}{\partial x} + W^2\frac{\partial}{\partial y} + W^3\frac{\partial}{\partial z} \,, \\
\eta(V, W) = \eta_{ab}V^aW^b = V^0W^0 - V^1W^1 - V^2W^2 - V^3W^3 \,, \\
\eta(V, V) = (V^0)^2 - (V^1)^2 - (V^2)^2 - (V^3)^2 \,.
\end{gathered} \tag{2.3}$$

Remark 2.1. *Note that the tangent space to $\mathbb{M}$ at a given point p is $\mathbb{R}^4$ endowed with the Minkowski metric, but as a vector space. Minkowski space has the structure of an affine space. The tangent space at any given point will be referred to as Minkowski vector space. We shall see in Section 2.4 that it is the model for the tangent space to any space-time.*

We see that for each point $p \in \mathbb{M}$, (2.3) distinguishes three disjoint classes of tangent vectors.

Definition 2.1. *Let $p \in \mathbb{M}$, a vector $V \in T_p\mathbb{M}$ is said to be*

- *spacelike if $\eta(V,V) < 0$ (the projection of V on the space directions is longer than its time component),*
- *null if $\eta(V,V) = 0$ (the time and space parts of the vector are of equal length),*
- *timelike if $\eta(V,V) > 0$ (the time part of the vector is longer than its space part).*

Also V is said to be causal if it is either timelike or null, i.e. $\eta(V,V) \geq 0$.

Remark 2.2. *This gives us a local classification of curves (at least differentiable) as timelike, spacelike or null according to the classification of their tangent vector at a point.*

Remark 2.3. *Let us consider on $\mathbb{M}$ the trajectory of a particle whose "experience" of time is described by the variable t. This is a curve $\gamma(t) = (t, x(t), y(t), z(t))$. Its tangent vector is*

$$\dot{\gamma}(t) = \frac{\partial}{\partial t} + \dot{x}(t)\frac{\partial}{\partial x} + \dot{y}(t)\frac{\partial}{\partial y} + \dot{z}(t)\frac{\partial}{\partial z}$$

and

$$\eta(\dot{\gamma}(t), \dot{\gamma}(t)) = 1 - \dot{x}(t)^2 - \dot{y}(t)^2 - \dot{z}(t)^2 \,.$$

In the framework of classical mechanics, the vector

$$V(t) = \dot{x}(t)\frac{\partial}{\partial x} + \dot{y}(t)\frac{\partial}{\partial y} + \dot{z}(t)\frac{\partial}{\partial z}$$

is understood as describing the speed of the particle at time t. At a given time t, we know that the particle goes faster than, slower than, or at the speed of light, depending whether $|V(t)|^2 = \dot{x}(t)^2 + \dot{y}(t)^2 + \dot{z}(t)^2 > 1$, $|V(t)|^2 < 1$ or $|V(t)|^2 = 1$. However there is nothing unique about the choice of time parameter t, it is relative to the observer. A change of time parameter t will change the value of the time component of $\dot{\gamma}$ and the length of the space part of the tangent vector will then need to be compared to some quantity other than 1 (in fact the length of the time part) to compare the speed of the particle with that of light. As a matter of fact, even the notion of time and space part is not well defined, many other choices are possible corresponding to different choices of coordinates.

In relativity, the notion that replaces that of speed vector is that of 4-velocity vector, it is $\dot{\gamma}(t)$, the tangent vector to the trajectory of the particle. This is still a non unique notion since its "length" changes with a change of parameter of the curve. Its direction however is an intrinsic notion. And this gives us an intrinsic way of comparing the speed of a particle with that of light : a particle at a given point moves faster than, slower than, or at the speed of light depending whether the tangent vector field to its trajectory at that point (measured for any choice of parameter that is not singular at that point) is spacelike, timelike or null.

A massive particle will move along a timelike curve, a massless particle will move along a null curve.

Definition 2.2. *Given $p \in \mathbb{M}$, the set of null vectors in $T_p\mathbb{M}$ is the cone*

$$C_p = \left\{ V = V^0 \frac{\partial}{\partial t} + V^1 \frac{\partial}{\partial x} + V^2 \frac{\partial}{\partial y} + V^3 \frac{\partial}{\partial z}\,;\ (V^0)^2 = (V^1)^2 + (V^2)^2 + (V^3)^2 \right\}.$$

It is called the lightcone at p.

There are some useful orthogonality properties between vectors in the spacelike, timelike and lightlike cases. They are worth writing and proving in details since the orthogonality for an indefinite symmetric 2-form is less intuitive than for a positive definite one. First, let us introduce some notations that will be used extensively in the following proofs. Let $U \in T_p\mathbb{M}$, we denote

$$U = U^0 \partial_t + U'\,,$$

where U' is the projection of U on the spatial directions, i.e.

$$U' = U^1 \partial_x + U^2 \partial_y + U^3 \partial_z\,.$$

We shall also denote $|U'|$ the euclidian norm of U'

$$|U'|^2 = |U^1|^2 + |U^2|^2 + |U^3|^2\,.$$

Let $U, V \in T_p\mathbb{M}$, we denote by $\langle U', V' \rangle$ the euclidian inner product of U' and V' :

$$\langle U', V' \rangle = U^1V^1 + U^2V^2 + U^3V^3\,.$$

Proposition 2.1 (Orthogonal to a timelike vector)**.** *Let T be a timelike vector at a point p and $V \in T_p\mathbb{M}$ such that $\eta(T, V) = 0$, then V is spacelike or zero.*

Proof. We assume that $V \neq 0$. We know that T is timelike, i.e.

$$|T^0| > |T'|\,.$$

Moreover,

$$\eta(T, V) = T^0V^0 - \langle T', V' \rangle = 0\,.$$

This implies in particular that $V' \neq (0, 0, 0)$, otherwise the equality above would imply also that $V_0 = 0$ and this would contradict $V \neq 0$. In addition, it follows that

$$|V^0| = \frac{\langle T', V' \rangle}{|T^0|} \leq \frac{|T'||V'|}{|T^0|} < |V'|\,.$$

This concludes the proof. □

Remark 2.4. *This means that the orthogonal in $T_p\mathbb{M}$ to a timelike vector at p for the metric η is a hyperplane in $T_p\mathbb{M}$ containing only spacelike vectors.*

A vector orthogonal to a spacelike vector is not necessarily timelike, a simple example is given by the vectors ∂_x and ∂_y, but if we restrict ourselves to a plane spanned by a timelike and a spacelike vector, then the result becomes true.

Proposition 2.2. *Consider at a point p in $\mathbb{M}$ a spacelike vector V and a timelike vector T. Let W be a vector in the plane spanned by T and V and that is orthogonal to V, i.e. $\eta(W,V)=0$, then W is timelike or zero.*

Proof. The restriction of η to the plane spanned by T and V is a quadratic form whose matrix in the basis $\{T,V\}$

$$A := \begin{pmatrix} \eta(T,T) & \eta(T,V) \\ \eta(T,V) & \eta(V,V) \end{pmatrix}$$

is real symmetric and has negative determinant

$$\det A = \eta(T,T)\eta(V,V) - \eta(T,V)^2 \,.$$

Hence A has one positive and one negative eigenvalue. In the basis $\{V\,,\ W\}$ (assuming of course $W \neq 0$), the matrix of the quadratic form is diagonal since $\eta(V,W)=0$. Since $\eta(V,V)<0$ and the determinant of the matrix must still be strictly negative, it follows that $\eta(W,W)>0$, i.e. W is timelike. □

When looking at the space of vectors orthogonal to a null vector field, the situation gets more unusual.

Proposition 2.3. *Let V be a non-zero null vector at a point p in $\mathbb{M}$. The subspace of $T_p\mathbb{M}$ of vectors orthogonal to V contains V ; except for the straight line generated by V, it is entirely composed of spacelike vectors ; it is the hyperplane tangent to the light-cone containing V.*

Proof. The fact that V is orthogonal to itself is trivial since V is assumed to be null. The vector V can be decomposed as follows

$$V = V^0\partial_t + V' \,.$$

We can find two linearly independent vectors U and W in the hyperplane spanned by ∂_x, ∂_y, ∂_z which are orthogonal to V' for the euclidian inner product on $\mathbb{R}^3$. Then U, V, W are three linearly independent vectors orthogonal to V and which consequently span the hyperplane orthogonal to V. Moreover they are mutually orthogonal and since V is null and U and W are spacelike, it follows that any linear combination of the three is spacelike unless it is parallel to V. □

Definition 2.3. *Let S be a $\mathcal{C}^1$ hypersurface in $\mathbb{M}$. We say that S is :*

- *spacelike if its normal vector at each point is a timelike vector, this means that its tangent plane at each point is entirely composed of spacelike vectors ;*
- *null if its normal vector at each point is a null vector, this means that its tangent plane at each point is composed of spacelike vectors and one null direction given by the normal vector ;*
- *achronal or weakly spacelike if its normal vector at each point is a causal vector ;*
- *timelike if its normal vector at each point is a spacelike vector, this means that its tangent plane at each point is generated by one timelike and two spacelike vectors ;*

2.2 Manifolds, tensors

We start with a short description of manifolds and their tangent structures; for more details, we refer the reader to the excellent book by do Carmo [15].

Definition 2.4 (Differentiable or $\mathscr{C}^k$ manifold)**.** *A real n-dimensional differentiable (resp. $\mathscr{C}^k$) manifold is a set $\mathcal{M}$ equipped with a differentiable (resp. $\mathscr{C}^k$) atlas, i.e. a family $\{\mathcal{U}_\alpha\}$ of open sets in $\mathbb{R}^n$ and a family of injective maps $\phi_\alpha : \mathcal{U}_\alpha \to \mathcal{M}$ such that*

1. *the family $\{\mathcal{V}_\alpha = \phi_\alpha(\mathcal{U}_\alpha)\}$ is a covering of $\mathcal{M}$, and*
2. *for any α and β such that $W = \phi_\alpha(\mathcal{U}_\alpha) \cap \phi_\beta(\mathcal{U}_\beta) \neq \emptyset$, $\phi_\alpha^{-1}(W)$ and $\phi_\beta^{-1}(W)$ are open sets of $\mathbb{R}^n$ and the map $\phi_\beta^{-1} \circ \phi_\alpha : \phi_\alpha^{-1}(W) \to \phi_\beta^{-1}(W)$ is differentiable (resp. $\mathscr{C}^k$).*

The pairs $(\mathcal{U}_\alpha, \phi_\alpha)$ are called local parametrisations of $\mathcal{M}$ or local charts.

From an atlas, we can easily construct a maximal atlas by adding all possible parametrisations that are compatible with the ones we already have, in the sense that they satisfy the second property above. A maximal atlas is called a differentiable structure.

A $\mathscr{C}^\infty$ manifold will also be called a smooth manifold.

Examples. For all $n \in \mathbb{N}^*$, $\mathbb{R}^n$ is an n-dimensional smooth manifold, with the 1-chart atlas $\{(\mathbb{R}^n, I)\}$. Parametrised surfaces in $\mathbb{R}^3$ are 2-dimensional manifolds and they are submanifolds of $\mathbb{R}^3$. For instance, the unit 2-sphere S^2 can be realised as a 2-dimensional manifold with an atlas with two charts

$$\mathcal{U}_1 =]0,\pi[\times]0,2\pi[\,,\ \phi_1(\theta,\varphi) = (\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)\,,$$
$$\mathcal{U}_2 =]0,\pi[\times]0,2\pi[\,,\ \phi_2(\theta,\varphi) = (-\cos\theta, -\sin\theta\cos\varphi, \sin\theta\sin\varphi).$$

Remark 2.5. *A smooth differentiable structure induces a topology on $\mathcal{M}$ and allows to define the notion of a differentiable function from $\mathcal{M}$ into another differentiable manifold, by requiring that the function conjugated by local parametrisations*

$$\psi_\beta^{-1} \circ f \circ \phi_\alpha \,:\, \mathcal{U}_\alpha \to \mathcal{V}_\beta$$

be differentiable. This induces the definition of the differential of a function. If we have a smooth differentiable structure, we have access to the notions of smooth functions and distributions (see Chapter A).

It turns out that any differentiable n-dimensional manifold $\mathcal{M}$ is in fact a submanifold of $\mathbb{R}^d$ for d large enough depending on $\mathcal{M}$ and not just on n, this is a theorem due to Whitney in 1936 [60]. A submanifold of $\mathbb{R}^d$ can be defined in a slightly different manner as follows, taking the ambient space into account.

Definition 2.5. *A $\mathscr{C}^k$ submanifold of $\mathbb{R}^d$ of dimension $n \in \{1, ..., d-1\}$ is a subset S of $\mathbb{R}^d$ such that, for any point $p_0 \in S$, there exists V a neighbourhood of p_0 in $\mathbb{R}^d$, U a neighbourhood of 0 in $\mathbb{R}^d$ and $\phi : U \to V$ a $\mathscr{C}^k$ diffeomorphism such that $\phi(0) = p_0$ and*

$$S \cap V = \left\{p = \phi(q)\,;\ q = (x^1, .., x^n, 0, ..., 0) \in U\right\}\,,$$

i.e. it is a subset of $\mathbb{R}^d$ that can locally be straightened as an n-dimensional plane.

The theorem by Whitney suggests that the notion of a submanifold is sufficient, but given an n-dimensional manifold, it is not always natural to describe it as a subset of a larger $\mathbb{R}^d$ and it is often much easier to consider it intrinsically as a manifold.

Definition 2.6 (Tangent and co-tangent bundles)**.** *Let $\mathcal{M}$ be an n-dimensional differentiable manifold and $p \in \mathcal{M}$. A differentiable function from $]-\varepsilon, \varepsilon[$ to $\mathcal{M}$ is called a differentiable curve. Consider α a differentiable curve such that $\alpha(0) = p$ and $\mathcal{F}$ a family of differentiable functions from $\mathcal{M}$ to $\mathbb{R}$. The tangent vector to the curve α at $t = 0$ is the map $\alpha'(0)$ from $\mathcal{F}$ to $\mathbb{R}$ defined by*

$$\alpha'(0)f = \frac{\mathrm{d}(f \circ \alpha)}{\mathrm{d}t}(0)\,.$$

All such maps are called tangent vectors to $\mathcal{M}$ at p, they form an n-dimensional vector space denoted $T_p\mathcal{M}$ and called the tangent space to $\mathcal{M}$ at p. Its dual $T_p^*\mathcal{M}$ is called the cotangent space to $\mathcal{M}$ at p, its elements are co-vectors at p. We denote by $T\mathcal{M}$ (resp. $T^*\mathcal{M}$) and call tangent bundle (resp. cotangent bundle) the set of pairs (p, X) where $X \in T_p\mathcal{M}$ (resp. $X \in T_p^*\mathcal{M}$). Both are smooth manifolds of dimension $2n$. The sections of $T\mathcal{M}$ (resp. $T^*\mathcal{M}$) are maps from $\mathcal{M}$ to $T\mathcal{M}$ (resp. $T^*\mathcal{M}$) that can be continuous, differentiable, $\mathscr{C}^k$, smooth or even distributional. The sections of $T\mathcal{M}$ are called vector fields and those of $T^*\mathcal{M}$ are called 1-forms. The spaces of sections of $T\mathcal{M}$ and $T^*\mathcal{M}$ are denoted $\Gamma T\mathcal{M}$ and $\Gamma T^*\mathcal{M}$. The differential is a linear map from $\mathscr{D}'(\mathcal{M})$ into ΓT^*M.

Definition 2.7 (Coordinate basis)**.** *Consider a differentiable manifold $\mathcal{M}$ and a local parametrisation $(\mathcal{U}, \phi)$. The coordinates functions $x^1, ..., x^n$ on $\phi(\mathcal{U})$ are the differentiable functions defined by $x^i = \pi^i \circ \phi^{-1}$ where π^i is the projection on the i^{th} axis on $\mathbb{R}^n$. Their differentials $\mathrm{d}x^1, ..., \mathrm{d}x^n$ at each point p are linear forms on $T_p\mathcal{M}$, they form a basis of $T_p^*\mathrm{M}$; its dual basis, denoted $\frac{\partial}{\partial x^1}, ..., \frac{\partial}{\partial x^n}$, or simply $\partial_{x^1}, ..., \partial_{x^n}$, is a basis of $T_p\mathcal{M}$ at each point p. These are called coordinate bases.*

A tensor bundle on $\mathcal{M}$ is a multiple tensor product of the tangent bundle $T\mathcal{M}$ and the cotangent bundle $T^*\mathcal{M}$. The sections of a tensor bundle are called tensor fields. The valence of a tensor bundle reflects the number of times the tangent and cotangent bundle appear in the tensor product. For instance,

$$T\mathcal{M} \otimes T^*\mathcal{M} \otimes T\mathcal{M} \otimes T^*\mathcal{M} \text{ and } T\mathcal{M} \otimes T^*\mathcal{M} \otimes T^*\mathcal{M}$$

are respectively tensor bundles of valence $\begin{bmatrix} 2 \\ 2 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. The valence does not give any information on the order in which we take the tangent and cotangent bundles in the tensor product, but changing the order is meaningless as the different permutations give bundles that can be canonically identified. Tensor fields of valence $\begin{bmatrix} 0 \\ p \end{bmatrix}$ are called p-forms. Differential forms are completely skew p-forms.

2.3 Abstract index formalism

Projecting tensors onto local bases is very useful for doing explicit calculations. The disadvantage of such calculations is that sometimes, they depend on the basis chosen. The intrinsic aspect of the result is therefore often a problem. However, in many cases, the advantage of a local basis is purely notational, in keeping track of the indices. This is what led Roger Penrose to developing the abstract index formalism. A complete axiomatic description of this set of notations is given in Spinors and space-time Vol.1 [45]. We simply intend to give a flavour of the essential idea here in order to be able to use this formalism for explicit calculations.

Abstract indices

Consider T a tensor field of valence $\begin{bmatrix} m \\ n \end{bmatrix}$. We shall denote T with indices, m up and n down, in order to be able to see the nature of this object purely from the way it is denoted. The indices used are always lower case lightface latin letters[1], possibly with indices themselves. Here for instance, we would do well to use a notation like

$$T^{a_1 a_2 \dots a_m}_{b_1 b_2 \dots b_n} .$$

For the moment, the respective position of an index that is up and another that is down is unimportant, for the reason already mentionned earlier that the product on $\mathbb{R}$ is commutative, therefore there is no reason a priori to distinguish between $\alpha \otimes V$ and $V \otimes \alpha$, where α is a 1-form and V a vector field. So we write the up and down indices above one another.

It is important to understand that the notation above does not refer to a collection of components in referrence to a basis. It is the intrinsic tensor field to which we have just put some stickers to see how many legs up and down it has[2]. The tensor T has m 1-form arguments and n vector field arguments. Suppose we wish to express

$$T(\alpha, \beta, ..., \gamma, U, V, ..., W) \tag{2.4}$$

with abstract indices, we shall denote the 1-forms with an index down, since they are tensor fields of valence $\begin{bmatrix} 0 \\ 1 \end{bmatrix}$ and the vector fields with an index up since they are tensor fields of valence $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. Then the notation for (2.4) will be

$$T^{a_1 a_2 \dots a_m}_{b_1 b_2 \dots b_n} \alpha_{a_1} \beta_{a_2} \dots \gamma_{a_m} U^{b_1} V^{b_2} \dots W^{b_n} .$$

There is no sum over indices of course since these are not indices that take numerical values, purely labels. The fact that an index is present once up and once down in the same expression means that a contraction has to take place, this denotes the action of one "leg" of the tensor on a vector or a 1-form. This is the abstract index version of the Einstein convention. The order of the factors in the above expression is irrelevent, the repeated indices simply tell us in what slot a vector or a 1-form should be contracted.

[1] In Penrose's abstract index formalism, indices denoted by upper case latin letters are for spinors and indices denoted by greek letters are for twistors. As for boldface indices, they are concrete indices with referrence to a basis.

[2] Indeed, a more abstract set of notations has been developed by Penrose, consisting purely of diagrams with legs. See [45] for a description of the "legged diagram" formalism.

The tensor bundles of a given valence can be denoted with abstract indices too, for example $T_a S$ denotes $T^* S$ and $T^{ab}_c S$ is the tensor bundle of valence $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$.

The link between the quantities with indices and without indices is formally realized by objects denoted $\mathrm{d}x^a$ and $\frac{\partial}{\partial x^a}$. For instance,

$$V = V^a \frac{\partial}{\partial x^a}\ ,\ \ \alpha = \alpha_a \mathrm{d}x^a\ ,\ \ T = T^a_{bc} \frac{\partial}{\partial x^a} \otimes \mathrm{d}x^b \otimes \mathrm{d}x^c\ .$$

This looks like a decomposition on a basis, but the indices are all abstract, this is purely a formal link between indexed and non indexed quantities. This type of link is required for the coherence of some expressions. Typically, if we integrate a 1-form on a curve, we wish to obtain a scalar, hence without an index, so it is clear that the expression

$$I = \int_{\mathscr{C}} \alpha_a\ ,$$

is inadequate. Instead, the following expression should be used

$$I = \int_{\mathscr{C}} \alpha_a \mathrm{d}x^a\ .$$

Another good reason for using these $\mathrm{d}x^a$ and $\frac{\partial}{\partial x^a}$ conventions is that most expressions should be the same with abstract indices or with concrete indices referring to a basis.

Symmetrizers and anti-symmetrizers

The symmetry operations on a tensor can now be expressed explicitely. If we swap two indices (they have to be both up or both down for this to be legitimate), this means that when applying the tensor to 1-forms and vectors, we shall swap the corresponding arguments. The symmetry operations known as symmetrizers are denoted by parentheses on each side of the group of indices it applies to, and anti-symmetrizers are denoted by square brackets. For example

$$\begin{aligned} T^a_{(bc)d} &= \frac{1}{2}\left(T^a_{bcd} + T^a_{cbd}\right)\ , \\ N^{a[bc]d}_{ef} &= \frac{1}{2}\left(N^{abcd}_{ef} - N^{acbd}_{ef}\right)\ , \\ K_{[abc]} &= \frac{1}{6}\left(K_{abc} + K_{bca} + K_{cab} - K_{bac} - K_{acb} - K_{cba}\right)\ . \end{aligned}$$

If we wish to exclude an index or a group of indices from a symmetry operation, we put them between vertical bars, such as

$$T^{ab}_{(c|de|f)g} = \frac{1}{2}\left(T^{ab}_{cdefg} + T^{ab}_{fdecg}\right)\ .$$

Concrete indices

Concrete indices refer to a given basis and label the components of tensors with respect to this basis, they take numerical values. They are denoted by boldface lower case latin letters. They also label the basis vectors and 1-forms. For instance, a frame $\{V_1, ..., V_k\}$ will be denoted $\{V_{\mathbf{a}}\}_{\mathbf{a}=1,...,k}$ and the dual basis of 1-forms $\{\alpha^{\mathbf{a}}\}_{\mathbf{a}=1,...,k}$. As indexed objects, the basis vectors are denoted $V_{\mathbf{a}}^a$ and the 1-forms $\alpha_a^{\mathbf{a}}$, i.e. we can write

$$V_{\mathbf{a}} = V_{\mathbf{a}}^a \frac{\partial}{\partial x^a} .$$

There is no contraction possible between a concrete index and an abstract index, they are objects of different natures.

The decomposition of a vector or a 1-form in the basis is written as

$$W^a = W^{\mathbf{a}} V_{\mathbf{a}}^a \text{ or } W = W^{\mathbf{a}} V_{\mathbf{a}} ,$$
$$\beta_a = \beta_{\mathbf{a}} \alpha_a^{\mathbf{a}} \text{ or } \beta = \beta_{\mathbf{a}} \alpha^{\mathbf{a}} .$$

For handwriting, boldface letters are not exactly natural, instead one can underline the indices to signify that they are concrete indices.

2.4 Space-time, metric, connection, curvature

The framework of general relativity is a space-time: it is a 4-dimensional manifold equipped with a *Lorentzian metric*. We start by defining this notion.

Definition 2.8 (metric). *Let $\mathcal{M}$ be a smooth manifold of dimension n. A metric on $\mathcal{M}$ is a symmetric 2-form on $\mathcal{M}$ that is non degenerate.*

Consider a metric g on $\mathcal{M}$, for any point p of $\mathcal{M}$ there exists a neighbourhood $\mathcal{U}$ of p and a local frame (i.e. a family of n linearly independent vector fields) $\{V_1, .., V_n\}$ that is orthonormal for g i.e.

$$g(V_{\mathbf{i}}, V_{\mathbf{j}}) = \begin{cases} 0 & \text{if} \quad \mathbf{i} \neq \mathbf{j} , \\ \pm 1 & \text{if} \quad \mathbf{i} = \mathbf{j} . \end{cases}$$

We say that a metric g on $\mathcal{M}$ has signature $+ ... + - ... -$ with k "+" and $n-k$ "−" if, for any point p of $\mathcal{M}$ there exists a neighbourhood $\mathcal{U}$ of p and an orthonormal local frame $\{V_1, .., V_n\}$ on $\mathcal{U}$ such that for exactly k values of $\mathbf{i} \in \{1, ..., n\}$ we have $g(V_{\mathbf{i}}, V_{\mathbf{i}}) = 1$ and for exactly $n-k$ values of $\mathbf{i} \in \{1, ..., n\}$ we have $g(V_{\mathbf{i}}, V_{\mathbf{i}}) = -1$. When the signature is $+ + ... +$, the metric is said to be Riemannian; when it is $+ - ... -$ (resp. $- + ... +$), it is said to be Lorentzian.

We can now give the definition of a space-time.

Definition 2.9. *A space-time is a pair $(\mathcal{M}, g)$ where $\mathcal{M}$ is a 4-dimensional real smooth manifold that is orientable and g is a Lorentzian metric on $\mathcal{M}$ of signature $+ - - -$. The tangent space at each point p of $\mathcal{M}$ equipped with the quadratic form g at p is isometric to Minkowski vector space.*

The relativity community is divided into two sub-communities according to the preference between the signatures $+---$ and $-+++$ for Lorentzian metrics. Each community has good reasons for their choice. By and large, the two conventions are purely a matter of taste, except when one deals with spinors using the 2-spinor formalism, in this case, the signature $+---$ is preferred. This is the convention we adopt.

Remark 2.6 (The metric as an index raising and lowering operator)**.** *Consider a space-time $(\mathcal{M}, g)$. To a vector V^a at a point we can associate a covector by contracting V^a into the metric at that point. We denote by V_a the covector thus obtained*

$$V_a = V^b g_{ab}\,.$$

Since the metric is a non degenerate symmetric 2-form, this operation is an isomorphism between vectors and covectors. We denote by g^{ab} the inverse operator, i.e.

$$V^a = g^{ab} V_b\,.$$

Then g^{ab} is a symmetric tensor of valence $\left[\begin{array}{c}2\\0\end{array}\right]$ and by construction we have

$$g_{ab} g^{bc} = \delta_a^c\,,$$

where δ_a^c is merely an operator that replaces the index a by the index c, i.e. it transforms a vector field into the same vector field but with the index denoted by another letter. In terms of concrete indices, $g_{\mathbf{ab}}$ will be the matrix of the metric in the chosen basis, $g^{\mathbf{ab}}$ will be the inverse matrix and $\delta_{\mathbf{a}}^{\mathbf{c}}$ is the usual Kronecker symbol, that is 1 is $\mathbf{a} = \mathbf{c}$ and 0 otherwise, i.e. the identity matrix.

Remark 2.7 (Musical isomorphisms)**.** *For vector fields and 1-forms, the raising and lowering of indices using the metric is sometimes denoted using the 'sharp' and 'flat' notations and referred to as the musical isomorphisms. So if α_a is a 1-form and V^a a vector field, the notations $\alpha^\sharp$ and $V^\flat$ refer to the vector-field α^a and the 1-form V_a.*

As soon as we start raising and lowering indices using the metric, we realize that the respective position of up and down indices may have some importance. Typically we want to avoid the following absurdity

$$g^{cf} T^{ab}_{cde} = T^{abf}_{de}\,,\ g_{fc} T^{abf}_{de} = T^{ab}_{dec} \text{ whence } T^{ab}_{cde} = T^{ab}_{dec}\,,$$

which looks like a symmetry property whereas it should just be $T = T$. Hence, in some cases where we wish to keep track of indices through raising and lowering operations, we will order all indices, irrespective of their position up or down. We will have notations like

$$g_{ai} T^a{}_{bc}{}^{de}{}_f = T_{ibc}{}^{de}{}_f\,.$$

We will need to differentiate tensor fields (sections of tensor bundles), for this we need a connection, we will use the Levi-Civita connection.

The Levi-Civita connection on a Lorentzian manifold is defined exactly as in the Riemannian case, in fact the definition and uniqueness of the Levi-Civita connection are independent of the signature of the metric.

Definition 2.10. *Let $\mathcal{M}$ be a smooth manifold. A connection ∇_a is an extension of the differential to arbitrary tensor fields, such that:*

1. *it is linear from any tensor bundle F to $T^*\mathcal{M} \otimes F$;*
2. *it satisfies the Leibniz rule.*

Theorem 2.1 (Levi-Civita connection)**.** *Let $(\mathcal{M}, g)$ be a space-time. There exists a unique connection ∇_a such that:*

1. *it is torsion-free, meaning that $[\nabla_a\,,\ \nabla_b]\, f = 0$ for any scalar field f, where $[\nabla_a\,,\ \nabla_b]$ is the commutator of ∇_a and ∇_b, $[\nabla_a\,,\ \nabla_b] = \nabla_a\nabla_b - \nabla_b\nabla_a$;*
2. *it commutes with the metric, i.e. $\nabla_a g_{bc} = 0$ and $\nabla_a g^{bc} = 0$.*

It is called the Levi-Civita connection. In a local coordinate basis $e_{\mathbf{a}} = \partial_{\mathbf{a}}$, $\mathrm{d}x^{\mathbf{a}}$, denoting

$$\nabla_{\mathbf{a}} = \nabla_{\partial_{\mathbf{a}}} = (\partial_{\mathbf{a}})^a \nabla_a\,,$$

its action on vector fields and 1-forms is given by:

$$\begin{aligned}(\nabla_{\mathbf{a}} V)^{\mathbf{b}} &= \mathrm{d}x^{\mathbf{b}}(\nabla_{\mathbf{a}} V) = \partial_{\mathbf{a}} V^{\mathbf{b}} + \Gamma_{\mathbf{ac}}{}^{\mathbf{b}} V^{\mathbf{c}}\,,\\ (\nabla_{\mathbf{a}}\omega)_{\mathbf{b}} &= (\nabla_{\mathbf{a}}\omega)(e_{\mathbf{b}}) = \partial_{\mathbf{a}}\omega_{\mathbf{b}} - \Gamma_{\mathbf{ab}}{}^{\mathbf{c}}\omega_{\mathbf{c}}\end{aligned}$$

and for a tensor field of valence $\begin{bmatrix} p \\ q \end{bmatrix}$,

$$\begin{aligned}(\nabla_{\mathbf{a}} K)^{\mathbf{i}_1\ldots\mathbf{i}_p}{}_{\mathbf{j}_1\ldots\mathbf{j}_q} = \partial_{\mathbf{a}} K^{\mathbf{i}_1\ldots\mathbf{i}_p}{}_{\mathbf{j}_1\ldots\mathbf{j}_q} &- \Gamma_{\mathbf{a}\mathbf{j}_1}{}^{\mathbf{b}} K^{\mathbf{i}_1\ldots\mathbf{i}_p}{}_{\mathbf{b}\ldots\mathbf{j}_q} - \ldots - \Gamma_{\mathbf{a}\mathbf{j}_q}{}^{\mathbf{b}} K^{\mathbf{i}_1\ldots\mathbf{i}_p}{}_{\mathbf{j}_1\ldots\mathbf{b}}\\ &+ \Gamma_{\mathbf{ab}}{}^{\mathbf{i}_1} K^{\mathbf{b}\ldots\mathbf{i}_p}{}_{\mathbf{j}_1\ldots\mathbf{j}_q} + \ldots + \Gamma_{\mathbf{ab}}{}^{\mathbf{i}_p} K^{\mathbf{i}_1\ldots\mathbf{b}}{}_{\mathbf{j}_1\ldots\mathbf{j}_q}\,,\end{aligned} \tag{2.5}$$

where the Christoffel symbols are defined by

$$\Gamma_{\mathbf{ab}}{}^{\mathbf{c}} = \frac{1}{2} g^{\mathbf{cd}} \left(\partial_{\mathbf{a}} g_{\mathbf{bd}} + \partial_{\mathbf{b}} g_{\mathbf{ad}} - \partial_{\mathbf{d}} g_{\mathbf{ab}}\right) \tag{2.6}$$

and satisfy

$$\Gamma_{\mathbf{ab}}{}^{\mathbf{c}} = \Gamma_{(\mathbf{ab})}{}^{\mathbf{c}}\,.$$

Remark 2.8. *It is important to note that the Christoffel symbols $\Gamma_{\mathbf{ab}}{}^{\mathbf{c}}$ are not a tensor field: it is very easy to see that they depend on the choice of local coordinates. For instance, on Minkowski space-time in cartesian coordinates, the Christoffel symbols are all zero, but this is not the case in spherical coordinates. However, the connection is an intrinsic object independent of the coordinate system.*

Proposition 2.4. *When the commutator of two covariant derivatives acts on tensor fields of arbitrary valence, it involves a new tensor field: the Riemann curvature tensor R_{abcd}. More precisely,*

$$\begin{aligned}[\nabla_a\,,\ \nabla_b]\, K^{i_1\ldots i_p}{}_{j_1\ldots j_q} &= R_{abc}{}^{i_1} K^{c\ldots i_p}{}_{j_1\ldots j_q} + \ldots + R_{abc}{}^{i_p} K^{i_1\ldots c}{}_{j_1\ldots j_q}\\ &\quad - R_{abj_1}{}^{d} K^{i_1\ldots i_p}{}_{d\ldots j_q} - \ldots - R_{abj_q}{}^{d} K^{i_1\ldots i_p}{}_{j_1\ldots d}\,.\end{aligned} \tag{2.7}$$

In a local coordinate basis, its expression in terms of the Christoffel symbols is given by

$$R_{\mathbf{abc}}{}^{\mathbf{d}} = \partial_{\mathbf{b}}\left(\Gamma_{\mathbf{ac}}{}^{\mathbf{d}}\right) - \partial_{\mathbf{a}}\left(\Gamma_{\mathbf{bc}}{}^{\mathbf{d}}\right) + \Gamma_{\mathbf{bc}}{}^{\mathbf{e}}\Gamma_{\mathbf{ae}}{}^{\mathbf{d}} - \Gamma_{\mathbf{ac}}{}^{\mathbf{e}}\Gamma_{\mathbf{be}}{}^{\mathbf{d}}\,. \tag{2.8}$$

Although the Christoffel symbols and the Riemann tensor can be calculated by hand, it is much faster and often safer to use a formal calculus software (we strongly recommend SageMaths [50] which is free and open source) for such calculations.

Theorem 2.2. *The Riemann tensor has the following symmetries:*

1. $R_{(ab)cd} = 0$;
2. $R_{ab(cd)} = 0$;
3. $R_{[abc]}{}^{d} = 0$; *it is the first Bianchi identity which, using* $R_{(ab)cd} = 0$, *becomes*
$$R_{abc}{}^{d} + R_{bca}{}^{d} + R_{cab}{}^{d} = 0\,;$$
4. $\nabla_{[a} R_{bc]d}{}^{e} = 0$ *(second Bianchi identity).*

Definition 2.11. *We define some important curvature quantities that are special parts of the full Riemann tensor:*

- *the Ricci tensor* R_{ab} *is the trace of the Riemann tensor in its second and fourth indices*
$$R_{ab} := R_{acb}{}^{c} = g^{cd} R_{acbd}\,;$$
- *the scalar curvature* R *is the trace of the Ricci tensor*
$$R := R_a{}^{a} = g^{ab} R_{ab}$$
and it is often denoted by Scal_g;
- *the Einstein tensor* G_{ab} *is defined as*
$$G_{ab} := R_{ab} - \frac{1}{2} R g_{ab}\,;$$
- *the Weyl tensor* C_{abcd} *is the trace-free part of the Riemann tensor*
$$C_{abcd} = R_{abcd} - \frac{1}{2}\left(g_{a[c}R_{d]b} - g_{b[c}R_{d]a}\right) + \frac{1}{3} R g_{a[c} g_{d]b}\,.$$

Proposition 2.5. *We have the following properties of the Ricci and Einstein tensors:*

1. $R_{ab} = R_{(ab)}$ *(which implies immediately* $G_{ab} = G_{(ab)}$*)*;
2. $\nabla^a G_{ab} = 0$.

The **Einstein vacuum equations** that characterize the geometry of an empty universe are simply
$$G_{ab} = 0\,. \tag{2.9}$$

In the case of a universe containing energy or matter, the Einstein equations will become

$$G_{ab} = 8\pi T_{ab}$$

where T_{ab} is a tensor (referred to as the stress-energy tensor) describing the distribution of matter and energy in the universe.

Considered as an equation on the metric, Einstein's equations are a system of non linear second order partial differential equations. Taking the trace of G_{ab}, we obtain

$$G_a{}^a = R_a{}^a - \frac{1}{2}Rg_a{}^a = R - 2R = -R\,,$$

whence (2.9) is equivalent to

$$R_{ab} = 0\,. \tag{2.10}$$

Einstein vacuum space-times are also referred to as Ricci-flat space-times.

There is a modified version of the Einstein equation, due to Einstein himself in 1917, involving a constant Λ called the "cosmological constant". It has the following form

$$G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}\,. \tag{2.11}$$

A usual description of the reason why Einstein introduced this modification is that, for religious reasons, Einstein favoured a static universe and the original form of the theory did not allow for it unless it is also flat. The cosmological constant induces a repulsive force which Einstein adjusted so that it would counterbalance gravitation exactly. His new version of the theory thus allows for a static universe: the Einstein cylinder which we shall encounter again later. In fact, the reason is completely different and of a scientific nature. Einstein was trying to describe the movement of galaxies using statistical mechanics and trivial topology led to inconsistencies in their behaviour. This is what led him to change the topology and a natural manner of doing this was to introduce the cosmological constant. This unfortunately prevented him from discovering the expansion of the universe which Hubble proved in 1929. He subsequently declared that this was his greatest mistake. It is interesting to notice that observations made from 1993 to 2005 show that the expansion of the universe is now faster than we would expect. A well accepted explanation is that a repulsive force induced by a cosmological constant is responsible for it: in the early stages of the universe, the expansion from the big bang was slowed down by gravity, but as the universe expanded, the effects of gravity weakened and this repulsive force (referred to as dark energy) accelerated the expansion. The universe would appear to have a small but strictly positive cosmological constant. It is regrettable that Einstein never knew that his greatest mistake was just another brilliant idea.

Taking the trace of (2.11), we see that in the vacuum case, i.e. for $T_{ab} = 0$, the cosmological constant is a multiple of the scalar curvature:

$$\Lambda = \frac{1}{4}R\,.$$

2.5 Causality, global hyperbolicity

If $(\mathcal{M}, g)$ is a space-time, then we can find in the neighbourhood of each point an orthonormal basis. In such a basis, the metric g is described by the matrix

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

The tangent space at each point is therefore a copy of Minkowski vector space. This gives us natural definitions of timelike, spacelike, null and causal vectors and a similar classification for curves and hypersurfaces.

Definition 2.12. *Let $p \in \mathrm{M}$, a vector $V \in T_p\mathcal{M}$ is said to be*

- *spacelike if $g(V, V) < 0$,*
- *null if $g(V, V) = 0$,*
- *timelike if $g(V, V) > 0$,*
- *causal (or also non-spacelike) if $g(V, V) \geq 0$.*

The definitions of timelike, spacelike, etc... for curves and hypersurfaces follow exactly as they do in Minkowski space.

Definition 2.13. *A time orientation on a space-time $(\mathcal{M}, g)$ is a globally defined nowhere vanishing continuous timelike vector field on $\mathcal{M}$. If a time orientation exists on $(\mathcal{M}, g)$, the space-time is said to be time orientable.*

Definition 2.14. *Let $(\mathcal{M}, g)$ be a time orientable space-time and T^a a time orientation. A causal vector V at a point is then said to be future oriented (resp. past oriented) if $g_{ab}V^aT^b > 0$ (resp. $g_{ab}V^aT^b < 0$).*

The following proposition establishes, among other things, that the definition above is meaningful.

Proposition 2.6. *Let $(\mathcal{M}, g)$ be a time orientable space-time on which we consider T^a and τ^a two time orientations. Then one of the two following assertions is true:*

($\imath$) for any causal vector V at a given point, the signs of $g_{ab}V^aT^b$ and $g_{ab}V^a\tau^b$ are the same; the orientations are then said to be the same or compatible;

($\imath$) for any causal vector V at a given point, the signs of $g_{ab}V^aT^b$ and $g_{ab}V^a\tau^b$ are opposite; the orientations are then said to be opposite.

Proof. On Minkowski space-time, we have the following result.

Lemma 2.1. *Consider a timelike vector $T = T^0\partial_t + T^1\partial_x + T^2\partial_y + T^3\partial_z$ and a non-zero causal vector $V = V^0\partial_t + V^1\partial_x + V^2\partial_y + V^3\partial_z$ at a point p in $\mathbb{M}$. Then the sign of $\eta(T, V)$ is that of T^0V^0.*

Proof. We have $T^0 \neq 0$ since T is timelike and $V^0 \neq 0$ since V is non-zero and causal. Then denoting $T = T^0\partial_t + T'$ and $V = V^0\partial_t + V'$,

$$\begin{aligned}\eta(T, V) &= T^0V^0 - \langle T', V'\rangle \\ &= T^0V^0\left(1 - \frac{\langle T', V'\rangle}{T^0V^0}\right).\end{aligned}$$

Now

$$\left|\frac{\langle T', V'\rangle}{T^0V^0}\right| \leq \frac{|T'|}{|T^0|}\frac{|V'|}{|V^0|} < 1$$

since T is timelike and V is non-zero an causal. Hence the result. □

Lemma 2.1 entails the result of Proposition 2.6 by choosing orthonormal frames in an open covering of $\mathcal{M}$. □

Proposition 2.7. *Let $(\mathcal{M}, g)$ be a time orientable space-time. A spacelike vector has no time orientation. More precisely, given V a spacelike vector at a point p, there exist two choices $T_\pm$ of timelike vectors at p with the same time orientation such that $g(T_\pm, V)$ have opposite signs.*

Proof. Consider W a timelike vector at p orthogonal to V. For $\varepsilon > 0$ small enough, the vectors

$$T_\pm := W \pm \varepsilon V$$

are both timelike and belong to the same component of the light-cone at p, moreover the quantities

$$g(T_\pm, V) = \pm\varepsilon g(V, V)$$

are non zero and have opposite signs. □

An important notion is that of the domain of dependence of a set (we also define the domain of influence):

Definition 2.15. *Let $(\mathcal{M}, g)$ be a time orientable space-time on which a time orientation has been chosen. We consider a set A in $\mathcal{M}$.*

- **Domain of influence.** *The future (resp. past) domain of influence of A in $(\mathcal{M}, g)$ is the set of points of $\mathcal{M}$ that can be reached from a point of A along a future (resp. past) oriented causal curve. These are often merely referred to as the future or the past of A. The domain of influence of A is the union of its future and past domains of influence.*
- **Domain of dependence.** *The future (resp. past) domain of dependence of A in $(\mathcal{M}, g)$ is the set of points of $\mathcal{M}$ such that all inextendible oriented causal curves containing M intersect A in the past (resp. future) of M. The domain of dependence of A is the union of its future and past domains of dependence.*

This is related to the concepts of Cauchy hypersurfaces and global hyperbolicity. Of all the equivalent definitions that have been proposed for a globally hyperbolic space-time, the first one being due to Leray, the clearest is certainly that which R.P. Geroch put forward in 1970 [18]. The fundamental definition is that of a Cauchy hypersurface.

Definition 2.16 (Cauchy hypersurface)**.** *Let $(\mathcal{M}, g)$ be a time orientable space-time. A Cauchy hypersurface on $(\mathcal{M}, g)$ is a hypersurface Σ satisfying:*

1. *Σ is spacelike;*
2. *every inextendible timelike curve intersects Σ at exactly one point (in particular, this entails that the domain of dependence of Σ is $\mathcal{M}$).*

We see that this is an adequate surface on which to impose initial data for covariant equations (a covariant equation on a Lorentzian space-time will necessarily be a generalization to the case of a curved space-time of covariant equations on Minkowski space, which are hyperbolic equations), since they propagate the information at finite speed lower than or equal to the speed of light; the condition that the domain of dependence of Σ should be the whole space-time is exactly what ensures that by specifying some data on Σ, we have enough information to propagate the solution to the whole space-time. Moreover, the second condition is here to guarantee that the information propagated along causal geodesics does not come back to a point where the solution is already determined, thus creating some possible incompatibility. A globally hyperbolic space-time as defined by Geroch is simply a space-time that admits a Cauchy hypersurface.

Definition 2.17. *A time-orientable space-time $(\mathcal{M}, g)$ is said to be globally hyperbolic if it admits a Cauchy hypersurface.*

Remark 2.9. *The original definition of global hyperbolicity is more complicated. It is the property for a time orientable space-time to be causal (i.e. to admit no causal loop) and to be such that the intersection of the past of one point and the future of another, if non empty, is compact. Geroch [18] established that this is equivalent to the existence of a Cauchy hypersurface. This is also equivalent to saying that the set of C^1 causal future-oriented paths between any two points is compact in the C^1 topology. For a thorough presentation of all the notions of causality one can define on a space-time, see [36].*

So globally hyperbolic space-times are essentially the space-times for which the Cauchy problem makes sense. The space-times in which it is hardest to make any sense at all of the Cauchy problem are called totally vicious space-times, they are such that any point can be reached from any other point in the space-time along a future oriented timelike curve. An example of a totally vicious part of a space-time is the inner part of a rotating black hole.

In fact, global hyperbolicity has stronger consequences: the existence of a smooth time function t whose level hypersurfaces Σ_t are all Cauchy hypersurfaces and are diffeomorphic to a fixed 3-surface Σ. For a long time, the only available proof of this result was due to Geroch and his construction only guaranteed the existence of a continuous time function whose level hypersurfaces were homeomorphic to a fixed hypersurface. The work of Bernal and Sanchez [3, 4] proved that the time function can be chosen smooth when the metric is smooth. Their result in fact gives a $\mathscr{C}^k$ time function when the metric is $\mathscr{C}^k$.

Examples.

1. Minkowski space-time is a globally hyperbolic space-time and t is a natural global time function on it whose level hypersurfaces are all diffeomorphic to $\mathbb{R}^3$ and are Cauchy hypersurfaces.

2. Removing a single point p from Minkowski space-time deprives it of its global hyperbolicity. Any inextendible timelike curve γ in $\mathbb{M}$ going through p will be split in the new space-time into two inextendible timelike curves, γ_1 lying in the future of p and γ_2 lying in its past. Any spacelike hypersurface on Minkowski space-time that does not contain p, fails to intersect either γ_1 or γ_2. Therefore any spacelike hypersurface in the new space-time fails to intersect either γ_1 or γ_2. It follows that the new space-time does not admit a Cauchy hypersurface.

3. Another less obvious example can be obtained on $\mathbb{R} \times \mathbb{R}^n$ with a variable speed of light as follows
$$g = v(r)^2 \mathrm{d}t^2 - \mathrm{d}r^2 - r^2 \mathrm{d}\omega^2 \,,$$
where $\mathrm{d}\omega^2$ is the euclidean metric on the $(n-1)$-sphere. The lightcone of the origin will exist globally and be generated by integral curves of the vector field
$$V = \frac{\partial}{\partial t} + v(r) \frac{\partial}{\partial r} \,.$$
For any such integral line starting from the origin and for any $r_1 > 0$, the time t_1 at which it reaches the value $r = r_1$, if it does, is given by
$$t_1 = \int_0^{r_1} \frac{1}{v(r)} \mathrm{d}r \,.$$
Take for example
$$v(r) = 1 + r^2 \,.$$
Then the lightcone of the origin reaches infinity in the future in finite time, at $t = \frac{\pi}{2}$ precisely since
$$\int_0^{+\infty} \frac{1}{1+r^2} \mathrm{d}r = \frac{\pi}{2} \,.$$
Moreover the lightcone of the origin is time symmetric.

   Any spacelike slice going through the origin will be caught between the two components of the lightcone and will therefore be contained in the strip
$$\Omega =]-\frac{\pi}{2}, \frac{\pi}{2}[\times \mathbb{R}^n \,.$$
Consider now an integral curve of the timelike future-oriented vector field
$$U = \frac{\partial}{\partial t} + \frac{v(r)}{2} \frac{\partial}{\partial r}$$
going through the point $(2\pi, 0) \in \mathbb{R} \times \mathbb{R}^n$. These curves start at infinity at time π and end up at infinity at time 3π. Therefore they are inextendible timelike curves that never

meet any spacelike slice going through the origin. For any spacelike slice, let $(t_0, 0)$ be the coordinates of its intersection with the time axis, then by translating the previous construction in time by t_0 we also obtain families of inextendible timelike curves that do not touch the slice. This space-time therefore does not admit any Cauchy hypersurface; it is not globally hyperbolic.

4. Another classic example is the universal covering of anti-de Sitter space-time. It is given in stereographic coordinates on $\mathbb{R}_t \times \mathbb{R}_r^+ \times S_\omega^2$ by the metric

$$g = (1+r^2)\mathrm{d}t^2 - \frac{1}{1+r^2}\mathrm{d}r^2 - r^2\mathrm{d}\omega^2 \,.$$

We have a family of null vectors at each point (t, r, ω) given by

$$V = \frac{\partial}{\partial t} + (1+r^2)\frac{\partial}{\partial r} \,.$$

We have exactly the same effect of widening of the light-cones as in the previous example. The light-cone of the origin reaches infinity at time $t = \pi/2$.

2.6 Differential forms and conservation laws

Recall that the bundle of differential 1-forms on $\mathcal{M}$ is simply $\Lambda^1(\mathcal{M}) = T^*\mathcal{M}$ and the bundle of differential p-forms is the p^{th} exterior power of $\Lambda^1(\mathcal{M})$, i.e.

$$\Lambda^p(\mathcal{M}) = \underbrace{\Lambda^1(\mathcal{M}) \wedge \Lambda^1(\mathcal{M}) \wedge \ldots \wedge \Lambda^1(\mathcal{M})}_{p \text{ times}} \,,$$

it is the totally skew part of

$$\underbrace{T^*\mathcal{M} \otimes T^*\mathcal{M} \otimes \ldots \otimes T^*\mathcal{M}}_{p \text{ times}} \,.$$

Definition 2.18 (Volume form). *The volume-form on $(\mathcal{M}, g)$ is the 4-form e whose expression in a coordinate basis is given by (the ordering of coordinates being chosen in agreement with the orientation of $\mathcal{M}$)*

$$e = \sqrt{|g|}\mathrm{d}x^0 \wedge \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3 \,. \tag{2.12}$$

Equivalently, it is defined as follows : consider an ortonormal basis $\mathcal{B} = \{e_0, e_1, e_2, e_3\}$, for any set of 4 vectors $\{U, V, W, Z\}$, denoting U^0, U^1, U^2, U^3 the components of U in $\mathcal{B}$, etc..., we have

$$e_{abcd}U^aV^bW^cZ^d = \det \begin{pmatrix} U^0 & V^0 & W^0 & Z^0 \\ U^1 & V^1 & W^1 & Z^1 \\ U^2 & V^2 & W^2 & Z^2 \\ U^3 & V^3 & W^3 & Z^3 \end{pmatrix} \,. \tag{2.13}$$

We shall often simply denote the volume form dVol*, or* dVol_g *to make the relation to the metric explicit.*

The volume form has the following useful properties :

Proposition 2.8. *The volume form is covariantly constant, i.e.*

$$\nabla_i e_{abcd} = 0 \,.$$

Moreover,

$$e_{abcd}e^{pqrs} = -24 g_a^{[p} g_b^{q} g_c^{r} g_d^{s]}\,,\ e_{abcd}e^{pqrd} = -6 g_a^{[p} g_b^{q} g_c^{r]}\,,\ e_{abcd}e^{pqcd} = -4 g_a^{[p} g_b^{q]}\,,$$
$$e_{abcd}e^{pbcd} = -6 g_a^{p}\,,\ e_{abcd}e^{abcd} = -24\,,\ e_{ab}{}^{cd}e_{cd}{}^{pq} = -4 g_a^{[p} g_b^{q]}\,.$$

Proof. The proof of the covariant constancy follows easily from the expression of the volume form in terms of the spinorial symplectic forms (see [45], Vol. 1, p. 138, eq. (3.3.31)). The proof of the other properties is merely a matter of counting permutations. □

We shall essentially use differential 1-forms and differential 3-forms (often simply referred to as 1-forms and 3-forms) in the context of conservation laws (exact or approximate). Hence we will often make use of the Hodge duality and of Stokes' theorem.

Definition 2.19 (Hodge dual)**.** *Let $\omega \in \Gamma(\Lambda^p(\mathcal{M}))$, $0 \le p \le 4$, the Hodge dual of ω is the $(4-p)$-form defined by*

$$*\omega := \frac{1}{p!} e \underbrace{\lrcorner \ldots \lrcorner}_{p \text{ times}} \omega \,, \tag{2.14}$$

where e is the volume-form on $(\mathcal{M}, g)$. More explicitely,

- *for a 0-form f*
$$(*f)_{abcd} = f e_{abcd}\,;$$
- *for a 1-form α*
$$(*\alpha)_{abc} = e_{abcd}\alpha^d\,;$$
- *for a 2-form β*
$$(*\beta)_{ab} = \frac{1}{2} e_{abcd}\beta^{cd}\,;$$
- *for a 3-form γ*
$$(*\gamma)_a = \frac{1}{6} e_{abcd}\gamma^{bcd}\,;$$
- *for a 4-form ϵ*
$$(*\epsilon) = \frac{1}{24} e_{abcd}\epsilon^{abcd}\,.$$

The Hodge star is an isomorphism between p-forms and $(4-p)$-forms, as the following property, which is a direct consequence of Proposition 2.8, shows :

Proposition 2.9. *For a p-form α, we have*

$$*(*\alpha) = (-1)^{p+1}\alpha\,.$$

Proof. This is obvious for a 0-form, let us check the property for the other types of forms.

- $p = 1$:

$$\begin{aligned} *(*\alpha)_a &= \frac{1}{6} e_{abcd} e^{bcdi} \alpha_i \\ &= -\frac{1}{6} e_{abcd} e^{bcdi} \alpha_i = -\frac{1}{6}(-6 g^i_a)\alpha_i = \alpha_a \,. \end{aligned}$$

- $p = 2$:

$$\begin{aligned} *(*\alpha)_{ab} &= \frac{1}{2} e_{abcd} e^{cdij} \alpha_{ij} \\ &= \frac{1}{2} e_{abcd} \frac{1}{2} e^{ijcd} \alpha_{ij} \\ &= \frac{1}{4}(-4 g^i_a g^j_b)\alpha_{ij} \text{ since } \alpha \text{ is skew,} \\ &= -\alpha_{ab} \,. \end{aligned}$$

- $p = 3$:

$$\begin{aligned} *(*\alpha)_{abc} &= e_{abcd} \frac{1}{6} e^{dijk} \alpha_{ijk} \\ &= -\frac{1}{6} e_{abcd} \frac{1}{2} e^{ijkd} \alpha_{ijk} \\ &= -\frac{1}{6}(-6 g^{[i}_a g^j_b g^{k]}_c)\alpha_{ijk} \\ &= g^i_a g^j_b g^k_c \alpha_{ijk} \text{ since } \alpha \text{ is skew,} \\ &= \alpha_{abc} \,. \end{aligned}$$

- $p = 4$. In this case we have $\alpha_{abcd} = f e_{abcd}$, hence

$$\begin{aligned} *(*\alpha)_{abcd} &= e_{abcd} \frac{1}{24} e^{ijkl} \alpha_{ijkl} \\ &= \frac{1}{24} e_{abcd} e^{ijkl} f e_{ijkl} \\ &= \frac{1}{24}(-24) f e_{abcd} = -\alpha_{abcd} \,. \end{aligned}$$

This proves the proposition. □

The Hodge $*$ operator has the following property, that entirely characterizes it :

Theorem 2.3. *For any two p-forms α, β, $1 \leq p \leq 3$,*

$$\alpha \wedge *\beta = (-1)^p \frac{(4-p)!}{4!} \langle \alpha \,,\ \beta \rangle_g \, e \,, \tag{2.15}$$

where

$$\langle \alpha \,,\ \beta \rangle_g = \alpha_{a_1 \dots a_p} \beta^{a_1 \dots a_p} \,.$$

Proof. We write the proof for each value of p.

- $p = 1$. Since the quantity $\alpha \wedge *\beta$ is a 4-form, it is necessarily a multiple of the volume form, all we need to do is work out the proportionality factor. We proceed as follows : since
$$\alpha \wedge *\beta = \alpha_{[a}\,(*\beta)_{bcd]} = f e_{abcd}$$
then
$$\alpha_{[a}\,(*\beta)_{bcd]}e^{abcd} = -24f\,.$$
We calculate
$$\begin{aligned}\alpha_{[a}\,(*\beta)_{bcd]}e^{abcd} &= \alpha_a\,(*\beta)_{bcd}e^{abcd}\\ &= \alpha_a e_{bcdi}\beta^i e^{abcd}\\ &= -e_{ibcd}e^{abcd}\alpha_a\beta^i = 6g_i^a\alpha_a\beta^i = 6\alpha_i\beta^i\,.\end{aligned}$$
Whence
$$(\alpha \wedge *\beta)_{abcd} = -\frac{1}{4}\alpha_i\beta^i e_{abcd}\,.$$

- $p = 2$. We proceed similarly :
$$\begin{aligned}\alpha_{[ab}\,(*\beta)_{cd]}e^{abcd} &= \alpha_{ab}\,(*\beta)_{cd}e^{abcd}\\ &= \alpha_{ab}\frac{1}{2}e_{cdij}\beta^{ij}e^{abcd}\\ &= \frac{1}{2}e_{ijcd}e^{abcd}\alpha_{ab}\beta^{ij}\\ &= \frac{1}{2}(-4g_i^{[a}g_j^{b]})\alpha_{ab}\beta^{ij}\\ &= -2g_i^a g_j^b\alpha_{ab}\beta^{ij} \text{ since } \alpha \text{ is skew,}\\ &= -2\alpha_{ij}\beta^{ij}\,.\end{aligned}$$
Whence
$$(\alpha \wedge *\beta)_{abcd} = \frac{1}{12}\alpha_{ij}\beta^{ij}e_{abcd}\,.$$

- $p = 3$:
$$\begin{aligned}\alpha_{[abc}\,(*\beta)_{d]}e^{abcd} &= \alpha_{abc}\,(*\beta)_d e^{abcd}\\ &= \alpha_{abc}\frac{1}{6}e_{dijk}\beta^{ijk}e^{abcd}\\ &= -\frac{1}{6}e_{ijkd}e^{abcd}\alpha_{abc}\beta^{ijk}\\ &= -6(-\frac{1}{6})g_i^{[a}g_j^b g_k^{c]}\alpha_{abc}\beta^{ijk}\\ &= g_i^a g_j^b g_k^c\alpha_{abc}\beta^{ijk} \text{ since } \alpha \text{ is skew,}\\ &= \alpha_{ijk}\beta^{ijk}\,.\end{aligned}$$
Whence
$$(\alpha \wedge *\beta)_{abcd} = -\frac{1}{24}\alpha_{ijk}\beta^{ijk}e_{abcd}\,.$$

This concluded the proof. □

If instead of taking the exterior product with another 1-form, we take the exterior derivative of the Hodge dual of a 1-form, we recover a 4-form which is $(-1/4)$ times the dual of the divergence of the associated vector field:

Proposition 2.10. *Let α be a 1-form, we have*

$$(\mathrm{d}(*\alpha))_{abcd} = -\frac{1}{4}\nabla_i \alpha^i e_{abcd}\,.$$

Proof. The proof is analogous to that of the case $p=1$ in the above theorem, with just an additional ingredient : the covariant constancy of the volume form. □

We now turn to conservation laws, i.e. to Stokes' theorem for the Hodge dual of a 1-form. Recall Stokes' theorem for a 3-form :

Theorem 2.4. *Let Ω a bounded open subset of $\mathcal{M}$ with piecewise $\mathcal{C}^1$ boundary S. Let $\omega \in \Gamma(\Lambda^3(\mathcal{M}))$, $\mathcal{C}^1$ on $\bar{\Omega}$, then*

$$\int_S \omega = \int_\Omega \mathrm{d}\omega\,.$$

In the case where ω is the Hodge dual of a 1-form α the above equality gives the Lorentzian generalization of the divergence theorem.

Theorem 2.5 (The divergence theorem)**.** *Let Ω be a bounded open subset of $\mathcal{M}$ with piecewise $\mathscr{C}^1$ boundary S, n^a an outgoing normal vector field to S and l^a a vector field transverse to S such that $l_a n^a = 1$. Let α be a 1-form $\mathscr{C}^1$ on $\bar{\Omega}$, then*

$$\int_S \alpha_a n^a (l\lrcorner \mathrm{dVol}) = \int_\Omega \nabla_a \alpha^a \mathrm{dVol}\,.$$

Proof. We have essentially already proved the result above. Take the 3-form ω to be

$$\omega = *\alpha\,,$$

then

$$\mathrm{d}\omega = -\frac{1}{4}\nabla_a\alpha^a \mathrm{dVol}\,.$$

Moreover, denoting by $n^\flat$ the 1-form $(n^\flat)_a = n_a$ (see Remark 2.7),

$$\begin{aligned} \int_S *\alpha &= \int_S \langle l, n\rangle_g * \alpha \\ &= \int_S l\lrcorner(n^\flat \wedge *\alpha) - \underbrace{\int_S n^\flat \wedge (l\lrcorner * \alpha)}_{=0 \text{ since } n\perp S} \\ &= \int_S (-\frac{1}{4})\alpha_a n^a (l\lrcorner \mathrm{dVol}). \end{aligned} \tag{2.16}$$

This concludes the proof. □

The left-hand side of the equality in Theorem 2.5 is the outgoing flux of the vector field $\alpha^a = g^{ab}\alpha_b$, also denoted $\alpha^\sharp$. Generally, we can define the flux of a vector field across an oriented hypersurface as follows.

Definition 2.20. *Let S be an oriented, piecewise $\mathscr{C}^1$ hypersurface and J a vector field defined in the neighbourhood of S. Let n be a normal vector field to S whose orientation is compatible with that of S and l a transverse vector field to S such that $g(l,n) = 1$. The flux of J across S is defined by the two formulae whose equality is established in* (2.16)

$$\begin{aligned} \mathcal{F}_S(J) &= \int_S J_a n^a \,(l \lrcorner \mathrm{dVol}) \\ &= -4 \int_S *J^\flat . \end{aligned}$$

2.7 Flow of a vector field, Lie derivative, Killing vectors

Beside the covariant derivative along a vector field, there is an important type of directional derivative called the Lie derivative. It is independent of a choice of connection and is a derivation along the flow of a vector field. We start by defining the flow of a vector fields and its action on tensors.

Consider on a space-time $(\mathcal{M}, g)$ a $\mathcal{C}^1$ vector field V, i.e. a $\mathcal{C}^1$ section of $T\mathcal{M}$.

Definition 2.21 (Integral curve)**.** *An integral curve of V is a curve in $\mathcal{M}$ that is a maximal solution to the equation*

$$\gamma'(s) = V(\gamma(s)) \,. \tag{2.17}$$

By the Cauchy-Lipschitz theorem (used in open sets of $\mathbb{R}^n$ through local charts), we have existence and uniqueness of maximal solutions to the Cauchy problem for (2.17). This allows us to define the associated propagator or flow of the vector field. A more detailed use of the machinery of the theory of ordinary differential equations shows that it is a local 1-parameter group of diffeomorphisms.

Definition 2.22 (Flow)**.** *The flow of the vector field V is a family of mappings $\Phi_V(s)$ that to a point p in $\mathcal{M}$ associate $\gamma_p(s)$, where γ_p is the unique maximal solution to the Cauchy problem*

$$\gamma_p'(s) = V(\gamma_p(s)) \,,\ \gamma_p(0) = p \,.$$

Remark 2.10. *Since the maximal solution does not necessarily exist for all values of s, the mapping $\Phi_V(s)$ is not usually globally defined, except of course $\Phi_V(0)$ which is the identity. However, if $\Phi_V(s)$ is well defined at a point $p \in \mathcal{M}$, it is defined in a neighbourhood of p.*

Proposition 2.11. *Let V be a $\mathcal{C}^k$ vector field, then its flow Φ_V is a local 1-parameter group of $\mathcal{C}^k$ diffeomorphisms, i.e. it has the following properties:*

1. *given $s \in \mathbb{R}$ and an open set $\mathcal{U}$ of $\mathcal{M}$ on which $\Phi_V(s)$ is well defined, $\Phi_V(s)$ is a $\mathcal{C}^k$ diffeomorphism from $\mathcal{U}$ onto $\mathcal{V} = \Phi_V(s)(\mathcal{U})$;*

2. for any $s_1, s_2 \in \mathbb{R}$, we have $\Phi_V(s_1)\Phi_V(s_2) = \Phi_V(s_1+s_2)$ wherever all quantities are defined.

We omit the proof of this result and refer to the classic theory of ordinary differential equations for it. A good reference in french is the book by Zuily and Queffélec [62], see also Teschl's monograph [58]. It is important to understand that the second property as well as the invertibility of $\Phi_V(t)$ are trivial consequences of the uniqueness of maximal solutions of the Cauchy problem. The delicate part of the proof is the regularity of Φ_V. This amounts to proving the regular dependence of the solution with respect to the initial data.

The Lie derivative along V, denoted $\mathcal{L}_V$, is a way of differentiating tensor fields along the flow of V. We first define it on scalar functions simply as the action of V on the function

$$\mathcal{L}_V f := V f. \tag{2.18}$$

Then, we extend it to tensor fields by requiring that it satisfies the Leibnitz rule. For a differentiable vector field X and a scalar function f on $\mathcal{M}$, we have:

$$\begin{aligned} (\mathcal{L}_V X)f &= \mathcal{L}_V(Xf) - X\mathcal{L}_V f \\ &= VXf - XVf = [V,X]f. \end{aligned}$$

where $[V,X]$ is the Lie bracket of the two vector fields V and X, i.e. their commutator as differential operators acting on functions. It follows that

$$\mathcal{L}_V X = [V,X]. \tag{2.19}$$

Note that this can be expressed in terms of covariant derivatives as

$$(\mathcal{L}_V X)^a = V^b\nabla_b X^a - X^b\nabla_b V^a = \nabla_V X^a - \nabla_X V^a. \tag{2.20}$$

In particular, the Lie derivative of a differentiable 1-form on $\mathcal{M}$ can be obtained using the fact that for a differentiable vector field X^a, $\omega_a X^a$ is a differentiable scalar function and that we know the Lie derivatives of both vector fields and scalar functions. Indeed, we have on the one hand

$$\mathcal{L}_V\left(\omega_a W^a\right) = \nabla_V\left(\omega_a W^a\right) = W^a\nabla_V\omega_a + \omega_a\nabla_V W^a$$

and on the other hand

$$\begin{aligned} \mathcal{L}_V\left(\omega_a W^a\right) &= W^a\mathcal{L}_V\omega_a + \omega_a\mathcal{L}_V W^a \\ &= W^a\mathcal{L}_V\omega_a + \omega_a\nabla_V W^a - \omega_a\nabla_W V^a. \end{aligned}$$

Putting the two together, we obtain

$$W^a\mathcal{L}_V\omega_a = W^a\nabla_V\omega_a + \omega_a\nabla_W V^a = W^a\left(\nabla_V\omega_a + \omega_b\nabla_a V^b\right).$$

We have therefore proved that

$$\mathcal{L}_V\omega_a = V^b\nabla_b\omega_a + \omega_b\nabla_a V^b \tag{2.21}$$

The formula for a general tensor field is then

$$\begin{aligned}\mathcal{L}_V T^{i_1 \dots i_p}{}_{j_1 \dots j_q} &= V^a \nabla_a T^{i_1 \dots i_p}{}_{j_1 \dots j_q} - T^{a i_2 \dots i_p}{}_{j_1 \dots j_q} \nabla_a V^{i_1} - \dots - T^{i_1 \dots i_{p-1} a}{}_{j_1 \dots j_q} \nabla_a V^{i_p} \\ &\quad + T^{i_1 \dots i_p}{}_{a j_2 \dots j_q} \nabla_{j_1} V^a + \dots + T^{i_1 \dots i_p}{}_{j_1 \dots j_{q-1} a} \nabla_{j_q} V^a . \end{aligned} \tag{2.22}$$

Of particular interest is the expression of the Lie derivative of the metric along a vector field. It is obtained using (2.22) and the fact that the Levi-Civita connection commutes with the metric:

$$\mathcal{L}_V g_{ab} = g_{cb} \nabla_a V^c + g_{ac} \nabla_b V^c = 2 \nabla_{(a} V_{b)} \, . \tag{2.23}$$

Proposition 2.12. *The Lie derivative is independent of the connection, i.e. it can be expressed using any connection, it will remain the same.*

Proof. This is clear for its action on vector fields and scalars. Now given a vector field X and a 1-form ω,

$$\mathcal{L}_V (\omega_a X^a) = \omega_a \mathcal{L}_V X^a + X^a \mathcal{L}_V \omega_a \, ,$$

whence

$$X^a \mathcal{L}_V \omega_a = \mathcal{L}_V (\omega_a X^a) - \omega_a \mathcal{L}_V X^a$$

is the sum of two terms independent of the connection. This extends to all types of tensors by the Leibnitz rule. □

Definition 2.23 (Killing vector)**.** *A Killing vector field on a manifold $\mathcal{M}$ equipped with a metric g is a differentiable vector field K^a on $\mathcal{M}$ such that its flow leaves the metric invariant, i.e. $\mathcal{L}_K g_{ab} = 0$. As a consequence of (2.23), a differentiable vector field K^a on $(\mathcal{M}, g)$ is Killing if and only if K^a satisfies the Killing equation*

$$\nabla_{(a} K_{b)} = 0 \, . \tag{2.24}$$

Definition 2.24 (Stationarity, staticity)**.** *A space-time is said to be stationary if it admits a global timelike Killing vector field. It is said to be static if it admits a global timelike Killing vector field that is orthogonal to a family of spacelike hypersurfaces (equivalently, orthogonal to a Cauchy hypersurface).*

As an example, the symmetry group of Minkowski space-time (preserving the metric, orientation and time-orientation) is the Poincaré group. It is the 10-dimensional group generated by the four Cartesian coordinate translations, the three space rotations and the three boosts or hyperbolic rotations. The infinitesimal generators of these transformations provide the 10 independent Killing vector fields of Minkowski space-time:

translations: ∂_t, ∂_x, ∂_y, ∂_z;

rotations: $x\partial_y - y\partial_x$, $y\partial_z - z\partial_y$, $z\partial_x - x\partial_z$;

boosts: $x\partial_t + t\partial_x$, $y\partial_t + t\partial_y$, $z\partial_t + t\partial_z$, which are sometimes viewed as generating rotations in the planes (it, x), (it, y) and (it, z) (hyperbolic rotations).

2.8 Geodesics

It is a classic notion that the most direct path between two points is the straight line. The notion of straight line however only has a meaning in affine spaces. We of course do not live in an affine space, so this classic image is in fact wrong and even meaningless. It is however true to a very good degree of accuracy provided the two points are not too far from each other (which may mean arbitrarily close to each other if the curvature is arbitrarily large). In an affine space, a useful notion is that of a "freely falling object", i.e. an object that is not accelerated. The trajectories of such objects are of course exactly the straight lines. The advantage is that the notion of an object that is not accelerated can be extended to a general manifold, its trajectory is then a particular type of curve referred to as a geodesic. We have some freedom in the way we define the acceleration, i.e. on how we differentiate the speed vector along the curve. We choose a way of differentiating along the curve that transforms a tensor of a given valence into another tensor of the same valence, it is the so-called absolute derivative

$$\frac{D}{Ds} := \nabla_{\dot{\gamma}(s)}$$

i.e. the covariant derivative along the speed vector.

This provides us with the following definition of a geodesic.

Definition 2.25 (Geodesics)**.** *A geodesic on a space-time $(\mathcal{M}, g)$ is a $\mathcal{C}^2$ curve on $\mathcal{M}$ (i.e. the data of a pair (I, γ) where I is an interval and $\gamma\ : I \to \mathcal{M}$ is a $\mathcal{C}^2$ function such that $\dot{\gamma}(s)$ does not vanish on I) such that its acceleration, defined by $\frac{D}{Ds}\dot{\gamma}(s) = \nabla_{\dot{\gamma}(s)}\dot{\gamma}(s) = 0$. Expressing the covariant derivative in a coordinate basis, this immediately gives the equation of a geodesic*

$$\frac{\mathrm{d}^2\gamma^{\mathbf{a}}}{\mathrm{d}s^2} + \Gamma^{\mathbf{a}}_{\mathbf{bc}}\frac{\mathrm{d}\gamma^{\mathbf{b}}}{\mathrm{d}s}\frac{\mathrm{d}\gamma^{\mathbf{c}}}{\mathrm{d}s} = 0\,.$$

If we consider a differentiable vector field T^a that is propagated parallel along itself, i.e. such that $T^a\nabla_a T^b$ is colinear to T^a its integral curves are geodesics. Indeed, modulo reparametrization, we can assume that $T^a\nabla_a T^b = 0$; the parameters of the integral curves that give a tangent vector field satisfying this are called affine parameters, because if we know one such parameter, all the others are obtained from it by affine transformations.

The geodesic equation is a differential equation whose coefficients are the Christoffel symbols, i.e. involve first order derivatives of the metric. Therefore, the metric needs to be such that its derivative is locally Lipschitz in order to ensure the existence and uniqueness of maximal solutions by the Cauchy-Lipschitz theorem. For a $\mathcal{C}^2$ metric, this is naturally guaranteed.

Remark 2.11. *In euclidian space or Minkowski space-time in cartesian coordinates, the Christoffel symbols all vanish and the geodesics are the straight lines.*

Remark 2.12. *In Riemannian signature, a geodesic between two points can be understood as a length minimizing curve. There is no such property in Lorentzian signature (see figure 2.1).*

The definition of a geodesic entails the existence of a conserved quantity along such a curve. Moreover, any Killing vector field will give another conserved quantity along a geodesic.

