Aspects of (quantum) field theory on curved spacetimes, particularly in the presence of boundaries

Umberto Lupo

Doctor of Philosophy

University of York
Mathematics

September 2015

Abstract

This thesis has two main themes: on the one hand, in Chapters 3 and 5 we study some effects of the presence of *timelike* boundaries on linear classical and quantum field theories; the second theme is the analysis of technical issues with the 1991 paper [KW91], which is carried out in parts of Chapter 2 and in Chapter 4. Chapter 2 contains a novel result on the characteristic initial value problem on globally hyperbolic spacetimes. In Chapter 3, we conjecture that (when the notion of a *Hadamard* state is suitably adapted to spacetimes with timelike boundaries) there is no isometry-invariant Hadamard state for the Klein–Gordon equation defined on the region of the Kruskal spacetime 'to the left of' a surface of constant Schwarzschild radius in the right Schwarzschild wedge, if Dirichlet boundary conditions are imposed there. We also prove that, under a suitable notion for 'boost-invariant Hadamard state' which also takes into account the special infra-red pathology of massless fields in 1+1 dimensions, there is no such state for the massless 1+1 wave equation on the region of Minkowski space to the left of an eternally uniformly accelerating mirror – with Dirichlet boundary conditions at the mirror. Chapter 5 collects and extends results of Solis [Sol06] about the causal structure of spacetimes with timelike boundaries, and deals with algebraic aspects of the interplay between *Green hyperbolicity* and boundary conditions in classical field theory. It also outlines a plan for generalizing the established work on wave-like equations from globally hyperbolic spacetimes to 'globally hyperbolic spacetimes-with-timelike-boundaries'. Appendix B contains a non-existence result for boost-invariant Hadamard states of a massless scalar field in (1+1)-dimensional Minkowski spacetime.

Contents

List of Figures

Acknowledgements

I would like to thank the University of York for providing me with the opportunity and means to carry out the research on which this thesis is based. In particular, my gratitude goes to Prof. Bernard S. Kay for supervising my PhD, providing useful ideas and suggesting various ways of improving my work along the way.

Further special thanks go to all other members of the Mathematical Physics research group in York, and in particular to Prof. Christopher J. Fewster for the many interesting discussions and for the technical help he provided on several occasions.

I would also like to thank the University of Göttingen for giving me the opportunity to present my research at the LQP meeting in Goslar in November 2014, the University of Trento for its kind hospitality during my visit there in January 2015, and the University of Birmingham for allowing me to give a short talk at the event "Britgrav 15" in April 2015.

I am especially grateful to Dr. Igor Khavkine for his enlightening explanations of a lot of beautiful mathematics and for his collaboration on interesting research, some of which is still in embryonic form but which I have nonetheless tried to include in this thesis as part of Chapter 5. I am also thankful to Prof. Valter Moretti and Dr. Claudio Dappiaggi for their support and encouragement at various stages.

Declaration of Authorship

I, Umberto Lupo, declare that the work for this thesis was done wholly during my candidature for a PhD at this University. Where I have consulted the published work of others, this is always clearly attributed.

Parts of Chapters 3 and 4 are taken from a paper [KL16] which was written jointly with my PhD supervisor Prof. Bernard S. Kay. The intuition, and basic structure of the argument supporting the claim that there should be no isometry-invariant Hadamard states either for a generic Klein–Gordon field on the portion of the Kruskal spacetime described in Chapter 3, or for a massless scalar field on the portion of (1+1)-dimensional Minkowski spacetime also described there, is due to Prof. Kay. However, the general philosophy adopted towards the Hadamard notion in the presence of timelike boundaries, and towards infrared issues in 1+1 dimensions, together with the precise statement and proof of the no-go result for massless fields in the (1+1)-dimensional case, are my own. The discovery of the gap in the arguments of the paper [KW91], analysed and partially filled in Chapter 4, occurred during a conversation with Prof. Kay. Finally, the idea that analytic elliptic regularity could be used to fill this gap in some cases is due to Robert M. Wald (private communication). However, the precise technical setup of Chapter 4 [in particular, the definition and relevance of the spaces $(\hat{S}^k, \hat{\sigma})$ defined there], all proofs in Section 4.2, the idea that recent decay results for the wave equation in a variety of spacetimes of interest could provide an alternative strategy for filling the gap, and all the version of the argument given for the case of Kerr in Section 4.4, are my own original contribution. Section 5.1 mainly collects results already proved in [Sol06], but some additional novel results are my own.

All original results in Chapter 2, and their proofs, are entirely my own.

The exact sequence in Section 5.2.1 is a generalisation of the well-known exact sequence pertaining Green hyperbolic operators on globally hyperbolic spacetimes, and its precise form is my own conception. However, the importance of obtaining an exact sequence in this context was stressed to me by Dr. Igor Khavkine. Finally, the content of Appendix B is entirely my own.

For my parents, who have always given everything.
For my sister, to whom I wish all the best.

Chapter 1

Introduction and Motivation

Modern fundamental physics can be confidently said to rest on the two building blocks of Einstein's general relativity (GR) and quantum (field) theory [Q(F)T], in the sense that both theories have by now been extensively tested, and vindicated, on their respective domains of applicability.

To wit, GR makes correct (so far) predictions about the *large scale* structure of the universe, and in particular on the medium-to-long-range interactions of its constituents due to gravity. When complemented with suitable assumptions,[1] it also provides a satisfactory understanding of the universe's past state of affairs, provided at least that one does not attempt to reconstruct the latter's very early history. GR interprets gravity geometrically by positing that gravitational interactions are compound effects of the local *curvature* of spacetime (i.e., of the universe including its past and future history). By means of Einstein's field equations, it describes and constrains where, when and how geometry is curved in relation to where, when and how matter is localised. Thus, geometry in GR is a dynamical entity which cannot be prescribed *a priori*: in the words of J. A. Wheeler, "Spacetime tells matter how to move; matter tells spacetime how to curve".

It is perhaps curious that a theory with such spectacular predictions of large-scale phenomena should be mathematically formulated in terms of the infinitesimal (differential) properties of a continuum. In analogy with the example of fluid mechanics, it is tempting to conclude that the existence of a successful continuum description is just a 'lucky' coincidence, and that one should perhaps be willing to accept that a more complete theory, in which GR would be embedded, may well describe the universe in different terms. One does not, in fact, need to leave the realm of GR to obtain indications that this expectation is well-founded: rigorous *singularity theorems* in cosmology prove that GR leads to the occurrence of spacetime singularities – i.e. of 'places' at which relevant physical quantities blow-up preventing further predictability – in a generic way [HP70].

[1] For instance, with initial conditions as the ones used in the Lambda-CDM model [EMM12, Pla16].

Ordinary Q(F)T, on the other hand, describes matter and its non-gravitational interactions in regimes in which gravity can be safely ignored and in terms of the behaviour of its fundamental, *microscopic*, constituents. The strong and weak nuclear forces, as well as the electromagnetic force, can all be described[2] in terms of appropriate quantum (field) theories. Furthermore, QFT fully and unproblematically embraces the principles of Einstein's *Special* Relativity, so that processes involving relative speeds close to the speed of light also fit in the theoretical framework. No experimental disagreements with the predictions of the *Standard Model* of particle physics, which is based on QFT, have so far been identified. Now, any quantum-mechanical theory radically departs from the *classical* point of view on physics according to which natural phenomena should admit a deterministic description – that is, one in which the only limitations to our ability to predict the future, given the present, have their origin in the *practical* impossibility to attain complete knowledge of the present itself. In quantum mechanics, the evolution of physical entities is governed by probabilistic laws whose status is that of fundamental axioms and not simply of 'effective' descriptions. By contrast, the mathematical formalism of GR is that of a classical, deterministic theory.

Now suppose that an ultimate theory of physics can in principle be found which is valid in all possible regimes and is able to reproduce the already established predictions of GR and QFT. It is hardly believable that, in such a theory, gravitational phenomena and the geometry of space-time would still be described by deterministic laws while non-gravitational ones would remain fundamentally probabilistic, *if the evolution of spacetime (geometry) is to be inextricably linked to the evolution of matter*. Since, of the two building blocks we described, only Q(F)T makes any claim about the character of microscopic phenomena, the most conservative approach towards obtaining the ultimate theory seems to be to reformulate GR in a quantum mechanical language – that is, to obtain a *quantum theory of gravity* (and presumably, if one of the fundamental tenets of GR is to be kept, of geometry). Aside from technical difficulties which arise in naïvely trying to do so [GS86], there is then another issue: Although QFT makes probabilistic statements about nature's constituents, its probabilistic laws are formulated in terms of a known spacetime 'stage' which, traditionally, is modelled by a continuum. So if the geometric spacetime stage is to be replaced, in a theory of quantum gravity, by something entirely different, then the rules governing the behaviour of the matter 'actors' will require substantial modification as a result. Currently, experiments provide no obvious hints as to how to proceed.

The quest for a full theory of quantum gravity thus remains open. But it is possible and fruitful to seek a mathematical formulation of, and to then extract physical predictions from, an intermediate model of the interaction between gravitation/geometry

[2]Sometimes, e.g. in the case of the *electroweak* model, in a unified manner.

and the quantum mechanical matter fields. The starting point is the observation that the geometric setup of Special Relativity (the *Minkowski spacetime*), in which ordinary QFT is formulated, is just a particular instance of the class of geometries which can be described by GR. In this intermediate model, commonly referred to as *QFT on curved spacetime*, one views spacetime as a fixed background, classically described in the language of GR, and formulates a suitable generalisation of the laws of propagation of quantum matter fields from Minkowski spacetime to more general, possibly curved, geometries. No attempt is made to quantize gravity/geometry itself or even to account for the back-reaction of quantum matter fields on the geometric fabric of spacetime, i.e. quantum fields are treated as 'test' objects and only the first half of Wheeler's slogan is implemented, replacing 'matter' by 'quantum matter'.

In the absence of a unified quantum treatment of gravity together with matter, one of course ought to ask the question of what the *expected* range of applicability of this hybrid approximate theory might be. We certainly should not expect QFT on curved spacetimes to produce accurate predictions when the spacetime curvature approaches Planck scales, i.e. when it is of the order of $\ell_{\mathrm{P}}^{-2} \sim 10^{69}\,\mathrm{m}^{-2}$. This presumed range of validity is, however, sufficiently large not to preclude meaningful applications of the theory to many relevant cosmological and astrophysical phenomena. In this context, a basic physical prediction of QFT on curved spacetimes is that strong gravitational fields (i.e., in GR, the *metric* of the spacetime under consideration) can 'polarize' the vacuum of a quantum matter field in a way analogous to the way in which the vacuum of the electron-positron field is polarised under an external time-dependent electromagnetic field in conventional QFT. A related prediction is that, when the gravitational field is time-dependent in a precise sense, the theory predicts pair creation of particles associated to the quantum field. As a result, QFT on popular cosmological models such as the Friedman–Lemaître–Robertson–Walker (FLRW) spacetime was quickly recognised [TU57, Par68] to predict particle creation as a result of the expansion of the universe.

The most surprising prediction of QFT on curved spacetimes to this date, however, is undoubtedly the *Hawking effect* [Haw75], according to which a classical black hole of mass M arising from gravitational collapse will emit thermal radiation at the *Hawking temperature* given (in units in which G, c, $\hbar$ and Boltzmann's constant k_{B} are all taken to be 1), for a spherically symmetric (Schwarzschild) black hole, by

$$T_{\mathrm{H}} = \frac{1}{8\pi M} = \frac{1}{4\pi R} = \frac{1}{\sqrt{4\pi A}}$$

where $R = 2M$ is the *Schwarzschild radius* of the black hole, and $A = 4\pi R^2$ is the surface area of the black hole's *event horizon*. The existence of such a counterintuitive prediction already at this level of approximation thus provides a strong reason for studying related effects in QFT on curved spacetimes in idealised situations. Surprises arise as a result of doing so, such as the famous discovery by Unruh (the *Unruh effect* [Unr76], but see also the previous related work by Fulling and Davies [Ful73, Dav75]) that a uniformly accelerated detector in Minkowski space will detect a thermal spectrum of particles when the global state of the quantum field is the Minkowski vacuum.[3]

The discovery of the Hawking and Unruh effects catalysed research in the field, leading to a vast number of research programs. The work in this thesis is related to, and sheds some light on, at least three such programs, which we now recall.

Black hole thermodynamics. Hawking's discovery that black holes behave exactly like thermodynamic objects possessing a temperature related in a simple way to their mass, and therefore also to their geometry, was of course tantalising. Prior to Hawking's derivation, Bekenstein [Bek73] had already argued, on general grounds not specifically tied to QFT on curved spacetimes, that black holes should possess an entropy and that this entropy ought in fact to be proportional to the area of the event horizon. Hawking's result shed light on the origin of this entropy in the context of simple quantum-field-theoretic models, and revealed the exact value of the constant of proportionality. The *Bekenstein–Hawking entropy* of a Schwarzschild black hole (in units as above) is

$$S_{\mathrm{BH}} = \frac{1}{4}A = 4\pi M^2 = \pi R^2.$$

Further clarifying the origin of this entropy naturally led to investigating idealised situations in which Schwarzschild black holes are imagined to be placed in a spherical container – a 'box' – situated beyond the event horizon [Haw76]. In these models, black holes are surrounded by black-body radiation – a so-called 'thermal atmosphere' – at the Hawking temperature, with which the black hole (viewed as a gravitational object) is in thermal equilibrium [GP76, Haw76, GH93]. A natural question is whether a black hole in equilibrium in such a box[4] has a semi-classical description in terms of a fixed classical spacetime together with a suitable state of a quantum field defined on it – with equilibrium between the two systems being modelled by both the classical spacetime metric and the quantum state being *invariant* under the appropriate groups of transformations describing time-invariance.

Curved spacetime generalisations of the Unruh effect; QFT on spacetimes with bifurcate Killing horizons. Analogs of the Unruh effect were soon recognised to occur on more

[3]The Unruh effect is actually a result in ordinary Minkowski spacetime QFT. That it was discovered as a result of an effort to better understand the Hawking effect for black holes is, we believe, an indication that the challenge to generalise Minkowski spacetime QFT to the curved spacetime scenario can actually help to clarify which physical notions and mathematical structures are really important in QFT and which are merely convenient assumptions.

[4]Here we leave aside the issue that a Schwarzschild black hole in equilibrium in a box is believed to be thermodynamically unstable [Haw76].

complicated geometries than the Minkowski spacetime. Notable examples were: (i) the maximal extension of the Schwarzschild spacetime (i.e. the *Kruskal* spacetime), with the rôle of the Minkowski vacuum played by the *Hartle–Hawking–Israel* state [HH76, Isr76]; (ii) the *de Sitter* spacetime modelling an isotropic, homogeneous, closed and expanding universe, with the rôle of the Minkowski vacuum played by a quantum state first constructed by Chernikov and Tagirov in [CT68] (but nowadays known as the *Bunch–Davies vacuum* after [BD78]), as discussed e.g. in [Isr76, GH77]. It was soon understood, thanks both to the aforementioned explicit investigations and to general abstract reasoning involving conventional (axiomatic) QFT [Sew82], that the core of the argument leading to the Unruh effect both on Minkowski spacetime and on such other geometries lay in the fact that in all cases there exists a one-parameter group of metric-preserving transformations (i.e. of *isometries*), acting as 'time' translations[5] on a *double-wedge* portion of the spacetime [Kay85]. The culmination of this program, in the case of linear spin-0 quantum fields, was a seminal 1991 paper by Kay and Wald [KW91] – but see also [Kay93] for some technical refinements – in which, amongst other major achievements:

- It was shown that a double-wedge structure naturally arises on a large class of spacetimes comprising many of the spacetimes of interest in quantum field theory on curved spacetimes, including in particular the Minkowski, Kruskal and de Sitter spacetimes on which the Unruh effect was already known to occur. Spacetimes in this class are required to enjoy 'good' causal properties (specifically *global hyperbolicity*), and to admit a one-parameter family of isometries yielding a *bifurcate Killing horizon* structure (most of Chapter 2 in this work will be concerned with reviewing this geometric background).

⋆ Under some technical caveats, it was 'proved' that quantized (real, spin-0) linear fields on such spacetimes admit a *unique* state which is both invariant under the spacetime isometries and exhibits unpathological short-distance behaviour (respectively, the state is *isometry-invariant* and *Hadamard*).