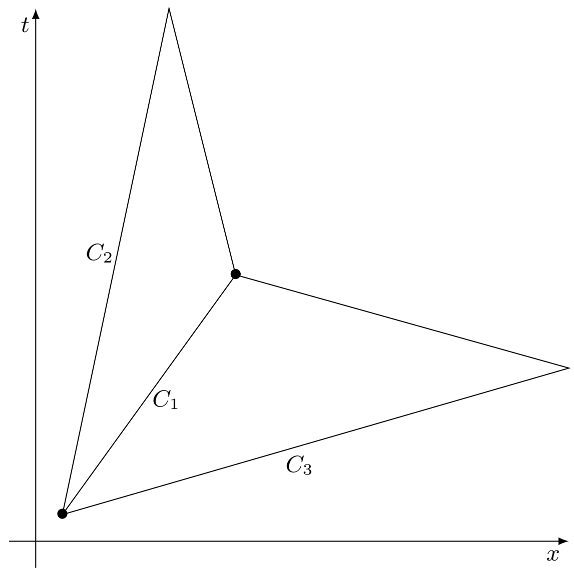

Figure 2.1: In Lorentzian signature, geodesics between two points are not extrema of the arc length. Here we consider curves on Minkowski space-time lying in the (t, x) plane. The curve $C1$ is a geodesic but $C2$ and $C3$ are not. The length of $C2$ is larger than that of $C1$ which is larger than that of $C3$. Also, $C2$ and $C3$ can be continuously deformed to $C1$ whilst retaining the same ordering of lengths.

Proposition 2.13. *Consider a space-time $(\mathcal{M}, g)$ whose metric is $\mathcal{C}^2$ (or has locally Lipschitz first derivative), let γ be a geodesic. Then we have the following two properties.*

1. *The quantity*
$$g(\dot\gamma(s), \dot\gamma(s)) = g_{ab}(\gamma(s))\dot\gamma^a(s)\dot\gamma^b(s)$$
is conserved along the curve.

2. *If K is a Killing vector field on $(\mathcal{M}, g)$ or on an open neighbourhood of γ, then*
$$g(\dot\gamma(s), K) = g_{ab}(\gamma(s))\dot\gamma^a(s)K^b(\gamma(s))$$
is conserved along γ.

Proof. For the first quantity, we have

$$\frac{\mathrm{d}}{\mathrm{d}s} g_{ab}\dot\gamma^a(s)\dot\gamma^b(s) = \left(\nabla_{\dot\gamma(s)} g_{ab}\right)\dot\gamma^a(s)\dot\gamma^b(s) + 2g_{ab}\dot\gamma^a(s)\nabla_{\dot\gamma(s)}\dot\gamma^b(s) = 0$$

since the connection is metric compatible and the curve γ is a geodesic.

Now for K^a a Killing vector field on $(\mathcal{M}, g)$,

$$\frac{\mathrm{d}}{\mathrm{d}s} g_{ab}K^a\dot\gamma^b(s) = \left(\nabla_{\dot\gamma(s)} g_{ab}\right)K^a\dot\gamma^b(s) + g_{ab}\left(\nabla_{\dot\gamma(s)}K^a\right)\dot\gamma^b(s) + g_{ab}K^a\nabla_{\dot\gamma(s)}\dot\gamma^b(s)\,.$$

The first term is zero since the connection is metric compatible and the third since γ is a geodesic. As for the second term, it can be written as

$$\begin{aligned} g_{ab}\dot{\gamma}^b(s)\nabla_{\dot{\gamma}(s)}K^a &= g_{ab}g_{cd}\dot{\gamma}^b(s)\dot{\gamma}^c(s)\nabla^d K^a \\ &= g_{ab}g_{cd}\dot{\gamma}^b(s)\dot{\gamma}^c(s)\nabla^{[d}K^{a]} \text{ since } K^a \text{ is Killing} \\ &= -g_{db}g_{ca}\dot{\gamma}^b(s)\dot{\gamma}^c(s)\nabla^{[d}K^{a]} \\ &= -g_{ca}\dot{\gamma}^c(s)\nabla_{\dot{\gamma}(s)}K^a \\ &= -g_{ab}\dot{\gamma}^b(s)\nabla_{\dot{\gamma}(s)}K^a \text{ by symmetry of } g_{ab}\,. \end{aligned}$$

This concludes the proof. □

The first part of the proposition above has the following consequence.

Corollary 2.1. *Under the assumptions of Proposition 2.13, the tangent vectors $\dot{\gamma}(s)$ and $\dot{\gamma}(t)$ at any two points of the curve γ have the same causal type. In other words, a geodesic that is timelike (resp. spacelike, resp. null) at one point is timelike (resp. spacelike, resp. null) everywhere.*

Some functions referred to as optical functions have the interesting property that their gradients generate geodesics.

Definition 2.26. *An optical function on a space-time $(\mathcal{M}, g)$ is a scalar function u on $\mathcal{M}$ such that its gradient is everywhere null.*

Lemma 2.2. *Let u be an optical function on a space-time $(\mathcal{M}, g)$, then denoting $l^a = \nabla^a u$, the integral curves of l are a congruence of affinely parametrised null geodesics, i.e.*

$$\nabla_l l = 0\,.$$

Proof. The proof is a simple calculation:

$$\begin{aligned} \nabla_l l^b &= \nabla_{\nabla u}\nabla^b u\,, \\ &= \nabla_a u \nabla^a \nabla^b u\,, \\ &= \nabla_a u \nabla^b \nabla^a u \text{ since the connection is torsion-free,} \\ &= \nabla^b\left(\nabla_a u \nabla^a u\right) - \left(\nabla^b \nabla_a u\right)\nabla^a u\,, \\ &= 0 - \nabla_a u \nabla^a \nabla^b u \text{ since } \nabla u \text{ is null and the connection torsion-free,} \\ &= -\nabla_{\nabla u}\nabla^b u\,. \end{aligned}$$

□

Note that for a general null congruence, the more complete Propositions (7.1.60) and (7.1.61) in Penrose and Rindler Vol. 2 [45] state that the following three properties are equivalent :

1. it is hypersurface-orthogonal;
2. it is hypersurface-forming;
3. it is geodetic and twist-free.

2.9 Conformal transformations and classes

A conformal transformation consists of multiplying the space-time metric by a positive function. The new rescaled metric has the same light-cone structure as the original one, but it may have different asymptotic properties. The whole family of metrics generated from the original one by such transformations is called its conformal class.

Definition 2.27 (Conformal class)**.** *Consider a space-time* $(\mathcal{M}, g)$*. We say that a metric* $\hat{g}$ *on* $\mathcal{M}$ *is conformally equivalent to* g *if there exists a positive nowhere vanishing smooth function* Ω *on* $\mathcal{M}$ *such that* $\hat{g} = \Omega^2 g$*. We also say that* $\hat{g}$ *is a conformal rescaling of* g*. The conformal class* $[g]$ *of* g *is the set of all metrics on* $\mathcal{M}$ *that are conformally equivalent to* g*.*

The property that conformally equivalent metrics have the same lightcones is in fact a characterisation of conformal equivalence.

Theorem 2.6. *On a manifold* $\mathcal{M}$*, two Lorentzian metrics* g *and* h *are conformally equivalent if and only if they have the same lightcones.*

Proof. One of the two implications is trivial. Let us prove the other. Consider two metrics g and h that have the same lightcones. Let p be any given point in $\mathcal{M}$ and e_0 be a unit timelike vector for the metric g, i.e. $g(e_0, e_0) = 1$. We consider V the orthogonal subspace to e_0 in $T_p\mathcal{M}$ and S its unit 2-sphere, which is the set of $v \in V$ such that $g(v, v) = -1$. Let $e_1 \in S$, then the vectors $e_0 \pm e_1$ are both null for g and therefore also for h. Hence we have

$$\begin{aligned} h(e_0 - e_1, e_0 - e_1) &= h(e_0, e_0) + h(e_1, e_1) - 2h(e_0, e_1) = 0\,, \\ h(e_0 + e_1, e_0 + e_1) &= h(e_0, e_0) + h(e_1, e_1) + 2h(e_0, e_1) = 0\,. \end{aligned}$$

It follows that $h(e_0, e_1) = 0$ and $h(e_1, e_1) = -h(e_0, e_0)$. Moreover, the metrics g and h have the same timelike vectors, since the timelike vectors for a Lorentzian metric at a given point are exactly given by the sums of pairs of null vectors belonging to the same component of the light-cone. Hence, $h(e_0, e_0) > 0$. It follows that the quadratic forms g and $\frac{g(e_0,e_0)}{h(e_0,e_0)}h$ coincide on e_0 and on S. By homogeneity and the polarisation identity, the two metrics coincice on V and the line spanned by e_0, but since the two are orthogonal for both g and h, then the two metrics coincide on $T_p\mathcal{M}$. This is true at any point, so $h = \Omega^2 g$ where $\Omega = \sqrt{h(e_0, e_0)}$. $\square$

Under a conformal rescaling, the connection changes in a rather simple way. Recall the expression of the Christoffel symbols in a coordinate basis

$$\Gamma_{\mathbf{ab}}{}^{\mathbf{c}} = \frac{1}{2} g^{\mathbf{cd}} \left(\partial_{\mathbf{a}} g_{\mathbf{bd}} + \partial_{\mathbf{b}} g_{\mathbf{ad}} - \partial_{\mathbf{d}} g_{\mathbf{ab}} \right) ,$$

then for a metric $\hat{g} = \Omega^2 g$, $\Omega > 0$ on $\mathcal{M}$ and smooth, we have the Christoffel symbols

$$\begin{aligned} \hat{\Gamma}_{\mathbf{ab}}{}^{\mathbf{c}} &= \frac{1}{2} \hat{g}^{\mathbf{cd}} \left(\partial_{\mathbf{a}} \hat{g}_{\mathbf{bd}} + \partial_{\mathbf{b}} \hat{g}_{\mathbf{ad}} - \partial_{\mathbf{d}} \hat{g}_{\mathbf{ab}} \right) \\ &= \Gamma_{\mathbf{ab}}{}^{\mathbf{c}} = \frac{1}{2} \Omega^{-2} g^{\mathbf{cd}} \left(\partial_{\mathbf{a}}(\Omega^2 g_{\mathbf{bd}}) + \partial_{\mathbf{b}}(\Omega^2 g_{\mathbf{ad}}) - \partial_{\mathbf{d}}(\Omega^2 g_{\mathbf{ab}}) \right) \\ &= \Gamma_{\mathbf{ab}}{}^{\mathbf{c}} + 2 g^{\mathbf{a}}_{(\mathbf{b}} \nabla_{\mathbf{c})} \ln \Omega - g_{\mathbf{bc}} \nabla^{\mathbf{a}} \ln \Omega\,. \end{aligned}$$

We denote by $C_{bc}{}^a$ the difference between the two Christoffel symbols ; note that this is a true tensor field (contrary to the Christoffel symbols) expressing the difference between the Levi-Civita connections ∇ and $\hat{\nabla}$ of the two metrics g and $\hat{g}$:

$$C_{\mathbf{bc}}{}^{\mathbf{a}} := \hat{\Gamma}_{\mathbf{ab}}{}^{\mathbf{c}} - \Gamma_{\mathbf{ab}}{}^{\mathbf{c}}\,,\ C_{bc}{}^a = 2g^a_{(b}\nabla_{c)}\ln\Omega - g_{bc}\nabla^a\ln\Omega\,,\ C_{bc}{}^a = C_{(bc)}{}^a\,. \tag{2.25}$$

This tensor can be used to express the difference between the Riemann tensors for g and $\hat{g}$: first we write

$$\begin{aligned}\hat{\nabla}_a\hat{\nabla}_b\omega_c \;=\;& \nabla_a(\nabla_b\omega_c - C_{bc}{}^d\omega_d)\\ & -C_{ab}{}^e(\nabla_e\omega_c - C_{ec}{}^d\omega_d)\\ & -C_{ac}{}^e(\nabla_b\omega_e - C_{be}{}^d\omega^d)\,.\end{aligned}$$

This gives

$$\begin{aligned}-\hat{R}_{abc}{}^d\omega_d \;=\;& (\hat{\nabla}_a\hat{\nabla}_b - \hat{\nabla}_b\hat{\nabla}_a)\omega_c\\ =\;& -R_{abc}{}^d\omega_d - 2\nabla_{[a}C_{b]c}{}^d\omega_d\\ & -0\\ & -2C_{c[a}{}^e\nabla_{b]}\omega_e + 2C_{c[a}{}^eC_{b]e}{}^d\omega_d\,,\end{aligned}$$

whence

$$\hat{R}_{abc}{}^d - R_{abc}{}^d = 2(\nabla_{[a}C_{b]c}{}^d) - 2C_{c[a}{}^eC_{b]e}{}^d\,. \tag{2.26}$$

By taking the trace of (2.26), we can obtain the relation between $\mathrm{Scal}_{\hat{g}}$ and Scal_g. The trace of the second covariant derivative will appear ; it is referred to as the d'Alembertian

Definition 2.28 (d'Alembertian)**.** *On a given space-time* $(\mathcal{M}, g)$ *the d'Alembertian operator is defined by*

$$\Box_g = \nabla_a\nabla^a\,. \tag{2.27}$$

It is easy to check that in a local coordinate basis, its expression is given by

$$\Box_g = \frac{1}{\sqrt{|g|}}\partial_{\mathbf{a}}(\sqrt{|g|}g^{\mathbf{ab}}\partial_{\mathbf{b}})\,. \tag{2.28}$$

Note that one must be careful. Taking the trace of (2.26) means using a metric to raise an index, $\hat{g}$ for $\hat{R}$ and g for R. The index d is already raised, so we just need to contract with $\hat{g}^b_d = g^b_d$, but we also need to contract with $\hat{g}^{ac} = \Omega^{-2}g^{ac}$. We get

$$\mathrm{Scal}_{\hat{g}} = \hat{R}_{ab}{}^{ab} = \Omega^{-2}(R_{ab}{}^{ab} + g^{ac}2(\nabla_{[a}C_{b]c}{}^b) - 2g^{ac}C_{c[a}{}^eC_{b]e}{}^b)$$

and after a long but straightforward calculation, provided we are careful and do not make mistakes, we find the result[3].

[3] A more detailed study of the modification of the different parts of the curvature under conformal rescalings is given in [59], with a different sign convention for Lorentzian metrics though, so some conversions are necessary, and in the formalism of Weyl spinors in [45].

Theorem 2.7. *Consider a space-time $(\mathcal{M}, g)$ and a metric $\hat{g}$ in the conformal class of g with conformal factor Ω, i.e. $\hat{g} = \Omega^2 g$, then*

$$\mathrm{Scal}_{\hat{g}} = \Omega^{-2}\mathrm{Scal}_g + 6\Omega^{-3}\Box_g\Omega\,.$$

The Weyl tensor C_{abcd} has conformal weight 2, i.e. under the conformal rescaling $\hat{g} = \Omega^2 g$, it changes as follows

$$\hat{C}_{abcd} = \Omega^2 C_{abcd}$$

and if we raise one index, we have

$$\hat{C}_{abc}{}^{d} = C_{abc}{}^{d}.$$

Definition of a conformal Killing vector field. Conformal Killing equation. Conformal Killing vector fields of Minkowski space-time.

2.10 Spinors

Chapter 3

Classic space-times and their conformal compactification

Conformal rescalings can be used in some cases to bring infinity to a finite distance and to extend the original manifold by adding a boundary to it that describes infinity for the original metric. This is called a conformal compactification. We present a simple example of such a construction, well-known from undergraduate geometry courses but not always presented from the point of view of metric rescaling : the stereographic projection from the North pole of the unit 2-sphere to its equatorial plane. The formula relating the points on the sphere in spherical coordinates (θ, φ) to those on the plane in polar coordinates (r, ψ) are

$$\psi = \varphi \,, \ \theta = 2\arctan(1/r) \,.$$

Let us write the euclidean metric on the 2-sphere in terms of the variables r and ψ :

$$\begin{aligned}
e_{S^2} &= \mathrm{d}\theta^2 + \sin^2\theta \mathrm{d}\varphi^2 \\
&= \left(2\frac{-1}{r^2}\frac{1}{1+\frac{1}{r^2}} \right)^2 \mathrm{d}r^2 + \sin^2\theta \mathrm{d}\psi^2 \\
&= \frac{4}{(1+r^2)^2}\mathrm{d}r^2 + \frac{4r^2}{(1+r^2)^2}\mathrm{d}\psi^2 \,, \text{ using the identity } \sin t = \frac{2\tan(t/2)}{1+\tan^2(t/2)} \,, \\
&= \frac{4}{(1+r^2)^2}\left(\mathrm{d}r^2 + r^2\mathrm{d}\psi^2\right) \\
&= \frac{4}{(1+r^2)^2} e_{\mathbb{R}^2} \,,
\end{aligned}$$

where $e_{\mathbb{R}^2}$ is the euclidean metric on $\mathbb{R}^2$. So we see that by multiplying the euclidean metric on $\mathbb{R}^2$ by Ω^2, where

$$\Omega = \frac{2}{1+r^2} \,,$$

we turn it into the Euclidean metric on S^2. The thus rescaled metric is defined only away from the North pole, but it can be extended analytically to the whole 2-sphere. This is the conformal compactification of $\mathbb{R}^2$, which is the "metric" version of the usual Alexandroff compactification. It is called conformal because, since the metric is merely multiplied by a positive function, the angles, as measured using the metric, are unchanged.

Can we perform a compactification of a space-time by rescaling its metric, just as we did with the euclidean metric on $\mathbb{R}^2$?

3.1 Minkowski space-time

3.1.1 The conformal embedding in the Einstein cylinder

The contents of this section, and much more, can be found in [44]. The Minkowski metric in spherical coordinates is expressed as

$$\eta = \mathrm{d}t^2 - \mathrm{d}r^2 - r^2 \mathrm{d}\omega^2 \,,\ \mathrm{d}\omega^2 = \mathrm{d}\theta^2 + \sin^2\theta \, \mathrm{d}\varphi^2 \,.$$

We choose the advanced and retarded coordinates

$$u = t - r \,,\ v = t + r \,. \tag{3.1}$$

The metric η in terms of these new coordinates takes the form

$$\eta = \mathrm{d}u\mathrm{d}v - \frac{(v-u)^2}{4}\mathrm{d}\omega^2 \,.$$

We now introduce new null coordinates that allow us to describe the whole of Minkowski space as a bounded domain :

$$p = \arctan u \,,\ q = \arctan v \,. \tag{3.2}$$

We obtain

$$\eta = (1+u^2)(1+v^2)\mathrm{d}p\,\mathrm{d}q - \frac{(v-u)^2}{4}\mathrm{d}\omega^2 \,.$$

Finally coming back to time and space coordinates as follows,

$$\begin{aligned} \tau &= p + q = \arctan(t-r) + \arctan(t+r) \,, \\ \zeta &= q - p = \arctan(t+r) - \arctan(t-r) \,, \end{aligned} \tag{3.3}$$

we get

$$\eta = \frac{(1+u^2)(1+v^2)}{4}\left(\mathrm{d}\tau^2 - \mathrm{d}\zeta^2\right) - \frac{(v-u)^2}{4}\mathrm{d}\omega^2 \,.$$

Remark 3.1. *It can be useful to express partial derivatives in the (t, r, ω) coordinate systems in terms of those in (τ, ζ, ω) coordinates and vice versa. We have the following relations*

$$\begin{aligned} \partial_t + \partial_r &= \frac{2}{1+(t+r)^2}(\partial_\tau + \partial_\zeta), \\ \partial_t - \partial_r &= \frac{2}{1+(t-r)^2}(\partial_\tau - \partial_\zeta), \end{aligned}$$

which yield

$$\partial_t = \frac{1}{1+(t+r)^2}(\partial_\tau + \partial_\zeta) + \frac{1}{1+(t-r)^2}(\partial_\tau - \partial_\zeta)\,, \tag{3.4}$$

$$\partial_r = \frac{1}{1+(t+r)^2}(\partial_\tau + \partial_\zeta) - \frac{1}{1+(t-r)^2}(\partial_\tau - \partial_\zeta)\,. \tag{3.5}$$

These relations shall be important in particular at $t = 0$ (which corresponds to $\tau = 0$) when considering the rescaling of initial data through a conformal transformation:

$$\partial_\tau|_{\tau=0} = \frac{1+r^2}{2}\partial_t|_{t=0}\,,\ \ \partial_\zeta|_{\tau=0} = \frac{1+r^2}{2}\partial_r|_{t=0}\,. \tag{3.6}$$

Choosing the conformal factor

$$\Omega = \sqrt{\frac{4}{(1+u^2)(1+v^2)}} = \sqrt{\frac{4}{(1+\tan^2 p)(1+\tan^2 q)}} = 2\cos p \cos q\,, \tag{3.7}$$

we obtain the rescaled metric

$$\begin{aligned}
\mathfrak{e} \ &:= \ \Omega^2\eta = \mathrm{d}\tau^2 - \mathrm{d}\zeta^2 - \frac{(v-u)^2}{(1+u^2)(1+v^2)}\mathrm{d}\omega^2 \\
&= \ \mathrm{d}\tau^2 - \mathrm{d}\zeta^2 - \left((\tan q - \tan p)\cos p \cos q\right)^2 \mathrm{d}\omega^2 \\
&= \ \mathrm{d}\tau^2 - \mathrm{d}\zeta^2 - \left(\sin q \cos p - \sin p \cos q\right)^2 \mathrm{d}\omega^2 \\
&= \ \mathrm{d}\tau^2 - \mathrm{d}\zeta^2 - \left(\sin(q-p)\right)^2 \mathrm{d}\omega^2 \\
&= \ \mathrm{d}\tau^2 - \mathrm{d}\zeta^2 - \left(\sin \zeta\right)^2 \mathrm{d}\omega^2 \\
&= \ \mathrm{d}\tau^2 - \sigma^2_{S^3}\,,
\end{aligned}$$

where $\sigma^2_{S^3}$ is the euclidian metric on the 3-sphere. Minkowski space-time is now described as the diamond

$$\mathbb{M} = \{|\tau| + \zeta \leq \pi\,,\ \zeta \geq 0\,,\ \omega \in S^2\}\,.$$

The metric $\mathfrak{e}$ is the Einstein metric, it extends analytically to the whole Einstein cylinder

$$\mathfrak{E} = \mathbb{R}_\tau \times S^3_{\zeta,\theta,\varphi}.$$

The full conformal boundary of Minkowski space can be defined in this framework. It is described as

$$\partial\mathbb{M} = \{|\tau| + \zeta = \pi\,,\ \zeta \geq 0\,, \omega \in S^2\}\,.$$

Several parts can be distinguished (see Figure 3.1).

- Future and past null infinities :

$$\begin{aligned}
\mathscr{I}^+ \ &= \ \left\{(\tau\,,\ \zeta\,,\ \omega)\,;\ \tau + \zeta = \pi\,,\ \zeta \in]0,\pi[\,,\ \omega \in S^2\right\}\,, \\
\mathscr{I}^- \ &= \ \left\{(\tau\,,\ \zeta\,,\ \omega)\,;\ \zeta - \tau = \pi\,,\ \zeta \in]0,\pi[\,,\ \omega \in S^2\right\}\,.
\end{aligned}$$

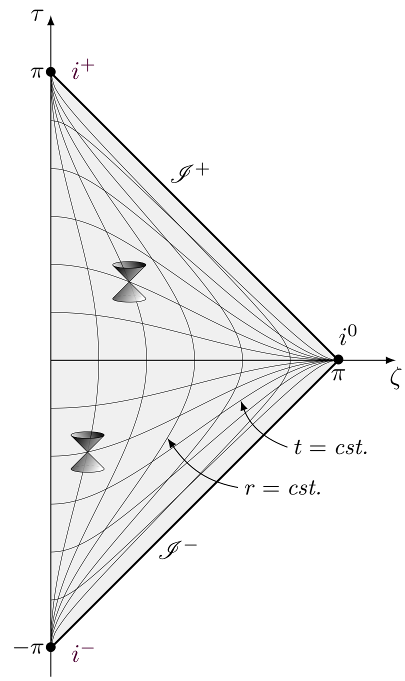

Figure 3.1: Penrose diagram of compactified Minkowski space-time

- Future and past timelike infinities :

$$i^{\pm} = \left\{ (\tau = \pm\pi\,,\ \zeta = 0\,,\ \omega)\,;\ \omega \in S^2 \right\}\,.$$

They are smooth points for $\mathfrak{e}$ (2-spheres whose area is zero because they correspond to $\zeta = 0$).

- Spacelike infinity :

$$i^{0} = \left\{ (\tau = 0\,,\ \zeta = \pi\,,\ \omega)\,;\ \omega \in S^2 \right\}\,.$$

It is also a smooth point for $\mathfrak{e}$.

Proposition 3.1. *The hypersurfaces $\mathscr{I}^{\pm}$ are smooth null hypersurfaces for $\mathfrak{e}$. Their null generators are respectively the vector fields*

$$\partial_\tau - \partial_\zeta \text{ for } \mathscr{I}^{+} \text{ and } \partial_\tau + \partial_\zeta \text{ for } \mathscr{I}^{-}\,.$$

Proof. They are clearly smooth hypersurfaces. The vector fields $\partial_\tau - \partial_\zeta$ and $\partial_\tau + \partial_\zeta$ are null and tangent respectively to $\mathscr{I}^{+}$ and $\mathscr{I}^{-}$. They are orthogonal to the two other generators of $\mathscr{I}^{\pm}$: ∂_θ and ∂_φ. They are therefore normal to $\mathscr{I}^{+}$ and $\mathscr{I}^{-}$ respectively. This proves the proposition. □

We defined compactified Minkowski space-time as

$$\overline{\mathbb{M}} := \mathbb{M} \cup \partial\mathbb{M}\,. \tag{3.8}$$

An interesting property of the conformal factor Ω is that its gradient is everywhere smooth on $\overline{\mathbb{M}}$, is tangent to $\mathscr{I}$ and vanishes at $i^\pm$ and i^0 (it also vanishes at the origin $\{t = r = 0\}$), as can be seen from its explicit expression in (τ, ζ, ω) coordinates:

$$\begin{aligned}
\nabla_{\mathfrak{e}}\Omega &= \mathfrak{e}^{-1}(\mathrm{d}\Omega) \\
&= \mathfrak{e}^{-1}\left(-\sin\tau \mathrm{d}\tau - \sin\zeta \mathrm{d}\zeta\right) \\
&= -\sin\tau \frac{\partial}{\partial\tau} - \sin\zeta \frac{\partial}{\partial\zeta} \,.
\end{aligned}$$

We also observe that the coordinate vector fields ∂_t and ∂_r extend smoothly to the boundary and we have

$$\partial_t|_{\mathscr{I}^+} = -\partial_r|_{\mathscr{I}^+} = \frac{1}{1+u^2}\left(\partial_\tau - \partial_\zeta\right) = \frac{1}{1+\tan\left(\tau - \frac{\pi}{2}\right)^2}\left(\partial_\tau - \partial_\zeta\right), \tag{3.9}$$

$$\partial_t|_{\mathscr{I}^-} = \partial_r|_{\mathscr{I}^-} = \frac{1}{1+v^2}\left(\partial_\tau + \partial_\zeta\right) = \frac{1}{1+\tan\left(\tau + \frac{\pi}{2}\right)^2}\left(\partial_\tau + \partial_\zeta\right). \tag{3.10}$$

In particular, ∂_t and ∂_r both vanish at $i^\pm$ and i^0. The different parts of the conformal boundary can be realised as endpoints of different families of geodesics. Recall that the geodesics of Minkowski space-time are the straight lines (see Remark 2.11).

- **Timelike infinities are the endpoints in the future and the past of all timelike geodesics.**
  A timelike geodesic is a curve described in cartesian coordinates as

  $$\gamma(t) = (t + t_0, tv + x)$$

  where $t_0 \in \mathbb{R}$, $x, v \in \mathbb{R}^3$ are fixed parameters such that $\|v\| < 1$ and $\langle x, v\rangle = 0$ (we denote by $\langle ., .\rangle$ and $\|.\|$ the usual inner product and associated norm on $\mathbb{R}^3$). Along such a curve, as $t \to +\infty$ we have

  $$\begin{aligned}
  \tau &= \arctan\left(t + t_0 + \|tv + x\|\right) + \arctan\left(t + t_0 - \|tv + x\|\right) \to \pi \,, \\
  \zeta &= \arctan\left(t + t_0 + \|tv + x\|\right) - \arctan\left(t + t_0 - \|tv + x\|\right) \to 0 \,.
  \end{aligned}$$

- **Spacelike infinity is the enpoint of all spacelike geodesics.**
  A spacelike geodesic can be described as

  $$\gamma(t) = (t + t_0, tv + x) \text{ where } t_0 \in \mathbb{R}\,,\ \|v\| > 1 \text{ and } \langle x, v\rangle = 0\,.$$

  Along such a curve, as $t \to \pm\infty$ we have

  $$\begin{aligned}
  \tau &= \arctan\left(t + t_0 + \|tv + x\|\right) + \arctan\left(t + t_0 - \|tv + x\|\right) \to 0 \,, \\
  \zeta &= \arctan\left(t + t_0 + \|tv + x\|\right) - \arctan\left(t + t_0 - \|tv + x\|\right) \to \pi \,.
  \end{aligned}$$

- **Null infinities are the endpoints in the future and the past of all null geodesics.** A future oriented null geodesic can be described as

$$\gamma(t) = (t+t_0, tv+x) \text{ where } t_0 \in \mathbb{R}\,,\ \|v\| = 1 \text{ and } \langle x, v\rangle = 0\,. \tag{3.11}$$

  Along such a curve, we have

$$\begin{aligned} \tau &= \arctan\left(t+t_0+\|tv+x\|\right) + \arctan\left(t+t_0-\|tv+x\|\right)\,, \\ \zeta &= \arctan\left(t+t_0+\|tv+x\|\right) - \arctan\left(t+t_0-\|tv+x\|\right)\,. \end{aligned}$$

  Since

$$\left\| v + \frac{1}{t}x \right\|^2 = 1 + \frac{1}{t^2}\|x\|^2\,,$$

  it follows that

$$\|tv+x\| = t + O\left(\frac{1}{t}\right)\,,$$

  whence as $t \to +\infty$,

$$t+t_0-\|tv+x\| \to t_0\,.$$

  It follows that as $t \to +\infty$,

$$\tau \to \frac{\pi}{2} + \arctan(t_0),\ \zeta \to \frac{\pi}{2} - \arctan(t_0)$$

  and

$$\omega = \frac{tv+x}{\|tv+x\|} \to \frac{v}{\|v\|} = v\,.$$

  As $t \to -\infty$, we have

$$\tau \to -\frac{\pi}{2} + \arctan(t_0) \text{ and } \zeta \to \frac{\pi}{2} + \arctan(t_0)$$

  and

$$\omega = \frac{tv+x}{\|tv+x\|} \to \frac{-v}{\|v\|} = -v\,.$$

  The end-point of γ as $t \to -\infty$ is the point on $\mathscr{I}^-$

$$(\zeta_0 - \pi, \zeta_0, \omega_0) \text{ with } \zeta_0 = \frac{\pi}{2} + \arctan(t_0) \text{ and } \omega_0 = -v \tag{3.12}$$

  and as $t \to +\infty$ it is the point in $\mathscr{I}^+$

$$(\zeta_0, \pi - \zeta_0, -\omega_0)\,. \tag{3.13}$$

Note that if ω_0 has spherical coordinates (θ_0, φ_0), then $-\omega_0$ has coordinates $(\pi - \theta_0, \pi + \varphi_0)$; the point $(\pi - \zeta_0, -\omega_0)$ is the point antipodal to (ζ_0, ω_0) on S^3.

Remark 3.2. *The terminology "null infinities" comes from the fact that $\mathscr{I}^\pm$ are the end points in the future and the past of the null geodesics of Minkowski space-time. The fact that $\mathscr{I}^\pm$ are null hypersurfaces of the compactified metric is an independent property. For example, on de Sitter space-time, null infinities are spacelike hypersurfaces.*

We see from (3.12) and (3.13) that the end points of the curve (3.11) depend only on t_0 and v. The family of null geodesics (3.11) for fixed t_0 and v forms a null hyperplane on $\mathbb{M}$ whose equation is given by

$$\mathcal{N} = \{p \in \mathbb{M}\,,\ \eta_{ab} p^a V^b = t_0\} \tag{3.14}$$

where $V = (1, v)$; note that V is null and future-oriented. The light-cone from the point $(\zeta_0 - \pi, \zeta_0, \omega_0)$ on $\mathscr{I}^-$ is made of the null hyperplane (3.14) and of one light-ray that is tangent to $\mathscr{I}^-$ until it reaches i^0 and then continues inside $\mathscr{I}^+$; its equation in $\mathscr{I}^-$ is

$$\gamma(\zeta) = (\zeta - \pi, \zeta, \omega_0)\,,\ \zeta_0 \leq \zeta \leq \pi \tag{3.15}$$

and once it has crossed i^0 the equation becomes

$$\gamma(\zeta) = (\pi - \zeta, \zeta, -\omega_0)\ ,\ \text{with } \zeta \text{ decreasing from } \pi\,; \tag{3.16}$$

there it will meet the future focusing point of $\mathcal{N}$: $(\zeta_0, \pi - \zeta_0, -\omega_0)$. This is a remarkable property of compactified Minkowski space-time: the lightcones from points on $\mathscr{I}^-$ refocus at the antipodal point on $\mathscr{I}^+$. The same is true for the lightcones of i^- and i^0. Note that the following theorem can be found in a similar form in Spinors and space-times [45] Vol. 2 Section 9.2.

Theorem 3.1. *Consider any point $p_0 = (\zeta_0 - \pi, \zeta_0, \omega_0) \in \mathscr{I}^- \cup \{i^-\} \cup \{i^0\}$. Then the future lightcone of p_0 for the metric $\mathfrak{e}$ refocuses on $\mathscr{I}^+ \cup \{i_0\} \cup \{i^+\}$ at the point $p_1 = (\zeta_0, \pi - \zeta_0, -\omega_0)$ where in spherical coordinates $\omega_0 = (\theta_0, \varphi_0)$, $-\omega_0 = (\pi - \theta_0, \pi + \varphi_0)$.*

The scalar curvature of $\mathfrak{e}$ can be calculated easily using the result of Theorem 2.7:

$$\frac{1}{6}\mathrm{Scal}_{\mathfrak{e}} = \Omega^{-3}\Box_\eta \Omega = 1\,. \tag{3.17}$$

The Killing vectors for $\mathfrak{e}$ are $\frac{\partial}{\partial \tau}$ and the generators of the rotation group $SO(4)$ and the symmetry group of the Einstein cylinder is $\mathbb{R} \times SO(4)$.

3.1.2 A less complete compactification

We can also perform an incomplete compactification of Minkowski space-time for which we only construct null infinities. One may wonder what the point is when we can have the full compactification we just described. This incomplete compactification is interesting for two reasons at least. First, the conformal factor is simpler, does not decay in timelike directions and decays less in spacelike directions ; this will have the advantage of allowing larger classes of initial data in some of our studies of the asymptotic behaviour of solutions to field equations. Second, this compactification can be performed essentially identically on many asymptotically flat space-times, among which Schwarzschild, Reissner-Nordström, Kerr, Kerr-Newman ; the complete compactification however, fails for all black hole space-times and in fact as soon as the universe contains energy.

We start again from the expression of the Minkowski metric in spherical coordinates

$$\eta = \mathrm{d}t^2 - \mathrm{d}r^2 - r^2 \mathrm{d}\omega^2\,,\ \mathrm{d}\omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\, \mathrm{d}\varphi^2\,,$$

and we replace the time variable t by the retarded (resp. advanced) null variable $u = t - r$ (resp. $v = t + r$) as above. In the (u, r, θ, φ) coordinates, the metric reads

$$\eta = \mathrm{d}u^2 + 2\mathrm{d}u\mathrm{d}r - r^2\mathrm{d}\omega^2$$

and in (v, r, θ, φ) coordinates,we have

$$\eta = \mathrm{d}v^2 - 2\mathrm{d}v\mathrm{d}r - r^2\mathrm{d}\omega^2 \,.$$

Then we invert the coordinate r by putting $R = 1/r$ and we rescale the metric by R^2. We obtain

$$\begin{aligned} \tilde{\eta} := R^2\eta &= \frac{1}{r^2}\left(\mathrm{d}t^2 - \mathrm{d}r^2\right) - \mathrm{d}\omega^2 \,, && (3.18)\\ &= R^2\mathrm{d}u^2 - 2\mathrm{d}u\mathrm{d}R - \mathrm{d}\omega^2 \,, && (3.19)\\ &= R^2\mathrm{d}v^2 + 2\mathrm{d}v\mathrm{d}R - \mathrm{d}\omega^2 \,. && (3.20)\end{aligned}$$

Both expressions of $\tilde{\eta}$ extend analytically to $R = 0$ but the locus $\{R = 0\}$ is different for (3.19) and (3.20). In the first case, $\{R = 0\} = \mathbb{R}_u \times \{0\}_R \times S^2_\omega$ is the set of end-points of the lines of constant u and ω, i.e. of outgoing radial null geodesics, as $r \to +\infty$; this is therefore $\mathscr{I}^+$. In the second case, the lines of constant v and ω are the incoming radial null geodesics and the boundary $\{R = 0\} = \mathbb{R}_v \times \{0\}_R \times S^2_\omega$ is consequently $\mathscr{I}^-$. We denote by $\tilde{\mathbb{M}}$ the compactified manifold that we obtain with this choice of conformal factor. We have

$$\tilde{\mathbb{M}} = \mathbb{M} \cup \mathscr{I}^{\pm} \,. \tag{3.21}$$

The scalar curvature of the metric $\tilde{\eta}$ can be calculated using Theorem 2.7

$$\mathrm{Scal}_{\tilde{\eta}} = R^{-3}\Box_\eta R = r^3\left(\frac{\partial^2}{\partial r^2} + \frac{2}{r}\frac{\partial}{\partial r}\right)\frac{1}{r} = 0 \,. \tag{3.22}$$

The Killing vectors for the metric $\tilde{\eta}$ are first the generators of rotations. There is also a causal Killing vector field given by ∂_u in the expression (3.19), by ∂_v for (3.20). This vector field that we denote by T^a is timelike in the bulk of the space-time and coincides with ∂_t in cartesian or spherical coordinates – which is natural since the conformal factor is independent of t – and extends to null infinities as their future oriented null generator. The metric $\tilde{\eta}$ has another Killing vector field referred to as the Morawetz vector field. It was first used by Kathleen Morawetz in 1962 [38] to establish the decay properties of solutions to the wave equation[1]. In spherical coordinates, it is given by

$$T = (r^2 + t^2)\partial_t + 2tr\partial_r \,.$$

It has the simplest expression in the coordinate system (u, v, ω)

$$T = u^2\partial_u + v^2\partial_v$$

[1] Cathleen Morawetz used multipliers in energy estimates to establish decay properties for the wave equation on flat space-time. Her multipliers can be understood as vector fields acting on the solution with an additional zero order term. In [37], she used the generator of dilations $r\partial_r + t\partial_t$ and in [38] she introduced the vector field that now bears her name. The best place to read about this vector field is in the third appendix of the book by Lax and Phillips [31], written by Morawetz. In there, she explains its construction as the image of the timelike Killing vector ∂_t through a light-cone inversion or Kelvin transform.

and in (u, R, ω) coordinates it is given by

$$T = u^2 \partial_u - 2(1 + uR)\partial_R \,.$$

The Morawetz vector field is timelike everywhere except on the lightcone of the origin where it is null; it is future oriented and vanishes only at the origin. It is interesting to note that the Killing vector $K = \partial_\tau$ on the fully compactified Minkowski spacetime, corresponding to the time translation on the Einstein cylinder, is simply a half of the sum of T and the timelike Killing vector $\kappa = \partial_t$:

$$2K = \kappa + T.$$

3.2 The Schwarzschild metric

The Schwarzschild metric is an exact solution to the Einstein vacuum equations describing the gravitational field of a static point mass. It was discovered by Karl Schwarzschild in 1916 (see the original paper [51] or its English translation from 2003 [52]) and is the first non trivial solution to Eintein's equations. In his third paper on general relativity in November 1915 [16], Einstein derived the advance of the perihelion of Mercury's orbit using perturbative methods. Schwarzschild, in December 1915, wrote in a letter to Einstein that he had reworked his proof using an exact expression for the gravitational field of a point mass instead of the original perturbative approach. Schwarzschild subsequently sent a manuscript to Einstein that the latter presented to the Prussian academy in January 1916 and the paper was published a month later. In the coordinates in which it is usually expressed, justly called the Schwarzschild coordinates since they were first introduced in the last equation in [51], Schwarzschild's space-time appears to have a singularity at a sphere whose radius is proportional to the mass of the central body. This was certainly considered as an inconvenience but in 1923, Birkhoff's theorem [5] established that the Schwarzschild metric also describes the gravitational field around a static spherically symmetric extended massive body. In concrete physical situations, the Schwarzschild radius would be deeply buried within the star in which the Einstein vaccum equations would not hold, so the associated singularity was not considered as a serious worry by astronomers. However, some people still tried to clarify its true geometrical nature, if only as an abstract exercice. Eddington was the first to introduce the coordinates that would allow this, but did not use them for this purpose. It was Lemaître [32] who first understood the nature of the event horizon, which was then worked out in details by Synge [54] and rediscovered by Finkelstein in 1958 [17]. The coordinate system that allowed to realise the horizon as a smooth null hypersurface is now called the Eddington-Finkelstein coordinates. In 1960, Kruskal [30] and Szekeres [55] completed the picture independently and built the maximal analytic extension of the Schwarzschild metric.

The Schwarzschild space-time is a reference model for all asymptotically flat universes containing energy/matter. The metric describing any such universe, when restricted to the leaves of a foliation by asymptotically flat spacelike hypersurfaces, is generically a short-range perturbation (i.e. a perturbation in $1/r^2$, r being for example the geodesic distance to a given point on the slice) of the Schwarzschild metric.

Asymptotically simple space-times are an attempt, due to Roger Penrose, at defining generic cosmological models of asymptotically flat space-times. A special class of asymptotically simple space-times, which will be of particular interest to us, coincides with Schwarzschild's space-time in a neighbourhood of infinity.

The Schwarzschild metric is expressed in Schwarzschild coordinates (t, r, ω) as

$$g = F(r)\mathrm{d}t^2 - F(r)^{-1}\mathrm{d}r^2 - r^2\mathrm{d}\omega^2\,,\ \mathrm{d}\omega^2 = \mathrm{d}\theta^2 + \sin^2\theta\mathrm{d}^2\varphi\,,\ F(r) = 1 - \frac{2M}{r}\,, \tag{3.23}$$

on $\mathbb{R}_t \times]0, +\infty[_r \times S^2_\omega$, where m is the mass of the black hole and $\mathrm{d}\omega^2$ is the euclidian metric on the 2-sphere. Expressed in the form (3.23), this metric appears to have two singularities corresponding to $r = 2M$ and $r = 0$. The sphere $\{r = 2M\}$, referred to as the event horizon, is merely a coordinate singularity, the metric can be extended analytically through it, while the origin $\{r = 0\}$ which is a true curvature singularity. The horizon separates the space-time in two domains :

- the exterior of the black hole $\{r > 2M\}$ is a static domain where $\partial/\partial t$ is timelike and $\partial/\partial r$ spacelike ;
- the interior of the black hole $\{r < 2M\}$, is a dynamic region where $\partial/\partial t$ is spacelike, $\partial/\partial r$ timelike, so r should be thought of as a time variable inside the black hole, it is therefore oriented ; the usual understanding of a black hole says that things can fall into it but not come out of it ; this would correspond to the inertial frames in the interior being dragged towards the singularity at $\{r = 0\}$, i.e. $-\partial/\partial r$ being future oriented, but one may just as well consider the reverse time orientation which would correspond to a white hole ; nothing at this point indicates that one orientation is preferable to the other.

The two domains are globally hyperbolic. The hypersurfaces

$$\{t\} \times]2M, +\infty[\times S^2_{\theta,\varphi}$$

are Cauchy hypersurfaces for the exterior and

$$\mathbb{R}_t \times \{r\} \times S^2_{\theta,\varphi}$$

are Cauchy hypersurfaces for the interior.

The shape of the lightcones outside and inside the black-hole is well described by the respective position of the null vectors

$$V^\pm := \frac{\partial}{\partial t} + F(r)\frac{\partial}{\partial r}\,.$$

The vectors V^+ and V^- get closer to each other as one approaches the horizon from the inside or the outside. The situation is however very different on either side of the horizon : outside the black hole, the light cones get narrower as one approaches the horizon, whereas inside they get wider (see figure 3.2). The intuitive description of a black hole tells us that the more we approach the horizon from the exterior, the harder it becomes to escape the attraction, until at the horizon, even a photon cannot escape anymore. But it is easier and easier to go towards the black hole. In terms of light-cones, this seems to indicate a picture where the lightcones are tilted towards the horizon and become tangent to the horizon as we reach it. When representing the lightcones in the Schwarzschild coordinates however, this does not appear to be correct after

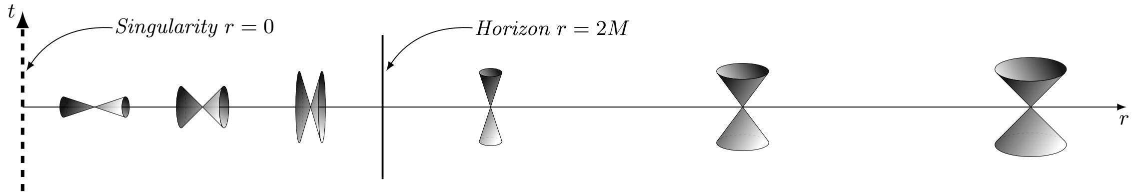

Figure 3.2: The light cones outside and inside the black-hole in Schwarzschild coordinates (t, r).

all. How do we solve this canondron? We will see that the intuitive picture has some degree of realism when we build the maximal analytic extension of the Schwarzschild space-time, which gives the correct picture of the horizon.

Schwarzschild's space-time is asymptotically flat. This can be seen in the fact that as $r \to +\infty$, the metric g approaches the Minkowski metric in spherical coordinates. It is also apparent in the property that the curvature tends to zero as $r \to +\infty$ (see next paragraph). Note that asymptotically flat means asymptotically flat in space, certainly not in time, we have a curved space-time that is static, therefore the curvature does not die out as time tends to infinity.

3.2.1 Connection and curvature

In the Schwarzschild coordinates (t, r, θ, φ), the non zero Christoffel symbols of the Levi-Civita connection are

$$\Gamma^0_{01} = \frac{M}{r(r-2M)}\,,\ \Gamma^1_{00} = \frac{M(r-2M)}{r^3}\,,\ \Gamma^1_{11} = -\frac{M}{r(r-2M)}\,,$$
$$\Gamma^1_{22} = -(r-2M)\,,\ \Gamma^1_{33} = -(r-2M)\sin^2\theta\,,$$
$$\Gamma^2_{12} = \Gamma^3_{13} = \frac{1}{r}\,,\ \Gamma^2_{33} = -\sin\theta\cos\theta\,,\ \Gamma^3_{23} = \cot\theta\,,$$

and the non-zero components of the Riemann tensor

$$R_{0101} = -\frac{M(r-2M)}{r^2}\sin^2\theta\,,\ R_{0202} = \frac{2M}{r^3}\,,\ R_{0303} = -\frac{M(r-2M)}{r^2}\,,$$
$$R_{1212} = \frac{M}{r-2M}\sin^2\theta\,,\ R_{1313} = -2Mr\sin^2\theta\,,$$
$$R_{2323} = \frac{M}{r-2M}\,.$$

If, instead of the Schwarzschild coordinate basis, we evaluate the components of the Riemann tensor with respect to an orthonormal basis with vectors proportional to the coordinate basis vectors, namely (adopting Chandrasekhar's notations for frame indices between brackets)

$$e_{(0)}{}^a\partial_a = \frac{1}{\sqrt{F}}\frac{\partial}{\partial t}\,,\ e_{(1)}{}^a\partial_a = \sqrt{F}\frac{\partial}{\partial r}\,,\ e_{(2)}{}^a\partial_a = \frac{1}{r}\frac{\partial}{\partial\theta}\,,\ e_{(3)}{}^a\partial_a = \frac{1}{r\sin\theta}\frac{\partial}{\partial\varphi}\,,$$

we find

$$R_{1010} = -R_{3232} = \frac{2M}{r^3}\,,\ R_{3131} = R_{1212} = R_{3030} = -R_{2020} = \frac{M}{r^3}\,,$$

and we see that the curvature, expressed in this frame, blows up at $\{r = 0\}$ but not at the horizon.

Of course it is hard to infer the existence of a singularity from the behaviour of the components of the curvature in a frame, however well-chosen we think it is. What we need is to find an intrinsic quantity that blows up at $r = 0$. Let us calculate the curvature scalar, also referred to as the Kretschmann scalar, which is defined as follows:

$$\mathrm{Kr} = R_{abcd} R^{abcd} \, .$$

It is easily calculated in the orthonormal frame above using the symmetries of the Riemann tensor:

$$\mathrm{Kr} = 32 \frac{M^2}{r^6} \, .$$

This proves that the "origin" $\{r = 0\}$ is a true curvature singularity. As for the locus $\{r = 2M\}$ we shall see in details how it is to be interpreted.

The Ricci tensor of the Schwarzschild metric is zero, Schwarzschild's space-time is a solution to the Einstein vacuum equations.

3.2.2 Symmetries, Killing vectors

Schwarzschild's space-time has a four-dimensional space of global Killing vector fields, generated by

$$\partial_t \, , \;\; \sin\varphi \partial_\theta + \cot\theta \cos\varphi \partial_\varphi \, , \;\; \cos\varphi \partial_\theta - \cot\theta \sin\varphi \partial_\varphi \, , \;\; \partial_\varphi \, ,$$

which are the timelike (outside the black hole) Killing vector field ∂_t already mentionned above and the three generators of the rotation group. In other words, the symmetry group of Schwarzschild's space-time is $\mathbb{R} \times SO(3)$.

3.2.3 The exterior of the black hole

We first consider the Schwarzschild geometry from the point of view of an observer static with respect to infinity. Such observers only see the exterior of the black hole and their perception of space-time is described by the time function t of the Schwarzschild coordinates outside the black hole. To their eyes, light rays falling into the black hole slow down infinitely as they approach the horizon and never cross it. One way of seeing this is to calculate the radial null geodesics.

Indeed, the fastest way of falling into the black hole, since the space-time is spherically symmetric (i.e. in particular without rotation), is to go towards it radially and at the speed of light. Let us first evaluate the radial null directions. A radial vector at a given point (t, r, θ, φ) is of the form

$$V = \alpha \partial_t + \beta \partial_r \, .$$

For it to be null, α and β must satisfy

$$\frac{\beta}{\alpha} = F$$

since

$$g(V, V) = \alpha^2 F - \beta^2 F^{-1} \, .$$

So the two future oriented[2] radial null directions at a given point outside the black hole are those of the vectors

$$V^{\pm} = \partial_t \pm F\partial_r\,.$$

The apparent radial speed of these vectors for an observer static at infinity and measured using the variable r is $\pm F(r)$, it is ± 1 at infinity and slows down continuously to zero as one considers points closer and closer to the black hole horizon. Moreover, their integral curves are geodesics:

Proposition 3.2. *The radial null vectors $V^{\pm}$ satisfy*

$$\nabla_{V^+}V^+ = \frac{2M}{r^2}V^+\,,\ \nabla_{V^-}V^- = -\frac{2M}{r^2}V^-\,.$$

Proof. Let us check this property for V^+. Dropping the "+" superscript for simplicity, using the values of the Christoffel symbols given above, we have

$$\begin{aligned}
\nabla_V V^{\mathbf{a}}\partial_{\mathbf{a}} &= V^{\mathbf{b}}\nabla_{\mathbf{b}}V^{\mathbf{a}}\partial_{\mathbf{a}}\\
&= V^0\nabla_0 V^{\mathbf{a}}\partial_{\mathbf{a}} + V^1\nabla_1 V^{\mathbf{a}}\partial_{\mathbf{a}}\\
&= \partial_t(V^{\mathbf{a}})\partial_{\mathbf{a}} + \Gamma^{\mathbf{a}}_{0\mathbf{b}}V^{\mathbf{b}}\partial_{\mathbf{a}} + F\partial_r(V^{\mathbf{a}})\partial_{\mathbf{a}} + F\Gamma^{\mathbf{a}}_{1\mathbf{b}}V^{\mathbf{b}}\partial_{\mathbf{a}}\\
&= 0 + \Gamma^0_{01}V^1\partial_t + \Gamma^1_{00}V^0\partial_r + F\partial_r(V^1)\partial_r + F\Gamma^0_{10}V^0\partial_t + F\Gamma^1_{11}V^1\partial_r\\
&= \frac{MF^{-1}}{r^2}F\partial_t + \frac{MF}{r^2}\partial_r + F\frac{2M}{r^2}\partial_r + F\frac{MF^{-1}}{r^2}\partial_t - F\frac{MF^{-1}}{r^2}F\partial_r\\
&= \frac{2M}{r^2}V\,.
\end{aligned}$$

The calculation is absolutely similar for V^- and left as an exercice. □

Note that there is another, less explicit, way of seeing that $V^{\pm}$ generate geodesics. We have

$$\mathrm{d}u = \mathrm{d}t - F^{-1}\mathrm{d}r \text{ and } \mathrm{d}v = \mathrm{d}t + F^{-1}\mathrm{d}r\,,$$

whence

$$\nabla^a u\partial_a = g^{-1}(\mathrm{d}u) = F^{-1}\partial_t + \partial_r \text{ and } \nabla^a v\partial_a = g^{-1}(\mathrm{d}v) = F^{-1}\partial_t - \partial_r.$$

Therefore,

$$\nabla u = \frac{1}{F}V^+ \text{ and } \nabla v = \frac{1}{F}V^-.$$

Hence the functions u and v are optical functions and their gradients generate affinely parametrised null geodesics (see Definition 2.26 and Lemma 2.2). It is interesting to remark that r is an affine parameter for these two families of curves, indeed

$$\mathrm{d}r(\nabla u) = -\mathrm{d}r(\nabla v) = 1. \tag{3.24}$$

We note that the t,r-speed of radial light rays slows down as they approach the horizon. The question is whether this slowing down is strong enough to make t non-integrable along their worldlines. The answer is clearly yes since

$$\int_{2M}^R \frac{\mathrm{d}r}{F(r)} = \int_{2M}^R \frac{r\mathrm{d}r}{r-2M} = +\infty \text{ for any } R > 2M\,.$$

[2] Future-oriented provided we choose outside the black hole the time orientation given by ∂_t.

This can be seen explicitly by introducing the Regge-Wheeler variable

$$r_* = r + 2M\mathrm{Log}(r - 2M) \tag{3.25}$$

which varies from $-\infty$ to $+\infty$ as r varies from $2M$ to $+\infty$. It satisfies

$$\frac{dr_*}{dr} = F^{-1}$$

and the metric g takes the form

$$g = F\left(\mathrm{d}t^2 - \mathrm{d}r_*^2\right) - r^2\mathrm{d}\omega^2.$$

The radial null vectors take the expression

$$V^{\pm} = \partial_t \pm \partial_{r_*}$$

and their integral lines parametrized by r_* are the straight lines

$$\gamma^{\pm}_{C,\omega_0}(r_*) = \{(t, r_*, \omega)\,;\ \omega = \omega_0\,,\ t = \pm r_* + C\}\ ,\ C \in \mathbb{R}\,,\ \omega_0 \in S^2\,.$$

The horizon $\{r = 2M\}$ (corresponding to $r_* \to -\infty$) is reached in infinite time t. A remarkable consequence of this property is that if we choose for a covariant field equation (Dirac, Maxwell, or the wave equation for instance) some initial data at time $t = 0$ whose support is contained in $\{r \geq 2M + \varepsilon\}$, $\varepsilon > 0$, then the support of the solution will only reach the horizon when t becomes infinite. An important consequence of this remark is that the interior of the black hole and the exterior should not be considered as co-existing simultaneously for the time variable t, in other words, a $t =$ constant slice for $r \in]0, +\infty[$ has no physical meaning whatsoever. Such hypersurfaces will be represented and put in their proper perspective once we have constructed the maximal extension of Schwarzschild's space-time.

The spacelike geometry of the exterior of the black hole

The exterior of the black hole is globally hyperbolic. We consider the foliation by Cauchy hypersurfaces induced by the time function t, i.e. the slices are

$$\Sigma_t = \{t\} \times]2M, +\infty[_r \times S^2_\omega,\ t \in \mathbb{R},$$

with the induced Riemannian metric

$$h = F^{-1}\mathrm{d}r^2 + r^2\mathrm{d}\omega^2. \tag{3.26}$$

The 3+1 decomposition of the geometry is given by (calling $\mathcal{M}$ the exterior of the black hole) :

$$\mathcal{M} = \mathbb{R}_t \times \Sigma\,,\ \Sigma =]2M, +\infty[_r \times S^2_\omega\,,\ g = F\mathrm{d}t^2 - h = \frac{N^2}{2}\mathrm{d}t^2 - h \tag{3.27}$$

with the lapse function $N = \sqrt{2}F^{1/2}$. The exterior of the black hole is static : $\frac{\partial}{\partial t}$ is a Killing vector field (since g does not depend on t), is timelike outside the black hole and is everywhere orthogonal to the Cauchy hypersurfaces Σ_t. The time orientation is chosen by deciding that $\frac{\partial}{\partial t}$ is future pointing and the normalized vector field T^a is then

$$T^a \partial_a = \sqrt{2}\, F^{-1/2} \frac{\partial}{\partial t} = \frac{2}{N} \frac{\partial}{\partial t} \,.$$

We consider a generic spacelike slice (Σ, h). The metric h appears singular at $r = 2M$. This is merely due to the choice of coordinates ; introducing as the new radial variable $u(r)$ the h-distance to the horizon, we show that (Σ, h) is a smooth manifold and that the horizon $H = \{2M\}_r \times S^2_{\theta,\varphi}$ is a smooth boundary.

Given $p = (r, \omega) \in \Sigma$, the h-distance from p to the horizon is given by

$$u(r) = \int_{[2M,r]} F^{-1/2}(s) \mathrm{d}s = \int_{[2M,r]} \frac{\sqrt{s}}{\sqrt{s - 2M}} \mathrm{d}s \,. \tag{3.28}$$

This distance is finite and H thus appears as the boundary of (Σ, h). Since

$$\frac{\mathrm{d}u}{\mathrm{d}r} = F^{-1/2},$$

the metric h can be written as

$$h = \mathrm{d}u^2 + r^2 \mathrm{d}\omega^2 \tag{3.29}$$

and

$$\Sigma =]0, +\infty[_u \times S^2_\omega \,.$$

The function $u(r)$ is continuous and strictly increasing from $[2M, +\infty[$ onto $[0, +\infty[$, it is $\mathcal{C}^\infty$ on $]2M, +\infty[$ but it is not differentiable at $2M$. However, the inverse function satisfies

Lemma 3.1. *The function $u \longmapsto r(u)$ is $\mathcal{C}^\infty$ on $[0, +\infty[$ and all its derivatives are uniformly bounded on $[0, +\infty[$. In particular, the first derivative $\frac{\mathrm{d}r}{\mathrm{d}u} = F^{1/2}$ (and therefore also the lapse function) is uniformly bounded as well as all its derivatives on $[0, +\infty[$.*

Proof of lemma 3.1 : the first and second derivatives $F^{1/2}$ and M/r^2 are continuous on $[0, +\infty[_u$ whence r is $\mathcal{C}^2$ on $[0, +\infty[_u$. If r is $\mathcal{C}^k$ on $[0, +\infty[_u$, then so is the second derivative and the lemma is thus proved by induction. □

This entails that h is smooth on $\bar{\Sigma} = [0, +\infty[_u \times S^2_\omega$; $(\bar{\Sigma}, h)$ is a smooth manifold with boundary. Moreover

Theorem 3.2. *The metric h is uniformly equivalent to the euclician metric on the exterior of the unit ball in $\mathbb{R}^3$*

$$\mathrm{d}u^2 + (1 + u)^2 \mathrm{d}\omega^2 \,.$$

Proof. We see that

$$\frac{1+u}{r} \to \frac{1}{2M} \text{ as } r \to 2M \,,$$

$$\frac{1+u}{r} \to 1 \text{ as } r \to +\infty \text{ since } F(r) \to 1$$

and moreover $(1 + u)/r$ is continuous on $[2M, +\infty[_r$, hence, there exists $C > 0$ such that

$$C < \frac{1+u}{r} < \frac{1}{C} \text{ for } 2M \le r < +\infty \,.$$

This proves the theorem. □

Bending of light-rays : the photon sphere

We have an extreme example of bending of light rays by gravity in the schwarzschild geometry : the photon sphere, which is a sphere of trapped geodesics around the black hole. Let us consider in the equatorial plane a null vector that is purely rotational, i.e. of the form $V = a\partial_t + b\partial_\varphi$, for example, we can take

$$V = r\partial_t + \sqrt{1 - \frac{2M}{r}}\partial_\varphi \,.$$

The integral curves of this vector field describe circles in the equatorial plane (more correctly helices if we consider the time as well as space variables) whose tangent vectors are null. What is the acceleration of such curves? This is the following simple calculation:

$$\begin{aligned}
\nabla_V V = V^a \nabla_a V^b \partial_b &= \left(V^{\mathbf{a}} \nabla_{\mathbf{a}} V^{\mathbf{b}} + \Gamma^{\mathbf{b}}_{\mathbf{ac}} V^{\mathbf{c}} \right) \partial_{\mathbf{b}} \\
&= \left(V^0 \partial_t V^{\mathbf{b}} + V^3 \partial_\varphi V^{\mathbf{b}} + V^0 \Gamma^{\mathbf{b}}_{0\mathbf{c}} V^{\mathbf{c}} + V^3 \Gamma^{\mathbf{b}}_{3\mathbf{c}} V^{\mathbf{c}} \right) \partial_{\mathbf{b}} \\
&= \left(V^0 \Gamma^{\mathbf{b}}_{0\mathbf{c}} V^{\mathbf{c}} + V^3 \Gamma^{\mathbf{b}}_{3\mathbf{c}} V^{\mathbf{c}} \right) \partial_{\mathbf{b}} \\
&= r \left(\Gamma^0_{01} V^1 \partial_t + \Gamma^1_{00} V^0 \partial_r \right) \\
&\quad + \sqrt{1 - \frac{2M}{r}} \left(\Gamma^1_{33} V^3 \partial_r + \Gamma^3_{31} V^1 \partial_\varphi + \Gamma^3_{32} V^2 \partial_\varphi \right) \\
&= r \Gamma^1_{00} V^0 \partial_r + \sqrt{1 - \frac{2M}{r}} \Gamma^1_{33} V^3 \partial_r \\
&= \left(r^2 \frac{M}{r^3} (r - 2M) + \left(1 - \frac{2M}{r} \right) (-r) \left(1 - \frac{2M}{r} \right) \right) \partial_r \\
&= \left(1 - \frac{2M}{r} \right) (3M - r) \partial_r \,.
\end{aligned}$$

As could be expected, the acceleration is purely radial. It points towards the black hole if $r > 3M$, away from the black hole if $r < 3M$ and it is zero if $r = 3M$. This means that the integral curves of V for $r = 3M$ are geodesics : there are some "photon trajectories" orbiting the black hole at $r = 3M$. This is a very strong effect of light bending which requires a black hole or a very dense body of radius lower than three times its mass.

3.2.4 Maximal extension

After having adopted, in the previous section, the point of view of an observer static with respect to infinity, and thus limited our study to the exterior of the black hole foliated using Schwarzschild's time coordinate, we describe here briefly the global geometry of Schwarzschild's space-time. We define the Eddington-Finkelstein and the Kruskal-Szekeres coordinates inside and outside the black hole. These will allow us to show that the horizon is not a singularity of the metric. The maximal analytic extension of Schwarzschild's space-time will then appear naturally. Most of the material of this section is standard, it can be found under various forms in [6] and [23] for example.

Eddington-Finkelstein coordinates

There are two types of Eddington-Finkelstein coordinates respectively referred to as advanced and retarded, or, more to the point, incoming and outgoing. They are based on the incoming (resp. outgoing) radial null geodesics.

The incoming Eddington-Finkelstein coordinates are

$$v = t + r_*, r, \theta, \varphi,$$

where $r_* = r + 2M \log(r - 2M)$ is the Regge-Wheeler coordinate. The Schwarzschild metric, in these coordinates, reads

$$g = \left(1 - \frac{2M}{r}\right) \mathrm{d}v^2 - 2\mathrm{d}v\mathrm{d}r - r^2 \mathrm{d}\omega^2. \tag{3.30}$$

This is fine outside the black hole but not inside where the expression of r_* is no longer valid. If we define r_* inside the black hole as

$$r_* = r + 2M \log(2M - r), \tag{3.31}$$

r_* varies from $-\infty$ to $2M \log(2M)$ as r varies from $2M$ to 0. We keep the definition $v = t + r_*$ inside the black hole and we obtain the same expression (3.30) of the metric g. This is analytic on $\mathbb{R}_v \times]0, +\infty[_r \times S^2_\omega$ and does not degenerate anywhere (apart from the usual problem due to spherical coordinates) as we can see from the determinent of g:

$$\det g = -r^4 \sin^2 \theta.$$

The whole of Schwarzschild's space-time is represented by the incoming Eddington-Finkelstein coordinates and we can wonder how to interpret the space-time, and more particularly the horizon, physically.

A v =constant line is a curve

$$(t = -r_* + v_0, r_*, \omega = \omega_0),$$

with v_0 and ω_0 fixed; i.e. this is an integral curve of the vector field $V^- = \partial_t - \partial_{r_*}$, in other words, a null geodesic. Outside the black hole, this is clearly the incoming radial null geodesic γ_{v_0,ω_0}. If we parametrize this curve by r, which is an affine parameter (see (3.24)), then it is an analytic curve in all positive values of r, in particular we see that the incoming null geodesic γ_{v_0,ω_0} outside the black hole extends analytically inside the black hole as the same $v = v_0$ line. As we follow the geodesic from infinity inwards, we move towards the future and r decreases (with r_* decreasing from $+\infty$ to $-\infty$ as r decreases from $+\infty$ to $2M$), the geodesic then crosses the horizon $\{r = 2M\}$ and keeps going towards the singularity at the origin (r_* increasing from $-\infty$ to $2M \log(2M)$ as r decreases from $2M$ to 0). The interior of the black hole is thus understood as lying in the future of the exterior. The correct time orientation of the interior of the black hole, consistent with that given by ∂_t outside the black hole, would appear to be given by $-\partial_r$.

The horizon is seen as the hypersurface $\mathbb{R}_v \times \{2M\}_r \times S^2_\omega$ and separates the exterior from the interior. Moreover, the horizon appears as a null hypersurface. Indeed, the metric does not degenerate there, but its restriction to the horizon is the 2-metric

$$-(2M)^2 \mathrm{d}\omega^2 \,,$$

whereas the horizon is a 3-surface. This means that one of the tangent vectors to the horizon is null. At each point of the hypersurface $\{r = 2M\}$, the space of tangent vectors is spanned by ∂_v, ∂_θ and ∂_φ. The "squared norm" of ∂_v for the metric g is given by

$$g\left(\frac{\partial}{\partial v}, \frac{\partial}{\partial v}\right) = \left(1 - \frac{2M}{r}\right) .$$

So ∂_v is null for $r = 2M$. A picture of Schwarzschild's space-time in incoming Eddington-Finkelstein coordinates is given by (Figure 3.3) and we see that once inside the black hole, we cannot come back out of it.

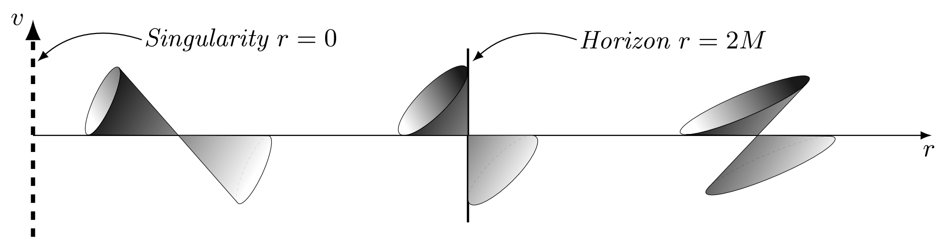

Figure 3.3: The light cones outside and inside the black-hole in incoming Eddington-Finkelstein coordinates (v, r).

We now perform a similar construction based on the outgoing Eddington-Finkelstein coordinates :

$$u = t - r_*, r, \theta, \varphi ,$$

and the Schwarzschild metric in these coordinates takes the expression

$$g = \left(1 - \frac{2M}{r}\right) \mathrm{d}u^2 + 2\mathrm{d}u\mathrm{d}r - r^2\mathrm{d}\omega^2 . \tag{3.32}$$

Similarly to the incoming case, this is analytic on $\mathbb{R}_u \times]0, +\infty[_r \times S^2_\omega$ and does not degenerate anywhere. The whole of Schwarzschild's space-time is again represented, but the physical picture is different. Following an outgoing radial null geodesic (a u =constant line) towards the future, we emerge from the singularity at $r = 0$, cross the interior of the "black hole", the horizon, emerge from the "black hole" and go towards infinity. The black hole does not appear to be so black in this case since light rays emerge from it. The horizon is again a null hypersurface but this time it cannot be crossed from the exterior to the interior. This is a very different description of Schwarzschild's space-time corresponding not to a black hole, but to a white hole (see figure OutgoEF). The time orientation of the interior consistent with the one given by ∂_t outside the black hole would now seem to correspond to ∂_r.

What we have constructed using the incoming and the outgoing Eddington-Finkelstein coordinates are similar objects but with the opposite time orientation. We shall see in the next section that the two descriptions are both present in the most complete picture of Schwarzschild's space-time : the maximal analytic extension, also known as the Kruskal manifold.

Kruskal-Szekeres coordinates

Outside the black hole, Kruskal Szekeres coordinates (T, X, ω), ω denoting the angular variables of the Schwarzschild coordinate system, are defined by

$$T = \frac{1}{2} e^{\frac{r_*}{4M}} \left(e^{\frac{t}{4M}} - e^{-\frac{t}{4M}} \right) \, , \; X = \frac{1}{2} e^{\frac{r_*}{4M}} \left(e^{\frac{t}{4M}} + e^{-\frac{t}{4M}} \right) \, , \tag{3.33}$$

where r_* is the Regge-Wheeler variable outside the black hole given by (3.25)

$$r_* = r + 2M \mathrm{Log}(r - 2M) \, .$$

This coordinate system maps the exterior of the black hole $\mathbb{R}_t \times]2M, +\infty[_r \times S^2_\omega$ onto the quadrant $\{X > |T|\}$ of $\mathbb{R}_T \times \mathbb{R}_X \times S^2_\omega$. The horizon now appears as the hypersurface

$$\{(T, X, \omega); \; |T| = X > 0, \; \omega \in S^2\},$$

its future connected component corresponding to the horizon as described in incoming Eddington-Finkelstein coordinates and its past component to the horizon crossed by the outgoing Eddington-Finkelstein coordinates. The outgoing (resp. incoming) radial null geodesics, represented in (t, r_*, ω) coordinates as the straight lines $\{(t, r_* = t + s, \omega); \; t \in \mathbb{R}\}$ (resp. $\{(t, r_* = -t + s, \omega); \; t \in \mathbb{R}\}$) for fixed $s \in \mathbb{R}$ and $\omega \in S^2$, are described in Kruskal-Szekeres coordinates as the straight lines $\{(T, X = T + S, \omega)\}$ (resp. $\{(T, X = -T + S, \omega)\}$) for fixed S and ω.

Inside the black hole, the definition is very similar. We consider the Regge-Wheeler coordinate adapted to this domain (given by (3.31))

$$r_* = r + 2M \mathrm{Log}|r - 2M| = r + 2M \mathrm{Log}(2M - r) \, ,$$

the expression of the variables T and X in terms of t and r_* is then given by

$$T = \frac{1}{2} e^{\frac{r_*}{4M}} \left(e^{-\frac{t}{4M}} + e^{\frac{t}{4M}} \right) \, , \; X = \frac{1}{2} e^{\frac{r_*}{4M}} \left(e^{-\frac{t}{4M}} - e^{\frac{t}{4M}} \right) \, . \tag{3.34}$$

The interior of the black hole $\mathbb{R}_t \times]0, 2M[_r \times S^2_\omega$ is mapped onto the domain

$$\{(T, X, \omega) \in \mathbb{R} \times \mathbb{R} \times S^2 \, ; \; |X| < T < \sqrt{X^2 + 2M}\}$$

and the singularity at $r = 0$ is represented as the product of S^2_ω with the hyperbola in the (T, X)-plane: $\{(T, X) \, ; \; T^2 - X^2 = 2M \, , \; T > 0\}$.

The expression of the metric in Kruskal-Szekeres coordinates is the same inside and outside the black hole

$$g = \frac{16M^2}{X^2 - T^2} \left(1 - \frac{2M}{r} \right) \left(\mathrm{d}T^2 - \mathrm{d}X^2 \right) - r^2 \mathrm{d}\omega^2 \, .$$

This can be simplified using the fact that

$$X^2 - T^2 = (r - 2M) e^{\frac{r}{2M}} \tag{3.35}$$

and we obtain

$$g = \frac{16M^2}{r} e^{-\frac{r}{2M}} \left(\mathrm{d}T^2 - \mathrm{d}X^2 \right) - r^2 \mathrm{d}\omega^2 \tag{3.36}$$

where r is determined implicitely in terms of T and X by (3.35). The function $(r - 2M) e^{\frac{r}{2M}}$ is analytic in r and strictly increasing from $]0, +\infty[$ onto $] - 2M, +\infty[$. It follows that r is an analytic function of $X^2 - T^2$, and therefore of (T, X), on $-2M < X^2 - T^2 < +\infty$.

Maximal Schwarzschild space-time

As we have seen above, the metric (3.36) can be extended analytically on the region

$$\mathcal{M}^{\mathcal{K}} = \left\{ (T, X, \omega) \in \mathbb{R} \times \mathbb{R} \times S^2_\omega \, ; \; X^2 - T^2 > -2M \right\} .$$

We obtain a new space-time $(\mathcal{M}^{\mathcal{K}}, g)$ called the Kruskal extension, or maximal analytic extension, of Schwarzschild's space-time (see figure 3.4). It contains four blocks separated by a bifurcate horizon $\{|T| = |X|\}$:

$$\begin{aligned}
\mathrm{I} &:= \left\{ (T, X, \omega) \, , \; X > |T| \, , \; \omega \in S^2 \right\} , \\
\mathrm{II} &:= \left\{ (T, X, \omega) \, , \; |X| < T < \sqrt{2M + X^2} \, , \; \omega \in S^2 \right\} , \\
\mathrm{III} &:= \left\{ (T, X, \omega) \, , \; X < -|T| \, , \; \omega \in S^2 \right\} , \\
\mathrm{IV} &:= \left\{ (T, X, \omega) \, , \; -|X| > T > -\sqrt{2M + X^2} \, , \; \omega \in S^2 \right\} .
\end{aligned}$$

Blocks I and III are exteriors (corresponding to $r > 2M$) and the blocks II and IV are interiors (corresponding to $0 < r < 2M$). The realisation of the Schwarzschild manifold that we constructed using the incoming (resp. outgoing) Eddington-Finkelstein coordinates is the union of blocks I and II (resp. I and IV) with the part of the horizon between them.