- It was shown that, when restricted to either one of the two 'wedge regions', this unique state is necessarily a thermal state (technically, a *KMS* state) with respect to the isometries restricted to that wedge, and at a temperature which, in the case of the Kruskal spacetime and when the isometries are rescaled 'at infinity' in a natural way, coincides with the Hawking temperature T_{H}. In the case of more general spacetimes with a bifurcate Killing horizon structure, this temperature is obtained by replacing the area A of the event horizon in Hawking's formula by

[5]'Time' here is meant in a sense which generalises the usual global time translation of Minkowski spacetime. In more mathematical language, the requirement is that the flow resulting from the isometries occurs in a *timelike* direction.

the quantity π/κ^2, where κ is a constant of geometric origin, determined solely by the bifurcate Killing horizon structure and called the *surface gravity.* That is, this general temperature is given in terms of κ by the simple formula

$$T = \frac{\kappa}{2\pi}.$$

⋆ For some notable cases, such as appropriate portions of the maximal extensions of the *Kerr* spacetime (which models an isolated, electrically neutral, rotating black hole in an open universe) and of the *Schwarzschild–de Sitter* spacetime (which models a nonrotating, electrically neutral black hole in an otherwise de Sitter universe), the authors claimed to provide rigorous proofs that there is *no* such state. As a result, in particular, analogs of the Unruh effect for linear, real, spin-0 quantum fields cannot exist on such spacetimes. In the case of the Kerr spacetime, the argument for non-existence was based on the fact that some of the rotational kinetic energy of a rotating black hole can be extracted by scattering into it waves of sufficiently low frequency and high angular momentum [Zel72] – a phenomenon called *superradiance.* For Schwarzschild–de Sitter, one argument for the no-go result was based on the fact that, should such a state exist, it would simultaneously have two different Hawking temperatures, one associated with the black hole horizon and the other with the de Sitter cosmological horizon. Another argument relied on what, in quantum information theory, is now known as 'monogamy' of entanglement (although this notion had not yet been coined at the time).

Quantization in the presence of moving mirrors. As a result of yet another struggle to capture the essence of Hawking's original argument for particle creation by black holes [Haw75], by separating it from the technical clutter associated with curvature and four-dimensionality, and by reducing the prominence of global considerations, some authors (see [DF77] and references therein) studied radiation processes for quantum fields in simplified models which, they argued, simulate most of the important features of spherically symmetric gravitational collapse models as the ones considered by Hawking. Specifically, in these models the quantum fields were linear, massless and of spin 0, and the kind of backwards-in-time wave scattering which, in the case of four-dimensional collapse models, was at the heart of Hawking's derivation via a somewhat mysterious application of the *geometric optics approximation*, was achieved in the context of a <u>flat</u> and <u>two-dimensional</u> spacetime by introducing a spatial boundary and reflecting conditions for the field at that boundary – that is, by introducing a mirror. What was an overall effect of non-trivial 'bulk' geometry in Hawking's derivation was now the result of a disturbance concentrated along the trajectory of the mirror.

Aside from the motivation coming from the Hawking effect in the manner just recalled, these low-dimensional mirror scenarios were extensively studied as worthwhile models in their own right. For instance, the discovery of the *Casimir effect* [Cas48] had already shown that the presence of reflecting mirrors can produce non-zero and physically significant energy density and pressure associated with the vacuum state. In [FD76, DF77] (see also [BD84, Sec. 4.4 and Sec. 7.1] for an overview), it was pointed out that a mirror which starts out inertial – with the state of the field the initial vacuum state – and later undergoes uniform acceleration does *not* radiate during the period of uniform acceleration.

1.1 Relation with previous work and main results

The work in this thesis was initially motivated by the first of the three research programs described above. Specifically, we sought an answer to the already mentioned natural question of whether or not regular quantum field theory on curved spacetimes might reasonably 'simulate' black holes in boxes when these are supposed to be in thermal equilibrium with a thermal atmosphere described by ordinary quantum-field–theoretic matter. Insofar as such systems *can* ever have a semi-classical description in which the geometry is described by an ordinary general-relativistic spacetime with a boundary describing the 'box', what might this spacetime look like?

As far as the treatment of the matter quantum fields is concerned, it seemed natural to begin our investigation with quantized, linear, spin-0 fields (i.e., Klein–Gordon fields) and to model their being 'confined' to a box by prescribing boundary conditions at the classical level – i.e., before quantization. The choice of boundary conditions must be, ultimately, physically motivated. And there must be an understanding that 'sharp' boundary conditions are perhaps never too physical – for instance, while the effect of some real physical devices (e.g., mirrors) on fields may be described very well by a 'sharp' boundary condition if the incident fields carry low energy, this description typically fails to be an accurate one at very high energies (e.g., in the case of mirrors, absorption inevitably begins to occur then). We will not touch on this inherent limiting aspect of our model any further in this thesis. We opted for Dirichlet boundary conditions as they are elementary to treat, but with a view towards allowing more general boundary conditions (and, concurrently, more general fields) in future research.

In this context, the work of B. S. Kay in [Kay15] suggested that, in the case of a Schwarzschild black hole, a description of such thermal equilibrium in which the geometry is modelled by the Kruskal spacetime with portions of the two exterior regions (symmetrically) removed, as in the Penrose diagram of Figure 1.1, runs into serious difficulties.

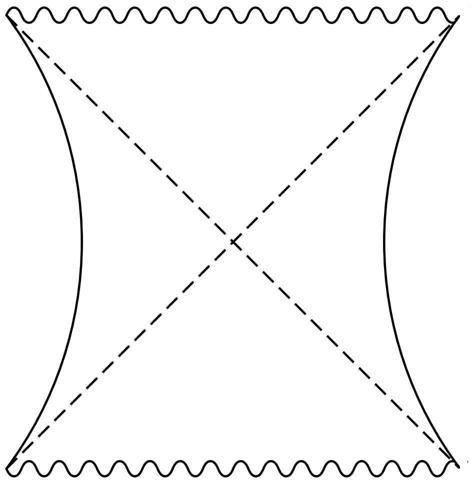

FIGURE 1.1: Penrose diagram for the region of the Kruskal spacetime considered in [Kay15]. See also Figure 3.3.

We therefore decided to examine the viability of another possible geometry, obtained by 'cutting off' the maximally extended Kruskal spacetime only in the right wedge. More precisely, as in [Kay15] we remove a portion of the spacetime in which the Schwarzschild radial coordinate exceeds some fiducial value larger than the Schwarzschild radius. But this time we only do so on the *right* exterior wedge, leaving the left one intact. The result is a spacetime which possesses a timelike boundary defining the surface of the imagined 'confining box' – see Section 3.1.

Unlike in the case of globally hyperbolic spacetimes without boundaries, a general, mathematically rigorous, theory of classical and quantum fields in spacetimes with timelike boundaries, and in the presence of boundary conditions, is still missing. In order to be able to produce physical arguments despite the lack of such a theory, we initially proceeded under the assumption that a certain minimal set of *desiderata* for classical fields still hold true in the context of certain spacetimes with timelike boundaries, and in particular in the context of our portion of the Kruskal spacetime. These requirements are minimal in the sense that they are analogous versions of properties which are known to hold for physically relevant fields on globally hyperbolic spacetimes and which are essential ingredients for carrying out rigorous quantization procedures there.

Making these assumptions allowed us to argue for and thus conjecture a non-existence result for regular (i.e. of 'Hadamard' type in a sense which we will clarify) quantum states which are additionally invariant under the isometries of our spacetime. Supporting evidence for the validity of our conjecture is provided by means of a no-go *theorem*, Theorem 3.4.7, which we are able to rigorously establish, for an analog toy model in 1+1 dimensions. This latter theorem is interesting in its own right as it sheds some further light on the analysis of quantization in the presence of moving mirrors, see our brief review on p. 18.

The content of the above paragraph, and a discussion of the physical significance of the conjectures and results there, is the topic of Chapter 3. Both our no-go conjecture and our no-go theorem may be regarded as non-existence results in the style of those proved by Kay and Wald in [KW91] – see the last item in the list beginning on p. 17. Yet another non-existence result in that style – for the wave equation on (1+1)-dimensional Minkowski spacetime – is proved in Appendix B.

In view of the close connection of our work with the setup and techniques in [KW91], we decided to clarify some technical aspects of the analysis in [KW91]. Chapter 2, as well as introducing necessary mathematical background, provides some such clarifications. Specifically: in Section 2.2 we deal with purely Lorentzian-geometric aspects, while in Section 2.5 we rigorously prove some significant results in hyperbolic PDE theory which were assumed true, without proof, in [KW91].

Another, more severe, issue with the analysis in [KW91] was also discovered in the process of tackling the problems in Chapter 3. Namely, an important gap was discovered which affected in particular the arguments for the items labelled '$\star$' in the list beginning on p. 17. We attempted, achieving partial success, to fill this gap and thus repair the affected arguments in [KW91]. Chapter 4 presents the results of our investigation and highlights the situations in which we were *not* able to fill the gap.

Finally, in Chapter 5 we initiate the treatment in earnest of the classical theory of fields and boundary-value problems on spacetimes possessing timelike boundaries, with a view to eventually establishing, as rigorous results in PDE theory, the *desiderata* mentioned five paragraphs above the current one. Some preliminary results are obtained, and a suggested road map for a complete resolution is provided in Section 5.4.

Chapter 2

Mathematical Background and Preliminaries

2.1 Review of relevant Lorentzian geometry

2.1.1 Differential topology generalities

Our main reference for definitions and results in differential topology is [Lee13]. In particular, the underlying topological space of a smooth or topological manifold will always be required to be Hausdorff and second-countable. However, unlike that reference, we will use the notation $T_p\phi : T_pM \to T_pN$ for the differential (or 'tangent map') of a map $\phi : M \to N$ at a point $p \in M$; the bundle map resulting from the collection $\{T_p\phi\}_{p\in M}$ will be denoted by $T\phi : TM \to TN$. For brevity, from now on we will often simply write 'manifold' to mean 'smooth manifold'. In general, we adopt the convention that terms such as 'map (of manifolds)', 'chart', 'immersion', 'submersion', 'embedding', 'submanifold', 'metric', 'vector/covector/tensor field', '(vector) (sub)bundle (map/morphism)', 'local/global trivialisation', 'section', 'local/global basis' etc., if used without further qualifiers, will refer to their smooth versions. One notable exception to this practice will be the fact that the word 'curve', if unqualified, will simply refer to what is sometimes called a 'continuous parametrised curve' or a 'path', i.e. a continuous function $\gamma : J \to M$ where $J \subseteq \mathbb{R}$ is an interval with non-empty interior.

We will write $A = B \uplus C$ to mean that $A = B \cup C$ and $B \cap C = \emptyset$. $\mathbb{N}$ will denote the non-zero natural numbers, while $\mathbb{N}_0 \coloneqq \mathbb{N} \cup \{0\}$. The topological closure, interior and boundary of a subset A of a topological space X will be denoted by $\overline{A}$, $\mathring{A}$ and $\dot{A}$ respectively, while for a smooth manifold with boundary M we reserve the notations ∂M and $\operatorname{Int} M$ to the boundary and interior of M in the sense of manifolds with boundary.

Throughout, if M is a manifold then $C^k(M)$ for $k \in \mathbb{N}_0 \cup \{\infty\}$ will denote the space of k times continuously differentiable *real-valued* functions on M – with '0 times continuously differentiable' meaning continuous. If N is another manifold, $C^k(M;N)$ will denote the space of k times continuously differentiable functions from M to N. If $E \to M$ is a smooth fiber bundle, then $\Gamma^k(E)$ for $k \in \mathbb{N}_0 \cup \{\infty\}$ will denote the space of k times continuously differentiable sections of $E \to M$, so that in particular $C^k(M) = \Gamma^k(M \times \mathbb{R})$ where $M \times \mathbb{R} \to M$ is the trivial vector bundle. We may occasionally use the notation $\mathfrak{X}(M)$ for the set $\Gamma^\infty(TM)$ of smooth vector fields on M. If M and N are real analytic manifolds and $E \to M$ is a real analytic fiber bundle, then we can define $C^\omega(M)$, $C^\omega(M;N)$ and $\Gamma^\omega(E)$ as above, with 'ω times continuously differentiable' meaning real analytic. Finally, if $E \to M$ is a vector bundle then $\Gamma^k_0(E)$ denotes the subspace of $\Gamma^k(E)$ consisting of sections $\phi : M \to E$ whose *support*, defined by

$$\operatorname{supp} \phi \coloneqq \overline{\{x \in M \mid \phi(x) \neq 0\}} \subseteq M,$$

is compact. Thus, in particular, $C^k_0(M)$ is the space of k times differentiable real-valued functions on M with compact support. Appendix C reviews some other notable function spaces which will be of use to us.

Since the technical distinction between *immersed* and *embedded* submanifolds will become important later, we recall these definitions below.

Definition 2.1.1 ([Topological] immersions and embeddings)**.** Let P and M be manifolds, and $i : P \to M$ be a map. Then i is an *immersion* if all its differentials are injective linear maps; it is an *embedding* if it is an immersion and also a topological embedding, i.e. a homeomorphism onto its image $i(P) \subseteq M$ in the subspace topology. It P and M are just topological spaces, then $i : P \to M$ is a *topological embedding* if it is a homeomorphism onto its image equipped with the subspace topology, and a *topological immersion* if every $x \in P$ has a neighbourhood U such that $i\restriction_U$ is a topological embedding.

Definition 2.1.2 (Immersed and embedded [topological] submanifolds)**. (i)** We say that a subset $S \subseteq M$ can be given the structure of an *immersed submanifold* of M if there exists an injective immersion $i : P \to M$ whose image is S. There exists then a unique choice of topology and smooth structure on S, depending on i, turning S into a manifold and $i : P \to S$ into a diffeomorphism. However, if an i exists such that $i : P \to S$ is also a homeomorphism when S is equipped with the subspace topology, then all choices of injective immersion yield the same topology (i.e. the subspace topology) and smooth structure, relative to which the inclusion map of S into M is an embedding [Lee13, Thm. 5.31]. We then say that S is an *embedded submanifold* of M. **(ii)** Stripping M of its smooth structure and viewing M as a topological manifold, we say that S can be given the structure of an *immersed topological submanifold* if there is a topology on S (not necessarily the subspace topology) which makes it into a topological manifold such that the inclusion is a topological immersion. If the inclusion map is a topological embedding then S is an *embedded topological submanifold* of M.

Definition 2.1.3. Throughout this thesis, by a *hypersurface* in a manifold M we will mean a codimension-1 embedded submanifold of M according to Definition 2.1.2. A codimension-1 embedded topological submanifold of M will be called a *topological hypersurface.*

2.1.2 Lorentzian manifolds and elements of causal theory

We set the scene for semi-Riemannian and, in particular, Lorentzian geometry. Our principal reference is [O'N83]. Doing this will allow us to give the general definition of a spacetime (without boundary) which will be adhered to throughout this thesis.

Definition 2.1.4. Let V be a finite-dimensional real vector space, and $b : V \times V \to \mathbb{R}$ be a symmetric bilinear form. We call the *index* of b the largest integer which is the dimension of a subspace $W \subseteq V$ on which $b\restriction_W$ is positive definite.

Definition 2.1.5 (Semi-Riemannian and Lorentzian manifolds)**.** Given a manifold M, a *metric tensor* – or simply a 'metric' – is a smooth tensor field of type $(0, 2)$ – that is, an element of $\Gamma^\infty(T^*M \otimes T^*M)$ – which is everywhere symmetric and non-degenerate and has constant index throughout M. By a *semi-Riemannian manifold* we mean a pair (M, g) where M is a manifold and g is a metric tensor. If, in addition, $\dim M \geq 2$ and g has index equal to 1 then (M, g) is a *Lorentzian manifold.*

Whenever convenient, we will adopt the standard summation convention and abstract index notation for tensor indices. As is customary, we will denote the covariant derivative arising from g by ∇ and, for any smooth function f, the vector field $\nabla^a f$ by $\operatorname{grad} f$. Our convention for the Riemann curvature tensor is

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)v^c = R_{abd}{}^c v^d,$$

the Ricci tensor is $R_{ab} := R^c{}_{acb}$, and the Ricci scalar curvature is $R := R^a{}_a$. For us, a *geodesic* is always what some authors refer to as an 'affinely parametrised geodesic'; i.e., a (necessarily smooth) curve $\gamma : I \to M$ satisfying the geodesic equation $\nabla_{\dot\gamma}\dot\gamma = 0$. We will instead refer to curves $\gamma : I \to M$ parametrised so that $\nabla_{\dot\gamma}\dot\gamma \propto \dot\gamma$ as *pre*geodesics.

Definition 2.1.6 (Types of submanifolds of Lorentzian manifolds)**.** If (M, g) is a Lorentzian manifold, P is a manifold, and $i : P \to M$ is an injective immersion, the resulting immersed submanifold $S = i(P)$ is called *spacelike* (resp. *timelike*) if the pullback i^*g

defines a Riemannian (resp. Lorentzian) metric on P. If i^*g is everywhere degenerate then S is said to be *null*.

Definition 2.1.7 (Normal bundle of a submanifold)**.** Given an arbitrary embedded submanifold $S \overset{i}{\hookrightarrow} M$ of a Lorentzian manifold (M, g), the *normal bundle* of S, denoted by NS, is the vector subbundle of the pullback bundle $TM|_S = i^*TM$ consisting of vectors normal to S using the metric g. I.e., as a set and viewing TS as a subset of TM,

$$NS = \{n \in TM|_S \mid g(n, X) = 0 \; \forall \; X \in TS\}.$$

The standard Riemannian exponential map will be used in the sequel. For a proof of its smoothness on the maximal domain $\mathcal{D}$ defined below, we refer to [Lee97, Prop. 5.7]. Here and throughout this thesis, given $V \in TM$ we will denote by γ_V the (unique) *maximally extended* geodesic with initial tangent vector V – whose domain of definition is then necessarily an open interval containing $0 \in \mathbb{R}$ and depending on V.

Proposition 2.1.8. *Let (M, g) be a semi-Riemannian manifold, and let*

$\mathcal{D} := \{V \in TM :$ *the maximal geodesic γ_V is defined on an interval containing* $[0, 1]\}$.

Then $\mathcal{D}$ is an open subset of TM and the exponential map *of (M, g),* $\exp : \mathcal{D} \to M$ *defined by* $\exp(V) = \gamma_V(1)$*, is smooth.* □

Note that, if $f : (M_1, g_1) \to (M_2, g_2)$ is a local isometry, then $Tf(\mathcal{D}_1) \subseteq \mathcal{D}_2$ where $\mathcal{D}_1$ (resp. $\mathcal{D}_2$) denotes the domain of the exponential map on (M_1, g_1) [resp. (M_2, g_2)]. This will be used in the proof of Theorem 2.2.6. If S is an embedded submanifold then the *normal exponential map* $\exp_S^\perp$ is defined as the restriction of exp to the normal bundle NS; it immediately follows that $\exp_S^\perp$ is defined and smooth on an open submanifold $\mathcal{D}_S^\perp$ of NS. The following stronger result is Lemma 7.25 in [O'N83].

Proposition 2.1.9. *If (M, g) is a semi-Riemannian manifold and S is an embedded submanifold such that the pullback of g to S is everywhere non-degenerate, then S has a* normal neighbourhood $\mathcal{U}$ *in M. This means that there exists an open neighbourhood $\mathcal{W}$ of the set of zero vectors in NS, with $\mathcal{W} \subseteq \mathcal{D}_S^\perp$, such that $\exp_S^\perp$ carries $\mathcal{W}$ diffeomorphically onto $\mathcal{U}$.* □

If $S = \{p\}$ is a one-point subset, then NS is just T_pM and Proposition 2.1.9 adapts in the following sense. Letting $\exp_p := \exp \restriction_{T_pM \cap \mathcal{D}}$, a *normal neighbourhood* of p is, by definition, an open neighbourhood $\mathcal{U} \ni p$ such that there exists an open neighbourhood $\mathcal{W} \subseteq T_pM$, starshaped about the zero vector in T_pM, such that $\exp_p : \mathcal{W} \to \mathcal{U}$ is a diffeomorphism. With these definitions, any point in a semi-Riemannian manifold has a normal neighbourhood. Indeed, more is true. We call a neighbourhood of p *convex* or *geodesically convex* if it is a normal neighbourhood of each of its points.

Lemma 2.1.10 (Existence of convex neighbourhoods)**.** *A semi-Riemannian manifold has arbitrarily small convex neighbourhoods of any of its points.* □

We now review some aspects of the *causal theory* of Lorentzian manifolds. That is, we concern ourselves with relations between points in Lorentzian manifolds which arise from the *causal character* of vectors and curves there.

Definition 2.1.11 (Causal character of vectors)**.** The *causal character* of a vector $X \in T_pM$ in a Lorentzian manifold (M, g) is defined as follows: X is called *timelike* if $g_p(X, X) > 0$, *spacelike* if $X = 0$ or $g_p(X, X) < 0$, *null* or *lightlike* if $g_p(X, X) = 0$ but $X \neq 0$, and *causal* (or *nonspacelike*) if either timelike or null.

If (M, g) is a Lorentzian manifold and $p \in M$, the subset of T_pM consisting of timelike vectors forms an open cone in T_pM [1] with two connected components. The same is true – minus the openness property – of the subset consisting of null vectors, and of the one consisting of causal vectors. In all these cases, a timelike vector X at p selects one of the two connected components by intersecting with the set $\{Y \in T_pM \mid g_p(X, Y) > 0\}$ – in particular, in the timelike or causal case, X is contained in this connected component – and any other timelike vector in the same connected component would select the same connected component. Therefore, any (everywhere) timelike vector *field* on M, even a discontinuous one, induces a global assignment, to each point in M, of one of the two connected components of the set of timelike/null/causal vectors at that point.

Definition 2.1.12 (Time-orientability)**.** A Lorentzian manifold (M, g) for which there exists a global continuous (but see the remark below) timelike vector field is called *time-orientable*[2], and the resulting global choice of connected components of the sets of causal vectors is called a *time orientation.* If a choice of time orientation is made, a causal vector at $p \in M$ belonging to the connected component selected at p is called *future-directed*, while one belonging to the remaining connected component is called *past-directed*.

Remark. Some authors would impose, in Definition 2.1.12, the seemingly stricter requirement that there exist a global *smooth* timelike vector field. It is perhaps rarely explicitly pointed out that this is in fact not a restriction at all in the following sense: If a continuous timelike vector field T_0 on M exists then there exists a smooth timelike vector field T which is everywhere future-directed according to T_0. The proof is simple using a fact which can be easily extracted from much more general results given in e.g. [Hir76] [see in particular Exercise 3(b) on p. 56 there]: If M, N are smooth manifolds, $p : N \to M$ is a submersion, and there exists a continuous $f_0 : M \to N$ such that $p \circ f_0 = \mathrm{id}_M$, then there exists a smooth $f : M \to N$ with $p \circ f = \mathrm{id}_M$. Our claim then follows by applying this fact to our Lorentzian manifold M, defining

$$N := \{X \in TM \mid g(X, X) > 0\} \cap \{Y \in TM \mid g(T_0, Y) > 0\},$$

[1] A *cone* in a vector space is a set invariant under multiplication by positive real numbers.
[2] We also say that g is a time-orientable metric for M.

which is an open subset of TM, and letting $p : N \to M$ be the restriction to N of the standard bundle projection $TM \to M$. Incidentally, these considerations also apply if g is merely a continuous metric.

Definition 2.1.13 (Spacetime)**.** By a *spacetime* we mean a triple $\mathscr{M} = (M, g, \mathfrak{t})$, in which (M, g) is a connected[3] and time-orientable Lorentzian manifold, and $\mathfrak{t}$ is a choice of time orientation.

Definition 2.1.14 (Causal character of curves)**.** A C^1 curve $\gamma : I \to M$ in a Lorentzian manifold is called *timelike* if all its tangent vectors are timelike, *null* if all its tangent vectors are null, and *spacelike* if they are all spacelike.[4] Finally, it is called *causal* or sometimes *nonspacelike*[5] if all its tangent vectors are causal. If the Lorentzian manifold is given a time orientation, then γ is said to be future/past-directed if all its tangent vectors are future/past-directed in the chosen time orientation. A timelike/null/causal curve $\gamma : [a, b] \to M$ with $-\infty < a < b < +\infty$ such that $\gamma(a) = p$ and $\gamma(b) = q$ is called a (timelike/null/causal curve) curve *from p to q*.

We can easily extend Definition 2.1.14 to curves which are merely *piecewise-C^1* provided that the following holds: at any break point of the curve, the left-sided and right-sided tangents to the curve at that point belong to the same connected component of the set of causal vectors. We can even extend it to continuous curves which are only differentiable on a dense subset of their interval of definition $I \subseteq \mathbb{R}$, provided that the Lorentzian manifold is time-orientable: in that case, the tangent vectors to the curve, whenever they exist, are required to be either always future-directed or always past-directed once a choice of time orientation is made. This will become important below.

Definition 2.1.15 (Causal relations between points in a spacetime)**.** Let $\mathscr{M} = (M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold, and $p, q \in M$. **(1)** We say that p *chronologically precedes* q in $\mathscr{M}$ [equivalently, that q chronologically *follows* p in $\mathscr{M}$], and write $p \ll_{\mathscr{M}} q$ [equivalently, $q \gg_{\mathscr{M}} p$], if there exists a future-directed and piecewise-smooth timelike curve from p to q [equivalently, a past-directed and piecewise-smooth timelike curve from q to p]. We then call

$$I^+_{\mathscr{M}}(p) := \{q \in M \mid p \ll_{\mathscr{M}} q\} \quad \text{and, for an arbitrary } A \subseteq M, \quad I^+_{\mathscr{M}}(A) := \bigcup_{p \in A} I^{\pm}_{\mathscr{M}}(p)$$

the *chronological future* of p and A, respectively. The *chronological past* of p or of A, denoted by $I^-_{\mathscr{M}}(p)$ and $I^-_{\mathscr{M}}(A)$ respectively, can be defined dually by using $\gg_{\mathscr{M}}$ instead of $\ll_{\mathscr{M}}$. **(2)** We say that p *causally* precedes q in $\mathscr{M}$ [equivalently, that q causally follows p in $\mathscr{M}$], and write $p \leq_{\mathscr{M}} q$ [equivalently, $q \geq_{\mathscr{M}} p$], if *either* $p = q$ *or* there

[3] It is well-known [Ger68, Mar72] that if a locally Euclidean, connected, Hausdorff topological space equipped with a smooth structure admits a (smooth) Lorentzian metric, then it is automatically paracompact. By connectedness, it is then also second-countable.

[4] In particular, under this convention the constant curves are always spacelike.

[5] But we will prefer the word ‘causal’ as we find it less ambiguous.

exists a future-directed and piecewise-smooth causal curve from p to q [equivalently, a past-directed and piecewise-smooth causal curve from q to p]. We then call

$$J^{+}_{\mathscr{M}}(p) \coloneqq \{q \in M \mid p \leq_{\mathscr{M}} q\} \quad \text{and, for an arbitrary } A \subseteq M, \quad J^{+}_{\mathscr{M}}(A) \coloneqq \bigcup_{p\in A} J^{\pm}_{\mathscr{M}}(p)$$

the *causal future* of p and A, respectively, and use $\geq_{\mathscr{M}}$ instead of $\leq_{\mathscr{M}}$ to define the *causal past* of p or of A, denoted by $J^{-}_{\mathscr{M}}(p)$ and $J^{-}_{\mathscr{M}}(A)$ respectively. Finally, for $A \subseteq M$, we write

$$J_{\mathscr{M}}(A) \coloneqq J^{+}_{\mathscr{M}}(A) \cup J^{-}_{\mathscr{M}}(A).$$

Remark. We will often drop references to $\mathscr{M}$ in e.g. the subscripts in Definition 2.1.15 whenever this is unlikely to cause confusion.

It is well-known that, for any $p \in M$, $I^{\pm}_{\mathscr{M}}(p)$ is an open set, and thus that so is $I^{\pm}_{\mathscr{M}}(A)$ for any $A \subseteq M$. In Chapter 5, we will explicitly prove a generalisation of this result in the case in which M is allowed to be a manifold with (smooth) boundary.

The reader will have noticed that we used curves which were piecewise-<u>smooth</u>, and not just (piecewise-)$\underline{C^1}$, in Definition 2.1.15. One naturally wonders if the resulting relations and sets would change upon using piecewise-C^1 curves instead. Fortunately, this is not the case and, actually, it is well-known that more is true, as we now explain. Recall that, if (X_1, d_1) and (X_2, d_2) are metric spaces, then a function $f : X_1 \to X_2$ is called *locally Lipschitz* if for every compact subset K of X_1 there exists a constant $C_K \geq 0$ such that

$$d_2(f(x), f(y)) \leq C_K d_1(x, y) \quad \forall\ x, y \in K.$$

If we are given two connected Riemannian manifolds (M_1, h_1) and (M_2, h_2) we can therefore speak of a locally Lipschitz map between them, meaning one which is locally Lipschitz with respect to the associated distance functions d_{h_1} and d_{h_2}. Then, we have the important *theorem of Rademacher*, which states that (amongst other things, see e.g. [Chr11, Thm. 2.3.2] and references therein) any such map has a classical derivative almost everywhere with respect to the Lebesgue measure in local coordinates on M_1 – in particular, it has a classical derivative on a dense subset of M_1.

Rademacher's theorem allows us to considerably extend the class of curves which might deserve to be called timelike/causal/null, whenever the underlying Lorentzian manifold is time-orientable. After a preliminary result, which is stated and proved as Proposition 2.3.1 in [Chr11], we will be able to provide this extended definition.

Proposition 2.1.16. *Let h_1 and h_2 be two complete Riemannian metrics on M. Then a curve $\gamma : I \to M$ – with $I \subseteq \mathbb{R}$ equipped with the standard Euclidean distance – is locally Lipschitz with respect to h_1 if and only if it is locally Lipschitz with respect to h_2.* $\square$

Definition 2.1.17 (Locally Lipschitz causal curves)**.** Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold. Then a curve $\gamma : I \to M$ which is locally Lipschitz with respect to one (and therefore any other) complete Riemannian metric is called *future-directed causal* [resp. *past-directed causal*] if, whenever it exists, $\dot{\gamma}$ is a causal and future-directed [resp. past-directed] vector. Replacing 'causal' with 'timelike' or 'null' yields the notion of future- or past-directed timelike or null locally Lipschitz curves.

Thus, at least in the case of time-oriented Lorentzian manifolds, a much wider class of causal/timelike/null curves is potentially available than that defined by requiring either piecewise-C^1 or piecewise-C^∞ regularity. The main reason for practitioners in Lorentzian geometry to require enlargements of the notion of causal and timelike curves, such as this one or the one reviewed below, is that piecewise-C^1 or piecewise-C^∞ curves do not form a class which is 'closed' under limiting procedures which often arise in applications, see Subsection 2.1.3. It is however useful to know that such enlargements do not also enlarge the causal relations and sets introduced in Definition 2.1.15. The following result is (itself a corollary of) Corollary 2.4.11 in [Chr11] – we refer also to [KSSV14, Cor. 3.10] for an even stronger statement.

Theorem 2.1.18. *Let $\mathscr{M} = (M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold, and suppose that the expression 'locally Lipschitz' were to be used everywhere in Definition 2.1.15 in place of 'piecewise-smooth'. Then the 'new' resulting causal relations $\ll_{\mathscr{M}}$ and $\leq_{\mathscr{M}}$ are the same as the original ones. Therefore, the sets $I^{\pm}_{\mathscr{M}}(A)$ and $J^{\pm}_{\mathscr{M}}(A)$ do not change. More concretely: if there exists a locally Lipschitz future-directed timelike [resp. causal] curve from p to q then there exists a piecewise-smooth future-directed timelike [resp. causal] curve from p to q.* $\square$

The rather weak notion of timelike and causal curve we just described is not universally used in the literature on Lorentzian geometry; an example of a remarkable application of this definition is the work in [FS12, Fat15]. An alternative way to extend the notion of causal curve dates back at least to the monograph [HE73, p. 184] and has been extensively used since. One obtains an enlargement of the class of piecewise-C^1 causal curves which, as we shall see, is strictly contained in the locally Lipschitz enlargement which we reviewed above.

Definition 2.1.19 (Continuous causal curves)**.** Let (M, g) be a Lorentzian manifold. A curve $\gamma : I \to M$ is a *continuous causal curve* if it is continuous and, for every convex set U and any two $t_1, t_2 \in I$ with $t_1 < t_2$ and $\gamma[t_1, t_2] \subset U$, it is the case that there is a piecewise-C^1 causal curve[6] from $\gamma(t_1)$ to $\gamma(t_2)$ which is entirely contained in U. If $(M, g, \mathfrak{t})$ is time-oriented, then we can define *future- or past-directed* continuous causal curves in the obvious way.

[6]But, of course, piecewise-smooth would do just as well!

As was already noted in [Pen72, Rk. 2.26] and [BEE96, Eq. (3.14)], the local chart expressions of continuous causal curves must satisfy local Lipschitz conditions. The topic of the exact degree of regularity of continuous causal curves according to the definition above is discussed in [CFS08, App. A, particularly Rk. A.4], and some of the results there are generalised in [Min15b, Thm. 7]. In particular, it holds that every future-directed [resp. past-directed] continuous causal curve $\gamma : I \to M$, when reparametrised with respect to the arc-length of an arbitrary Riemannian metric on M, becomes a future-directed locally Lipschitz causal curve. It follows at once from Theorem 2.1.18 that the causal precedence relation $\leq$ and notion of causal future/past of a set is not changed by using continuous causal curves instead of piecewise-smooth ones.

We shall therefore not have to worry, in what follows, about exactly what (extended) notion of causal or timelike curve is used. But it is healthy to keep in mind that most of the literature which we will cite adopts the notion in Definition 2.1.19 for causal curves. The following result, which is Corollary 14.1 in [O'N83] and which can be established rigorously using variational methods [O'N83, Prop. 10.46], is of the utmost importance.

Lemma 2.1.20 ('Push-up Lemma')**.** *If either $x \ll y \leq z$ or $x \leq y \ll z$, then $x \ll z$. As a result, $I^+(J^+(A)) = I^+(A)$ for any subset $A \subseteq M$.* □

An important consequence of this lemma will be used at various point in this thesis. It may also be proved by different methods, as done for instance in [BEE96, Cor. 4.14]. O'Neill's version of this result [O'N83, Cor. 14.5] also contains a statement about *conjugate points* which we will omit here.

Corollary 2.1.21. *Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold, and $A \subseteq M$ be an arbitrary subset. If γ is a future-pointing causal curve from A to a point $q \in J^+(A) \setminus I^+(A)$, then, up to reparametrisation, γ is a (smooth) null geodesic entirely contained in $J^+(A) \setminus I^+(A)$. A similar statement holds for $J^-(A)$ and $I^-(A)$.* □

With the setup of Corollary 2.1.21, since $\operatorname{im}\gamma \subseteq J^+(A)$ then $I^+(\operatorname{im}\gamma) \subseteq I^+(J^+(A)) = I^+(A)$. But then $\operatorname{im}\gamma \cap I^+(\operatorname{im}\gamma) \subseteq \operatorname{im}\gamma \cap I^+(A)$ which is empty. That is, $\operatorname{im}\gamma$ is an *achronal* (sub)set as in the following definition.