The union of blocks III and IV with the part of the horizon between them is also a realization of the Schwarzschild manifold; it is isometric to the union of blocks I and II with the adequate part of the horizon with the time orientation reversed. More explicitely, blocks III and IV are the image of the Schwarzschild space-time, described in Schwarzschild coordinates, by the transformations (3.33) and (3.34) with the signs of T and X reversed. In the maximal extension, the horizon can be understood as the union of two null hypersurfaces describes as $\{T = X\}$ and $\{T = -X\}$. They intersect at the 2-sphere at $\{T = X = 0\}$ that is the common boundary of all the level hypersurfaces of the function t. This 2-sphere is a smooth submanifold of the extended space-time, referred to as the crossover of the horizons or crossing sphere.

Note that $(\mathcal{M}^{\mathcal{K}}, g)$ is globally hyperbolic, the hypersurface $\{T = 0\}$ is a Cauchy hypersurface.

3.2.5 Conformal compactification

Schwarzschild's space-time contains mass. This is apparent in the asymptotic behaviour of the metric : some terms are proportional to the mass M of the black hole and fall off in $1/r$ at infinity. These terms prevent the construction of a complete regular compactification similar to what can be done with Minkowski space-time. A partial compactification however is possible and yields in the limit $M \to 0$ a partial compactification of Minkowski space-time where only $\mathscr{I}^\pm$ are defined but neither $i^\pm$ nor i^0. This compactification is performued using the variables $u = t - r_*$ and $v = t + r_*$. The lines of constant (u, ω), resp. (v, ω), are outgoing, resp. incoming, radial null geodesics. They are referred to as the principal null geodesics because their tangent vectors are double roots of the Weyl tensor (see [45] for more details).

In terms of variables $u = t - r_*$, $R = 1/r$, θ and φ, the Schwarzschild metric g takes the form

$$g = (1 - 2MR)\mathrm{d}u^2 - \frac{2}{R^2}\mathrm{d}u\mathrm{d}R - \frac{1}{R^2}\mathrm{d}\omega^2 \, .$$

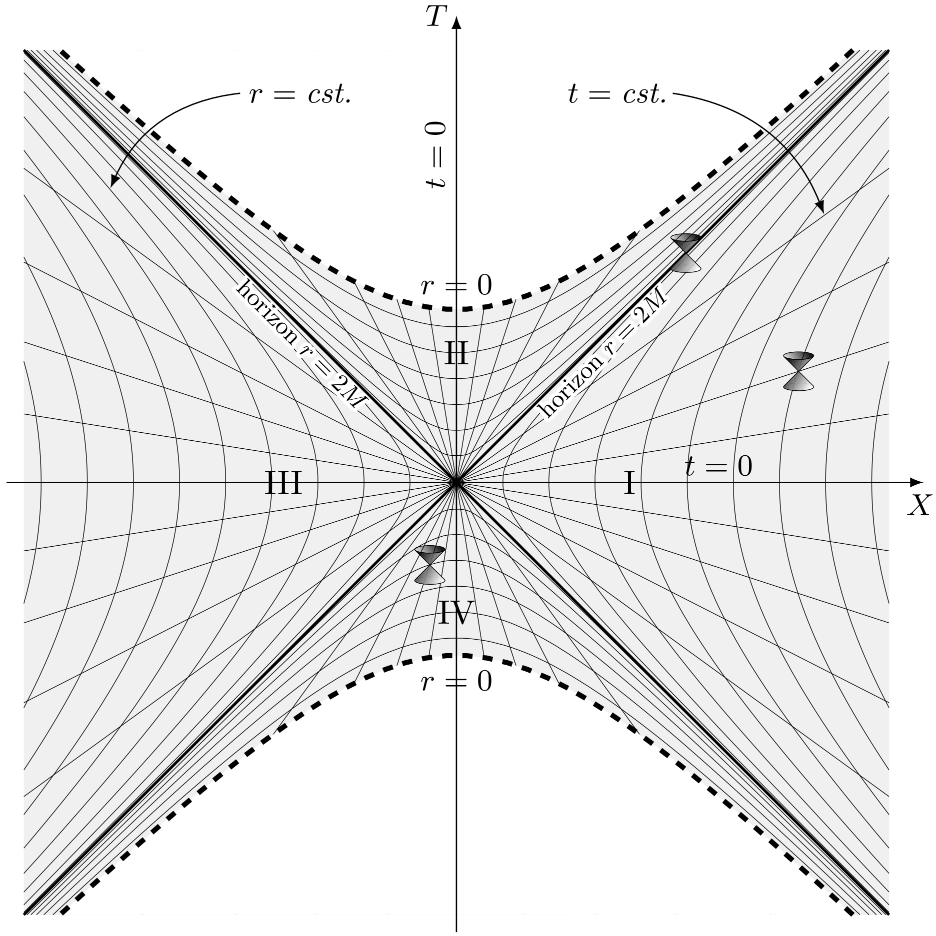

Figure 3.4: The maximal analytic extension of Schwarzschild's space-time in Kruskal-Szekeres coordinates: domains I and III correspond to $r > 2M$, domain II represents the interior of the black hole and domain IV the interior of the white hole.

Rescaling the metric with the conformal factor $\Omega = R = 1/r$, we obtain

$$\hat{g} = R^2 g = R^2(1 - 2MR)\mathrm{d}u^2 - 2\mathrm{d}u\mathrm{d}R - \mathrm{d}\omega^2 ,$$

which extends as an analytic metric on the domain $\mathbb{R}_u \times \left[0, \frac{1}{2M}\right[_R \times S^2_{\theta,\varphi}$. Similarly to the Minkowski case, we can add a boundary to the exterior of the black hole : the hypersurface $\mathbb{R}_u \times \{0\}_R \times S^2_{\theta,\varphi}$. A point $(u_0, 0, \theta_0, \varphi_0)$ on the boundary is reached along the outgoing radial null geodesic

$$\begin{aligned} \gamma_{u_0,\theta_0,\varphi_0}(r) &= (t = r + 2M\mathrm{Log}(r - 2M) + u_0\, ,\ r\, ,\ \theta = \theta_0\, ,\ \varphi = \varphi_0) \\ &= \left(u = u_0\, ,\ R = \frac{1}{r}\, ,\ \theta = \theta_0\, ,\ \varphi = \varphi_0 \right) \end{aligned}$$

as $r \to +\infty$ and there is a one to one correspondence between the points on the boundary and the

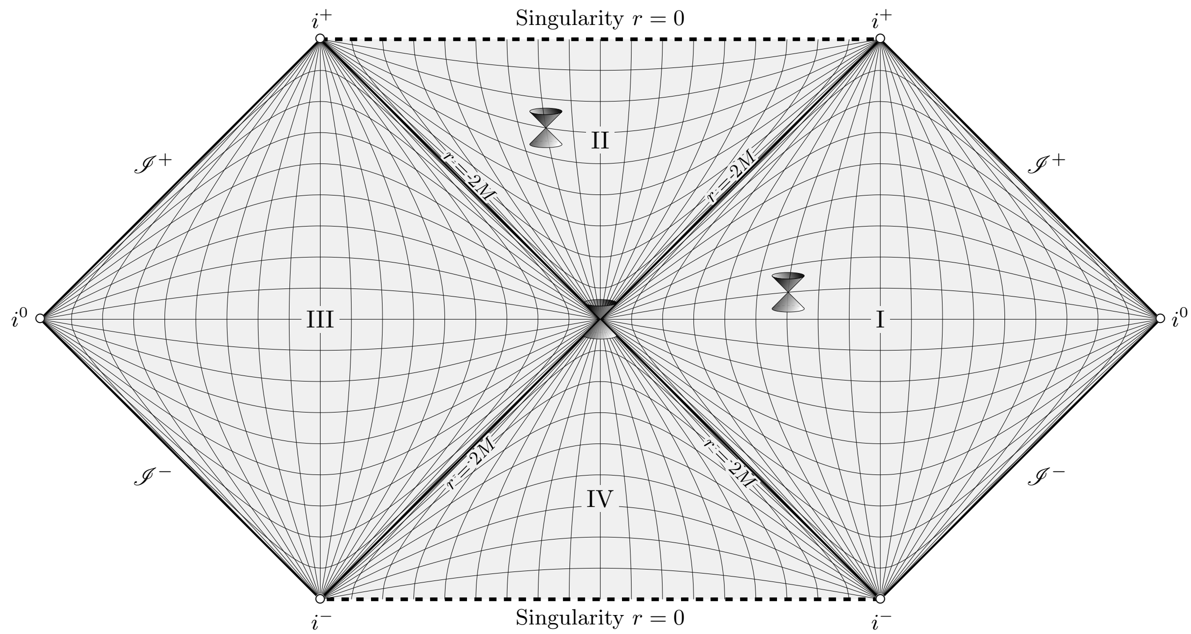

Figure 3.5: The Penrose diagram of maximal Schwarzschild space-time. The $t = cst$ lines all go through the crossing sphere and the $r = cst$ lines end at the timelike infinities.

outgoing radial null geodesics. The hypersurface therefore represents future null infinity, $\mathscr{I}^+$, for the Schwarzschild metric. A similar construction can be done using the variables v, R, θ, φ and allows to construct $\mathscr{I}^-$. These two hypersurfaces can be added as boundaries to blocks I and III. We obtain the Penrose diagram of maximal Schwarzschild space-time (Figure 3.5). Using Theorem 2.7 and the fact that the scalar curvature of the Schwarzschild metric is zero, we can calculate the scalar curvature of the rescaled metric $\hat{g} = R^2 g$ and we find

$$\frac{1}{6}\mathrm{Scal}_{\hat{g}} = 2MR\,.$$

The gradient of Ω on the rescaled space-time is given by

$$\nabla_{\hat{g}}\,\Omega = \hat{g}^{-1}(\mathrm{d}\Omega) = \hat{g}^{-1}(\mathrm{d}R) = -\frac{\partial}{\partial u}$$

is everywhere timelike and past-oriented outside the black hole and becomes null at $\mathscr{I}$; it is also a Killing vector field. Compactified Schwarzschild's space-time has the same Killing vector fields as the original space-time.

3.3 Other spherically symmetric black holes

The Schwarzschild metric has extensions that still retain the feature of spherical symmetry :

- the Reissner-Nordström metric describes a charged spherical static black hole in an asymptotically flat universe ; it is no longer a solution of the Einstein vacuum equations but of the Einstein-Maxwell system, i.e. the Einstein equations with the stress-energy tensor of an electromagnetic field as a source, coupled to the Maxwell system ;
- the de Sitter-Schwarzschild metric describes a spherical eternal uncharged black hole in a universe with a positive cosmological constant.

In fact the two extensions are part of the de Sitter-Reissner-Nordström family. This is the two-parameter family of metrics describing a spherical, charged, eternal black hole in a universe with a non negative cosmological constant. It is defined on $\mathbb{R}_t \times]0,+\infty[_r \times S^2_\omega$ by

$$g = F(r)\mathrm{d}t^2 - F(r)^{-1}\mathrm{d}r^2 - r^2\mathrm{d}\omega^2 \, , \; F(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2} - \Lambda r^2 \, , \tag{3.37}$$

where $M > 0$ is the mass of the black hole, Q its charge and $\Lambda > 0$ the cosmological constant. In the case where $\Lambda = 0$, g is the Reissner-Nordström metric and when $Q = 0$, g is the De Sitter-Schwarzschild metric. When $M = Q = 0$ and $\Lambda > 0$, the geometry we obtain is known as De Sitter space-time.

3.3.1 Reissner-Nordström metrics

We consider the metric given by (3.37) with $\Lambda = 0$, i.e. with

$$F(r) = 1 - \frac{2M}{r} + \frac{Q^2}{r^2} \, .$$

Similarly to the case of the Schwarzschild metric, $\{r = 0\}$ is a curvature singularity and the zeros of the function F are the radii of the horizons, which are fictitious singularities that can be understood as smooth null hypersurfaces by means of Kruskal-Szekeres-type coordinates ; except now we may have two horizons. There are three types of Reissner-Nordström metrics, depending on the respective importance of M and Q.

1. For $M > |Q|$, the function F has two roots
$$r_\pm := M \pm \sqrt{M^2 - Q^2} \, , \tag{3.38}$$
so the space-time has two horizons. The horizon $\{r = r_+\}$ will be called the outside horizon, or horizon of the black hole, while $\{r = r_-\}$ will be called the inner horizon.

2. For $M = |Q|$, $r_+ = r_- = M$ is the only root of F and there is only one horizon. The corresponding black hole is referred to as an extreme Reissner-Nordström black hole.

3. For $M < |Q|$, the function F has no real root. There are no horizons in this case, the space-time contains no black hole and the singularity $\{r = 0\}$ is naked (i.e. not hidden beyond a horizon).

Sub-extremal case : $M > |Q|$

The two horizons decompose the Reissner-Nordström manifold into three regions called blocks.

- *Block I* is the exterior of the black hole $\{r > r_+\}$. It is a static and globally hyperbolic region : the Killing vector ∂_t is timelike and orthogonal to the Cauchy hypersurfaces $\{t\}\times]r_+, +\infty[_r \times S^2_\omega$.
- *Block II* is the region between the two horizons $\{r_- < r < r_+\}$: it is a dynamic region where ∂_r is timelike and ∂_t is spacelike, as inside a Schwarzschild black hole. It is also globally hyperbolic.
- *Block III* is the region beyond the inner horizon $\{r < r_-\}$. It is another static region where ∂_t is Killing and timelike and orthogonal to the level hypersurfaces of t that are spacelike. But block III is not globally hyperbolic because of the singularity at $r = 0$. If we take any smooth connected spacelike hypersurface Σ in block III, there are inextendible timelike geodesics ending in the singularity and not meeting Σ. The singularity is timelike since the vector field ∂_t is timelike in block III.

The spacelike geometry of block I is similar to the Schwarzschild case in that the outer horizon is at finite spacelike distance from any point outside the black hole. This is a straightforward consequence of the fact that for any $r_0 > r_+$, the integral

$$\int_{r_+}^{r_0} \frac{1}{\sqrt{F(r)}} \mathrm{d}r = \int_{r_+}^{r_0} \frac{r}{\sqrt{(r-r_+)(r-r_-)}} \mathrm{d}r < \infty \,.$$

Similarly, in block III, the inner horizon is at finite spacelike distance from any point in block III. And so is the singularity from any point in block III.

We can define a Regge-Wheeler-type coordinate r_* in each of the three blocks easily. It needs to satisfy

$$\frac{\mathrm{d}r_*}{\mathrm{d}r} = \frac{1}{F(r)} = \frac{r^2}{(r-r_+)(r-r_-)} = 1 + \frac{r_+^2}{r_+ - r_-}\frac{1}{r-r_+} + \frac{r_-^2}{r_- - r_+}\frac{1}{r-r_-} \,,$$

i.e.

$$r_* = r + \frac{r_+^2}{r_+ - r_-} \log|r - r_+| + \frac{r_-^2}{r_- - r_+} \log|r - r_-| + R_0$$

where R_0 is an arbitrary constant. In each block, the metric is expressed in terms of the variables (t, r_*, ω) as

$$g = F(r)\left(\mathrm{d}t^2 - \mathrm{d}r_*^2\right) - r^2 \mathrm{d}\omega^2 \,.$$

This allows to construct Eddington-Finkelstein-type coordinates in each block $u = t - r_*$ and $v = t + r_*$. The corresponding expressions of the metric will be

$$\begin{aligned} g &= F(r)\mathrm{d}u^2 + 2\mathrm{d}u\mathrm{d}r - r^2\mathrm{d}\omega^2 \\ &= F(r)\mathrm{d}v^2 - 2\mathrm{d}v\mathrm{d}r - r^2\mathrm{d}\omega^2 \,. \end{aligned}$$

We can then glue blocks together as we did in the Schwarzschild case :

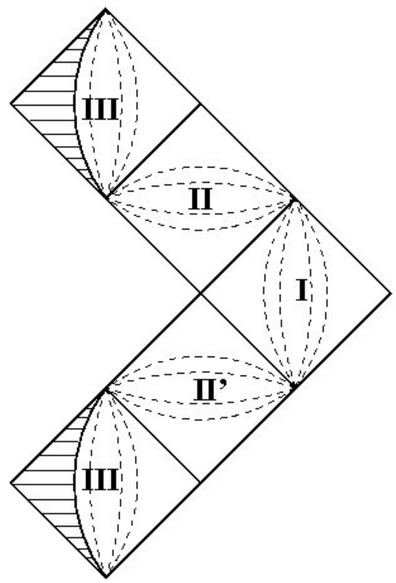

Figure 3.6: A first extension of the sub-extremal Reissner-Nordtröm space-time drawn as a Penrose diagram. The dotted lines represent the r =cont. hypersurfaces. The thick continuous curved line is the singularity $r = 0$. The shaded regions are not part of the space-time.

- using the incoming coordinates (v, r, ω), we glue block II to the future of block I via a future outer horizon and block III to the future of block II via a future inner horizon ;
- using the incoming coordinates (u, r, ω), we glue block II to the past of block I via a past outer horizon and block III to the past of block II via a past inner horizon.

We obtain a first extension of the Reissner-Nordström manifold shown in figure 3.6. We notice that some radial null geodesics are incomplete in this picture and therefore the extension is not maximal. We can then construct the maximal analytic extention of the sub-extremal Reissner-Nordström black hole by extending the incomplete radial null geodesics : the additional blocks to be glued are found by smoothness of the function r over the whole extension and by observations of time orientation. In total 6 types of blocks will be used in the construction of the maximal analytic extension : I, II, III (II here is understood as having the time orientation given by $-\partial_r$) and the same blocks with their time orientation reversed I', II' and III'.

The Penrose diagram of the maximal analytic extension of the sub-extremal Reissner-Nordtröm space-time is shown in figure 3.7.

Extreme case : $M = |Q|$

Block II no longer exists in this case and only blocks I and III remain. The Regge-Wheeler coordinate r_* is now given as a primitive of the function

$$\frac{1}{F(r)} = \frac{r^2}{(r-M)^2} = 1 + \frac{2M}{r-M} + \frac{M^2}{(r-M)^2} \, ,$$

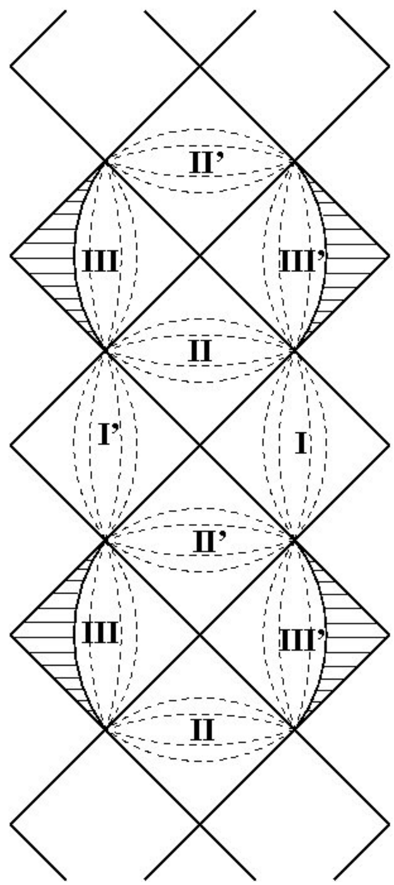

Figure 3.7: The Penrose diagram of the maximal analytic extension of the sub-extremal Reissner-Nordtröm space-time.

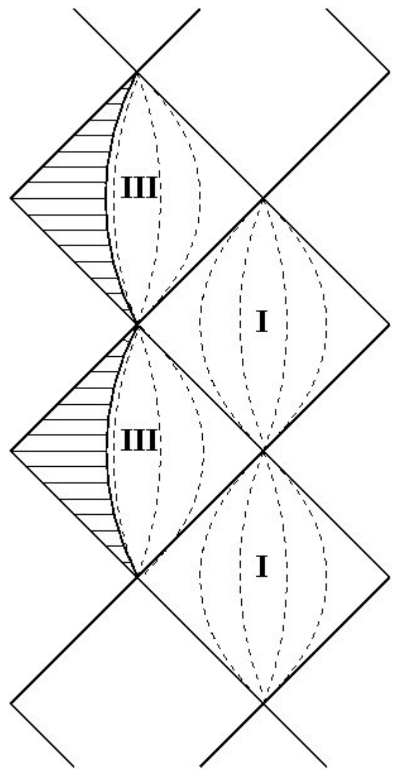

Figure 3.8: The Penrose diagram of the maximal analytic extension of the extreme Reissner-Nordtröm space-time.

i.e.

$$r_* = r + 2M \log |r - M| - \frac{M^2}{r - M} .$$

This has a very different behaviour at the horizon from the sub-extremal case.

A similar difference appears in the spacelike geometry of extreme Reissner-Nordström black holes. We note that

$$\int_M^{r_0} \frac{1}{\sqrt{F(r)}} \mathrm{d}r = \int_M^{r_0} \frac{r}{\sqrt{(r - M)^2}} \mathrm{d}r = \infty .$$

The horizon is at infinite spacelike distance from any point in block I and from any point in block III (the singularity remains of course at finite spacelike distance from points in block III).

This seems to indicate that the maximal analytic extension of an extreme Reissner-Nordström black hole should be more puzzling than in the sub-extremal case. The Penrose diagram of the maximal analytic extension of the extremal Reissner-Nordtröm space-time is shown in figure 3.8. The structure is not so surprising, but the "spacelike horizon", i.e. $r = M$ for finite values of t is not a part of the space-time.

Super-extremal case : $M < |Q|$

There is only one block with a singularity at $r = 0$. The space-time is not extendible.

3.3.2 De Sitter-Schwarzschild metrics

We now study the metric (3.37) with $Q = 0$ and $\Lambda > 0$. The function $F(r)$ has the expression

$$F(r) = 1 - \frac{2M}{r} - \Lambda r^2 = -\frac{\Lambda r^3 - r + 2M}{r}\,.$$

The derivative $1 - 3\Lambda r^2$ of the numerator only vanishes on $]0, +\infty[$ for $r = 1/\sqrt{3\Lambda}$; it is positive in $]0, 1/\sqrt{3\Lambda}[$ and negative for $r > 1/\sqrt{3\Lambda}$. The value of the numerator at $r = 1/\sqrt{3\Lambda}$ is

$$\frac{1}{\sqrt{3\Lambda}} - \frac{1}{3\sqrt{3\Lambda}} - 2M = 2\left(\frac{1}{\sqrt{27\Lambda}} - M\right)\,.$$

So there are three distinct situations.

- $27\Lambda M^2 < 1$: the function F has two zeros $0 < r_- < r_+ < +\infty$, there are two horizons and three blocks ; since F is positive between r_- and r_+, block II (the domain between the horizons) is static with ∂_t Killing, timelike and orthogonal to the Cauchy hypersurfaces of constant t. Block I ($r \in]0, r_-[$) is dynamic, it is the inside of the black hole or of the white hole. Block III ($r \in]r_+, +\infty[$) is also dynamic, either in expansion or in contraction. The horizon $r = r_+$ is called the cosmological horizon. The singularity at $r = 0$ is spacelike as in the Schwarzschild case.

- $27\Lambda M^2 = 1$: the function F has only one double zero at $r = 1/\sqrt{3\Lambda} = 3M$ and F is negative on either side. Block II vanishes and we only have blocks I and III. The singularity at $r = 0$ is spacelike.

- $27\Lambda M^2 > 1$: F has no zero and is negative everywhere, there is a naked singularity in a dynamic universe. The singularity at $r = 0$ is spacelike.

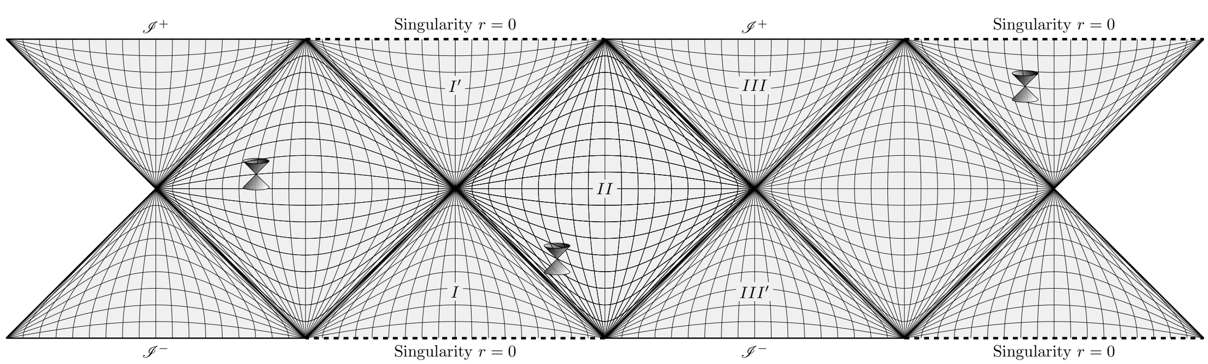

Figure 3.9: Penrose diagram of part of the infinite strip of the maximal analytic extension of de Sitter-Schwarzschild space-time in the case $27\Lambda M^2 < 1$.

We give in figures 3.9, 3.10 and 3.11 the Penrose diagrams of the De Sitter-Schwarzschild space-time in the cases $27\Lambda M^2 < 1$, $27\Lambda M^2 = 1$ and $27\Lambda M^2 > 1$.

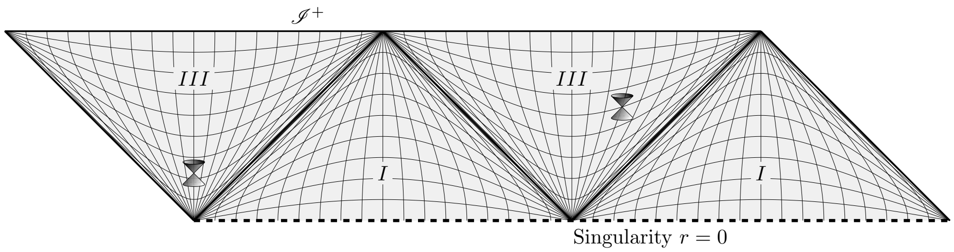

Figure 3.10: Penrose diagram of part of the infinite strip of the maximal analytic extension of de Sitter-Schwarzschild space-time in the case $27\Lambda M^2 = 1$.

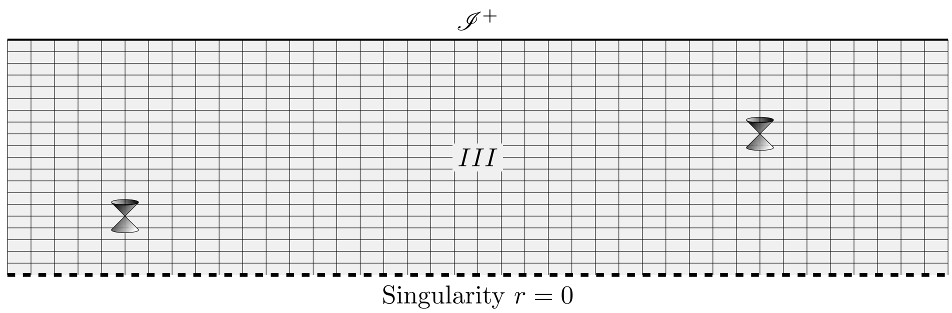

Figure 3.11: Penrose diagram of De Sitter-Schwarzschild space-time in the case $27\Lambda M^2 > 1$.

3.3.3 Anti-de Sitter-Schwarzschild metrics

3.4 The Kerr metric

The Kerr metric is another extension of the Schwarzschild metric : it is no longer spherically symmetric. The additional parameter is the angular momentum per unit mass. It is a solution of the Einstein vacuum equations that describes a rotating uncharged black hole ; in Boyer-Lindquist coordinates on $\mathbb{R}_t \times \mathbb{R}_r \times S^2_\omega$, it takes the form

$$\begin{aligned} g_{\mu\nu}\mathrm{d}x^\mu \mathrm{d}x^\nu = \quad & \left(1 - \frac{2Mr}{\rho^2}\right)\mathrm{d}t^2 + \frac{2a\sin^2\theta\left(r^2+a^2-\Delta\right)}{\rho^2}\mathrm{d}t\mathrm{d}\varphi - \frac{\rho^2}{\Delta}\mathrm{d}r^2 \\ & -\rho^2\mathrm{d}\theta^2 - \left(\frac{\left(r^2+a^2\right)\rho^2 + 2Mra^2\sin^2\theta}{\rho^2}\right)\sin^2\theta\,\mathrm{d}\varphi^2, \end{aligned} \tag{3.39}$$

$$\rho^2 = r^2 + a^2\cos^2\theta,\ \Delta = r^2 - 2Mr + a^2,$$

where a is the angular momentum per unit mass and $M > 0$ is the mass of the black hole. The black hole rotates around the axis going through its North and South poles. This results into a non-zero coefficient $g_{t\varphi}$ that couples the variables t and φ. The function Δ is the analogue of $r^2(1-2M/r)$ in Schwarzschild's space-time ; it defines the horizons as the sets of points where $\Delta = 0$. These horizons appear as singularities in the expression (3.39) above, but they are merely coordinate singularities, the metric can be extended smoothly through them. The only true curvature singularity of the metric is the equatorial ring defined by $\rho^2 = 0$, i.e. $r = 0$ and $\theta = \pi/2$. There are three types of Kerr space-times depending on the respective importance of the rotation and the mass :

- Slow Kerr space-time for $0 < |a| < M$ (the case $a = 0$ reduces to the Schwarzschild metric). Δ has two real roots r_- and r_+ :

$$0 < r_- = M - \sqrt{M^2 - a^2} < M < r_+ = M + \sqrt{M^2 - a^2} < 2M, \qquad (3.40)$$

  so there are two horizons on either side of the sphere $\{r = M\}$.

- Extreme Kerr space-time for $|a| = M$. M is then a double root for Δ and the sphere $\{r = M\}$ is the only horizon.

- Fast Kerr space-time for $|a| > M$. Δ has no real root and the space-time has no horizon. There is no black hole in this case, the ring singularity is a naked singularity.

We consider only the case of slow Kerr metrics. Horizons separate the space-time in connected regions called Boyer-Lindquist blocks :

Block I is the exterior of the black hole $\{r > r_+\}$. It is the simplest of all three blocks. In this region, the vectors $\partial/\partial r$, $\partial/\partial\theta$, $\partial/\partial\varphi$ are spacelike and, for $r >> 1$, $\partial/\partial t$ is timelike. However, block I contains a region called the ergosphere in which $g_{tt} < 0$ which means that $\partial/\partial t$ is spacelike. We shall see this in more details below. Inside $\mathcal{E}$, the effects of the rotation are extreme and along every future-oriented non spacelike curve, the quantity $a\varphi$ is strictly increasing.

Block I, like any Boyer-Lindquist block, is not stationary, i.e. there is no timelike Killing vector field globally defined on it. However, the exterior of the ergosphere is stationary, and even absolutely stationary, since $\partial/\partial t$ is the unique (up to multiplication by a constant) timelike Killing vector field globally defined there. Also, every point in block I, even inside the ergosphere, has a stationary neighbourhood.

Block II is the region between the outer and inner horizons $\{r_- < r < r_+\}$; it only exists in the slow case. $\partial/\partial r$ is timelike there and $\partial/\partial t$, $\partial/\partial\theta$, $\partial/\partial\varphi$ are spacelike. It is a dynamic domain where the inertial frames are dragged towards the inner horizon (the time orientation implicit in this description is such that $\partial/\partial r$ is past pointing).

Block III lies beyond the inner horizon $\{-\infty < r < r_-\}$. It contains another ergosphere

$$\mathcal{E}' = \left\{ (t, r, \theta, \varphi) \ ; \ M - \sqrt{M^2 - a^2 \cos^2\theta} < r < r_- \right\},$$

the ring singularity and a time machine (being the only region where $\partial/\partial\varphi$ is timelike) which allows any two points in block III to be joined by a future-oriented timelike curve. Hence, not only is block III not stationary, it is not causal either.

The ergosphere is in fact not restricted to Block I; it is defined as the region where $g_{tt} < 0$, i.e. where $r^2 - 2Mr + a^2\cos^2\theta < 0$. It is therefore given by

$$\mathcal{E} = \left\{(t,r,\theta,\varphi)\ ;\ M - \sqrt{M^2 - a^2\cos^2\theta} < r < M + \sqrt{M^2 - a^2\cos^2\theta}\right\}.$$

It covers the whole of Block II and overlaps Blocks I and III. Its intersection with Block I can be used to devise a process – the Penrose process (see below) – that allows to extract energy from the underlying geometry (the usual way of understanding this from the physical point of view is that one extracts energy from the black hole).

For a detailed description of the geometry of Kerr black holes, see [41].

3.4.1 The exterior of the black hole

In this section, we study block I from the point of view of an observer who is static with respect to infinity. The perception of such observers is limited to block I and is described by the time function t of the Boyer-Lindquist coordinates. Just as in the Schwarzschild case, light rays in block I can only reach the horizon when t becomes infinite. To illustrate this property, we consider the principal null geodesics which play a role similar to the radial null geodesics on the Schwarzschild space-time. They are defined by

$$\dot{r} = \pm 1,\ \dot{\theta} = 0\,,\ \dot{\varphi} = \frac{a}{\Delta}\,,\ \dot{t} = \frac{r^2 + a^2}{\Delta}\,.$$

Introducing a new coordinate r_* such that

$$\frac{\mathrm{d}r_*}{\mathrm{d}r} = \frac{r^2 + a^2}{\Delta} > 0 \text{ on }]r_+, +\infty[$$

we get

$$\dot{r}_* = \pm \dot{t}$$

and therefore, along a principal null geodesic we must have

$$t = \pm r_* + C\,.$$

The horizon $r = r_+$ corresponds to $r_* \to -\infty$ and is consequently reached only when t becomes infinite.

We study the geometry of $\{t = \text{constant}\}$ slices ; their extrinsic geometry is non trivial and even singular at the horizon.

We denote by $\mathcal{M}$ the space-time outside the black hole and we choose the foliation of $\mathcal{M}$ by the level hypersurfaces of the time-function t :

$$\Sigma_t = \{t\}\times]r_+, +\infty[_r \times S^2_{\theta,\varphi}\,. \tag{3.41}$$

For each t, the hypersurface Σ_t is spacelike since at each point, its tangent plane is spanned by the three spacelike vectors $\frac{\partial}{\partial r}$, $\frac{\partial}{\partial \theta}$, $\frac{\partial}{\partial \varphi}$. This shows that t is indeed a time function, i.e. its gradient $\nabla^a t$ is a timelike vector field, in spite of the fact that in Boyer-Lindquist coordinates, $\frac{\partial}{\partial t}$ is not everywhere timelike in block I. The time orientation is fixed by deciding that $\nabla^a t$ is future pointing, which is equivalent to saying that $\frac{\partial}{\partial t}$ is future pointing in the region of block I where it is timelike.

3.4.2 The $3+1$ decomposition of the Kerr metric in block I

We perform the 3+1 decomposition of the metric g relative to the foliation $\{\Sigma_t\}_{t\in\mathbb{R}}$. We calculate the expression of the vector

$$T^a = \frac{\nabla^a t}{|\nabla t|}$$

in Boyer-Lindquist coordinates. To do this, we look for a future pointing timelike vector field U^a orthogonal to Σ_t at each point and we normalize it to obtain T^a (we could also calculate the inverse metric and apply it to $\mathrm{d}t$). The time orientation yields that t increases along all timelike future pointing curves, hence we choose U^a of the form

$$U^a\partial_a = \frac{\partial}{\partial t} + A\frac{\partial}{\partial r} + B\frac{\partial}{\partial \theta} + C\frac{\partial}{\partial \varphi}$$

and imposing that U^a should be everywhere g-orthogonal to $\frac{\partial}{\partial r}$, $\frac{\partial}{\partial \theta}$ and $\frac{\partial}{\partial \varphi}$, we obtain

$$U^a\partial_a = \frac{\partial}{\partial t} - \frac{g_{t\varphi}}{g_{\varphi\varphi}}\frac{\partial}{\partial \varphi} = \frac{\partial}{\partial t} + \frac{2aMr}{(r^2+a^2)\,\rho^2 + 2Mra^2\sin^2\theta}\frac{\partial}{\partial\varphi}\,. \tag{3.42}$$

We put

$$\alpha(r,\theta) = -\frac{g_{t\varphi}}{g_{\varphi\varphi}} = \frac{2aMr}{(r^2+a^2)\,\rho^2 + 2Mra^2\sin^2\theta}\,. \tag{3.43}$$

The norm of U^a is then given by

$$|U|^2 = U_aU^a = g_{tt} - \frac{(g_{t\varphi})^2}{g_{\varphi\varphi}} = \frac{-\Delta\sin^2\theta}{g_{\varphi\varphi}} = \frac{\Delta\rho^2}{(r^2+a^2)\,\rho^2 + 2Mra^2\sin^2\theta} > 0 \text{ in block I}\,,$$

and the vector T^a is

$$T^a = \frac{U^a}{|U|}\,.$$

If we introduce the vector fields r^a, θ^a, φ^a defined as

$$r^a\partial_a = |g_{rr}|^{-1/2}\frac{\partial}{\partial r}\,,\ \theta^a\partial_a = |g_{\theta\theta}|^{-1/2}\frac{\partial}{\partial \theta}\,,\ \varphi^a\partial_a = |g_{\varphi\varphi}|^{-1/2}\frac{\partial}{\partial\varphi}\,,$$

then $\{T^a, r^a, \theta^a, \varphi^a\}$ is a local orthonormal Lorentz frame in block I ; the metric can therefore be written as

$$g_{ab} = T_aT_b - h_{ab}\,,\ h_{ab} = r_ar_b + \theta_a\theta_b + \varphi_a\varphi_b$$

and the 1-forms T_a, r_a, θ_a and φ_a are given by

$$T_a\mathrm{d}x^a = |U|\,\mathrm{d}t = \sqrt{g_{tt} - \frac{(g_{t\varphi})^2}{g_{\varphi\varphi}}}\mathrm{d}t\,,\ r_a\mathrm{d}x^a = -|g_{rr}|^{1/2}\mathrm{d}r\,,\ \theta_a\mathrm{d}x^a = -|g_{\theta\theta}|^{1/2}\mathrm{d}\theta\,,$$

$$\varphi_a\mathrm{d}x^a = |g_{\varphi\varphi}|^{-1/2}\left(g_{t\varphi}\mathrm{d}t + g_{\varphi\varphi}\mathrm{d}\varphi\right) = -|g_{\varphi\varphi}|^{1/2}\left(\mathrm{d}\varphi - \alpha\,\mathrm{d}t\right)\,.$$

This gives the expression of the lapse function

$$N = |U| = \left(g_{tt} - \frac{(g_{t\varphi})^2}{g_{\varphi\varphi}} \right)^{1/2} = \left(\frac{\Delta\rho^2}{(r^2+a^2)\,\rho^2 + 2Mra^2\sin^2\theta} \right)^{1/2}.$$

In Boyer-Lindquist coordinates, the product structure is associated to the Killing vector field $\frac{\partial}{\partial t}$. If we wish our decomposition of the metric to be useful, we must interpret h_{ab} as a (time dependent) metric on

$$\Sigma \ :=]r_+, +\infty[_r \times S^2_{\theta,\varphi}\,.$$

This requires to choose the product structure associated with T^a. An explicit way of doing this is to define the new coordinates τ, R, Θ, Φ :

$$\tau = t\,,\ \ R = r\,,\ \ \Theta = \theta\,,\ \ \Phi = \varphi - (t-t_0)\alpha(r,\theta)\ (\mathrm{mod}\ 2\pi)$$

for a given $t_0 \in \mathbb{R}$. We obtain the following expression of g :

$$\begin{aligned} g(\tau) &= N^2 \mathrm{d}\tau^2 - h(\tau) \\ &= \left(g_{tt} - \frac{(g_{t\varphi})^2}{g_{\varphi\varphi}} \right) \mathrm{d}\tau^2 + g_{rr}\mathrm{d}R^2 + g_{\theta\theta}\mathrm{d}\Theta^2 + g_{\varphi\varphi}\left(\mathrm{d}\Phi + (\tau - t_0)\frac{\partial\alpha}{\partial R}\mathrm{d}R + (\tau-t_0)\frac{\partial\alpha}{\partial\Theta}\mathrm{d}\Theta \right)^2 \\ &= \left(g_{tt} - \frac{(g_{t\varphi})^2}{g_{\varphi\varphi}} \right) \mathrm{d}\tau^2 + \left(g_{rr} + (\tau-t_0)^2 \left(\frac{\partial\alpha}{\partial R}\right)^2 g_{\varphi\varphi} \right) \mathrm{d}R^2 \\ &\quad + \left(g_{\theta\theta} + (\tau-t_0)^2 \left(\frac{\partial\alpha}{\partial\Theta}\right)^2 g_{\varphi\varphi} \right) \mathrm{d}\Theta^2 + g_{\varphi\varphi}\mathrm{d}\Phi^2 \\ &\quad + 2(\tau-t_0)^2 \frac{\partial\alpha}{\partial R}\frac{\partial\alpha}{\partial\Theta} g_{\varphi\varphi}\mathrm{d}R\mathrm{d}\Theta + 2(\tau-t_0)\frac{\partial\alpha}{\partial R}g_{\varphi\varphi}\mathrm{d}R\mathrm{d}\Phi + 2(\tau-t_0)\frac{\partial\alpha}{\partial\Theta}g_{\varphi\varphi}\mathrm{d}\Theta\mathrm{d}\Phi\,. \end{aligned} \tag{3.44}$$

Note that for these new variables, we have

$$\frac{\partial}{\partial\tau} = U^a\partial_a\,,\ \frac{\partial}{\partial R} = \frac{\partial}{\partial r}\,,\ \frac{\partial}{\partial\Theta} = \frac{\partial}{\partial\theta}\,,\ \frac{\partial}{\partial\Phi} = \frac{\partial}{\partial\varphi}\,,$$

$$T^a\partial_a = \frac{\sqrt{2}}{|U|}\frac{\partial}{\partial\tau} = \frac{2}{N}\frac{\partial}{\partial\tau}\,.$$

Remark 3.3. *The quantity α is the local rotation speed of the space-time. We see that the function α has no singularity at r_+ (in fact it is only singular at the boundary of the time machine). And for $r = r_+$ the function α no longer depends on θ, indeed*

$$\begin{aligned} \alpha(r_+,\theta) &= \frac{2aMr_+}{(r_+^2+a^2)(r_+^2+a^2\cos^2\theta) + 2Mr_+a^2\sin^2\theta} \\ &= \frac{2aMr_+}{(2Mr_+)(r_+^2+a^2\cos^2\theta+a^2\sin^2\theta)} \quad \textit{since } r_+^2+a^2 = 2Mr_+\,, \\ &= \frac{a}{r_+^2+a^2}\,. \end{aligned}$$

The rotation speed of the outer horizon is the same everywhere on the horizon, it does not depend on the latitude. The same is true of the inner horizon with r_+ replaced by r_-.

The intrinsic and extrinsic geometry of the slices

All slices Σ_τ, $\tau \in \mathbb{R}$ have the same geometry (both intrinsic and extrinsic) since in Boyer-Lindquist coordinates, the metric g is independent of t ($\frac{\partial}{\partial t}$ is a Killing vector field). We consider a generic slice $(\Sigma, h(\tau_0))$ and we choose $t_0 = \tau_0$ in order to simplify the expression of $h(\tau_0)$:

$$h(\tau_0) = -g_{rr}\mathrm{d}R^2 - g_{\theta\theta}\mathrm{d}\Theta^2 - g_{\varphi\varphi}\mathrm{d}\Phi^2$$

$$= \frac{\rho^2}{\Delta}\mathrm{d}R^2 + \rho^2\mathrm{d}\Theta^2 + \left[\frac{(R^2+a^2)\rho^2 + 2MRa^2\sin^2\Theta}{\rho^2}\right]\sin^2\Theta\,\mathrm{d}\Phi^2,$$

$$\rho^2 = R^2 + a^2\cos^2\Theta\,,\ \Delta = R^2 - 2MR + a^2\,.$$

The coefficient ρ^2/Δ is singular at the horizon $H = \{r_+\}_R \times S^2_{\Theta,\Phi}$; we introduce a new radial coordinate to show that the metric $h(\tau_0)$ can be extended smoothly through H. Putting

$$F(R)\ := \frac{\Delta}{R^2} = 1 - \frac{2M}{R} + \frac{a^2}{R^2} = \frac{(R-r_+)(R-r_-)}{R^2}\,,$$

we define $u(R)$ for $R \in [r_+, +\infty[$ by

$$u(R)\ := \int_{r_+}^{R} F^{-1/2}(s)\mathrm{d}s\,.$$

(Note that for extreme Kerr space-time, we would have $r_+ = r_- = M$ and consequently, the integral defining $u(R)$ would diverge. Hence, the h-distance to the horizon would be everywhere infinite in block I.) The function u of R is continuous strictly increasing from $[r_+, +\infty[$ onto $[0, +\infty[$, it is $\mathcal{C}^\infty$ on $]r_+, +\infty[$ but is not differentiable at r_+. As in the Schwarzschild case, we easily show the following result ; the proof is identical to that of lemma 3.1 and we do not repeat it here :

Lemma 3.2. *The inverse function $u \longmapsto R(u)$ is smooth from $[0, +\infty[$ onto $[r_+, +\infty[$ and all its derivatives are uniformly bounded on $[0, +\infty[$.*

Lemma 3.2 will allow us to prove that each slice is a smooth manifold with boundary H and that the lapse function is smooth on $\bar{\Sigma}$. The following corollary expresses these properties as well as the fact that $h(\tau)$ depends regularly on τ :

Corollary 3.1. *The manifold*

$$\left(\bar{\Sigma} = [0, +\infty[_u \times S^2_{\Theta,\Phi}, h(\tau_0)\right)$$

is a smooth manifold with boundary. The lapse function N, which is independent of τ, is regular and uniformly bounded on $\bar{\Sigma}$ as well as all its derivatives. Moreover, the metric $h(\tau)$ is a smooth function of τ ; to be more explicit, we have

$$h_{ab} \in \mathcal{C}^\infty\left(\mathbb{R}_\tau\ ; \mathcal{C}^\infty_b\left(\bar{\Sigma}\ ; T_{ab}\mathcal{M}\right)\right)\,,\ h^{ab} \in \mathcal{C}^\infty\left(\mathbb{R}_\tau\ ; \mathcal{C}^\infty_b\left(\bar{\Sigma}\ ; T^{ab}\mathcal{M}\right)\right)\,.$$

Remark 3.4. *The extrinsic curvature*

$$K_{ab} = -\mathcal{L}_T(h_{ab})$$

is singular at the horizon, however,

$$NK_{ab} \in \mathcal{C}^\infty\left(\mathbb{R}_\tau\ ; \mathcal{C}^\infty_b\left(\bar\Sigma\ ; T_{ab}\mathcal{M}\right)\right)\ .$$

<u>*Proof of corollary 3.1 :*</u> We write the metric $h(\tau_0)$ in the form

$$h(\tau_0) = \frac{\rho^2}{R^2}\mathrm{d}u^2 + \frac{\rho^2}{(1+u)^2}(1+u)^2\mathrm{d}\Theta^2 + \left[\frac{(R^2+a^2)\rho^2 + 2MRa^2\sin^2\Theta}{\rho^2(1+u)^2}\right](1+u)^2\sin^2\Theta\,\mathrm{d}\Phi^2.$$

The functions

$$\frac{\rho^2}{R^2}\,,\ \frac{\rho^2}{(1+u)^2}\,,\ \frac{(R^2+a^2)\rho^2 + 2MRa^2\sin^2\Theta}{\rho^2(1+u)^2}$$

are smooth on $\bar\Sigma$, positive, uniformly bounded as well as all their derivatives and uniformly bounded away from zero. Hence, $h(\tau_0)$ is a smooth, symmetric, positive definite 2-form on $\bar\Sigma$, uniformly controlled below and above by the euclidian metric on $\bar\Sigma$ considered as $\mathbb{R}^3 \setminus B(0,1)$:

$$\mathrm{d}u^2 + (1+u)^2\mathrm{d}\Theta^2 + (1+u)^2\sin^2\Theta\,\mathrm{d}\Phi^2.$$

This shows in particular that $(\bar\Sigma, h(\tau_0))$ is a smooth Riemannian manifold with boundary H. Given a regular coordinate system on $\bar\Sigma$, say the underlying euclidian coordinates on $\mathbb{R}^3\setminus B(0,1)$, the 3×3 matrices h_{ij} and h^{ij}, representing the metric $h(\tau_0)$ and its inverse in this coordinate basis, are smooth and bounded on $\bar\Sigma$ as well as all their derivatives. This is expressed more intrinsically by

$$h_{ab}(\tau_0) \in \mathcal{C}^\infty_b\left(\bar\Sigma\ ; T_{ab}\mathcal{M}\right)\ ,\ \ h^{ab}(\tau_0) \in \mathcal{C}^\infty_b\left(\bar\Sigma\ ; T^{ab}\mathcal{M}\right)\ .$$

The lapse function N is given by

$$N(R,\Theta) = \left(\frac{2R^2\rho^2}{(R^2+a^2)\,\rho^2 + 2MRa^2\sin^2\Theta}\right)^{1/2} F^{1/2}.$$

It is the result of the multiplication of $F^{1/2}$ by a smooth function on $\bar\Sigma$, uniformly bounded as well as all its derivatives and uniformly bounded away from zero. Therefore, as a trivial consequence of lemma 3.2 and $\frac{\mathrm{d}R}{\mathrm{d}u} = F^{1/2}$, we have

$$N \in \mathcal{C}^\infty_b(\bar\Sigma)\,.$$

We now study the regularity of $h(\tau)$ with respect to τ. Let us consider the expressions of $h(\tau)$ and $h(\tau_0)$ in the coordinate system R, Θ, Φ with $t_0 = \tau_0$:

$$h(\tau) = -g_{rr}\mathrm{d}R^2 - g_{\theta\theta}\mathrm{d}\Theta^2 - g_{\varphi\varphi}\left(\mathrm{d}\Phi + (\tau-\tau_0)\frac{\partial\alpha}{\partial R}\mathrm{d}R + (\tau-\tau_0)\frac{\partial\alpha}{\partial\Theta}\mathrm{d}\Theta\right)^2,$$

$$h(\tau_0) = -g_{rr}\mathrm{d}R^2 - g_{\theta\theta}\mathrm{d}\Theta^2 - g_{\varphi\varphi}\mathrm{d}\Phi^2.$$

Putting

$$\tilde{\Phi} = \Phi + (\tau - \tau_0)\alpha(R,\Theta) \pmod{2\pi}\,,$$

we have

$$h(\tau) = -g_{rr}\mathrm{d}R^2 - g_{\theta\theta}\mathrm{d}\Theta^2 - g_{\varphi\varphi}\mathrm{d}\tilde{\Phi}^2.$$

$h(\tau)$ is obtained from $h(\tau_0)$ by a rotation around the axis of the black hole whose angle (depending on τ, R and Θ) is

$$(\tau - \tau_0)\alpha(R,\Theta) = -(\tau - \tau_0)\frac{g_{t\varphi}(R,\Theta)}{g_{\varphi\varphi}(R,\Theta)}\,.$$

The function $\alpha(R,\Theta)$ is smooth on $\bar{\Sigma}$ and bounded as well as all its derivatives. Denoting by $G(\tau-\tau_0)$ the $\mathcal{C}^\infty$-diffeomorphism of $\bar{\Sigma}$

$$G(\tau - \tau_0) \;:\; (R,\Theta,\Phi) \longmapsto (R,\Theta,\Phi + (\tau-\tau_0)\alpha(R,\Theta))\,,$$

we have

$$h_{ab}(\tau) = h_{ab}(\tau_0)\circ G(\tau-\tau_0)\,,\; h^{ab}(\tau) = h^{ab}(\tau_0)\circ G(\tau-\tau_0)\,.$$

This entails

$$h_{ab} \in \mathcal{C}^\infty\left(\mathbb{R}_\tau\,;\mathcal{C}^\infty_b\left(\bar{\Sigma}\,;T_{ab}\mathcal{M}\right)\right)\,,\; h^{ab} \in \mathcal{C}^\infty\left(\mathbb{R}_\tau\,;\mathcal{C}^\infty_b\left(\bar{\Sigma}\,;T^{ab}\mathcal{M}\right)\right)$$

and concludes the proof of corollary 3.1. $\square$

The Penrose process

If we consider a geodesic γ in block I, its energy as perceived by an observer static at infinity is

$$E_\gamma(s) := \langle\dot{\gamma}(s)\,\partial_t\rangle\,.$$

It is conserved since ∂_t is Killing, but since ∂_t is spacelike inside the ergosphere, the energy of a timelike or null geodesic is allowed to be negative there. This has led Roger Penrose to imagine a process by which one can extract energy from the ergoregion: a particle is sent towards the black hole; its energy is of course positive; once inside the ergosphere, it disintegrates into two particles in such a way that one of them has negative energy (i.e. its 4-velocity is oriented so that its inner product with ∂_t is negative); the one with negative energy cannot leave the ergosphere and we assume that it falls into the black hole and that the other one comes out of the ergosphere; by conservation of the total energy, the energy of the particle that comes back out is larger than that of the particle we originally sent in. This is called the Penrose process.

This has given Subramanian Chandrasekhar the idea of an industrial city built in orbit around a black hole and drawing all its energy from the ergoregion. Shuttles are sent inside the ergosphere loaded with the city's litter. Once inside, they eject the litter with an angular speed such that the energy of the litter bags is negative. The shuttles then come back lighter than they left but with more energy. A wheel slowing down the shuttles on their return transforms the added energy into electricity.

How can we, at any given point of the ergosphere, choose a timelike direction τ^a such that $g(\tau,\partial_t)<0$? Recall that for any spacelike vector, we can find two future oriented timelike vectors such that their inner products with the spacelike vector have opposite signs. In particular, this guarantees that there is a future timelike direction τ^a (and by continuity an open set of future timelike directions) such that $g(\tau,\partial_t)<0$.

Superradiance

The phenomenon of superradiance is the analogue of the Penrose process at the level of fields.

$$\begin{aligned}\mathcal{F}_{\Sigma_t}(J) &= \frac{1}{2}\int_{\Sigma_t}\left(|\partial_t\phi|^2+|\partial_{r_*}\phi|^2+\frac{F}{r^2}|\nabla_{S^2}\phi|^2\right.\\ &\qquad\left.+\left(\frac{FF'}{r}+Fm^2-\frac{q^2Q^2}{r^2}\right)|\phi|^2\right)\sin\theta \mathrm{d}r_*\mathrm{d}\theta\mathrm{d}\varphi\end{aligned}$$

TO BE CONTINUED...

3.4.3 Maximal extension of Kerr's space-time

The global geometry of Kerr's space-time (and in particular slow Kerr) is far more complex than that of Schwarzschild's space-time. An entire chapter of B. O'Neill's book [41] is devoted to the construction of the maximal extension. Our purpose in this section is to describe this construction schematically and to point out so-called Kruskal domains in maximal slow Kerr space-time.

Kerr-star and star-Kerr coordinates

Just as we did in the Schwarzschild case, we choose a coordinate system which will allow us to represent globally the whole of Kerr's space-time. This choice is guided by the following physical consideration : if a particle is to pass from block I to block II across the outer horizon and then from block II to block III across the inner horizon, its most direct course is to follow an incoming principal null geodesic. The whole idea of the Kerr-star coordinate system is to turn incoming principal null geodesics into coordinate lines. Such geodesics are defined on all three blocks in Boyer-Lindquist coordinates by

$$\dot{t}=\frac{r^2+a^2}{\Delta}\,,\ \dot{r}=-1\,,\ \dot{\theta}=0\,,\ \dot{\varphi}=\frac{a}{\Delta}\,.$$

Keeping the coordinates r and θ, we introduce two new coordinates t^* and φ^* of the form

$$t^*=t+T(r)\,,\ \varphi^*=\varphi+A(r)$$

where the functions T and A are required to satisfy

$$\frac{\mathrm{d}T}{\mathrm{d}r}=\frac{r^2+a^2}{\Delta}\,,\ \frac{\mathrm{d}A}{\mathrm{d}r}=\frac{a}{\Delta}\,.$$

(t^*,r,θ,φ^*) defines a coordinate system in each Boyer-Lindquist block[3], called Kerr-star coordinates, in which the incoming principal null geodesics are described by

$$\dot{r}=-1\,,\ \dot{\theta}=0\,,\ \dot{t^*}=\dot{t}+\frac{\mathrm{d}T}{\mathrm{d}r}\dot{r}=0\,,\ \dot{\varphi^*}=\dot{\varphi}+\frac{\mathrm{d}A}{\mathrm{d}r}\dot{r}=0\,,$$

[3] with the exception of the axis ($\theta=0$ and $\theta=\pi$) ; this coordinate singularity can be dealt with simply (see [41] lemma 2.2.2), we shall systematically ignore it.

i.e. they are the r coordinate curves parametrized by $s=-r$ (or $-r+C$). The expression of the Kerr metric in Kerr-star coordinates is given by

$$g = g_{tt}\mathrm{d}t^{*2} + 2g_{t\varphi}\mathrm{d}t^*\mathrm{d}\varphi^* + g_{\varphi\varphi}\mathrm{d}\varphi^{*2} - \rho^2\mathrm{d}\theta^2 - 2\mathrm{d}t^*\mathrm{d}r + 2a\sin^2\theta\,\mathrm{d}\varphi^*\mathrm{d}r\,, \tag{3.45}$$

where g_{tt}, $g_{t\varphi}$, $g_{\varphi\varphi}$ and $g_{\theta\theta}=-\rho^2$ are as defined in (3.39), i.e.

$$g_{tt} = \left(1-\frac{2Mr}{\rho^2}\right)\,,\; g_{t\varphi} = \frac{a\sin^2\theta\left(r^2+a^2-\Delta\right)}{\rho^2}\,,$$

$$g_{\varphi\varphi} = -\left(\frac{\left(r^2+a^2\right)\rho^2+2Mra^2\sin^2\theta}{\rho^2}\right)\sin^2\theta\,,\; \rho^2=r^2+a^2\cos^2\theta\,.$$

We see from (3.45) that the metric g is smooth on all three blocks, with the exception of the ring singularity $\{\rho^2=0\}=\{r=0 \text{ and } \theta=\pi/2\}$ in block III, and across both horizons (the component g_{rr} in Boyer-Lindquist coordinates was the only component of g to be singular at the horizons and it does not appear in (3.45)).

Kerr-star space-time is defined as the manifold

$$\mathcal{M}^* = \mathbb{R}_{t^*}\times\mathbb{R}_r\times S^2_{\theta,\varphi^*}\setminus\left\{(t^*,r,\theta,\varphi^*)\,;\; r=0 \text{ and } \theta=\frac{\pi}{2}\right\}$$

equipped with the smooth metric (3.45) and with the time orientation such that the null coordinate vector field $-\frac{\partial}{\partial r}$, defined and smooth on the whole of $\mathcal{M}^*$ and whose integral lines are the incoming principal null geodesics, be future oriented. This time orientation is consistent with the fact that, in Boyer-Lindquist coordinates, the Killing vector field $\frac{\partial}{\partial t}$ is future oriented outside the ergosphere in block I and also with the description of block II given at the beginning of the chapter, with $-\frac{\partial}{\partial r}$ (in Boyer-Lindquist coordinates) future pointing. This space-time contains all three blocks, glued smoothly at the horizons by the requirement that incoming principal null geodesics should cross horizons smoothly and that their orientation defines the time orientation. Block II is thus glued to block I in such a way that it lies in the future of block I and similarly, block III lies in the future of block II. The horizons $\{r=r_+\}$ and $\{r=r_-\}$ are smooth null hypersurfaces of $(\mathcal{M}^*,g)$. The fact that they are null is easily shown considering the metric induced by g on hypersurfaces of constant r

$$g_r = g_{tt}\mathrm{d}t^{*2} + 2g_{t\varphi}\mathrm{d}t^*\mathrm{d}\varphi^* + g_{\varphi\varphi}\mathrm{d}\varphi^{*2} - \rho^2\mathrm{d}\theta^2\,.$$

This induced metric has determinent

$$\det(g_r) = -\rho^2\left(g_{tt}g_{\varphi\varphi}-(g_{t\varphi})^2\right) = \rho^2\Delta\sin^2\theta$$

and thus degenerates for $\Delta=0$, i.e. at the horizons. See figure 3.12 for a Penrose diagram of Kerr-star space-time.

This construction is similar to what we did in Schwarzschild's space-time, when we first used Kruskal-Szekeres coordinates to show that the metric could be extended smoothly across the horizon. In the Schwarzschild case, the maximal extension of the space-time followed naturally by extending the domain of definition of the Kruskal-Szekeres coordinate system. This we cannot do here since the domain of definition of Kerr-star coordinates is already maximal. We shall need to use other coordinate systems which will allow us to glue Boyer-Lindquist blocks in different manners.

Kerr-star coordinates were defined by modifying Boyer-Lindquist coordinates so that incoming principal null geodesics could become coordinate lines. Using outgoing principal null geodesics instead of the incoming ones, we obtain the star-Kerr coordinate system. These geodesics are defined on all three blocks in Boyer-Lindquist coordinates by

$$\dot{t} = \frac{r^2+a^2}{\Delta}\,,\ \dot{r} = 1\,,\ \dot{\theta} = 0\,,\ \dot{\varphi} = \frac{a}{\Delta}\,.$$

Keeping r and θ, we introduce the new coordinates

$$^*t = t - T(r)\,,\ ^*\varphi = \varphi - A(r)$$

where the functions T and A are the same used to define t^* and φ^*. In the star-Kerr coordinate system $(^*t, r, \theta, ^*\varphi)$, the outgoing principal null geodesics are the r coordinate lines parametrized by $s = r$ and the Kerr metric takes the form

$$\begin{aligned} g = g_{tt}\mathrm{d}(^*t)^2 + 2g_{t\varphi}\mathrm{d}(^*t)\mathrm{d}(^*\varphi) + g_{\varphi\varphi}\mathrm{d}(^*\varphi)^2 - \rho^2\mathrm{d}\theta^2 \\ +2\mathrm{d}(^*t)\mathrm{d}r - 2a\sin^2\theta\,\mathrm{d}(^*\varphi)\mathrm{d}r\,. \end{aligned} \tag{3.46}$$

This gives rise to star-Kerr space-time which is the manifold

$$^*\mathcal{M} = \mathbb{R}_{^*t} \times \mathbb{R}_r \times S^2_{\theta, ^*\varphi} \setminus \left\{ (^*t, r, \theta, ^*\varphi)\,;\ r = 0 \text{ and } \theta = \frac{\pi}{2} \right\}$$

equipped with the smooth metric (3.46) and time orientation such that, in star-Kerr coordinates, the null coordinate vector field $\frac{\partial}{\partial r}$, which is defined and smooth all over $^*\mathcal{M}$ and whose integral lines are the outgoing principal null geodesics, is future pointing. This space-time contains all three blocks, glued together at the horizons which appear as regular null hypersurfaces. The gluing is done by requiring that the outgoing principal null geodesics should cross the horizons smoothly. The time orientation reflects this choice ; it is consistent with the fact that in Boyer-Lindquist coordinates $\frac{\partial}{\partial t}$ is future pointing outside the ergosphere in block I, but incompatible with $-\frac{\partial}{\partial r}$ future oriented in block II : in star-Kerr space-time, the inertial frames in bloc II are dragged outwards from the inner horizon to the outer horizon. There is a canonical isometry between star-Kerr and Kerr-star space-times. This isometry preserves the time orientation of blocks I and III but reverses that of block II. Star-Kerr space-time can be seen as a block I, to the past of which is glued a block II with its time orientation reversed, to the past of which is glued a block III : it describes a "slow Kerr white hole". See figure 3.12 for the Penrose diagram of star-Kerr space-time (II′ refers to a block II with reversed time orientation).

Maximal slow Kerr space-time

The maximal analytic extension of slow Kerr space-time is constructed using both Kerr-star and star-Kerr space-times. We start with Kerr-star space-time : all the incoming principal null geodesics are complete but the outgoing ones are not. The idea is to glue other blocks so as to

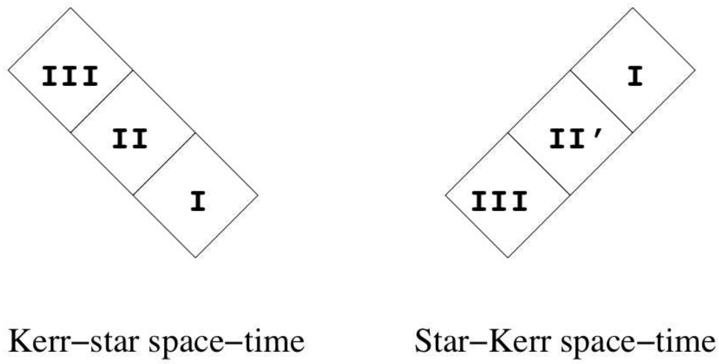

Figure 3.12: Penrose diagrams of Kerr-star and star-Kerr space-times

make the outgoing principal null geodesics complete. The solution for blocks I and III is simple : we consider them as belonging to star-Kerr space-times, i.e. we glue to the future of block III a block II′ followed by a new block I and to the past of block I a block II′ preceded by a new block III. For block II, the situation is trickier ; we also wish to understand block II as part of a star-Kerr space-time, but this is incompatible with the time orientation of block II. The solution is to reverse the time orientation of the whole star-Kerr space-time. We are thus led to gluing to the future of block II a block III′ (block III with its time orientation reversed) and to its past a block I′ (block I with reversed time orientation). The resulting space-time is shown in figure 3.14. We keep on extending this new space-time wherever a family of principal null geodesics is incomplete. The extension is done step by step and is based on the same simple principle : if a family of principle null geodesics is incomplete, it means that the Kerr-star (in the incoming case) or star-Kerr (in the outgoing case) space-time which it generates lacks one or two blocks ; this is cured by gluing the lacking blocks, bearing in mind the consistency of the time orientation of the whole space-time. In this manner, we construct maximal slow Kerr space-time (see figure 3.13) as a reunion of four types of space-times : Kerr-star space-times, Kerr-star with their time orientation reversed, star-Kerr and star-Kerr with their time orientation reversed. Important objects in this maximal extension are the so-called Kruskal domains. They are "diamond shaped" reunions of four contiguous blocks. At their "centre" lies a 2-sphere, referred to as a crossing sphere, where the horizons intersect. Building this crossing sphere rigorously and extending the metric over it are important difficulties in the construction of maximal slow Kerr space-time. This is done by means of Kruskal-Boyer-Lindquist coordinates (see [41] for a fully detailed account). There are two types of Kruskal domains, as shown in figure 3.15. Type II-III contains two copies of block III ; it is not causal, therefore not globally hyperbolic, and contains two timelike singularities (the ring singularity of each block III). Because of the lack of causality, the notion of Cauchy problem is not even meaningful on type II-III domains. Type I-II domains are much more gentle. They are globally hyperbolic and contain no singularity. They can be treated in exactly the same manner as maximal Schwarzschild space-time.

For a type I-II Kruskal domain, we consider a foliation $\{S_\tau\}_{\tau\in\mathbb{R}}$ (see figure 3.16) by Cauchy

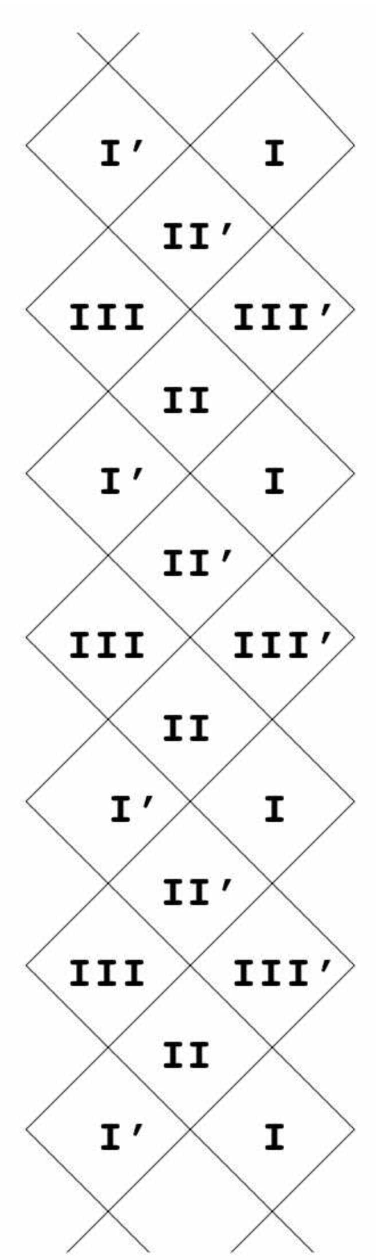

Figure 3.13: Maximal slow Kerr space-time

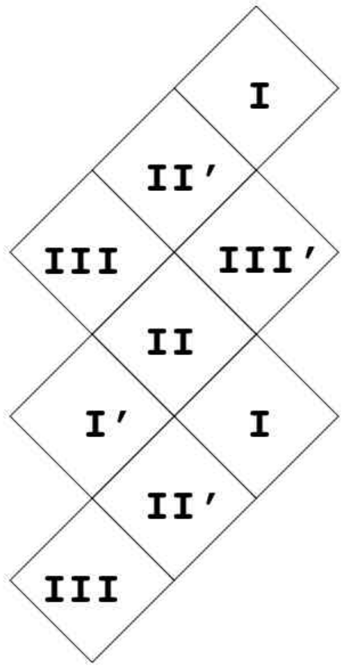

Figure 3.14: First step in the construction of maximal slow Kerr space-time

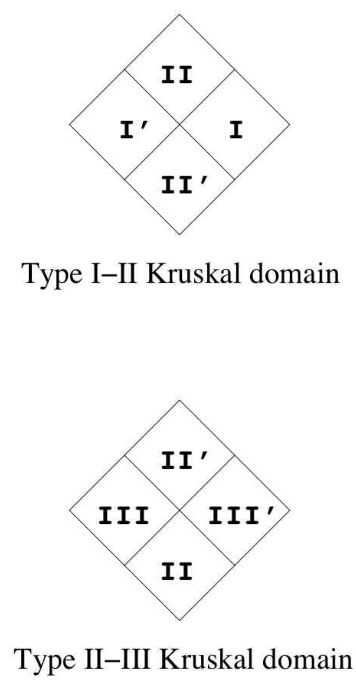

Figure 3.15: The two different types of Kruskal domains

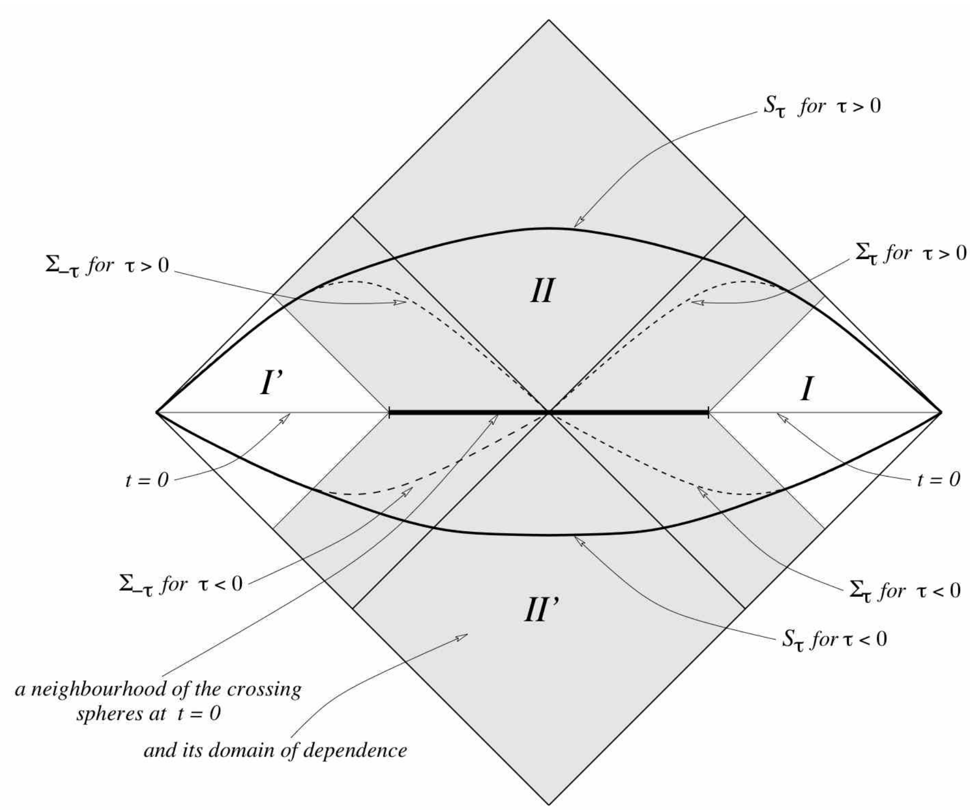

Figure 3.16: Foliation of a type I-II Kruskal domain

hypersurfaces such that, outside the domain of dependence of a neighbourhood of the crossing sphere, for each $\tau \in \mathbb{R}$ the hypersurface S_τ coincides in block I with the level hypersurface $\Sigma_\tau = \{t = \tau\}$ of the time coordinate t of Boyer-Lindquist coordinates and in block I′ with $\Sigma_{-\tau}$ (suffice it to say that the Boyer-Lindquist coordinates in blocks I, II, I′ and II′ are defined unambiguously from the Kruskal-Boyer-Lindquist coordinates defined on the whole domain).

3.5 Asymptotically simple space-times

Asymptotically simple space-times were introduced by Roger Penrose (see for example [44]) as generic models of asymptotically flat space-times. Here, the notion of asymptotic flatness is to be understood in a stronger sense as in the case of Schwarzschild's space-time ; the curvature falls off to zero at infinity in all directions : timelike, spacelike or null. Their definition is formulated in terms of conformal compactification and the "degree of flatness at infinity" is expressed in terms of the regularity of the conformally rescaled metric at the conformal boundary. We shall not worry about the quantitative fall-off of the curvature at infinity here and we therefore only consider asymptotically simple space-times for which the metric is $\mathcal{C}^\infty$ at the conformal boundary.

Definition 3.1. *A space-time $(\mathcal{M}, g)$ is said to be asymptotically simple if $\mathcal{M}$ is diffeomorphic to $\mathbb{R}^4$ and there exist a positive function Ω on $\mathcal{M}$ and a smooth space-time with boundary $(\bar{\mathcal{M}}, \hat{g})$, such that :*

1. *$\hat{g} = \Omega^2 g$ on $\mathcal{M}$, Ω vanishes at the boundary of $\bar{\mathcal{M}}$ but $\mathrm{d}\Omega$ is nowhere zero there ;*
2. *$\mathcal{M}$ is the interior of $\bar{\mathcal{M}}$, the boundary of $\bar{\mathcal{M}}$ is the union of two points $i^\pm$, the past light-cone of i^+ (denoted $\mathscr{I}^+$) and the future light-cone of i^- (denoted $\mathscr{I}^-$) ;*
3. *every inextendible null geodesic acquires a future end-point on $\mathscr{I}^+$ and a past end-point on $\mathscr{I}^-$.*

The place where $\mathscr{I}^+$ and $\mathscr{I}^-$ meet, i.e. spacelike infinity i^0, is a in general a singularity of the conformal structure, it is not part of the "compactified space-time" $\bar{\mathcal{M}}$ which is therefore not compact.