Definition 2.1.22 (Achronal sets)**.** A subset $A \subseteq M$ is *achronal* if, equivalently, $A \cap I^+(A) = \emptyset$ or $A \cap I^-(A) = \emptyset$. That is, there are no timelike curves from A to itself.

There is an obvious modification of this definition which uses $J^\pm$ instead of $I^\pm$. The result is the notion of an *acausal* (sub)set.

Definition 2.1.23 (Acausal sets)**.** A subset $A \subseteq M$ is *acausal* if, equivalently, $A \cap J^+(A) = \emptyset$ or $A \cap J^-(A) = \emptyset$. That is, there are no causal curves from A to itself.

We would now like to review some of the most notable conditions in the *causal hierarchy of spacetimes* [MS08], the strongest of which ('global hyperbolicity') will play a pivotal role throughout this thesis. Only one preliminary definition is still required.

Definition 2.1.24 (Causally convex subsets)**.** Let (M, g) be a Lorentzian manifold and U, V be open subsets of M with $U \subseteq V$. U is said to be *causally convex in* V if any causal curve between two points in U is entirely contained in V if and only if it is entirely contained in U. When $V = M$ we just say that U is causally convex.

The first is a mild condition which prevents 'back in time travel' via causal curves.

Definition 2.1.25 (Causality)**.** A spacetime is said to be *causal* if it admits no closed causal loops. Equivalently, if $J^+(p) \cap J^-(p) = \{p\}$ for any point p.

Further up on the 'causal ladder', a (strictly) stronger condition also rules out the existence of 'almost closed' causal loops.

Definition 2.1.26 (Strong causality)**.** A spacetime is said to be *strongly causal* if any of its points has an arbitrarily small causally convex neighbourhood.
Remark. Lemma 3.21 in [MS08] provides an equivalent definition of strong causality not explicitly phrased in terms of causally convex sets.

The definitions so far come with no guarantees as to whether they still hold if the spacetime metric is perturbed, even very slightly, in such a way as to 'widen' the cones of causal vectors in some places. The notion of 'stable causality' is (strictly) stronger than the previous two and enjoys this stability property. We let $\mathrm{Con}(M)$ be the set of conformal equivalence classes of Lorentzian metrics on M: two Lorentzian metrics g_1 and g_2 on M are *conformally equivalent* if there exists a smooth $\Omega : M \to (0, +\infty)$ such that $g_2 = \Omega g_1$. We denote the conformal equivalence class of g by $\boldsymbol{g}$. Then the partial ordering

$$g \prec g' \text{ if and only if all the causal vectors for } g \text{ are timelike for } g'$$

descends to a partial ordering on $\mathrm{Con}(M)$ which we still denote by $\prec$. We can then define stable causality in a number of equivalent ways – the equivalence between some of these definitions being a major recent achievement – we refer to [MS08] for further details. In fact, we will omit another definition given in terms of a topology on $\mathrm{Con}(M)$ (the *quotient* C^0 or *interval* topology) which can be characterised in a several natural ways [MS08, pp. 334–335].

Definition 2.1.27 (Stable causality)**.** A spacetime $(M, g, \mathfrak{t})$ is said to be *stably causal* if, equivalently:

(i) there exists a $\boldsymbol{g'} \in \mathrm{Con}(M)$ such that $\boldsymbol{g} \prec \boldsymbol{g'}$ and $\boldsymbol{g'}$ is causal;

(ii) it admits a continuous function $t : M \to \mathbb{R}$ which is strictly increasing along any future-directed (continuous, locally Lipschitz) causal curve;

(iii) it admits a smooth function $\tilde{t} : M \to \mathbb{R}$ whose gradient is future-directed and timelike.

Definition 2.1.28 (Time and temporal functions)**.** Given a spacetime, a *time function* is one with the properties described in item **(ii)** of Definition 2.1.27, while a *temporal function* is one with the properties described in item **(iii)** of Definition 2.1.27.

Remark. Note that a smooth time function need not be a temporal function, while a temporal function is necessarily a time function. Since a spacetime is connected, the image of a time function is an interval in $\mathbb{R}$, and since we can always travel slightly to the causal future or past of any point, this interval must be open in $\mathbb{R}$. By composing with an orientation-preserving diffeomorphism onto $\mathbb{R}$, given a time or temporal function we can obtain another one whose image is $\mathbb{R}$.

Definition 2.1.29 (Causal simplicity)**.** A spacetime $(M, g, \mathfrak{t})$ is said to be *causally simple* if it is causal and if any of the following equivalent properties [MS08, Lem. 3.67] holds:

(i) $J^{\pm}(p)$ is closed for every $p \in M$;

(ii) $J^{\pm}(K)$ is closed for every compact set K.

Causal simplicity is stronger than stable causality in the category of spacetimes – which have no manifold boundary – which we are going to deal with in this chapter. We finally give the strongest causal condition of all – which is also the most useful one for applications to quantum field theory on curved spacetimes (see Section 2.4).

Definition 2.1.30 (Global hyperbolicity)**.** A spacetime $(M, g, \mathfrak{t})$ is *globally hyperbolic* if it is causal and, for any two $p, q \in M$, the *causal diamond* $J^+(p) \cap J^-(q)$ is compact.

Remark. Until recently, strong causality was used in place of causality in this definition. It is a recent observation [BS07] that, since causality and compactness of the causal diamonds implies strong causality, one in fact obtains the same notion by using the definition given above. Clearly then, we could also have used <u>stable</u> (!) causality instead of just causality; while this would make no sense in the current context, it is often what needs to be done when attempting to generalise the notion of global hyperbolicity to make it an attribute of structures generalising the field of causal cones on a Lorentzian manifold, as was done in [FS12, Fat15].

One of the main reasons for the usefulness of global hyperbolicity is its equivalence with the existence of a *Cauchy surface*.

Definition 2.1.31 (Inextendible curves)**.** Let M be a topological space and $I \subseteq \mathbb{R}$ be an interval with extremes a and b, $-\infty \leq a < b \leq +\infty$. Let $\gamma : I \to M$ be a continuous curve. Then $p \in M$ is a *left endpoint* of γ if $\lim_{t \to a} \gamma(t) = p$, and a *right endpoint* of γ if

$\lim_{t\to b}\gamma(t) = p$. γ is *right-inextendible* if it has no right endpoint, and *left-inextendible* if it has no left endpoint. If $(M, g, \mathfrak{t})$ is a time-oriented Lorentzian manifold and $\gamma : I \to M$ is a future-directed causal curve, then a right (resp. left) endpoint of γ is called a *future (resp. past) endpoint* of γ, and γ is said to be *future (resp. past) inextendible* if it has no future future (resp. past) endpoint. A curve which is both future and past inextendible is simply called inextendible.[7]

Remark. Another notion of (in)extendibility is that of *geodesic* (in)extendibility. A geodesic $c : (a, b) \to M$ is right (resp. left) geodesically inextendible if it has no extension to a geodesic $\tilde{c} : (c, d) \to M$ with $(c, d) \supset (a, b)$ and $d > b$ (resp. $c < a$). One can prove [O'N83, Lem. 5.8] that c is right (resp. left) geodesically inextendible if and only if it is right (resp. left) inextendible as a continuous curve.

Definition 2.1.32 (Cauchy [hyper]surfaces)**.** If $(M, g, \mathfrak{t})$ is a spacetime then a subset $\mathscr{C} \subset M$ is a *Cauchy surface* if any inextendible (continuous, locally Lipschitz) timelike curve intersects $\mathscr{C}$ exactly once.

Proposition 2.1.33 (Some properties of Cauchy surfaces)**.** *The following hold for a Cauchy surfaces $\mathscr{C}$ of a spacetime $(M, g, \mathfrak{t})$, whenever it exists:*

- **(i)** *$\mathscr{C}$ is achronal, closed and an embedded topological hypersurface in M;*
- **(ii)** *$J^{\pm}(K) \cap \mathscr{C}$ is compact whenever $K \subseteq M$ is compact;*
- **(iii)** *$M = I^-(\mathscr{C}) \uplus \mathscr{C} \uplus I^+(\mathscr{C})$;*
- **(iv)** *any inextendible (continuous, locally Lipschitz) causal curve intersects $\mathscr{C}$;*
- **(v)** *if $\mathscr{C}$ is a C^1 spacelike submanifold, then it is acausal.* □

Definition 2.1.34 (Cauchy [time or temporal] functions)**.** A function $t : M \to \mathbb{R}$ is called a *Cauchy function* if all its level sets are Cauchy surfaces. A time (resp. temporal) function (see Definition 2.1.28) which is Cauchy is called a Cauchy time (resp. temporal) function.

Theorem 2.1.35 (Global hyperbolicity in terms of Cauchy surfaces and Cauchy functions)**.** *For a spacetime $\mathscr{M} = (M, g, \mathfrak{t})$, the following are equivalent:*

- **(a)** *$\mathscr{M}$ is globally hyperbolic according to Definition 2.1.30;*
- **(b)** *$\mathscr{M}$ admits a Cauchy surface $\mathscr{C}$;*
- **(c)** *$\mathscr{M}$ admits a Cauchy time function;*
- **(d)** *$\mathscr{M}$ admits a smooth and spacelike Cauchy surface $\widetilde{\mathscr{C}}$;*
- **(e)** *$\mathscr{M}$ admits a Cauchy temporal function τ onto $\mathbb{R}$.* □

In case **(b)***, it also follows that any other Cauchy surface is homeomorphic to $\mathscr{C}$ and that M is homeomorphic to $\mathbb{R} \times \mathscr{C}$. In case* **(d)***, it also follows that any other smooth Cauchy surface (spacelike or not) is diffeomorphic*[8] *to $\widetilde{\mathscr{C}}$, and that there exists a Cauchy temporal function such that $\widetilde{\mathscr{C}}$ is one of its level sets. In case* **(e)***, it also follows that,*

[7] The word 'endless' is sometimes used in the literature in place of 'inextendible'.
[8] If the spacetime is orientable, even oriented-diffeomorphic.

with the notation $\mathscr{C}_t := \tau^{-1}\{t\}\ \forall\ t \in \mathbb{R}$, *there exists an isometry* $\phi_0 : \mathbb{R} \times \mathscr{C}_0 \to M$ *such that* $\mathcal{T} := \tau \circ \phi_0 : \mathbb{R} \times \mathscr{C}_0 \to \mathbb{R}$ *is just the natural projection, and the pullback metric has the form*

$$(\phi_0)^* g = \mathrm{d}\mathcal{T}^2 - g_\tau$$

where $\beta : \mathbb{R} \times \mathscr{C}_0 \to \mathbb{R}$ *is a positive smooth function, and* g_τ *is a smooth symmetric* $(0,2)$*-tensor field on* $\mathbb{R} \times \mathscr{C}_0$ *whose radical at each point is precisely the span of* $\partial_\mathcal{T}$ *at that point, and whose pullback under each inclusion* $\mathcal{T}^{-1}\{t\} \hookrightarrow \mathbb{R} \times \mathscr{C}_0$ *is a Riemannian metric on* $\mathcal{T}^{-1}\{t\} = \phi^{-1}(\mathscr{C}_t)$. □

Causal completeness and past/future compactness

We shall make use of the following notion of *(future/past) causal completeness* of a subset, formally introduced in [Gal86].

Definition 2.1.36. A subset $A \subseteq M$ in a time-oriented Lorentzian manifold $(M, g, \mathfrak{t})$ is called future causally complete if, for each $q \in J^+(A)$, the closure of $J^-(q) \cap A$ <u>in A</u> is compact. We define past causal completeness by exchanging the roles of $+$ and $-$. A subset that is both future and past causally complete will simply be called *causally complete.*

Remark. The subtly different notion of *future/past compactness* of a subset A was first introduced by Leray in [Ler53], and used extensively in subsequent work, e.g. in [Fri75, BGP07]. Namely, $A \subseteq M$ is future [resp. past] compact if $J^+(p) \cap A$ [resp. $J^-(p) \cap A$] is compact for all $p \in M$. A is *temporally compact* if it is both future and past compact. In general, future [resp. past] causal completeness as per Definition 2.1.36 is a weaker condition than past [resp. future] compactness. However, if $(M, g, \mathfrak{t})$ is globally hyperbolic, a set is future [resp. past] causally complete if and only if it is past [resp. future] compact, and thus is causally complete if and only if it is temporally compact.

Clearly, the union and intersection of two future (resp. past) causally complete subsets is future (resp. past) causally complete. Compact sets are always causally complete. More interestingly, as already remarked in [Gal86], Cauchy surfaces (when they exist) are also causally complete [Tre11, p. 114], and thus so are their closed subsets. For future reference, we also record some useful results the (simple) proofs of which can be found in [GV14, Sec. 3.1] or in [Bär15, Sec. 1.2]. In all three statements, $(M, g, \mathfrak{t})$ is a time-oriented Lorentzian manifold.

Lemma 2.1.37. *If* $A \subseteq M$ *is either past or future causally complete, then it is closed.* □

Lemma 2.1.38. *Suppose that* $(M, g, \mathfrak{t})$ *is globally hyperbolic. If* $A \subseteq M$ *is future causally complete, then* $J^+(A)$ *is closed and thus* $J^+(A) = \overline{I^+}(A)$. *Furthermore,* $J^+(A)$ *too is* *future causally complete (and so is any of its closed subsets). Analogous statements hold for $J^-(A)$ if A is past causally complete.* $\square$

If $(M, g, \mathfrak{t})$ is globally hyperbolic, we can describe the future/past causal completeness of closed sets using Cauchy surfaces. A subset $A \subseteq M$ is said to be *future [resp. past] bounded* if there exists a Cauchy surface $\mathscr{C}^+$ [resp. $\mathscr{C}^-$] such that $A \subseteq J^-(\mathscr{C}^+)$ [resp. $A \subseteq J^+(\mathscr{C}^-)$].

Lemma 2.1.39. *Suppose that $(M, g, \mathfrak{t})$ is globally hyperbolic, and that $A \subseteq M$ is closed. Then A is future (resp. past) causally complete if and only if it is past (resp. future) bounded.* $\square$

Having introduced future and past causally complete subsets of a generic time-oriented Lorentzian manifold $(M, g, \mathfrak{t})$, given a vector bundle $F \to M$ we may define subspaces of $\Gamma^\infty(F)$, or even of the space $\Gamma^{-\infty}(F) := [\Gamma_0^\infty(F^*)]'$ of distributional sections of F (denoted $\mathscr{D}'(M, F)$ in [BGP07, Bär15]), with causally restricted supports.

Definition 2.1.40. If $(M, g, \mathfrak{t})$ is a time-oriented Lorentzian manifold and $F \to M$ is a vector bundle, we call

$$\Gamma^\infty_{\mathrm{ret}}(F) := \big\{u \in \Gamma^\infty(F) \,\big|\, \operatorname{supp} u \subseteq J^+(K) \text{ for a compact } K \subseteq M\big\}$$

the space of smooth sections with *retarded support*,

$$\Gamma^\infty_{\mathrm{adv}}(F) := \big\{u \in \Gamma^\infty(F) \,\big|\, \operatorname{supp} u \subseteq J^-(K) \text{ for a compact } K \subseteq M\big\}$$

the space of smooth sections with *advanced support*,

$$\Gamma^\infty_{\mathrm{sc}}(F) := \{u \in \Gamma^\infty(F) \mid \operatorname{supp} u \subseteq J(K) \text{ for a compact } K \subseteq M\} \supseteq \Gamma^\infty_{\mathrm{ret}}(F) \cup \Gamma^\infty_{\mathrm{adv}}(F)$$

the space of smooth sections with *spatially compact support*,

$$\Gamma^\infty_{\mathrm{fcc}}(F) := \{u \in \Gamma^\infty(F) \mid \operatorname{supp} u \text{ is future causally complete}\}$$

the space of smooth sections with future causally complete support,

$$\Gamma^\infty_{\mathrm{pcc}}(F) := \{u \in \Gamma^\infty(F) \mid \operatorname{supp} u \text{ is past causally complete}\}$$

the space of smooth sections with past causally complete support,

$$\Gamma^\infty_{\mathrm{cc}}(F) := \{u \in \Gamma^\infty(F) \mid \operatorname{supp} u \text{ is causally complete}\} = \Gamma^\infty_{\mathrm{fcc}}(F) \cap \Gamma^\infty_{\mathrm{pcc}}(F)$$

the space of smooth sections with causally complete support. We define and denote corresponding subspaces of distributional sections (cf. Appendix C) analogously. If $(M, g, \mathfrak{t})$ is globally hyperbolic we may say that sections in $\Gamma^\infty_{\mathrm{fcc}}(F)/\Gamma^\infty_{\mathrm{pcc}}(F)/\Gamma^\infty_{\mathrm{cc}}(F)$ have past/future/temporally compact support, and use the alternative notations $\Gamma^\infty_{\mathrm{pc}}(F)/\Gamma^\infty_{\mathrm{fc}}(F)/\Gamma^\infty_{\mathrm{tc}}(F)$.

Corollary 2.1.41. *The following hold in the presence of global hyperbolicity:*

(i) $\Gamma^\infty_{\mathrm{ret}}(F) \cap \Gamma^\infty_{\mathrm{adv}}(F) = \Gamma^\infty_0(F)$;

(ii) $\Gamma^\infty_{\mathrm{ret}}(F) \subseteq \Gamma^\infty_{\mathrm{fcc}}(F)$ *and* $\Gamma^\infty_{\mathrm{adv}}(F) \subseteq \Gamma^\infty_{\mathrm{pcc}}(F)$. $\square$

2.1.3 Limit curve theorems

We now wish to record a general result, commonly referred to as the *Limit Curve Lemma*, which is an essential analytical tool of Lorentzian geometry. A corollary thereof (Corollary 2.1.46), which holds in the presence of global hyperbolicity, will be used in the proof of Theorem 2.2.9. Although several different versions of the Limit Curve Lemma exist in the literature, we will recall the formulation given in [Gal86] (see also [BEE96, Lem. 14.2] and the recent review [Min08a]), which best suits our purposes. We first define the notion of *h-uniform convergence (on compact subsets)* for sequences of curves in a Riemannian manifold (M, h) [Min08a, Def. 2.1], as it plays a central role in the statement of our version of the Limit Curve Lemma. Then, we will recall a certain 'sequential closedness' property of the space of continuous causal curves under this notion of convergence [Min08a, Lem. 2.7]. This will be enough background to state our version of the Limit Curve Lemma, i.e. Theorem 2.1.44.

Definition 2.1.42. Let (M, h) be a Riemannian manifold and $d^h : M \times M \to \mathbb{R}$ be the associated Riemannian distance function. Let I_1 and I_2 be intervals in $\mathbb{R}$; given two curves $\gamma_1 : I_1 \to M$ and $\gamma_2 : I_2 \to M$, we write

$$d^h_\infty(\gamma_1, \gamma_2) \coloneqq \sup\{d^h(\gamma_1(t), \gamma_2(t)) \mid t \in I_1 \cap I_2\}.$$

Given a sequence of intervals $I_n \subseteq \mathbb{R}$, we write $I_n \to I$ to mean that the boundary points of the I_n converge to the boundary points of I.
Let $\gamma_n : I_n \to \mathbb{R}$ be a sequence of curves. We say that the sequence converges *h-uniformly* to the curve $\gamma : I \to \mathbb{R}$ if $I_n \to I$ and $\lim_{n\to\infty} d^h_\infty(\gamma_n, \gamma) = 0$. We say that the sequence converges *h-uniformly on compact subsets* to the curve $\gamma : I \to \mathbb{R}$ if for every compact interval $J \subseteq I$ there exist intervals $J_n \subseteq I_n$ such that the sequence given by the curves $\gamma_n\restriction_{J_n}$ converges h-uniformly to $\gamma\restriction_J$.

Since, when a time-oriented Lorentzian manifold $(M, g, \mathfrak{t})$ is equipped with an arbitrary Riemannian metric h, any continuous causal curve can be shown to satisfy a local Lipschitz condition with respect to the resulting distance function d^h [BEE96, Eq. (3.14)], it straightforwardly follows that any continuous causal curve admits an h–arc-length reparametrisation.

Lemma 2.1.43. *Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold and let h be a Riemannian metric on M. In the following, $I_n, I \subseteq \mathbb{R}$ are intervals with non-empty interior. If $\gamma_n : I_n \to M$ is a sequence of continuous causal curves parametrised with respect to h–arc-length, which converges h-uniformly on compact subsets to $\gamma : I \to M$, then γ is a continuous causal curve.* $\square$

A judicious application of the Arzelà-Ascoli Theorem, together with Lemma 2.1.43, lead to our version of the Limit Curve Lemma.

Theorem 2.1.44 (Limit Curve Lemma)**.** *Let* $(M, g, \mathfrak{t})$ *be a time-oriented Lorentzian manifold of arbitrary dimension, and* h *be a complete Riemannian metric on* M *with associated distance function* $d^h : M \times M \to \mathbb{R}$*. Further let* $\gamma_n : \mathbb{R} \to M$ *be a sequence of continuous causal curves, parametrised with respect to* h*–arc-length. If* $p \in M$ *is an accumulation point of the sequence* $\gamma_n(0)$*, then there is a continuous causal curve* $\gamma : \mathbb{R} \to M$ *(not necessarily parametrised with respect to* h*–arc-length) such that* $\gamma(0) = p$*, and a subsequence* γ_{n_k} *which converges* h*-uniformly on compact subsets to* γ*.* □

As shown in [Min08a, Thm. 3.1], although Theorem 2.1.44 is stated only for sequences parametrised (with respect to h–arc-length) over the entire real line,[9] it is in fact possible to generalise the Limit Curve Lemma to obtain analogous statements concerning sequences of causal curves parametrised (with respect to h–arc-length) over generic intervals, even allowing for up to countably many *limit* points.

In proving Theorem 2.2.9 in the next section, we will be faced with the following issue:

Issue. For a generic time-oriented Lorentzian manifold $(M, g, \mathfrak{t})$, a complete Riemannian metric h, and a sequence $\gamma_n : [0, b_n] \to \mathbb{R}$ of h–arc-length–parametrised, continuous, future-directed, causal curves with future and past endpoints, if $\gamma_n(0) \to p \in M$ and $\gamma_n(b_n) \to q \in M$ then, in the instance where the sequence does not collapse to the single point $p = q$,[10] there exist *two* continuous, future-directed, causal curves which (in a sense made precise in statement (2) of Theorem 3.1 in [Min08a]) deserve to be called 'limit curves' for our sequence. There are two possibilities: (a) both curves have future and past endpoints p and q respectively, in which case they are reparametrisations of each other; (b) one curve has past endpoint at p but no future endpoint, while the other has future endpoint at q but no past endpoint.

Only by providing additional assumptions can possibility (b) be ruled out. One such assumption can be given straightforwardly, and was indeed provided in [Min08a, Thm. 3.1 (2)]: Suppose that $(M, g, \mathfrak{t})$ is *nontotally (future and past) imprisoning*: I.e. that given any compact set $K \subseteq M$ and any continuous future-directed causal curve $\gamma : I \to \mathbb{R}$ with no future (resp. past) endpoint, if $\gamma(t) \in K$ for some $t \in I$ then there exist $t' \in I$ with $t' > t$ (resp. $t' < t$) such that $\gamma(t') \notin K$. Then, clearly, possibility (b) is ruled out if it is also known that the curves γ_n are contained in a common compact set. The following Lemma demonstrates a situation in which this is guaranteed to be the case.

[9] In particular, since h is assumed complete, these curves are necessarily past and future inextendible, i.e. they don't have future or past endpoints [Min08a, Lem. 2.6].

[10] More precisely, if there exists a subsequence of curves whose h–arc-lengths tend to a positive number.

Lemma 2.1.45. *Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold such that, for all $p, q \in M$, the causal diamonds $J^+(p) \cap J^-(q)$ are (empty or) compact. Further let there be sequences $p_n \to p$ and $q_n \to q$. Then $\bigcup_{n\in\mathbb{N}}[J^+(p_n) \cap J^-(q_n)]$ (is empty or) has compact closure.*

Proof. Pick $p' \in I^-(p)$ and $q' \in I^+(q)$. Then $I^+(p')$ is an open neighbourhood of p, so it contains all but a finite number of the p_n. Therefore, $J^+(p_n) \subseteq I^+(p')$ for all but finitely many p_n. Similarly, $I^-(q')$ is an open neighbourhood of q, and contains all but a finite number of the q_n, so that $J^-(q_n) \subseteq I^-(p')$ for all but finitely many q_n. Hence there exists $N \in \mathbb{N}$ such that

$$\begin{aligned}\bigcup_{n\in\mathbb{N}}[J^+(p_n) \cap J^-(q_n)] &\subseteq \bigcup_{n<N}[J^+(p_n) \cap J^-(q_n)] \cup [I^+(p') \cap I^-(q')] \\ &\subseteq \bigcup_{n<N}[J^+(p_n) \cap J^-(q_n)] \cup [J^+(p') \cap J^-(q')],\end{aligned}$$

and the result follows since the last line is the union of finitely many compact sets. □

Hence, if a time-orientable Lorentzian manifold is nontotally imprisoning and has compact causal diamonds, we can rule out possibility (b) in our Issue above. Now, the property of being nontotally imprisoning trivially implies causality since any closed causal path can be periodically extended, thus yielding an inextendible causal curve which lies entirely on a compact set (for further discussion, see [Min08b]). Since causality together with compactness of the causal diamonds implies global hyperbolicity [BS07], any Lorentzian manifold as described in the first sentence of this paragraph is therefore necessarily globally hyperbolic. Conversely, global hyperbolicity is a stronger condition than being nontotally imprisoning – and of course has compactness of the causal diamonds as part of its definition. Therefore, this resolution of our Issue is tantamount to requiring global hyperbolicity. We record this conclusion for future reference, and note that this result was previously stated and proved (in a more direct way without relying on the results in [Min08a]) as Theorem 3.2.11 in [Tre11].

Corollary 2.1.46. *Let $(M, g, \mathfrak{t})$ be a globally hyperbolic Lorentzian manifold and h a complete Riemannian metric on M. Let $\gamma_n : [0, b_n] \to M$ be a sequence of continuous, future-directed, causal curves parametrised with respect to h–arc-length. If $\gamma_n(0) \to p \in M$ and $\gamma_n(b_n) \to q \in M$ and the sequence of curves does not collapse to the point $p = q$, there exists $b \in (0, \infty)$ and a subsequence γ_{n_k} which converges h-uniformly to a future-directed causal curve $\gamma : [0, b] \to M$ from p to q. An analogous statement holds for past-directed causal curves.* □

2.2 Spacetimes with bifurcate Killing horizons

The geometry of embedded, two-dimensional, orientable and spacelike submanifolds of a given (orientable) spacetime will play a special role in this section. A starting point is provided by the following (well-known) fact.

Proposition 2.2.1. *Let* $\Sigma \overset{\iota}{\hookrightarrow} M$ *be an embedded, codimension-2, orientable, spacelike submanifold of an orientable spacetime* $\mathscr{M} = (M, g, \mathfrak{t}, \mathfrak{o})$*. Then the normal bundle* $N\Sigma$ *is (smoothly) trivial and admits a global basis* $\{l, n\}$ *of future-directed null vectors.*

Proof. Since Σ has codimension 2, $N\Sigma$ is a rank-2 vector bundle; since Σ is spacelike, the fiberwise restrictions of g to $N\Sigma$ define a (symmetric and non-degenerate) vector bundle metric on $N\Sigma$, which we will denote by $\hat{g}$. Since both M and Σ are orientable, $N\Sigma$ is orientable [O'N83, p. 214]. Therefore, there exists a nowhere vanishing section ω of the line bundle $\bigwedge^2 N\Sigma^* = \coprod_{x\in\Sigma} \bigwedge^2 N_x\Sigma^*$. On the other hand, time-orientability of M implies that there exists a global future directed and timelike vector field $\Theta \in \Gamma(TM)$. For all $x \in \Sigma$, the metric g allows us to project each vector $\Theta|_x$ to $N_x\Sigma$. Since Θ is never tangent to Σ, the result is a non-vanishing global section $\hat{\Theta}$ of $N\Sigma$ such that each $\hat{\Theta}|_x$ is future-directed and timelike. Using the musical isomorphism $\hat{\flat} : N\Sigma^* \to N\Sigma$ provided by $\hat{g}$, we can define another global nowhere vanishing section of $N\Sigma$ by

$$\Xi|_x := [\omega_x(\hat{\Theta}|_x, \cdot)]^{\hat{\flat}} \Longrightarrow \hat{g}_x(\Xi|_x, \hat{\Theta}|_x) = \omega_x(\hat{\Theta}|_x, \hat{\Theta}|_x) = 0.$$

It follows that $\{\hat{\Theta}, \Xi\}$ is a global basis of $N\Sigma$, and that each $\Xi|_x$ is a spacelike vector. Letting

$$l := \frac{\hat{\Theta}}{\hat{g}(\hat{\Theta}, \hat{\Theta})} - \frac{\Xi}{\hat{g}(\Xi, \Xi)} \quad \text{and} \quad n := \frac{\hat{\Theta}}{\hat{g}(\hat{\Theta}, \hat{\Theta})} + \frac{\Xi}{\hat{g}(\Xi, \Xi)}$$

yields the desired global null basis of $N\Sigma$. □

The null basis in Proposition 2.2.1 is not unique: it can be partially fixed by the requirement that $g(l, n) = f$ for a specified smooth $f : S \to (0, +\infty)$,[11] leaving a residual freedom consisting of the transformations $l \mapsto Fl$, $n \mapsto F^{-1}n$ where $F \in C^\infty(S, \mathbb{R}^+)$.

Specialising further, we now assume that **(A1)** $\mathscr{M}$ possesses a non-trivial Killing vector field ξ whose set of zeros contains a spacelike, two-dimensional embedded submanifold $\Sigma \subset M$. It follows that each point of Σ is a fixed point under the maximal flow generated by ξ. Below, we will discuss in depth how this setup, together with some additional mild assumptions, gives rise to a 'bifurcate Killing horizon' structure around the set Σ.

Remark. It is well-known that, if a four-dimensional spacetime possesses a non-trivial Killing vector field, then any connected component of the set of zeros of the field is either a singleton set or a closed, embedded, 2-codimensional and totally geodesic submanifold

[11] A typical choice, with our metric convention, would be $f \equiv 2$.

of M with normals everywhere of the same causal type. See e.g. [Kob58], [Boy69], [Hal90, Sec. III C] or [Hal04, Sec. 10.3]; see also [KN63, Ex. 8.1] for a result in a more general context.

Henceforth, we will also assume that **(A2)** the surface Σ is orientable, and that **(A3)** the Killing field ξ is complete, so that it generates a global one-parameter group of isometries which we denote by β_τ ($\tau \in \mathbb{R}$). Let l and n be null, future directed vector fields along Σ, and normal to Σ, constructed as explained above. For every $x \in \Sigma$, the isometries β_τ leave x fixed. Thus, the respective tangent maps $T_x\beta_\tau$ are Lorentz transformations of T_xM, and leave $N_x\Sigma \subset T_xM$ invariant. Since they are continuously connected to the identity via the group parameter τ, they are actually proper and orthochronous. It then immediately follows that $T_x\beta_\tau$ acts on the line generated by l_x as a scaling by $\mu(x,\tau)$, and on the line generated by n_x as a scaling by $\mu(x,\tau)^{-1}$, where $\mu(x,\tau)$ is positive and smoothly varying in x and τ. In fact, we can say more: by the group properties of $\{\beta_\tau\}_{\tau\in\mathbb{R}}$, $\mu(x,0) = 1$ and $\mu(x,\tau_1+\tau_2) = \mu(x,\tau_1)\mu(x,\tau_2)\ \forall\ \tau_1,\tau_2 \in \mathbb{R}$. I.e., for any $x \in \Sigma$ $\mu(x,\cdot)$ is a homomorphism from the additive group $\mathbb{R}$ to the multiplicative group $\mathbb{R}^+$. Since it is continuous, it must equal $e^{\kappa(x)\tau}$ for some $\kappa(x) \in \mathbb{R}$ smoothly varying in x; i.e.

$$(T_x\beta_\tau)l_x = e^{\kappa(x)\tau}l_x \quad \text{and} \quad (T_x\beta_\tau)n_x = e^{-\kappa(x)\tau}n_x. \tag{2.1}$$

As a result of this scaling property, the images of the maximally extended geodesics γ_{l_x} and γ_{n_x} with initial tangent vectors l_x and n_x respectively (see the paragraph above Proposition 2.1.8) must be invariant under each β_τ. This in turn implies that ξ is everywhere tangent to γ_{l_x} and γ_{n_x}. That is, denoting by I_{l_x} (resp. I_{n_x}) the interval of definition of γ_{l_x} (resp. γ_{n_x}), there exist functions $f_{l_x} \in C^\infty(I_{l_x})$ and $g_{n_x} \in C^\infty(I_{n_x})$ such that

$$\xi \circ \gamma_{l_x} = f_{l_x}\dot{\gamma}_{l_x} \quad \text{and} \quad \xi \circ \gamma_{n_x} = g_{n_x}\dot{\gamma}_{n_x}. \tag{2.2}$$

The dot denotes differentiation with respect to the affine parameter chosen so that $\gamma_{l_x}(0) = \gamma_{n_x}(0) = x$, and therefore also $\dot{\gamma}_{l_x}(0) = l_x$, $\dot{\gamma}_{n_x}(0) = n_x$, and $f_{l_x}(0) = g_{n_x}(0) = 0$. Our notation for the functions f_{l_x} and g_{n_x} may appear heavy-handed: we keep the subscripts to emphasise that different choices of the vectors l_x and n_x lead to different functions. The effect of making a new choice of vectors is, however, a simple one to describe: if $l'_x = cl_x$ for some $c > 0$, then $f_{l'_x}(U) = c^{-1}f_{l_x}(cU)$, whence $\dot{f_{l'_x}}(U) = \dot{f_{l_x}}(cU)$. The functions g_{n_x} behave analogously under scalings of n_x. In particular, the values of $\dot{f_{l_x}}(0)$ and $\dot{g_{n_x}}(0)$ are independent of the choices of l_x and n_x. This is also true of the value of $\kappa(x)$ in Equation (2.1), and the following Lemma illustrates that this is not a coincidence.