Such space-times were initially considered by many people as rather empty idealizations of "real" asymptotically flat space-times because it was not known whether the Einstein vacuum equations admitted any asymptotically simple solutions. Recently, the works of Chrusciel-Delay, Corvino-Schoen and Klainerman-Nicolò [7, 8, 29] established the existence of a large class of such Einstein space-times. They are generically time-dependent, which prevents the use of standard analytic methods for the construction of a scattering theory.

As already mentionned, the Schwarzschild metric is a model for asymptotically flat space-times containing energy/matter in the following sense : at first order, any such space-time differs at infinity from the flat one by a Schwarzschild-type contribution that falls-off like $1/r$, r being, say, the spacelike geodesic distance from a given timelike curve for a choice of spacelike asymptotically flat slicing. The space-times constructed by Klainerman and Nicolò in [29] are asymptotic to the Schwarzschild metric on each slice in this sense. Such structures are delicate to manipulate analytically because all conservation laws fail and are replaced by "approximate conservation laws" that, essentially, work in the same manner, but require some care and painful calculations. The space-times of Corvino/Schoen-Chrusciel/Delay are simplified versions of the Klainerman-Nicolò universes in that they are exactly diffeomorphic to Schwarzschild's space-time in a neighbourhood of spacelike infinity (in the simplest case, other versions are diffeomorphic to the Kerr space-time near i^0). This means that near spacelike infinity at least, we have the luxury of the symmetries of the Schwarzschild metric which grants us access to exact conservation laws. This makes the conformal scattering construction more clearcut and avoids cumbersome estimates.

CONSTRUCT A COMPACTIFICATION WITH SCALAR CURVATURE VANISHING ON SCRI

Chapter 4

Linear hyperbolic equations on globally hyperbolic space-times

Hyperbolic equations are essentially equations for which the Cauchy problem is locally well-posed. The notion of hyperbolicity was introduced by Jacques Hadamard in the 1920's in his lecture notes on the Cauchy problem for linear partial differential equations (this was later published by Dover in 1952 [19]). Considering a linear partial differential operator of second order with real constant coefficients on $\mathbb{R}^n$,

$$P(D) = \sum_{i,j=1}^{n} a_{ij} D_i D_j + \sum_{i=1}^{n} b_i D_i + c,$$

where the matrix $A = (a_{ij})$ is symmetric, and $D_j = \frac{1}{i}\frac{\partial}{\partial x^j}$, its principal symbol is defined as the quadratic form

$$P_2(\xi) = \sum_{i,j=1}^{n} a_{ij} \xi^i \xi^j = {}^{\mathrm{t}}\xi A \xi,$$

which is the image of the highest order part of $P(D)$ through a Fourier transform. The operator $P(D)$ is said to be

- **elliptic** if $P_2(\xi)$ is positive definite or negative definite,
- **parabolic** if $P_2(\xi)$ is positive or negative semi-definite with exactly one zero eigenvalue,
- **hyperbolic** if $P_2(\xi)$ is indefinite and non degenerate,
- **normal-hyperbolic** if $P_2(\xi)$ has signature $(1, n-1)$ or $(n-1, 1)$.

We see that the level sets of the polynomial $P(\xi)$ are ellipsoids in the elliptic case, paraboloids in the parabolic case and hyperboloids in the hyperbolic case. Note also that when the coefficients depend on x, the operator $P(D)$ can be of different types in different regions.

The notion of normal hyperbolicity is nowadays called hyperbolicity whereas what Hadamard defined as hyperbolic operators are now referred to as ultra-hyperbolic when they are not normal-hyperbolic. The wave equation is part of the normal hyperbolic family.

For scalar operators of higher order, a classic definition is the following. A scalar differential operator of order k on $\mathbb{R}_t \times \mathbb{R}^n_x$

$$P = \sum_{|\alpha|+j\leq k} a_{\alpha,j}(t,x) D_t^j D_x^\alpha$$

where $\alpha \in \mathbb{N}^n$ and $a_{\alpha,j}$ are smooth functions, is said to be *strictly hyperbolic* at (t,x) with respect to ∂_t if its principal symbol

$$P_k(t,x,\tau,\xi) = \sum_{|\alpha|+j=k} a_{\alpha,j}(t,x)\tau^j \xi^\alpha$$

satisfies the following property: $P_k(t,x,1,0) \neq 0$ *and for all nonzero* $\xi \in \mathbb{R}^n$*, the polynomial* $p(\tau) = P_k(t,x,\tau,\xi)$ *admits* k *distinct real roots.*

A typical example of this is the wave equation on flat space-time for which we have

$$P(\tau,\xi) = P_2(\tau,\xi) = -\tau^2 + |\xi|^2.$$

The two distinct real roots here are $\tau = \pm|\xi|$.

Remark 4.1. *The wave equation is generally considered as the fundamental model of hyperbolic equations, although in some sense Dirac's equation is more fundamental and in many ways exhibits simpler features, but the wave equation is a scalar equation and therefore does not require to deal with spinors. Because of this simplicity of structure, we shall follow the usual trend and start with the wave equation in this chapter, before tackling other examples.*

In the definition above, the whole geometrical structure is given. However, when considering partial differential equations with variable coefficients on $\mathbb{R}^{n+1}$ or on curved manifolds, one needs to find a splitting between time and space with respect to which the equation may be hyperbolic. An intrinsic definition can be given.

Definition 4.1 (Hyperbolicity and strict hyperbolicity)**.** *Let* $P(x,D)$ *be a differential operator of order* k *on* $\mathbb{R}^n$

$$P(x,D) = \sum_{|\alpha|\leq k} a_\alpha(x) D^\alpha$$

where the a_α*'s are smooth functions on* $\mathbb{R}^n$*. The principal symbol of* P *is the polynomial in* ξ

$$P_k(x,\xi) = \sum_{|\alpha|=k} a_\alpha(x)\xi^\alpha.$$

The operator $P(x,D)$ *is said to be hyperbolic at* x *with respect to a non zero real vector* N *if* $P_k(x,N) \neq 0$ *and there exists* s_0 *such that, for all* $s < s_0$*,*

$$P(x,\xi + isN) \neq 0 \text{ for all } \xi \in \mathbb{R}^n \text{ and } s < s_0.$$

It is said to be strictly hyperbolic at x *with respect to* N *if* $P_k(x,N) \neq 0$ *and* $P_k(x,\xi + sN)$ *has* k *distinct real roots as a function of* s *for every* $\xi \in \mathbb{R}^n$ *that is not proportional to* N*.*

Remark 4.2. *The zeros of $P_k(x, \xi + sN)$ depend continuously on ξ, so in the strictly hyperbolic case, when ξ is proportional to N, the zeros are still real but they are allowed not to be simple anymore. This means in particular that strict hyperbolicity implies hyperbolicity for the principal part P_k.*

Strict hyperbolicity is a property of the principal symbol and is independent of the lower order terms. In fact, strict hyperbolicity implies hyperbolicity for the whole polynomial independently of the lower order terms.

Let us check that the differential operator for the wave equation on $\mathbb{R} \times \mathbb{R}^n$

$$P(D) = \partial_t^2 - \Delta_x \tag{4.1}$$

satisfies this property with respect to the vector field ∂_t. The principal symbol is (denoting $(\tau, \xi) \in \mathbb{R} \times \mathbb{R}^n$ the dual variables of (t, x))

$$P(\tau, \xi) = -\tau^2 + |\xi|^2.$$

We have $P(\partial_t) = 1$ and for any $(\tau, \xi) \in \mathbb{R} \times \mathbb{R}^n$ such that $\xi \neq 0$ (i.e. (τ, ξ) not proportional to ∂_t)

$$P((\tau, \xi) + s\partial_t) = -(\tau + s)^2 + |\xi|^2$$

has exactly two real roots given by $s = \pm|\xi| - \tau$. Note that when $\xi = 0$, we have a real double root $s = -\tau$. This implies that the function of s

$$P((\tau, \xi) + is\partial_t) = -(\tau + is)^2 + |\xi|^2$$

never vanishes, except at $s = 0$ in the cases where $\tau = \pm|\xi|$.

Remark 4.3. *The operator (4.1) is in fact strictly hyperbolic with respect to any timelike vector.*

All these notions can be found in the books by Lars Hörmander [24] volume 2 for operators with constant coefficients and volume 3 for the case of varying coefficients. For such operators, the Cauchy problem is well-posed in Sobolev spaces. We give here the main theorem to this effect. It is Theorem (23.2.2) in volume 3, p. 393, for strictly hyperbolic pseudo-differential operators, that we have expressed for differential polynomials on $\mathbb{R}^{n+1}$ where the variables are denoted (t, x), $x \in \mathbb{R}^n$.

Theorem 4.1. *Let*

$$P = \sum_{j=1}^{k} Q_j(t, x, D_x) D_t^j$$

be a differential operator of order k on $\mathbb{R}^{n+1}$, i.e.

$$Q_j(t, x, D_x) = \sum_{|\alpha| \leq k-j} a_{j,\alpha}(t, x) D_x^\alpha,$$

such that $Q_k = 1$. Assume that

1. *the coefficients $a_{j,\alpha}$ are uniformly bounded on $\mathbb{R}^{n+1}$ as well as their successive derivatives,*

2. *the principal symbol P_k is stricly hyperbolic at all points with respect to ∂_t,*

3. *P_k is such that the zeros of $\tau \mapsto P_k(t,x,\tau,\xi)$ are uniformly simple for $\xi \neq 0$ in the sense that there exists a constant $C > 0$ such that*

$$\left|\frac{\partial P_k}{\partial \tau}\right| \geq C|\xi|^{k-1}, \quad \textit{if } P_k(t,x,\tau,\xi) = 0.$$

Then, for any $s \in \mathbb{R}$ and $\phi_j \in H^{s+k-1-j}(\mathbb{R}^n)$, $j = 0, 1, ..., k-1$, the Cauchy problem

$$P\phi = 0, \quad \left.\frac{\partial^j \phi}{\partial t^j}\right|_{t=0} = \phi_j, \ j = 0, 1, ..., k-1,$$

has a unique solution

$$\phi \in \bigcap_{j=0}^{k-1} \mathscr{C}^j(\mathbb{R}_t;\ H^{s+k-1-j}(\mathbb{R}^n)).$$

In particular, for smooth, compactly supported data, we have a unique solution in $\mathscr{C}^\infty(\mathbb{R}^{n+1})$.

Note that the third property is also clearly satisfied by the wave equation on $\mathbb{R}^{n+1}$. Indeed, for $P(\tau,\xi) = -\tau^2 + |\xi|^2$, we have

$$\left|\frac{\partial P}{\partial \tau}\right| = 2|\tau| = 2|\xi| \text{ if } P(\tau,\xi) = 0.$$

Sofar, we have only considered scalar operators. For systems, the notion of hyperbolicity is more delicate but there is a class that will be sufficient for us and for which the theory is quite simple: the so-called symmetric hyperbolic systems.

Definition 4.2 (Symmetric hyperbolic system on $\mathbb{R} \times \mathbb{R}^n$)**.** *A first order $d \times d$ system of partial differential equations on $\mathbb{R} \times \mathbb{R}^n$ of the form*

$$\frac{\partial u}{\partial t}(t,x) + \sum_{j=1}^{n} A^j(t,x)\frac{\partial u}{\partial x^j}(t,x) + B(t,x)u(t,x) = 0,$$

where u is a function on $\mathbb{R} \times \mathbb{R}^n$ with values in $\mathbb{C}^d$, is said to be symmetric hyperbolic if the matrices A^j are hermitian.

In the case of constant coefficients, when the matrix B is zero, the system reads

$$\frac{\partial u}{\partial t}(t,x) = -\sum_{j=1}^{n} A^j \frac{\partial u}{\partial x^j}(t,x). \tag{4.2}$$

Note that such a system has an obvious finite propagation speed given by

$$c = \max\{\rho(A^j),\ 1 \leq j \leq n\}$$

where $\rho(A^j)$ denotes the maximum of the absolute values of the eigenvalues of A^j.

Taking the Fourier transform in space, we obtain

$$\frac{\partial \hat{u}}{\partial t}(t,\xi) = -i \sum_{j=1}^{n} A^j \xi_j \hat{u}(t,\xi). \tag{4.3}$$

The matrix

$$A(\xi) = \sum_{j=1}^{n} A^j \xi_j$$

is Hermitian for all $\xi \in \mathbb{R}^n$, therefore, for any $s \in \mathbb{R}$ and $u_0 \in H^s(\mathbb{R}^n)$, the exists a unique $u \in \mathscr{C}(\mathbb{R}_t;\ H^s(\mathbb{R}^n))$ solution to (4.2) such that $u(0) = u_0$. It is given by

$$\hat{u}(t) = e^{-itA(\xi)}\widehat{u_0}.$$

Since the matrix $e^{itA(\xi)}$ is unitary for all t, ξ, the H^s norm of the solution is preserved through time. When B is non zero, we can obtain a priori exponential bounds on the H^s norm of the solution and solve the Cauchy problem by a fixed point method.

In the case with variable coefficients, we also have a well-posed Cauchy problem in all Sobolev spaces (see the books by Benzoni-Gavage and Serre [2] or Racke [46]).

Theorem 4.2. *Consider the system*

$$\frac{\partial u}{\partial t}(t,x) = \sum_{i=1}^{n} A^i(t,x) \frac{\partial u}{\partial x^i}(t,x) + B(t,x)u(t,x), \tag{4.4}$$

where the matrices A^i are Hermitian and A^i and B are smooth and bounded as well as all their derivatives on $\mathbb{R} \times \mathbb{R}^n$. Then for any $s \in \mathbb{R}$ and $u_0 \in H^s(\mathbb{R}^n)$, there exists a unique solution $u \in \mathscr{C}(\mathbb{R}_t;\ H^s(\mathbb{R}^n))$ of (4.4) such that $u(0) = u_0$. In particular, for smooth, compactly supported data, we have a unique solution in $\mathscr{C}^\infty(\mathbb{R}^{n+1})$.

Remark 4.4. *Using local charts and finite speed propagation (see section 4.2.1), Theorems 4.1 and 4.2 remain valid without modification for globally hyperbolic space-times with compact or asymptotically flat spacelike slices. For more general globally hyperbolic space-times, the weaker versions, with smooth compactly supported data and smooth global solutions, are always valid.*

This chapter first deals with hyperbolic equations on Minkowski space-time, starting with the wave equation, then moving on to some particular symmetric hyperbolic system: zero rest-mass field equations.

4.1 The scalar wave equation on flat space-time

The scalar wave equation, in the flat case, is the strictly hyperbolic evolution equation on $\mathbb{R}_t \times \mathbb{R}^n_x$

$$\Box\phi = 0,\ \text{where } \Box = \partial_t^2 - \Delta_x \text{ is the d'Alembertian on } \mathbb{R} \times \mathbb{R}^n. \tag{4.5}$$

We first take a look at integral formulae describing solutions from different viewpoints in all space dimensions. We know from Theorem 4.1 that the Cauchy problem in well-posed in Sobolev spaces but we shall describe more explicit approaches in this simple case as well as derive basic properties of the solutions, such as finite speed propagation.

4.1.1 Integral formulae

There are integral formulae giving either the general solution of equation (4.5) or of the Cauchy problem

$$\partial_t^2 \phi - \Delta_x \phi = 0 \text{ on } \mathbb{R}_t \times \mathbb{R}_x^n \,,\ \phi(0,.) = f \,,\ \partial_t \phi(0,.) = g \,. \tag{4.6}$$

• $\mathbf{n = 1.}$ The first well-known formula giving solutions to the wave equation is due to d'Alembert; it provides the general solution of (4.5) for $n = 1$. It was given by d'Alembert in 1747 [9] in the following form:

$$\phi(t,x) = F(x+t) + G(x-t) \,. \tag{4.7}$$

The proof is a simple exercise using a change of variables. Let us put $u = t - x$ and $v = t + x$, then

$$\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} = \frac{\partial^2}{\partial u \partial v} \,,$$

whence, if $\phi \in \mathcal{C}^2(\mathbb{R}^2)$, we have

$$\begin{aligned} \left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} \right) \phi = 0 \quad & \Leftrightarrow \quad \frac{\partial^2 \phi}{\partial u \partial v} = 0 \,, \\ & \Leftrightarrow \quad \frac{\partial \phi}{\partial v} = \psi_1(v) \,, \\ & \Leftrightarrow \quad \phi = \psi_2(v) + \psi_3(u) \,, \end{aligned}$$

where ψ_1 is a $\mathcal{C}^1$ function on $\mathbb{R}$, ψ_2 is a primitive of ψ_1 and ψ_3 is a $\mathcal{C}^2$ function on $\mathbb{R}$. This gives d'Alembert's formula (4.7) where F and G are both $\mathcal{C}^2$ on $\mathbb{R}$.

Remark 4.5. *It is indeed an integral formula, it can be written as follows*

$$\phi(t,x) = \int_{\{-1,1\}} \Phi(u, x+ut) \mathrm{d}(\delta_{-1} + \delta_1)(u) = \int_{S^0} \Phi(u, x+ut) \mathrm{d}\sigma(u) \,,$$

where $\Phi(-1,.) = G$ *and* $\Phi(1,.) = F$.

There is an apparent limitation attached to this formula. The functions F and G need to be $\mathcal{C}^2$, or at least twice differentiable, for equation (4.5) to be satisfied in the strong sense. However, the formula can be reformulated as :

$$\phi = \tau_{-t} F + \tau_t G \,, \tag{4.8}$$

where τ_t is the translation with respect to t defined for continuous functions by

$$(\tau_t f)(x) = f(x-t)$$

The family of translations $\{\tau_t\}_{t \in \mathbb{R}}$ is a smooth one parameter group of linear isometries on all $\mathcal{C}^k$ spaces on $\mathbb{R}$. In particular, for any $\phi \in \mathcal{D}(\mathbb{R})$, if we define $\psi(t,x) := (\tau_t \phi)(x)$, we have $\psi \in \mathcal{C}^\infty(\mathbb{R}_t \,;\ \mathcal{D}(\mathbb{R}))$. The translation operator can be extended to distributions by duality as follows. For $T \in \mathcal{D}'(\mathbb{R})$, the distribution $\tau_t T$ is defined by

$$\forall \phi \in \mathcal{D}(\mathbb{R}) \,,\ \langle \tau_t T, \phi \rangle_{\mathcal{D}', \mathcal{D}} = \langle T, \tau_{-t} \phi \rangle_{\mathcal{D}', \mathcal{C}_0^\infty} \,,$$

which generalises

$$\int_{\mathbb{R}} f(x-t)g(x)\mathrm{d}x = \int_{\mathbb{R}} f(x)g(x+t)\mathrm{d}x \text{ for } f,g \in L^2(\mathbb{R})\,.$$

This construction ensures that for any $T \in \mathcal{D}(\mathbb{R})$, if we consider the distribution $S \in \mathcal{D}'(\mathbb{R}_t \times \mathbb{R}_x)$ defined by

$$\forall \psi \in \mathcal{D}(\mathbb{R}_t \times \mathbb{R}_x)\,,\ \langle S\,,\ \psi\rangle_{\mathcal{D}'(\mathbb{R}^2),\mathcal{D}(\mathbb{R}^2)} := \int_{\mathbb{R}} \langle \tau_t T\,,\ \psi(t,.)\rangle_{\mathcal{D}'(\mathbb{R}),\mathcal{D}(\mathbb{R})}\mathrm{d}t\,,$$

then S is well-defined and $S \in \mathcal{C}^\infty(\mathbb{R}_t\,;\ \mathcal{D}'(\mathbb{R}))$. The properties

$$\frac{\partial}{\partial x}(\tau_t f) = (\tau_t f')\,,\ \frac{\partial}{\partial t}(\tau_t f) = -(\tau_t f')\,,$$

that are valid for differentiable functions f on $\mathbb{R}$, extend by density to distributions on $\mathbb{R}$. Hence d'Alembert's formula (4.7) provides solutions to (4.5) for $n = 1$ in $\mathcal{C}^\infty(\mathbb{R}_t\,,\ \mathcal{D}'(\mathbb{R}_x))$. For $n = 1$, t and x can be interchanged without changing the equation, therefore we also have that d'Alembert's formula provides solutions in $\mathcal{C}^\infty(\mathbb{R}_x\,,\ \mathcal{D}'(\mathbb{R}_t))$. Moreover, the proof of d'Alembert's formula is in fact valid for distributions since any distribution on $\mathbb{R}$ admit primitives, two of which differ by a constant. Therefore we have the following theorem.

Theorem 4.3. *All solutions of (4.5) for $n = 1$ in $\mathcal{D}'(\mathbb{R}^2)$ are given by d'Alembert's formula, written using an abusive notation for clarity,*

$$\begin{aligned} T(t,x) &= (\tau_t S_1)(x) + (\tau_{-t} S_2)(x)\,, S_1\,,\ S_2 \in \mathcal{D}'(\mathbb{R})\,,\\ &= (\tau_x \check{S}_1)(t) + (\tau_{-x} S_4)(t)\,, \end{aligned}$$

where " ˇ " denotes symmetry on the real axis. They all belong to $\mathcal{C}^\infty(\mathbb{R}_t\,,\ \mathcal{D}'(\mathbb{R}_x))$ as well as $\mathcal{C}^\infty(\mathbb{R}_x\,,\ \mathcal{D}'(\mathbb{R}_t))$.

When one is interested in solving the Cauchy problem (4.6) for $n = 1$, the following alternative expression of d'Alembert's formula is more adapted :

$$\phi(t,x) = \frac{1}{2}(f(x-t) + f(x+t)) + \frac{1}{2}\int_{x-t}^{x+t} g(s)\mathrm{d}s\,. \tag{4.9}$$

This too makes sense for distributions, provided we express it using translations and a primitive of g.

It is easy to go from (4.7) to (4.9). This simply requires to express the initial data for the Cauchy problem in terms of the functions F and G

$$f(x) := \phi(0,x) = F(x) + G(x)\,,$$
$$g(x) := \frac{\partial \phi}{\partial t}(0,x) = F'(x) - G'(x)\,.$$

Then,

$$\begin{aligned} F(x+t)+G(x-t) &= \frac{1}{2}[(F(x+t)+G(x+t))+(F(x+t)-G(x+t))] \\ &\quad +\frac{1}{2}[(F(x-t)+G(x-t))-(F(x-t)-G(x-t))] \\ &= \frac{1}{2}(f(x-t)+f(x+t))+\frac{1}{2}\int_{x-t}^{x+t} g(s)\mathrm{d}s\,. \end{aligned}$$

Note also that (4.9) gives us the expression of F and G in terms of the Cauchy data :

$$F(x)=\frac{1}{2}\left(f(x)+\int_0^x g(s)\mathrm{d}s\right)\,,\ G(x)=\frac{1}{2}\left(f(x)-\int_0^x g(s)\mathrm{d}s\right)\,.$$

Of course we do not have uniqueness as a constant can be added to F and subtracted to G without changing the solution.

D'Alembert's formula shows that spacelike regularity is preserved by the evolution for solutions of the wave equation in one space dimension. In terms of the Cauchy problem, we have the following results.

Proposition 4.1. *If the initial data satisfy for $k\in\mathbb{N}$, $k\geq 2$,*

$$f\in\mathcal{C}^k(\mathbb{R})\,,\ g\in\mathcal{C}^{k-1}(\mathbb{R})\,,$$

then the solution satisfies

$$\phi\in\mathcal{C}^k(\mathbb{R}^2)\,.$$

Also, if the initial data satisfy for $s\in\mathbb{R}$,

$$f\in H^s(\mathbb{R})\,,\ g\in H^{s-1}(\mathbb{R})\,,$$

then the solution satisfies for all $k\in\mathbb{N}$

$$\phi\in\mathcal{C}^k(\mathbb{R}\,;\ H^{s-k}(\mathbb{R}))\,.$$

• **n = 1 with symmetry.**

Proposition 4.2. *The form (4.9) of 's formula shows that some symmetries of the data are transferred to the solution :*

1. *if f and g are even, then the solution (4.9) will be even in x for all t ;*

2. *if f and g are odd, then the solution (4.9) will be odd in x for all t.*

Proof. The proof simply uses the fact that if g is odd (resp. even), then its primitive which vanishes at the origin is even (resp. odd). The rest is an obvious direct calculation. □

• **Kirchhoff's formula (n = 3).** In 3 space dimensions, the classic formula is the Kirchhoff formula (or D'Adhémar-Fresnel-Kirchhoff formula), dating from the late XIXth century. It is in the spirit of the second form of d'Alembert's formula and gives the solution (provided it is $\mathscr{C}^2$) to the Cauchy problem (4.6) for $n = 3$ in terms of integrals of the initial data. For each $x \in \mathbb{R}^3$ and $r > 0$, we denote by $S(x,r)$ the sphere in $\mathbb{R}^3$ of center x and radius r. For $t > 0$ and $x \in \mathbb{R}^3$, the solution ϕ is given in terms of the initial data $f = \phi|_{t=0}$ and $g = \partial_t \phi|_{t=0}$ by

$$\begin{aligned} \phi(t,x) &= \frac{1}{4\pi t^2}\int_{S(x,t)} \left(f(y) + (y-x).\nabla f(y) + tg(y)\right) \mathrm{d}\sigma(y) \\ &= \frac{1}{4\pi}\int_{S^2} \left(f(x+t\omega) + t\omega.(\nabla f)(x+t\omega) + tg(x+t\omega)\right) \mathrm{d}\omega. \end{aligned} \tag{4.10}$$

Remark 4.6. *We observe from (4.10) that if the initial data satisfy $f \in \mathscr{C}^{k+1}(\mathbb{R}^3)$ and $g \in \mathscr{C}^k(\mathbb{R}^3)$, then the corresponding solution can only be expected to be $\mathscr{C}^k$ on $\mathbb{R} \times \mathbb{R}^3$. However, as stated in Theorem 4.1 and as we have shall see later on by explicit methods, Sobolev regularity is preserved by the evolution.*

Proof of the formula. It is done by the method of spherical means. We write the proof for $t > 0$, it can be extended to $t < 0$ simply by a time reflexion (meaning changing t into $-t$ and g into $-g$). Let $\phi \in \mathcal{C}^2(\mathbb{R}_t \times \mathbb{R}^3_x)$ a solution of (4.5) for $n = 3$. We define the average $U(x,t,r)$ of $\phi(t,y)$ on the sphere $S^2(x,r)$ in $\mathbb{R}^3$:

$$U(x,t,r) = \frac{1}{4\pi r^2}\int_{S^2(x,r)} \phi(t,y)\mathrm{d}\sigma(y) = \frac{1}{4\pi}\int_{|\xi|=1} \phi(t, x+r\xi)\mathrm{d}\sigma(\xi). \tag{4.11}$$

The fact that ϕ satisfies the wave equation implies that U satisfies a partial differential equation purely in the variables t and r. Let us start by evaluating the derivative of U with respect to r:

$$\begin{aligned} \frac{\partial U}{\partial r}(x,t,r) &= \frac{1}{4\pi}\int_{|\xi|=1} \xi^i \frac{\partial \phi}{\partial x^i}(t,x+r\xi)\mathrm{d}\sigma(\xi) \text{ (outgoing flux of the gradient)} \\ &= \frac{1}{4\pi}\int_{|\xi|<1} \mathrm{div}_\xi(\nabla\phi(t,x+r\xi))\mathrm{d}\xi \\ &= \frac{r}{4\pi}\int_{|\xi|<1} \Delta\phi(t,x+r\xi)\mathrm{d}\xi \\ &= \frac{r}{4\pi}\int_{|\xi|<1} \frac{\partial^2\phi}{\partial t^2}(t,x+r\xi)\mathrm{d}\xi \\ &= \frac{r}{4\pi}\frac{\partial^2}{\partial t^2}\int_{|\xi|<1} \phi(t,x+r\xi)\mathrm{d}\xi \\ &= \frac{1}{4\pi r^2}\frac{\partial^2}{\partial t^2}\int_{|y-x|<r} \phi(t,y)\mathrm{d}y. \end{aligned} \tag{4.12}$$

Now we express the volume integral in terms of the spherical average function U by a simple application of Fubini's Theorem:

$$\frac{1}{4\pi}\int_{|y-x|<r} \phi(t,y)\mathrm{d}y = \int_{]0,r[} \rho^2 U(x,t,\rho)\mathrm{d}\rho.$$

It follows that

$$r^2\frac{\partial U}{\partial r}(x,t,r)=\frac{\partial^2}{\partial t^2}\int_{]0,r[}\rho^2 U(x,t,\rho)\mathrm{d}\rho=\int_{]0,r[}\rho^2\frac{\partial^2}{\partial t^2}U(x,t,\rho)\mathrm{d}\rho\,.$$

Taking the derivative with respect to r and dividing by r^2, we get

$$\frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial U}{\partial r}\right)=\frac{\partial^2 U}{\partial r^2}+\frac{2}{r}\frac{\partial U}{\partial r}=\frac{\partial^2 U}{\partial t^2}\,.$$

Hence $\psi(x,t,r)=rU(x,t,r)$ satisfies

$$\frac{\partial^2\psi}{\partial t^2}-\frac{\partial^2\psi}{\partial r^2}=0$$

which is the wave equation in one space dimension. The function ψ, for a given x, is a priori only defined on $\mathbb{R}_t\times\mathbb{R}_x^+$ but the second expression of U in (4.11) is defined for any $r\in\mathbb{R}$, is even in r and is $\mathcal{C}^2$ in (x,t,r) for $r>0$. Moreover, (4.12) shows that

$$\frac{\partial U}{\partial r}\to 0 \text{ as } r\to 0\,.$$

So ψ extends naturally for each x as a function that is $\mathcal{C}^2$ on $\mathbb{R}_t\times\mathbb{R}_r$ (in fact in $\mathcal{C}^2(\mathbb{R}_x\times\mathbb{R}_t\times\mathbb{R}_r)$) and odd in r, which consequently satisfies the wave equation on the whole (t,r)-plane for each x. We denote by $\tilde\psi$ the extension of ψ to $\mathbb{R}_x\times\mathbb{R}_t\times\mathbb{R}_r$ that is odd in r. Hence, putting

$$\tilde f(x,r)=\tilde\psi(x,0,r)\,,\ \tilde g(x,r)=\frac{\partial\tilde\psi}{\partial t}(x,0,r)\,,$$

using 's formula we get

$$\tilde\psi(x,t,r)=\frac12(\tilde f(x,r+t)+\tilde f(x,r-t))+\frac12\int_{r-t}^{r+t}\tilde g(x,s)\mathrm{d}s\,.$$

Using the fact that $\tilde f$ and $\tilde g$ are odd in r, we deduce the following expressions for $\psi(x,t,r)$ in the domains $r\geq t\geq 0$ and $t\geq r\geq 0$ respectively :

$$\begin{aligned}\psi(x,t,r)&=\frac12(\tilde f(x,r+t)+\tilde f(x,r-t))+\frac12\int_{r-t}^{r+t}\tilde g(x,s)\mathrm{d}s\quad\text{for } r\geq t>0\,,\\ \psi(x,t,r)&=\frac12(\tilde f(x,r+t)-\tilde f(x,t-r))+\frac12\int_{t-r}^{t+r}\tilde g(x,s)\mathrm{d}s\quad\text{for } t>r\geq 0\end{aligned}$$

since

$$\int_{r-t}^{t-r}\tilde g(x,s)\mathrm{d}s=0\quad\text{for } t>r\geq 0\,.$$

Now we divide by r and take the limit as $r \to 0$. So only the form for $t > r > 0$ is useful and we obtain

$$\begin{aligned}\lim_{r\to 0}\left(\frac{\tilde f(x,r+t)-\tilde f(x,t-r)}{2r}+\frac{1}{2r}\int_{t-r}^{t+r}\tilde g(x,s)\mathrm{d}s\right)\\ =\frac{\partial \tilde f}{\partial r}(x,t)+\tilde g(x,t)\\ =\frac{\partial\tilde\psi}{\partial r}(x,0,r=t)+\frac{\partial\tilde\psi}{\partial t}(x,0,r=t)\\ =t\frac{\partial U}{\partial r}(x,0,r=t)+U(x,0,r=t)+t\frac{\partial U}{\partial t}(x,0,r=t)\end{aligned}$$

and using the second expression of U in (4.11) we obtain the formula. □

This formula has a remarkable consequence which is often referred to as the Huygens principle (or strong Huygens principle in contrast with some weak versions we shall encounter later on) and stating that for the wave equation in 3 space dimensions, the information travels exactly at speed 1 :

Theorem 4.4 (Huygens principle)**.** *For $n = 3$, if the data for the wave equation (4.5) at $t = 0$ are supported in the ball $B(0, R)$, then the associated solution ϕ satisfies*

$$\phi(t,x)=0 \text{ for } |x|\le |t|-R \text{ and for } |x|\ge |t|+R\,.$$

For space dimensions $n = 2, 4, 5, \mathrm{etc}...$ we can deduce integral formulae from the $n = 3$ case. In even space dimensions, the propagation exhibits a very different behaviour from the odd dimensional cases, namely, the Huygens principle is valid only for space dimensions $n \ge 3$ and odd. We will only be interested in the $n = 3$ case.

• **Whittaker's formula (n = 3).** In 1903, Whittaker [61] obtained another integral formula which is comparable to the first form of d'Alembert's formula. It provides "general" solutions of the wave equation on $\mathbb{R}_t \times \mathbb{R}^3$ with no obvious relation to the Cauchy problem :

$$\phi(t,x)=\int_{S^2}\Phi(x.\omega-t,\omega)\mathrm{d}\omega\,. \tag{4.13}$$

It is quite different from the Kirchhoff formula, the most striking aspect being the presence of only one arbitrary function on $\mathbb{R}\times S^2$ instead of two functions on $\mathbb{R}^3$. We shall see that its natural interpretation is in terms of scattering theory. Whittaker's proof was a direct calculation, similar to the one he used to obtain an integral formula characterizing harmonic functions on $\mathbb{R}^3$. We will not give his proof here. We shall see later a different and much less direct proof of this formula in the Lax-Phillips version of scattering theory. It will however have the advantage of giving a clear-cut geometrical interpretation of (4.13).

4.1.2 The Cauchy problem on flat space-time

The Cauchy problem (4.6) can be made sense of and solved in very general function spaces, provided we keep some sort of time regularity to allow for a meaningful initial data condition.

Several methods can be used to solve (4.6). Some will be adapted to these very general function spaces, others will bring their own sets of function spaces, which, although less general, will provide ideal frameworks for developing scattering theories.

Spectral approach

In this section we work with $n \geq 3$ (see footnote 1 below). We write the wave equation in its Hamiltonian form, i.e. as a Schrödinger equation:

$$\partial_t U = iAU\,,\ U := \begin{pmatrix} \phi \\ \partial_t \phi \end{pmatrix}\,,\ A = -i \begin{pmatrix} 0 & 1 \\ \Delta & 0 \end{pmatrix}\,. \tag{4.14}$$

Theorem 4.5. *The operator A is self-adjoint on $\mathscr{H} = \dot{H}^1(\mathbb{R}^n) \times L^2(\mathbb{R}^n)$, defined as the completion of $\mathscr{C}_0^\infty(\mathbb{R}^n) \times \mathscr{C}_0^\infty(\mathbb{R}^n)$ in the norm*

$$\|U\|^2_{\mathscr{H}} := \int_{\mathbb{R}^n} (|\nabla \phi_1|^2 + |\phi_2|^2) \mathrm{d}x\,.$$

Proof. First for $U \in \mathscr{C}_0^\infty(\mathbb{R}^n) \times \mathscr{C}_0^\infty(\mathbb{R}^n)$, we have $AU \in \mathscr{C}_0^\infty(\mathbb{R}^n) \times \mathscr{C}_0^\infty(\mathbb{R}^n) \subset \mathscr{H}$, so the domain of A contains $\mathscr{C}_0^\infty(\mathbb{R}^n) \times \mathscr{C}_0^\infty(\mathbb{R}^n)$ and is therefore dense in $\mathscr{H}$. Let us prove that A is symmetric. Let

$$U = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}\,,\ V = \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \in \mathscr{C}_0^\infty(\mathbb{R}^n) \times \mathscr{C}_0^\infty(\mathbb{R}^n)\,,$$

$$\begin{aligned}
\langle AU, V \rangle_{\mathscr{H}} &= -i \langle \begin{pmatrix} \phi_2 \\ \Delta\phi_1 \end{pmatrix}, \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \rangle_{\mathscr{H}} \\
&= -i \int_{\mathbb{R}^n} \left(\nabla\phi_2 . \nabla\overline{\psi_1} + \Delta\phi_1 \overline{\psi_2} \right) \mathrm{d}x \\
&= i \int_{\mathbb{R}^n} \left(\phi_2 . \Delta\overline{\psi_1} + \nabla\phi_1 . \nabla\overline{\psi_2} \right) \mathrm{d}x \\
&= \langle \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, -i \begin{pmatrix} \psi_2 \\ \Delta\psi_1 \end{pmatrix} \rangle_{\mathscr{H}} = \langle U, AV \rangle_{\mathscr{H}}\,.
\end{aligned}$$

The symmetry on $D(A)$ follows by density. It remains to show that $D(A^*) \subset D(A)$. For $n \geq 3$, $\dot{H}^1(\mathbb{R}^n)$ is a space of distributions[1] so it is easy to understand A^* as a differential operator and to determine its domain. Let

$$U = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix} \in \mathscr{H}\,,$$

then $U \in D(A^*)$ if and only if the map

$$V \in D(A) \mapsto \langle AV, U \rangle_{\mathscr{H}}$$

[1]This is not the case for $n = 1$ and $n = 2$, see Soga 1983 [53], p. 732.

extends as a linear continuous map on $\mathscr{H}$. This map restricted to $\mathscr{C}_0^\infty(\mathbb{R}^n) \times \mathscr{C}_0^\infty(\mathbb{R}^n)$ is a distribution which we can evaluate in terms of ϕ_1 and ϕ_2 in the usual manner. Consider

$$V = \left(\begin{array}{c} \psi_1 \\ \psi_2 \end{array} \right) \in \mathscr{C}_0^\infty(\mathbb{R}^n) \times \mathscr{C}_0^\infty(\mathbb{R}^n) \,,$$

then

$$\begin{aligned} \langle AV, U \rangle_{\mathscr{H}} &= \langle -i \left(\begin{array}{c} \psi_2 \\ \Delta\psi_1 \end{array} \right), \left(\begin{array}{c} \phi_1 \\ \phi_2 \end{array} \right) \rangle_{\mathscr{H}} \\ &= -i \int_{\mathbb{R}^n} \left(\overline{\phi_2}.\Delta\psi_1 + \nabla\overline{\phi_1}.\nabla\psi_2 \right) \mathrm{d}x \\ &= -i \langle \overline{\phi_2}, \Delta\psi_1 \rangle_{\mathscr{D}',\mathscr{D}} - i \langle \nabla\overline{\phi_1}, \nabla\psi_2 \rangle_{\mathscr{D}',\mathscr{D}} \\ &= i \langle \nabla\overline{\phi_2}, \nabla\psi_1 \rangle_{\mathscr{D}',\mathscr{D}} + i \langle \Delta\overline{\phi_1}, \psi_2 \rangle_{\mathscr{D}',\mathscr{D}} \,. \end{aligned}$$

This extends as a continuous linear map on $\mathscr{H}$ if and only if

$$\nabla\overline{\phi_2} \text{ and } \Delta\overline{\phi_1} \text{ are in } L^2(\mathbb{R}^n) \,,$$

which is equivalent to $AU \in \mathscr{H}$, i.e. to $U \in D(A)$. Therefore $D(A^*) = D(A)$ and the proof is complete. $\square$

Consequently, by Theorem A.12 and Remark A.7, the Cauchy problem (4.6) is well-posed in $\mathscr{H}$ and all the successive domains of A in $\mathscr{H}$. Let us express and prove the theorem explicitly.

Theorem 4.6. *Let $U_0 \in \mathscr{H}$, there exists a unique solution $U \in \mathscr{C}(\mathbb{R}_t; \mathscr{H})$ of (4.14), in the sense of distributions, such that $U(0) = U_0$. It is given by*

$$U(t) = e^{itA} U_0$$

and satisfies

$$\| U(t) \|_{\mathscr{H}} = \| U_0 \|_{\mathscr{H}} \ \forall t \in \mathbb{R}.$$

Moreover, if $U_0 \in D(A^k)$ for $k \geq 1$, then

$$U \in \bigcap_{p=0}^{k} \mathscr{C}^p(\mathbb{R}_t; D(A^{k-p})).$$

Proof. Theorem A.12 gives us well-posedness in $D(A)$ as well as the additional regularity condition. We simply need to prove the minimum regularity well-posedness. Let $U_0 \in \mathscr{H}$, we consider a sequence $(U_0^n)_n$ in $D(A)$ that converges towards U_0 in $\mathscr{H}$ and, for each n, the corresponding solution $U^n \in \mathscr{C}(\mathbb{R}_t; D(A))$ of (4.14) such that $U^n(0) = U_0^n$. The properties of the one-parameter group e^{itA} imply that $U^n \to U$ in $\mathscr{C}(\mathbb{R}_t; \mathscr{H})$ and since this is a distribution space and A is a differential operator, if follows that $\partial_t U^n - iAU^n$ converges towards $\partial_t U - iAU$ in $\mathscr{D}'(\mathbb{R} \times \mathbb{R}^n)$. Therefore, U is a solution of (4.14) in the sense of distributions and satisfies the initial data condition.

As for uniqueness in $\mathscr{C}(\mathbb{R}_t; \mathscr{H})$, thanks to the linearity of the equation, it amounts to showing that for zero initial data, zero is the only solution to the Cauchy problem in $\mathscr{C}(\mathbb{R}_t; \mathscr{H})$. We consider $U \in \mathscr{C}(\mathbb{R}_t;\mathscr{H})$ a solution of (4.14) such that $U(0)=0$. Let $t_0 \in \mathbb{R}$, $V_0 \in \mathscr{D}(\mathbb{R}^n)\times\mathscr{D}(\mathbb{R}^n)$ and $V \in \mathscr{C}(\mathbb{R}_t;D(A))\cap\mathscr{C}^1(\mathbb{R}_t,,\mathscr{H})$ solution of (4.14) such that $V(t_0)=V_0$. One can see from the Kirchhoff formula, or from Theorem 4.8 below, that $V\in\mathscr{C}^\infty(\mathbb{R}_t;\mathscr{D}(\mathbb{R}^n)\times\mathscr{D}(\mathbb{R}^n))$. Then, considering the following inner products as hermitian duality brackets between $\mathscr{D}'(\mathbb{R}^n)\times\mathscr{D}'(\mathbb{R}^n)$ and $\mathscr{D}(\mathbb{R}^n)\times\mathscr{D}(\mathbb{R}^n)$, we have

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\langle U,V\rangle_{\mathscr{H}} &= \langle \partial_t U, V\rangle + \langle U,\partial_t V\rangle\\ &= \langle iAU,V\rangle + \langle U, iAV\rangle = 0\end{aligned}$$

since iA is skew for this duality bracket. Il follows that $\langle U(t),V(t)\rangle$ is constant in time and therefore vanishes for all times. In particular we have $\langle U(t_0),V_0\rangle = 0$. This is true for any $V_0\in\mathscr{D}(\mathbb{R}^n)\times\mathscr{D}(\mathbb{R}^n)$ and $t_0\in\mathbb{R}$. It follows that $U(t)=0$ as a distribution for all t. □

Remark 4.7. *One might feel that the purely spectral approach for the wave equation is somewhat unsatisfactory, the space $\mathscr{H}$ being a little awkward because of the homogeneous Sobolev space $\dot H^1(\mathbb{R}^n)$. This would be particularly true for $n=1$ and $n=2$ since $\dot H^1(\mathbb{R})$ is not even a distribution space (see footnote 1 above). But even for $n\geq 3$, the domains of A are not easy to understand because of the lack of L^2 control on the first component of U. However, such function spaces are natural for constructing a scattering theory for the wave equation on asymptotically flat backgrounds, whether one uses a spectral approach or a conformal approach (see the example of the Schwarzschild metric with the spectral approach [12, 13, 14] and the conformal approach [40]).*

Fourier transform

Assuming $f,g\in\mathscr{S}'(\mathbb{R}^n)$ (resp. $\mathscr{S}(\mathbb{R}^n)$), we look for a solution $\phi\in\mathscr{C}^1(\mathbb{R}_t\,;\,\mathscr{S}'(\mathbb{R}^n))$ (resp. $\mathscr{C}^1(\mathbb{R}_t\,;\,\mathscr{S}(\mathbb{R}^n))$) of (4.6). If ϕ is in such a distribution space, then ϕ satisfies (4.6) if and only if its Fourier transform in space $\hat\phi(t,\xi)$ satisfies

$$\partial_t^2\hat\phi + |\xi|^2\hat\phi = 0 \text{ on } \mathbb{R}_t\times\mathbb{R}^n_x\,,\ \hat\phi(0,.)=\hat f\,,\ \partial_t\hat\phi(0,.)=\hat g\,. \tag{4.15}$$

The problem (4.15) has a unique solution in $\mathscr{C}^1(\mathbb{R}_t\,;\,\mathscr{S}'(\mathbb{R}^n))$ (resp. $\mathscr{C}^1(\mathbb{R}_t\,;\,\mathscr{S}(\mathbb{R}^n))$) given by

$$\hat\phi(t,\xi) = \hat f(\xi)\cos(t|\xi|) + \hat g(\xi)\frac{\sin(t|\xi|)}{|\xi|}. \tag{4.16}$$

This is due to the fact that the functions $\cos(t|\xi|)$ and $\frac{\sin(t|\xi|)}{|\xi|}$ are analytic in t with values in analytic functions in ξ with moderate growth at infinity, and are therefore continuous multipliers of both spaces $\mathscr{C}^1(\mathbb{R}_t\,;\,\mathscr{S}'(\mathbb{R}^n))$ and $\mathscr{C}^1(\mathbb{R}_t\,;\,\mathscr{S}(\mathbb{R}^n))$. This proves the following theorem.

Theorem 4.7. *Given $f,g\in\mathscr{S}'(\mathbb{R}^n)$ (resp. $\mathscr{S}(\mathbb{R}^n)$), the Cauchy problem (4.6) for the wave equation has a unique solution in $\mathscr{C}^1(\mathbb{R}_t\,;\,\mathscr{S}'(\mathbb{R}^n))$ (resp. $\mathscr{C}^1(\mathbb{R}_t\,;\,\mathscr{S}(\mathbb{R}^n))$) given by*

$$\phi(t,.) = \mathcal{F}_\xi^{-1}\left(\hat f(\xi)\cos(t|\xi|) + \hat g(\xi)\frac{\sin(t|\xi|)}{|\xi|}\right).$$

This solution is in fact in $\mathscr{C}^\infty(\mathbb{R}_t\,;\,\mathscr{S}'(\mathbb{R}^n))$ (resp. $\mathscr{C}^\infty(\mathbb{R}_t\,;\,\mathscr{S}(\mathbb{R}^n))$).

Moreover it is immediate to check explicitly, without resorting to Theorem 4.1 but using the characterization of Sobolev spaces via the Fourier transform, that the Cauchy problem is also well-posed in any Sobolev space:

Corollary 4.1. *Given $s \in \mathbb{R}$, $f \in H^s(\mathbb{R}^n)$, $g \in H^{s-1}(\mathbb{R}^n)$, the associated solution ϕ of the Cauchy problem (4.6) satisfies:*

$$\phi \in \mathscr{C}(\mathbb{R}_t\,;\; H^s(\mathbb{R}^n)) \cap \mathscr{C}^1(\mathbb{R}_t\,;\; H^{s-1}(\mathbb{R}^n))\,.$$

Fundamental solutions

G the backward fundamental solution ($G(x,y) = \delta((x'-y')^2)\theta(-x_0+y_0)$ essentially). Then write

$$\phi(x) = \int_{\mathbb{R}^4} (\phi(y)\Box G(x,y) - G(x,y)\Box\phi(y))\,\mathrm{d}y = \int_{\mathbb{R}^4} \nabla.\,(\phi(y)\nabla G(x,y) - G(x,y)\nabla\phi(y))\,\mathrm{d}y\,.$$

Integrate by parts on the future of some initial data hypersurface (null or spacelike) and since the fundamental solution has support limited in the future, we get a boundary term that is purely on the data hypersurface and formally reads

$$\phi(x) = \int_{\Sigma} (\phi(y)\nabla_n G(x,y) - G(x,y)\nabla_n\phi(y))\,\mathrm{d}^3y\,.$$

4.2 Energy estimates

In this section, we present the method of energy estimates for the wave equation on flat space-time.

The Fourier transform approach provides the existence of solutions in spaces of very weak regularity (tempered distributions) and very strong regularity (the Schwartz class of smooth rapidly decreasing functions). The well-posedness in the Schwartz class together with energy estimates allow us to recover, for $s \in \mathbb{N}^*$, the results of Corollary 4.1. A general method to obtain similar results on general curved space-times is to combine energy estimates with the weak version of Theorem 4.1 for smooth compactly supported data. This will be described in the next section.

There are two essentially equivalent ways of understanding the principle of energy estimates. The first, which is familiar to most PDE analysts, is to multiply equation (4.5) by a well-chosen directional derivative of the solution ϕ and to integrate the result by parts on a domain of $\mathbb{R}^{n+1}$ with piecewise $\mathscr{C}^1$ boundary. The second is much more geometrical and comes from the physical/geometrical invariance properties of the equation.

4.2.1 Analytical approach: finite propagation speed

We explain this description of energy estimate on a particular example and use it to establish the finite propagation speed of the solution. Consider $\phi \in \mathscr{C}^\infty(\mathbb{R}\times\mathbb{R}^n)$ solution of (4.5): we have

$$0 = \partial_t\phi\left(\partial_t^2\phi - \Delta_x\phi\right)\,.$$

We integrate this by parts on the domain

$$\Omega_{R,T} = \{(t,x) \in \mathbb{R} \times \mathbb{R}^n\,;\; 0 \leq t \leq T\,,\; |x| \leq R+T-t\} \tag{4.17}$$

and denote the three pieces of the boundary of $\Omega_{R,T}$ by

$$\begin{aligned} \Sigma_T &= \{(t,x) \in \mathbb{R} \times \mathbb{R}^n\,;\; t=T\,,\; |x| \leq R\}\,, & (4.18)\\ \Sigma_0 &= \{(t,x) \in \mathbb{R} \times \mathbb{R}^n\,;\; t=0\,,\; |x| \leq R+T\}\,, & (4.19)\\ S &= \{(t,x) \in \mathbb{R} \times \mathbb{R}^n\,;\; 0 \leq t \leq T\,,\; |x| = R+T-t\}\,. & (4.20) \end{aligned}$$

We obtain

$$\begin{aligned} 0 &= \frac{1}{2}\int_{\Sigma_T} \left(|\partial_t\phi|^2 + |\nabla\phi|^2\right) \mathrm{d}^3x \\ &\quad -\frac{1}{2}\int_{\Sigma_0} \left(|\partial_t\phi|^2 + |\nabla\phi|^2\right) \mathrm{d}^3x \\ &\quad +\frac{1}{2\sqrt{2}}\int_S \left(|\partial_t\phi|^2 + |\nabla\phi|^2 + 2\partial_t\phi\frac{x}{|x|}.\nabla\phi\right) \mathrm{d}\sigma\,. \end{aligned}$$

The last integral being non negative, we see that

$$\frac{1}{2}\int_{\Sigma_T} \left(|\partial_t\phi|^2 + |\nabla\phi|^2\right) \mathrm{d}^3x \leq \frac{1}{2}\int_{\Sigma_0} \left(|\partial_t\phi|^2 + |\nabla\phi|^2\right) \mathrm{d}^3x\,. \tag{4.21}$$

In particular, if the solution is zero for $|x| \leq R+T$ at $t=0$, this implies that it must also be zero at $t=T$ for $|x| \leq R$, i.e. the information propagates at most at speed 1.

Remark 4.8. *Note that this result is weaker than the Huygens principle which gives an exact propagation speed. But it is also more general: unlike the Huygens principle, this property will be valid for perturbations of the wave equation by first or zero order terms and it can also be extended, using the same method, to similar equations on curved backgrounds.*

Finite propagation speed and Theorem 4.7 entail the following result

Theorem 4.8. *Given $f, g \in \mathscr{C}_0^\infty(\mathbb{R}^n)$, the Cauchy problem (4.6) for the wave equation has a unique solution in $\mathscr{C}^1(\mathbb{R}_t\,;\; \mathscr{C}_0^\infty(\mathbb{R}^n))$. This solution is in fact in $\mathscr{C}^\infty(\mathbb{R}_t\,;\; \mathscr{C}_0^\infty(\mathbb{R}^n))$.*

By duality, we deduce well-posedness with general distribution data:

Corollary 4.2. *Given $f, g \in \mathcal{D}'(\mathbb{R}^n)$, the Cauchy problem (4.6) for the wave equation has a unique solution in $\mathscr{C}^1(\mathbb{R}_t\,;\; \mathcal{D}'(\mathbb{R}^n))$. This solution is in fact in $\mathscr{C}^\infty(\mathbb{R}_t\,;\; \mathcal{D}'(\mathbb{R}^n))$.*

4.2.2 Geometrical approach

From now on, we consider only the case $n=3$ corresponding to the framework of special relativity. We denote by $(\mathbb{M}, \eta)$ the 4-dimensional Minkowski space-time.

The wave equation (4.5) has a conserved stress-energy tensor. It is the symmetric 2-tensor

$$T_{ab} = \partial_a\phi\partial_b\phi - \frac{1}{2}\langle\nabla\phi, \nabla\phi\rangle_\eta \eta_{ab}\,, \tag{4.22}$$

where η is the Minkowski metric

$$\eta = \mathrm{d}t^2 - e_{\mathbb{R}^3}\,,\ e_{\mathbb{R}^3} \text{ being the euclidean metric on } \mathbb{R}^3\,,$$

and $\langle .\,,.\rangle$ is the indefinite inner product induced by η, i.e.

$$\langle \nabla\phi\,,\nabla\phi\rangle_\eta = \eta^{ab}\nabla_a\phi\nabla_b\phi = |\partial_t\phi|^2 - |\nabla_x\phi|^2\,.$$

The stress-energy tensor (4.22) satisfies the following fundamental property.

Proposition 4.3.

$$\nabla^a T_{ab} = (\nabla_b\phi)\,\Box\phi \tag{4.23}$$

and therefore

$$\nabla^a T_{ab} = 0 \tag{4.24}$$

whenever ϕ satisfies the wave equation.

Proof. It is a direct calculation:

$$\begin{aligned}
\nabla^a T_{ab} &= \nabla^a\left(\nabla_a\phi\nabla_b\phi - \frac{1}{2}\nabla_c\phi\nabla^c\phi\eta_{ab}\right)\\
&= (\Box\phi)\nabla_b\phi + \nabla_a\phi\nabla^a\nabla_b\phi - (\nabla^a\nabla_c\phi)\,(\nabla^c\phi)\eta_{ab}\\
&= (\Box\phi)\nabla_b\phi + \nabla_c\phi\nabla_b\nabla^c\phi - (\nabla^a\nabla_c\phi)\,(\nabla^c\phi)\eta_{ab} \text{ no torsion}\\
&= (\Box\phi)\nabla_b\phi + (\nabla^c\phi)(\nabla_b\nabla_c\phi) - (\nabla_b\nabla_c\phi)\,(\nabla^c\phi)\\
&= (\Box\phi)\nabla_b\phi\,.\quad\Box
\end{aligned}$$

A "family of local observers" is described by a timelike vector field. From a stress-energy tensor, an **energy current** can be inferred by contracting it with a timelike vector field. Note that it is also sometimes interesting to consider energy currents associated with spacelike or null vector fields, they do not correspond to a physical measurement of energy current by a realistic observer but they can give useful information nonetheless, like local dispersion of energy for instance.

The conservation law (4.24) is not directly usable because it does not readily provide a conserved current (i.e. a divergence-free vector field). However, the symmetries of flat space-time will allow us to infer many conserved currents from T_{ab}.

Proposition 4.4. *Let K be a Killing vector field on $\mathbb{M}$, then the vector field*

$$J^a = K^b T_b^a$$

is divergence-free.

Proof. It is a direct consequence of the Killing equation and the symmetry of the stress-energy tensor:

$$\nabla^a J_a = K^b\nabla^a T_{ab} + T_{ab}\nabla^a K^b = K^b\nabla^a T_{ab} + T_{ab}\nabla^{(a}K^{b)}$$

and this is zero since K is Killing and by the conservation law (4.24). $\Box$

Recall that Minkowski space-time has a 10-dimensional group of isometries: the Poincaré group. Its associated Lie algebra is the 10-dimensional vector space of all Killing vector fields of $\mathbb{M}$, a basis of which is made of:

- ∂_t, ∂_{x^1}, ∂_{x^2}, ∂_{x^3}, generating translations;
- $x^1\partial_{x^2} - x^2\partial_{x^1}$, $x^2\partial_{x^3} - x^3\partial_{x^2}$, $x^3\partial_{x^1} - x^1\partial_{x^3}$, generating spatial rotations;
- $t\partial_{x^1} + x^1\partial_t$, $t\partial_{x^2} + x^2\partial_t$, $t\partial_{x^3} + x^3\partial_t$, generating boosts.

This gives us 10 independent conserved currents.

An important property of the stress-energy tensor for the wave equation is that when we contract it with a future-oriented timelike vector field and calculate its flux across a spacelike hypersurface with future-oriented normal, we obtain a positive energy. This is the so-called "dominant energy condition".

Proposition 4.5 (Dominant energy condition)**.** *The stress-energy tensor T_{ab} satisfies the dominant energy condition: for every future-oriented causal vector field V, the vector field $T^a_b V^b$ is itself causal and future-pointing. Another equivalent way of stating the dominant energy condition is the following: for all future-oriented causal vector fields U, V, we have $T_{ab}U^aV^b \geq 0$. In other words, for any future-oriented causal vector field V, the vector field $W^a = T^a_b V^b$ is causal and future-oriented.*

Proof. Let V be a future-oriented causal vector field, i.e.

$$V = V^0\partial_t + V', \ V^0 \geq |V'|,$$

put

$$W^a = T^a_b V^b = \nabla^a\phi\nabla_V\phi - \frac{1}{2}\langle\nabla\phi, \ \nabla\phi\rangle V^a.$$

We have

$$\begin{aligned}
W^0 &= g^{00}T_{0b}V^b = \partial_t\phi\nabla_V\phi - \frac{1}{2}((\partial_t\phi)^2 - |\nabla_x\phi|^2)V^0 \\
&= \frac{1}{2}V^0((\partial_t\phi)^2 + |\nabla_x\phi|^2) + \partial_t\phi\nabla_{V'}\phi \\
&\geq \frac{1}{2}V^0((\partial_t\phi)^2 + |\nabla_x\phi|^2) - |\partial_t\phi|\,|V'|\,|\nabla_x\phi| \\
&\geq \frac{1}{2}((\partial_t\phi)^2 + |\nabla_x\phi|^2)(V^0 - |V'|) \geq 0.
\end{aligned}$$

So if W^a is causal, it is future-oriented. Let us check the causality:

$$\begin{aligned}
g_{ab}W^aW^b &= g_{ab}(\nabla^a\phi\nabla_V\phi - \frac{1}{2}\langle\nabla\phi, \ \nabla\phi\rangle V^a)(\nabla^b\phi\nabla_V\phi - \frac{1}{2}\langle\nabla\phi, \ \nabla\phi\rangle V^b) \\
&= (\nabla_V\phi)^2\langle\nabla\phi, \ \nabla\phi\rangle - (\nabla_V\phi)^2\langle\nabla\phi, \ \nabla\phi\rangle + \frac{1}{4}\langle\nabla\phi, \ \nabla\phi\rangle^2\langle V, \ V\rangle \\
&= \frac{1}{4}\langle\nabla\phi, \ \nabla\phi\rangle^2\langle V, \ V\rangle \geq 0 \text{ since } V^a \text{ is causal.}
\end{aligned}$$

This proves the proposition. ∎

This property allows to establish estimate (4.21) by a different approach using more geometrical ingredients. We consider the energy current

$$J^a = T^a_b(\partial_t)^b = T^a_0\,,$$

corresponding to the perception of a static observer (whose velocity 4-vector is given by ∂_t); recall that the current is conserved because ∂_t is Killing. We integrate the divergence of J over the domain $\Omega_{R,T}$ with boundary made of Σ_T, Σ_0 and S as defined in (4.17), (4.18), (4.19) and (4.20). Denoting by E_{Σ_T} and E_{Σ_0} the energy fluxes across Σ_T and Σ_0 oriented by ∂_t and by E_S the outgoing energy fluxes across S, we get

$$E_{\Sigma_T} + E_S - E_{\Sigma_0} = 0\,,$$

all fluxes being calculated using the expression in the divergence Theorem (Theorem 2.5). For the first two fluxes, we take $l = n = \partial_t$:

$$E_{\Sigma_T} = \frac{1}{2}\int_{\Sigma_T}\left(|\partial_t\phi|^2 + |\nabla\phi|^2\right)\mathrm{d}^3x\,, \tag{4.25}$$

$$E_{\Sigma_0} = \frac{1}{2}\int_{\Sigma_0}\left(|\partial_t\phi|^2 + |\nabla\phi|^2\right)\mathrm{d}^3x\,. \tag{4.26}$$

As for E_S, taking

$$n = \frac{1}{\sqrt{2}}(\partial_t - \partial_r)\,,\ l = \frac{1}{\sqrt{2}}(\partial_t + \partial_r)\,,$$

we have

$$E_S = \int_S T_{ab}(\partial_t)^a n^b l \lrcorner \mathrm{dVol} \geq 0$$

by the dominant energy condition. This gives (4.21).

When considering smooth compactly supported initial data for the wave equation, taking R large enough, the same calculation gives us the energy identity

$$E_{\Sigma_t} = E_{\Sigma_0} \text{ for all } t \in \mathbb{R}.$$

This allows us to prove the well-posedness of the Cauchy problem in the finite energy space using a simple density argument on top of the smooth version of Theorem 4.1.

Theorem 4.9. *Given* $(\phi_0, \phi_1) \in \dot{H}^1(\mathbb{R}^3) \times L^2(\mathbb{R}^3)$*, the Cauchy problem*

$$\left\{\begin{array}{l}\partial_t^2\phi - \Delta\phi = 0 \text{ on } \mathbb{R}\times\mathbb{R}^3\,,\\ \phi|_{t=0} = \phi_0\,, \partial_t\phi|_{t=0} = \phi_1\,,\end{array}\right.$$

admits a unique solution $\phi \in \mathscr{C}(\mathbb{R}_t \times \dot{H}^1(\mathbb{R}^3)) \cap \mathscr{C}^1(\mathbb{R}_t\,;\ L^2(\mathbb{R}^3))$*. Its energy*

$$E_{\Sigma_t}(\phi) = \|\phi(t)\|^2_{\dot{H}^1} + \|\partial_t\phi(t)\|^2_{L^2}$$

is constant in time.

4.2.3 Energy estimates on a general space-time

The multiplier technique can be used in a general curved framework just as in the flat case using a local coordinate system. We present here the geometrical method involving a stress-energy tensor and a choice of observer (or vector field in general), for the wave equation

$$\Box_g \phi = 0\,, \tag{4.27}$$

on a space-time $(\mathcal{M}, g)$. Equation (4.27) has a conserved stress-energy tensor

$$T_{ab} = \partial_a\phi\partial_b\phi - \frac{1}{2}\langle\nabla\phi\,,\ \nabla\phi\rangle_g g_{ab}\,, \tag{4.28}$$

satisfying

$$\nabla^a T_{ab} = (\nabla_b\phi)\,\Box_g\phi\,,$$

the proof being identical to the flat case.

Once again, we have a conservation law that cannot be used directly and we must contract T_{ab} with a vector field V^a (usually timelike but not always) in order to get an energy current

$$J^a = K^b T^a_b\,.$$

In a general situation, we have no Killing vector field and we get the following expression for the divergence of the energy current:

$$\nabla^a J_a = \nabla_V\phi\,\Box_g\phi + T_{ab}\nabla^{(a}V^{b)}\,,$$

which, for ϕ solution to (4.27), simplifies to

$$\nabla^a J_a = T_{ab}\nabla^{(a}V^{b)}\,, \tag{4.29}$$

Remark 4.9. *Note that T_{ab} satisfies the dominant energy condition, the proof being identical to the flat case using an orthonormal basis at each point.*

Now consider S a closed hypersurface whose interior we denote by Ω, S being oriented by the outgoing normal. We have the following equality from the divergence theorem:

$$E_S = \int_\Omega T_{ab}\nabla^{(a}V^{b)}\mathrm{dVol}\,.$$

If V^a is causal and future-oriented, then we know that on parts of S where the outgoing normal is also causal and future-oriented, the flux is non-negative.

4.3 Lars Hörmander's treatment of the characteristic Cauchy problem

A crucial ingredient of a conformal scattering theory is the resolution of the characteristic Cauchy problem, or Goursat problem, with data set on null infinity. It is essential to understand that this is a global problem; the local Goursat problem, in the neighbourhood of a point on a null hypersurface, is ill-posed. Several methods are available that provide solutions to a Goursat problem for a wide class of hyperbolic equations. A classic approach is to find an integral formula for the solution using a Green function, i.e. a two-point function $G(p,q)$ such that, when the operator is applied to G in the variable p, it gives the Dirac distribution at the point q. Another no less classic approach uses energy estimates: it has been formulated very neatly by Lars Hörmander in a short paper in 1990 [25]. We present here the details of the proof.

The geometrical framework chosen by Hörmander is the following. Let X be a smooth compact manifold without boundary of dimension n, $n \geq 1$, and $\tilde{X} := \mathbb{R} \times X$. For $t \in \mathbb{R}$, denote $X_t = \{t\} \times X$ and endow X_t with a smooth Riemannian metric $g(t)$ that has a smooth dependence in t. Also consider a smooth density measure ν on X, $\mathrm{d}\nu = \gamma \mathrm{d}x$. The equation considered is then

$$\left(\partial_t^2 - \gamma^{-1}\partial_i\left(\gamma g^{ij}(t)\partial_j\right) + L_1\right)\phi = 0$$

where L_1 is a smooth first order differential operator. This is in fact equivalent to studying the wave equation with a smooth first order perturbation on a smooth globally hyperbolic space-time that is spatially compact[2] and this is how we shall present his result. Note that the choice of working on a spatially compact framework is for convenience but Hörmander's result can be extended to general globally hyperbolic space-times using domain of dependence arguments.

Let $\tilde{X}$ be a smooth globally hyperbolic $(n+1)$-dimensional space-time, $n \geq 1$, that is spatially compact. Then, using the results of Bernal and Sanchez [3] we have a smooth time function t on $\tilde{X}$ whose level hypersurfaces are Cauchy hypersurfaces and are diffeomorphic to a fixed n-dimensional compact manifold X without boundary. Using a global timelike vector field, we can therefore realize $\tilde{X}$ as $\mathbb{R} \times X$. With this identification, the level-hypersurfaces of t are simply $X_t = \{t\} \times X$. We chose the gradient of t for our timelike vector field. Since it is orthogonal to the level hypersurfaces of t, we can perform an orthogonal decomposition of the metric g into parts along ∇t and along the level hypersurfaces of t. We have

$$g = N^2 \mathrm{d}t^2 - h \tag{4.30}$$

where h is a smooth time-dependent Riemannian metric on X, but we see it here as a symmetric 2-form on $\tilde{X}$ such that $h_{ab}\nabla^a t = 0$, in fact a projector onto the tangent bundle of the foliation $\{X_t\}_t$, and N is the lapse-function, given by

$$N = \frac{1}{\sqrt{g(\nabla t, \nabla t)}}\,. \tag{4.31}$$

Indeed, using abstract indices, $\mathrm{d}t$ corresponds to $\nabla_a t$, ∇t to $\nabla^a t$ and we have

$$\mathrm{d}t(\nabla t) = \nabla_a t \nabla^a t = g_{ab}\nabla^a t \nabla^b t = g(\nabla t, \nabla t).$$

Hence, the expression (4.30) and the fact that $h_{ab}\nabla^a t = 0$ imply

$$g(\nabla t, \nabla t) = N^2\left(g(\nabla t, \nabla t)\right)^2,$$

[2]Meaning that each closed spacelike hypersurface is compact.

which gives (4.31). Identifying the slices X_t along the integral lines of ∇t also defines the vector field ∂_t as being proportional to ∇t and such that $\mathrm{d}t(\partial_t) = 1$. Hence, we have

$$\frac{\partial}{\partial t} = \alpha \nabla t,\ 1 = \mathrm{d}t\left(\frac{\partial}{\partial t}\right) = \alpha \mathrm{d}t(\nabla t) = \alpha g(\nabla t, \nabla t) = \frac{\alpha}{N^2}.$$

We obtain

$$\frac{\partial}{\partial t} = N^2 \nabla t = N\tau \tag{4.32}$$

where

$$\tau = \frac{1}{\sqrt{g(\nabla t, \nabla t)}} \nabla t \tag{4.33}$$

is the unit future oriented vector field proportional to ∇t.

On $\tilde{X}$, we consider a perturbed wave equation of the form

$$\square_g u + u + L_1 u = 0 \tag{4.34}$$

where L_1 is a general first order differential operator

$$L_1 = b^a \nabla_a + c$$

with smooth coefficients b^a and c This allows in particular to work with the wave equation or the Klein-Gordon equation by choosing $L_1 = m^2 - 1$ for $m = 0$ or $m > 0$ respectively.

Hörmander chose to specify the initial data on a hypersurface that can be a spacelike Cauchy hypersurface, a light cone, or anything in between. It is defined as follows

$$\Sigma = \{(\varphi(x), x)\, ;\ x \in X\}\ ,\ \varphi :\ X \longrightarrow \mathbb{R}\,, \tag{4.35}$$

where φ is simply a Lipschitz function on X, to allow for singularities such as the vertex of a light cone, and Σ is assumed to be "weakly spacelike", by which we mean that

$$g^{ab}(\varphi(x), x)\nabla_a(t - \varphi(x))\nabla_b(t - \varphi(x)) \geq 0 \text{ almost everywhere on } X\,. \tag{4.36}$$

Remark 4.10. *Here, we consider that the function φ is not merely defined on X but on the whole of $\tilde{X}$ and constant along the integral lines of ∇t. Condition (4.36) is meaningful, since Lipschitz functions are differentiable almost everywhere, and it simply says that Σ is allowed to be locally spacelike or null but not timelike, i.e. its normal vector field is required to be causal where it is defined.*

If Σ is spacelike everywhere, we are studying a standard Cauchy problem, with the difference that the hypersurface on which we set the data is quite rough. If Σ is almost everywhere null, we are looking at a Goursat problem. Hörmander's approach is to study both problems at the same time as part of a more general class of problems by allowing the hypersurface Σ to be locally spacelike or null.

The well-posedness in all Sobolev spaces of the Cauchy problem, with data set on X_0, for Equation (4.34) is given by Theorem 4.1 and Remark 4.4. It is the H^1 version that will be useful for us. In order to express it properly, we define an energy on the slices X_t inherited from an energy current. We consider the stress-energy tensor for the Klein-Gordon equation

$$\square_g \phi + \phi = 0$$

given by

$$T_{ab} = \nabla_a \phi \nabla_b \phi - \frac{1}{2} \langle \nabla \phi \, , \; \nabla \phi \rangle_g g_{ab} + \frac{1}{2} \phi^2 g_{ab} \, .$$

It satisfies, for ϕ a solution of (4.34),

$$\nabla^a T_{ab} = (\square_g \phi + \phi) \nabla_b \phi = -L_1 \phi \nabla_b \phi \, . \tag{4.37}$$

We define the energy current associated with the unit vector field τ defined in (4.33)

$$J^a := T^a_b \tau^b \, .$$

This current is of course not conserved but it satisfies an "approximate conservation law"

$$\nabla^a J_a = T_{ab} \nabla^{(a} \tau^{b)} - L_1 \phi \nabla_\tau \phi \, . \tag{4.38}$$

The energy on each slice X_t is the flux of J across X_t, it is given by

$$E_{X_t}(\phi) = \int_{X_t} *J_a \mathrm{d}x^a = \int_{X_t} J_a \tau^a (\tau \lrcorner \mathrm{dVol}).$$

We define the H^1 norm on X_t as follows

$$\| . \|_{H^1(X_t)} := \sqrt{E_{X_t}(.)}. \tag{4.39}$$

They are equivalent to the H^1 norm on X defined for any given Riemannian metric on X, this equivalence being locally uniform in time. A similar property holds for the $L^2(X_t)$ norms induced on each X_t by the metric g. We define the spaces $H^1(X)$ and $L^2(X)$ to be $H^1(X_0)$ and $L^2(X_0)$. We have the following result.

Theorem 4.10. *For any $(\phi_0 \, , \; \phi_1) \in H^1(X) \times L^2(X)$, there exists a unique solution of* (4.34)

$$\phi \in \mathscr{C}(\mathbb{R}_t \, ; \; H^1(X)) \cap \mathscr{C}^1(\mathbb{R}_t \, ; \; L^2(X))$$

such that

$$\phi|_{X_0} = \phi_0 \text{ and } \partial_t \phi|_{X_0} = \phi_1 \, .$$

In addition, for any $T > 0$, there is a constant $C > 0$ such that for each $\phi \in \mathscr{C}(\mathbb{R}_t \, ; \; H^1(X)) \cap \mathscr{C}^1(\mathbb{R}_t \, ; \; L^2(X))$ solution of (4.34),

$$E_{X_t}(\phi) \leq E_{X_s}(\phi) e^{C|t-s|} \; \forall t, s \in [-T, T]. \tag{4.40}$$

The last part of the theorem is established using Grönwall's inequality (Theorem A.14).