Lemma 2.2.2. *Assume that* **A1**, **A2** *and* **A3** *hold. Then the map* $\kappa : \Sigma \to \mathbb{R}$ *in (2.1) and the functions* f_{l_x} *and* g_{n_x} *in (2.2) are related by*

$$\kappa(x) = \dot{f}_{l_x}(0) = -\dot{g}_{n_x}(0).$$

Proof. Let $x \in \Sigma$. We pick a basis (e_μ) of T_xM such that $e_0 = l_x$, and use it to construct a geodesic normal coordinate chart $(\mathcal{U}, x^\mu)$ centred at x. In these coordinates, the Christoffel symbols of the metric g vanish at x, and the geodesic with the initial condition $v = a^\mu e_\mu$ is expressed by $x^\mu(t) = a^\mu t$ [KN63, Prop. 8.3]. Applied to $v = l_x$, and indicating γ_{l_x} simply by γ, this readily implies that

$$\dot{\gamma}(s) = \left.\frac{\partial}{\partial x^0}\right|_{\gamma(s)} \qquad \text{for small enough } s,$$

which in turn, by (2.2), also yields

$$\xi|_{\gamma(s)} = \xi^0(\gamma(s)) \left.\frac{\partial}{\partial x^0}\right|_{\gamma(s)} = f(s) \left.\frac{\partial}{\partial x^0}\right|_{\gamma(s)} \qquad \text{for small enough } s.$$

The ODE system describing the integral curve of ξ starting at an arbitrary point $\gamma(s) \in \mathcal{U}$ reduces in these coordinates to the single differential equation

$$\frac{\mathrm{d}\theta_{\gamma(s)}}{\mathrm{d}\tau}(\tau) = f(\theta_{\gamma(s)}(\tau)) \qquad \text{for small enough } s \text{ and } \tau,$$

together with the initial condition $\theta_{\gamma(s)}(0) = x^0(\gamma(s)) = s$. Using (2.1), and interchanging τ and s derivatives as is certainly allowed by the joint smoothness of the flow of ξ [Lee13, Ch. 9], we can now compute

$$\begin{aligned}
\kappa(x) l_x = \left.\frac{\mathrm{d}}{\mathrm{d}\tau}\right|_{\tau=0} [(T\beta_\tau) l_x] &= \left.\frac{\mathrm{d}}{\mathrm{d}\tau}\right|_{\tau=0} \left[\left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} (\beta_\tau \circ \gamma)(s) \right] \\
&= \left\{ \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} \left[\left.\frac{\mathrm{d}}{\mathrm{d}\tau}\right|_{\tau=0} \theta_{\gamma(s)}(\tau) \right] \right\} l_x \\
&= \left\{ \left.\frac{\mathrm{d}}{\mathrm{d}s}\right|_{s=0} f(\theta_{\gamma(s)}(0)) \right\} l_x \\
&= \dot{f}(0) l_x,
\end{aligned}$$

and the result follows. The second equality in the statement of the Lemma follows by exchanging the roles of the vectors l_x and n_x, and using Equation (2.1). □

The following is a simple but important corollary of the equations in (2.2).

Lemma 2.2.3. *Denoting by $U \in I_{l_x}$ the affine parameter of γ_{l_x}, and by D_U the covariant derivative along γ_{l_x} defined to act on vector fields along γ_{l_x},*

$$D_U(\xi \circ \gamma_{l_x}) = \nabla_{\dot{\gamma}_{l_x}} \xi = \dot{f}_{l_x} \dot{\gamma}_{l_x}. \tag{2.3}$$

In particular, $\nabla_{l_x} \xi = \dot{f}_{l_x}(0) l_x = \kappa(x) l_x$. Furthermore,

$$-\frac{1}{2} \operatorname{grad} g(\xi, \xi) \,|_{\gamma_{l_x}(U)} = \nabla_\xi \xi \,|_{\gamma_{l_x}(U)} = \dot{f}_{l_x}(U)\, \xi|_{\gamma_{l_x}(U)}. \tag{2.4}$$

Analogous results hold in the case of γ_{n_x}.

Proof. We give the proof in the case of γ_{l_x}. The geodesic equation is $D_U \dot{\gamma}_{l_x} = 0$. By (2.2) and the Leibniz rule then

$$D_U(\xi \circ \gamma_{l_x}) = D_U(f_{l_x} \dot{\gamma}_{l_x}) = \dot{f}_{l_x} \dot{\gamma}_{l_x} + f_{l_x} D_U \dot{\gamma}_{l_x} = \dot{f}_{l_x} \dot{\gamma}_{l_x},$$

giving (2.3). The second equality in (2.4) then follows by multiplying the second equality in (2.3) by f_{l_x} and making use of (2.2) again, while the first equality there follows by a well-known and straightforward property of Killing fields: namely,

$$\begin{aligned} \nabla_a(\xi^b \xi_b) &= \xi^b \nabla_a \xi_b + \xi_b \nabla_a \xi^b \\ &= 2\xi^b \nabla_a \xi_b \\ &= -2\xi^b \nabla_b \xi_a, \end{aligned}$$

and the result follows by index raising. □

Killing's equation $\nabla_{(a}\xi_{b)} = 0$ implies that at each $p \in M$ the endomorphism $(\nabla^a \xi_b)_p$ of T_pM is represented, in an orthonormal basis of T_pM, by an antisymmetric matrix. Therefore, its kernel has even codimension. Since ξ vanishes identically on Σ which has codimension 2, and since a non-trivial Killing vector field in a (connected) spacetime may never vanish together with its first (covariant) derivative, it follows that, at any $x \in \Sigma$, $\ker(\nabla^a \xi_b)_x = T_x\Sigma$. As l and n are linearly independent and transverse to Σ, $0 \neq (\nabla^a \xi_b) l^b = -l^b \nabla_b \xi^a$ and $0 \neq (\nabla^a \xi_b) n^b = -n^b \nabla_b \xi^a$ everywhere on Σ. By Lemma 2.2.3 and Lemma 2.2.2, it follows that

$$0 \neq \dot{f}_{l_x}(0) = \kappa(x) = -\dot{g}_{n_x}(0). \tag{2.5}$$

Thus, at least locally around any $x \in \Sigma$, ξ cannot vanish along either γ_{l_x} or γ_{n_x} except at x, and the functions f_x and g_x in Equation (2.2) change sign at 0. A more careful analysis of the eigenvector-eigenvalue problem for the tensor $\nabla^a \xi_b$ actually shows that under our assumptions – and in particular since Σ is spacelike – if f_x is positive immediately after 0 (i.e., if ξ is future-directed along γ_{l_x} immediately to the future of x) then g_x is negative there, and vice-versa. Details can be found in the references [Boy69, Hal90, Hal04] already given in the remark above. In fact, the conclusion that ξ is non-zero locally along γ_{l_x} and γ_{n_x} is also an immediate corollary of the results mentioned in that Remark. This is not a real surprise, since we have in effect reproduced some of the arguments entering the proofs of those results.

We now turn to global considerations. Specifically, we will show that the geodesics γ_{l_x} and γ_{n_x} (for any $x \in \Sigma$) are future/past geodesically complete (Proposition 2.2.5). We will also establish that the sets

$$\mathcal{H}_A \coloneqq \bigcup_{x\in\Sigma} \operatorname{im}\gamma_{l_x} \quad \text{and} \quad \mathcal{H}_B \coloneqq \bigcup_{x\in\Sigma} \operatorname{im}\gamma_{n_x}, \tag{2.6}$$

are always codimension-1 null immersed submanifolds of M under the minimal requirements **A1**, **A2** and **A3** (Theorem 2.2.6). Later, we will provide a reasonable set of additional topological and causal requirements on M and Σ under which we are able to rigorously prove that these sets are actually embedded submanifolds of M, i.e. null hypersurfaces in M (Theorem 2.2.9).[12] Under these more restrictive hypotheses, it will follow that $\mathcal{H}_A$ and $\mathcal{H}_B$ are even *achronal*; this observation will be of crucial importance in view of our applications to the characteristic initial value problem for linear hyperbolic PDEs in Section 2.5.

We begin by returning to (2.4). All the following arguments apply with obvious small modifications to the case of the geodesics γ_{n_x}. To stress this, and to simplify notation, we drop all subscript l_x from the calculations below. We may act on (2.4) with the derivative (bundle) maps $T\beta_\tau : TM \to TM$ (for each $\tau \in \mathbb{R}$). Since β_τ is an isometry, and using the fact that any vector field is invariant under its own flow, $\beta_{\tau*}\nabla_\xi\xi = \nabla_{\beta_{\tau*}\xi}(\beta_{\tau*}\xi) = \nabla_\xi\xi$ where, for any diffeomorphism F, F_* denotes the pushforward operator on vector *fields*, related to TF by $F_*X = TF \circ X \circ F^{-1}\ \forall\ X \in \Gamma(TM)$. Recalling further that $\operatorname{im}\gamma$ is invariant under β_τ, we conclude that for any $U \in I$ and any $\tilde{U} \in I$ such that $\beta_\tau(\gamma(U)) = \gamma(\tilde{U})$,

$$T\beta_\tau \left(\nabla_\xi\xi\,|_{\gamma(U)}\right) = \nabla_\xi\xi\,|_{\beta_\tau(\gamma(U))} = \nabla_\xi\xi\,|_{\gamma(\tilde{U})} = \dot{f}(\tilde{U})\,\xi|_{\gamma(\tilde{U})},$$

where (2.4) was used in the last step. On the other hand,

$$T\beta_\tau \left(\dot{f}(U)\,\xi|_{\gamma(U)}\right) = \dot{f}(U)\xi|_{\gamma(\tilde{U})}$$

[12] Note that, while all of the hypotheses in Theorem 2.2.9 make an appearance in [KW91], some of our global arguments were left implicit there, or claimed without proof. We leave open the question of whether the hypotheses in our Theorem 2.2.9 can in fact be weakened.

by linearity. Combining this with the previous line, and observing that β_τ may not map a point at which $\xi \neq 0$ to one at which $\xi = 0$, it follows that $\dot{f}(U) = \dot{f}(\tilde{U})$ whenever $U \in J \setminus \{0\}$ with J a suitable interval – which always exists as shown above – such that $\xi \circ \gamma\restriction_{J\setminus\{0\}}$ never vanishes. Clearly then, $\dot{f}$ is locally constant on $J \setminus \{0\}$ and, by smoothness, it is actually constant on J. We record the (obvious) global version of this result, and its immediate consequences, below.

Lemma 2.2.4. *Assume that* **A1**, **A2** *and* **A3** *hold. Let* $x \in \Sigma$ *and define*

$$\begin{aligned} U_+ &:= \sup\{U > 0 \mid U \in I_{l_x},\ \xi(\gamma_{l_x}(U)) \neq 0\ \forall\ U \in (0, U_+)\} \\ \text{and}\quad U_- &:= \inf\{U < 0 \mid U \in I_{l_x},\ \xi(\gamma_{l_x}(U)) \neq 0\ \forall\ U \in (U_-, 0)\}. \end{aligned}$$

Then $\dot{f}_{l_x}$ *is constant and equal to* $\kappa(x)$ *on* (U_-, U_+), *so that* $f_{l_x}(U) = \kappa(x)U\ \forall\ U \in (U_-, U_+)$. *Furthermore,*

$$D_U(\xi \circ \gamma_{l_x}) = \nabla_{\dot{\gamma}_{l_x}}\xi = \kappa(x)\dot{\gamma}_{l_x} \tag{2.7}$$

$$-\frac{1}{2}\,\mathrm{grad}\, g(\xi, \xi)\,|_{\gamma_{l_x}(U)} = \nabla_\xi \xi\,|_{\gamma_{l_x}(U)} = \kappa(x)\,\xi|_{\gamma_{l_x}(U)}. \tag{2.8}$$

Analogous statements hold for the geodesics γ_{n_x}.

Proof. That the constant value of $\dot{f}_{l_x}$ must equal $\kappa(x)$ follows from Lemma 2.2.2. (2.7) and (2.8) then follow immediately from (2.3) and (2.4). □

Now suppose that γ – which is future and past geodesically inextendible by assumption – is future or past geodesically *incomplete*. Then γ is not periodic, and by Equation (2.2) it may only intersect itself at points where the Killing field ξ vanishes. It follows that $\gamma\restriction_{(U_-, U_+)}$ is a regular injective curve, with either U_+ or U_- finite. Without loss of generality, assume $U_+ < \infty$. We may pull back ξ along $\gamma\restriction_{(0, U_+)}$ to define a vector field $\hat{\xi}$ on $(0, U_+)$, with

$$\hat{\xi}(U) = f(U)\frac{\partial}{\partial U} = \kappa(x)U\frac{\partial}{\partial U}$$

by Lemma 2.2.4. The completeness assumption **A3** on ξ implies that $\hat{\xi}$ too is complete. Thus, picking a $U_0 \in (0, U_+)$, the unique integral curve θ_0 of $\hat{\xi}$ through U_0 is defined on all $\mathbb{R}$:

$$\begin{aligned} \theta_0 : \mathbb{R} \to (0, U_+) \quad \text{solving} \quad & \frac{\mathrm{d}\theta_0}{\mathrm{d}\tau}(\tau) = \hat{\xi}^U(\theta_0(\tau)) = \kappa(x)\theta_0(\tau) \\ \text{with} \quad & \theta_0(0) = U_0. \end{aligned}$$

The solution is

$$\theta_0(\tau) = U_0 e^{\kappa(x)\tau}\ \forall\ \tau \in \mathbb{R},$$

whose image [since $\kappa(x) \neq 0$] is $(0, +\infty)$, contradicting the assumed finiteness of U_+. A similar argument shows that U_- cannot be finite. Furthermore, obvious modifications extend these results to the geodesics emanating from Σ in the direction of the vectors n_x. We thus conclude:

Proposition 2.2.5. *Assume that* **A1***,* **A2** *and* **A3** *hold. For all $x \in \Sigma$, the maximal geodesics γ_{l_x} and γ_{n_x} are future and past geodesically complete. The Killing field ξ never vanishes along these geodesics, except on Σ. Each geodesic never intersects itself, nor does it intersect any other geodesic in the collection $\{\gamma_{l_x}, \gamma_{n_x}\}_{x\in\Sigma}$, except possibly at zero affine parameter.* □

By Proposition 2.2.5, for any $x \in \Sigma$ the one-dimensional subspaces $\langle l_x \rangle$ and $\langle n_x \rangle$ of $N_x\Sigma$, generated by l_x and n_x respectively, are contained in the maximal domain $\mathcal{D}_\Sigma^\perp$ of the normal exponential map $\exp_\Sigma^\perp$. The total spaces of the line bundles

$$\langle l \rangle := \coprod_{x\in\Sigma} \langle l_x \rangle \quad \text{and} \quad \langle n \rangle := \coprod_{x\in\Sigma} \langle n_x \rangle \tag{2.9}$$

are therefore embedded submanifolds of $\mathcal{D}_\Sigma^\perp$. It is clear that $\exp_\Sigma^\perp \langle l \rangle = \mathcal{H}_A$ and $\exp_\Sigma^\perp \langle n \rangle = \mathcal{H}_B$, with $\mathcal{H}_A$ and $\mathcal{H}_B$ defined as in (2.6). Similarly, we let $\langle l_x \rangle_\pm := \{\rho l_x \mid \rho \gtrless 0\}$, $\langle n_x \rangle_\pm := \{\rho n_x \mid \rho \gtrless 0\}$,

$$\langle l \rangle_\pm := \coprod_{x\in\Sigma} \langle l_x \rangle_\pm \quad \text{and} \quad \langle n \rangle_\pm := \coprod_{x\in\Sigma} \langle n_x \rangle_\pm .$$

Then $\langle l \rangle_\pm$ and $\langle n \rangle_\pm$ are also embedded submanifolds of $\mathcal{D}_\Sigma^\perp$, and we define

$$\mathcal{H}_A^{R/L} := \exp_\Sigma^\perp \langle l \rangle_\pm \quad \text{and} \quad \mathcal{H}_B^{L/R} := \exp_\Sigma^\perp \langle n \rangle_\pm . \tag{2.10}$$

Clearly, $\mathcal{H}_A = \mathcal{H}_A^L \uplus \Sigma \uplus \mathcal{H}_A^R$ and $\mathcal{H}_B = \mathcal{H}_B^L \uplus \Sigma \uplus \mathcal{H}_B^R$. We are now in a position to study the submanifold properties of $\mathcal{H}_A$ and $\mathcal{H}_B$, as anticipated in the paragraph below (2.6).

In what follows, if c is a smooth curve in M whose parameter is denoted by σ, then D_σ will denote the operator of covariant differentiation along c acting on arbitrary tensor fields along c, and acting on c itself as $D_\sigma c = \dot{c} = Tc$.