We denote by $\mathcal{E}$ the space of finite energy solutions of (4.34), i.e. the set of solutions of (4.34) in $\mathscr{C}(\mathbb{R}_t\,;\ H^1(X))\cap\mathscr{C}^1(\mathbb{R}_t\,;\ L^2(X))$. Note that this space can be canonically identified with $H^1(X)\times L^2(X)$ by taking the initial data for each solution. Using the energy on Σ of a solution ϕ, we also define a function space on Σ. It will be our natural space of data for the Cauchy problem on Σ. Let ϕ be a solution of (4.34) in $\mathcal{E}$, recall that the energy flux of ϕ across Σ is given by

$$E_\Sigma(\phi)=\int_\Sigma *J_a\mathrm{d}x^a=\int_\Sigma \tau^aT_{ab}\nu^b(\lambda\lrcorner\mathrm{dVol})\,,$$

where ν is a future-oriented normal vector field to Σ and λ a future-oriented transverse vector field to Σ such that $g(\lambda,\nu)=1$. Note that the measure $\lambda\lrcorner\mathrm{dVol}$ will be uniformly equivalent to the measure μ_Σ obtained from the measure induced by g on X_0 via the parametrization (4.35) of Σ. In our case, we have an explicit choice of ν given by

$$\nu=N\nabla(t-\varphi(x))=\tau-N\nabla\varphi.$$

We have $g(\nu,\tau)=1$ and we can therefore take $\lambda=\tau$. At points where $\nabla\varphi=0$, we have $\nu=\tau$ and we obtain the usual energy density on the X_t slices. At points where $\nabla\varphi\neq 0$, things will be different. If the vector ν remains timelike, the energy density will still be equivalent to that on the slices X_t, if ν becomes null however, the equivalence will be lost. Let us see this in details. The energy density is given by

$$\mathscr{E}=\tau^aT_{ab}\nu^b=\nabla_\tau\phi\nabla_\nu\phi-\frac12 g(\tau,\nu)g(\nabla\phi,\nabla\phi)+\frac12 g(\tau,\nu)\phi^2=\nabla_\tau\phi\nabla_\nu\phi-\frac12 g(\nabla\phi,\nabla\phi)+\frac12\phi^2$$

since $g(\tau,\nu)=g(\tau,\tau)=1$. Moreover, since $\nabla\varphi$ is orthogonal to ∇t, $g(\nabla\phi,\nabla\phi)=-h(\nabla\phi,\nabla\phi)$ and

$$\begin{aligned}\mathscr{E}&=(\nabla_\tau\phi)^2-\nabla_\tau\phi N\nabla_{\nabla\varphi}\phi-\frac12(\nabla_\tau\phi)^2+\frac12 h(\nabla\phi,\nabla\phi)+\frac12\phi^2\\&=\frac12\left((\nabla_\tau\phi)^2-2g(\nabla_\tau\phi N\nabla\phi,\nabla\varphi)+h(\nabla\phi,\nabla\phi)+\phi^2\right).\end{aligned}$$

Using again the same property of $\nabla\varphi$, we have

$$-g(\nabla_\tau\phi N\nabla\phi,\nabla\varphi)=h(\nabla_\tau\phi N\nabla\phi,\nabla\varphi)=h(\nabla\phi,\nabla_\tau\phi N\nabla\varphi).$$

Consequently,

$$\begin{aligned}\mathscr{E}&=\frac12\left((\nabla_\tau\phi)^2+2h(\nabla\phi,\nabla_\tau\phi N\nabla\varphi)+h(\nabla\phi,\nabla\phi)+\phi^2\right)\\&=\frac12\left((\nabla_\tau\phi)^2-h(\nabla_\tau\phi N\nabla\varphi,\nabla_\tau\phi N\nabla\varphi)+h(\nabla\phi+\nabla_\tau\phi N\nabla\varphi,\nabla\phi+\nabla_\tau\phi N\nabla\varphi)+\phi^2\right)\\&=\frac12\left(g(\tau-N\nabla\varphi,\tau-N\nabla\varphi)(\nabla_\tau\phi)^2+h(\nabla\phi+\nabla_\tau\phi N\nabla\varphi,\nabla\phi+\nabla_\tau\phi N\nabla\varphi)+\phi^2\right).\end{aligned}$$

If we consider a coordinate basis (t,x^1,x^2,x^3), denoting $e_i=\frac{\partial}{\partial x^i}$, the vectors

$$\frac{\partial}{\partial x^i}+\frac{\partial\varphi}{\partial x^i}\frac{\partial}{\partial t}=e_i+(N\nabla_{e_i}\varphi)\tau,\ \ i=1,2,3,$$

are a basis of the tangent space to Σ, so

$$h_{ab}\left(\nabla^b\phi + \nabla_\tau\phi N \nabla^b\varphi\right)$$

is a projection of $\nabla\phi$ onto the tangent space to Σ. We denote

$$h(\nabla\phi + \nabla_\tau\phi N\nabla\varphi, \nabla\phi + \nabla_\tau\phi N\nabla\varphi) = |\nabla\phi|_\Sigma|^2$$

and the energy density becomes

$$\mathscr{E} = \frac{1}{2}\left(g(\nu,\nu)(\nabla_\tau\phi)^2 + |\nabla\phi|_\Sigma|^2 + \phi^2\right).$$

Denoting by

$$\mathrm{dVol}_\Sigma = \lambda \lrcorner \mathrm{dVol} = \tau \lrcorner \mathrm{dVol}\,,\ \mu^0_\Sigma = g(\nu,\nu)\mathrm{dVol}_\Sigma, \tag{4.41}$$

the energy on Σ has the following form

$$E_\Sigma(\phi) = \frac{1}{2}\int_\Sigma (\nabla_\tau\phi)^2 \mathrm{d}\mu^0_\Sigma + \frac{1}{2}\int_\Sigma \left(|\nabla\phi|_\Sigma|^2 + \phi^2\right)\mathrm{dVol}_\Sigma\,. \tag{4.42}$$

Considering ϕ and $\nabla_\tau\phi$ as independent functions on Σ, the energy (4.42) induces a semi-norm on the pairs $(\phi, \nabla_\tau\phi)$, denoted by $\|(\phi,\nabla_\tau\phi)\|_{\mathcal{E}_\Sigma}$. It only fails to be a norm at points where Σ is null, by losing control over $\nabla_\tau\phi$. We denote by $\mathcal{E}_\Sigma$ the completion of the space of pairs of smooth functions[3] on Σ in the semi-norm $\|(.,.)\|_{\mathcal{E}_\Sigma}$. If Σ is null (almost) everywhere, the measure μ^0_Σ vanishes and $\mathcal{E}_\Sigma$ is simply a natural H^1 space on Σ, it really is a space of real valued functions, not pairs of them. In general, the space $\mathcal{E}_\Sigma$ can be understood as follows

$$\mathcal{E}_\Sigma \simeq H^1(\Sigma) \oplus L^2(\Sigma; \mathrm{d}\nu^0_\Sigma)\,. \tag{4.43}$$

where $H^1(\Sigma)$ is simply defined as $H^1(X)$ considered as a set of functions on Σ via the parametrisation (4.35).

Remark 4.11. *Since Σ is merely assumed to be Lipschitz, using such a definition, we only have access to $H^s(\Sigma)$ for $0 \leq s \leq 1$ and by duality this is naturally extended to $s \in [-1, 1]$.*

The case of a light-cone at infinity on a conformally compactified space-time is what we are interested in for conformal scattering theory, so we express and prove Hörmander's theorem in this particular case. The Cauchy problem is then fully characteristic and is also called a Goursat problem. More precisely, we assume that $n = 3$ and that Σ satisfies the property

(P) Σ is null and smooth except at isolated points where φ is not differentiable.

Definition 4.3. *We denote by $\dot{\mathscr{C}}^\infty(\Sigma)$ the space of smooth fonctions on Σ supported away from the isolated points at which φ is not differentiable.*

[3]Smooth functions on Σ can be defined as restrictions to Σ of smooth functions on $\tilde{X}$, their regularity is in fact limited by that of Σ.

Since Σ is of dimension 3, the space $\dot{\mathscr{C}}^\infty(\Sigma)$ is dense in $H^s(\Sigma)$ for $s \in [-1,1]$. The Goursat problem on Σ is solved by the following theorem.

Theorem 4.11. (Hörmander, 1990) *Assuming the property* **(P)***, the map*

$$\begin{array}{rccc} \mathbb{T}_\Sigma : & \mathcal{E} & \longrightarrow & H^1(\Sigma) \\ & \phi & \longmapsto & \phi_{|\Sigma}, \end{array} \tag{4.44}$$

which is well defined for smooth solutions, extends as an isomorphism. In particular, there exists a constant $C>0$ such that, for any $\phi \in \mathcal{E}$, we have

$$\|\mathbb{T}_\Sigma \phi\|^2_{H^1(\Sigma)} \leq C(\|\phi|_{t=0}\|^2_{H^1(X_0)} + \|\nabla_\mathcal{T}\phi|_{t=0}\|^2_{L^2(X_0)})$$

and

$$\|\phi|_{t=0}\|^2_{H^1(X_0)} + \|\nabla_\mathcal{T}\phi|_{t=0}\|^2_{L^2(X_0)} \leq C\,\|\mathbb{T}_\Sigma \phi\|^2_{H^1(\Sigma)}\ ,$$

or equivalently (adapting the constant C to take (4.43) into account)

$$E_\Sigma(\phi) \leq C E_{X_0}(\phi) \tag{4.45}$$

and

$$E_{X_0}(\phi) \leq C E_\Sigma(\phi)\,. \tag{4.46}$$

Proof. It is organised in two steps. First we establish energy estimates both ways for smooth solutions of (4.34), these are exactly estimates (4.45) and (4.46). This will entail that the trace operator $\mathbb{T}_\Sigma$ extends as a bounded linear operator from $\mathcal{E}$ to $H^1(\Sigma)$, that is one-to-one and with closed range. Then we construct solutions to the Goursat problem on Σ for smooth data supported away from the points where φ is not differentiable. This entails the density of the range of $\mathbb{T}_\Sigma$ in $H^1(\Sigma)$ and proves the theorem.

Step 1: energy estimates. We first establish them for smooth solutions of (4.34) using Grönwall's Lemma A.14. Let

$$t_0 = \min\{\varphi(x); x \in X\},\ t_1 = \max\{\varphi(x); x\in X\}.$$

We consider $T > t_1$ and the bounded domain

$$\Omega_T := \{(t,x) \in \tilde{X};\ \varphi(x) < t < T\}$$

whose boundary is made of Σ in the past and X_T in the future. Using the divergence Theorem 2.4, the outgoing flux of the energy current on the boundary of Ω_T is given by

$$E_{X_T} - E_\Sigma = \int_{\Omega_T} \nabla^a J_a \mathrm{dVol}. \tag{4.47}$$

The divergence of J is a quadratic form in ϕ and its first order derivatives, with smooth coefficients. It is therefore locally controlled by the energy density and this control is uniform on any bounded time interval. So from (4.47), we infer

$$E_{X_T} \leq E_\Sigma + \int_{t_0}^T C(t) E_{X_t} \mathrm{d}t \tag{4.48}$$

where $C(t)$ is a positive continuous function on $\mathbb{R}$. Using Grönwall's Lemma A.14, this implies the existence of a continuous positive function K on $[t_0, +\infty[$ such that for all $T > t_0$,

$$E(X_T) \leq K(T)E_\Sigma$$

and estimate (4.40) then gives (4.46).

For the converse estimate, for $t \geq t_0$ we define the hypersurfaces

$$S_t := \{(s,x);\ s = \max\{t, \varphi(x)\},\ x \in X\}.$$

In particular, we have $S_{t_0} = \Sigma$ and $S_t = X_t$ for $t \geq t_1$. For $t_0 \leq \sigma \leq T$ we consider the domain $\Omega_{\sigma,T}$ lying in the future of S_σ and in the past of S_T and we write an identity similar to (4.47)

$$E(S_T) - E(S_\sigma) = \int_{\Omega_{\sigma,T}} \nabla^a J_a \mathrm{dVol}.$$

The domain $\Omega_{\sigma,T}$ can be foliated by the part of the hypersurfaces X_t contained in it, on which the divergence of J is uniformly controlled by the energy density. Since for $\sigma \leq t \leq T$, $X_t \cap \Omega_{\sigma,T} \subset S_t$, this gives us the estimate, where C is a positive continuous function on $\mathbb{R}$,

$$E(S_\sigma) \leq E(S_T) + \int_\sigma^T C(t)E(S_t)\mathrm{d}t\,.$$

Using Grönwall's Lemma again, this entails that there exists $K > 0$ such that for any $t \in [t_0, T]$,

$$E(S_t) \leq KE(S_T),$$

which for $t = t_0$ gives (4.45) using (4.40). The estimates (4.45) and (4.46) then extend to finite energy solutions by density.

Step 2: existence of solutions to the Goursat problem. We consider some data $\phi_0 \in \dot{\mathscr{C}}^\infty(\Sigma)$ (see Definition 4.3). The method chosen by Hörmander is to slow down the propagation speed in order to make Σ uniformly spacelike. This can be done by replacing the metric g by

$$g_\lambda = \lambda N^2 \mathrm{d}t^2 - h \tag{4.49}$$

where $\lambda \in [\frac{1}{2}, 1[$. For this new metric, the hypersurface Σ is uniformly spacelike and there is a unique solution

$$\phi_\lambda \in \mathscr{C}(\mathbb{R}_t; H^1(X)) \cap \mathscr{C}^1(\mathbb{R}_t; L^2(X))$$

of equation

$$\Box_{g_\lambda}\phi = 0, \tag{4.50}$$

such that $\phi_\lambda|_\Sigma = \phi_0$ and $\partial_t\phi_\lambda|_\Sigma = 0$. Of course, Σ is not smooth, but the initial data are supported away from its singular points, so this makes no difference.

Performing energy estimates for (4.50), one easily checks that the family $\{\phi_\lambda;\ \lambda \in [\frac{1}{2}, 1[\}$ is bounded in $\mathscr{C}([-T,T]; H^1(X)) \cap \mathscr{C}^1([-T,T]; L^2(X))$ for any given $T > 0$ such that

$\Sigma \subset]-T,T[\times X$. More precisely, there exists a constant $C>0$ such that for all $\lambda \in [\frac{1}{2},1[$ and for all $t \in [-T,T]$,

$$\|\phi_\lambda(t)\|^2_{H^1(X)} + \|\partial_t \phi_\lambda(t)\|_{L^2(X)} \leq C. \tag{4.51}$$

This means that we can extract a sequence $\lambda_k \in [\frac{1}{2},1[$ that converges to 1 such that we have the following convergences:

$$\begin{aligned}
\phi_{\lambda_k} &\to \phi \text{ in } H^1(]-T,T[\times X) \text{ weak}, &(4.52)\\
\phi_{\lambda_k} &\to \phi \text{ in } H^s(]-T,T[\times X) \text{ strong, for any } s<1, &(4.53)\\
\partial_t\phi_{\lambda_k} &\to \partial_t\phi \text{ in } L^2(]-T,T[\times X) \text{ weak}, &(4.54)\\
\phi_{\lambda_k} &\to \phi \text{ in } L^\infty([-T,T];H^1(X)) \text{ weak}-*, &(4.55)\\
\partial_t\phi_{\lambda_k} &\to \partial_t\phi \text{ in } L^\infty([-T,T];L^2(X)) \text{ weak}-*, &(4.56)\\
\phi_{\lambda_k}|_{X_T} &\to \phi|_{X_T} \text{ in } H^1(X_T) \text{ weak}, &(4.57)\\
\partial_t\phi_{\lambda_k}|_{X_T} &\to \partial_t\phi|_{X_T} \text{ in } L^2(X_T) \text{ weak} &(4.58)
\end{aligned}$$

where (4.53) follows from (4.52) via the Rellich-Kondrachov Theorem and (4.55) and (4.56) are a consequence of the Banach-Alaoglou Theorem. From (4.53), we infer that ϕ_n converges towards ϕ in $\mathscr{D}'(]-T,T[\times X)$, which entails that ϕ is a solution of (4.34) in the sense of distributions. The same property plus a trace theorem for Sobolev spaces implies that the restriction of ϕ_n to Σ converges towards the restriction of ϕ to Σ in $H^s(\Sigma)$ for $0 \leq s < 1/2$. In particular, we see that the restriction of ϕ to Σ is equal to ϕ_0. The function ϕ is therefore a solution to the Goursat problem on Σ with data ϕ_0 but we still need to prove that $\phi \in \mathcal{E}$ and $\mathbb{T}_\Sigma \phi = \phi_0$. To do this, we first neet to gain some continuity in time for ϕ and $\partial_t\phi$, then we shall use the H^{-1} version of Theorem 4.1.

We have $\phi \in H^s(]-T,T[\times X)$ for any $s<1$ and for $s>1/2$

$$H^s(]-T,T[\times X) \hookrightarrow \mathscr{C}([-T,T];L^2(X)).$$

In addition, using the equation, we have that

$$\partial_t^2\phi \in L^\infty([-T,T];H^{-1}(X)).$$

Sofar, we have therefore established

$$\begin{aligned}
\phi &\in L^\infty([-T,T];H^1(X)), &(4.59)\\
\partial_t\phi &\in L^\infty([-T,T];L^2(X)), &(4.60)\\
\partial_t^2\phi &\in L^\infty([-T,T];H^{-1}(X)), &(4.61)\\
\phi &\in \mathscr{C}([-T,T];L^2(X)). &(4.62)
\end{aligned}$$

We can gain some time continuity for $\partial_t\phi$ using the principle of intermediate derivatives by Lions and Magenes. This is the following theorem that can be found in a very nice paper by Jérémie Joudioux [27] and is adapted from the general theorems 2.3 and 3.1 in [34].

Theorem 4.12. *If* $\phi \in L^2(]-T,T[;H^1(X))$ *and* $\partial_t^2\phi \in L^2(]-T,T[;H^{-1}(X))$*, then*

$$\phi \in \mathscr{C}([-T,T];L^2(X)) \text{ and } \partial_t\phi \in \mathscr{C}([-T,T];H^{-1/2}(X)). \tag{4.63}$$

So we have

$$\phi \in \mathscr{C}([-T,T];L^2),\ \partial_t\phi \in \mathscr{C}([-T,T];H^{-1/2}),\ \phi(T) \in H^1(X) \text{ and } \partial_t\phi(T) \in L^2(X).$$

Using Theorem 4.1 and Remark 4.4, the Cauchy problem is well-posed for equation (4.34) in any Sobolev space. The function ϕ is a solution in $\mathscr{C}([-T,T];L^2(X))\cap\mathscr{C}^1([-T,T];H^{-1/2}(X))$ of the Cauchy problem for (4.34) with data at $t=T$ in $H^1(X)\times L^2(X)$. Using uniqueness of solutions in $\mathscr{C}(\mathbb{R}_t;L^2(X))\cap\mathscr{C}^1(\mathbb{R}_t;H^{-1}(X))$ and existence of a solution in $\mathscr{C}(\mathbb{R}_t;H^1(X))\cap\mathscr{C}^1(\mathbb{R}_t;L^2(X))$, we infer that ϕ belongs to $\mathscr{C}([-T,T];H^1(X))\cap\mathscr{C}^1([-T,T];L^2(X))$. □

4.4 Higher spin equations on flat space-time

4.5 Linear wave equations, conformal invariance

Some covariant equations (i.e. equations whose structure comes entirely from the metric) have a property called conformal invariance, which means that their solutions, after some appropriate rescaling, satisfy the same equation for the conformally rescaled metric.

Definition 4.4 (Conformal invariance)**.** *The conformal invariance of a covariant equation means that there exists* $s \in \mathbb{R}$ *such that a field* ϕ *satisfies the equation for the metric* g *if and only if* $\Omega^s\phi$ *satisfies the equation for* $\hat{g} = \Omega^2 g$.

The wave equation itself is not conformally invariant, however, a slight modification of it is. This modified equation involves the scalar curvature and we shall refer to it as the **conformal wave equation**:

$$\Box_g\phi + \frac{1}{6}\mathrm{Scal}_g\phi = 0\,. \tag{4.64}$$

We have the following fundamental result that is a corollary of Theorem 2.7.

Corollary 4.3. *We consider a space-time* $(\mathcal{M},g)$ *and a metric* $\hat{g}$ *in the conformal class of* g *with conformal factor* Ω*, i.e.* $\hat{g}=\Omega^2 g$*. Then we have the equality of operators acting on scalar fields on* $\mathcal{M}$*:*

$$\Box_g + \frac{1}{6}\mathrm{Scal}_g = \Omega^3\left(\Box_{\hat{g}} + \frac{1}{6}\mathrm{Scal}_{\hat{g}}\right)\Omega^{-1}\,. \tag{4.65}$$

Proof. We express the right-hand side of (4.65) in terms of the conformal factor Ω and the metric g, using the relation between the scalar curvatures of $\hat{g}$ and g given in Theorem 2.7:

$$\begin{aligned}\Omega^3\left(\Box_{\hat{g}} + \frac{1}{6}\mathrm{Scal}_{\hat{g}}\right)\Omega^{-1} &= \Omega^3\left(\Box_{\hat{g}} + \frac{1}{6}\Omega^{-2}\mathrm{Scal}_g + \Omega^{-3}\Box_g\Omega\right)\Omega^{-1}\\ &= \Omega^3\Box_{\hat{g}}\Omega^{-1} + \frac{1}{6}\mathrm{Scal}_g + \Omega^{-1}(\Box_g\Omega)\,.\end{aligned}$$

We apply the first term of the right-hand side to a scalar field ϕ and develop the expression in a local coordinate basis

$$\begin{aligned}
\Omega^3\Box_{\hat{g}}\Omega^{-1}\phi &= \Omega^3\frac{1}{\sqrt{|\hat{g}|}}\partial_{\mathbf{a}}\sqrt{|\hat{g}|}\,\hat{g}^{\mathbf{ab}}\partial_{\mathbf{b}}\Omega^{-1}\phi \\
&= \Omega^{-1}\frac{1}{\sqrt{|g|}}\partial_{\mathbf{a}}\Omega^2\sqrt{|g|}\,g^{\mathbf{ab}}\partial_{\mathbf{b}}\Omega^{-1}\phi \\
&= 2(\partial_{\mathbf{a}}\Omega)g^{\mathbf{ab}}\partial_{\mathbf{b}}(\Omega^{-1}\phi)+\Omega\Box_g(\Omega^{-1}\phi) \\
&= 2\langle\nabla\Omega\,,\ \nabla(\Omega^{-1}\phi)\rangle_g+\Omega(\Box_g\Omega^{-1})\phi+2\Omega\langle\nabla\Omega^{-1}\,,\ \nabla\phi\rangle_g+\Box_g\phi \\
&= \Box_g\phi+2\Omega^{-1}\langle\nabla\Omega\,,\ \nabla\phi\rangle_g-2\Omega^{-2}\langle\nabla\Omega\,,\ \nabla\Omega\rangle_g\phi-2\Omega^{-1}\langle\nabla\Omega\,,\ \nabla\phi\rangle_g \\
&\quad -\Omega\phi\nabla_a(\Omega^{-2}\nabla^a\Omega) \\
&= \Box_g\phi-2\Omega^{-2}\langle\nabla\Omega\,,\ \nabla\Omega\rangle_g\phi+2\Omega^{-2}\langle\nabla\Omega\,,\ \nabla\Omega\rangle_g\phi-\Omega^{-1}(\Box_g\Omega)\phi \\
&= \Box_g\phi-\Omega^{-1}(\Box_g\Omega)\phi\,.
\end{aligned}$$

Putting things together gives (4.65) and proves the theorem. □

This has the immediate consequence:

Corollary 4.4. *Let $\phi\in\mathcal{D}'(\mathcal{M})$, the following conditions are equivalent:*

1. *ϕ satisfies (4.64) in the sense of distributions on $\mathcal{M}$;*
2. *$\hat{\phi}:=\Omega^{-1}\phi$ satisfies*
$$\Box_{\hat{g}}\hat{\phi}+\frac{1}{6}\mathrm{Scal}_{\hat{g}}\hat{\phi}=0$$
in the sense of distributions on $\mathcal{M}$.

The higher spin zero rest-mass field equations are also conformally invariant.

Chapter 5

The Lax-Phillips approach

We present here an approach to scattering that is based on spectral analysis and makes contact with very geometrical structures: the Lax-Phillips theory [31]. Its essential ingredient is a translation representer of the evolution. We first explain the construction of the translation representation on the simple example of a differential system, then describe Lax and Phillips's treatment of the wave equation on $\mathbb{M}$ and its relation to the Whittaker formula. The Lax-Phillips theory is really about scattering by an obstacle, but we describe it here as a means to describe the asymptotic behaviour of solutions to the wave equation on Minkowski space-time. We shall see in Chapter 8 that it gives precise and complete results.

5.1 Finite dimensional case: translation representation

Consider the equation for a time-dependent vector in $\mathbb{C}^n$:

$$\partial_t V(t) = iAV(t)$$

where A is an $n \times n$ hermitian matrix A with n distinct eigenvalues σ_1, ..., σ_n. Let $\{e_1\,,\ \dots\,,\ e_n\}$ be an orthonormal basis of eigenvectors of A. A vector $V \in \mathbb{R}^n$ can be described as the function $\tilde{V}$ from $\mathbb{R}$ to itself that is zero everywhere except for

$$\tilde{V}(\sigma_i) := \langle V, e_i \rangle\,.$$

The vector AV is then simply represented as the function $\sigma \tilde{V}(\sigma)$, i.e. the action of A is represented as the multiplication by the spectral parameter σ. Similarly, the unitary group e^{itA} is described as the multiplication by $e^{it\sigma}$. This is a **spectral representation** of the matrix A and its associated unitary group.

A Fourier transform in σ then gives naturally a **translation representation** of the group:

$$\mathcal{F}_\sigma(\widetilde{e^{itA}V})(s) = \mathcal{F}_\sigma(e^{it\sigma}\tilde{V})(s) = \hat{\tilde{V}}(s-t)\,.$$

Remark 5.1. *Of course for it all to make sense, the Fourier transform must be understood on $\mathscr{S}'(\mathbb{R})$ or on a discrete L^2 space over the spectrum of A.*

5.2 The wave equation: spectral representation

Consider the wave equation on Minkowski space-time in its Hamiltonian form (4.14). Recall that the operator A is self-adjoint on $\mathscr{H} = \dot{H}^1(\mathbb{R}^3) \times L^2(\mathbb{R}^3)$, completion of $\mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$ in the norm

$$\|U\|^2 := \frac{1}{2}\int_{\mathbb{R}^3} (|\nabla_x u_1|^2 + |u_2|^2)\mathrm{d}x, \tag{5.1}$$

that corresponds to the energy on a spacelike slice. Let us first try to define a spectral representer for solutions to (4.14) following the same approach as in the finite dimensional case.

We look for the eigenvalues of A, i.e. $\sigma \in \mathbb{R}$ such that the equation $AU = \sigma U$ admits non trivial solutions in $\mathscr{H}$. This equation reads:

$$\left\{ \begin{array}{rcl} u_2 &=& i\sigma u_1\,, \\ \Delta u_1 &=& i\sigma u_2\,, \\ &=& -\sigma^2 u_1\,. \end{array} \right. \tag{5.2}$$

A first observation is that there are in fact no eigenvalues in the spectrum of A.

Lemma 5.1. *The system (5.2) has no solution in $\mathscr{H}$, i.e. the point-spectrum of A is empty.*

Proof. Taking the Fourier transform of (5.2) gives

$$(\sigma^2 - |\xi|^2)\hat{u}_1 = 0\,. \tag{5.3}$$

Hence, $\mathrm{supp}(\hat{u}_1) \subset \{|\xi| = |\sigma|\}$. The same is therefore true of the support of $\widehat{\nabla u_1}$, i.e. the support of $\widehat{\nabla u_1}$ is negligeable and since $\widehat{\nabla u_1}$ is an element of $L^2(\mathbb{R}^3)$, it follows that $\widehat{\nabla u_1} = 0$, which in turn implies that u_1 is constant. Now using the fact that $u_1 \in \dot{H}^1(\mathbb{R}^3)$, which means not only that $\nabla u_1 \in L^2(\mathbb{R}^3)$ but also that there exists a sequence of smooth compactly supported functions whose gradient converges to that of u_1 in $L^2(\mathbb{R}^3)$, we get that $u_1 = 0$, using for example Hardy's inequality that states that in dimension $n \geq 3$, we have

$$\left\| \frac{f}{\|x\|} \right\|_{L^2(\mathbb{R}^n)} \leq \frac{2}{n-2} \|\nabla f\|_{L^2(\mathbb{R}^n)}\,.$$

We also have $u_2 = 0$ by the first equation. □

However (5.3), and therefore also (5.2), has solutions in $\mathscr{S}'(\mathbb{R}^3)$. Among these are the distributions whose Fourier transform is equal to the Dirac measure at $\sigma\omega$ for any $\omega \in S^2$. These are given, up to a multiplicative constant, by

$$e_{\sigma,\omega}(x) = \left(\begin{array}{c} e^{-i\sigma x.\omega} \\ i\sigma e^{-i\sigma x.\omega} \end{array} \right),\ \omega \in S^2.$$

We see that for each $\sigma \in \mathbb{R}$, we have a whole 2-sphere of solutions (except for $\sigma = 0$ in which case the sphere degenerates to a point). These are the initial data for the plane wave solutions to the wave equation

$$\phi(t,x) = e^{i\sigma(t - x.\omega)}.$$

Let us assume for now that what we have here is a continuous analogue of the basis in the finite dimensional toy model, we shall verify this shortly. For $U \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$, we put

$$\tilde{U}(\sigma,\omega) := \frac{1}{(2\pi)^{3/2}} \langle U, e_{\sigma,\omega}\rangle_{\mathscr{H}} \tag{5.4}$$

where the inner product is the one associated to the norm (5.1), i.e.

$$\langle U, V\rangle_{\mathscr{H}} = \frac{1}{2}\int_{\mathbb{R}^3} \left(\nabla u_1 \overline{\nabla v_1} + u_2 \overline{v_2}\right) \mathrm{d}x. \tag{5.5}$$

The right-hand side in (5.4) does not completely make sense since $e_{\sigma,\omega} \notin \mathscr{H}$ but the formal expression of the inner product is meaningful thanks to the compact support of U. More appropriately, we can define[1]

$$\tilde{U}(\sigma,\omega) := \frac{1}{(2\pi)^{3/2}} \langle U, \chi e_{\sigma,\omega}\rangle_{\mathscr{H}} \tag{5.6}$$

where $\chi \in \mathscr{C}_0^\infty(\mathbb{R}^3)$ and $\chi \equiv 1$ on $\mathrm{supp}(U)$. The definition (5.6) is independent of the choice of χ and

$$\begin{aligned}
\tilde{U}(\sigma,\omega) &= \frac{1}{2}\frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} (\nabla u_1 \overline{\nabla(\chi e^{-i\sigma x.\omega})} + \chi u_2 \overline{i\sigma e^{-i\sigma x.\omega}}) \mathrm{d}x \\
&= \frac{1}{2}\frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} (\nabla u_1 \overline{\nabla e^{-i\sigma x.\omega}} + u_2 \overline{i\sigma e^{-i\sigma x.\omega}}) \mathrm{d}x
\end{aligned}$$

since $\chi \equiv 1$ on the support of U. Then, we have

$$\begin{aligned}
\tilde{U}(\sigma,\omega) &= \frac{1}{2}\frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} (u_1 (\overline{-\Delta e^{-i\sigma x.\omega}}) + u_2 \overline{i\sigma e^{-i\sigma x.\omega}}) \mathrm{d}x \\
&= \frac{1}{2}\frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} (\sigma^2 u_1 - i\sigma u_2) e^{i\sigma x.\omega} \mathrm{d}x \\
&= \frac{1}{2}(\sigma^2 \hat{u}_1(-\sigma\omega) - i\sigma \hat{u}_2(-\sigma\omega)) .
\end{aligned}$$

We now have a result which proves that we have found enough tempered distribution solutions of (5.2) to represent all elements of $\mathscr{H}$.

Proposition 5.1. *The definition for $U \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$ of*

$$\tilde{U}(\sigma,\omega) = \frac{1}{2}(\sigma^2 \hat{u}_1(-\sigma\omega) - i\sigma \hat{u}_2(-\sigma\omega)) , \tag{5.7}$$

extends to $U \in \mathscr{H}$ and the map that to U associates $\tilde{U}$ is then an isometry from $\mathscr{H}$ onto $L^2(\mathbb{R}_\sigma \times S^2_\omega)$.

[1] We could also make sense of this inner product as a duality bracket.

Proof. We put for $U \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$,

$$\tilde{U}(\sigma, \omega) := \frac{1}{2}(\sigma^2 \hat{u}_1(-\sigma\omega) - i\sigma\hat{u}_2(-\sigma\omega)) .$$

Let us show that $U \mapsto \tilde{U}$ is a linear continuous map from $\mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$ to $L^2(\mathbb{R}_\sigma \times S^2_\omega)$ for the norms $\|.\|_{\mathscr{H}}$ and $\|.\|_{L^2(\mathbb{R}_\sigma \times S^2_\omega)}$. For $U \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$, we have

$$4|\tilde{U}(\sigma,\omega)|^2 = \sigma^4 |\hat{u}_1(-\sigma\omega)|^2 + \sigma^2|\hat{u}_2(-\sigma\omega)|^2 + i\sigma^3 \hat{u}_1(-\sigma\omega)\bar{\hat{u}}_2(-\sigma\omega) - i\sigma^3 \bar{\hat{u}}_1(-\sigma\omega)\hat{u}_2(-\sigma\omega) .$$

When changing the signs of σ and ω, the last two terms change sign and therefore their integral over $\mathbb{R}_\sigma \times S^2_\omega$ vanishes. Whence,

$$\begin{aligned}
\|\tilde{U}\|^2_{L^2(\mathbb{R}_\sigma \times S^2_\omega)} &= \frac{1}{4}\int_{\mathbb{R}\times S^2} \left(\sigma^2|\hat{u}_1(-\sigma\omega)|^2 + |\hat{u}_2(-\sigma\omega)|^2\right)\sigma^2 \mathrm{d}\sigma\mathrm{d}\omega \\
&= \frac{1}{2}\int_{\mathbb{R}^3} \left(|\widehat{\nabla_x u_1}(\xi)|^2 + |\hat{u}_2(\xi)|^2\right)\mathrm{d}\xi \\
&= \|U\|^2_{\mathscr{H}} .
\end{aligned}$$

It follows that $U \mapsto \tilde{U}$ extends as a linear continuous map from $\mathscr{H}$ to $L^2(\mathbb{R} \times S^2)$ and that map is one-to-one and has closed range. We therefore only need to prove that its range is dense in order to prove that it is an isometry. Let us consider the dense subspace of $L^2(\mathbb{R} \times S^2)$

$$\mathcal{K} := \left\{\phi \in \mathscr{C}_0^\infty(\mathbb{R} \times S^2)\,;\ \mathrm{supp}(\phi) \cap \{\sigma = 0\} = \emptyset\right\} .$$

Let $F \in \mathcal{K}$, let us find $f \in \mathscr{H}$ such that $F = \tilde{f}$. We must have

$$\begin{aligned}
F(\sigma,\omega) &= \frac{1}{2}(\sigma^2 \hat{f}_1(-\sigma\omega) - i\sigma\hat{f}_2(-\sigma\omega)) , \\
F(-\sigma,-\omega) &= \frac{1}{2}(\sigma^2 \hat{f}_1(-\sigma\omega) + i\sigma\hat{f}_2(-\sigma\omega))
\end{aligned}$$

and therefore

$$\begin{aligned}
\hat{f}_1(-\sigma\omega) &= \frac{F(\sigma,\omega) + F(-\sigma,-\omega)}{\sigma^2} , \\
\hat{f}_2(-\sigma\omega) &= i\frac{F(\sigma,\omega) - F(-\sigma,-\omega)}{\sigma} .
\end{aligned}$$

Putting $\xi = -\sigma\omega$, we have

$$\begin{aligned}
\hat{f}_1(\xi) &= \frac{F\left(|\xi|, -\frac{\xi}{|\xi|}\right) + F\left(-|\xi|, \frac{\xi}{|\xi|}\right)}{|\xi|^2} , \\
\hat{f}_2(\xi) &= i\frac{F\left(|\xi|, -\frac{\xi}{|\xi|}\right) - F\left(-|\xi|, \frac{\xi}{|\xi|}\right)}{|\xi|} ,
\end{aligned}$$

and these two functions belong to $\mathscr{C}_0^\infty(\mathbb{R}^3)$ because $F \in \mathcal{K}$ and its support therefore remains away from $\{\sigma = 0\}$. The functions f_1 and f_2 are hence defined by

$$\begin{aligned} f_1 &= \mathcal{F}_\xi^{-1}\left(\frac{F\left(|\xi|, -\frac{\xi}{|\xi|}\right) + F\left(-|\xi|, \frac{\xi}{|\xi|}\right)}{|\xi|^2}\right) \in \mathscr{S}(\mathbb{R}^3)\,, \\ f_2 &= \mathcal{F}_\xi^{-1}\left(i\frac{F\left(|\xi|, -\frac{\xi}{|\xi|}\right) - F\left(-|\xi|, \frac{\xi}{|\xi|}\right)}{|\xi|}\right) \in \mathscr{S}(\mathbb{R}^3) \end{aligned}$$

and $f = (f_1, f_2) \in \mathscr{H}$. This concludes the proof. □

Proposition 5.2. *This provides a spectral representation of A and its propagator, i.e.*

$$\forall U \in D(A)\,,\ \widetilde{AU} = \sigma \tilde{U}\,, \tag{5.8}$$

$$\forall U \in \mathscr{H}\,,\ \widetilde{e^{itA}U} = e^{it\sigma}\tilde{U}\,. \tag{5.9}$$

Proof. We prove (5.8). Let $U \in D(A)$, we choose a sequence $(U^n)_{n\in\mathbb{N}} \subset \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$ such that $U^n \to U$ in $D(A)$ as $n \to +\infty$. For $n \in \mathbb{N}$, consider $\chi \in \mathscr{C}_0^\infty(\mathbb{R}^3)$ that is identically equal to 1 on an open neighbourhood of the support of U^n. We have

$$\begin{aligned} \widetilde{AU^n}(\sigma, \omega) &= \langle AU^n, \chi e_{\sigma,\omega}\rangle_{\mathscr{H}} \\ &= \langle U^n, A\chi e_{\sigma,\omega}\rangle_{\mathscr{H}} \text{ since } A \text{ is self-adjoint on } \mathscr{H}\,, \\ &= \langle U^n, \chi A e_{\sigma,\omega}\rangle_{\mathscr{H}} \text{ thanks to the definition of } \chi\,, \\ &= \langle U^n, \chi\sigma e_{\sigma,\omega}\rangle_{\mathscr{H}} = \sigma\widetilde{U^n}(\sigma,\omega)\,. \end{aligned}$$

Taking the limit as $n \to +\infty$ and using the continuity of $U \mapsto \tilde{U}$ on $\mathscr{H}$, we get

$$\widetilde{AU} = \sigma\tilde{U}\,.$$

The property (5.9) can be proved using the same method. □

5.3 The wave equation: translation representation

The translation representer is obtained from the spectral representer by taking the Fourier transform in σ. We denote

$$\mathcal{R}U(s,\omega) := \mathcal{F}_\sigma(\tilde{U}(.,\omega))(s)\,. \tag{5.10}$$

Then

$$\mathcal{R}(e^{itA}U)(s,\omega) = (\mathcal{R}U)(s-t,\omega)\,. \tag{5.11}$$

This representation is of course also an isometry from $\mathscr{H}$ onto $L^2(\mathbb{R}\times S^2)$.

5.4 Link with the Radon transform and asymptotic profiles

Sofar, we have constructed a translation representer that encodes completely the initial data for a solution to the wave equation, but it is not yet clear what this has to do with a scattering theory. In this section, we start by giving the explicit inverse map to the translation representer, which turns out to be the Whittaker formula. Then we use this formula to give a description of the translation representer as the scattering data for the solution.

Definition 5.1. *Let $f \in \mathscr{C}_0^\infty(\mathbb{R}^3)$, we define its Radon transform as the function of $s \in \mathbb{R}$ and $\omega \in S^2$:*

$$Rf(s,\omega) = \int_{x.\omega = s} f(x) \mathrm{d}\sigma(x)\,,$$

i.e. $Rf(s,\omega)$ is the integral of f on the plane with normal ω containing the point $s\omega$.

Proposition 5.3. *The Radon transform has the following properties:*

1. *If* $\mathrm{supp} f \subset B(0,r)$ *then* $\mathrm{supp}(Rf) \subset [-r,r] \times S^2$*;*
2. $Rf \in \mathscr{C}^\infty(\mathbb{R} \times S^2)$*;*
3. $Rf(s,\omega) = Rf(-s,-\omega)$*;*
4. $R(\partial_{x^k} f)(s,\omega) = \omega^k \partial_s (Rf)(s,\omega)$*, whence* $R(\Delta f) = \partial_s^2 (Rf)$*;*
5. *if* $g \in \mathscr{C}_0^\infty(\mathbb{R} \times S^2)$*, then*
$$\langle Rf, g \rangle_{L^2(\mathbb{R}\times S^2)} = \langle f, R^* g\rangle_{L^2(\mathbb{R}^3)}\,;$$
R^* *is the formal adjoint of* R *given by*
$$R^*\phi(x) = \int_{S^2} \phi(x.\omega,\omega)\mathrm{d}\omega\,.$$

The definition and proposition above allow us to express the translation representation in a simple manner in terms of the Radon transform as well as to find an explicit formula for its inverse.

Theorem 5.1. *For $U \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$, the translation representation has the following simple expression in terms of the Radon transform:*

$$\mathcal{R}U(s,\omega) = \frac{1}{4\pi}(-\partial_s^2 R u_1 + \partial_s R u_2)(s,\omega)\,.$$

Moreover, the map

$$\begin{aligned} \mathcal{I} \,:\, & \mathscr{C}_0^\infty(\mathbb{R}\times S^2) \to \mathscr{C}^\infty(\mathbb{R}^3) \times \mathscr{C}^\infty(\mathbb{R}^3)\,,\\ (\mathcal{I}k)(x) &= \left(\frac{1}{2\pi} R^* k, -\frac{1}{2\pi} R^* \partial_s k \right)\,, \end{aligned} \tag{5.12}$$

extends as an isometry from $L^2(\mathbb{R}\times S^2)$ onto $\mathscr{H}$ which is the inverse of $\mathcal{R}$, i.e.

$$\mathcal{R}\mathcal{I} = \mathrm{Id}_{L^2(\mathbb{R}\times S^2)}\,,\ \mathcal{I}\mathcal{R} = \mathrm{Id}_{\mathscr{H}}\,.$$

Proof. Let $U \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$, we have

$$\begin{aligned} \mathcal{R}U(s,\omega) &= \mathcal{F}_\sigma(\tilde{U}(\sigma,\omega))(s) \\ &= \mathcal{F}_\sigma(\frac{1}{2}(\sigma^2\hat{u}_1(-\sigma\omega) - i\sigma\hat{u}_2(-\sigma\omega)))(s) \\ &= \frac{1}{2}\left(-\partial_s^2\mathcal{F}_\sigma(\hat{u}_1(-\sigma\omega))(s) + \partial_s\mathcal{F}_\sigma(\hat{u}_2(-\sigma\omega))(s)\right) . \end{aligned} \tag{5.13}$$

Now, given $f \in \mathscr{C}_0^\infty(\mathbb{R}^3)$, $\sigma \in \mathbb{R}$ and $\omega \in S^2$,

$$\begin{aligned} \hat{f}(-\sigma\omega) &= \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}^3} e^{i\sigma\langle\omega,x\rangle} f(x)\mathrm{d}x \\ &= \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}} \int_{\{x.\omega=s\}} e^{i\sigma\langle\omega,x\rangle} f(x)\mathrm{d}\sigma(x)\mathrm{d}s \\ &= \frac{1}{(2\pi)^{3/2}} \int_{\mathbb{R}} e^{i\sigma s} Rf(s,\omega)\mathrm{d}s = \frac{1}{2\pi}\mathcal{F}_s^{-1}(Rf(s,\omega))(\sigma) . \end{aligned} \tag{5.14}$$

From (5.13) and (5.14), we infer

$$\mathcal{R}U(s,\omega) = \frac{1}{4\pi}(-\partial_s^2 Ru_1 + \partial_s Ru_2)(s,\omega) .$$

Let us now turn to the inverse of the translation representation. Since $\mathcal{R}$ is an isometry and $\mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$ is dense in $\mathscr{H}$, it follows that $\mathcal{R}(\mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3))$ is dense in $L^2(\mathbb{R} \times S^2)$. In addition, thanks to properties *1.* and *2.* of the Radon transform,

$$\mathcal{R}(\mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)) \subset \mathscr{C}_0^\infty(\mathbb{R} \times S^2) .$$

Let $k \in \mathcal{R}(\mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3))$, there exists a unique $F \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$ such that

$$k = \mathcal{R}F . \tag{5.15}$$

Using the fact that $\mathcal{R}$ is an isometry, (5.15) is equivalent to

$$\langle F, G\rangle_{\mathscr{H}} = \langle k, \mathcal{R}G\rangle_{L^2(\mathbb{R}\times S^2)} \text{ for all } G \in \mathscr{H} .$$

If we restrict ourselves to $G \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$, the left-hand side can be written as a distribution acting on a test function on $\mathbb{R}^3$

$$\begin{aligned} \langle F, G\rangle_{\mathscr{H}} &= \frac{1}{2}\int_{\mathbb{R}^3} \left(\nabla f_1 \overline{\nabla g_1} + f_2\overline{g_2}\right)\mathrm{d}x \\ &= \frac{1}{2}\langle -\Delta f_1, g_1\rangle_{\mathscr{D}',\mathscr{D}} + \frac{1}{2}\langle f_2, g_2\rangle_{\mathscr{D}',\mathscr{D}} . \end{aligned}$$

We endeavour to do the same with the right-hand side in order to interpret the equality above as the equality of two distributions on $\mathbb{R}^3$

$$\begin{aligned} \langle k, \mathcal{R}G\rangle_{L^2(\mathbb{R}\times S^2)} &= \frac{1}{4\pi}\left\langle k(s,\omega), \left(-\partial_s^2 Rg_1 + \partial_s Rg_2\right)\right\rangle_{L^2(\mathbb{R}\times S^2)} \\ &= \frac{1}{4\pi}\left\langle -\partial_s^2 k(s,\omega), Rg_1\right\rangle_{L^2(\mathbb{R}\times S^2)} + \frac{1}{4\pi}\left\langle -\partial_s k, Rg_2\right\rangle_{L^2(\mathbb{R}\times S^2)} \\ &= -\frac{1}{4\pi}\langle R^*(\partial_s^2 k), g_1\rangle_{L^2(\mathbb{R}^3)} - \frac{1}{4\pi}\langle R^*(\partial_s k), g_2\rangle_{L^2(\mathbb{R}^3)} \\ &= -\frac{1}{4\pi}\langle R^*(\partial_s^2 k), \overline{g_1}\rangle_{\mathscr{D}',\mathscr{D}} - \frac{1}{4\pi}\langle R^*(\partial_s k), \overline{g_2}\rangle_{\mathscr{D}',\mathscr{D}} . \end{aligned}$$

We infer that (5.15) is equivalent to the following two equalities in $\mathscr{D}'(\mathbb{R}^3)$

$$-\frac{1}{2}\Delta f_1 = -\frac{1}{4\pi}R^*(\partial_s^2 k)\,, \tag{5.16}$$

$$\frac{1}{2}f_2 = -\frac{1}{4\pi}R^*(\partial_s k)\,. \tag{5.17}$$

Since the inverse Radon transform has the following property

$$\begin{aligned} \Delta(R^*k) &= \Delta_x \int_{S^2} k(x.\omega,\omega)\mathrm{d}\omega \\ &= \int_{S^2} |\omega|^2 \partial_s^2 k(x.\omega,\omega)\mathrm{d}^2\omega = R^*(\partial_s^2 k)\,, \end{aligned}$$

it follows that (5.16) can be written as

$$\Delta\left(f_1 - \frac{1}{2\pi}R^*k\right) = 0. \tag{5.18}$$

The function f_1 is smooth and compactly supported and although R^*k may not have compact support, it is easy to show that it tends to zero at infinity. Indeed, as $|x| \to \infty$, $|x.\omega| \to \infty$ for almost every $\omega \in S^2$ and we also have that k being smooth and compactly supported on $\mathbb{R}\times S^2$, $|k(x.\omega,\omega)|$ can be bounded by the sup of $|k|$ on $\mathbb{R}\times S^2$, which is finite and therefore integrable on S^2. As a consequence of Lebesgue's Theorem of dominated convergence, it follows that

$$R^*k(x) \to 0 \text{ as } |x| \to \infty\,.$$

Hence, the function

$$f_1 - \frac{1}{2\pi}R^*k$$

is harmonic and tends to zero at infinity; it is therefore identically zero. Hence, (5.17) and (5.18) give exactly $f = \mathcal{I}k$. Theorem 5.1 then follows by density. □

Remark 5.2. *Given a solution ϕ of the wave equation, formula (5.12) gives in particular $\phi(0,x)$ in terms of the translation representer k of ϕ constructed from the pair $(\phi(0,x),\partial_t\phi(0,x))$:*

$$\phi(0,x) = \frac{1}{2\pi}\int_{S^2} k(x.\omega,\omega)\mathrm{d}\omega\,,$$

and using the property (5.11), we get

$$\phi(t,x) = \frac{1}{2\pi}\int_{S^2} k(x.\omega - t,\omega)\mathrm{d}\omega\,,$$

which is exactly Whittaker's formula (4.13).

This can be used to establish the **asymptotic profile** property.

Theorem 5.2 (Asymptotic profiles)**.** *Assuming that the data ϕ_0, ϕ_1 are smooth and compactly supported, denoting*

$$k(s,\omega) = \mathcal{R}U(s,\omega)\,,$$

we have

$$k(s,\omega) = -\lim_{r\to+\infty} r\partial_t\phi(r-s,r\omega)\,. \tag{5.19}$$

Proof. First note that since the data are smooth and compactly supported, k is also smooth and compactly supported and there exists $R>0$ such that $\operatorname{supp} k \subset [-R,R]\times S^2$. We have

$$\phi(t,x) = \frac{1}{2\pi}\int_{S^2} k(x.\zeta - t,\zeta)\mathrm{d}\zeta$$

and since k is smooth, we can differentiate under the integral

$$\partial_t\phi(t,x) = -\frac{1}{2\pi}\int_{S^2}\partial_s k(x.\zeta - t,\zeta)\mathrm{d}\zeta\,.$$

In particular, we have

$$\begin{aligned} -r\partial_t\phi(r-s,r\omega) &= \frac{r}{2\pi}\int_{S^2}(\partial_s k)(r\omega.\zeta - r + s,\zeta)\mathrm{d}\zeta \\ &= \frac{r}{2\pi}\int_{S^2}(\partial_s k)(r(\omega.\zeta-1)+s,\zeta)\mathrm{d}\zeta\,. \end{aligned} \tag{5.20}$$

Thanks to the compact support of k, the integration will be done only over a neighbourhood of ω on S^2. As $r\to+\infty$, this neighbourhood will shrink and asymptotically reduce to ω itself. Indeed,

$$(r(\omega.\zeta-1)+s,\zeta)\notin \operatorname{supp} k \Leftrightarrow |r(\omega.\zeta-1)+s| > R$$

and since

$$|r(\omega.\zeta-1)+s| \geq |r(\omega.\zeta-1)| - |s|,$$

it follows that that if

$$|1-\omega.\zeta| > \frac{R+|s|}{r}$$

then $(r(\omega.\zeta-1)+s,\zeta)\notin\operatorname{supp} k$. Therefore, the integral (5.20) is in fact localised on the domain

$$V_{Rsr} = \left\{\zeta\in S^2\,;\ 1-\omega.\zeta \leq \frac{R+|s|}{r}\right\},$$

i.e.

$$-r\partial_t\phi(r-s,r\omega) = \frac{r}{2\pi}\int_{V_{Rsr}}(\partial_s k)(r(\omega.\zeta-1)+s,\zeta)\mathrm{d}\zeta\,.$$

We now add and subtract a term where the dependence in ζ is frozen

$$\begin{aligned} -r\partial_t\phi(r-s,r\omega) &= \frac{r}{2\pi}\int_{V_{Rsr}}\left[(\partial_s k)(r(\omega.\zeta-1)+s,\zeta) - (\partial_s k)(r(\omega.\zeta-1)+s,\omega)\right]\mathrm{d}\zeta \\ &\quad +\frac{r}{2\pi}\int_{V_{Rsr}}(\partial_s k)(r(\omega.\zeta-1)+s,\omega)\mathrm{d}\zeta\,. \end{aligned} \tag{5.21}$$

The area of V_{Rsr} is easy to calculate. Indeed, given $0 < \varepsilon < 1$, defining

$$V_\varepsilon := \{\zeta \in \mathbb{S}^2 \,;\; 1 - \zeta_3 \leq \varepsilon\}\,,$$

the elements of V_ε can be described in spherical coordinates as

$$\zeta = (\sin\theta\cos\varphi\,,\;\; \sin\theta\sin\varphi\,,\;\; \cos\theta)\,,\;\; 0 \leq \theta \leq \arccos(1-\varepsilon)\,.$$

Therefore the area of V_ε is given by

$$\mathcal{A}(V_\varepsilon) = \int_0^{2\pi}\int_0^{\arccos(1-\varepsilon)} \sin\theta \mathrm{d}\theta\mathrm{d}\varphi = 2\pi\left[-\cos\theta\right]_0^{\arccos(1-\varepsilon)} = 2\pi\varepsilon\,.$$

The area of V_{Rst} is therefore exactly given by

$$\mathcal{A}(V_{Rsr}) = 2\pi\frac{R+|s|}{r}\,.$$

This implies that the absolute value of the first term in (8.3) can be estimated by a constant times

$$r \times \frac{1}{r}\sup_{1-\omega.\zeta\leq\frac{R+|s|}{r}} |(\partial_s k)(r(\omega.\zeta-1)+s,\zeta) - (\partial_s k)(r(\omega.\zeta-1)+s,\omega)|$$

which, since k is smooth, in particular $\mathscr{C}^1$, tends to zero as $r \to +\infty$. We now do an explicit calculation for the second term that we denote by I. We use spherical coordinates (θ,φ) based on the direction ω: i.e. $\omega.\zeta = \cos\theta$. Putting

$$\theta_r = \arccos\left(1 - \frac{R+|s|}{r}\right)\,,$$

we have

$$\begin{aligned} I &= \frac{r}{2\pi}\int_0^{2\pi}\int_0^{\theta_r} (\partial_s k)(r(\cos\theta-1)+s,\omega)\sin\theta\mathrm{d}\theta\mathrm{d}\varphi \\ &= \int_0^{\theta_r}\left(-\frac{\mathrm{d}}{\mathrm{d}\theta}\left(k(r(\cos\theta-1)+s,\omega)\right)\right)\mathrm{d}\theta \\ &= k(s,\omega) - k(-(R+|s|)+s,\omega). \end{aligned}$$

Since

$$|-(R+|s|)+s| \geq R,$$

it follows that

$$k(-(R+|s|)+s,\omega) = 0,$$

whence

$$-\lim_{r\to+\infty} r\partial_t\phi(r-s,r\omega) = k(s,\omega).$$

This proves the theorem. $\square$

This allows us to interpret the translation representer as a scattering datum in the future. Indeed the translation representer k for the data (ϕ_0, ϕ_1) is interpreted by Theorem 5.2 as the restriction to $\mathscr{I}^+$ of $\partial_t \tilde{\phi}$, where $\tilde{\phi} = r\phi$. A natural way of defining scattering data is as

$$\tilde{\phi}^{\pm} := \tilde{\phi}|_{\mathscr{I}^{\pm}}$$

which is exactly what we do in our conformal scattering theory in Section 7.6. Since the Cauchy data are smooth and compactly supported and ∂_t extends to $\mathscr{I}^+$ as its future oriented null generator, there is an equivalence between the knowledge of k and that of $\tilde{\phi}^+$. It is given by (in variables (u, ω) on $\mathscr{I}^+$)

$$\tilde{\phi}^+(u, \omega) = \int_{-\infty}^{u} k(s, \omega)\mathrm{d}s$$

and

$$\|\tilde{\phi}^+\|_{\dot{H}^1(\mathbb{R}\times S^2)} = \|k\|_{L^2(\mathbb{R}\times S^2)}.$$

If we consider the data $(\phi_0, -\phi_1)$, the associated solution is $\phi(-t, x)$ and its translation representer therefore contains the information of $\tilde{\phi}^-$. The map that to the past translation representer associates the future translation representer is a scattering operator for the wave equation on Minkowski space-time. We shall see a more precise description of it in Chapter 8.

Chapter 6

Analytic time-dependent scattering

The principle of analytic time-dependent scattering theory is to analyse the behaviour of solutions to a field equation in an asymptotic region (also referred to as a scattering channel) by finding a simplification of the equation in the region considered and proving that the field approaches solutions to the simplified system in this region. A complete scattering theory will also state that the field is completely described by its associated simplified solutions. This is described in details in Reed and Simon vol. 3 [49] and Derezinski and Gérard [11]. We give here a brief description of the two main classes of problems and present some techniques that allow to establish such scattering theories.

6.1 Two classes of problems

The simplest situation is when we try to compare two Hamiltonians that are self-adjoint on the same Hilbert space. Another very common case is when the two Hamiltonians are self-adjoint on different Hilbert spaces; in this case, identification operators are required. We describe here the abstract principles of the comparison, no scattering channel is involved. We shall see in examples how the comparison is concretely localised to a particular asymptotic region.

6.1.1 The one-space scattering

Let A and A_0 be two unbounded self-adjoint operators on the same separable Hilbert space $\mathscr{H}$. We consider the two evolution equations

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = iA\phi, \tag{6.1}$$

$$\frac{\mathrm{d}\psi}{\mathrm{d}t} = iA_0\psi, \tag{6.2}$$

where the unknowns ϕ and ψ are functions on $\mathbb{R}$ with values in $\mathscr{H}$. The propagator e^{itA} for equation (6.1) will be considered as the full dynamics and e^{itA_0} as the simplified dynamics. A concrete way of comparing the two dynamics for large times is to define the wave operators. If they exist, they are defined as the following strong limits of unitary operators

$$W^{\pm} := \mathrm{s} - \lim_{t \to \pm\infty} e^{-itA} e^{itA_0}, \tag{6.3}$$

$$\tilde{W}^{\pm} := \mathrm{s} - \lim_{t \to \pm\infty} e^{-itA_0} e^{itA}. \tag{6.4}$$

The operators $W^{\pm}$ are called the future and past direct wave operators and $\tilde{W}^{\pm}$ the future and past inverse wave operators.

Remark 6.1. *This type of construction is valid in a relatively simple case where the spectra of both A_0 and A are purely absolutely continuous. In more general situations, the definition of the wave operators requires to apply first a projector onto the absolutely continuous spectrum of the first operator.*

If in a certain sense the operators A and A_0 are asymptotically close, then one can establish that the direct and inverse wave operators exist, are isometries on $\mathscr{H}$ and satisfy

$$\tilde{W}^{\pm} = (W^{\pm})^{-1} = (W^{\pm})^* \,. \tag{6.5}$$

The existence of the direct and inverse wave operators is all there is to prove, their unitarity follows automatically from the definition and the unitarity of the propagators. Equation (6.5) also follows directly since for $t \in \mathbb{R}$

$$e^{-itA} e^{itA_0} e^{-itA_0} e^{itA} = \mathrm{Id}_{\mathscr{H}} \,.$$

For some initial data $\phi_0 \in \mathscr{H}$ for the full equation (6.1), the scattering data are given as the images $\phi^{\pm}$ of ϕ_0 under the future and past inverse wave operators $\tilde{W}^{\pm}$. The scattering operator summarises the whole evolution of the field by associating the future scattering data to the past scattering data. It is the isometry of $\mathscr{H}$ defined by

$$S = \tilde{W}^{+} W^{-}.$$

Why do we say that the existence of direct and inverse wave operators provides a scattering theory for the full dynamics? Let us consider some initial data $\phi \in \mathscr{H}$ for (6.1) and put

$$\phi^{+} := \tilde{W}^{+} \phi = \lim_{t \to +\infty} e^{-itA_0} e^{itA} \phi.$$

Then we have

$$0 = \lim_{t \to +\infty} \| \phi^{+} - e^{-itA_0} e^{itA} \phi \|_{\mathscr{H}} = \lim_{t \to +\infty} \| e^{itA_0} \phi^{+} - e^{itA} \phi \|_{\mathscr{H}},$$

which means that the solution to the full equation approaches in the distant future the solution to the simplified equation associated to the datum ϕ^{+}. The existence of direct and inverse wave operators thus shows that any solution to the full equation approaches in the distant future and past a solution to the simplified equation and is entirely characterised by each of its asymptotic solution. The future and past data for the simplified equation thus play the role of scattering data.

6.1.2 The two-space scattering

Let $\mathscr{H}$ and $\mathscr{H}_0$ be two separable Hilbert spaces and A and A_0 be self-adjoint operators respectively on $\mathscr{H}$ and $\mathscr{H}_0$. Let $\mathcal{J} \in \mathscr{L}(\mathscr{H}_0\,;\ \mathscr{H})$ be an identification operator. There is no systematic definition of such an object. The idea is that $\mathcal{J}$ should realise an "approximate embedding", by which me mean that at least for a large class of functions in the first space, the identification operator should transform them into elements of the second space with as little modification as possible. As we shall see on examples, this can still mean non trivial changes. The choice of identification operator will be guided by the properties of the full and simplified equations. We consider the same equations (6.1) and (6.2) as before and the associated propagators e^{itA_0} and e^{itA} are now strongly continuous one parameter groups of unitary operators on $\mathscr{H}_0$ and $\mathscr{H}$ respectively. We define direct and inverse wave operators as follows

$$W^{\pm} \ := \ s - \lim_{t\to\pm\infty} e^{-itA}\mathcal{J}e^{itA_0}, \tag{6.6}$$

$$\tilde{W}^{\pm} \ := \ s - \lim_{t\to\pm\infty} e^{-itA_0}\mathcal{J}^* e^{itA}, \tag{6.7}$$

where the adjoint of $\mathcal{J}$ is defined by

$$\langle \mathcal{J}\phi, \psi\rangle_{\mathscr{H}} = \langle \phi, \mathcal{J}^*\psi\rangle_{\mathscr{H}_0}\,,\ \forall \phi \in \mathscr{H}_0\,,\ \psi \in \mathscr{H}\,.$$

Provided we can establish the existence of both direct and inverse wave operators, the interpretation of the asymptotic behaviour of solutions to the full equation will then involve the operator $\mathcal{J}$, but we should always have that for a dense subclass of data, the solution approaches solutions to the simplified equation in the past and future and each of these completely characterises the solution. When the operators do not have purely absolutely continuous spectrum, we also need to apply a projector onto the absolutely continuous spectrum of the first Hamiltonian before applying the first propagator.

6.2 Cook's method and the importance of the Huygens principle

In most cases, the proof of existence of wave operators uses Cook's method. It is a simple argument based on the fact that if the derivative of a $\mathscr{C}^1$ function on $\mathbb{R}$ is integrable on $[a, +\infty[$ for a certain $a \in \mathbb{R}$, then the function admits a finite limit at $+\infty$, as is obvious from the fundamental theorem on integral calculus:

$$f(t) = f(a) + \int_{[a,t]} f'(s)\mathrm{d}s.$$

We quote it here from Theorem XI.4 in the third volume of Reed and Simon's book [49]. It concerns the one-space scattering but can easily be adapted to the two-space case. It involves a projector onto the absolutely continuous subspace of the first operator, which, in the cases we shall consider, reduces to the identity.

Theorem 6.1 (Cook's method). *Let A and B be two self-adjoint operators on a Hilbert space $\mathscr{H}$. Suppose that there is a set $\mathscr{D} \subset D(B) \cap P_{ac}(\mathscr{H})$ that is dense in $P_{ac}(\mathscr{H})$ such that, for each $\phi \in \mathscr{D}$, there exists $t_0 \geq 0$ satisfying*

1. *for* $|t| \geq t_0$, $e^{-itB}\phi \in D(A)$, *and,*
2. $\int_{t_0}^{+\infty} \left(\|(B-A)e^{-itB}\phi\| + \|(B-A)e^{itB}\phi\|\right) \mathrm{d}t < +\infty$.

Then the wave operators

$$\Omega^{\pm}(A,B) := s - \lim_{t \to \pm\infty} e^{-itA}e^{itB}P_{ac}(\mathscr{H})$$

exist.

Proof. It is very simple. We first need to establish the existence of the strong limit on the subset $\mathscr{D}$ and since it is part of the absolutely continuous subspace of B the projector $P_{ac}(\mathscr{H})$ acts trivially on $\mathscr{D}$ and we can remove it. Now for $\phi \in \mathscr{D}$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{-itA}e^{itB}\phi = e^{-itA}(B-A)e^{itB}\phi$$

which makes sense for $|t| > t_0$ thanks to assumption *1.* Taking into account the unitarity of e^{itA}, assumption *2* then states that

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{-itA}e^{itB}\phi \in L^1(]-\infty - t_0] \cap [t_0, +\infty[; \mathscr{H})$$

which entails the existence of

$$\lim_{t \to \pm\infty} e^{-itA}e^{itB}\phi.$$

It is then easy to prove that the strong limit exists globally. Let us consider $\phi \in \mathscr{H}$. Let $\varepsilon > 0$, there exists $\phi_\varepsilon \in \mathscr{D}$ such that $\|\phi_\varepsilon - P_{ac}(\mathscr{H})\phi\|_{\mathscr{H}} < \varepsilon/4$. Moreover, there exists $a > 0$ such that for all $t_1 > a$ and $t_2 > a$ we have

$$\left\|e^{-itA}e^{itB}\phi_\varepsilon\right\|_{\mathscr{H}} < \varepsilon/2$$

since we know that the limit

$$\lim_{t \to +\infty} e^{-itA}e^{itB}\phi_\varepsilon$$

exists. Therefore we have that for all $t_1 > a$ and $t_2 > a$,

$$\begin{aligned}
\left\|\left(e^{-it_2A}e^{it_2B} - e^{-it_1A}e^{it_1B}\right)P_{ac}(\mathscr{H})\phi\right\|_{\mathscr{H}} \quad &\leq \quad \left\|\left(e^{-it_2A}e^{it_2B} - e^{-it_1A}e^{it_1B}\right)\left(P_{ac}(\mathscr{H})\phi - \phi_\varepsilon\right)\right\|_{\mathscr{H}} \\
&\quad + \left\|\left(e^{-it_2A}e^{it_2B} - e^{-it_1A}e^{it_1B}\right)\phi_\varepsilon\right\|_{\mathscr{H}} \\
&\quad < 2\frac{\varepsilon}{4} + \frac{\varepsilon}{2} = \varepsilon.
\end{aligned}$$

This establishes the existence of the limit

$$\lim_{t \to +\infty} e^{-itA}e^{itB}P_{ac}(\mathscr{H})\phi$$

and concludes the proof. $\square$

How does one apply Cook's method to establish the existence of waves operators in the one-Hilbert space scattering for instance? It is obvious that the method relies on a direct comparison of the operators A and A_0 along the evolution. A simple situation is when $\mathscr{H}$ is a Hilbert space on $\mathbb{R}^3$. Let us take $\mathscr{H} = L^2(\mathbb{R}^3; \mathbb{C}^n)$ for example, and

$$A = A_0 + V$$

where V is simply a multiplication operator, i.e. a function on $\mathbb{R}^3$ with values in Hermitian $n \times n$ matrices. Such a perturbation is called a potential and we can assume that it is short-range, which means that there exists $\alpha > 0$ and $C > 0$ such that, for all $x \in \mathbb{R}^3$,

$$\|V(x)\| \leq \frac{C}{\|x\|^{1+\alpha}}.$$

In this case V is bounded on $\mathscr{H}$ and A is relatively A_0-bounded with $\alpha_{A_0}(A) = 0$. So by the Kato-Rellich Theorem A.8, A is self-adjoint on $\mathscr{H}$ with the same domain as A_0. If the dynamics e^{itA_0} satisfies the Huygens principle, we can consider the set $\mathscr{D} = \mathscr{C}_0^\infty(\mathbb{R}^3; \mathbb{C}^n)$ and for each $\phi_0 \in \mathscr{D}$ there exists $R > 0$ such that

$$(e^{itA_0}\phi_0)(x) = 0 \text{ for } \|x\| \leq |t| - R.$$

This implies that

$$\|Ve^{itA_0}\phi_0\|_{\mathscr{H}} \leq \|V\|_{L^\infty(\|x\| \geq |t|-R)} \|\phi_0\|_{\mathscr{H}} \leq \frac{C}{(|t| - R)^{1+\alpha}} \|\phi_0\|_{\mathscr{H}}$$

and this is integrable over $[-\infty, -R-1] \cap [R+1, +\infty[$.

We see here that the Huygens principle plays a fundamental role in transforming the spacelike decay of $H - H_0$ into a decay in time for

$$\left\| \frac{\mathrm{d}}{\mathrm{d}t} (e^{-itA} e^{itA_0} \phi) \right\|_{\mathscr{H}}.$$

It is not unusual that the simplified dynamics satisfies the Huygens principle, but this is extremely rare of the full dynamics. When the first dynamics does not satisfy a Huygens principle, we still need a similar mechanism to make Cook's method run. This is where time-dependent scattering theory gets more technical. We shall see some examples of weak forms of the Huygens principle in Sections 6.4 and 6.5.

6.3 Scalar fields on flat space-time

We start by expressing the wave equation on $\mathbb{M}$ in spherical coordinates

$$\partial_t^2 \phi - \partial_r^2 \phi - \frac{2}{r} \partial_r \phi - \frac{1}{r^2} \Delta_{S^2} \phi = 0\,. \tag{6.8}$$

In scattering theory, there is a crucial difference between short-range perturbations which fall-off like $r^{-\alpha}$ with $\alpha > 1$ and long-range perturbations which fall off like $r^{-\alpha}$ with $\alpha \leq 1$. This is of course a question of integrability of these quantities at infinity, but the space dimension is irrelevant, it is always a matter of integrability in 1-dimension, as we have seen in the description of Cook's method above. Short-range perturbations can be treated naturally provided we have some weak version of the Huygens principle, but long-range perturbations require an in depth modification of the construction which reveals the profound change they induce in the asymptotic behaviour. So it is crucial to understand, when long-range terms are present, whether they are genuine or artificial. Here the term $\frac{2}{r}\partial_r\phi$ is artificially long-range since it can be eliminated by a simple rescaling of the unknown function. Putting

$$\tilde{\phi} = r\phi \,, \tag{6.9}$$

we get that $\phi \in \mathcal{D}'(\mathbb{R}^4)$ satisfies the wave equation on $\mathbb{M}$ (equivalently (6.8)) if and only if $\tilde{\phi}$ is a solution of the simplified equation

$$\partial_t^2\tilde{\phi} - \partial_r^2\tilde{\phi} - \frac{1}{r^2}\Delta_{S^2}\tilde{\phi} = 0 \tag{6.10}$$

We can explain this a little more systematically. Recall that the operator A (here expressed in spherical coordinates)

$$A = -i\begin{pmatrix} 0 & 1 \\ \partial_r^2 + \frac{2}{r}\partial_r + \frac{1}{r^2}\Delta_{S^2} & 0 \end{pmatrix}$$

is self-adjoint on $\mathscr{H} = \dot{H}^1(\mathbb{R}^3) \times L^2(\mathbb{R}^3)$ and putting

$$h = -(\partial_r^2 + \frac{2}{r}\partial_r + \frac{1}{r^2}\Delta_{S^2})\,, \text{ i.e. } h = -\Delta_{\mathbb{R}^3}\,,$$

the $\mathscr{H}$ inner product is given by[1]

$$\langle \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}, \begin{pmatrix} \zeta_1 \\ \zeta_2 \end{pmatrix} \rangle_{\mathscr{H}} = \langle h\phi_1, \zeta_1 \rangle_{L^2(\mathbb{R}^3\,,\; r^2 \mathrm{d}r\mathrm{d}\omega)} + \langle \phi_2, \zeta_2 \rangle_{L^2(\mathbb{R}^3\,,\; r^2 \mathrm{d}r\mathrm{d}\omega)}\,.$$

We now consider the unitary operator

$$\mathcal{R} \,:\; L^2(\mathbb{R}^3\,,\; r^2\mathrm{d}r\mathrm{d}\omega) \to L^2(\mathbb{R}^3\,,\; \mathrm{d}r\mathrm{d}\omega)\,,\; \mathcal{R}\phi = r\phi\,.$$

Then by conjugation by $\mathcal{R}$, we have

$$\mathcal{R}h\mathcal{R}^* = -\partial_r^2 - \frac{1}{r^2}\Delta_{S^2}\,,$$

the operator $\mathcal{R}^*$ being simply the multiplication by $1/r$ from $L^2(\mathbb{R}^3\,,\; \mathrm{d}r\mathrm{d}\omega)$ to $L^2(\mathbb{R}^3\,,\; r^2\mathrm{d}r\mathrm{d}\omega)$, and

$$\mathcal{R}A\mathcal{R}^* = -i\begin{pmatrix} 0 & 1 \\ \partial_r^2 + \frac{1}{r^2}\Delta_{S^2} & 0 \end{pmatrix} =: B\,,$$

[1] We have expressed the inner product under a form that makes sense only on a dense subspace of $\mathscr{H}$, but has the advantage of having the operator h appearing explicitely.

where $\mathcal{R}$ and $\mathcal{R}^*$ are understood as acting on each component. It follows that B is self-adjoint on $\tilde{\mathscr{H}} = \mathcal{R}\mathscr{H}$, equipped with the inner product

$$\begin{aligned}\langle \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix}, \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} \rangle_{\tilde{\mathscr{H}}} &= \langle (-\partial_r^2 - \frac{1}{r^2}\Delta_{S^2})\psi_1, \xi_1 \rangle_{L^2(\mathbb{R}^3\,,\ \mathrm{d}r\mathrm{d}\omega)} + \langle \psi_2, \xi_2 \rangle_{L^2(\mathbb{R}^3\,,\ \mathrm{d}r\mathrm{d}\omega)}, \\ \text{i.e. } \| \begin{pmatrix} \psi_1 \\ \psi_2 \end{pmatrix} \|^2_{\tilde{\mathscr{H}}} &= \int_{\mathbb{R}_r^+ \times S^2} (|\partial_r \psi_1|^2 + \frac{1}{r^2} |\nabla_{S^2} \psi_1|^2 + |\psi_2|^2)\, \mathrm{d}r\mathrm{d}\omega. \qquad (6.11)\end{aligned}$$

Moreover, $\tilde{\mathscr{H}}$ is the completion of $\left(\mathcal{R}\mathscr{C}_0^\infty(\mathbb{R}^3)\right)^2$ in the norm (6.11). Let us now give the steps of the scattering construction.

- **Comparison dynamics.** The asymptotic region is $r \to +\infty$. In this region, the equation (6.10) simplifies to
$$\partial_t^2 \upsilon - \partial_r^2 \upsilon = 0\,. \qquad (6.12)$$
The Hamiltonian forms of equations (6.10) and (6.12) are given by
$$\begin{gathered}\partial_t U = iBU\,,\ \partial_t V = iB_0 V\,, \\ U := \begin{pmatrix} \psi \\ \partial_t \psi \end{pmatrix},\ B = -i \begin{pmatrix} 0 & 1 \\ \partial_r^2 + \frac{1}{r^2}\Delta_{S^2} & 0 \end{pmatrix}, \\ V := \begin{pmatrix} \upsilon \\ \partial_t \upsilon \end{pmatrix},\ B_0 = -i \begin{pmatrix} 0 & 1 \\ \partial_r^2 & 0 \end{pmatrix}.\end{gathered}$$
The operator B is self-adjoint on $\tilde{\mathscr{H}}$ and B_0 is self-adjoint on
$$\tilde{\mathscr{H}_0} = \dot{H}^1(\mathbb{R}\,;\ L^2(S^2)) \times L^2(\mathbb{R} \times S^2).$$
Note that the range of the variable r in $\tilde{\mathscr{H}}$ is $\mathbb{R}^+$ whereas it is the whole real axis for $\tilde{\mathscr{H}_0}$. The propagators e^{itB} and e^{itB_0} are strongly continuous 1-parameter groups of unitary operators on $\tilde{\mathscr{H}}$ and $\tilde{\mathscr{H}_0}$ respectively.

- **Free outgoing and incoming data.** Since (6.12) is the wave equation on $\mathbb{R}_t \times \mathbb{R}_r$, we know that every solution is a sum of an incoming and an outgoing progressive wave in the (t, r) variables. Each of these two types of finite energy solutions is characterized by a special subspace of initial data:
$$\tilde{\mathscr{H}_0}^{\pm} := \{ V = (\upsilon_1, \upsilon_2) \in \tilde{\mathscr{H}_0}\,,\ \upsilon_2 = \mp \partial_r \upsilon_1 \}\,.$$
This is obvious from the fact that
$$\frac{\partial}{\partial t}(f(r-t)) = -\frac{\partial}{\partial r}(f(r-t))$$
and $f(r-t)$ is simply $f(r)$ translated by t. We have $\tilde{\mathscr{H}_0} = \tilde{\mathscr{H}_0}^+ \oplus \tilde{\mathscr{H}_0}^-$ and for any $V \in \tilde{\mathscr{H}_0}^{\pm}$, $(e^{itB_0}V)(r) = V(r \pm t)$. Moreover, $\tilde{\mathscr{H}_0}^+$ and $\tilde{\mathscr{H}_0}^-$ are closed subspaces of $\tilde{\mathscr{H}_0}$.

- **Inverse wave operators.** The situation we are studying here is very special in that the full dynamics satisfies a strong Huyghens principle. This means that the scattering theory is essentially trivial, apart from purely formal difficulties related to function spaces. Since the function spaces on which the two dynamics act are different, we need an identifying operator between the two spaces. Consider a cut-off function

$$\chi \in \mathscr{C}^\infty(\mathbb{R}^+)\,,\ \chi \equiv 0 \text{ on } [0,1/2]\,,\ \chi \equiv 1 \text{ on } [1,+\infty[$$

and define the bounded operator

$$\mathcal{J} \,:\, \tilde{\mathscr{H}} \to \tilde{\mathscr{H}}_0\,,\ \mathcal{J}U = \left\{ \begin{array}{c} \chi U \text{ on } \mathbb{R}^+ \times S^2\,, \\ 0 \text{ on } \mathbb{R}^- \times S^2\,. \end{array} \right. \tag{6.13}$$

Theorem 6.2. *The inverse wave operators*

$$\tilde{W}^\pm = s - \lim_{t\to\pm\infty} e^{-itB_0}\mathcal{J}e^{itB} \tag{6.14}$$

are well defined for smooth compactly supported data and extend as partial isometries from $\tilde{\mathscr{H}}$ *to* $\tilde{\mathscr{H}}_0$.

Proof. We prove the theorem for $\tilde{W}^+$, the proof is similar for $\tilde{W}^-$. We consider smooth compactly supported data $V \in (\mathcal{R}\mathscr{C}_0^\infty(\mathbb{R}^3))^2$. In this case

$$e^{-itB_0}\mathcal{J}e^{itB}V \in \mathscr{C}^\infty(\mathbb{R}^+;\tilde{\mathscr{H}}_0)\,.$$

By the strong Huyghens principle, taking $R>0$ such that $\mathrm{supp}(V) \subset B(0,R)$, for $t>R$ we have

$$e^{itB}V \equiv 0 \text{ for } 0 \le r \le t-R \text{ and for } r \ge R+t.$$

Therefore in particular $\chi e^{itB}V = e^{itB}V$ for $t > R+1$. In order to prove the existence of the limit

$$\lim_{t\to+\infty} e^{-itB_0}\mathcal{J}e^{itB}V, \tag{6.15}$$

we use Cook's method and prove that

$$\frac{\mathrm{d}}{\mathrm{d}t}e^{-itB_0}\mathcal{J}e^{itB}V \in L^1(\mathbb{R}^+\,;\ \tilde{\mathscr{H}}_0)\,. \tag{6.16}$$

We have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}e^{-itB_0}\mathcal{J}e^{itB}V &= -ie^{-itB_0}(B_0\mathcal{J} - \mathcal{J}B)e^{itB}V \\ &= -ie^{-itB_0}\mathcal{J}\left(\begin{array}{cc} 0 & 0 \\ \frac{1}{r^2}\Delta_{S^2} & 0 \end{array} \right) e^{itB}V \text{ for } t > R+1\,. \end{aligned}$$

We calculate the norm of this quantity in $\tilde{\mathscr{H}}_0$ for $t > R+1$, denoting $\tilde{\phi}(t,r,\omega)$ the first component of $e^{itB}V$:

$$\begin{aligned}
\| -ie^{-itB_0}\mathcal{J}\begin{pmatrix} 0 & 0 \\ \frac{1}{r^2}\Delta_{S^2} & 0 \end{pmatrix} e^{itB}V\|^2_{\tilde{\mathscr{H}}_0} &= \|\mathcal{J}\begin{pmatrix} 0 & 0 \\ \frac{1}{r^2}\Delta_{S^2} & 0 \end{pmatrix} e^{itB}V\|^2_{\tilde{\mathscr{H}}_0} \\
&= \int_{\mathbb{R}^+\times S^2} |\frac{1}{r^2}\Delta_{S^2}\tilde{\phi}(t,r,\omega)|^2 \mathrm{d}r\mathrm{d}\omega \\
&= \int_{[t-R,t+R]\times S^2} |\frac{1}{r^2}\Delta_{S^2}\tilde{\phi}(t,r,\omega)|^2 \,\mathrm{d}r\mathrm{d}\omega \\
&\leq \frac{1}{(t-R)^4}\|\Delta_{S^2}\tilde{\phi}(t)\|^2_{L^2([t-R,t+R]\times S^2\,;\,\mathrm{d}r\mathrm{d}\omega)}\,.
\end{aligned}$$

The last expression can be estimated by a norm in $\tilde{\mathscr{H}}$ of an angular derivative of $e^{itB}V$ which in turn can be estimated by the same quantity at $t=0$. To show this, we use first a Poincaré estimate:

Lemma 6.1. *Let $f \in H^1(\mathbb{R})$ supported in $[R_1, R_2]$, then*

$$\|f\|_{L^2(\mathbb{R})} \leq (R_2 - R_1)\|f'\|_{L^2(\mathbb{R})}\,.$$

Proof. We prove the result in the case where f is smooth, using the fundamental theorem of integral calculus

$$\begin{aligned}
\|f\|^2_{L^2(\mathbb{R})} &= \int_{R_1}^{R_2} |f(t)|^2 \mathrm{d}t \\
&= \int_{R_1}^{R_2} \left| \int_{R_1}^{t} f'(x)\mathrm{d}x \right|^2 \mathrm{d}t \\
&\leq \int_{R_1}^{R_2} (R_2 - R_1) \int_{R_1}^{R_2} \left|f'(x)\right|^2 \mathrm{d}x\mathrm{d}t \leq (R_2-R_1)^2\|f'\|^2_{L^2(\mathbb{R})}\,.
\end{aligned}$$

This gives the result in the smooth case and the inequality extends by density to functions in H^1 that are supported in $[R_1, R_2]$. □

It follows that

$$\left\| -ie^{-itB_0}\mathcal{J}\begin{pmatrix} 0 & 0 \\ \frac{1}{r^2}\Delta_{S^2} & 0 \end{pmatrix} e^{itB}V \right\|^2_{\tilde{\mathscr{H}}_0} \leq \frac{4R^2}{(t-R)^4}\|\partial_r\Delta_{S^2}\psi(t)\|^2_{L^2([t-R,t+R]\times S^2\,;\,\mathrm{d}r\mathrm{d}\omega)}\,.$$

Since Δ_{S^2} commutes with the equation (6.10) and therefore also with e^{itB}, $\Delta_{S^2}\tilde{\phi}$ is a solution of (6.10) with data $\Delta_{S^2}V$. Hence

$$\begin{aligned}
\left\| -ie^{-itB_0}\mathcal{J}\begin{pmatrix} 0 & 0 \\ \frac{1}{r^2}\Delta_{S^2} & 0 \end{pmatrix} e^{itB}V \right\|^2_{\tilde{\mathscr{H}}_0} &\leq \frac{4R^2}{(t-R)^4}\|e^{itB}\Delta_{S^2}V\|^2_{\tilde{\mathscr{H}}} \\
&= \frac{4R^2}{(t-R)^4}\|\Delta_{S^2}V\|^2_{\tilde{\mathscr{H}}}.
\end{aligned}$$

Hence,

$$\left\| \frac{\mathrm{d}}{\mathrm{d}t} e^{-itB_0} \mathcal{J} e^{itB} V \right\|_{\tilde{\mathscr{H}}_0} \leq \frac{2R}{(t-R)^2} \|\Delta_{S^2} V\|_{\tilde{\mathscr{H}}} \in L^1([R+1,+\infty[).$$

This implies that (6.16) is true and the limit (6.15) exists for all smooth and compactly supported V. Since the operator $\mathcal{J}$ is bounded, this suffices to define the operator $\tilde{W}^+$ on $\tilde{\mathscr{H}}$. Indeed for $V \in (\mathcal{R}\mathscr{C}_0^\infty(\mathbb{R}^3))^2$, we have for all $t \in \mathbb{R}$

$$\left\| e^{-itB_0} \mathcal{J} e^{itB} V \right\|_{\tilde{\mathscr{H}}_0} \leq \|\mathcal{J}\|_{\mathscr{L}(\tilde{\mathscr{H}};\tilde{\mathscr{H}}_0)} \|V\|_{\tilde{\mathscr{H}}}$$

and therefore $\tilde{W}^+$ can be extended to the whole of $\tilde{\mathscr{H}}$ as a bounded operator with

$$\|\tilde{W}^+\|_{\mathscr{L}(\tilde{\mathscr{H}};\tilde{\mathscr{H}}_0)} \leq \|\mathcal{J}\|_{\mathscr{L}(\tilde{\mathscr{H}};\tilde{\mathscr{H}}_0)} .$$

But it is not clear that it is defined as the strong limit (6.14) on the whole energy space $\tilde{\mathscr{H}}$. Let us prove that the limit (6.15) exists for all $V \in \tilde{\mathscr{H}}$. We follow the same reasoning as in the proof of Cook's method (Theorem 6.1). Let $V \in \tilde{\mathscr{H}}$, consider $\{V_n\}_{n\in\mathbb{N}}$ a sequence in $(\mathscr{C}_0^\infty(\mathbb{R}^3))^2$ converging towards V in $\tilde{\mathscr{H}}$. For $\varepsilon > 0$, let $n_0 \in \mathbb{N}$ be such that for all $n \geq n_0$, $\|V - V_n\|_{\tilde{\mathscr{H}}} < \varepsilon$. Then for all $t \in \mathbb{R}$ we have

$$\|e^{-itB_0} \mathcal{J} e^{itB} (V - V_{n_0})\|_{\tilde{\mathscr{H}}_0} < \varepsilon \|\mathcal{J}\|_{\mathscr{L}(\tilde{\mathscr{H}};\tilde{\mathscr{H}}_0)} . \tag{6.17}$$

Now we use the fact that the limit

$$\lim_{t\to\pm\infty} e^{-itB_0} \mathcal{J} e^{itB} V_{n_0}$$

exists: we take $a > 0$ large enough so that for all $t_1 > t_0 > a$,

$$\|e^{-it_1B_0} \mathcal{J} e^{it_1B} V_{n_0} - e^{-it_0B_0} \mathcal{J} e^{it_0B} V_{n_0}\|_{\tilde{\mathscr{H}}_0} < \varepsilon.$$

This gives that for all $t_1 > t_0 > a$,

$$\begin{aligned} \|e^{-it_1B_0} \mathcal{J} e^{it_1B} V - e^{-it_0B_0} \mathcal{J} e^{it_0B} V\|_{\tilde{\mathscr{H}}_0} &\leq \|e^{-it_1B_0} \mathcal{J} e^{it_1B} (V - V_{n_0})\|_{\tilde{\mathscr{H}}_0} \\ &\quad + \|e^{-it_0B_0} \mathcal{J} e^{it_0B} (V - V_{n_0})\|_{\tilde{\mathscr{H}}_0} \\ &\quad + \|e^{-it_1B_0} \mathcal{J} e^{it_1B} V_{n_0} - e^{-it_0B_0} \mathcal{J} e^{it_0B} V_{n_0}\|_{\tilde{\mathscr{H}}_0} \\ &< \left(2 + \|\mathcal{J}\|_{\mathscr{L}(\tilde{\mathscr{H}};\tilde{\mathscr{H}}_0)}\right) \varepsilon. \end{aligned}$$

This proves the existence of the limit

$$\lim_{t\to\pm\infty} e^{-itB_0} \mathcal{J} e^{itB} V.$$

The operator $\tilde{W}^+$ is thus defined by the strong limit (6.14) on the whole of $\tilde{\mathscr{H}}$.

Let us now show that $\tilde{W}^{\pm}$ preserve the norm. It suffices to prove this for $V \in (\mathcal{R}\mathscr{C}_0^\infty(\mathbb{R}^3))^2$ as it will extend by density to all $V \in \tilde{\mathscr{H}}$. Let $V \in (\mathcal{R}\mathscr{C}_0^\infty(\mathbb{R}^3))^2$ and $R > 0$ such that $\mathrm{supp} V \subset B(0,R)$. Denoting by ψ the first component of $e^{itB}V$, we have

$$\begin{aligned}
\|e^{-itB_0}\mathcal{J}e^{itB}V\|^2_{\tilde{\mathscr{H}}_0} &= \|\mathcal{J}e^{itB}V\|^2_{\tilde{\mathscr{H}}_0} \\
&= \int_{\mathbb{R}^+\times S^2}(|\partial_r\psi(t,r,\omega)|^2 + |\partial_t\psi(t,r,\omega)|^2)\mathrm{d}r\mathrm{d}\omega \text{ for } t > R+1\,, \\
&= \|e^{itB}V\|^2_{\tilde{\mathscr{H}}} - \int_{\mathbb{R}^+\times S^2}\frac{1}{r^2}|\nabla_{S^2}\psi(t,r,\omega)|^2\,\mathrm{d}r\mathrm{d}\omega \\
&= \|e^{itB}V\|^2_{\tilde{\mathscr{H}}} - \int_{[t-R,t+R]\times S^2}\frac{1}{r^2}|\nabla_{S^2}\psi(t,r,\omega)|^2\,\mathrm{d}r\mathrm{d}\omega\,.
\end{aligned}$$

We show that the last integral tends to zero using again a Poincaré inequality:

$$\begin{aligned}
&\int_{[t-R,t+R]\times S^2}\frac{1}{r^2}|\nabla_{S^2}\psi(t,r,\omega)|^2\,\mathrm{d}r\mathrm{d}\omega \\
&\qquad = \int_{[t-R,t+R]\times S^2}\frac{1}{r^2}\left(-\Delta_{S^2}\psi(t,r,\omega)\right)\overline{\psi(t,r,\omega)}\,\mathrm{d}r\mathrm{d}\omega \\
&\qquad \le \frac{1}{2(t-R)^2}\int_{[t-R,t+R]\times S^2}\frac{1}{r^2}|\Delta_{S^2}\psi(t,r,\omega)|^2 + |\psi(t,r,\omega)|^2\,\mathrm{d}r\mathrm{d}\omega \\
&\qquad \le \frac{4R^2}{2(t-R)^2}\int_{[t-R,t+R]\times S^2}\frac{1}{r^2}|\partial_r\Delta_{S^2}\psi(t,r,\omega)|^2 + |\partial_r\psi(t,r,\omega)|^2\,\mathrm{d}r\mathrm{d}\omega \\
&\qquad \le \frac{4R^2}{2(t-R)^2}(\|e^{itB}\Delta_{S^2}V\|^2_{\tilde{\mathscr{H}}} + \|e^{itB}V\|^2_{\tilde{\mathscr{H}}}) \\
&\qquad \le \frac{4R^2}{2(t-R)^2}(\|\Delta_{S^2}V\|^2_{\tilde{\mathscr{H}}} + \|V\|^2_{\tilde{\mathscr{H}}}) \to 0 \text{ as } t\to+\infty\,.
\end{aligned}$$

It follows that

$$\|\tilde{W}^+V\|_{\tilde{\mathscr{H}}_0} = \|V\|_{\tilde{\mathscr{H}}}.$$

The proof is similar for $\tilde{W}^-$. □

This establishes an important property of the solutions of (6.10).

Corollary 6.1. *For any solution $\tilde{\phi}$ of (6.10), there exist $v^{\pm}$ solutions of (6.12) such that*

$$\lim_{t\to\pm\infty}\left\|\mathcal{J}\begin{pmatrix}\tilde{\phi}(t)\\ \partial_t\tilde{\phi}(t)\end{pmatrix} - \begin{pmatrix}v^{\pm}(t)\\ \partial_t v^{\pm}(t)\end{pmatrix}\right\|_{\tilde{\mathscr{H}}_0} = 0\,.$$

Moreover, $v^{\pm}$ are respectively an outgoing and an incoming solution of (6.12), i.e.

$$\begin{pmatrix}v^{\pm}|_{t=0}\\ \partial_t v^{\pm}|_{t=0}\end{pmatrix} \in \tilde{\mathscr{H}}_0^{\pm}\,.$$

In other words, $\mathcal{F}_0^{\pm} := \mathrm{Ran}(\tilde{W}^{\pm}) \subset \tilde{\mathscr{H}}_0^{\pm}$; $\mathcal{F}_0^+$ (resp. $\mathcal{F}_0^-$) is a closed subspace of $\tilde{\mathscr{H}}_0^+$ (resp. $\tilde{\mathscr{H}}_0^-$) and $\tilde{W}^+$ (resp. $\tilde{W}^-$) is an isomorphism between $\tilde{\mathscr{H}}$ and $\mathcal{F}_0^+$ (resp. $\mathcal{F}_0^-$).

Proof. The first part is a direct consequence of the theorem, let us prove the second part. Let $V \in \tilde{\mathscr{H}}$, put $Y^\pm := \tilde{W}^\pm V$, we have

$$Y^\pm = \lim_{t\to\pm\infty} e^{-itB_0}\mathcal{J}e^{itB}V\,,$$

whence

$$\lim_{t\to\pm\infty} \|e^{-itB_0}\mathcal{J}e^{itB}V - Y^\pm\|_{\tilde{\mathscr{H}_0}} = 0$$

and since e^{-itB_0} is a unitary operator on $\tilde{\mathscr{H}_0}$,

$$\|e^{-itB_0}\mathcal{J}e^{itB}V - Y^\pm\|_{\tilde{\mathscr{H}_0}} = \|\mathcal{J}e^{itB}V - e^{itB_0}Y^\pm\|_{\tilde{\mathscr{H}_0}}\,.$$

Now the fact that $\tilde{W}^\pm$ are partial isometries implies that they have closed range and that they are isomorphisms from $\tilde{\mathscr{H}}$ onto their range. For $V \in (\mathcal{R}\mathscr{C}_0^\infty(\mathbb{R}^3))^2$ with $\mathrm{supp}(V) \subset \bar{B}(0,R)$, $e^{itB}V$ vanishes for $r < t - R$, whence if $\mathcal{J}e^{itB}V$ approaches a solution $e^{itB_0}Y^+$ of (6.12) as $t \to +\infty$, we must have $Y^+ \in \tilde{\mathscr{H}_0}^+$. By density and using the fact that $\tilde{\mathscr{H}_0}^+$ is a closed subspace of $\tilde{\mathscr{H}_0}$, this must then be true of $\tilde{W}^+V$ for any $V \in \tilde{\mathscr{H}}$. We can argue similarly in the past. □

- **Direct wave operators and asymptotic completeness.** What remains to be done is to prove the surjectivity of the inverse wave operators, i.e. $\mathcal{F}_0^\pm = \tilde{\mathscr{H}_0}^\pm$. This is referred to as asymptotic completeness. This can be done by constructing the direct wave operators on $\tilde{\mathscr{H}_0}^\pm$
$$W^\pm := s - \lim_{t\to\pm\infty} e^{-itB}\mathcal{J}^* e^{itB_0}, \tag{6.18}$$
where the adjoint of $\mathcal{J}$ is the operator
$$\mathcal{J}^* : \tilde{\mathscr{H}_0} \to \tilde{\mathscr{H}}\,,\ \mathcal{J}^*V = \chi(r)V|_{r\geq 0}. \tag{6.19}$$
We have the following result.

Theorem 6.3. *The direct wave operators (6.18) are well defined on $\tilde{\mathscr{H}_0}^\pm \cap \mathscr{C}_0^\infty(\mathbb{R}\times S^2)$ and satisfy*

$$\tilde{W}^\pm W^\pm V = V \text{ for all } V \in \tilde{\mathscr{H}_0}^\pm \cap \mathscr{C}_0^\infty(\mathbb{R}\times S^2). \tag{6.20}$$

This implies that $\tilde{W}^\pm$ are isometries from $\tilde{\mathscr{H}}$ onto $\tilde{\mathscr{H}_0}^\pm$, that $W^\pm$ extend as isometries from $\tilde{\mathscr{H}_0}^\pm$ onto $\tilde{\mathscr{H}}$ and that

$$W^\pm = (\tilde{W}^\pm)^{-1} = (\tilde{W}^\pm)^*.$$

Moreover, we have the intertwining relations

$$W^\pm B_0 = BW^\pm\,,\ B_0\tilde{W}^\pm = \tilde{W}^\pm B\,.$$

The wave operators exchange the domains of B_0 and B.

Proof. We give the details for W^+ and $\tilde{W}^+$, the arguments are completely analogous for W^- and $\tilde{W}^-$.

The fact that W^+ is well-defined on $\tilde{\mathscr{H}}_0^+ \cap \mathscr{C}_0^\infty(\mathbb{R}\times S^2)$ is easily established by Cook's method. Given $V \in \tilde{\mathscr{H}}_0^+ \cap \mathscr{C}_0^\infty(\mathbb{R}\times S^2)$ with support in $[-R,R]\times S^2$,

$$\tilde{W}^+ W^+ V = \lim_{t\to+\infty} e^{-itB_0}\mathcal{J}e^{itB}e^{-itB}\mathcal{J}^* e^{itB_0}V = \lim_{t\to+\infty} e^{-itB_0}\mathcal{J}\mathcal{J}^* e^{itB_0}V.$$

Since $e^{itB_0}V \equiv 0$ for $r \leq t-R$, it follows that for $t > R+1$

$$\mathcal{J}\mathcal{J}^* e^{itB_0}V = e^{itB_0}V,$$

which implies (6.20).

Next we need to prove the intertwining relations. Given $V \in \left(\mathcal{R}\mathscr{C}_0^\infty(\mathbb{R}^3)\right)^2$, we have for $t>0$

$$B_0 e^{-itB_0}\mathcal{J}e^{itB} - e^{-itB_0}\mathcal{J}e^{itB}BV = -\frac{1}{it}\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-itB_0}\mathcal{J}e^{itB}BV\right).$$

This quantity has a limit as $t\to+\infty$ given by

$$(B_0\tilde{W}^+ - \tilde{W}^+B)V.$$

From (6.16), we infer that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-itB_0}\mathcal{J}e^{itB}BV\right) \in L^1(\mathbb{R}^+;\ \tilde{\mathscr{H}}_0)$$

and the limit must therefore be zero lest we contradict integrability, i.e.

$$(B_0\tilde{W}^+ - \tilde{W}^+B)V = 0.$$

The intertwining relation is therefore proved for smooth compactly supported functions. By density, it extends to $V \in D(B)$ and this shows in particular that $\tilde{W}^+$ sends the domain of B to the domain of B_0. The other intertwining relations are established in the same manner. □

Remark 6.2. *It is important to observe that although $W^\pm$ extend as isometries from $\tilde{\mathscr{H}}_0^\pm$ onto $\tilde{\mathscr{H}}$, they are no longer defined by the strong limit (6.18). This is simply because for $V \in \tilde{\mathscr{H}}_0^\pm$, $\mathcal{J}^* e^{itB_0}V \notin \tilde{\mathscr{H}}$ in general and therefore the quantities in the limit (6.18) are not even well-defined.*

- **Scattering operator.** It is the operator that to the past scattering data associates the future scattering data and thus summarises the full evolution of the field:

$$S = \tilde{W}^+ W^-.$$

It is an isometry from $\tilde{\mathscr{H}}_0^-$ to $\tilde{\mathscr{H}}_0^+$.

6.4 Trace-class perturbation methods

In most cases, the full equation does not satisfy the Huygens principle. This is what happens for instance as soon as we perturb the wave equation on flat space-time with a potential or for the wave equation on non conformally flat space-times. Historically, the first approach to this type of problem was to compare to simplified equations such that the difference between the full and simplified Hamiltonians was trace-class or at least involved trace-class operators. The idea is to understand the integrability that is required for Cook's method, at the level of the spectrum. This is the technique that Dimock [12] and Dimock and Kay [13, 14] used to construct a scattering theory for scalar fields on the Schwarzschild metric. Most of the material in the first part of this section is present in Reed and Simon vol. 3 [49].

We start by defining the notion of trace class operators. We consider a separable Hilbert space $\mathscr{H}$. If $A \in \mathscr{L}(\mathscr{H})$ is a positive operator, then we can define the trace of A by choosing an orthonormal basis $\{e_n\}_{n\in\mathbb{N}}$ and putting

$$\mathrm{tr}(A) := \sum_{n\in\mathbb{N}} \langle Ae_n, e_n\rangle. \tag{6.21}$$

The result, finite or infinite, is independent of the choice of orthonormal basis. Moreover, the trace satisfies properties that generalise those of the trace in finite dimension:

- $\mathrm{tr}(A+B) = \mathrm{tr}(A) + \mathrm{tr}(B)$,
- $\mathrm{tr}(\lambda A) = \lambda \mathrm{tr}(A)$ for any $\lambda \geq 0$,
- $\mathrm{tr}(UAU^{-1}) = \mathrm{tr}(A)$ for any unitary operator U (which is the same as saying that the trace is independent of the choice of orthonormal basis),
- if $0 \leq A \leq B$ then $\mathrm{tr}(A) \leq \mathrm{tr}(B)$.

If we consider merely a bounded operator $A \in \mathscr{L}(\mathscr{H})$ without assuming that it is positive, we can always define its modulus as

$$|A| := \sqrt{A^*A}. \tag{6.22}$$

This makes sense observing that A^*A is a positive self-adjoint operator and then using the functional calculus (Theorem A.6) to define its square root. We can now define the trace class.

Definition 6.1 (Trace class). *The trace class $\mathscr{I}_1$ is the set of $A \in \mathscr{L}(\mathscr{H})$ such that*

$$\mathrm{tr}(|A|) < +\infty.$$

Trace class operators are compact, which means that for any bounded sequence $(u_n)_n$ in $\mathscr{H}$, we can extract a subsequence $(u_{n_k})_k$ such that Au_{n_k} converges in $\mathscr{H}$. Moreover, the trace class is an ideal, in other words, if $A \in \mathscr{I}_1$ and $B \in \mathscr{L}(\mathscr{H})$, then $AB \in \mathscr{I}_1$.

Another trace ideal of compact operators is the Hilbert-Schmidt family.

Definition 6.2 (Hilbert-Schmidt operators). *A bounded operator A is said to be Hilbert-Schmidt if $\mathrm{tr}(A^*A) < +\infty$. The set of Hilbert-Schmidt operators is denoted $\mathscr{I}_2$.*

Of course the product of two Hilbert-Schmidt operators is trace-class and we have a useful class of operators that belongs to the Hilbert-Schmidt family.

Theorem 6.4 (Theorem XI-20, Reed and Simon vol. 3 [49]). *Let $f, g \in L^2(\mathbb{R}^n)$, then the operator $f(x)g(D)$ belongs to $\mathscr{I}_2$ (recall that $D = -i\nabla$) in $\mathscr{H} = L^2(\mathbb{R}^n)$.*

The basic result for trace class perturbation scattering is the following.

Theorem 6.5 (Kato-Rosenblum Theorem). *Let A and B be self-adjoint operators on $\mathscr{H}$. If we suppose that $B - A$ is trace class, then the direct and inverse wave operators* (6.3) *and* (6.4) *exist.*

Unfortunately, this result is too weak to be directly useful because the operator $B - A$ may not be bounded as we have seen with the example of the wave equation on flat space-time and even if it is (in the case of a potential for instance), it is not compact in general. The following theorem turns out to be more useful.

Theorem 6.6 (Kuroda-Birman). *Let A and B be self-adjoint operators on $\mathscr{H}$ such that*

$$(A + i)^{-1} - (B + i)^{-1} \in \mathscr{I}_1,$$

then the direct and inverse wave operators (6.3) *and* (6.4) *exist.*

Another important result is the striking invariance principle. To express it, we need the notion of admissible function.

Definition 6.3 (Admissible function). *Let T be an open subset of $\mathbb{R}$, a function $f : T \mapsto \mathbb{R}$ is called admissible if $f'' \in L^1_{\mathrm{loc}}(T)$ and there is a subdivision of T such that in the interior of any of its intervals, f' is either strictly positive or strictly negative.*

Theorem 6.7 (Invariance principle). *Let f be an admissible function on an open set T. Let A and b be self-adjoint operators on $\mathscr{H}$ with $\sigma(A), \sigma(B) \subset \bar{T}$. Assume moreover that at each boundary point of T, either f has a finite limit, or both A and B do not have point spectrum at that point. Suppose also that $A - B$ is trace-class. Then $\Omega^\pm(f(A), f(B))$ exist and*

$$\Omega^\pm(f(A), f(B)) = \Omega^\pm(A, B)\chi_{T_1}(B) + \Omega^\mp(A, B)\chi_{T_2}(B)$$

where T_1 and T_2 are the unions of sub-intervals of T in which $f' > 0$ (resp. $f' < 0$) and χ_{T_1} and χ_{T_2} are their characteristic functions (i.e. $\chi_{T_1}(x) = 1$ if $x \in T_1$ and 0 otherwise).

This is quite a remarquable result that states that the wave operators are independent of the exact function f and only depend on the intervals on which the function is increasing or decreasing.

6.4.1 Application to potential scattering

Let us consider the wave equation on Minkowski space-time, with a short-range non-negative potential

$$V \in \mathscr{C}^0(\mathbb{R}^3),\ V \geq 0, V(x) = O(r^{-1-\alpha}),\ \alpha > 0.$$

i.e.

$$\partial_t^2 \phi - \Delta\phi + V\phi = 0. \tag{6.23}$$

Under Hamiltonian form it is expressed as

$$\frac{\mathrm{d}U}{\mathrm{d}t} = iAU,\ A = -i\begin{pmatrix} 0 & 1 \\ \Delta - V & 0 \end{pmatrix} \tag{6.24}$$

where A is self-adjoint on the Hilbert space $\mathscr{H}$, completion of $(\mathscr{C}_0^\infty(\mathbb{R}^3))^2$ in the norm

$$\|(u_1, u_2)\|^2_{\mathscr{H}} = \int_{\mathbb{R}^3} \left(|\nabla u_1|^2 + V|u_1|^2 + |u_2|^2\right) \mathrm{d}x.$$

We compare with the free wave equation

$$\frac{\mathrm{d}U}{\mathrm{d}t} = iA_0U,\ A_0 = -i\begin{pmatrix} 0 & 1 \\ \Delta & 0 \end{pmatrix} \tag{6.25}$$

where A_0 is self-adjoint on the Hilbert space $\mathscr{H}_0 = \dot{H}^1(\mathbb{R}^3) \times L^2(\mathbb{R}^3)$, completion of $(\mathscr{C}_0^\infty(\mathbb{R}^3))^2$ in the norm

$$\|(u_1, u_2)\|^2_{\mathscr{H}_0} = \int_{\mathbb{R}^3} \left(|\nabla u_1|^2 + |u_2|^2\right) \mathrm{d}x.$$

Equation (6.25) satisfies the Huygens principle but not (6.24). The two Hilbert spaces are different and so we will have to use an identifying operator. We put

$$h = -\Delta + V \text{ and } h_0 = -\Delta$$

and we consider the operators

$$\mu = (-\Delta + V)^{1/2} \text{ and } \mu_0 = (-\Delta)^{1/2}.$$

They both are well-defined by the functional calculus since h and h_0 are self-adjoint non-negative operators on $L^2(\mathbb{R}^3)$. The operator $\mu \otimes 1$ is an isometry from $\mathscr{H}$ onto $(L^2(\mathbb{R}^3))^2$ and similarly $\mu_0 \otimes 1$ is an isometry from $\mathscr{H}_0$ onto $(L^2(\mathbb{R}^3))^2$. We introduce the operator

$$\mathcal{J} := (\mu^{-1} \otimes 1)(\mu_0 \otimes 1) = \mu^{-1}\mu_0 \otimes 1 \in \mathscr{L}(\mathscr{H}_0, \mathscr{H})$$

that is in fact an isometry from $\mathscr{H}_0$ onto $\mathscr{H}$. We aim to define the direct and inverse wave operators as

$$W^{\pm} := s - \lim_{t\to\pm\infty} e^{-itA}\mathcal{J}e^{itA_0}, \tag{6.26}$$

$$\tilde{W}^{\pm} := s - \lim_{t\to\pm\infty} e^{-itA_0}\mathcal{J}^*e^{itA}. \tag{6.27}$$

The method here is quite different from the flat space-time construction of the previous section. We shall establish a link between these wave operators and those that compare the operators μ and μ_0. To this purpose, we take advantage of the fact that the equation is real and that it suffices to consider real solutions. In other words, we consider $\mathscr{H}$ and $\mathscr{H}_0$ are real Hilbert spaces. We define the *unitary* operators

$$T: \mathscr{H} \to L^2(\mathbb{R}^3;\mathbb{C}),\ TU = \mu u_1 + iu_2 \text{ and } T_0: \mathscr{H}_0 \to L^2(\mathbb{R}^3),\ T_0U = \mu_0 u_1 + iu_2.$$

Then we have

$$e^{iAt} = T^{-1}e^{i\mu t}T,\ e^{iA_0t} = T_0^{-1}e^{i\mu_0 t}T_0,\ \mathcal{J} = T^{-1}T_0 \text{ and } \mathcal{J}^* = \mathcal{J}^{-1} = T_0^{-1}T.$$

It follows that for any $t \in \mathbb{R}$,

$$e^{-iAt}\mathcal{J}e^{iA_0t} = T^{-1}e^{-i\mu t}e^{i\mu_0 t}T_0$$

and

$$e^{-iA_0t}\mathcal{J}^*e^{iAt} = T_0^{-1}e^{-i\mu_0 t}e^{i\mu t}T.$$

So it turns out that all we need to do is establish the existence of the direct and inverse wave operators

$$\begin{aligned} w^{\pm} &:= s - \lim_{t\to\pm\infty} e^{-i\mu t}e^{i\mu_0 t}, \\ \tilde{w}^{\pm} &:= s - \lim_{t\to\pm\infty} e^{-i\mu_0 t}e^{i\mu t}. \end{aligned}$$

In order to prove this, we shall use the invariance principle. We start by establishing that the difference of the resolvents of h and h_0 is trace-class:

$$\begin{aligned} (h+1)^{-1} - (h_0+1)^{-1} &= (h+1)^{-1}(h_0+1)(h_0+1)^{-1} - (h+1)^{-1}(h+1)(h_0+1)^{-1} \\ &= -(h+1)^{-1}V(h_0+1)^{-1} \\ &= -\left[(h+1)^{-1}(h_0+1)\right]\left[(h_0+1)^{-1}V^{1/2}\right]\left[V^{1/2}(h_0+1)^{-1}\right]. \end{aligned}$$

The idea is then to say that this is the product of a bounded operator and two Hilbert-Schmidt operators and is therefore trace-class. For this to work using Theorem 6.6, we must be able to ensure that

$$\frac{1}{1+|\xi|^2} \in L^2(\mathbb{R}^3)\,,$$

which is fine, and that

$$|V(x)|^{1/2} \in L^2(\mathbb{R}^3),\ \text{i.e. } V(x) \in L^1(\mathbb{R}^3),$$

which requires $\alpha > 2$ and is therefore a lot more restrictive than just V being short-range. If we wish to assume merely that $\alpha > 0$, we need to assume in addition that V is spherically symmetric. Then we can use a decomposition of the solution into spherical harmonics. This has the advantage that we then work for each angular momentum on a $1+1$-dimensional problem and we merely need to ensure that $V(r) \in L^1([0,+\infty[)$ which is satisfied for any $\alpha > 0$.

Remark 6.3. *We see here the major weakness of trace-class methods and the reason why in the early 2000's, when people started to extend time-dependent scattering to the Kerr metric (see for example [20, 21]), that is not spherically symmetric, they had to use Mourre theory instead. Another drawback of trace-classe perturbation methods is that the reason for the existence of wave operators is to be found in spectral properties of the difference of the Hamiltonians and is not easily interpreted as a weak version of the Huygens principle.*

In such situations, we then have that the wave operators $\Omega^{\pm}(h,h_0)$ and $\Omega^{\pm}(h_0,h)$ exist. Since μ and μ_0 are non negative, have purely continuous spectrum and $h=\mu^2$ and $h_0=\mu_0^2$, we may use the invariance principle with $T=]0,+\infty[$ and $f(x)=\sqrt{x}$ that is strictly increasing. We infer that $w^{\pm}$ and $\tilde{w}^{\pm}$ exist and this gives us our complete scattering theory. We can go further and compare to a simpler dynamics than the wave equation on flat space-time. This can be done by combining this result with the time-dependent scattering construction on flat space-time that we have described in Section 6.3. The chain rule applies when concatenating wave operators with a common intermediate dynamics and this gives us a scattering theory for the wave equation with a short-range potential, with a comparison dynamics that is simply given by the $1+1$-dimensional wave equation, i.e. by the radial null geodesic flow. We must however be careful to re-express the scattering theory of Section 6.3 using the propagator for the wave equation and the isometry $\mathcal{R}$ instead of the propagator for the rescaled equation, otherwise we will not have a common dynamics.

6.4.2 Scattering for the wave equation on the Schwarzschild metric

This is a historical result obtained by John Dimock in 1985 [12], partially based on joint work with Bernard Kay that was to appear later [13, 14]. It was the first scattering theory on a non trivial Einstein vacuum space-time. It is entirely based on trace-class perturbation methods.

Recall that the Schwarzschild metric in Schwarzschild coordinates (t,r,ω) is expressed as

$$g=F(r)\mathrm{d}t^2-\frac{1}{F(r)}\mathrm{d}r^2-r^2\mathrm{d}\omega^2,\ F(r)=1-\frac{2M}{r},$$

where $\mathrm{d}\omega^2$ is the Euclidean metric on the round 2-sphere. We work outside the black hole in the region $\{r>2M\}$. The d'Alembertian operator associated with g can be calculated using formula (2.28). It has the form

$$\Box_g=\frac{1}{F}\frac{\partial^2}{\partial t^2}-\frac{1}{r^2}\frac{\partial}{\partial r}r^2F\frac{\partial}{\partial r}-\frac{1}{r^2}\Delta_{S^2}\,. \tag{6.28}$$

We express it in terms of the Regge-Wheeler tortoise coordinate r_*. Recall that

$$r_*=r+2M\log(r-2M),\ \frac{\mathrm{d}r_*}{\mathrm{d}r}=\frac{1}{F}.$$

The d'Alembertian becomes

$$\Box_g=\frac{1}{F}\left(\frac{\partial^2}{\partial t^2}-\frac{\partial^2}{\partial r_*^2}-\frac{2F}{r}\frac{\partial}{\partial r_*}\right)-\frac{1}{r^2}\Delta_{S^2}\,. \tag{6.29}$$

The first order term remaining in (6.29) is not very pleasant in the sense that it has a slow fall-off at spacelike infinity, it is a long-range term. We use a rescaling by r as we did on Minkowski space-time. This will not completely eliminate this term, contrary to the flat case, but it will replace it with a short range potential. Indeed, we have

$$r\,\square_g \frac{1}{r} = \frac{1}{F}\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial r_*^2} + \frac{FF'}{r}\right) - \frac{1}{r^2}\Delta_{S^2}\,, \tag{6.30}$$

where the left-hand side is to be understood as the composition of three operators. Hence the wave equation outside a Schwarzschild black hole can be expressed in terms of the unknown

$$\tilde{\phi} = r\phi \tag{6.31}$$

as

$$\left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial r_*^2} + F\left(\frac{-\Delta_{S^2}}{r^2} + \frac{F'}{r}\right)\right)\tilde{\phi} = 0\,,\ \text{on } \mathbb{R}_t \times \mathbb{R}_{r_*} \times S^2_\omega\,, \tag{6.32}$$

which is notably simpler than the equation satisfied by the physical field ϕ and no longer contains any long-range term.

Remark 6.4. *Once again we see the rescaling by r appearing as a fundamental step in the construction of the scattering theory, just as in the time-dependent scattering theory for the wave equation on Minkowski space-time, or the conformal scattering theory using the incomplete compactification. It is this rescaling that will allow us to compare the full equation with the wave equation on Minkowski space-time in a simple manner and also to introduce an even simpler comparison dynamics that would not be accessible if the first order term remained.*

The Hamiltonian for Equation (6.32) is

$$A = -i\begin{pmatrix} 0 & 1 \\ \partial^2_{r_*} - F(\frac{-\Delta_{S^2}}{r^2} + \frac{F'}{r}) & 0 \end{pmatrix}. \tag{6.33}$$

It is self-adjoint on the space $\mathscr{H}$, completion of $(\mathscr{C}_0^\infty(\mathbb{R}_{r_*}\times S^2))^2$ in the norm

$$\|U\|^2_{\mathcal{H}} = \int_{\mathbb{R}_{r_*}\times S^2}\left\{|\partial_{r_*}u_1|^2 + F\left[\frac{|\nabla_{S^2}u_1|^2}{r^2} + \left(m^2 + \frac{1}{r}\frac{dF}{dr}\right)|u_1|^2\right] + |u_2|^2\right\}dr_*d\omega. \tag{6.34}$$

For more details, see [12].

There are two asymptotic regions, the horizon ($r_* \to -\infty$) and infinity ($r_* \to +\infty$), and we need to introduce a comparison dynamics for each. At the horizon, the function F falls-off exponentially fast in r_*. As a consequence a natural comparison dynamics in this region is given by

$$A_0 = -i\begin{pmatrix} 0 & 1 \\ \partial^2_{r_*} & 0 \end{pmatrix}. \tag{6.35}$$

At infinity, two approaches can be adopted. The first one consists in considering that the region is asymptotically flat and it is therefore natural to chose the Hamiltonian for the wave equation on Minkowski space-time as a comparison dynamics. There is however a freedom in the manner in which we choose to glue Minkowski space-time onto the Schwarzschild geometry in the neighbourhood of infinity; this can be expressed in a simple manner as a choice of coordinates on Minkowski space-time in terms of the coordinates on Schwarzschild's space-time. We choose to use the coordinates (t, r_*, ω) as spherical coordinates in flat space-time. We obtain the following comparison dynamics

$$A_1 = -i \begin{pmatrix} 0 & 1 \\ \partial^2_{r_*} + \frac{\Delta_{S^2}}{r_*^2} & 0 \end{pmatrix} . \tag{6.36}$$

Using the spherical symmetry, this can be understood as a short range perturbation of $\partial^2_{r_*}$. This is done by decomposing the field on the basis of spherical harmonics. For each given angular dependence the perturbation reduces to a short-range scalar potential. The eigenvalue $l(l+1)$ of the spherical Laplacian appears as a factor in the potential, as a reminder of the fact that Δ_{S^2} is not a bounded operator. The unitarity of the propagators allows to recombine the full solution and to infer the existence of the strong limit. Note that using r instead of r_*, however natural it may seem, would lead to an artificially long range perturbation that would make the scattering theory much more complicated and less meaningful. Another approach is to say that since the perturbation $F(\frac{-\Delta_{S^2}}{r^2} + \frac{F'}{r})$ is short-range at infinity, we can choose A_0 defined above in (6.35) as the comparison dynamics at infinity. Both the choices (6.35) and (6.36) are valid and they provide two different and in a sense inequivalent descriptions of the asymptotic behaviour of massless scalar fields on Minkowski space-time. This is because the two Hamiltonians A_0 and A_1 are self-adjoint on two different Hilbert spaces, namely

$$\mathscr{H}_0 = \dot{H}^1(\mathbb{R}; L^2(S^2)) \times L^2(\mathbb{R} \times S^2) , \tag{6.37}$$

$$\mathscr{H}_1 = \dot{H}^1(\mathbb{R}^3) \times L^2(\mathbb{R}^3) , \tag{6.38}$$

that are the completion of $\mathcal{C}_0^\infty(\mathbb{R} \times S^2)^2$ (resp. $\mathcal{C}_0^\infty(\mathbb{R}^3)^2$) in the respective norms

$$\|(f,g)\|^2_{\mathscr{H}_0} = \int_{\mathbb{R}\times S^2} \left(|\partial_{r_*} f(r_*,\omega)|^2 + |g(r_*,\omega)|^2 \right) \mathrm{d}u \mathrm{d}^2\omega , \tag{6.39}$$

$$\|(f,g)\|^2_{\mathscr{H}_1} = \int_{\mathbb{R}^3} \left(|\partial_{r_*} f(r_*,\omega)|^2 + \frac{1}{r_*^2} |\nabla_{S^2} f(r_*,\omega)|^2 + |g(r_*,\omega)|^2 \right) \mathrm{d}r_* \mathrm{d}^2\omega . \tag{6.40}$$

The corresponding wave operators take into account the existence of the two asymptotic regions and can be defined as follows for $j = 0, 1$ depending on the comparison dynamics chosen at infinity

$$W^\pm := \left(\mathrm{s} - \lim_{t\to\pm\infty} e^{-itA} \overleftarrow{\mathcal{J}}_0 e^{itA_0} \, , \; \mathrm{s} - \lim_{t\to\pm\infty} e^{-itA} \overrightarrow{\mathcal{J}}_j e^{itH_j} \right) , \tag{6.41}$$

$$\tilde{W}^\pm := \mathrm{s} - \lim_{t\to\pm\infty} e^{-itA_0} \overleftarrow{\mathcal{J}}_0^{\,*} e^{itA} + \mathrm{s} - \lim_{t\to\pm\infty} e^{-itA_j} \overrightarrow{\mathcal{J}}_j^{\,*} e^{itH} , \tag{6.42}$$

where the identification operators $\mathcal{J}_0 \in \mathscr{L}(\mathscr{H}; \mathscr{H}_0)$ and $\mathcal{J}_1 \in \mathscr{L}(\mathscr{H}; \mathscr{H}_1)$ are defined using cut-off functions

$$\overleftarrow{\chi} \in \mathscr{C}_0^\infty(\mathbb{R}) , \;\; \overleftarrow{\chi}(x) \equiv 1 \text{ for } x \leq -R , \;\; \overleftarrow{\chi}(x) \equiv 0 \text{ for } x \geq R ,$$

$$\overrightarrow{\chi} \in \mathscr{C}_0^\infty(\mathbb{R}) , \;\; \overrightarrow{\chi}(x) \equiv 1 \text{ for } x \geq R , \;\; \overrightarrow{\chi}(x) \equiv 0 \text{ for } x \leq \varepsilon$$

and the operators

$$\begin{aligned}
\mu &:= \left(-\partial_{r_*}^2 + F\left(\frac{-\Delta_{S^2}}{r^2} + \frac{F'}{r}\right)\right)^{1/2} , \\
\mu_1 &:= \left(-\partial_{r_*}^2 - \frac{\Delta_{S^2}}{r_*^2}\right)^{1/2} , \\
\mu_0 &:= \left(-\partial_{r_*}^2\right)^{1/2} ,
\end{aligned}$$

by

$$\overleftarrow{\mathcal{J}_0} := \mu^{-1} \overleftarrow{\chi} \mu_0 \oplus \overleftarrow{\chi} = \left(\mu^{-1} \otimes 1\right) \overleftarrow{\chi} \left(\mu_0 \otimes 1\right) , \tag{6.43}$$

$$\overrightarrow{\mathcal{J}_j} = \mu^{-1} \overrightarrow{\chi} \mu_{\mathrm{j}} \oplus \overrightarrow{\chi} = \left(\mu^{-1} \otimes 1\right) \overrightarrow{\chi} \left(\mu_{\mathrm{j}} \otimes 1\right) . \tag{6.44}$$

The proof of the existence and completeness of the wave operators for the comparison dynamics A_1 at infinity has been established by Dimock in [12]. The proof with the choice A_0 is not explicitly present in the literature in the exact form given here but an equivalent result is proven by the author in [40].

The idea of the construction by Dimock is very clearly explained in his paper. One would naturally like to use mere cut-off functions as identification operators, but they are not bounded operators between the various energy spaces. However they are bounded at the level of the corresponding L^2 spaces. So one composes $\overleftarrow{\chi}$ or $\overrightarrow{\chi}$ with the operators or their inverses

$$\begin{aligned}
\mu \otimes 1 &\in \mathscr{L}\left(\mathscr{H}; L^2(\mathbb{R} \times S^2, \mathrm{d}r_* \mathrm{d}^2\omega)^2\right) , \\
\mu_0 \otimes 1 &\in \mathscr{L}\left(\mathscr{H}_0; L^2(\mathbb{R} \times S^2, \mathrm{d}r_* \mathrm{d}^2\omega)^2\right) , \\
\mu_1 \otimes 1 &\in \mathscr{L}\left(\mathscr{H}_1; L^2(\mathbb{R}^3, \mathrm{d}r_* \mathrm{d}^2\omega)^2\right) ,
\end{aligned}$$

that are *isometries*. The existence of the wave operators then translates into the comparison between solutions to evolution equations associated to A and A_0 or A_1 seen through these isometries. Considering $\Phi \in \mathscr{H}_0$ for instance and putting $\Psi := \mathrm{s} - \lim_{t\to+\infty} e^{-itA} \overleftarrow{\mathcal{J}_0} e^{itA_0}\Phi$, we have

$$\begin{aligned}
0 &= \lim_{t\to+\infty} \left\| e^{-itA} \left(\mu^{-1} \otimes 1\right) \overleftarrow{\chi} \left(\mu_0 \otimes 1\right) e^{itA_0}\Phi - \Psi \right\|_{\mathscr{H}} \\
&= \lim_{t\to+\infty} \left\| \left(\mu^{-1} \otimes 1\right) \overleftarrow{\chi} \left(\mu_0 \otimes 1\right) e^{itA_0}\Phi - e^{itA}\Psi \right\|_{\mathscr{H}} \\
&= \lim_{t\to+\infty} \left\| \left(\mu \otimes 1\right) \left(\mu^{-1} \otimes 1\right) \overleftarrow{\chi} \left(\mu_0 \otimes 1\right) e^{itA_0}\Phi - \left(\mu \otimes 1\right) e^{itA}\Psi \right\|_{L^2(\mathbb{R}^3;\mathrm{d}r_*\mathrm{d}^2\omega)} \\
&= \lim_{t\to+\infty} \left\| \overleftarrow{\chi} \left(\mu_0 \otimes 1\right) e^{itA_0}\Phi - \left(\mu \otimes 1\right) e^{itA}\Psi \right\|_{L^2(\mathbb{R}^3;\mathrm{d}r_*\mathrm{d}^2\omega)} .
\end{aligned}$$

The proof is essentially similar to the one we gave for potential scattering in the previous section, modulo the additional feature of the cut-off functions. Just as we saw in the potential case, a spherical harmonics decomposition needs to be used in order to work in lower dimensions and to be able to apply the Kuroda-Birman Theorem.

6.5 Minimal velocity estimates

Mourre theory is inspired by quantum mechanics in the Heisenberg formalism. Considering a given evolution equation, the idea is to construct for the hamiltonian a so-called conjugate operator that resembles a position operator or a generator of dilations and is increasing along the evolution. This property is usually derived locally in energy by means of pseudo-differential cut-offs. It entails among other things a weak version of the Huygens principle that is called a minimal velocity estimate. It is more flexible than trace-class methods, in particular, it has successfully been applied to the Kerr metric (see for example [21]). Another major advantage is that the minimal velocity estimate is easy to interpret geometrically and easily replaces the Huygens principle in the construction of wave operators. The drawback is that the theory is extremely technical and requires great care in its application. We give here however a very simple example in which one can obtain a minimal velocity estimate without any of the analytic complications: Dirac's equation in $1+1$-dimensions. The equation is

$$\partial_t \phi + iH\phi = 0\,,\ H = -i\gamma\partial_x + VI_2 \tag{6.45}$$

where

$$\gamma = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}\,,\ I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\,,\ V(x) = O(|x|^{-1-\alpha})\,,\ \alpha > 0\,.$$

We consider a function $f \in \mathcal{C}_0^\infty(\mathbb{R})$ such that, for $0 < \varepsilon < 1$ given, $\mathrm{supp} f \subset [-1+\varepsilon, 1-\varepsilon]$, and we define the function F by

$$F(t) = \int_{-\infty}^{t} (f(s))^2 \mathrm{d}s\,.$$

Next, we introduce the observable

$$\Phi(t) := F\left(\gamma\frac{x}{t}\right)\,.$$

Then

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{itH}\Phi(t)e^{-itH}\right) = e^{itH}D\Phi(t)e^{-itH}$$

where $D\Phi(t)$ is the Heisenberg derivative of Φ:

$$D\Phi(t) = \frac{\mathrm{d}\Phi}{\mathrm{d}t} + i\left[H, \Phi\right]\,.$$

This can be expressed explicitly as follows

$$\begin{aligned} D\Phi(t) &= \frac{\mathrm{d}\Phi}{\mathrm{d}t} + \gamma\frac{\mathrm{d}\Phi}{\mathrm{d}x} \\ &= \gamma\frac{1}{t}\left(\gamma - \frac{x}{t}\right) f^2\left(\gamma\frac{x}{t}\right) \\ &= \frac{1}{t}\left(1 - \gamma\frac{x}{t}\right) f^2\left(\gamma\frac{x}{t}\right)\,. \end{aligned}$$

Given the support of f, it follows that as a quadratic form,

$$D\Phi(t) \geq \frac{\varepsilon}{t} f^2\left(\gamma\frac{x}{t}\right)\,.$$

Using this inequality, we obtain for $T > 0$

$$\begin{aligned} \int_1^T \left\| f\left(\gamma\frac{x}{t}\right) e^{-itH}\phi \right\|^2 \frac{\mathrm{d}t}{t} &= \int_1^T \left\langle f^2\left(\gamma\frac{x}{t}\right) e^{-itH}\phi\,,\ e^{-itH}\phi \right\rangle \frac{\mathrm{d}t}{t} \\ &\leq \frac{1}{\varepsilon}\int_1^T \left\langle D\Phi(t)e^{-itH}\phi\,,\ e^{-itH}\phi\right\rangle \mathrm{d}t \end{aligned}$$

and

$$\begin{aligned} \int_1^T \left\langle e^{itH}D\Phi(t)e^{-itH}\phi\,,\ \phi\right\rangle \mathrm{d}t &= \frac{\mathrm{d}}{\mathrm{d}t}\int_1^T \left\langle e^{itH}\Phi(t)e^{-itH}\phi\,,\ \phi\right\rangle \mathrm{d}t \\ &= \left(\left\langle e^{iTH}\Phi(T)e^{-iTH}\phi\,,\ \phi\right\rangle - \left\langle e^{iH}\Phi(1)e^{-iH}\phi\,,\ \phi\right\rangle\right)\,. \end{aligned}$$

It follows that

$$\int_1^T \left\| f\left(\gamma\frac{x}{t}\right) e^{-itH}\phi \right\|^2 \frac{\mathrm{d}t}{t} \leq \frac{2}{\varepsilon}\|\Phi\|_{L^\infty(\mathbb{R}\,;\ \mathscr{L}(\mathscr{H}))}\|\phi\|^2 \tag{6.46}$$

with

$$\|\Phi\|_{L^\infty(\mathbb{R}\,;\ \mathscr{L}(\mathscr{H}))} = \|F\|_{L^\infty} = \|f\|_{L^2}^2\,.$$

Taking the limit of (6.46) as $T \to +\infty$ and using the fact that the bound depends only on the L^2 norm of f and the localisation of its support, we obtain a minimal velocity estimate : for any $\varepsilon \in]0,1[$ there exists $C > 0$ such that for all $\phi \in \mathscr{H}$,

$$\int_1^{+\infty} \left\| \mathbf{1}_{[-1+\varepsilon,1-\varepsilon]}\left(\gamma\frac{x}{t}\right) e^{-itH}\phi \right\|^2 \frac{\mathrm{d}t}{t} \leq C\|\phi\|^2\,. \tag{6.47}$$

Chapter 7

Conformal scattering on flat space-time

Constructing a conformal scattering theory for a field equation on flat space-time consists in showing that one can associate to the field some restriction in some sense, at the future and past conformal boundaries, and that the field is entirely described by these restrictions. This will imply that we can define a "scattering operator" that to the past restriction associates the future restriction and is invertible. In the case of Minkowski space-time, we have the luxury of being able to embed it in a compact manifold with boundary. So our conformal boundary can be chosen to be quite complete. But only null infinity is relevant for conformal scattering, so it would seem natural to work instead with the incomplete compactification that we have described in Chapter 3. We establish conformal scattering theories for both the complete compactification and the incomplete one and we compare the two.

7.1 Consequences of conformal invariance

In the case of Minkowski space-time, we see that $\phi \in \mathcal{D}'(\mathbb{R}^4)$ satisfies (4.5) if and only if $\hat{\phi} := \Omega^{-1}\phi$ (Ω defined by (3.7)) satisfies

$$\Box_{\mathfrak{e}} \hat{\phi} + \hat{\phi} = 0 \,, \tag{7.1}$$

where

$$\Box_{\mathfrak{e}} = \partial_\tau^2 - \Delta_{S^3} \,.$$

The Cauchy data for both equations are related by simple relations. Let us denote

$$\phi_0 := \phi|_{t=0} \,, \;\; \phi_1 := \partial_t \phi|_{t=0} \,, \;\; \hat{\phi}_0 := \hat{\phi}|_{\tau=0} \,, \;\; \hat{\phi}_1 := \partial_\tau \hat{\phi}|_{\tau=0} \,.$$

We can express the hatted data in terms of the unhatted ones. We start from the expression (3.7) of Ω and the relations (3.6) between the partial derivatives with respect to (t, r) and to (τ, ζ) at $t = 0$. These entail

$$\begin{aligned} \Omega^{-1}|_{\tau=0} &= \frac{1+(r)^2}{2} \,, \\ \partial_\tau \Omega^{-1}|_{\tau=0} &= 0 \,, \end{aligned}$$

whence

$$\begin{aligned}
\hat{\phi}|_{\tau=0} &= \frac{1+r^2}{2}\phi|_{t=0}\,, \\
\partial_\tau\hat{\phi}|_{\tau=0} &= \left(\Omega^{-1}|_{\tau=0}\right)\left(\frac{1+r^2}{2}\partial_\tau\phi|_{t=0}\right) + \left(\partial_\tau\Omega^{-1}|_{\tau=0}\right)\left(\phi|_{t=0}\right) \\
&= \frac{(1+r^2)^2}{4}\partial_\tau\phi|_{t=0}\,.
\end{aligned}$$

Therefore, we have

$$\hat{\phi}_0 = \frac{1+r^2}{2}\phi_0\,,\ \hat{\phi}_1 = \frac{(1+r^2)^2}{4}\phi_1\,. \tag{7.2}$$

On the Einstein cylinder, the Cauchy problem for (7.1) can be solved in any Sobolev space on S^3. This is a consequence of Theorem 4.1 but can also be established explicitly, for Sobolev spaces of non-negative integral order, by the spectral approach, which gives us some conserved norms in addition.

Theorem 7.1. *The operator*

$$A = -i\begin{pmatrix} 0 & 1 \\ \Delta_{S^3}-1 & 0 \end{pmatrix}$$

is self-adjoint on

$$\mathscr{H} = H^1(S^3)\times L^2(S^3)$$

with domain $D(A) = H^2(S^3)\times H^1(S^3)$. As a consequence, for any $k\in\mathbb{N}$ and any $(\hat{\phi}_0,\hat{\phi}_1)\in H^{k+1}(S^3)\times H^k(S^3)$, there exists a unique solution

$$\hat{\phi}\in\bigcap_{j=0}^{k+1}\mathscr{C}^j(\mathbb{R}_\tau;H^{k+1-j}(S^3))$$

of (7.1) such that $\hat{\phi}|_{\tau=0}=\hat{\phi}_0$ and $\partial_\tau\hat{\phi}|_{\tau=0}=\hat{\phi}_1$. Moreover for all $j\in\{0,1,...,k\}$ the norm of $(\hat{\phi}(\tau),\partial_\tau\hat{\phi}(\tau))$ in $D(A^j)$ is constant in time and is equivalent to the $H^{j+1}\times H^j$-norm on S^3.

In particular, for data $\hat{\phi}|_{\tau=0}\,,\ \partial_\tau\hat{\phi}|_{\tau=0}\in\mathscr{C}^\infty(S^3)$, the associated solution of (7.1) is smooth on the whole Einstein cylinder. Hence, we have the following result.

Theorem 7.2. *Assume that the data ϕ_0,ϕ_1, for the Cauchy problem for the wave equation (4.5) on Minkowski space-time, are such that the corresponding data for the rescaled field*

$$\hat{\phi}_0 = \frac{1+r^2}{2}\phi_0\,,\ \hat{\phi}_1 = \frac{(1+r^2)^2}{4}\phi_1\,,$$

extend as smooth functions on S^3. Then the rescaled solution $\hat{\phi}=\Omega^{-1}\phi$ extends as a smooth function on the Einstein cylinder $\mathfrak{E}$.

Remark 7.1. *Implicit in the hypotheses of Theorem 7.2 are some requirements on the fall-off of initial data for ϕ. Indeed, the smoothness of $\hat{\phi}_0$ and $\hat{\phi}_1$ on S^3 entails that*

$$\lim_{r\to+\infty} r^2\phi(0,r,\omega) = 2\hat{\phi}_0(i^0)\,,$$
$$\lim_{r\to+\infty} r^4\partial_t\phi(0,r,\omega) = 4\hat{\phi}_1(i^0)\,.$$

Hence, ϕ_0 decays as $1/r^2$ and ϕ_1 as $1/r^4$. We can compare this with a naive description of the requirement of finite physical energy in terms of fall-off in r. The physical energy on a t =constant slice is given by

$$\int_{\mathbb{R}^3}\left(|\partial_t\phi|^2 + |\partial_r\phi|^2 + \frac{1}{r^2}|\nabla_{S^2}\phi|^2\right) r^2\mathrm{d}r\mathrm{d}\omega.$$

If we consider only integral powers of $1/r$, the finiteness of the energy would correspond to

$$\nabla_{S^2}\phi = O(1/r) \text{ and } \partial_t\phi,\ \partial_r\phi = O(1/r^2) \text{ as } r\to+\infty.$$

It appears that the assumptions that allow us to extend the rescaled fields smoothly to the Einstein cylinder are stronger than the finiteness of the energy of the data.

7.2 Local and global decay

If the assumptions of Theorem 7.2 are satisfied, this implies that $\hat{\phi}$ as well as all its successive derivatives $\partial_\tau^k\partial_\zeta^l\nabla_{S^2}^n\hat{\phi}$, $k,l,n\in\mathbb{N}$, extend as smooth function on the Einstein cylinder and are therefore uniformly bounded on $\mathbb{M}$. This can be used to infer many local and global decay results for such solutions. We start with a simple power-law behaviour along timelike and null directions.

Proposition 7.1. *Under the assumptions of Theorem 7.2, the solution ϕ of (4.5) associated to the data ϕ_0, ϕ_1 at $t=0$ satisfies the following properties.*

1. **Decay along null directions.** *There exist smooth functions $\hat{\phi}^\pm\in\mathcal{C}^\infty(\mathbb{R}\times S^2)$ given as the restrictions of $\hat{\phi}$ to future and past null infinities described respectively in (u,ω) and (v,ω) coordinates, such that*

$$\lim_{r\to+\infty} r\phi(t=r+u,r,\omega) = \frac{1}{\sqrt{1+u^2}}\,\hat{\phi}^+(u,\omega)\,, \tag{7.3}$$
$$\lim_{r\to+\infty} r\phi(t=-r+v,r,\omega) = \frac{1}{\sqrt{1+v^2}}\,\hat{\phi}^-(v,\omega)\,. \tag{7.4}$$

The two functions in the right-hand side of (7.3) and (7.4) are the future and past radiation fields of ϕ.

2. **Decay along timelike directions.** *For all $(r,\omega)\in\mathbb{R}^+\times S^2$,*

$$\lim_{t\to\pm\infty} t^2\phi(t,r,\omega) = 2\hat{\phi}(i^\pm)\,.$$

In other words, the physical solution ϕ decays like $1/r$ along radial null geodesics and like $1/t^2$ along the integral lines of ∂_t.

Proof. This is a simple consequence of Theorem 7.2. Let us first consider the decay in null directions. For $(u,\omega) \in \mathbb{R} \times S^2$, we have

$$\begin{aligned}\phi(t = r+u, r, \omega) &= \left(\Omega|_{\{(t=r+u,r,\omega)\}}\right) \hat{\phi}(t = r+u, r, \omega) \\ &= \frac{2}{\sqrt{(1+u^2)(1+v^2)}} \hat{\phi}(t = r+u, r, \omega) \\ &= \frac{2}{\sqrt{(1+u^2)(1+(u+2r)^2)}} \hat{\phi}(t = r+u, r, \omega)\,.\end{aligned}$$

It follows that

$$r\phi(t = r+u, r, \omega) = \frac{r}{\sqrt{(1+(u+2r)^2)}} \frac{2}{\sqrt{(1+u^2)}} \hat{\phi}(t = r+u, r, \omega)\,.$$

Hence, using the (u, R, ω) coordinate system to describe the right-hand side, we have

$$\lim_{r \to +\infty} r\phi(t = r+u, r, \omega) = \frac{\hat{\phi}(u, R=0, \omega)}{\sqrt{1+u^2}}$$

and this is equation (7.3). The proof for (7.4) is completely analogous.

We perform a similar calculation on an integral curve of ∂_t. Let $(r,\omega) \in \mathbb{R}^+ \times S^2$ be given,

$$\begin{aligned}\phi(t, r, \omega) &= \left(\Omega|_{\{(t,r,\omega)\}}\right) \hat{\phi}(t, r, \omega) \\ &= \frac{2}{\sqrt{(1+(t-r)^2)(1+(t+r)^2)}} \hat{\phi}(t, r, \omega)\,.\end{aligned}$$

Therefore

$$\lim_{t \to \pm\infty} t^2 \phi(t, r, \omega) = 2\hat{\phi}(i^\pm)\,.$$

This concludes the proof. □

We now turn to global decay properties, first in terms of the time variable t, then along a family of lightcones accumulating at i^+.

Proposition 7.2 (Decay in t)**.** *Under the assumptions of Theorem 7.2, the solution ϕ decays uniformly like $1/t$, i.e. there exists a constant $C > 0$ (depending on ϕ) such that for all (t, r, ω),*

$$|\phi(t, r, \omega)| \leq \frac{C}{\sqrt{1+t^2}}\,. \tag{7.5}$$

Moreover, the L^2 norm of ϕ on Σ_t for the measure $\mathrm{d}r\mathrm{d}\omega$ tends to zero as $t \to +\infty$

$$\lim_{t \to +\infty} \int_{\mathbb{R}^+ \times S^2} |\phi(t, r, \omega)|^2 \mathrm{d}r\mathrm{d}\omega = 0\,. \tag{7.6}$$

Due to the fact that the generators of rotations and ∂_t are Killing vector fields, properties (7.5) and (7.6) are also valid for $\partial_t^k \nabla_{S^2}^n \phi$ for all $k, n \in \mathbb{N}$.

Proof. Recall that

$$\phi = \Omega\hat{\phi} = \frac{2}{\sqrt{(1+(t+r)^2)(1+(t-r)^2)}}\hat{\phi}$$

and

$$(1+(t+r)^2)(1+(t-r)^2) = 1+(t^2-r^2)^2+2t^2+2r^2 \geq 1+2t^2\,.$$

Therefore

$$|\phi(t,r,\omega)| \leq \frac{2}{\sqrt{1+2t^2}}\|\hat{\phi}\|_{L^\infty(\mathbb{M})}\,.$$

Let us now prove (7.6). We can obviously improve the estimate above as

$$(1+(t+r)^2)(1+(t-r)^2) \geq 1+2t^2+2r^2\,.$$

This implies that

$$|\phi(t,r,\omega)|^2 \leq \frac{4}{1+2t^2+2r^2}\|\hat{\phi}\|^2_{L^\infty(\mathbb{M})}$$

and (7.6) follows by the dominated convergence Theorem. □

We also consider the lightcones C_T for $T \geq 0$ defined by

$$C_T := \{(t,r,\omega)\ ;\ t = T+r\}\,. \tag{7.7}$$

We have the following uniform decay result in terms of T.

Proposition 7.3 (Decay in T)**.** *Under the assumptions of Theorem 7.2, the solution ϕ satisfies for all $T>0$ and (r,ω),*

$$|\phi(T+r,r,\omega)| \leq \frac{2}{\sqrt{(1+T^2)(1+T^2+4r^2)}}\|\hat{\phi}\|_{L^\infty(\mathbb{M})}\,. \tag{7.8}$$

This implies that the L^2 norm of ϕ on C_T for the measure $\mathrm{d}r\mathrm{d}\omega$ *decays faster than $1/T$, i.e.*

$$\lim_{T\to+\infty}(1+T^2)\int_{\mathbb{R}^+\times S^2}|\phi(T+r,r,\omega)|^2\mathrm{d}r\mathrm{d}\omega = 0\,. \tag{7.9}$$

Properties (7.8) and (7.9) are also valid for $\partial_t^k\nabla^n_{S^2}\phi$ for all $k,n\in\mathbb{N}$.

Proof. Property (7.8) is a direct consequence of $\phi = \Omega\hat{\phi}$, of Theorem 7.2 and of the fact that for $t = T+r$, $T\geq 0$, we have

$$\Omega^2 = \frac{4}{(1+T^2)(1+(T+2r)^2)}\,.$$

Then (7.9) follows by the dominated convergence Theorem. □

Remark 7.2. *The lightcones C_T are singular at their tip. If one prefers to work with smooth hypersurfaces, C_T can be replaced by the hyperboloid*

$$S_T = \left\{ (t,r,\omega) \, , \; t = T + \sqrt{1+r^2} \right\} \, .$$

The decay rates in terms of T obtained for C_T will also be valid for S_T.

For integrations by parts, the singularity at the tip does not matter since Stokes' Theorem can be applied to closed bounded hypersurfaces that are piecewise $\mathcal{C}^1$. Moreover the energy functionals on C_T will have much simpler expressions that those on S_T. This is why we prefer to work with the lightcones.

It is interesting to infer decay rates for derivatives of ϕ as well. This can be done using the smoothness of $\hat{\phi}$ and of the coordinates τ and ζ on the Einstein cylinder. First, we have

$$\begin{aligned}
\partial_t &= \frac{\partial \tau}{\partial t}\frac{\partial}{\partial \tau} + \frac{\partial \zeta}{\partial t}\frac{\partial}{\partial \zeta} \\
&= \frac{2(1+t^2+r^2)}{(1+(t+r)^2)(1+(t-r)^2)}\frac{\partial}{\partial \tau} + \frac{4tr}{(1+(t+r)^2)(1+(t-r)^2)}\frac{\partial}{\partial \zeta} ,
\end{aligned}$$

$$\begin{aligned}
\partial_r &= \frac{\partial \tau}{\partial r}\frac{\partial}{\partial \tau} + \frac{\partial \zeta}{\partial r}\frac{\partial}{\partial \zeta} \\
&= \frac{4tr}{(1+(t+r)^2)(1+(t-r)^2)}\frac{\partial}{\partial \tau} + \frac{2(1+t^2+r^2)}{(1+(t+r)^2)(1+(t-r)^2)}\frac{\partial}{\partial \zeta} .
\end{aligned}$$

Then,

$$\begin{aligned}
\partial_t \phi &= (\partial_t \Omega)\, \hat{\phi} + \Omega \partial_t \hat{\phi} \\
&= \left(\frac{-2(t+r)}{(1+(t-r)^2)^{1/2}(1+(t+r)^2)^{3/2}} + \frac{-2(t-r)}{(1+(t-r)^2)^{3/2}(1+(t+r)^2)^{1/2}} \right) \hat{\phi} \\
&\quad + \frac{2}{\sqrt{(1+(t-r)^2)(1+(t+r)^2)}} \frac{2(1+t^2+r^2)}{(1+(t+r)^2)(1+(t-r)^2)} \frac{\partial \hat{\phi}}{\partial \tau} \\
&\quad + \frac{2}{\sqrt{(1+(t-r)^2)(1+(t+r)^2)}} \frac{4tr}{(1+(t+r)^2)(1+(t-r)^2)} \frac{\partial \hat{\phi}}{\partial \zeta} , \\
\partial_r \phi &= (\partial_r \Omega)\, \hat{\phi} + \Omega \partial_r \hat{\phi} \\
&= \left(\frac{-2(t+r)}{(1+(t-r)^2)^{1/2}(1+(t+r)^2)^{3/2}} + \frac{2(t-r)}{(1+(t-r)^2)^{3/2}(1+(t+r)^2)^{1/2}} \right) \hat{\phi} \\
&\quad + \frac{2}{\sqrt{(1+(t-r)^2)(1+(t+r)^2)}} \frac{4tr}{(1+(t+r)^2)(1+(t-r)^2)} \frac{\partial \hat{\phi}}{\partial \tau} \\
&\quad + \frac{2}{\sqrt{(1+(t-r)^2)(1+(t+r)^2)}} \frac{2(1+t^2+r^2)}{(1+(t+r)^2)(1+(t-r)^2)} \frac{\partial \hat{\phi}}{\partial \zeta} .
\end{aligned}$$

We obtain the following decay rates in terms of t and T.