Theorem 2.2.6. *Assume that* **A1***,* **A2** *and* **A3** *hold.*

- **(a)** *The restrictions of $\exp_\Sigma^\perp$ to the line bundles $\langle l \rangle$ and $\langle n \rangle$ defined in (2.9) provide $\mathcal{H}_A$ and $\mathcal{H}_B$ (respectively) with the structure of immersed, codimension-1 and null submanifolds of M.*
- **(b)** *All of the subsets $\mathcal{H}_A^{R/L}$, $\mathcal{H}_B^{L/R}$, $\mathcal{H}_A \setminus \Sigma$, $\mathcal{H}_B \setminus \Sigma$, and $(\mathcal{H}_A \cup \mathcal{H}_B) \setminus \Sigma$, are embedded, null, codimension-1 submanifolds of M, i.e. null hypersurfaces in M.*

Proof of Theorem 2.2.6. **(a)** Let $\exp_A$ and $\exp_B$ denote the (clearly, smooth) restrictions of $\exp_\Sigma^\perp$ to $\langle l \rangle$ and $\langle n \rangle$ respectively. By Proposition 2.2.5, they are also injective. It therefore remains to show that they are immersions. To this end, note first that, since Σ is spacelike, Proposition 2.1.9 may be applied to our situation, yielding an open subset $\mathcal{W}$ of $N\Sigma$ containing the image of the zero section, and an open subset $\mathcal{U}$ of M, such that $\exp_\Sigma^\perp \restriction_{\mathcal{W}} : \mathcal{W} \to \mathcal{U}$ is a diffeomorphism. In particular, it follows by the inverse function theorem that the differentials of $\exp_\Sigma^\perp$ at any point in $\mathcal{W}$ are isomorphisms. This already tells us that $\exp_A$ and $\exp_B$ are immersions at points (i.e. vectors) sufficiently close to the zero section of $N\Sigma$. We now use the isometries to 'transport' the result that $\exp_\Sigma^\perp$ is a local diffeomorphism to all other vectors in $\langle l \rangle$ and $\langle n \rangle$. Namely, since the one-parameter group elements β_τ are isometries, they 'commute' with the exponential map in the sense that

$$\exp \circ T\beta_\tau = \beta_\tau \circ \exp \quad \text{on } \mathcal{D}.$$

(Indeed, an analogous equation holds true in the context of more general *affine* maps of manifolds equipped with linear connections [KN63, Ch. 6].) Restricting to $\mathcal{D}_\Sigma^\perp$, using the fact that each $T\beta_\tau$ maps $N\Sigma$ (diffeomorphically) to itself, and taking differentials, it follows that

$$[T_{T\beta_\tau(V)} \exp_\Sigma^\perp] \circ [T_V(T\beta_\tau \restriction_{\mathcal{D}_\Sigma^\perp})] = [T_{\exp(V)}\beta_\tau] \circ [T_V \exp_\Sigma^\perp] \quad \forall\, V \in \mathcal{D}_\Sigma^\perp.$$

Since $T\beta_\tau \restriction_{\mathcal{D}_\Sigma^\perp} : \mathcal{D}_\Sigma^\perp \to N\Sigma$ is a diffeomorphism onto its image, its differentials $T_V(T\beta_\tau \restriction_{\mathcal{D}_\Sigma^\perp})$ are isomorphisms for any $V \in \mathcal{D}_\Sigma^\perp$. Hence, for all $\tau \in \mathbb{R}$, $T_V \exp_\Sigma^\perp$ is an isomorphism if and only if $T_{T\beta_\tau(V)} \exp_\Sigma^\perp$ is. By Equation (2.1) and the fact that $\kappa : \Sigma \to \mathbb{R}$ never vanishes, for any $V \in \langle l \rangle$ or $\langle n \rangle$ it is possible to find a $\tau \in \mathbb{R}$ such that $T\beta_\tau(V)$ is arbitrarily close to the zero vector, and in particular such that $T\beta_\tau(V) \in \mathcal{W}$. This completes the proof that $\exp_A$ and $\exp_B$ give $\mathcal{H}_A$ and $\mathcal{H}_B$ (respectively) the structure of immersed submanifolds of M. We next show that these immersed submanifolds are null, considering for definiteness and without loss of generality the case of $\mathcal{H}_A$. The argument is a straightforward and well-known extension of the methods used in e.g. the proof of Proposition 8.6 in [O'N83]. Any $p \in \mathcal{H}_A$ is of the form $p = \gamma_{l_x}(s) = \exp_A(sl_x)$ for some $x \in \Sigma$ and $s \in \mathbb{R}$. If $p \in \mathcal{H}_A \setminus \Sigma$ then, by rescaling l_x and/or changing the time orientation if necessary, we can assume without loss of generality that $s = 1$. By continuity, it suffices to prove that $(\exp_A)^* g$ is degenerate at all such points. We will do so by showing that the preimage X, under $T_{l_x} \exp_A$, of $\dot\gamma_{l_x}(1) \in T_p\mathcal{H}_A$ is orthogonal to every vector in the tangent space $T_{l_x}\langle l \rangle$. I.e. we aim to show that, for any $Y \in T_{l_x}\langle l \rangle$,

$$[{\exp_A}^* g](X, Y) = g(T_{l_x} \exp_A(X), T_{l_x} \exp_A(Y)) = g(\dot\gamma_{l_x}(1), T_{l_x} \exp_A(Y)) = 0. \tag{2.11}$$

Let $I \subseteq \mathbb{R}$ be an open interval around 0 and $\alpha : I \ni z \mapsto \alpha(z) \in \langle l \rangle$ be a smooth curve with $\alpha(0) = l_x$ and $D_z\alpha(0) = Y$. Then the two-parameter map

$$[0,1] \times I \ni (t,z) \mapsto \boldsymbol{x}(t,z) = \exp_A(t\alpha(z))$$

is a variation through null geodesics, i.e. each $\boldsymbol{x}(\cdot, z) : [0,1] \to M$ is a null geodesic. It follows that the vector field Υ along $\gamma_{l_x}\restriction_{[0,1]}$ defined by

$$\Upsilon(t) = D_z\boldsymbol{x}(t,0)$$

is a *Jacobi vector field* [O'N83, Ch. 8]. The reason for introducing these objects is given by the fact that

$$\Upsilon(1) = D_z\boldsymbol{x}(1,z)\big|_{z=0} = [D_z(\exp_A \circ\alpha)](0) = (T_{l_x}\exp_A \circ D_z\alpha)(0) = T_{l_x}\exp_A(Y).$$

To wit, we will show that $\Upsilon(t)$ is orthogonal to γ_{l_x} for all $t \in [0,1]$, so that, in particular, (2.11) holds. Letting $\pi : \langle l \rangle \to \Sigma$ denote the bundle projection, $\Upsilon(0) = T\pi \circ \alpha_z(0)$ is tangent to Σ, and thus normal to γ_{l_x}. By [O'N83, Prop. 4.44], covariant derivatives commute when acting on two-parameter maps such as $\boldsymbol{x}$. Therefore,

$$D_t\Upsilon(0) = D_tD_z\boldsymbol{x}(0,0) = D_zT(0),$$

where $T(z) = D_t\boldsymbol{x}(0,z)$ defines a vector field along the curve $\pi \circ \alpha$ on Σ. For all z, $T(z)$ is null as it is tangent to the null curve $t \mapsto \exp_A(t\alpha(z))$. In particular, $T(0) = \dot{\gamma}_{l_x}(0)$. It follows that

$$g(\dot{\gamma}_{l_x}(0), D_t\Upsilon(0)) = g(T(0), D_t\Upsilon(0)) = g(T(0), D_zT(0)) = \frac{1}{2}D_zg(T,T)\big|_{z=0} = 0. \quad (2.12)$$

So far we have shown that both $\Upsilon(0)$ and $D_t\Upsilon(0)$ are orthogonal to the geodesic γ_{l_x}. Consider the function

$$h(t) := g(\dot{\gamma}_{l_x}(t), \Upsilon(t)).$$

Then $h(0) = 0$ and, since $\dot{h}(t) = g(\dot{\gamma}_{l_x}(t), D_t\Upsilon(t))$ by the geodesic equation, we also know that $\dot{h}(0) = 0$. But for Jacobi vector fields along geodesics, as argued in [O'N83, Lem. 8.7], it also holds that $\ddot{h}(t) = 0 \ \forall\ t$. So it follows that $h(t) = 0 \ \forall\ t \in [0,1]$, and (2.11) is proved.

(b) The proof of this part is identical for any of the subsets given, so let S denote any one of these subsets. It is clear that analogous arguments as the ones used to prove **(a)** show that suitable restrictions of $\exp_\Sigma^\perp$ provide S with the structure of an immersed, codimension-1 and null submanifold of M. To prove that S is actually embedded we will show that any $p \in S$ has an open neighbourhood $\mathcal{U}$ in M such that $\mathcal{U} \cap S = h^{-1}\{0\}$, where

$h : \mathcal{U} \to \mathbb{R}$ is smooth and 0 is a regular value of h. Recall that any immersion is locally an embedding [Lee13, Prop. 4.25]: this implies that there is a subset T of S, containing p, which is an embedded submanifold of M. We pick a slice chart $(\mathcal{V}, \varphi = (x^\mu))$ for T around p, such that $T' := T \cap \mathcal{V} = (x^0)^{-1}\{0\}$, and the coordinate vector field $\partial/\partial x^0$ is transverse to T'. Without loss of generality, we may assume that T' is connected. Now define $f : \mathcal{V} \to \mathbb{R}$ to be the pseudo-norm of the Killing field, i.e.

$$f(q) := g_q(\xi|_q, \xi|_q) \quad \forall\, q \in \mathcal{V}.$$

ξ is null on S, so f vanishes on T'. Since ξ is non-zero everywhere on S, (2.4) implies that $\mathrm{d}f$ is non-zero and parallel to $\xi^\flat := g(\xi, \cdot)$ on T'. As we saw in in the proof of part **(a)** of this theorem, $\xi^\flat$ is (co)normal to S. Therefore, $\mathrm{d}f(X) \neq 0$ for any vector X transverse to T'. Applying this to $X = \partial/\partial x^0$, we see that

$$\mathrm{d}f\left(\frac{\partial}{\partial x^0}\right) = \frac{\partial f}{\partial x^0} \neq 0 \quad \text{on } T'.$$

Since T' was assumed connected, actually either $\partial f/\partial x^0 > 0$ or $\partial f/\partial x^0 < 0$ everywhere on T'. We will assume for definiteness that the first of these options occurs – the other one can be treated similarly. We now let

$$\mathcal{V}' := \left\{ q \in \mathcal{V} \;\middle|\; \frac{\partial f}{\partial x^0}(q) > 0 \right\}.$$

Then $\mathcal{V}'$ is open and $T' \subset \mathcal{V}'$. By the flowout theorem [Lee13, Thm. 9.20] applied to the manifold $\mathcal{V}'$, codimension-1 embedded submanifold T' and vector field $V := \partial/\partial x^0\restriction_{\mathcal{V}'}$, the maximal flow domain of V can be used to define an open submanifold $\mathcal{U}$ of $\mathcal{V}'$ containing T' and with the property that each point in $\mathcal{U}$ can be reached from T' by following an integral curve of V. In particular, on $\mathcal{U}$, f – that is, the pseudo-norm of the Killing field ξ – can only vanish on T'. It follows that $\mathcal{U} \cap S = T'$. Thus, letting $h := f\restriction_{\mathcal{U}}$, h is defined on an open neighbourhood $\mathcal{U}$ of p in M, it is smooth, has 0 as a regular value, and satisfies $\mathcal{U} \cap S = h^{-1}\{0\}$. This completes the proof of **(b)**. □

In passing, we present a remarkable result which was already proved in [KW91] (see also [Heu96]). When Σ is connected, it implies that $\kappa : \Sigma \to \mathbb{R}$ is actually a non-zero *constant*, in which case it is called the *surface gravity* of the structure $(\mathscr{M}, \xi, \Sigma)$. We provide a (to the best of our knowledge) new, direct proof of this result. Our proof avoids, in particular, the need to first obtain an explicit expression for κ^2 – as is done e.g. in the literature just mentioned.

Theorem 2.2.7. *Assume that* **A1**, **A2** *and* **A3** *hold. Then* $\kappa : \Sigma \to \mathbb{R}$ *is locally constant, and so constant on each connected component of* Σ.

Proof. We begin by observing that, taking covariant derivatives of (2.8) (omitting l_x subscripts for simplicity), and using (2.7), one obtains

$$D_U(\nabla_\xi \xi \circ \gamma) = \kappa(x) D_U(\xi \circ \gamma) = \kappa(x)^2 \dot{\gamma}. \tag{2.13}$$

As explained in the remark on p. 40, Σ is totally geodesic, and thus is *autoparallel* – meaning that parallel transport along curves in Σ preserves tangency to Σ [KN69, Sec. VII.8]. Together with the fact that, for any curve $\delta : I \to M$ contained in Σ and with $\delta(0) = x$, the parallel transport maps ${}^{\delta}P_0^z : T_xM \to T_{\delta(z)}M$ are linear isometries, it follows that the vector field along δ defined by parallely translating l_x along δ, viz.

$$\tilde{l}(z) := {}^{\delta}P_0^z(l_x),$$

always lies in $\langle l \rangle$. By construction, it satisfies

$$D_z \tilde{l} = 0. \tag{2.14}$$

We can now define a two-parameter map $\boldsymbol{y}$ by

$$\boldsymbol{y}(t, z) = \gamma_{\tilde{l}(z)}(t), \quad (t, z) \in [0, 1] \times I,$$

where $\gamma_{\tilde{l}(z)} : [0, 1] \to M$ is (a portion of) the unique geodesic with initial tangent vector $\tilde{l}(z)$. In terms of $\boldsymbol{y}$ and its variables (t, z), and letting $\tilde{\kappa} := \kappa \circ \delta$, (2.13) becomes

$$D_t[\nabla_\xi \xi \circ \boldsymbol{y}](t, z) = \tilde{\kappa}(z)^2 D_t \boldsymbol{y}(t, z). \tag{2.15}$$

We first examine the right-hand side of Equation (2.15). Upon evaluating at $t = 0$, it is clearly equal to $\tilde{\kappa}(z)^2 \tilde{l}(z)$. Further taking z-covariant derivatives, and using Equation (2.14), we obtain

$$D_z(\tilde{\kappa}^2 \tilde{l}) = (D_z \tilde{\kappa}^2)\tilde{l} + \tilde{\kappa}^2 D_z \tilde{l} = (D_z \tilde{\kappa}^2)\tilde{l}. \tag{2.16}$$

We now turn to the left-hand side of Equation (2.15). First we observe that, using the fact that Killing vector fields satisfy the 'integrability condition' $\nabla_a \nabla_b \xi^c = R_b{}^c{}_{ad} \xi^d$ (see e.g. [O'N83, p. 259]), one calculates

$$\begin{aligned}
\nabla_a(\xi^c \nabla_c \xi^b) &= (\nabla_a \xi^c)(\nabla_c \xi^b) + \xi^c \nabla_a \nabla_c \xi^b = (\nabla_a \xi^c)(\nabla_c \xi^b) + \xi^c R_c{}^b{}_{ad} \xi^d \\
\implies \nabla_e \nabla_a(\xi^c \nabla_c \xi^b) &= (\nabla_e \nabla_a \xi^c)(\nabla_c \xi^b) + (\nabla_a \xi^c)(\nabla_e \nabla_c \xi^b) + R_c{}^b{}_{ad}[(\nabla_e \xi^c)\xi^d + (\nabla_e \xi^d)\xi^c] \\
&\quad + (\nabla_e R_c{}^b{}_{ad})\xi^c \xi^d \\
&= \xi^d [R_a{}^c{}_{ed}(\nabla_c \xi^b) + R_c{}^b{}_{ed}(\nabla_a \xi^c)] + R_c{}^b{}_{ad}[(\nabla_e \xi^c)\xi^d + (\nabla_e \xi^d)\xi^c] \\
&\quad + (\nabla_e R_c{}^b{}_{ad})\xi^c \xi^d.
\end{aligned}$$

This result, in particular, implies that whenever ξ vanishes, so must the second covariant derivative of $\nabla_\xi \xi$. But then, using the identity $[\nabla_X \nabla_Y Z]^c = X^a Y^b \nabla_a \nabla_b Z^c - [\nabla_{\nabla_X Y} Z]^a$ for vector fields X, Y, Z, and Equation (2.14) again, we have

$$D_z D_t [\nabla_\xi \xi \circ \boldsymbol{y}](0, \cdot) = \left[\frac{\mathrm{d}\delta}{\mathrm{d}z}\right]^a \tilde{l}^b \, \nabla_a \nabla_b (\nabla_\xi \xi) - \nabla_{D_z \tilde{l}} (\nabla_\xi \xi) = 0.$$

Combining this result with Equation (2.16) yields that $D_z \tilde{\kappa}^2 = 0$. Since $x \in \Sigma$ and the curve δ were arbitrary, the claim that κ is locally constant on Σ follows. □

The proof of Theorem 2.2.9 below will make use of the following result (which is elementary, but to which we were unable to find a reference in the existing literature).

Lemma 2.2.8. *Let M be a manifold and $S_1, \dots, S_N$ be embedded submanifolds of codimension m in M. Let $S := \bigcup_{i=1}^N S_i$. If*

$$\overline{S \setminus S_i} \cap S_i = \emptyset \quad \forall \; i = 1, \dots, N \tag{2.17}$$

then S is an embedded submanifold of codimension m in M.