Proposition 7.4. *Under the assumptions of Theorem 7.2, for all $t \geq 0$ and (r, ω)*

$$|\partial_t \phi(t, r, \omega)| \leq \frac{4}{\sqrt{1+(t+r)^2}} \left(\|\hat{\phi}\|_{L^\infty(\mathbb{M})} + \|\partial_\tau \hat{\phi}\|_{L^\infty(\mathbb{M})} + \|\partial_\zeta \hat{\phi}\|_{L^\infty(\mathbb{M})} \right), \tag{7.10}$$

$$|\partial_r \phi(t, r, \omega)| \leq \frac{4}{\sqrt{1+(t+r)^2}} \left(\|\hat{\phi}\|_{L^\infty(\mathbb{M})} + \|\partial_\tau \hat{\phi}\|_{L^\infty(\mathbb{M})} + \|\partial_\zeta \hat{\phi}\|_{L^\infty(\mathbb{M})} \right) \tag{7.11}$$

and for all $T \geq 0$ and (r, ω),

$$|\partial_t \phi(T+r, r, \omega)| \leq \frac{4}{(1+T^2)\sqrt{1+(T+2r)^2}} \left(\|\hat{\phi}\|_{L^\infty(\mathbb{M})} + \|\partial_\tau \hat{\phi}\|_{L^\infty(\mathbb{M})} + \|\partial_\zeta \hat{\phi}\|_{L^\infty(\mathbb{M})} \right), \tag{7.12}$$

$$|\partial_r \phi(T+r, r, \omega)| \leq \frac{4}{(1+T^2)\sqrt{1+(T+2r)^2}} \left(\|\hat{\phi}\|_{L^\infty(\mathbb{M})} + \|\partial_\tau \hat{\phi}\|_{L^\infty(\mathbb{M})} + \|\partial_\zeta \hat{\phi}\|_{L^\infty(\mathbb{M})} \right). \tag{7.13}$$

7.3 A first glance at peeling properties

What happens to the fall-off rate along null geodesics if instead of ϕ we consider $(\partial_t + \partial_r)\phi$ and $(\partial_t - \partial_r)\phi$? We proceed similarly, working with coordinates (u, v, ω) and (τ, ζ, ω) :

$$(\partial_t + \partial_r)\phi = \partial_v \phi = (\partial_v \Omega)\hat{\phi} + \Omega \frac{\partial \tau}{\partial v} \partial_\tau \hat{\phi} + \Omega \frac{\partial \zeta}{\partial v} \partial_\zeta \hat{\phi} .$$

We have

$$\Omega = \frac{2}{\sqrt{(1+u^2)(1+v^2)}} , \ \partial_v \Omega = \frac{-2v}{(1+u^2)^{1/2}(1+v^2)^{3/2}} , \ \frac{\partial \tau}{\partial v} = \frac{\partial \zeta}{\partial v} = \frac{1}{1+v^2} .$$

So along an outgoing null geodesic, where we have $v \simeq 2r$ and u is constant, $\partial_v \phi$ falls-off like $1/r^2$.

As for $\partial_u \phi$:

$$(\partial_t - \partial_r)\phi = \partial_u \phi = (\partial_u \Omega)\hat{\phi} + \Omega \frac{\partial \tau}{\partial u} \partial_\tau \hat{\phi} + \Omega \frac{\partial \zeta}{\partial u} \partial_\zeta \hat{\phi} .$$

We have

$$\partial_u \Omega = \frac{-2u}{(1+u^2)^{3/2}(1+v^2)^{1/2}} , \ \frac{\partial \tau}{\partial v} = -\frac{\partial \zeta}{\partial u} = \frac{1}{1+u^2} .$$

So along an outgoing null geodesic, $\partial_u \phi \simeq 1/r$.

This is one aspect of the peeling. The derivatives tangent to an outgoing null geodesic, when going out to infinity along that geodesic, behave better than the derivatives along null directions that are transverse to that geodesic. This property plays a very important role for nonlinear equations with Klainerman's "null condition" [28]. Another way of describing it is that $\hat{\phi}$ has a Taylor expansion in v near $\mathscr{I}^+$ at any order, this gives an asymptotic expansion into powers of $1/r$ for ϕ at any order along outgoing null geodesics. However this is a local description of the peeling, related to the asymptotic behaviour of the solution along a given outgoing null geodesic. This type of pointwise decay only contains a part of the information contained in the initial data; a loss is inevitable because pointwise decay estimates are obtained from energy estimates via Sobolev embeddings, that involve a loss of regularity. A more interesting description of the control of the transverse regularity of the solution at the conformal boundary in terms of functions spaces of initial data is provided in Section 7.5 by the energy equivalence (7.20). See the discussion in that section.

7.4 Conformal scattering using the complete compactification

We work with the compactification of Minkowski space-time described in Section 3.1.1. We consider the stress energy tensor for equation (7.1)

$$T_{ab} = T_{(ab)} = \partial_a\hat{\phi}\partial_b\hat{\phi} - \frac{1}{2}\mathfrak{e}_{ab}\mathfrak{e}^{cd}\partial_c\hat{\phi}\partial_d\hat{\phi} + \frac{1}{2}\hat{\phi}^2\mathfrak{e}_{ab} \tag{7.14}$$

and contract it with the Killing vector field $K = \partial_\tau$

$$J^a = K^b T_b^a .$$

This yields the conservation law

$$\nabla_a J^a = 0. \tag{7.15}$$

To calculate the flux of J across a hypersurface X_τ, we choose K for the normal vector field to X_τ as well as for the transverse vector field. We have

$$\begin{aligned} E_{X_\tau}(\phi) &= \int_{S^3} J_a K^a K \lrcorner \mathrm{dVol} = \int_{S^3} T_{00} K \lrcorner \mathrm{dVol} \\ &= \frac{1}{2}\int_{X_\tau}\left(\hat{\phi}_\tau^2 + \left|\nabla_{S^3}\hat{\phi}\right|^2 + \hat{\phi}^2\right)\mathrm{d}\mu_{S^3}. \end{aligned}$$

and for $\mathscr{I}^+$ we take $\partial_\tau - \partial_\zeta$ as the normal vector field and $\frac{1}{2}(\partial_\tau + \partial_\zeta)$ as the transverse vector field and, parametrising $\mathscr{I}^+$ as $\tau = \pi - \zeta$ over S^3, we get,

$$\begin{aligned} \mathcal{E}_{\mathscr{I}^+}(\hat{\phi}) &= \frac{1}{\sqrt{2}}\int_{\mathscr{I}^+}\left(-2\hat{\phi}_\tau\hat{\phi}_\zeta + \hat{\phi}_\tau^2 + \left|\nabla_{S^3}\hat{\phi}\right|^2 + \hat{\phi}^2\right)\mathrm{d}\mu_{S^3} \\ &= \frac{1}{\sqrt{2}}\int_{\mathscr{I}^+}\left(\left|\hat{\phi}_\tau - \hat{\phi}_\zeta\right|^2 + \frac{1}{\sin^2\zeta}\left|\nabla_{S^2}\hat{\phi}\right|^2 + \hat{\phi}^2\right)\mathrm{d}\mu_{S^3}\,. \end{aligned}$$

This is a natural H^1 norm of $\hat{\phi}$ on $\mathscr{I}^+$, involving only the tangential derivatives of $\hat{\phi}$ along $\mathscr{I}^+$.

Now consider a smooth solution $\hat{\phi}$ of (7.1). The conservation law (7.15) tells us in particular that the outgoing flux of J through the closed hypersurface made of the union of X_0 and $\mathscr{I}^+$ is zero, hence

$$\mathcal{E}_{\mathscr{I}^+}(\hat{\phi}) = \mathcal{E}_{X_0}(\hat{\phi}). \tag{7.16}$$

As a particular case of Theorem 4.11, we have the following theorem, with an additional property of the trace operators given by the energy equality (7.16).

Theorem 7.3. *The trace operators*

$$\mathbb{T}^{\pm} \,:\, (\hat{\phi}_0\,,\ \hat{\phi}_1) \longmapsto \hat{\phi}|_{\mathscr{I}^{\pm}}\,,$$

defined from $(\mathscr{C}^\infty(X_0))^2$ *to* $\mathscr{C}^\infty(\mathscr{I}^{\pm})$*, extend uniquely as isometries from* $H^1(X_0)\times L^2(X_0)$ *onto* $H^1(\mathscr{I}^{\pm})$ *equipped with the energy norms on* X_0 *and* $\mathscr{I}^{\pm}$*. The scattering operator, given by*

$$S := \mathbb{T}^+(\mathbb{T}^-)^{-1},$$

is an isometry from $H^1(\mathscr{I}^-)$ *onto* $H^1(\mathscr{I}^-)$*.*

The result of the theorem can be interpreted as follows. Given initial data ϕ_0 and ϕ_1 for equation (4.5) such that the corresponding rescaled data $\hat{\phi}_0$ and $\hat{\phi}_1$ extend as smooth functions on S^3 respectively, the corresponding solution ϕ divided by Ω has a limit along null geodesics, both in the past and future. These limits define functions on $\mathscr{I}^{\pm}$ that each completely characterises ϕ: they are the future and past scattering data. This extends to initial data such that $(\hat{\phi}_0, \hat{\phi}_1) \in H^1(S^3)\times L^2(S^3)$ and the corresponding scattering data span $H^1(\mathscr{I}^{\pm})$.

7.5 Higher order scattering

Let us consider $\hat{\phi}$ a smooth solution of (7.1). Since ∂_τ is a Killing vector, for any $k\in\mathbb{N}$, $\partial_\tau^k\hat{\phi}$ satisfies equation (7.1) and therefore also (7.16), i.e.

$$\mathcal{E}_{\mathscr{I}^+}(\partial_\tau^k\hat{\phi}) = \mathcal{E}_{X_0}(\partial_\tau^k\hat{\phi}). \tag{7.17}$$

For $k=2p$, we have

$$\begin{aligned} \mathcal{E}_{X_0}(\partial_\tau^k\hat{\phi}) &= \|\partial_\tau^{2p}\hat{\phi}\|^2_{H^1(X_0)} + \|\partial_\tau^{2p+1}\hat{\phi}\|^2_{L^2(X_0)} \\ &= \|(1-\Delta_{S^3})^p\hat{\phi}\|^2_{H^1(X_0)} + \|(1-\Delta_{S^3})^p\partial_\tau\hat{\phi}\|^2_{L^2(X_0)} \\ &\simeq \|\hat{\phi}\|^2_{H^{2p+1}(X_0)} + \|\partial_\tau\hat{\phi}\|^2_{H^{2p}(X_0)}\,, \end{aligned} \tag{7.18}$$

and for $k=2p+1$,

$$\begin{aligned} \mathcal{E}_{X_0}(\partial_\tau^k\hat{\phi}) &= \|\partial_\tau^{2p+1}\hat{\phi}\|^2_{H^1(X_0)} + \|\partial_\tau^{2p+2}\hat{\phi}\|^2_{L^2(X_0)} \\ &= \|(1-\Delta_{S^3})^p\partial_\tau\hat{\phi}\|^2_{H^1(X_0)} + \|(1-\Delta_{S^3})^{p+1}\hat{\phi}\|^2_{L^2(X_0)} \\ &\simeq \|\hat{\phi}\|^2_{H^{2p+2}(X_0)} + \|\partial_\tau\hat{\phi}\|^2_{H^{2p+1}(X_0)}\,. \end{aligned} \tag{7.19}$$

Hence, we have for each $k\in\mathbb{N}$:

$$\|\hat{\phi}\|^2_{H^{k+1}(X_0)} + \|\partial_\tau\hat{\phi}\|^2_{H^k(X_0)} \simeq \mathcal{E}_{X_0}(\partial_\tau^k\hat{\phi}) = \mathcal{E}_{\mathscr{I}^+}(\partial_\tau^k\hat{\phi}) \simeq \|\partial_\tau^k\hat{\phi}\|^2_{H^1(\mathscr{I}^+)}$$

and using the fact that the H^k norm controls all the lower Sobolev norms, this gives us the equivalence

$$\|\hat{\phi}\|^2_{H^{k+1}(X_0)} + \|\partial_\tau\hat{\phi}\|^2_{H^k(X_0)} \simeq \sum_{p=0}^{k}\|\partial_\tau^p\hat{\phi}\|^2_{H^1(\mathscr{I}^+)}\,. \tag{7.20}$$

This indicates that one can establish higher order scattering theories encoding the H^k propagation of more regular solutions. This is done by restricting the trace operators to the higher order energy spaces. There is however a slight difficulty: the corresponding energy spaces on $\mathscr{I}^\pm$ are defined using a derivative along ∂_τ that is transverse to $\mathscr{I}^\pm$. In order to understand the higher order energy spaces on $\mathscr{I}^\pm$, one must use the restriction of the equation to $\mathscr{I}^\pm$ to express the transverse derivative in terms of integrals along the null generators of tangential derivatives. This type of construction has been done in a paper by D. Häfner and J.-P. Nicolas [22] for the Goursat problem on a finite light-cone for Dirac's equation. Let us do the explicit calculation for a first order energy. We start from equality (7.17) for $k = 1$

$$\mathcal{E}_{\mathscr{I}^+}(\partial_\tau \hat{\phi}) = \mathcal{E}_{X_0}(\partial_\tau \hat{\phi}). \tag{7.21}$$

As we have seen in the previous section,

$$\mathcal{E}_{X_0}(\partial_\tau \hat{\phi}) \simeq \|\hat{\phi}\|^2_{H^2(X_0)} + \|\partial_\tau \hat{\phi}\|^2_{L^2(X_0)}.$$

We now evaluate $\mathcal{E}_{\mathscr{I}^+}(\partial_\tau \hat{\phi})$ for solutions of (7.1) that come from smooth compactly supported physical initial data. We do the construction for $\mathscr{I}^-$. Equation (7.1) can be written as

$$\partial_\tau^2 \hat{\phi} - \partial_\zeta^2 \hat{\phi} - \frac{1}{\sin^2 \zeta} \Delta_{S^2} \hat{\phi} + \hat{\phi} = 0,$$

i.e.

$$(\partial_\tau + \partial_\zeta) \left(\partial_\tau \hat{\phi} - \partial_\zeta \hat{\phi} \right) = \frac{1}{\sin^2 \zeta} \Delta_{S^2} \hat{\phi} - \hat{\phi}.$$

Since the physical initial data are compactly supported, the restriction of $\hat{\phi}$ to $\mathscr{I}^-$ vanishes in the neighbourhoods of i^- and i^0, whence, at any point $(\zeta_0 - \pi, \zeta_0, \omega_0)$ of $\mathscr{I}^-$, we have

$$\left(\partial_\tau \hat{\phi} - \partial_\zeta \hat{\phi} \right) (\zeta_0 - \pi, \zeta_0, \omega_0) = \sqrt{2} \int_0^{\zeta_0} \left(\frac{1}{\sin^2 \zeta} \Delta_{S^2} \hat{\phi}(\zeta - \pi, \zeta, \omega_0) - \hat{\phi}(\zeta - \pi, \zeta, \omega_0) \right) \sin^2 \zeta \mathrm{d}\zeta \mathrm{d}\omega.$$

This provides an expression for $\mathcal{E}_{\mathscr{I}^+}(\partial_\tau \hat{\phi})$ that involves only derivatives along directions tangent to $\mathscr{I}^-$. This is a norm on functions on $\mathscr{I}^-$ in which we can take the completion of the space of smooth functions supported away from i^- and i^0. This defines a function space of first order scattering data (we can also use $\sqrt{\mathcal{E}_{\mathscr{I}^+}(\hat{\phi}) + \mathcal{E}_{\mathscr{I}^+}(\partial_\tau \hat{\phi})}$ as an alternative choice of norm). The norm in this space is a little cumbersome, with a control on the regularity that is not particularly intuitive.

However, what is important to observe is that the energy equivalence (7.20) provides a complete description of the peeling at any order. Moreover, this description is optimal in the sense that for any given order of transverse regularity (choice of k), we have the exact class of initial data that entails this order of peeling for the rescaled solution, without loss of information. The thing that can be improved in this description is that it is based on the full compactification of Minkowski space-time. This does not generalise to meaningful asymptotically flat Einstein space-times and imposes some rather strong decay at spacelike infinity on the data in order to ensure regularity of the solution on the Einstein cylinder. A complete description of the peeling on Minkowski and Schwarzschild space-times is given in Section 9.5, based on the compactification with conformal factor $\Omega = 1/r$.

7.6 A stronger conformal scattering construction

Instead of using the full conformal embedding of Minkowski space-time into the Einstein cylinder, we now work with the partial compactification described in Section 3.1.2. The metric $\tilde{\eta}$ is also scalar-flat (see Equation (3.22)), so a scalar field ϕ on $\mathbb{M}$ satisfies the wave equation

$$\Box_\eta \phi = 0$$

if and only if the rescaled field

$$\tilde{\phi} = R^{-1}\phi = r\phi$$

satisfies the wave equation on $(\mathbb{M}, \tilde{\eta})$

$$\Box_{\tilde{\eta}} \tilde{\phi} = 0 \,. \tag{7.22}$$

In terms of coordinates (t, r, ω), this takes the form

$$r^2 \frac{\partial^2 \tilde{\phi}}{\partial t^2} - r^2 \frac{\partial^2 \tilde{\phi}}{\partial r^2} - \Delta_{S^2}\tilde{\phi} = 0 \,.$$

Equation (7.22) has a stress-energy tensor

$$\begin{aligned}
\tilde{T}_{ab} &= \nabla_a \tilde{\phi} \nabla_b \tilde{\phi} - \frac{1}{2}\left(\tilde{\eta}^{cd} \nabla_c \tilde{\phi} \nabla_d \tilde{\phi} \right) \tilde{\eta}_{ab} \,, \\
&= \nabla_a \tilde{\phi} \nabla_b \tilde{\phi} - \frac{1}{2}\left(r^2 |\partial_t \tilde{\phi}|^2 - r^2 |\partial_r \tilde{\phi}|^2 - |\nabla_{S^2} \tilde{\phi}|^2 \right) \tilde{\eta}_{ab} \,, \\
&= \nabla_a \tilde{\phi} \nabla_b \tilde{\phi} - \frac{1}{2}\left(-R^2 |\partial_R \tilde{\phi}|^2 - 2\partial_u \tilde{\phi} \partial_R \tilde{\phi} - |\nabla_{S^2} \tilde{\phi}|^2 \right) \tilde{\eta}_{ab} \,, \\
&= \nabla_a \tilde{\phi} \nabla_b \tilde{\phi} - \frac{1}{2}\left(-R^2 |\partial_R \tilde{\phi}|^2 + 2\partial_v \tilde{\phi} \partial_R \tilde{\phi} - |\nabla_{S^2} \tilde{\phi}|^2 \right) \tilde{\eta}_{ab} \,,
\end{aligned}$$

that is conserved, i.e.

$$\tilde{\nabla}^a \tilde{T}_{ab} = \Box_{\tilde{\eta}} \partial_b \tilde{\phi} = 0 \text{ if } \tilde{\phi} \text{ satisfies (7.22).}$$

Moreover, we have a future-oriented causal Killing vector field $T = \partial_t$ or ∂_u or ∂_v respectively in the coordinate systems (t, r, ω), (u, R, ω) and (v, R, ω), that is timelike in the interior of $\tilde{\mathbb{M}}$ and null on the boundary (see Section 3.1.2). We use it to define a conserved energy current

$$\tilde{J}_a := T^b \tilde{T}_{ab} \,, \; \tilde{\nabla}^a \tilde{J}_a = 0 \,. \tag{7.23}$$

The energy fluxes on the $\{t = 0\}$ hypersurface, denoted by Σ_0, and on $\mathscr{I}^{\pm}$, can be calculated as explained in the divergence Theorem 2.5 by choosing $n = l = r\partial_t$ for Σ_0, $n = \partial_u$, $l = -\partial_R$ for $\mathscr{I}^+$ and $n = \partial_v$, $l = \partial_R$ for $\mathscr{I}^-$. They are given by

$$\tilde{\mathcal{E}}_{\Sigma_0}(\tilde{\phi}) = \frac{1}{2} \int_{\mathbb{R}_r^+ \times S^2} \left(\left(\frac{\partial \tilde{\phi}}{\partial t} \right)^2 + \left(\frac{\partial \tilde{\phi}}{\partial r} \right)^2 + \frac{1}{r^2} \left| \nabla_{S^2} \tilde{\phi} \right|^2 \right) \mathrm{d}r \mathrm{d}\omega \,, \tag{7.24}$$

$$\tilde{\mathcal{E}}_{\mathscr{I}^+}(\tilde{\phi}) = \int_{\mathbb{R}_u \times S^2} \left(\frac{\partial \tilde{\phi}}{\partial u} \right)^2 \mathrm{d}u \mathrm{d}\omega \,, \tag{7.25}$$

$$\tilde{\mathcal{E}}_{\mathscr{I}^-}(\tilde{\phi}) = \int_{\mathbb{R}_v \times S^2} \left(\frac{\partial \tilde{\phi}}{\partial v} \right)^2 \mathrm{d}v \mathrm{d}\omega \,, \tag{7.26}$$

where

$$\left|\nabla_{S^2}\tilde{\phi}\right|^2 = \left(\frac{\partial\tilde{\phi}}{\partial\theta}\right)^2 + \frac{1}{\sin^2\theta}\left(\frac{\partial\tilde{\phi}}{\partial\varphi}\right)^2 \,,\ \mathrm{d}\omega = \sin\theta\,\mathrm{d}\theta\mathrm{d}\varphi\,.$$

Remark 7.3. *Note that the energy (7.24) may be quite close to the physical energy of the initial data, associated with the standard stress-energy tensor for the scalar wave equation on the Minkowski metric and the timelike Killing vector ∂_t*

$$\mathcal{E}_{\Sigma_0}(\phi) = \frac{1}{2}\int_{\mathbb{R}^+_r\times S^2}\left(\left(\frac{\partial\phi}{\partial t}\right)^2 + \left(\frac{\partial\phi}{\partial r}\right)^2 + \frac{1}{r^2}\left|\nabla_{S^2}\phi\right|^2\right) r^2\mathrm{d}r\mathrm{d}\omega\,,$$

but it differs from it by two terms,

$$\begin{aligned}\tilde{\mathcal{E}}_{\Sigma_0}(\tilde{\phi}) - \mathcal{E}_{\Sigma_0}(\phi) &= \frac{1}{2}\int_{\mathbb{R}^+_r\times S^2}\left(\left(\frac{\partial(r\phi)}{\partial r}\right)^2 - r^2\left(\frac{\partial\phi}{\partial r}\right)^2\right)\mathrm{d}r\mathrm{d}\omega\,,\\ &= \frac{1}{2}\int_{\mathbb{R}^+_r\times S^2}\left(\phi^2 + 2r\phi\frac{\partial\phi}{\partial r}\right)\mathrm{d}r\mathrm{d}\omega\,,\end{aligned}$$

the first one being proportional to the squared L^2 norm of ϕ on Σ_0, which is controlled by neither $\tilde{\mathcal{E}}_{\Sigma_0}(\tilde{\phi})$ nor $\mathcal{E}_{\Sigma_0}(\phi)$.

The energies (7.24)-(7.26) above define function spaces $\tilde{\mathcal{H}}_{\Sigma_0}$ and $\tilde{\mathcal{H}}_{\mathscr{I}^\pm}$ by completion of $\mathscr{C}_0^\infty(\mathbb{R}^3)\times\mathscr{C}_0^\infty(\mathbb{R}^3)$ and $\mathscr{C}_0^\infty(\mathbb{R}\times S^2)$ respectively, in the associated norms associated. These spaces are in fact standard homogeneous Sobolev spaces

$$\tilde{\mathcal{H}}_{\Sigma_0} \simeq \dot{H}^1(\mathbb{R}^3\,,\ \mathrm{d}r\mathrm{d}\omega)\times L^2(\mathbb{R}^3\,,\ \mathrm{d}r\mathrm{d}\omega)\,,\ \tilde{\mathcal{H}}_{\mathscr{I}^\pm} \simeq \dot{H}^1(\mathbb{R}\,;\ L^2(S^2))\,.$$

We see that we are dealing with a larger class of functions than when using the full compactification, hence we cannot apply here the results of Theorem 7.3. Moreover, since we are not on a spatially compact space-time, we cannot use directly the results of Theorem 4.11, we need to adapt the proof to our situation.

Remark 7.4. *As we have seen in Remark 4.7, homogeneous Sobolev spaces are delicate objects in low dimensions. The space $\dot{H}^k(\mathbb{R}^n)$ is defined as the completion of $\mathscr{C}_0^\infty(\mathbb{R}^n)$ in the norm*

$$\sqrt{\int_{\mathbb{R}^n}|\nabla f(x)|^2\mathrm{d}x}\,.$$

For $n=1$ and $n=2$ this is not a distribution space, this is because constants are not ruled out in the completion process and so in effect the completed space is a space of equivalence classes modulo constants. This is no longer the case for $n\geq 3$. In the case of $\tilde{\mathcal{H}}_{\mathscr{I}^\pm}$, we are dealing with $n=1$. As a consequence, many people prefer to define the scattering data (or the radiation field) as the restriction to $\mathscr{I}$ of $\partial_t\tilde{\phi}$, which is an honest function living in $L^2(\mathbb{R}\times S^2)$. In fact the two points of view are equivalent since the norm in $\tilde{\mathcal{H}}_{\mathscr{I}^\pm}$ of the restriction of $\tilde{\phi}$ to $\mathscr{I}^\pm$, for solutions for which this restriction is well defined, is exactly the norm in $L^2(\mathbb{R}\times S^2)$ of the restriction to

$\mathscr{I}^\pm$ of $\partial_t\tilde{\phi}$. Whittaker's formula (4.13) and its relation to the Lax-Phillips theory (see Chapter 5) give an additional reason for considering the restriction of $\partial_t\tilde{\phi}$ as the scattering data. However, for consistency with higher spin zero rest-mass field equations as well as for reasons of analogy with analytic time-dependent scattering theories, we have chosen to define the scattering data as the restrictions of $\tilde{\phi}$ itself.

With our choices of compactification, stress-energy tensor and observer, we can construct the following conformal scattering theory.

Theorem 7.4. *Let $T^\pm$ be the linear operators that to $(\tilde{\phi}_0, \tilde{\phi}_1) \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$ associate the restrictions of $\tilde{\phi}$ to $\mathscr{I}^\pm$, where $\tilde{\phi}$ is the solution in $\mathscr{C}^\infty(\mathbb{M})$ to the Cauchy problem*

$$\begin{cases} \square_{\tilde{\eta}}\tilde{\phi} = 0 \text{ on } \mathbb{M}\,, \\ \tilde{\phi}|_{\Sigma_0} = \tilde{\phi}_0\,,\ \partial_t\tilde{\phi}|_{\Sigma_0} = \tilde{\phi}_1\,. \end{cases}$$

Then $T^\pm$ take their values in $\tilde{\mathcal{H}}_{\mathscr{I}^\pm}$ and extend as isometries from $\tilde{\mathcal{H}}_{\Sigma_0}$ onto $\tilde{\mathcal{H}}_{\mathscr{I}^\pm}$. The scattering operator for the theory is then defined as

$$S := T^+(T^-)^{-1}\,.$$

Proof. A feature of the partial compactification of Section 3.1.2 is that spacelike and timelike infinities remain at infinity. This means that at least at first sight, energy estimates cannot be obtained by a direct application of Stokes' Theorem, since we shall not work with a compact hypersurface. However we only need to establish estimates for dense subclasses of data and we can restrict ourselves to smooth compactly supported functions. For the Cauchy problem, this is enough to ensure that Stokes's Theorem can be used. For the Goursat problem, we need to be a little more careful.

1. *The operators $T^\pm$ are partial isometries.* Let $(\phi_0, \phi_1) \in \mathscr{C}_0^\infty(\mathbb{R}^3) \times \mathscr{C}_0^\infty(\mathbb{R}^3)$ and $\tilde{\phi} \in \mathscr{C}^\infty(\mathbb{M})$ be the corresponding solution to the Cauchy problem. Our aim is to perform an energy estimate on the (neither closed nor bounded) hypersurface $\mathcal{S}$ that is the reunion of Σ_0 and $\mathscr{I}^+$. By finite propagation speed, ϕ vanishes identically in the neighbourhood of i^0, which means that in the spacelike directions, the fact that $\mathcal{S}$ is unbounded does not prevent us from using Stokes' Theorem. But of course the Huygens principle entails that the solution also vanishes in the neighbourhood of i^+, so we can in fact apply Stoke's Theorem on the hypersurface $\mathscr{I}^+ \cup \Sigma_0$ as if it were a closed bounded hypersurface and we obtain

   $$\tilde{\mathcal{E}}_{\mathscr{I}^+}(\tilde{\phi}) = \tilde{\mathcal{E}}_{\Sigma_0}(\tilde{\phi})\,.$$

   The same construction can be done in the past. Therefore the operators $T^\pm$ extend as bounded one-to-one operators with closed range between $\tilde{\mathcal{H}}_{\Sigma_0}$ and $\tilde{\mathcal{H}}_{\mathscr{I}^\pm}$.

2. *The operators $T^\pm$ are full isometries.* All we need to do is to prove that the range of $T^\pm$ is dense in $\tilde{\mathcal{H}}(\mathscr{I}^\pm)$. We write the proof for T^+, it is identical for T^-. Let us consider $\tilde{\phi}^+ \in \mathscr{C}_0^\infty(\mathbb{R} \times S^2)$ as characteristic data on $\mathscr{I}^+$. Let us consider a lightcone C_{T_1} that cuts $\mathscr{I}^+$ to the future of the support of $\tilde{\phi}^+$ and another C_{T_2} with $T_2 << -1$. We consider the part Ω_{T_1,T_2} of partially compactified Minkowski space-time that lies in the past of C_{T_1}

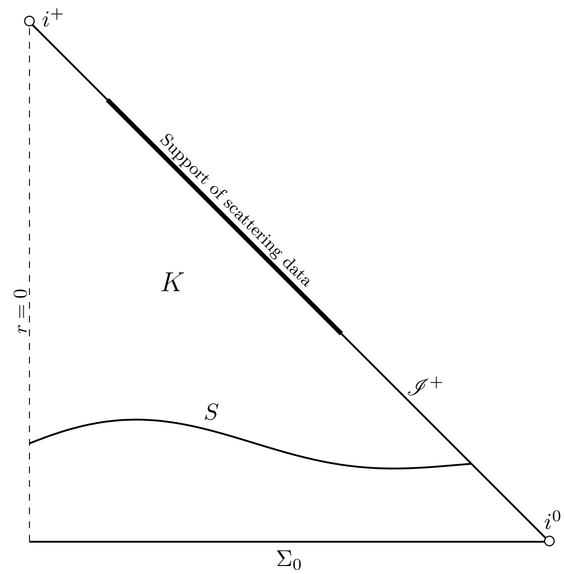

Figure 7.1: Representation of the intermediate surface S and the domain K. Note that i^0 and i^+ are represented as circles to signify that they are not part of the rescaled space-time since we are using the incomplete compactification. This is also why the domain K is not compact.

and in the future of both Σ_0 and C_{T_2}. Then we extend it as a smooth spatially compact space-time with manifold, say, diffeomorphic to $\mathbb{R} \times S^3$. Hörmander's Theorem for the Goursat problem entails that there exists a unique smooth solution $\tilde{\phi}$ to the wave equation on Ω_{T_1,T_2} whose restriction to $\mathscr{I}^+$ is equal to $\tilde{\phi}^+$. Since this is true for any T_1 and T_2 such as above, this is also true in the complete future of Σ_0. It now remains to establish that this solution has finite energy on Σ_0. To this purpose, let us consider a spacelike hypersurface S located entirely in the future of Σ_0 that cuts $\mathscr{I}^+$ to the past of the support of $\tilde{\phi}^+$. Let us denote by K the domain bounded by $\mathscr{I}^+$ and S (see Figure 7.1); this is not a compact set but since $\tilde{\phi}$ vanishes in the neighbourhood of i^+, $\operatorname{supp}\tilde{\phi} \cap K$ is compact and we can apply the divergence Theorem on K. This entails

$$\tilde{\mathcal{E}}_{\mathscr{I}^+}(\tilde{\phi}) = \tilde{\mathcal{E}}_S(\tilde{\phi})\,.$$

We observe that $\tilde{\phi}$ is smooth on K and its restriction to $\mathscr{I}^+ \cap S$ vanishes. Hence, we have

$$\tilde{\phi}|_S \in H^1_0(S)\,.$$

It follows that $\tilde{\phi}$ can be approached in $H^1(S)$ by a sequence $(\tilde{\phi}^n_S)$ of smooth and compactly supported functions. Moreover, if we consider ν^a the future oriented unit normal vector field to S, $\tilde{\nabla}_\nu\tilde{\phi}$ is in $L^2(S)$ and can be approached in $L^2(S)$ by a sequence $(\tilde{\phi}^n_{S,1})$ of smooth compactly supported functions. Let us denote by $\tilde{\phi}^n$ the solution to the wave equation

$$\Box_{\tilde{\eta}}\tilde{\phi}^n = 0 \text{ on } \mathbb{M}$$

such that

$$\tilde{\phi}^n|_S = \tilde{\phi}^n_S,\ \tilde{\nabla}_\nu \tilde{\phi}^n|_S = \tilde{\phi}^n_{S,1}$$

and $\tilde{\phi}^n \equiv 0$ in the neighbourhood of $\mathscr{I}^+ \cup i^0$ in the past of S. Since

$$\tilde{\phi}^n_S \to \tilde{\phi}|_S \text{ in } H^1(S) \text{ and } \tilde{\phi}^n_{S,1} \to \tilde{\nabla}_\nu \tilde{\phi}|_S \text{ in } L^2(S)\,,$$

we have that

$$\tilde{\mathcal{E}}_S(\tilde{\phi}^n - \tilde{\phi}) \to 0 \text{ as } n \to +\infty\,.$$

In other words, $(\tilde{\phi}^n)$ is a Cauchy sequence in energy norm on S and using the conservation of the energy, it follows that it is also a Cauchy sequence in energy norm on Σ_0, which entails that the energy of $\tilde{\phi}$ on Σ_0 is finite. Hence

$$\left(\tilde{\phi}|_{\Sigma_0}, \partial_t \tilde{\phi}|_{\Sigma_0}\right) \in \tilde{\mathcal{H}}_{\Sigma_0}$$

and

$$\tilde{\phi}^+ = T^+\left(\tilde{\phi}|_{\Sigma_0}, \partial_t \tilde{\phi}|_{\Sigma_0}\right)\,.$$

This concludes the proof of the theorem.

□

7.7 Perturbations of the wave equation

The conformal scattering method can be extended to non conformally invariant equations provided the lack of conformal invariance is controllable via energy estimates. A simple example is the equation

$$\square_\eta \phi + L\phi = 0, \tag{7.27}$$

on Minkowski space-time, where L is a first order differential operator with smooth coefficients depending on both time and space variables. We can try to express (7.27) as an equation on the rescaled unknown $\hat{\phi} = \Omega^{-1}\phi$ involving $\square_{\hat{\eta}}$. Using Corollary 4.3 and the fact that the scalar curvature of $\hat{\eta}$ is 6, we can write (7.27) as

$$\Omega^3\left(\square_{\hat{\eta}}\hat{\phi} + \hat{\phi}\right) + \Omega L\hat{\phi} + [L,\Omega]\hat{\phi} = 0,$$

or equivalently

$$\square_{\hat{\eta}}\hat{\phi} + \hat{\phi} + \Omega^{-2}L\hat{\phi} + \Omega^{-3}[L,\Omega]\hat{\phi} = 0. \tag{7.28}$$

If the coefficients of L, the differential operator being expressed using the coordinates $(\tau,\zeta,\theta,\varphi)$, decay sufficiently fast at the conformal boundary, namely as Ω^3 for the first order terms and as Ω^2 for the zero order term, then the operator $\Omega^{-2}L\hat{\phi} + \Omega^{-3}[L,\Omega]\hat{\phi}$ extends as a differential operator with smooth coefficients on the whole Einstein cylinder and we can construct a conformal scattering theory for (7.27) using exactly the same method as for the wave equation.

I STOPPED HERE (08/04/2025). ANALYTIC SCATTERING NOT REVIEWED YET.

Chapter 8

Some transversal constructions

8.1 A striking feature of scattering on Minkowski space-time

Let us consider massless scalar fields to start with. Let ϕ be a solution to the scalar wave equation on Minkowski space-time associated with smooth compactly supported initial data at $t = 0$. Let

$$\hat{\phi} = \Omega^{-1}\phi \text{ with } \Omega = \left(\frac{4}{(1+(t+r)^2)(1+(t-r)^2)} \right)^{1/2},$$

be the rescaled field satisfying the Klein-Gordon equation with mass 1 on the Einstein cylinder (see Subsection 3.1.1)

$$\partial_\tau^2 \hat{\phi} - \Delta_{S^3} \hat{\phi} + \hat{\phi} = 0\,. \tag{8.1}$$

We can decompose $\hat{\phi}$ on spherical harmonics on $\mathbb{S}^3$. The eigenvalues of the spherical laplacian on $\mathbb{S}^3$ are given by $l(l+2)$, $l \in \mathbb{N}$ and Equation (8.1) for solutions with angular momentum $l(l+2)$ becomes the ordinary differential equation

$$\partial_\tau^2 \hat{\phi} + (l+1)^2 \hat{\phi} = 0\,,$$

the solutions of which are given by

$$\hat{\phi} = \hat{\phi}_+ e^{i(l+1)\tau} + \hat{\phi}_- e^{-i(l+1)\tau}\,.$$

This shows in particular that $\hat{\phi}$ is 2π-periodic in τ. However, the refocusing of light-cones from $\mathscr{I}^-$ to $\mathscr{I}^+$ described in Theorem 3.1, has stronger consequences.

Theorem 8.1. *The scattering operator obtained in the previous two sections is in fact minus the identity composed with the antipodal map. More explicitly in the partial compactification picture, for a finite rescaled energy field, the scattering data satisfy*

$$\tilde{\phi}^+(u,\omega) = -\tilde{\phi}^-(v,-\omega)\,. \tag{8.2}$$

This property is established in Spinors and space-time [45] Vol. 2, using the Grgin index of zero rest-mass fields. We propose here two other proofs. The first one uses the geometrical reinterpretation of the translation representer in the Lax-Phillips theory and the second takes advantage of the symmetries of the spherical harmonics on the 3-sphere.

Proof of Theorem 8.1 using the Lax-Phillips theory. Let us assume that the data for the physical field ϕ are smooth and compactly supported. Given $s \in \mathbb{R}$ and $\omega \in \mathbb{S}^2$, the curve

$$\gamma(r) = (r - s, r\omega)$$

extends for $r < 0$ as (putting $\rho = -r$)

$$\tilde{\gamma}(\rho) = (-(\rho + s), \rho(-\omega)) .$$

In Theorem 5.2, we have established equation (5.19) that allows to reinterpret the translation representer $k(s, \omega)$ of our field as the limit along γ as $r \to +\infty$ of $-r\partial_t\phi$. Let us now study the limit

$$\lim_{\rho \to +\infty} \rho \partial_t \phi(-(\rho + s), \rho(-\omega)) ,$$

using a localisation near the direction $-\omega$. This is essentially identical to the proof of Theorem 5.2 and we therefore omit some of the details. Recall that k is compactly supported, there exists $R > 0$ such that $\operatorname{supp} k \subset [-R, R] \times S^2$. We have

$$\phi(t, x) = \frac{1}{2\pi} \int_{S^2} k(x.\zeta - t, \zeta) \mathrm{d}^2\zeta$$

and since k is $\mathcal{C}^1$ and compactly supported, we can differentiate under the integral

$$\partial_t \phi(t, x) = -\frac{1}{2\pi} \int_{S^2} \partial_s k(x.\zeta - t, \zeta) \mathrm{d}^2\zeta .$$

In particular, we have

$$\begin{aligned} \partial_t \phi(-(\rho + s), \rho(-\omega)) &= -\frac{1}{2\pi} \int_{S^2} (\partial_s k)(-\rho\omega.\zeta + \rho + s, \zeta) \mathrm{d}^2\zeta \\ &= -\frac{1}{2\pi} \int_{S^2} (\partial_s k)(\rho(1 - \omega.\zeta) + s, \zeta) \mathrm{d}^2\zeta . \end{aligned}$$

We now define a neighbourhood of the direction ω that becomes small as ρ becomes large,

$$V_{Rs\rho} = \left\{ \zeta \in S^2 ;\ 1 - \omega.\zeta \leq \frac{R + |s|}{\rho} \right\} ,$$

we have :

$$\begin{aligned} \partial_t \phi(-(\rho + s), \rho(-\omega)) &= -\frac{1}{2\pi} \int_{V_{Rs\rho}} (\partial_s k)(\rho(1 - \omega.\zeta) + s, \zeta) \mathrm{d}^2\zeta \\ &\quad -\frac{1}{2\pi} \int_{V_{Rs\rho}^C} (\partial_s k)(\rho(1 - \omega.\zeta) + s, \zeta) \mathrm{d}^2\zeta . \end{aligned}$$

When $\zeta \in V_{Rsr}^C$,

$$\frac{R + |s|}{\rho} < 1 - \omega.\zeta = |1 - \omega.\zeta| ,$$

so

$$R < \rho|1 - \omega.\zeta| - |s| \le |r(1 - \omega.\zeta) + s| \, .$$

The second integral is therefore zero and

$$\partial_t\phi(-(\rho + s), \rho(-\omega)) = -\frac{1}{2\pi}\int_{V_{Rs\rho}} (\partial_s k)(\rho(1 - \omega.\zeta) + s, \zeta)\mathrm{d}^2\zeta \, .$$

Adding and subtracting a term where ζ is replaced by ω, we obtain

$$\begin{aligned} \partial_t\phi(-(\rho + s), \rho(-\omega)) = & -\frac{1}{2\pi}\int_{V_{Rs\rho}} \left[(\partial_s k)(\rho(1 - \omega.\zeta) + s, \zeta) - (\partial_s k)(\rho(1 - \omega.\zeta) + s, \omega)\right] \mathrm{d}^2\zeta \\ & -\frac{1}{2\pi}\int_{V_{Rs\rho}} (\partial_s k)(\rho(1 - \omega.\zeta) + s, \omega)\mathrm{d}^2\zeta \, . \end{aligned} \tag{8.3}$$

The first term is of the form $\varepsilon(\rho)/\rho$, where $\varepsilon(\rho)$ tends to zero as $\rho \to +\infty$. We calculate explicitly the second term denoted by I, using spherical coordinates (θ, φ) based on the direction ω. Putting

$$\theta_\rho = \arccos\left(1 - \frac{R + |s|}{\rho}\right) ,$$

we have

$$\begin{aligned} I &= -\frac{1}{2\pi}\int_0^{2\pi}\int_0^{\theta_\rho} (\partial_s k)(\rho(1 - \cos\theta) + s, \omega)\sin\theta\mathrm{d}\theta\mathrm{d}\varphi \\ &= -\frac{1}{\rho}\int_0^{\theta_\rho}\left(\frac{\mathrm{d}}{\mathrm{d}\theta}\left(k(\rho(1 - \cos\theta) + s, \omega)\right)\right)\mathrm{d}\theta \\ &= -\frac{1}{\rho}\left(k(R + |s| + s, \omega) - k(s, \omega)\right) = \frac{1}{\rho}k(s, \omega) \, . \end{aligned}$$

Hence

$$\partial_t\phi(-(\rho + s), \rho(-\omega)) = \frac{1}{\rho}\varepsilon(\rho) + \frac{1}{\rho}k(s, \omega) \, .$$

and

$$k(s, \omega) = \lim_{\rho\to+\infty} \rho\partial_t\phi(-(\rho + s), \rho(-\omega)) \, . \tag{8.4}$$

The future and past scattering data for the partial compactification are defined as the restrictions of $\tilde{\phi} = r\phi$ to $\mathscr{I}^\pm$, i.e.

$$\begin{aligned} \tilde{\phi}^+(u, \omega) &= \lim_{r\to+\infty} r\phi(r + u, r\omega) \, , \\ \tilde{\phi}^-(v, \omega) &= \lim_{r\to+\infty} r\phi(-r + v, r\omega) \, . \end{aligned}$$

The variable s on $\mathscr{I}^+$ corresponds to $-u$ and on $\mathscr{I}^-$ to $-v$. Equations (5.19) and (8.4) then entail that

$$k(s, \omega) = \frac{\mathrm{d}}{\mathrm{d}s}\tilde{\phi}^+(-s, \omega) = -\frac{\mathrm{d}}{\mathrm{d}s}\tilde{\phi}^-(-s, -\omega) \, . \tag{8.5}$$

In other words

$$\partial_u\tilde{\phi}^+(u, \omega) = -\partial_v\tilde{\phi}^-(v, -\omega)$$

and since $\tilde{\phi}^\pm$ both have compact supports, (8.2) follows. This is established for smooth compactly supported data for the fields and extends to finite energy data by density. $\square$

8.2 Recovering the structures of analytic scattering theories

Although the scattering theories from Sections 7.4 and 7.6 have been constructed in a geometrical way, they retain some important analytic properties of the scattering constructions performed using the Lax-Phillips approach or Cook's method.

The wave operators can easily be interpreted as providing the comparison between two dynamics, the simplified dynamics being given by the flow of radial null geodesics. Even the identifying operator has a role to play : the simplified dynamics will associate to the asymptotic profile a function that is not well-defined on each level hypersurface of t, because of the rays focalising inside the space-time. This is solved easily using a smooth cut-off inside a fixed compact in space (in the physical space-time) and allows to interpret the inverse trace operators as direct wave operators defined exactly as in the section on Cook's method.

Provided we use the partial compactification of Section 3.1.2 (used in Section 7.4 and also in the case of the Schwarzschild metric in section 3.2.5), we see that the radiation fields are another type of translation representer for the solution. Indeed, in this compactification, the vector field ∂_t remains Killing and extends smoothly to null infinity as the null generator of $\mathscr{I}$. Taking as data the solution at time t instead of 0 means pulling back the full solution by a time $-t$ along the flow of ∂_t. This modifies the radiation fields (considered in the variables used for this compactification) by a translation of $-t$ along $\mathscr{I}$ (more precisely by pulling it back by $-t$ along the flow of the null generator of $\mathscr{I}$).

The translation representer is a feature associated with a timelike Killing vector that extends to $\mathscr{I}$ as its null generator. It will be present if we construct a (conformal) scattering theory on Schwarzschild's space-time. On time-dependent geometries however, we will lose this property. The interpretation of the trace operators as wave operators defined by comparing with a simplified dynamics will remain though, and the simplified dynamics will still be given as the flow of a congruence of null geodesics near $\mathscr{I}$.

Chapter 9

The wave equation on the Schwarzschild space-time

9.1 The Cauchy problem

Solved by the spectral approach.

9.2 Energy current and its conservation law

Provides an alternative way of solving the Cauchy problem.

9.3 Some decay results

Dafermos-Rodnianski, others?

9.4 The conformal scattering construction

9.5 Peeling

We study peeling properties in the neighbourhood of i^0 where the real problem is localized. If one wishes, results can be extended to the whole of $\mathscr{I}^+$ and the whole initial data surface in a straightforward manner. Our work deals with what happens near i^0.

We work with geometric energy estimates. We start by obtaining some basic estimates between $\mathscr{I}^+$ and the initial data surface $\{t = 0\}$. Then, by applying some well chosen differential operators to the equation, we get some higher order estimates.

For the basic energy estimates, we need to find some vector field that is close to being a Killing vector field for $\hat{g}$ and that is transverse to $\mathscr{I}$. We adapt the classic "Morawetz vector field" to Schwarzschild's space-time just as Inglese and Nicolò did in [26] and then Dafermos and Rodnianski in [10].

9.5.1 The Morawetz vector field

For $m=0$, the metric g is the Minkowski metric η and we have $u=t-r$. The Morawetz vector field is defined as the image of ∂_t, that is a Killing vector for η, by a light-cone inversion. Its simplest expression is in terms of variables $u=t-r$ and $v=t+r$:

$$T^a\partial_a = u^2\partial_u + v^2\partial_v \,,$$

which, in terms of variables u, R gives

$$T^a\partial_a := u^2\partial_u - 2(1+uR)\partial_R \,. \tag{9.1}$$

This vector field is a Killing vector for $\hat{g}$ for $m=0$, i.e. if we perform the time asymmetrical compactification for Minkowski space-time, we obtain a rescaled metric for which T is a Killing vector. It is interesting to note that the Killing vector ∂_τ corresponding to the time translation on the Einstein cylinder is a simple combination of T and ∂_u in this coordinate system :

$$2\partial_\tau = \partial_u + T^a\partial_a \,,$$

it is therefore also a Killing vector for $\hat{g}$ for $m=0$.

We keep the expression (9.1) in terms of variables u, R in the Schwarzschild case to define our approximate Killing vector field. We still refer to it as the Morawetz vector field and denote it T.

T^a is uniformly timelike in a neighbourhood of i^0 and can therefore be used for obtaining energy estimates with positive definite energies on spacelike hypersurfaces.

9.5.2 Stress energy tensor and energy density

We choose the stress-energy tensor for the free wave equation $\square_{\hat{g}}\phi = 0$

$$T_{ab} = T_{(ab)} = \partial_a\phi\partial_b\phi - \frac{1}{2}\hat{g}_{ab}\hat{g}^{cd}\partial_c\phi\partial_d\phi \,,$$

which, for ϕ solution of (**??**), satisfies $\nabla^a T_{ab} = \square\phi\partial_b\phi = -2mR\phi\partial_b\phi$. Contracting T_{ab} with T^a, we get the conservation law

$$\nabla^a\left(T^b T_{ab}\right) = T_{ab}\nabla^{(a}T^{b)} - 2mR\phi T^b\partial_b\phi \,. \tag{9.2}$$

The hope is then that the error terms (on the right hand side), have a sufficiently nice behaviour to allow a control via Gronwall-type arguments.

The energy density 3-form $E(\phi)$ associated with T^a is given by :

$$\begin{aligned} E(\phi) \;&:=\; T^a T_{ab}\mathrm{d}^3x^b = T^a T_a^b \partial_b \lrcorner \, \mathrm{dvol}^4 \\ &= \left[u^2\phi_u^2 + R^2(1-2mR)\left(u^2\phi_u\phi_R - (1+uR)\phi_R^2\right)\right. \\ &\quad \left. +(1+uR)|\nabla_{S^2}\phi|^2\right] \mathrm{d}u\wedge \mathrm{d}^2\omega \\ &\quad +\frac{1}{2}\left[\left((2+uR)^2 - 2mu^2R^3\right)\phi_R^2 + u^2|\nabla_{S^2}\phi|^2\right]\mathrm{d}R\wedge\mathrm{d}^2\omega \\ &\quad + \text{angular terms} \;. \end{aligned} \tag{9.3}$$

For a hypersurface $\mathcal{S}$, we denote

$$\mathcal{E}_{\mathcal{S}}(\phi) := \int_{\mathcal{S}} E(\phi)\,.$$

For instance,

$$\mathcal{E}_{\mathscr{I}^+}(\phi) = \int_{\mathscr{I}^+} \left[u^2\phi_u^2 + |\nabla_{S^2}\phi|^2\right] \mathrm{d}u \wedge \mathrm{d}^2\omega\,.$$

We foliate the domain $\{u < u_0 << -1\}$ by the spacelike (except for $s = 0$) hypersurfaces

$$\mathcal{H}_s := \{u = -sr_*\}\,,\ 0 \leq s \leq 1\,;$$

$\mathcal{H}_1$ is the $\{t = 0\}$ hypersurface and $\mathcal{H}_0$ corresponds to $\mathscr{I}^+$.

The energy on the surface $\mathcal{H}_s$ is given uniformly equivalent to

$$\mathcal{E}_{\mathcal{H}_s}(\phi) \quad \simeq \quad \int_{\mathcal{H}_s} \left(u^2\phi_u^2 + \frac{R}{|u|}\phi_R^2 + |\nabla_{S^2}\phi|^2\right) \mathrm{d}u \wedge \mathrm{d}^2\omega\,. \tag{9.4}$$

9.5.3 The fundamental energy estimates

For $m = 0$, T^a is a Killing vector field for our rescaled metric $\hat{g}$ (it is a conformal Killing vector field for g and in fact a Killing vector field for $\hat{g}$). For $m \neq 0$, the Killing form for T^a is given by

$$\nabla_{(a}T_{b)} = 4mR^2(3 + uR)\mathrm{d}u^2\,.$$

This gives

$$T_{ab}\nabla^a T^b = 4mR^2(3 + uR)\phi_R^2$$

which exhibits a nice fall-off near $\mathscr{I}^+$ and i_0.

We need to choose a vector field V that will identify the different hypersurfaces $\mathcal{H}_s$, and the conservation law (9.2) integrated over the domain $\{u < u_0\}$, $u_0 << -1$, gives (draw a picture of the domain $\{u < u_0\}$ and define the three hypersurfaces $\mathscr{I}^+_{u_0}$, $\mathcal{S}_{u_0}$ and $\mathcal{H}_{1,u_0}$)

$$\begin{aligned}
\int_{\{u<u_0\}} \nabla^a \left(T^b T_{ab}\right) \mathrm{d}^4\mathrm{Vol} \quad &= \quad \mathcal{E}_{\mathscr{I}^+_{u_0}} + \mathcal{E}_{\mathcal{S}_{u_0}} - \mathcal{E}_{\mathcal{H}_{1,u_0}} \\
&= \quad \int_{\{u<u_0\}} \left(T_{ab}\nabla^{(a}T^{b)} - 2mR\phi T^b\partial_b\phi\right) \mathrm{d}^4\mathrm{Vol} \\
&= \quad \int_0^1 \left(\int_{\mathcal{H}_{s,u_0}} \left(T_{ab}\nabla^{(a}T^{b)} - 2mR\phi T^b\partial_b\phi\right) V \lrcorner\, \mathrm{d}^4\mathrm{Vol}\right) \mathrm{d}s\,.
\end{aligned}$$

Another way of understanding this kind of energy estimates that is more familiar to PDE analysts is the following : we multiply equation (**??**) by $T\phi$ and integrate the result between $s = 0$ and $s = 1$. This is not quite precise enough however : it is implicitely assumed that we have a well defined product structure on the domain $\{u < u_0\}$, s being a "time" variable and the space variables being chosen on $\mathscr{I}^+$ or on $\mathcal{H}_1$ or any other $\mathcal{H}_s$. It is in order to define this product structure that we need an identifying vector field V. It will perform a splitting of the 4-volume measure into a part along the integral curves of V that is simply $\mathrm{d}s$ and a 3-volume measure on each $\mathcal{H}_s$ defined as $V \lrcorner\, \mathrm{d}^4\mathrm{Vol}$.

In this case, a natural identifying vector field ν [1] is given by

$$\nu = r_*^2 R^2 (1-2mR)|u|^{-1}\partial_R \,. \tag{9.5}$$

The corresponding 3-volume measure on each $\mathcal{H}_s$ is then

$$\nu \lrcorner \, \mathrm{dVol}^4|_{\mathcal{H}_s} = r_*^2 R^2 (1-2mR)|u|^{-1} \mathrm{d}u \mathrm{d}^2\omega|_{\mathcal{H}_s} \,.$$

Such a choice of identifying vector field ν^a, which is really associated with the choice of parameter s for the foliation, will lead to error terms that cannot be controlled by the energy density[2] and therefore to the impossibility of performing a priori estimates.

To be more precise, the 4-volume error terms will be multiplied by

$$T^a \partial_a \phi r_*^2 R^2 (1-2mR)|u|^{-1}$$

and then considered as terms on the 3-surface $\mathcal{H}_s$ that we shall try to control by the energy density. Supposing that the (squares of the) 4-volume error-terms are barely controlled by the energy density, we have a problem, indeed :

$$T^a \partial_a \phi \nu \lrcorner \, \mathrm{dVol}^4 \simeq u\phi_u - 2(1+uR)\frac{1}{u}\phi_R \,.$$

The first term is naturally controlled by the energy density on the $\mathcal{H}_s$ slices, but not the second, since u^{-1} is infinitely larger than Ru^{-1} near $\mathscr{I}^+$.

All we need to do in order to solve this problem is the following change of parameter :

$$\tau := -2\left(\sqrt{s}-1\right) \text{ , so that } \frac{\partial \tau}{\partial s} = -\frac{1}{\sqrt{s}}. \tag{9.6}$$

The change of sign and the -1 term are there purely for aesthetic reasons, the important part is $2\sqrt{s}$. This new parameter varies from 0 to 2 as s varies from 1 to 0. We denote

$$\Sigma_{\tau(s)} = \mathcal{H}_s \,. \tag{9.7}$$

The natural new identifying vector field is

$$V = -\sqrt{s}\,\nu = -\sqrt{\frac{|u|}{r_*}} r_*^2 R^2 (1-2mR)|u|^{-1}\partial_R = -\left(r_* R\right)^{3/2}(1-2mR)\sqrt{\frac{R}{|u|}}\,\partial_R \,. \tag{9.8}$$

---

[1] It needs to satisfy

$$\frac{\partial s}{\partial \nu} = -\frac{1}{r_*}\frac{\partial u}{\partial \nu} + \frac{u}{r_*^2}\frac{\partial r_*}{\partial \nu} = 1\,,$$

whence the expression (9.5) of ν obtained by imposing that it is colinear to ∂_R, i.e. choosing ν parallel to the level hypersurfaces of u ; this is natural given the shape of our domain $\{u < u_0\}$.

[2] To be more precise, this does not occur for the fundamental estimates, because the scalar curvature $2mR$ gives us some extra fall-off at $\mathscr{I}$. For higher order estimates, commuting ∂_R into the equation will give error terms without any fall-off. So the problem will occur as soon as we try to gain one extra degree of regularity from the fundamental estimates.

We get error terms on each Σ_τ that are equivalent to :

$$\begin{aligned}
(\nabla^a T^b) T_{ab} V \lrcorner \, \mathrm{dvol}^4 &\simeq R^2 \sqrt{\frac{R}{|u|}} \phi_R^2 \mathrm{d}u \mathrm{d}^2\omega \,, \\
\left(-2mR\phi T^b \partial_b \phi \right) V \lrcorner \, \mathrm{dVol}^4 &\simeq R\phi \sqrt{\frac{R}{|u|}} \left(u^2 \partial_u \phi - 2(1+uR)\partial_R \phi \right) \mathrm{d}u \mathrm{d}^2\omega \\
&\lesssim \left(R^2 \phi^2 + u^2 \phi_u^2 + \frac{R}{|u|} \phi_R^2 \right) \mathrm{d}u \mathrm{d}^2\omega \,.
\end{aligned}$$

The only difficulty comes from the zero order term. We solve it by proving (simply by integration by parts) the following estimate :

Lemma 9.1. *Given $u_0 < 0$, there exists a constant $C > 0$ such that for any $f \in \mathcal{C}_0^\infty(\mathbb{R})$, we have*

$$\int_{-\infty}^{u_0} (\phi(u))^2 \, \mathrm{d}u \leq C \int_{-\infty}^{u_0} u^2 \left(\phi'(u) \right)^2 \mathrm{d}u \,.$$

This entails that the energy density controls the L^2 norm on the $\mathcal{H}_s$ slices.

We see that this is much more than what we need, but for higher order estimates, we will use all that this lemma gives us.

This allows us to obtain, via simple Gronwall estimates, the following estimates

Theorem 9.1. *For $u_0 < 0$, $|u_0|$ large enough, there exists a constant $C > 0$ such that,*

$$\mathcal{E}_{\mathscr{I}^+_{u_0}}(\phi) \leq C \mathcal{E}_{\mathcal{H}_{1,u_0}}(\phi) \,,$$

$$\mathcal{E}_{\mathcal{H}_{1,u_0}}(\phi) \leq C \left(\mathcal{E}_{\mathscr{I}^+_{u_0}}(\phi) + \mathcal{E}_{\mathcal{S}_{u_0,s_0}}(\phi) \right) \,.$$

9.5.4 Higher order estimates

Contrary to what happens for $m = 0$, where T is a Killing vector field, we cannot use it here to raise the regularity, since

$$[T, [T, 2mR]] = 4mu(2 + uR) \,,$$

so differentiating the equation several times using T will introduce potentials which blow up near i^0. Although this may appear as unfortunate, it is a blessing in disguise.

Instead, we use the vector field ∂_R. We obtain

$$(\Box + 2mR) \, \phi_R = 2(1 - 3m) R \partial_R \phi_R - 2(1 - 6mR)\phi_R - 2m\phi \,. \tag{9.9}$$

All the terms in the right hand-side can be controlled by the energy density for ϕ_R or ϕ using when necessary the lemma above. It is clear that further differentiations using ∂_R will not raise any difficulty. We get the estimates

Theorem 9.2. *For each $k \in \mathbb{N}$, there exists a constant $C_k > 0$ such that, for any solution ϕ of (??) associated to smooth compactly supported initial data, we have for all $0 \leq s \leq 1$,*

$$\begin{aligned}\mathcal{E}_{\mathscr{I}_{u_0}}(\partial_R^k \phi) &\leq C \sum_{p=0}^{k} \mathcal{E}_{\mathcal{H}_{1,u_0}}(\partial_R^p \phi)\,,\\ \mathcal{E}_{\mathcal{H}_{1,u_0}}(\partial_R^k \phi) &\leq C \sum_{p=0}^{k} \left(\mathcal{E}_{\mathscr{I}^+_{u_0}}(\partial_R^p \phi) + \mathcal{E}_{\mathcal{S}_{u_0}}(\partial_R^p \phi) \right) .\end{aligned}$$

This gives us a simple characterization of the peeling

Definition 9.1. *We say that a solution ϕ of (??) peels at order $k \in \mathbb{N}$ if for all polynomial P in ∂_R and ∇_S^2 of order lower than or equal to k, we have $\mathcal{E}_{\mathscr{I}^+_{u_0}}(P\phi) < +\infty$. This means than for all $p \in \{0, 1, ..., k\}$ we have for all $q \in \{0, 1, ..., p\}$, $\mathcal{E}_{\mathscr{I}^+_{u_0}}(\partial_R^q \nabla_{S^2}^{p-q} \phi) < +\infty$.*

This is of course valid for Minkowski space as well. So in flat space-time, we have two definitions of the peeling :

- a first one obtained using the embedding in the Einstein cylinder and the time translation of the Einstein cylinder (quite close to the Morawetz cector field) ;
- a second one using the time asymmetric compactification and the vector field ∂_R.

It turns out these two definitions are different. The class of data given by the second definition is larger than that given by the first. This is due to the fact that the vector field we use for higher order estimates is characteristic (or null if one prefers this terminology), which gives less stringent conditions than a timelike vector field that controls all spacelike derivatives via the equation (the energies on $\mathcal{H}_1$ given by both constructions are equivalent).

So this is a complete verification of the peeling model at all orders for the wave equation on the Schwarzschild metric, as well a definition that is different and possibly more relevent definition of the sets of solutions admitting peeling at a given order. Why more relevent? Because we show that we merely need to control the null derivative in the direction of $\mathscr{I}^+$ instead of a timelike derivative. It is in a way a propagation estimate : the other null derivative does not give a contribution on $\mathscr{I}^+$ so even if it lies in very weakly regular spaces, it will not interfere with the regularity of the solution at $\mathscr{I}^+$.

Chapter 10

The wave equation on asymptotically simple space-times

10.1 The Goursat problem using a modified Hörmander approach

10.2 Conformal scattering for Corvino-Schoen/Chrusciel-Delay space-times

Appendix A

Elements of functional and spectral analysis

In this chapter, we give a brief construction of distributions and tempered distributions and Sobolev spaces on $\mathbb{R}^n$ and on regular bounded open sets, and give some bases of spectral theory. The material is presented as a quick review of useful notions for our purposes. Some proofs are given when they provide useful techniques for applications. The reader eager to explore these questions in depth can look at the first and second volumes of the book by Reed and Simon [47, 48], the first volume of the book by Taylor [57] and volumes 1, 2 and 3 of the book by Hörmander [24]. We only work with the canonical Euclidean norm on $\mathbb{R}^n$ and we denote it by $\|.\|$. From section A.4 onwards, we shall work on a Hilbert space $\mathscr{H}$ that will always be assumed to be separable. Its inner product and norm will be denoted by $\langle ., . \rangle$ and $\|.\|$. We assume a basic knowledge of Hilbert spaces. We also assume some familiarity with Lebesgue spaces, but we recall their definition here.

Definition A.1 (Lebesgue spaces)**.** *Let Ω be an open subset of $\mathbb{R}^n$ and $p \in [1, +\infty[$, the space $L^p(\Omega)$ is the space of Lebesgue measurable functions on Ω such that*

$$\int_\Omega |f(x)|^p \mathrm{d}x < +\infty \,.$$

It is a Banach space for the norm

$$\|f\|_p = \left(\int_\Omega |f(x)|^p \mathrm{d}x \right)^{1/p} \,.$$

The space $L^\infty(\Omega)$ is the space of functions on Ω such that

$$\operatorname*{ess\,sup}_{x \in \Omega} |f(x)| := \min\{b \in \mathbb{R} \,;\ f^{-1}(]b, +\infty[) \text{ negligible}\} < +\infty \,.$$

It is a Banach space for the norm

$$\|f\|_\infty = \operatorname*{ess\,sup}_{x \in \Omega} |f(x)| \,.$$

Remark A.1. *To be completely precise, the norms above do not distinguish between two functions that are equal except on a set of measure zero and for $p \in [1, \infty]$, the space $L^p(\Omega)$ is really a space of classes of functions under the equivalence relation "almost everywhere equal". We shall however ignore this and simply identify the classes with any of their representatives.*

The local Lebesgue spaces are also important. They are defined as follows.

Definition A.2 (Local Lebesgue spaces)**.** *Let Ω be an open subset of $\mathbb{R}^n$ and $p \in [1, \infty]$, $L^p_{\text{loc}}(\Omega)$ is the set of functions f such that for any relatively compact open subset $\mathcal{O}$ of Ω, we have $f \in L^p(\mathcal{O})$.*

A.1 Spaces of smooth functions and distributions on $\mathbb{R}^n$

Let Ω be an open subset of $\mathbb{R}^n$. We denote by $\mathscr{C}^\infty(\Omega)$ the space of infinitely differentiable functions on Ω. It is not a normed space but it is endowed with an infinite family of semi-norms defined for each compact subset K of Ω and $\alpha \in \mathbb{N}^n$ by

$$N_{K,\alpha}(f) = \max_{x \in K} \left| \frac{\partial^\alpha f}{\partial x^\alpha}(x) \right|$$

that define the topology. This means that a sequence of functions $(f_n)_{n \in \mathbb{N}}$ in $\mathscr{C}^\infty(\Omega)$ converges towards a function f in $\mathscr{C}^\infty(\Omega)$ if and only if for all compact subsets K of Ω and $\alpha \in \mathbb{N}^n$, we have

$$\lim_{n \to +\infty} N_{K,\alpha}(f_n - f) = 0\,.$$

The fundamental function space in the theory of distributions is the space of test functions. Its definition requires to introduce the notion of support of a continuous function.

Definition A.3. *Let Ω be an open subset of $\mathbb{R}^n$ and f a continuous function on Ω with real or complex values. The support of f, denoted* $\operatorname{supp} f$*, is the complement in Ω of the largest open subset of $\mathbb{R}^n$ on which f is identically zero. This is equivalently defined as the closure in Ω of the set on which f is non zero, i.e.*

$$\operatorname{supp} f = \left(\overset{\circ}{\overbrace{f^{-1}(\{0\})}} \right)^{\mathrm{C}} = \overline{\{x \in \mathbb{R}^n\,;\ f(x) \neq 0\}}\,.$$

Definition A.4. *We denote by $\mathscr{C}_0^\infty(\Omega)$ the space of smooth functions on Ω with compact support. In other words, $f \in \mathscr{C}_0^\infty(\Omega)$ if and only if $f \in \mathscr{C}^\infty(\Omega)$ and there exists a compact subset K of Ω such that $f(x) = 0$ for $x \in \Omega \setminus K$. The topology on $\mathscr{C}_0^\infty(\Omega)$ is the following: we say that a sequence $(\phi_n)_{n \in \mathbb{N}}$ converges towards ϕ in $\mathscr{C}_0^\infty(\Omega)$ if and only if the following two properties are satisfied*

1. *there exists a compact subset K of Ω such that* $\operatorname{supp} \phi \subset K$ *and for all $n \in \mathbb{N}$,* $\operatorname{supp} \phi_n \subset K$*, and*

2. *ϕ_n converges towards ϕ in $\mathscr{C}^\infty(\Omega)$.*

The space $\mathscr{C}_0^\infty(\Omega)$ will often be denoted by $\mathscr{D}(\Omega)$; its elements are called test functions.

This space is fundamental in the construction of many function spaces by completion under well-chosen norms. For example, for $p \in [1, +\infty[$, $L^p(\Omega)$ is the completion of $\mathscr{D}(\Omega)$ under the norm $\|.\|_p$. However the completion of $\mathscr{D}(\Omega)$ under the norm $\|.\|_\infty$ is not $L^\infty(\Omega)$ but $\mathscr{C}^0(\bar{\Omega})$.

There is another space of smooth functions that is of crucial importance when considering the Fourier transform, this is the space of Schwartz functions. It is not a "local" space defined on any open subset on $\mathbb{R}^n$ but it is globally defined on $\mathbb{R}^n$ as follows.

Definition A.5. *The space of Schwartz functions on $\mathbb{R}^n$, denoted $\mathscr{S}(\mathbb{R}^n)$ is the subspace of $\mathscr{C}^\infty(\mathbb{R}^n)$ of functions f that are rapidly decaying at infinity, in the sense that for each $\alpha, \beta \in \mathbb{N}^n$, the semi-norm*

$$N_{\alpha,\beta}(f) = \sup_{x \in \mathbb{R}^n} \left| x^\beta \frac{\partial^\alpha f}{\partial x^\alpha}(x) \right|$$

is finite. The semi-norms $N_{\alpha,\beta}$ define the topology on $\mathscr{S}(\mathbb{R}^n)$.

We now introduce the notion of distribution on Ω and its support. The theory of distributions was developed by the french mathematician Laurent Schwartz in the 1940's, on the basis of the works of Heaviside and Dirac. He was awarded the Fields medal for this in 1950.

Definition A.6. *Let Ω be an open subset of $\mathbb{R}^n$. A distribution on Ω is a linear form on $\mathscr{D}(\Omega)$ that satisfies the following property:*

($\mathscr{P}$) for any compact subset K of Ω, there exists $k \in \mathbb{N}$ and $C > 0$ such that for each $\phi \in \mathscr{D}(\Omega)$ with $\operatorname{supp}(\phi) \subset K$, we have

$$|T(\phi)| \leq C \sum_{|\alpha| \leq k} \max_{x \in \Omega} \left| \frac{\partial^\alpha \phi}{\partial x^\alpha}(x) \right|.$$

The action of a distribution $T \in \mathscr{D}'(\Omega)$ on a test function $\phi \in \mathscr{D}(\Omega)$ is a duality product denoted by $\langle T, \phi \rangle_{\mathscr{D}'(\Omega),\mathscr{D}(\Omega)}$ or $\langle T, \phi \rangle_{\mathscr{D}',\mathscr{D}}$ when the open set is clear, or even $\langle T, \phi \rangle$ when there is no ambiguity.

- *A distribution $T \in \mathscr{D}'(\Omega)$ is said to vanish on an open subset $\mathcal{O}$ of Ω is for all $\phi \in \mathscr{D}(\Omega)$ such that $\operatorname{supp}(\phi) \subset \mathcal{O}$ we have $\langle T, \phi \rangle = 0$.*

- *The support of a distribution $T \in \mathscr{D}'(\Omega)$ is the complement of the largest open subset $\mathcal{O}$ of Ω on which T vanishes.*

- *When in the property ($\mathscr{P}$) the integer k is independent of the choice of the compact subset K, the distribution T is said to be of finite order. The order of T is then the lowest integer k for which ($\mathscr{P}$) is valid.*

Here are some fundamental examples of distributions.

- **Distribution associated to a function.** Let Ω be an open subset of $\mathbb{R}^n$ and $f \in L^1_{\mathrm{loc}}(\Omega)$. The distribution T_f associated to f is a distribution of order 0 on Ω defined by

$$\langle T_f, \phi\rangle = \int_\Omega f(x)\phi(x)\mathrm{d}x \text{ for all } \phi \in \mathscr{D}(\Omega)\,.$$

  This realises the continuous embedding $L^1_{\mathrm{loc}}(\mathbb{R}^n) \hookrightarrow \mathscr{D}'(\mathbb{R}^n)$ and we have obviously for all $p \in [1,\infty]$,

$$L^p(\mathbb{R}^n) \hookrightarrow L^p_{\mathrm{loc}}(\mathbb{R}^n) \hookrightarrow L^1_{\mathrm{loc}}(\mathbb{R}^n) \hookrightarrow \mathscr{D}'(\mathbb{R}^n).$$

- **Dirac distribution at a point.** Let $a \in \mathbb{R}^n$, the Dirac distribution at a, denoted δ_a, is the distribution of order 0 on $\mathbb{R}^n$ defined by

$$\langle \delta_a, \phi\rangle = \phi(a) \text{ for all } \phi \in \mathscr{D}(\mathbb{R}^n)\,.$$

The construction of distributions by duality induces naturally some operations on distributions from their analogues on functions. Let Ω be an open subset of $\mathbb{R}^n$, $T \in \mathscr{D}'(\Omega)$ and $f \in \mathscr{C}^\infty(\Omega)$, then we have

- *multiplication by a smooth function*

$$\langle fT, \phi\rangle = \langle T, f\phi\rangle\,,$$

- *derivative of a distribution*

$$\left\langle \frac{\partial T}{\partial x^k}, \phi \right\rangle = -\left\langle T, \frac{\partial \phi}{\partial x^k}\right\rangle\,.$$

Two other important spaces of distributions are $\mathscr{E}'(\Omega)$, the space of distributions with compact support on Ω (the topological dual of $\mathscr{C}^\infty(\Omega)$) and $\mathscr{S}'(\mathbb{R}^n)$, the space of "tempered distributions" on $\mathbb{R}^n$ that is the topological dual of $\mathscr{S}(\mathbb{R}^n)$. All the Lebesgue spaces on $\mathbb{R}^n$ satisfy

$$L^p(\mathbb{R}^n) \hookrightarrow \mathscr{S}'(\mathbb{R}^n) \hookrightarrow \mathscr{D}'(\mathbb{R}^n)\,,\ p \in [1,\infty].$$

A.2 Fourier transform, Sobolev spaces on $\mathbb{R}^n$

The Fourier transform on $\mathbb{R}^n$, defined by

$$(\mathscr{F}f)(\xi) := \frac{1}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-ix.\xi} f(x)\mathrm{d}x,$$

also denoted $\hat{f}(\xi)$, is an isometry of $L^2(\mathbb{R}^n)$ and an automorphism of $\mathscr{S}(\mathbb{R}^n)$. It induces by duality an automorphism of $\mathscr{S}'(\mathbb{R}^n)$ still denoted by $\mathscr{F}$ and defined by

$$\langle \mathscr{F}S, \phi\rangle_{\mathscr{S}',\mathscr{S}} = \langle S, \mathscr{F}\phi\rangle_{\mathscr{S}',\mathscr{S}}\,, \quad \text{or also } \langle \hat{S}, \phi\rangle_{\mathscr{S}',\mathscr{S}} = \langle S, \hat{\phi}\rangle_{\mathscr{S}',\mathscr{S}}.$$

This follows from the well-known property of the Fourier transform

$$\int_{\mathbb{R}^n} \hat{f}(x) g(x)\mathrm{d}x = \int_{\mathbb{R}^n} f(x)\hat{g}(x)\mathrm{d}x$$

for functions f and g belonging to $L^1(\mathbb{R}^n)$. All the usual properties of the Fourier transform extend to $\mathscr{S}'(\mathbb{R}^n)$, in particular

$$\mathscr{F}\left(\frac{\partial S}{\partial x^k}\right) = i\xi_k \hat{S}\,, \ \frac{\partial \hat{S}}{\partial \xi_k} = -i\mathscr{F}\left(x^k S\right)\,.$$

Sobolev spaces on $\mathbb{R}^n$ are naturally defined using the Fourier transform.

Definition A.7. *For $s \in \mathbb{R}$, the Sobolev space of order s on $\mathbb{R}^n$, denoted by $H^s(\mathbb{R}^n)$, is defined by*

$$H^s(\mathbb{R}^n) = \left\{f \in \mathscr{S}'(\mathbb{R}^n)\,;\ (1+|\xi|^2)^{s/2}\hat{f} \in L^2(\mathbb{R}^n)\right\}\,,$$

which means that $(1+|\xi|^2)^{s/2}\hat{f}$ is the distribution associated to a function in $L^2(\mathbb{R}^n)$. It is a separable Hilbert space.

The space $H^s(\mathbb{R}^n)$ is also the completion of $\mathscr{D}(\mathbb{R}^n)$ in the norm

$$\|f\|_{H^s(\mathbb{R}^n)} = \left\|(1+|\xi|^2)^{s/2}\hat{f}\right\|_{L^2(\mathbb{R}^n)}\,.$$

When $k \in \mathbb{N}$, $H^k(\mathbb{R}^n)$ can be equivalently defined as the space of tempered distributions f on $\mathbb{R}^n$ such that $\partial^\alpha f \in L^2(\mathbb{R}^n)$ for $\alpha \in \mathbb{N}^n$, $|\alpha| \le k$ and a norm equivalent to the one above is

$$\|f\|_k = \left(\sum_{|\alpha|\le k} \|\partial^\alpha f\|^2_{L^2(\mathbb{R}^n)}\right)^{1/2}\,.$$

For $s > 0$, $H^{-s}(\mathbb{R}^n)$ is the topological dual of $H^s(\mathbb{R}^n)$ and we have the chain of continuous embeddings for $t > s > 0$:

$$\mathscr{D}(\mathbb{R}^n) \hookrightarrow \mathscr{S}(\mathbb{R}^n) \hookrightarrow H^t(\mathbb{R}^n) \hookrightarrow H^s(\mathbb{R}^n) \hookrightarrow L^2(\mathbb{R}^n),$$
$$L^2(\mathbb{R}^n) \hookrightarrow H^{-s}(\mathbb{R}^n) \hookrightarrow H^{-t}(\mathbb{R}^n) \hookrightarrow \mathscr{S}'(\mathbb{R}^n) \hookrightarrow \mathscr{D}'(\mathbb{R}^n).$$

A.3 Sobolev spaces on a regular bounded open set

When Ω is a regular bounded open set (i.e. a relatively compact open set whose boundary $\partial\Omega$ is a smooth compact hypersurface of $\mathbb{R}^n$), for $s \in \mathbb{R}$, we define the Sobolev space $H^s(\Omega)$ as the set of restrictions to Ω of elements of $H^s(\mathbb{R}^n)$, i.e.

$$H^s(\Omega) = \{u \in \mathscr{D}'(\Omega)\,;\ \exists U \in H^s(\mathbb{R}^n)\,;\ u = U|_\Omega\}.$$

It is a Banach space for the natural norm

$$\|u\|_{H^s(\Omega)} = \inf\{\|U\|_{H^s(\mathbb{R}^n)}\,,\ U \in H^s(\mathbb{R}^n) \text{ and } U|_\Omega = u\}.$$

This definition is in fact valid for any open set in $\mathbb{R}^n$ but for regular bounded open sets, we have the following property:

$$\text{for } k \in \mathbb{N}\,,\ H^k(\Omega) = \{u \in \mathscr{D}'(\Omega)\,;\ \partial^\alpha u \in L^2(\Omega) \text{ for } |\alpha| \le k\}$$

and it is endowed with an equivalent norm to $\|.\|_{H^s(\Omega)}$

$$\|u\|_{k,\Omega} = \left(\sum_{|\alpha|\leq k} \|\partial^\alpha u\|^2_{L^2(\Omega)} \right)^{1/2}$$

for which it is a separable Hilbert space.

If we define $\mathscr{D}(\bar{\Omega})$ as the set of restrictions to Ω of elements of $\mathscr{D}(\mathbb{R}^n)$, then for $s \in \mathbb{R}$, $H^s(\Omega)$ is the completion of $\mathscr{D}(\bar{\Omega})$ in the natural norm.

For $s > 1/2$, we have a notion of "trace", i.e. restriction to the boundary, for elements of $H^s(\Omega)$.

Theorem A.1 (Traces)**.** *Let Ω be a regular bounded open set with boundary $\partial\Omega$ and incoming unit normal ν. The trace of order j, for $j \in \mathbb{N}$, is the map $\gamma_j : \mathscr{D}(\bar{\Omega}) \to \mathscr{C}^\infty(\partial\Omega)$ that to ϕ associates its incoming normal derivative of order j at $\partial\Omega$, i.e. $\partial^j_\nu \phi$. For $s > j + 1/2$, γ_j extends by density to a continuous linear map from $H^s(\Omega)$ to $H^{s-j-1/2}(\partial\Omega)$.*

Remark A.2. *Note that the definition of Sobolev spaces on $\partial\Omega$ can be inferred from that of Sobolev spaces on a regular bounded open set on $\mathbb{R}^{n-1}$ via a choice of atlas on $\partial\Omega$.*

For $k \in N$, $k \geq 1$, the completion of $\mathscr{D}(\Omega)$ in the norm $\|.\|_{k,\Omega}$ is denoted by $H^k_0(\Omega)$. It is a closed subspace of $H^k(\Omega)$ and it has the following characterisation

$$H^k_0(\Omega) = \left\{ u \in H^k(\Omega)\,;\ \gamma_j(u) = 0\,;\ 0 \leq j \leq k-1 \right\}.$$

Understanding the origin of the loss of half a derivative in Theorem A.1. We show how to derive a weaker version of the theorem where we loose $\frac{1}{2} + \varepsilon$ in the order of the Sobolev space. The result of the theorem itself, with a loss of exactly half a degree of regularity, is not very complicated to prove but requires interpolation between Sobolev spaces. The weaker version is technically lighter and shows quite clearly the origin of the loss of half a regularity order in the trace process. This weaker version can be derived explicitly in a simple example that is not in the setting of the theorem but is in fact fundamental in the construction of the proof: the case where Ω is a half space. In order to understand this simple case, we are going to consider a function in $f \in H^s(\mathbb{R}^n)$, $s > 0$, and represent $\mathbb{R}^n$ as $\mathbb{R} \times \mathbb{R}^{n-1}$, the variables being denoted $z = (x, y) \in \mathbb{R} \times \mathbb{R}^n$ with dual variables through a Fourier transform $\zeta = (\xi, \eta)$. First, if $s_1, s_2 \geq 0$ are such that $s_1 + s_2 \leq s$, then we have

$$H^s(\mathbb{R}^n) \hookrightarrow H^{s_1}(\mathbb{R}\,;\ H^{s_2}(\mathbb{R}^{n-1}))\,.$$

This is due to the fact that

$$\begin{aligned}
\|f\|_{H^{s_1}(\mathbb{R};\ H^{s_2}(\mathbb{R}^{n-1}))} &= \left\| (1+|\xi|^2)^{s_1/2}(1+|\eta|^2)^{s_2/2}\hat{f} \right\|_{L^2(\mathbb{R}^n)} \\
&\leq \left\| (1+|\zeta|^2)^{s_1/2}(1+|\zeta|^2)^{s_2/2}\hat{f} \right\|_{L^2(\mathbb{R}^n)} \\
&\leq \left\| (1+|\zeta|^2)^{s/2}\hat{f} \right\|_{L^2(\mathbb{R}^n)} = \|f\|_{H^s(\mathbb{R}^n)}.
\end{aligned}$$

Now we look for the values of $s-1$ such that $H^{s_1}(\mathbb{R}\,;\ H^{s_2}(\mathbb{R}^{n-1})) \hookrightarrow \mathscr{C}^0(\mathbb{R}\,;\ H^{s_2}(\mathbb{R}^{n-1}))$, i.e. such that $H^{s_1}(\mathbb{R}) \hookrightarrow \mathscr{C}^0(\mathbb{R})$. The definition of $H^{s_1}(\mathbb{R})$ can be rewritten as

$$u \in H^{s_1}(\mathbb{R}) \Leftrightarrow \text{ there exists } v \in L^2(\mathbb{R}) \text{ such that } \hat{u} = \frac{v}{(1+|\xi|^2)^{s_1/2}}.$$

If this is the case, then u is given by the inverse Fourier transform of $\hat{u}$, i.e.

$$u(x) = (2\pi)^{-n/2} \int_{\mathbb{R}} \frac{1}{(1+|\xi|^2)^{s_1/2}} v(\xi) e^{ix.\xi} \mathrm{d}\xi\,.$$

The function

$$\frac{1}{(1+|\xi|^2)^{s_1/2}}$$

belongs to $L^2(\mathbb{R})$ if and only if $s_1 > 1/2$. In this case we can use Lebesgue's dominated convergence Theorem to infer the continuity of u. We have therefore established that for $s_1 > 1/2$ and $s \geq s_1 + s_2$, we have

$$H^s(\mathbb{R}^n) \hookrightarrow \mathscr{C}^0(\mathbb{R}\,;\ H^{s_2}(\mathbb{R}^{n-1}))\,.$$

In particular, in this case, the trace of order zero at the hypersurface $\{x = 0\} \simeq \mathbb{R}^{n-1}$ is well-defined as a continuous map from $H^s(\Omega)$ to $H^{s-s_1}(\mathbb{R}^{n-1})$ for $s > s_1 > 1/2$. Using an atlas of the boundary, this result can be used to prove the same result for a bounded regular open set. Extensions to higher order traces follow the same idea.

A.4 Spectral theory of bounded self-adjoint operators

A first fundamental notion is that of the adjoint of an operator.

Theorem A.2. *Let $A \in \mathscr{L}(\mathscr{H})$, there exists a unique operator $A^* \in \mathscr{L}(\mathscr{H})$ such that, for all $x, y \in \mathscr{H}$, we have*

$$\langle Ax, y \rangle = \langle x, A^* y \rangle\,.$$

The operator A^* is called the adjoint of A. It also satisfies $\|A^*\| = \|A\|$.

Proof. This is a direct consequence of Riesz's representation Theorem, which states that for any continuous linear form l on $\mathscr{H}$ (i.e. $l \in \mathscr{H}'$, the topological dual of $\mathscr{H}$), there exists a unique $z \in \mathscr{H}$ such that

$$l(x) = \langle x, z \rangle \text{ for all } x \in \mathscr{H}$$

and in addition $\|x\|_{\mathscr{H}} = \|l\|_{\mathscr{H}'}$. □

A useful property of the adjoint that is a direct consequence of its definition is that for any $A \in \mathscr{L}(\mathscr{H})$,

$$\mathscr{H} = \mathrm{Ker}(A) \overset{\perp}{\oplus} \overline{\mathrm{ran} A^*}\,, \tag{A.1}$$

where the symbol $\overset{\perp}{\oplus}$ denotes an orthogonal direct sum.

An operator $A \in \mathscr{L}(\mathscr{H})$ is said to be self-adjoint if $A = A^*$, skew-adjoint if $A = -A^*$. Note that A is skew-adjoint if and only if the operator iA is self-adjoint. We have a useful characterisation of self-adjointness.

Proposition A.1. *Let $A \in \mathscr{L}(\mathscr{H})$, then*

1. *A is self-adjoint if and only if*
$$\langle Ax, x\rangle \in \mathbb{R} \text{ for all } x \in \mathscr{H},$$

2. *if A is self-adjoint, then*
$$\|A\| = \sup_{x\in\mathscr{H}\,,\ \|x\|=1} |\langle Ax, x\rangle| \ .$$

Proof. The proof of the first point is a consequence of the polarisation identity. Let us prove point 2. A first inequality is immediate

$$\sup_{x\in\mathscr{H}\,,\ \|x\|=1} |\langle Ax, x\rangle| \leq \|A\| \, .$$

Let us show the converse inequality. We use the polarisation identity

$$\begin{aligned}\langle Ax, y\rangle &= \frac{1}{4}\left(\langle A(x+y), x+y\rangle - \langle A(x-y), x-y\rangle\right) \\ &\quad + \frac{i}{4}\left(\langle A(x+iy), x+iy\rangle - \langle A(x-iy), x-iy\rangle\right) \, .\end{aligned}$$

The first line of the right-hand side is real and the second is pure imaginary. We infer

$$\Re\langle Ax, y\rangle = \frac{1}{4}\left(\langle A(x+y), x+y\rangle - \langle A(x-y), x-y\rangle\right)$$

and putting

$$K = \sup_{x\in\mathscr{H}\,,\ \|x\|=1} |\langle Ax, x\rangle|$$

we have

$$|\Re\langle Ax, y\rangle| \leq \frac{K}{4}\left(\|x+y\|^2 + \|x-y\|^2\right) = \frac{K}{2}\left(\|x\|^2 + \|y\|^2\right) \tag{A.2}$$

since for $x \in \mathscr{H}$,

$$|\langle Ax, x\rangle| \leq K\|x\|^2 \, .$$

If $K = 0$, we choose $y = Ax$, and from (A.2) we get

$$|\Re\langle Ax, y\rangle| = \langle Ax, Ax\rangle = \|Ax\|^2 = 0$$

for all $x \in \mathscr{H}$, whence $\|A\| = 0$. If $K > 0$, we put $y = \frac{1}{K}Ax$. Inequality (A.2) yields

$$\frac{1}{K}\|Ax\|^2 \leq \frac{K}{2}\left(\|x\|^2 + \frac{1}{K^2}\|Ax\|^2\right) ,$$

whence

$$\frac{1}{2K}\|Ax\|^2 \leq \frac{K}{2}\|x\|^2$$

i.e. for $x \neq 0$,

$$\frac{\|Ax\|^2}{\|x\|^2} \leq K^2$$

and finally $\|A\| \leq K$. $\square$

Let $A \in \mathscr{L}(\mathscr{H})$, the resolvent set of A is defined as

$$\rho(A) = \{\lambda \in \mathbb{C}\,;\ \lambda I - A \text{ is an automorphism of } \mathscr{H}\}.$$

As a consequence of the fact that for $B \in \mathscr{L}(\mathscr{H})$ with $\|B\| < 1$, $I - B$ is invertible and

$$(I-B)^{-1} = \sum_{n=0}^{+\infty} B^n\,, \tag{A.3}$$

it is easy to show that $\rho(A)$ is an open set in $\mathbb{C}$ and that any $\lambda \in \mathbb{C}$ such that $|\lambda| > \|A\|$ is in $\rho(A)$. For $\lambda \in \rho(A)$ we denote

$$R_\lambda(A) = (\lambda I - A)^{-1},$$

it is called the resolvent of A at λ.

The spectrum of A is the complement of the resolvent set

$$\sigma(A) = \rho(A)^C.$$

It is a closed set and can be decomposed into three distinct parts

- **pure point spectrum**, this is the set of eigenvalues of A

$$\sigma_p(A) = \{\lambda \in \mathbb{C}\,;\ \ker(\lambda I - A) \neq \{0\}\}\,;$$

- **continuous spectrum**

$$\sigma_c(A) = \Big\{\lambda \in \mathbb{C}\,;\ \ker(\lambda I - A) = \{0\}\,,\ \operatorname{ran}(\lambda I - A) \neq \mathscr{H}\,,\ \overline{\operatorname{ran}(\lambda I - A)} = \mathscr{H}\Big\}\,;$$

- **residual spectrum**

$$\sigma_r(A) = \Big\{\lambda \in \mathbb{C}\,;\ \ker(\lambda I - A) = \{0\}\,,\ \overline{\operatorname{ran}(\lambda I - A)} \neq \mathscr{H}\Big\}\,.$$

The spectrum of an operator and that of its adjoint are related in the manner described by the following proposition, whose proof is a direct consequence of the definition of the adjoint.

Proposition A.2. *Let $A \in \mathcal{L}(H)$,*

1. $\lambda \in \sigma(A) \Leftrightarrow \bar{\lambda} \in \sigma(A^*)$,
2. *for $\lambda \in \rho(A)$, we have $(R_\lambda(A))^* = R_{\bar{\lambda}}(A^*)$,*
3. $\lambda \in \sigma_r(A) \implies \bar{\lambda} \in \sigma_p(A^*)$,
4. $\lambda \in \sigma_p(A) \implies \bar{\lambda} \in \sigma_p(A^*) \cup \sigma_r(A^*)$,
5. $\lambda \in \sigma_c(A) \Leftrightarrow \bar{\lambda} \in \sigma_c(A^*)$.

In addition, the spectrum of an operator is never empty.

Theorem A.3. *For $A \in \mathscr{L}(\mathscr{H})$, $\sigma(A) \neq \emptyset$.*

Proof. This is due to the fact that the resolvent $R_\lambda(A)$ is analytic on $\rho(A)$ as can easily be seen by developing it into an entire series in the neighbourhood of each point of $\rho(A)$, using nothing more than (A.3). If one assumes that $\sigma(A) = \emptyset$, then $\mathbb{R}_\lambda(A)$ is analytic on $\mathbb{C}$. Now for $|\lambda| > \|A\|$, we have

$$(\lambda I - A)^{-1} = \lambda^{-1}\left(I - \frac{A}{\lambda}\right)^{-1} = \lambda^{-1}\sum_{n=0}^{+\infty}\left(\frac{A}{\lambda}\right)^n$$

which yields

$$\|R_\lambda(A)\| \leq \frac{1}{|\lambda|}\sum_{n=0}^{+\infty}\left(\frac{\|A\|}{|\lambda|}\right)^n = \frac{1}{|\lambda|}\frac{1}{1-\frac{\|A\|}{|\lambda|}} \longrightarrow 0 \text{ as } |\lambda| \to \infty\,.$$

Consequently, for any continuous linear form l on $\mathscr{L}(\mathscr{H})$, $\lambda \mapsto l(R_\lambda(A))$ is holomorphic on $\mathbb{C}$ and tends to 0 at infinity, it is therefore bounded on $\mathbb{C}$; by Liouville's Theorem, this function is constant, the constant being necessarily 0 to ensure the zero limit at infinity. It follows that for all $l \in (\mathscr{L}(\mathscr{H}))'$, we have $l(R_\lambda(A)) \equiv 0$, whence $R_\lambda(A) \equiv 0$, which contradicts its invertibility. □

Moreover the spectrum can be localised in terms of the norm of the operator.

Definition A.8. *Let $A \in \mathcal{L}(H)$, we call spectral radius of A, and we denote $r(A)$, the quantity*

$$r(A) := \sup\{|\lambda|\,;\ \lambda \in \sigma(A)\}\ .$$

Theorem A.4. *Let $A \in \mathcal{L}(H)$, we have*

$$r(A) = \lim_{n\to+\infty}\|A^n\|^{1/n} \leq \|A\|\,. \tag{A.4}$$

This implies in particular that $\sigma(A)$ is compact.

Proof. Firstly, it is clear that

$$\|A^n\|^{1/n} \leq \|A\|$$

hence if the limit (A.4) exists, it is lower than $\|A\|$. Secondly, we prove that

$$r(A) = \limsup_{n\to+\infty}\|A^n\|^{1/n} \tag{A.5}$$

then we shall establish the existence of the limit.

The resolvent $R_\lambda(A)$ is well-defined and analytic on $\{|\lambda| > \|A\|\}$ and for all such λ,

$$R_\lambda(A) = (\lambda - A)^{-1} = \frac{1}{\lambda}\left(1 - \frac{A}{\lambda}\right)^{-1} = \frac{1}{\lambda}\sum_{n=0}^{+\infty}\frac{A^n}{\lambda^n}\,. \tag{A.6}$$

Moreover, by definition of $r(A)$, $\{|\lambda| > r(A)\} \subset \rho(A)$ and this is in fact the largest annulus of the form $\{|z| > C\}$ on which the resolvent is analytic. Since $R_\lambda(A)$ is analytic on $\{|\lambda| > r(A)\}$, it admits a unique expansion as a Laurent series on this domain, given by (A.6). Putting $z = 1/\lambda$, we have

$$R_\lambda(A) = \sum_{n=0}^{+\infty} z^{n+1} A^n$$

and the radius of convergence of this series is

$$R = \frac{1}{\limsup_{n\to+\infty} \|A^n\|^{1/n}} \geq \frac{1}{\|A\|} > 0\,. \tag{A.7}$$

It is in particular analytic in the neighbourhood of 0. We know in addition that it is analytic on $\{0 < |z| < r(A)^{-1}\}$. It is therefore analytic in the disc of center 0 and radius $r(A)^{-1}$ and by the remark above, this is the largest disc centred at the origin on which the series is analytic. It follows that $r(A)^{-1} = R$, i.e. $r(A) = \limsup_{n\to+\infty} \|A^n\|^{1/n}$.

It remains to show that the limit exists. We put $b_n = \log \|A^n\|$. We have $\|A^{n+m}\| \leq \|A^n\|\|A^m\|$ and therefore $b_{n+m} \leq b_n + b_m$. Let $m \in \mathbb{N}^*$ be given, for $n \geq m$, we can write the Euclidean division of n by m: $n = mq + r$, $0 \leq r \leq m-1$. Since $b_{mq} \leq qb_m$, it follows that $b_n \leq qb_m + b_r$ and dividing by n,

$$\frac{b_n}{n} \leq \frac{q}{n} b_m + \frac{b_r}{n}\,.$$

As $n \to +\infty$, $q/n \to 1/m$ and $b_r/n \to 0$ since b_r on takes a finite number of values, namely $b_0, ..., b_{m-1}$. We infer that

$$\limsup_{n\to+\infty} \frac{b_n}{n} \leq \limsup_{n\to+\infty} \left(\frac{q}{n} b_m + \frac{b_r}{n}\right) = \lim_{n\to+\infty} \left(\frac{q}{n} b_m + \frac{b_r}{n}\right) = \frac{b_m}{m}\,.$$

One can therefore write

$$\inf_{n>0} \frac{b_n}{n} \leq \liminf_{n\to+\infty} \frac{b_n}{n} \leq \limsup_{n\to+\infty} \frac{b_n}{n} \leq \frac{b_m}{m}$$

and since m is arbitrary, we have

$$\inf_{n>0} \frac{b_n}{n} \leq \liminf_{n\to+\infty} \frac{b_n}{n} \leq \limsup_{n\to+\infty} \frac{b_n}{n} \leq \inf_{m>0} \frac{b_m}{m}\,.$$

Hence, the limit of b_n/n as $n \to +\infty$ exists and since $b_n/n = \log(\|A^n\|^{1/n})$, the result follows. $\square$

In the case where A is a self-adjoint operator, the spectral radius can be characterised more explicitely

Corollary A.1. *Let $A \in \mathcal{L}(H)$ be a self-adjoint operator, then $r(A) = \|A\|$.*

Proof. Since the sequence $\|A^n\|^{1/n}$ tends to $r(A)$, we only need to show that one of its subsequences converges towards $\|A\|$. On the one hand,

$$\|A^*A\| = \sup_{\|x\|=1} \langle A^*Ax, x\rangle = \sup_{\|x\|=1} \|Ax\|^2 = \|A\|^2\,,$$

one the other hand A is self-adjoint and therefore $\|A^*A\| = \|A^2\|$. Hence, for all $n \in \mathbb{N}$, $\|A^{2^n}\| = \|A\|^{2^n}$. It follows that $r(A) = \|A\|$. $\square$

We have more precise informations on the spectrum of self-adjoint operators.

Theorem A.5. *Let $A \in \mathcal{L}(H)$ be self-adjoint, then*

1. $\sigma(A) \subset \mathbb{R}$,
2. $\sigma_r(A) = \emptyset$,
3. *the eigenfunctions of A associated to different eigenvalues are orthogonal.*

Proof.

- We begin by showing that $\sigma_p(A) \subset \mathbb{R}$. Let $\lambda \in \sigma_p(A)$ and $x \in H$, $x \neq 0$ such that $Ax = \lambda x$, then
$$\langle Ax, x\rangle = \lambda\langle x, x\rangle = \langle x, Ax\rangle = \bar{\lambda}\langle x, x\rangle$$
whence $\left(\lambda - \bar{\lambda}\right)\|x\|^2 = 0$ and $\|x\| \neq 0$, d'où $\lambda \in \mathbb{R}$.

- We infer easily that $\sigma_r(A) = \emptyset$, indeed, let $\lambda \in \sigma_r(A)$, then $\bar{\lambda} \in \sigma_p(A^*) = \sigma_p(A) \subset \mathbb{R}$. We therefore have $\lambda \in \mathbb{R}$ and $\lambda \in \sigma_p(A)$, which is absurd since $\sigma_r(A)$ and $\sigma_p(A)$ are disjoint. Therefore $\sigma_r(A) = \emptyset$.

- Let us now show that $\sigma_c(A) \subset \mathbb{R}$. Let $\lambda = \alpha + i\beta$ with $\beta \neq 0$. We assume that $\lambda \in \sigma_c(A)$. Then $\lambda I - A$ is injective and its range is dense in H but distinct from H. We show that in fact $\mathrm{Im}(\lambda I - A)$ is closed in H, which contradicts the assumptions. To this end, we first prove a useful inequality. Let $x \in H$,
$$\begin{aligned}\|(\lambda - A)x\|^2 &= \langle(\alpha + i\beta - A)x, (\alpha + i\beta - A)x\rangle \\ &= \|(\alpha - A)x\|^2 + \|\beta x\|^2 + 2\Re\langle i\beta x, (\alpha - A)x\rangle\end{aligned}$$
and,
$$\langle i\beta x, (\alpha - A)x\rangle = i\langle\beta, (\alpha - A)x\rangle = i\beta\alpha\|x\|^2 - i\beta\langle x, Ax\rangle \in i\mathbb{R}\,,$$
whence
$$\|(\lambda - A)x\|^2 = \|(\alpha - A)x\|^2 + \|\beta x\|^2 \geq \beta^2\|x\|^2\,. \tag{A.8}$$
Let us consider a Cauchy sequence $(x_n)_n$ in $\mathrm{Im}(\lambda I - A)$, denote x its limit H. For each n, there exists $y_n \in H$ such that $x_n = (\lambda I - A)y_n$ and using (A.8), we have
$$\|x_m - x_n\|^2 = \|(\lambda - A)(y_m - y_n)\|^2 \geq \beta^2\|y_m - y_n\|^2\,.$$
It follows that $\{y_n\}_n$ is a Cauchy sequence, hence convergent. Let y be its limit, by continuity of $(\lambda - A)$, we have $x = (\lambda - A)y \in \mathrm{Im}(\lambda - A)$. Hence the result. The assumption $\lambda \in \mathbb{C} \setminus \mathbb{R}$ is therefore incompatible with $\lambda \in \sigma_c(A)$, i.e. $\sigma_c(A) \subset \mathbb{R}$.

- Let $\lambda, \mu \in \sigma_p(A)$, $\lambda \neq \mu$. We consider $x, y \in H$, $x \neq 0$, $y \neq 0$ such that $Ax = \lambda x$, $Ay = \mu y$. Then, using the fact that $\mu \in \mathbb{R}$,
$$\begin{aligned}\langle Ax, y\rangle &= \langle\lambda x, y\rangle = \lambda\langle x, y\rangle \\ &= \langle x, Ay\rangle = \langle x, \mu y\rangle = \mu\langle x, y\rangle\end{aligned}$$
and since $\lambda \neq \mu$, it follows that $\langle x, y\rangle = 0$.

$\square$

The resolvent has a straightforward but important property.

Proposition A.3 (The resolvent identity). *Let $A \in \mathscr{L}(\mathscr{H})$, $\lambda, \mu \in \rho(A)$, we have*

$$R_\lambda(A) - R_\mu(A) = (\mu - \lambda) R_\lambda(A) R_\mu(A). \tag{A.9}$$

Proof. $R_\lambda(A) - R_\mu(A) = R_\lambda(A)\underbrace{(\mu I - A)R_\mu(A)}_{=I} - \underbrace{R_\lambda(A)(\lambda I - A)}_{=I}R_\mu(A) = (\mu - \lambda)R_\lambda(A)R_\mu(A).$ □

Remark A.3. *Note that the resolvent identity entails that for $\lambda, \mu \in \mathbb{C}$, $R_\lambda(A)$ and $R_\mu(A)$ commute. Indeed, taking the sum of the two equalities*

$$\begin{aligned} R_\lambda(A) - R_\mu(A) &= (\mu - \lambda) R_\lambda(A) R_\mu(A) \\ R_\mu(A) - R_\lambda(A) &= (\lambda - \mu) R_\mu(A) R_\lambda(A) \end{aligned}$$

we obtain

$$0 = (\mu - \lambda)\left(R_\lambda(A) R_\mu(A) - R_\mu(A) R_\lambda(A)\right)$$

and since $\mu - \lambda \neq 0$ we have the result.

Besides resolvents, we need to consider functions of operators as we shall see shortly. This is referred to as the functional calculus. Let $A \in \mathscr{L}(\mathscr{H})$ be a self-adjoint operator, for any polynomial P, the operator $P(A)$ is defined naturally by replacing the variable of the polynomial by A. We have the following important property.

Lemma A.1. *Le P be a polynomial with complex coefficients and $A \in \mathscr{L}(\mathscr{H})$ a self-adjoint operator, then $\sigma(P(A)) = P(\sigma(A))$ and*

$$\|P(A)\|_{\mathscr{L}(\mathscr{H})} = \sup_{\lambda \in \sigma(A)} |P(\lambda)|\,.$$

Proof. Let $\lambda \in \mathbb{C}$, we have

$$P(A) - \lambda = C(A - z_1)(A - z_2)...(A - z_n)$$

where $z_1, ..., z_n$ are the roots of $P(z) - \lambda$ counted with multiplicity. We see that $\lambda \in \sigma(P(A))$ if and only if one of the roots belongs to the spectrum of A, i.e. there exists $\mu \in \sigma(A)$ such that $P(\mu) - \lambda = 0$. This proves the first statement. For the second statement, recall that for any operator $B \in \mathscr{L}(\mathscr{H})$, we have

$$\|B\|^2 = \|B^* B\|\,.$$

Applying this to $P(A)$ and using the fact that A is self-adjoint, we have

$$P(A)^* = \bar{P}(A^*) = \bar{P}(A)\,,$$

whence

$$P(A)^* P(A) = \bar{P}P(A)\,.$$

Since $\bar{P}P(A)$ is self-adjoint, its norm is equal to its spectral radius, i.e.

$$\|P(A)\|^2 = \|\bar{P}P(A)\| = \sup_{\lambda \in \sigma(A)} |\bar{P}P(\lambda)| = \sup_{\lambda \in \sigma(A)} |\bar{P}(\lambda)P(\lambda)| = \sup_{\lambda \in \sigma(A)} |P(\lambda)|^2\,.$$

□

By the Stone-Weierstrass Theorem, we can approach uniformly any continuous function f on $\sigma(A)$ by a sequence of polynomials. Hence, the previous result extends to continuous functions. We have a more complete result, referred to as the spectral theorem for bounded self-adjoint operators in its continuous functional calculus form.

Theorem A.6 (Continuous functional calculus)**.** *Let $A \in \mathscr{L}(\mathscr{H})$ be self-adjoint, there exists a unique map $\Phi : \mathscr{C}(\mathbb{R}) \to \mathscr{L}(\mathscr{H})$ such that:*

1. *Φ is an algebraic $*$-homomorphism, i.e.*
   - $\Phi(f+g) = \Phi(f) + \Phi(g)$, $\Phi(\alpha f) = \alpha \Phi(f)$, $\alpha \in \mathbb{C}$,
   - $\Phi(fg) = \Phi(f) \circ \Phi(g)$,
   - $\Phi(1) = \mathrm{Id}_H$,
   - $\Phi(\bar{f}) = (\Phi(f))^*$;
2. *Φ is continuous and*
$$\|\Phi(f)\|_{\mathscr{L}(\mathscr{H})} = \sup_{\lambda \in \sigma(A)} |f(\lambda)| = \|f\|_{L^\infty(\sigma(A))};$$
3. *if f is the identity, i.e. $f(\lambda) = \lambda$ for all $\lambda \in \sigma(A)$, then $\Phi(f) = A$;*
4. *if $Ax = \lambda x$, then, $\Phi(f)x = f(\lambda)x$;*
5. *$\sigma(\Phi(f)) = f(\sigma(A))$, i.e. $\lambda \in \sigma(\Phi(f))$ if and only if there exists $\mu \in \sigma(A)$ such that $\lambda = f(\mu)$;*
6. *if $f \geq 0$, then the operator $\Phi(f)$ is non negative, i.e. $\langle \Phi(f)x, x\rangle \geq 0$ for all $x \in \mathscr{H}$.*

We shall denote $\Phi(f) =: f(A)$.

Proof. Let us start with uniqueness : *1.* and *3.* imply that for $f = P$ a polynomial, $\Phi(P)$ coincides with $P(A)$. Now let $f \in \mathcal{C}(\sigma(A))$, by the Stone-Weierstrass, there exists a sequence $\{P_n\}$ of polynomials such that $P_n \to f$ uniformly on $\sigma(A)$. Les us assume that we have two functions Φ_1 and Φ_2 that satisfy the theorem's list of properties. By the remark above, we have

$$\forall n \in \mathbb{N}\,,\ \Phi_1(P_n) = \Phi_2(P_n) = P(A)\,.$$

Moreover,

$$\|\Phi_1(f) - \Phi_2(f)\| \leq \|\Phi_1(f) - \Phi_1(P_n)\| + \|\Phi_2(P_n) - \Phi_2(f)\| = 2\|f - P_n\|_{\mathcal{C}(\sigma(A))} \to 0\,.$$

Hence $\Phi_1 = \Phi_2$.

We now turn to existence. Let $f \in \mathcal{C}(\sigma(A))$, and $\{P_n\}$ a sequence of polynomials that converges uniformly towards f on $\sigma(A)$. We put

$$\Phi(f) := \lim_{n \to +\infty} P_n(A)\,.$$

We must show that this is indeed a definition, i.e. that the limit is independent of the choice of the sequence $\{P_n\}$. Let us consider another such sequence $\{Q_n\}$, then

$$\|P_n(A) - Q_n(A)\| = \|P_n - Q_n\|_{\mathcal{C}(\sigma(A))} \to 0\,.$$

Moreover the limit exists in $\mathcal{L}(H)$, since

$$\|P_n(A) - P_m(A)\| = \|P_n - P_m\|_{\mathcal{C}(\sigma(A))}\,.$$

Since $\{P_n\}$ is a Cauchy sequence in $\mathcal{C}(\sigma(A))$, it follows that $\{P_n(A)\}$ is a Cauchy sequence in $\mathcal{L}(H)$.

Then *1.*, *2.*, *3.* are satisfied by construction and *4.* is clearly true for any polynomial and extends by density. Let us now prove *6.*: Let $f \in \mathcal{C}(\sigma(A))$, $f \geq 0$, then we can write f under the form $f = g^2$, where g is real valued and continuous on $\sigma(A)$. Since g is real valued, *1.* entails that $\Phi(g)$ is self-adjoint. Still using *1.*, we have $\Phi(f) = \Phi(g^2) = \Phi(g)^2$. Whence, for all $x \in H$

$$\langle \Phi(f)x, x\rangle = \langle \Phi(g)^2 x, x\rangle = \langle \Phi(g)x, \Phi(g)x\rangle = \|\Phi(g)x\|^2 \geq 0\,.$$

It remains to establish the spectral image property *5.*. We denote

$$F = \{f(\lambda)\,;\ \lambda \in \sigma(A)\}\ .$$

Considering $\lambda_0 \in \mathbb{C} \setminus F$ and the function $g \in \mathcal{C}(\sigma(A))$ defined by

$$g(\lambda) := \frac{1}{f(\lambda) - \lambda_0}\,.$$

We have $(f(\lambda) - \lambda_0)g(\lambda) = g(\lambda)(f(\lambda) - \lambda_0) \equiv 1$ and therefore

$$(f(A) - \lambda_0 I)g(A) = g(A)(f(A) - \lambda_0 I) = I$$

which proves that $\lambda_0 I - f(A)$ is invertible and that $\lambda_0 \notin \sigma(f(A))$. Hence, $\sigma(f(A)) \subset F$. We show the converse inclusion. Let $\lambda_0 \in F$, there exists $\mu \in \sigma(A)$ such that $\lambda_0 = f(\mu)$. There are two cases :

- $\ker(A - \mu I) \neq \{0\}$. Then there exists $x \neq 0$ such that $Ax = \mu x$. Then $f(A)x = f(\mu)x$ and therefore $\lambda_0 \in \sigma_p(f(A))$.

- $\mu \in \sigma_c(A)$ (remember that the residual spectrum of a self-adjoint operator is empty). Then the image of $A - \mu I$ is dense in H but distinct from H. The operator $(A - \mu I)^{-1}$ is defined on $\mathrm{Im}(A - \mu I)$ but is not bounded, otherwise, it would extend to a continuous operator on H that would be the inverse of $A - \mu I$ and this would contradict $\mu \in \sigma(A)$. Therefore, there exists a sequence $\{x_n\}$ in $\mathrm{Im}(A - \mu I)$, $\|x_n\| = 1$, such that $\|(A - \mu I)^{-1}x_n\| \to +\infty$. Hence, putting
$$y_n = \frac{(A - \mu I)^{-1}x_n}{\|(A - \mu I)^{-1}x_n\|}\,,$$

which has a meaning for n large enough, because $\|(A-\mu I)^{-1}x_n\| \to +\infty$ and therefore $(A-\mu I)^{-1}x_n \neq 0$ for n large enough, we see that $\|y_n\| = 1$ and moreover

$$\|(A-\mu I)y_n\| = \frac{\|x_n\|}{\|(A-\mu I)^{-1}x_n\|} = \frac{1}{\|(A-\mu I)^{-1}x_n\|} \to 0\ , \quad \text{lorsque } n \to +\infty\,.$$

For $p \in \mathbb{N}$ arbitrary, we have

$$\begin{aligned} \|(A^p-\mu^p)y_n\| &= \left\|\left(A^{p-1}+A^{p-2}\mu+\ldots+A\mu^{p-2}+\mu^{p-1}\right)(A-\mu)y_n\right\| \\ &\leq \left\|A^{p-1}+A^{p-2}\mu+\ldots+A\mu^{p-2}+\mu^{p-1}\right\| \|(A-\mu)y_n\| \to 0\,. \end{aligned}$$

We infer that for any polynomial P, we have

$$\|(P(A)-P(\mu))y_n\| \to 0\ , \quad \text{lorsque } n \to +\infty\,.$$

Let $\varepsilon > 0$, there exists a polynomial P such that $\|f(A)-P(A)\| = \|f-P\|_{\mathcal{C}(\sigma(A))} < \varepsilon/3$ and there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, we have $\|(P(A)-P(\mu))y_n\| < \varepsilon/3$. Then for $n \geq n_0$,

$$\begin{aligned} \|(f(A)-f(\mu))y_n\| &\leq \|(f(A)-P(A))y_n\| + \|(P(A)-P(\mu))y_n\| \\ &\quad + \|(P(\mu)-f(\mu))y_n\| < \varepsilon\,. \end{aligned}$$

We have constructed a sequence $\{y_n\}$, any element of which has norm 1 and whose image by $f(A)-f(\mu) = f(A)-\lambda_0$ tends towards 0. This cannot happen unless $\lambda_0 \in \sigma(f(A))$. Indeed, if $\lambda_0 \in \rho(f(A))$, then

$$1 = \|y_n\| = \left\|(f(A)-\lambda_0)^{-1}(f(A)-\lambda_0)y_n\right\| \leq \left\|(f(A)-\lambda_0)^{-1}\right\| \|(f(A)-\lambda_0)y_n\| \to 0$$

which is absurd.

□

A.5 Unbounded self-adjoint operators and perturbations thereof

Unbounded operators are the natural objects one has to deal with when studying linear partial differential equation. Some results that we have seen for bounded operators extend to unbounded operators without major modification, such as the functional calculus for self-adjoint operators (provided we work with bounded measurable functions), but the proofs become more involved and use some notions that we prefer not to describe here in order to keep the length of this appendix under control. So a few results in this section and the next will be given without proof. Once again, we refer the interested reader to the references mentioned above (mainly [47, 48]).

Definition A.9. *An unbounded operator on $\mathscr{H}$ is the data of a subspace $D(A)$ of $\mathscr{H}$ and of a linear map $A : D(A) \to \mathscr{H}$. It is usually denoted $(A, D(A))$. The space $D(A)$ is called the domain of A.*

Remark A.4. *Of course, if $D(A)$ is dense in $\mathscr{H}$ and A is bounded on $D(A)$ for the norm in $\mathscr{H}$, then A extends uniquely as a bounded operator on $\mathscr{H}$. The examples we are interested in are those where $D(A)$ is dense in $\mathscr{H}$ and A is not bounded on $D(A)$, i.e. not continuous for the topology of $\mathscr{H}$.*

Example. *Let*

$$\mathscr{H} = L^2(\mathbb{R})\,,\ D(A) = H^1(\mathbb{R})\,,\ A = \frac{d}{dx}\,.$$

This operator is unbounded, otherwise, there would exist $C > 0$ such that for all $f \in H^1(\mathbb{R})$,

$$\left\| f' \right\|_{L^2(\mathbb{R})} \leq C \left\| f \right\|_{L^2(\mathbb{R})}\,,$$

i.e. $L^2(\mathbb{R}) \hookrightarrow H^1(\mathbb{R})$, which is of course wrong.

Definition A.10. *An operator $(A, D(A))$ is called symmetric if for all $u, v \in D(A)$,*

$$\langle Au, v \rangle = \langle u, Av \rangle\,.$$

Definition A.11 (Adjoint)**.** *Let $(A, D(A))$ be an operator with dense domain (i.e. $D(A)$ dense in $\mathscr{H}$). The adjoint of $(A, D(A))$ is the operator $(A^*, D(A^*))$ defined by:*

$$D(A^*) = \{y \in \mathscr{H}\,;\ \exists z \in \mathscr{H}\,;\ \forall x \in D(A)\,,\ \langle Ax, y\rangle = \langle x, z\rangle\}$$

and

$$A^* y = z\,.$$

Remark A.5.

1. *The fact that $D(A)$ is dense in $\mathscr{H}$ ensures that if z exists, it is unique, for the difference of two such z's must be orthogonal to $D(A)$ and therefore has to vanish.*

2. *An operator is symmetric with dense domain if and only if*

$$D(A) \subset D(A^*) \text{ and } A^*|_{D(A)} = A\,. \tag{A.10}$$

   The property (A.10) exactly means that A^* is an extension of A. It is often written simply as

$$A \subset A^*.$$

3. *Let A be a symmetric operator with dense domain and let B be a symmetric extension of A, then we have*

$$A \subset B \subset A^*.$$

   Indeed, for all $x, y \in D(B)$, we have $\langle Bx, y\rangle = \langle x, By\rangle$ and in particular for all $x \in D(A)$ and $y \in D(B)$, $\langle Ax, y\rangle = \langle x, By\rangle$. By the definition of $D(A^*)$ and A^*, it follows that $y \in D(A^*)$ and that $A^* y = By$.

Definition A.12 (Self-adjoint operator)**.** *An operator $(A, D(A))$ is said to be self-adjoint if it satisfies the following three properties:*

1. *$D(A)$ is dense in $\mathscr{H}$;*
2. *A is symmetric;*
3. *$D(A^*) = D(A)$.*

Remark A.6. *By point 3. of Remark A.5, we see that a self-adjoint operator does not have any symmetric extension but itself.*

Examples.

1. The operator $D = \frac{1}{i}\frac{\mathrm{d}}{\mathrm{d}x}$ is self-adjoint on $L^2(\mathbb{R})$ with domain $H^1(\mathbb{R})$.
2. The operator
$$-\Delta = -\sum_{i=1}^{n} \frac{\partial^2}{\partial (x^i)^2}$$
is self-adjoint on $L^2(\mathbb{R}^n)$ with domain $H^2(\mathbb{R}^n)$ and is non negative, i.e. for all $u \in H^2(\mathbb{R}^n)$, $\langle -\Delta u, u\rangle \geq 0$.

As we have seen above, when an operator is symmetric with dense domain, we have $D(A) \subset D(A^*)$. However it is often difficult to calculate $D(A^*)$ explicitly. Luckily, there exists a fundamental criterion for the self-adjointness of such an operator that does not require a precise determination of $D(A^*)$. This criterion requires the notion of a closed operator.

Definition A.13 (Closed operator)**.** *Given an operator $(A, D(A))$, its graph Γ_A is defined as*
$$\Gamma_A := \{(x, Ax)\,;\ x \in D(A)\}\ .$$
We say that A is closed if Γ_A is closed in $\mathscr{H} \times \mathscr{H}$.

Note that for a closed operator, the domain of A, endowed with the graph norm
$$\|u\|^2_{D(A)} = \|u\|^2 + \|Au\|^2 \tag{A.11}$$
has the structure of a Hilbert space. We have the continuous embedding $D(A) \hookrightarrow \mathscr{H}$ and $\lambda I - A$ is continuous from $D(A)$ to $\mathscr{H}$. By Banach's isomorphism Theorem, $\lambda I - A$ is therefore an isomorphism from $D(A)$ to $\mathscr{H}$ if and only if it is bijective.

Any densely defined operator is such that its adjoint is a closed operator. This can be seen by considering the unitary isomorphism of $\mathscr{H} \times \mathscr{H}$ defined by
$$\mathscr{U}(x, y) = (-y, x)\,. \tag{A.12}$$
Since $\mathscr{U}$ is unitary (i.e. $\mathscr{U}\mathscr{U}^* = \mathscr{U}^*\mathscr{U} = I$), for any subspace F of $\mathscr{H}$, we have
$$(\mathscr{U}(F))^\perp = \mathscr{U}\left(F^\perp\right)\,.$$
Let $(x, y) \in \mathscr{H} \times \mathscr{H}$, then $(x, y) \in \mathscr{U}\left(\Gamma_A^\perp\right)$ if and only if for any $u \in D(A)$ we have
$$\langle (x, y), (-Au, u)\rangle = 0\,,\ \text{i.e.} \langle x, Au\rangle = \langle y, u\rangle\,,$$
which is equivalent to $x \in D(A^*)$ and $y = A^*x$. We have therefore proved the following result.

Proposition A.4. *Let A be a densely defined operator of $\mathscr{H}$, then*

$$\Gamma_{A^*} = (\mathscr{U}(\Gamma_A))^{\perp}$$

and is therefore a closed subspace of $\mathscr{H} \times \mathscr{H}$, i.e. the adjoint of A is a closed operator. If in addition A is symmetric, then A^* is a closed extension of A.

We are now ready to state and prove the basic criterion.

Theorem A.7 (Basic criterion for self-adjointness)**.** *Let $(A, D(A))$ be a symmetric operator with dense domain, the following properties are equivalent:*

1. $(A, D(A))$ *is self-adjoint;*
2. $(A, D(A))$ *is closed and* $\ker(A^* \pm iI) = \{0\}$*;*
3. $\operatorname{ran}(A \pm iI) = \mathscr{H}$*.*

Proof.

- *1. $\Rightarrow$ 2.* We have already established that if A is self-adjoint, then A is closed since it is equal to its adjoint. Now consider $x \in D(A^*) = D(A)$ such that $Ax = \pm ix$, then we have

$$\langle Ax, x \rangle = \pm i \|x\|^2;$$

the left-hand side is real and the right-hand side pure imaginary. It follows that $x = 0$.

- *2. $\Rightarrow$ 3.* Let us first establish that $\operatorname{ran}(A \pm iI)$ is closed. Let $(y_n)_n$ be a sequence in $\operatorname{ran}(A \pm iI)$ that converges in $\mathscr{H}$. For each n, there exists $x_n \in D(A)$ such that $y_n = (A \pm iI)x_n$ and

$$\|(A \pm iI)(x_m - x_n)\|^2 = \|A(x_m - x_n)\|^2 + \|x_m - x_n\|^2$$

since $\langle A(x_m - x_n), i(x_m - x_n) \rangle + \langle i(x_m - x_n), A(x_m - x_n) \rangle = 0$ by symmetry of A. This implies that both $(x_n)_n$ and $(Ax_n)_n$ are Cauchy sequences in $\mathscr{H}$, let x and z be their respective limits. Since the graph of A is closed, we have $(x, z) \in \Gamma_A$, i.e. $z = Ax$, in other words, the limit of y_n is $Ax \pm ix \in \operatorname{ran}(A \pm iI)$. The range of $A \pm iI$ is therefore closed in $\mathscr{H}$. We now need to prove that it is dense in $\mathscr{H}$, which is equivalent to proving that its orthogonal is trivial. Saying that $x \in (\operatorname{ran}(A \pm iI))^{\perp}$ means that for all $y \in D(A)$, $\langle Ay, x \rangle = \mp i \langle y, x \rangle = \langle y, \pm ix \rangle$ and this is exactly equivalent to saying that $x \in D(A^*)$ and $A^* x = \pm ix$. So we see that

$$(\operatorname{ran}(A \pm iI))^{\perp} = \ker(A^* \mp iI)$$

and this is trivial by assumption.

- *3. $\Rightarrow$ 1.* We only need to prove that $D(A^*) \subset D(A)$. Let $x \in D(A^*)$, since $\operatorname{ran}(A - iI) = \mathscr{H}$, there exists $y \in D(A)$ such that $Ay - iy = A^* x - ix$. Using the fact that $D(A) \subset D(A^*)$ and $A = A^*$ on $D(A)$, we have $y - x \in D(A^*)$ and $(A^* - iI)(y - x) = 0$, i.e.

$$y - x \in \ker(A^* - iI) = (\operatorname{ran}(A + iI))^{\perp} = \{0\}.$$

It follows that $x \in D(A)$ which concludes the proof.

$\square$

It is important, especially when studying scattering theory, to know under which conditions a perturbation of a self-adjoint operator remains self-adjoint. The Kato-Rellich Theorem provides a sufficient condition that is very useful. We need to introduce the notion of relative bound.

Definition A.14. *Let $(A, D(A))$ and $(B, D(B))$ be two operators. We say that B is relatively A-bounded if $D(B) \supset D(A)$ and there exist non-negative numbers a, b such that*

$$\|Bx\| \le a\|Ax\| + b\|x\| \ \forall x \in D(A)\,. \tag{A.13}$$

The infimum of all $a \ge 0$ for which there exists $b \ge 0$ such that (A.13) holds is called the relative A-bound of B and is denoted $\alpha_A(B)$.

Theorem A.8 (Kato-Rellich)**.** *Let $(A, D(A))$ be a self-adjoint operator. Let $(B, D(B))$ be a symmetric operator that is relatively A-bounded with $\alpha_A(B) < 1$, then the operator $(A+B, D(A))$ is self-adjoint.*

Example. Consider $V \in L^\infty(\mathbb{R}^n)$, then $-\Delta + V$ is self-adjoint on $L^2(\mathbb{R}^n)$ with domain $H^2(\mathbb{R}^n)$.

The spectrum of a closed operator is defined in terms of its restriction to its domain, endowed with the graph norm, as follows.

Definition A.15 (Spectrum)**.** *Let $(A, D(A))$ be a closed operator, the resolvent set of A, denoted $\rho(A)$, is the set of $\lambda \in \mathbb{C}$ such that*

$$\lambda I - A \,:\; D(A) \longrightarrow \mathscr{H}$$

is an isomorphism, where $D(A)$ is endowed with the graph norm (A.11). For $\lambda \in \rho(A)$, we denote $R_\lambda(A) := (\lambda I - A)^{-1}$, it is called the resolvent of A at λ. The spectrum of A is the complement of its resolvent set: $\sigma(A) := \mathbb{C} \setminus \rho(A)$.

Proposition A.5. *Let $(A, D(A))$ be a closed operator, for $\lambda \in \rho(A)$, $R_\lambda(A) \in \mathscr{L}(\mathscr{H})$ simply because $R_\lambda(A) \in \mathscr{L}(\mathscr{H}, D(A))$ and $D(A) \hookrightarrow \mathscr{H}$. Moreover, $\rho(A)$ is an open set in $\mathbb{C}$ (i.e. $\sigma(A)$ is a closed set in $\mathbb{C}$) in which $\lambda \mapsto R_\lambda(A)$ is analytic with values in $\mathscr{L}(\mathscr{H})$ and satisfies the resolvent identity.*

Let $(A, D(A))$ be a closed operator, similarly to a bounded operator, its spectrum can be decomposed into three distinct parts:

- **pure point spectrum**: $\sigma_p(A)$ is the set of $\lambda \in \mathbb{C}$ such that $\ker(\lambda I - A)\} \neq \{0\}$;
- **continuous spectrum**: $\sigma_c(A)$ is the set of $\lambda \in \mathbb{C}$ such that $\lambda I - A$ is injective and its range is dense in $\mathscr{H}$ but distinct from $\mathscr{H}$, i.e. $(\lambda I - A)^{-1}$ is an unbounded operator with dense domain $\mathrm{ran}(\lambda I - A)$;
- **residual spectrum**: $\sigma_r(A)$ is the set of $\lambda \in \mathbb{C}$ such that $\lambda I - A$ is injective and its range is not dense in $\mathscr{H}$, i.e. $(\lambda I - A)^{-1}$ is an unbounded operator with non dense domain $\mathrm{ran}(\lambda I - A)$.

Theorem A.9. *The spectrum of a self-adjoint operator is real.*

Proof. This follows easily from the basic criterion. Let $\lambda = \alpha + i\beta$, $\beta \neq 0$. We have

$$A - \lambda I = A - \alpha I - i\beta I = \beta \left(\frac{1}{\beta} (A - \alpha I) - iI \right).$$

By the Kato-Rellich Theorem, $B = A - \alpha I$ is self-adjoint with domain $D(A)$ and so is $\frac{1}{\beta}B$. Therefore $\frac{1}{\beta}B - iI$ is an isomorphism from $D(A)$ to $\mathscr{H}$ by the basic criterion. □

A.6 Spectral theory of unbounded self-adjoint operators

For unbounded self-adjoint operators, we have a functional calculus that ressembles that for bounded operators, but using bounded functions that ensure that the images of the operators are bounded.

Theorem A.10 (Spectral theorem for unbounded self-adjoint operators – bounded Borel functional calculus)**.** *Let A be a self-adjoint operator on $\mathscr{H}$, there exists a unique linear map Φ from bounded Borel functions on $\mathbb{R}$ to $\mathscr{L}(\mathscr{H})$ such that:*

1. *Φ is an algebraic $*$-homomorphism;*
2. *$\|\Phi(f)\|_{\mathscr{L}(\mathscr{H})} \leq \sup_{\lambda\in\sigma(A)} |f(\lambda)|$;*
3. *let $(f_n)_{n\in\mathbb{N}}$ be a sequence of bounded Borel functions such that for all $x \in \mathbb{R}$ $f_n(x) \to x$ and $|f_n(x)| \leq |x|$ for all $x \in \mathbb{R}$ and $n \in \mathbb{N}$, then for all $u \in D(A)$, $\Phi(f_n)u \to Au$;*
4. *let $(f_n)_{n\in\mathbb{N}}$ be a sequence of bounded Borel functions that converges pointwise to a bounded Borel function f and such that the sequence $\sup_{x\in\mathbb{R}} |f_n(x)|$ is bounded, then $\Phi(f_n) \to \Phi(f)$ strongly, i.e. $\Phi(f_n)u \to \Phi(f)u$ for all $u \in \mathscr{H}$;*
5. *if $Ax = \lambda x$, then, $\Phi(f)x = f(\lambda)x$;*
6. *$\sigma(\Phi(f)) = f(\sigma(A))$, i.e. $\lambda \in \sigma(\Phi(f))$ if and only if there exists $\mu \in \sigma(A)$ such that $\lambda = f(\mu)$;*
7. *if $f \geq 0$, then the operator $\Phi(f)$ is non negative, which means that for all $u \in H$, we have $\langle \Phi(f)u, u\rangle \geq 0$.*

We shall denote $\Phi(f) =: f(A)$.

An important function to consider as an example of application of the functional calculus is $f(x) = e^{itx}$.

Theorem A.11. *Let A be a self-adjoint operator on $\mathscr{H}$, then the family of operators*

$$\{\mathcal{U}(t) = e^{itA}\}_{t\in\mathbb{R}}$$

is a strongly continuous one-parameter group of unitary operators, i.e.

1. *$\mathcal{U}(t)$ is a unitary operator on $\mathscr{H}$ for all $t \in \mathbb{R}$ and $\mathcal{U}(0) = I_{\mathscr{H}}$,*

2. *for all* $\phi \in \mathscr{H}$, $\mathcal{U}(t)\phi \in \mathscr{C}(\mathbb{R}_t\,;\ \mathscr{H})$,

3. *for all* $t, s \in \mathbb{R}$, $\mathcal{U}(t)\mathcal{U}(s) = \mathcal{U}(t+s)$ *(and in particular* $\mathcal{U}(t)\mathcal{U}(s) = \mathcal{U}(s)\mathcal{U}(t)$*).*

Moreover, it satisfies the following properties

4. *for* $\psi \in D(A)$,
$$\lim_{t\to 0} \frac{\mathcal{U}(t)\psi - \psi}{t} = iA\psi\,,$$

5. *if*
$$\lim_{t\to 0} \frac{\mathcal{U}(t)\psi - \psi}{t}$$
exists, then $\psi \in D(A)$.

Proof. The first three properties follow easily from the functional calculus.

3. Since Φ is an algebraic $*$-homomorphism, we have
$$e^{itA}e^{isA} = \Phi(e^{itx})\Phi(e^{isx}) = \Phi(e^{itx}e^{isx}) = \Phi(e^{i(t+s)x}) = e^{i(t+s)A}.$$

1. Invoking the same property, we also have $e^{-itA} = \Phi(e^{-itx}) = \Phi(\overline{e^{itx}}) = (e^{itA})^*$. Putting the two together, we obtain
$$e^{itA}(e^{itA})^* = \Phi(e^{itx}e^{-itx}) = \Phi(1) = I_{\mathscr{H}}$$
and similarly for $(e^{itA})^*e^{itA}$.

2. Follows directly from Property *4.* of the functional calculus.

4. We have
$$\left|\frac{e^{itx}-1}{it}\right| \le |x|$$
and
$$\lim_{t\to 0} \frac{e^{itx}-1}{it} = x\,.$$
So for any sequence $t_n \to 0$, we are in the framework of Property *3.* of the functional calculus and this gives the result.

5. We define
$$\begin{aligned} D(B) &:= \left\{\psi \in \mathscr{H}\,;\ \lim_{t\to 0} \frac{\mathcal{U}(t)\psi - \psi}{t} \text{ exists}\right\}, \\ B\psi &:= \frac{1}{i}\lim_{t\to 0} \frac{\mathcal{U}(t)\psi - \psi}{t} \text{ for } \psi \in D(B). \end{aligned}$$
Since $\mathcal{U}(t)^* = \mathcal{U}(-t)$, it follows that the limit itself is a skew-symmetric operator, and B is therefore symmetric. Moreover, $D(B) \supset D(A)$ by Point *4.* above and B coincides with A on $D(A)$. Since A is self-adjoint, it follows from Remark A.6 that $B = A$.

□

We see that the family of operators $\mathcal{U}(t)$ provides solutions to the Cauchy problem for the Schrödinger equation

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = iA\phi$$

provided the initial data are in $D(A)$.

Theorem A.12. *Let A be a self adjoint operator on $\mathscr{H}$, consider the Schrödinger equation*

$$\frac{\mathrm{d}\phi}{\mathrm{d}t} = iA\phi. \tag{A.14}$$

Let $\phi_0 \in D(A)$, then (A.14) admits a unique solution in $\mathscr{C}(\mathbb{R}\,;\ D(A)) \cap \mathscr{C}^1(\mathbb{R}\,;\ \mathscr{H})$ given by $\phi(t) = e^{itA}\phi_0$. This solution satisfies

$$\|e^{itA}\phi_0\|_{\mathscr{H}} = \|\phi_0\|_{\mathscr{H}} \text{ for all } t \in \mathbb{R}.$$

Moreover, if $\phi_0 \in D(A^k)$, $k \in \mathbb{N}^*$, then

$$e^{itA}\phi_0 \in \bigcap_{p=0}^{k} \mathscr{C}^p(\mathbb{R}_t\,;\ D(A^{k-p}))\,.$$

Proof. First of all $e^{itA}\phi_0 \in \mathscr{C}(\mathbb{R}\,;\ \mathscr{H})$ and by point *4.* of Theorem A.11, we see that it is differentiable in time and that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{itA}\phi_0\right) = ie^{itA}A\phi_0 \in \mathscr{C}(\mathbb{R}\,;\ \mathscr{H}).$$

By the functional calculus, we also have that for all t, s,

$$\frac{e^{isA} - I}{s} e^{itA}\phi_0 = e^{itA}\frac{e^{isA} - I}{s}\phi_0$$

and since the right-hand side has a limit in $\mathscr{H}$ as $s \to 0$, it follows from point *5.* of Theorem A.11 that $e^{itA}\phi_0 \in D(A)$ and then point *4.* entails that the left-hand side tends to $iAe^{itA}\phi_0$. Hence $e^{itA}\phi_0$ satisfies (A.14) and $e^{itA}\phi_0 \in \mathscr{C}(\mathbb{R}\,;\ D(A))$. The norm in $\mathscr{H}$ of $e^{itA}\phi_0$ is preserved because e^{itA} is a unitary operator for all t. This concludes the existence part of the proof.

Now consider $\psi \in \mathscr{C}(\mathbb{R}\,;\ D(A)) \cap \mathscr{C}^1(\mathbb{R}\,;\ \mathscr{H})$ a solution to (A.14) such that $\psi(0) = 0$. We have

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle \psi, \psi\rangle = \langle iA\psi, \psi\rangle + \langle \psi, iA\psi\rangle = 0$$

since iA is skew adjoint. Since $\psi(0) = 0$, this entails that ψ is identically zero and establishes the uniqueness of solutions.

The additional regularity is established by a simple bootstrap argument. For $k = 1$ there is nothing new. Assuming $k \geq 2$, we have $A\phi_0 \in D(A^{k-1})$, with $k - 1 \geq 1$ and therefore

$$e^{itA}A\phi_0 \in \mathscr{C}(\mathbb{R}_t\,;\ D(A)) \cap \mathscr{C}^1(\mathbb{R}_t\,;\ \mathscr{H})\,.$$

But since A commutes with e^{itA} on $D(A)$, we get

$$e^{itA}A\phi_0 = Ae^{itA}\phi_0 = A\phi(t) \in \mathscr{C}(\mathbb{R}_t\,;\ D(A)) \cap \mathscr{C}^1(\mathbb{R}_t\,;\ \mathscr{H})\,.$$

This implies that

$$\phi \in \mathscr{C}(\mathbb{R}_t\,;\ D(A^2)) \cap \mathscr{C}^1(\mathbb{R}_t\,;\ D(A))\,. \tag{A.15}$$

And using the equation we also have

$$iA\phi = \frac{\mathrm{d}\phi}{\mathrm{d}t} \in \mathscr{C}(\mathbb{R}_t\,;\ D(A)) \cap \mathscr{C}^1(\mathbb{R}_t\,;\ \mathscr{H})\,. \tag{A.16}$$

Putting (A.15) and (A.16) together, we get

$$\phi \in \mathscr{C}(\mathbb{R}_t\,;\ D(A^2)) \cap \mathscr{C}^1(\mathbb{R}_t\,;\ D(A)) \cap \mathscr{C}^2(\mathbb{R}_t\,,\ \mathscr{H})\,.$$

The result follows by induction. □

Remark A.7. *In some important cases, $\mathscr{H}$ will be a distribution space ($\mathscr{H} = L^2(\mathbb{R}^n)$ for example), and A will be a differential operator. In such cases, it is often possible to gain the existence and uniqueness of solutions to* (A.14) *in $\in \mathscr{C}(\mathbb{R}_t\,;\ \mathscr{H})$ in the sense of distributions; such "minimum regularity" solutions will still have their norm in $\mathscr{H}$ conserved throughout time.*

In fact there is a converse result to that of the theorem above.

Theorem A.13 (Stone's Theorem)**.** *Let $\mathcal{U}(t)$ be a one-parameter group of strongly continuous operators on $\mathscr{H}$, then there exists a self-adjoint operator A such that $\mathcal{U}(t) = e^{itA}$. A is called the infinitesimal generator of $\mathcal{U}(t)$.*

A.7 Grönwall's Lemma

A priori estimates are among the most useful methods in analysis. In many cases, they rely on a fundamental inequality first established by the Swedish mathematician Thomas Hakon Grönwall in 1919 and then generalised by many people. A standard form of the result is the following.

Theorem A.14 (Grönwall's inequality)**.** *Let $\phi : [0,T] \to [0,+\infty]$ and $C : [0,T] \to [0,+\infty]$ be continuous functions and let $B > 0$. Assume that for $t \in [0,T]$ we have*

$$\phi(t) \leq B + \int_0^t C(s)\phi(s)\mathrm{d}s\,.$$

Then for all $t \in [0,T]$, we have

$$\phi(t) \leq B\exp\left(\int_0^t C(s)\mathrm{d}s\right).$$

Bibliography

[1] A. Antoci, D.-E. Liebscher, *On the gravitational field of a mass point according to Einstein's theory by Karl Schwarzschild*, General Relativity and Gravitation, **35** (2003), 5, 945–950.

[2] S. Benzoni-Gavage, D. Serre, *Multidimensional hyperbolic partial differential equations – First order systems and applications*, Oxford University Press, 2007.

[3] A.N. Bernal, M. Sánchez, *On smooth Cauchy hypersurfaces and Geroch's splitting theorem*, Commun. Math. Phys. **243** (2003), p. 461–470.

[4] A.N. Bernal, M. Sánchez, *Smoothness of time functions and the metric splitting of globally hyperbolic space-times*, Commun. Math. Phys. **257** (2005), p. 43–50.

[5] G.D. Birkhoff, *Relativity and modern physics*, Harvard University Press, 1923.

[6] S. Chandrasekhar, The mathematical theory of black holes, Oxford University Press, 1984.

[7] P. Chrusciel & E. Delay, *On mapping properties of the general relativistic constraints operator in weighted function spaces, with applications*, preprint Tours Univervity, 2003.

[8] J. Corvino & R.M. Schoen, *On the asymptotics for the vacuum Einstein constraint equations*, gr-qc 0301071, 2003.

[9] J. Le Rond d'Alembert, *Recherches sur la courbe que forme une corde tendue mise en vibration* (Researches on the curve that a tense cord forms [when] set into vibration), Histoire de l'académie royale des sciences et belles lettres de Berlin, vol. 3 (1747), p. 214–219.

[10] M. Dafermos & I. Rodnianski, *The redshift effect and radiation decay on black hole spacetimes*, Comm. Pure Appl. Math. **62** (2009), 7, 859–919.

[11] J. Dereziński, C. Gérard, *Scattering theory of classical and quantum N-particle systems*, Springer, 1997.

[12] J. Dimock *Scattering for the wave equation on the Schwarzschild metric*, Gen. Rel. Grav. **17** (1985), 4, 353–369.

[13] J. Dimock, B.S. Kay, *Classical and quantum scattering theory for linear scalar fields on the Schwarzschild metric. II*, J. Math. Phys. **27** (1986), 10, 2520–2525.

[14] J. Dimock, B.S. Kay, *Classical and Quantum Scattering theory for linear scalar fields on the Schwarzschild metric I*, Ann. Phys. **175** (1987), 366–426.

[15] M. P. do Carmo, *Differential geometry*, Birkhaüser, 1992.

[16] A. Einstein, *Erklärung der Perihelbewegung des Merkur aus der allgemeinen Relativitätstheorie*, Königlich Preussische Akademie der Wissenschaften (Berlin), Sitzungsberichte 831–835, 1915.

[17] D. Finkelstein, *Past-future asymmetry of the gravitational field of a point particle*, Physical Review **110** (1958), 4, 965–967.

[18] R.P. Geroch, *The domain of dependence*, J. Math. Phys. **11** (1970), 437–439.

[19] J. Hadamard, *Lectures on Cauchy's problem in linear partial differential equations*, Dover, 1952.

[20] D. Häfner, *Sur la théorie de la diffusion pour l'équation de Klein-Gordon dans la métrique de Kerr*, Dissertationes Mathematicae **421** (2003).

[21] D. Häfner, J.-P. Nicolas, *Scattering of massless Dirac fields by a slow Kerr black hole*, Reviews in Mathematical Physics **16**(1) (2004), 29–123.

[22] D. Häfner, J.-P. Nicolas, *The characteristic Cauchy problem for Dirac Fields on curved backgrounds*, J. Hyperbolic Differ. Equ. **8** (2011), 3, 437–483.

[23] S.W. Hawking & G.F.R Ellis, *The large scale structure of space-time*, Cambridge monographs in mathematical physics, Cambridge University Press 1973.

[24] L. Hörmander, *The analysis of linear partial differential operators*, Vols. 1 to 4, Classics in mathematics, reprints, Springer, 2003-2009.

[25] L. Hörmander, *A remark on the characteristic Cauchy problem*, J. Funct. Ana. **93** (1990), 270–277.

[26] W. Inglese, F. Nicolò, *Asymptotic properties of the electromagnetic field in the external Schwarzschild spacetime*, Ann. Henri Poincaré **1** (2000), 5, 895–944.

[27] J. Joudioux, *Hörmander's method for the characteristic Cauchy problem and conformal scattering for a nonlinear wave equation*, Letters in Mathematical Physics **110** (2020), 1391–1423.

[28] S. Klainerman, *The null condition and global existence to nonlinear wave equations*, Nonlinear systems of partial differential equations in applied mathematics, Part 1 (Santa Fe, N.M., 1984). Lectures in Appl. Math. **23** (1986), 293–326.

[29] S. Klainerman & F. Nicolò, *Peeling properties of asymptotically flat solutions to the Einstein vacuum equations*, Class. Quantum Grav. **20** (2003), 3215–3257.

[30] M.D. Kruskal, *Maximal extension of Schwarzschild metric*, Physical Review **119** (1960), 5, 1743–1745.

[31] P.D. Lax, R.S. Phillips, *Scattering theory*, Academic Press 1967.

[32] G.E. Lemaître, *L'Univers en expansion*, Ann. Soc. Scient. Bruxelles Series A **53** (1933), 51–85.

[33] J. Leray, *Hyperbolic differential equations*, Princeton Institute for Advanced Studies 1952.

[34] J.-L. Lions, E. Magenes, *Non-homogeneous boundary value problems and applications*, vol. I. Trans- lated from the French by P. Kenneth, Die Grundlehren der mathematischen Wissenschaften, Band 181. Springer, New York, p. xvi+357 (1972)

[35] L. Mason, J.-P. Nicolas, *Conformal scattering and the Goursat problem*, J. Hyperbolic Differ. Equ. **1** (2004), 2, 197–233.

[36] E. Minguzzi and M. Sánchez, *The causal hierarchy of space-times.* In Recent developments in pseudo-Riemannian geometry, ESI Lect. Math. Phys., 299–358. Eur. Math. Soc., Zürich, 2008.

[37] C.S. Morawetz, *The decay of solutions of the exterior initial-boundary value problem for the wave equation*, Comm. Pure Appl. Math. **14** (1961), p. 561–568.

[38] C.S. Morawetz, *The limiting amplitude principle*, Comm. Pure Appl. Math. **15** (1962), p. 349–361.

[39] J.-P. Nicolas, *On Lars Hörmander's remark on the characteristic Cauchy problem*, Ann. Inst. Fourier **56** (2006), 3, 517–543.

[40] J.-P. Nicolas, *Conformal scattering on the Schwarzschild metric*, arXiv:1312.1386, 2013.

[41] B. O'Neill, The geometry of Kerr black holes, A.K. Peters, Wellesley, 1995.

[42] R. Penrose, *Asymptotic properties of fields and space-time*, Phys. Rev. Lett. **10** (1963), 66–68.

[43] R. Penrose, *Conformal approach to infinity*, in Relativity, groups and topology, Les Houches 1963, ed. B.S. De Witt and C.M. De Witt, Gordon and Breach, New-York, 1964.

[44] R. Penrose, *Zero rest-mass fields including gravitation : asymptotic behaviour*, Proc. Roy. Soc. London **A284** (1965), 159–203.

[45] R. Penrose, W. Rindler, *Spinors and space-time*, Vol. I & II, Cambridge monographs on mathematical physics, Cambridge University Press 1984 & 1986.

[46] R. Racke, *Nonlinear evolution equations*, Vieweg, 1992.

[47] M. Reed, B. Simon, *Methods of modern mathematical physics*, vol. 1, Functional analysis, Academic press, 1972.

[48] M. Reed, B. Simon, *Methods of modern mathematical physics*, vol. 2, Fourier analysis and self-afjointness, Academic Press, 1975.

[49] M. Reed and B. Simon, *Methods of modern mathematical physics*, vol. 3, Scattering theory, Academic Press, 1979.

[50] SageMath: a free open-source mathematics software system, https://www.sagemath.org/

[51] K. Schwarzschild, *Über das Gravitationsfeld einer Kugel aus inkompressibler Flüssigkeit nach der Einsteinschen Theorie*, Berl. Ber. 1916, 424-434, 1916.

[52] K. Schwarzschild, *On the gravitational field of a mass point according to Einstein's theory*, English translation of the original article by S. Antoci and A. Loinger, General Relativity and Gravitation, **35** (2003), 5., 951–959.

[53] H. Soga, *Singularities of the scattering kernel for convex obstacles*, J. Math. Kyoto Univ. **22**-4 (1983), 729–765.

[54] J.L. Synge, *The gravitational field of a particle*, Proc. Roy. Irish Acad. A **53** (1950), 83–114.

[55] G. Szekeres, *Singularities of a Riemannian manifold*, Publicationes Mathematicae Debrecen **7** (1960), 285–301, reprinted as golden oldie, General Relativity and Gravitation **34** (2002), 2001–2016.

[56] J. Smulevici, *Lectures on Lorentzian geometry and hyperbolic PDEs*, https://www.ljll.math.upmc.fr/smulevici/lgpdes.pdf.

[57] M. E. Taylor, *Partial differential equations I. Basic theory*, Appl. Math. Sci., 115, Springer, 2023.

[58] G. Teschl, *Ordinary Differential Equations and Dynamical Systems*, Graduate Studies in Mathematics **40**, AMS, 2012.

[59] R. M. Wald, General relativity, University of Chicago Press, 1984.

[60] H. Whitney, *Differentiable functions defined in arbitrary subsets of Euclidean space*, Trans. Amer. Math. Soc. **40** (1936), no. 2, p. 309–317.

[61] E.T. Whittaker, *On the partial differential equations of mathematical physics*, Mathematische Annalen **57** (1903), p. 333-355.

[62] Queffélec, C. Zuily, *Analyse pour l'agrégation*, 5^{th} edition, Dunod, 2020.

Index