Proof. Recall that an arbitrary subset N of a manifold M is an embedded submanifold of codimension m in M if and only if it satisfies the following *local slice condition* [Lee13, Thm. 5.8]: each point in N is contained in the domain $\mathcal{U}$ of a coordinate map $\varphi = (x^\mu)$ for M such that, for some constants $c^0, \dots, c^{m-1} \in \mathbb{R}$,

$$N \cap \mathcal{U} = \{p \in \mathcal{U} \mid x^0(p) = c^1, \dots, x^{m-1}(p) = c^m\}.$$

We call the pair $(\mathcal{U}, \varphi)$ a slice chart for N (in M) around p. It follows that each S_i ($i = 1, \dots, N$) in the statement of this lemma satisfies the local slice condition. Pick a $p \in S$; then $p \in S_i$ for some i, and there is a slice chart $(\mathcal{U}_i, \varphi_i)$ for S_i around p. Now, Equation (2.17) guarantees that there exists an open neighbourhood $\mathcal{V}$ of p such that $S \cap \mathcal{V} = S_i \cap \mathcal{V}$. To see why, suppose instead that this were not the case. Then, by constructing a sequence $(\mathcal{V}_n)_{n \in \mathbb{N}}$ of open neighbourhoods of p with $\bigcap_{n=1}^\infty \mathcal{V}_n = \{p\}$, we would be able to find a sequence $(p_n)_{n \in \mathbb{N}}$ of points in $S \setminus S_i$ which converges to p. Thus p would simultaneously belong to the closure of $S \setminus S_i$ and to S_i, contradicting our hypothesis. The proof of the lemma is complete since if $\mathcal{V}_i := \mathcal{U}_i \cap \mathcal{V}$ then $S \cap \mathcal{V}_i = S_i \cap \mathcal{V}_i$ and thus $(\mathcal{V}_i, \varphi_i \restriction_{\mathcal{V}_i})$ is a slice chart for S around p. □

Theorem 2.2.9. *Assume that* **A1**, **A2** *and* **A3** *hold. Further assume that $\mathscr{M}$ is globally hyperbolic and that Σ is acausal and causally complete (Definition 2.1.36). Then $\mathcal{H}_A$ and $\mathcal{H}_B$ are embedded null submanifolds of M, i.e. null hypersurfaces.*

Proof. We prove this statement in the case of $\mathcal{H}_A$, the arguments for $\mathcal{H}_B$ being analogous. Let $\mathcal{W}$ be the open subset of $N\Sigma$ used in the proof of part **(a)**, and $\mathcal{W}_A := \mathcal{W} \cap \langle l \rangle$. Then the image of $\mathcal{W}_A$ under $\exp^{\perp}_{\Sigma}$ is an embedded submanifold of M containing Σ, which we denote by $\mathcal{H}_A^{\Sigma}$. It is clear that

$$\mathcal{H}_A = \mathcal{H}_A^L \cup \mathcal{H}_A^{\Sigma} \cup \mathcal{H}_A^R.$$

In virtue of part **(b)** of Theorem 2.2.6, this expresses $\mathcal{H}_A$ as a union of embedded, codimension-1 submanifolds. We now show that, under our hypotheses,

$$\emptyset = \overline{\mathcal{H}_A \setminus \mathcal{H}_A^L} \cap \mathcal{H}_A^L \tag{2.18}$$

$$= \overline{\mathcal{H}_A \setminus \mathcal{H}_A^R} \cap \mathcal{H}_A^R \tag{2.19}$$

$$= \overline{\mathcal{H}_A \setminus \mathcal{H}_A^{\Sigma}} \cap \mathcal{H}_A^{\Sigma}. \tag{2.20}$$

By Lemma 2.2.8, it will then follow that $\mathcal{H}_A$ is a hypersurface. The arguments proving Equation (2.18) and Equation (2.19) are identical up to a change in time orientation, so we will focus on Equations (2.19) and (2.20). We will make use of the fact that, since Σ is causally complete, it is closed by Lemma 2.1.37.

(2.19) $\mathcal{H}_A \setminus \mathcal{H}_A^R = \mathcal{H}_A^L \uplus \Sigma$ so that, since Σ is closed, $\overline{\mathcal{H}_A \setminus \mathcal{H}_A^R} = \overline{\mathcal{H}_A^L} \cup \Sigma$, and we need only check that

$$\overline{\mathcal{H}_A^L} \cap \mathcal{H}_A^R = \emptyset.$$

Since $\mathcal{H}_A^R$ and $\mathcal{H}_A^L$ are disjoint subsets of $\mathcal{H}_A \setminus \Sigma$ which are open in the subspace topology of $\mathcal{H}_A \setminus \Sigma$, this follows immediately from the following general fact: If M is a manifold, S is an embedded submanifold of codimension at least 1, and $S_1, S_2 \subseteq S$ are disjoint and open in the subspace topology of S, then there exist disjoint open sets $\mathcal{U}_1$ and $\mathcal{U}_2$ in M containing S_1 and S_2 respectively. (Proof: By the tubular neighbourhood theorem [Hir76, Sec. 4.5], there exists a vector bundle $E \xrightarrow{\pi} S$, and an embedding $f : E \to M$, such that (i) f composed with the zero section of E is the inclusion of S into M, and (ii) $f(E)$ is open in M. Then the sets $\mathcal{U}_1 := f[\pi^{-1}(S_1)]$ and $\mathcal{U}_2 := f[\pi^{-1}(S_2)]$ have the properties required.)

(2.20) We prove this in two parts, by first showing that (a) $\overline{\mathcal{H}_A \setminus \mathcal{H}_A^{\Sigma}}$ is disjoint from $\mathcal{H}_A^{\Sigma} \setminus \Sigma$, and then that (b) $\overline{\mathcal{H}_A \setminus \mathcal{H}_A^{\Sigma}} \cap \Sigma = \emptyset$. Global hyperbolicity of $\mathscr{M}$ and the causal assumptions on Σ will only be used in proving (b).

(a) For an arbitrary subset T of a topological space X, the following holds: if $(p_n)_{n\in\mathbb{N}} \subseteq T$ converges to $p \in T$ with respect to the topology of X, then it does so also with respect to the subspace topology of T. So suppose that $(p_n)_{n\in\mathbb{N}} \subset \mathcal{H}_A \setminus \mathcal{H}_A^{\Sigma}$ converges to $p \in \mathcal{H}_A^{\Sigma} \setminus \Sigma$ in the topology of M. Letting $X := M$ and $T := \mathcal{H}_A \setminus \Sigma$, we are in the general situation just described, so that $(p_n)_{n\in\mathbb{N}}$ converges with respect to the subspace topology of $\mathcal{H}_A \setminus \Sigma$, to a point assumed to be in $\mathcal{H}_A^\Sigma \setminus \Sigma$. But this contradicts the fact that (clearly) $\mathcal{H}_A^\Sigma \setminus \Sigma$ is an open subset of $\mathcal{H}_A \setminus \Sigma$ in the subspace topology of the latter.

(b) Suppose that $(p_n)_{n\in\mathbb{N}} \subset \mathcal{H}_A \setminus \mathcal{H}_A^\Sigma$ tends to $p \in \Sigma$. By passing to a subsequence if necessary, we may additionally assume that $(p_n)_{n\in\mathbb{N}}$ is entirely contained in either $\mathcal{H}_A^R \setminus \mathcal{H}_A^\Sigma$ or $\mathcal{H}_A^L \setminus \mathcal{H}_A^\Sigma$. We illustrate the first of these cases, as the second can be treated very similarly. Each p_n is the future endpoint of a future-directed null geodesic segment starting at a point $x_n \in \Sigma$. We pick a complete Riemannian metric h on M, whose distance function we denote by d^h, and let $\gamma_n : [0, b_n] \to M$ denote the h–arc-length reparametrisation of this null geodesic segment. Let $p' \in I^+(p)$, then $I^-(p')$ is a neighbourhood of p and we can assume without loss of generality that it contains all of the p_n, and thus also that $x_n \in J^-(p_n) \subseteq I^-(p')$. Hence $x_n \in J^-(p') \cap \Sigma$ which, by our hypotheses on Σ and by the global hyperbolicity of $\mathscr{M}$ (see the remark under Definition 2.1.36), is a compact set. We can therefore extract a subsequence[13] x_k which converges to a point $x \in \Sigma$. Associated with it is the subsequence $\gamma_k : [0, b_k] \to M$, which now has the property that $\gamma_k(0) \to x \in \Sigma$ and $\gamma_k(b_k) \to p \in \Sigma$. Since our Lorentzian manifold is assumed to be globally hyperbolic, according to Corollary 2.1.46 it follows that, unless the curves γ_k collapse, we can obtain a future-directed causal curve γ which connects two points of Σ. This contradicts the assumed acausality of Σ. Therefore, the proof is complete if we can argue that the curves γ_k do not collapse. Indeed, a (slightly) stronger property is already true of the original sequence γ_n. Namely, the h–arc-lengths b_n of the γ_n are bounded from below and away from zero. Indeed, if this were not the case, then we could find a subsequence γ_i such that the respective h–arc-lengths b_i tend to zero. We would then have

$$d^h(x_i, p_i) \leq b_i \to 0 \text{ as } i \to \infty$$

which in particular, by the triangle inequality, would imply that $x_i \to p$ and thus that $K := \{x_i \mid i \in \mathbb{N}\} \cup \{p\}$ is a compact subset of Σ. But this cannot be the case due to the way in which $\mathcal{H}_A^\Sigma$ was defined: Indeed, denoting by $\gamma_{l_x}^h$ (for any $x \in \Sigma$) the reparametrisation by h–arc-length of $\gamma_{l_x}\restriction_{[0,+\infty)}$, it is clear that, for any compact subset K of Σ, there is a uniform strictly positive lower bound to the h–arc-lengths of the portions of the $\gamma_{l_x}^h$ which remain entirely inside $\mathcal{H}_A^\Sigma$. More precisely, it holds that

$$\inf_{x\in K} \sup\{\ell \mid \gamma_{l_x}^h(t) \in \mathcal{H}_A^\Sigma \ \forall\, t \in [0, \ell]\} > 0.$$

[13] Abusing notation, passage to a subsequence of a sequence y_n will be denoted here and below simply by y_k and not by the more correct y_{n_k}.

Thus we have demonstrated that $b_n > c > 0$ for some c, ruling out collapse of the (sub)sequence γ_k above.

This completes the demonstration that $\mathcal{H}_A$ and $\mathcal{H}_B$ are hypersurfaces under the hypotheses of the Theorem. We already proved that they are null in Theorem 2.2.6. $\square$

Kay and Wald prove [KW91, pp. 60–61] that, under all the assumptions in Theorem 2.2.9, $\mathcal{H}^R_A \cup \Sigma \cup \mathcal{H}^L_A = \dot{J}^+(\Sigma)$, and similarly $\mathcal{H}^L_B \cup \Sigma \cup \mathcal{H}^R_B = \dot{J}^-(\Sigma)$. Since both of these sets are achronal boundaries, if one could show that there can be no timelike curve from $\mathcal{H}^L_A$ to $\mathcal{H}^R_A$ then it would immediately follow that $\mathcal{H}_A$ is achronal. Similar reasoning is of course applicable to $\mathcal{H}_B$. But that this is the case is proved in detail in [KW91] – the proof begins in the last paragraph of p. 60 there and we shall not repeat it here.

Finally, one can see that $\mathcal{H}_A$ and $\mathcal{H}_B$ are closed subsets of M. Indeed, since $\dot{J}^+(\Sigma) \cup \dot{J}^-(\Sigma)$ is closed, this reduces to checking that sequences in $\mathcal{H}^R_A \cup \mathcal{H}^L_A$ cannot have limit points in $\mathcal{H}^R_B \cup \mathcal{H}^L_B$ (and vice-versa). But this holds by essentially the same arguments in the proof of statement **(b)** in Theorem 2.2.6: sufficiently near $\mathcal{H}^{R/L}_{A/B}$, any point at which the Killing vector is null must belong to $\mathcal{H}^{R/L}_{A/B}$. Summarising, here and above we have shown:

Theorem 2.2.10. *Assume the hypotheses of Theorem 2.2.9 hold. Then $\mathcal{H}_A$ and $\mathcal{H}_B$ are achronal and closed.* $\square$

This completes our detailed study of the geometric setup introduced by Kay and Wald. Actually, for various technical reasons those authors ended up needing to make some further assumptions on the Killing vector. We now give the full definition of a *bifurcate Killing horizon structure* as it appeared in [KW91], which will be adhered to in the rest of this thesis.

Definition 2.2.11. By *spacetime with a bifurcate Killing horizon* in the sense of [KW91] we will mean the data $(\mathscr{M}, \xi, \Sigma)$, satisfying

- $\mathscr{M} = (M, g, \mathfrak{t}, \mathfrak{o})$ is an oriented, time-oriented four-dimensional globally hyperbolic spacetime;
- ξ is a non-trivial (smooth and) complete Killing vector field on M;
- $\Sigma \subset M$ is a closed, orientable two-dimensional (possibly disconnected) embedded submanifold of M on which $\xi = 0$;
- there exists a smooth spacelike Cauchy surface $\mathscr{C}$ for $\mathscr{M}$ such that $\Sigma \subset \mathscr{C}$ and ξ is timelike on $\mathscr{C} \setminus \Sigma$.

The submanifold Σ in Definition 2.2.11 is referred to as the *bifurcation surface*. The reason for this name is of course the fact that, as we have seen in this section, Σ

generates in a natural way the two transverse hypersurfaces $\mathcal{H}_A$ and $\mathcal{H}_B$, which we will henceforth refer to as the *A-horizon* and *B-horizon* respectively, and whose intersection is precisely Σ.

2.3 Null hypersurfaces and Gaussian null coordinates

As was shown in Section 2.2, null hypersurfaces occur naturally in the context of spacetimes with a bifurcate Killing horizon (Definition 2.2.11). Following [KW91], in view of applications to quantum field theory, we will later be concerned with solving initial value problems for linear (normally) hyperbolic partial differential equations when initial data is prescribed on the null hypersurfaces $\mathcal{H}_A$ and $\mathcal{H}_B$. As we will show, actually not all the properties enjoyed by these hypersurfaces will be relevant in deriving the basic PDE existence and uniqueness results needed. Indeed, we will be able to derive such results quite generically, if the ambient Lorentzian manifold is globally hyperbolic, for any hypersurface which is, null, achronal, and satisfies a certain completeness condition on the null geodesics tangent to it. This will be the topic of Section 2.5.

2.3.1 General properties of null hypersurfaces

It is therefore appropriate, at this stage, to collect a relevant selection of general facts about null hypersurfaces in Lorentzian manifolds. We follow the discussion in Kupeli [Kup87], which actually covers the more general case of null embedded submanifolds of any codimension. Unlike that reference, we don't require our submanifolds or hypersurfaces to be connected.

Let $\mathscr{N}$ be a null hypersurface in a Lorentzian manifold (M, g). Then [O'N83, Lem. 5.28], for any $p \in \mathscr{N}$ a vector in the tangent space $T_p\mathscr{N}$ is either spacelike or null, and the orthogonal (with respect to g) space $T_p\mathscr{N}^\perp \subseteq T_pM$ is actually a one-dimensional null subspace of $T_p\mathscr{N}$ containing all null vectors in $T_p\mathscr{N}$. There is therefore a canonical line bundle on $\mathscr{N}$, i.e. the *null line bundle* $K_\mathscr{N}$ of $\mathscr{N}$, viz.

$$K_\mathscr{N} := \coprod_{p\in\mathscr{N}} T_p\mathscr{N}^\perp \xrightarrow{\pi} \mathscr{N}.$$

From now on in this section, we assume that (M, g) is time-orientable, and let the global timelike vector field T define a time orientation $\mathfrak{t}$. View the metric as a map $g : TM \times_M TM \to \mathbb{R}$ (where $\times_M$ denotes fiber product), and define $\theta : K_\mathscr{N} \to \mathbb{R}$ by $\theta(n) = g(n, T|_{\pi(n)})$. Then θ is smooth, and $\theta(n) = 0 \iff n \in \mathbf{0}$ where $\mathbf{0}$ denotes the image of the zero section of $K_\mathscr{N}$. Clearly, θ has no critical points on $K_\mathscr{N} \setminus \mathbf{0}$.

Therefore, for any $\alpha \in \mathbb{R}\setminus 0$, the preimage $\theta^{-1}\{\alpha\} \subseteq K_{\mathscr{N}}$ is an embedded codimension-1 submanifold of $K_{\mathscr{N}}$ and the bundle projection $\pi : K_{\mathscr{N}} \to \mathscr{N}$ restricts to an injective smooth map onto $\mathscr{N}$. By e.g. expressing θ in a local trivialization of $K_{\mathscr{N}}$, it is also easy to see that $\pi\restriction_{\theta^{-1}\{\alpha\}} : \theta^{-1}\{\alpha\} \to \mathscr{N}$ is an immersion, and therefore a diffeomorphism. Its inverse defines a global nowhere-vanishing section of $K_{\mathscr{N}}$. Choosing α to be positive, we have proved the following result.

Proposition 2.3.1 ([Kup87, Prop. 4]). *Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold, and $\mathscr{N}$ a null hypersurface. There exists a global future-directed section of the null line bundle $K_{\mathscr{N}}$ of $\mathscr{N}$. Any two such sections n, n' are related by $n' = fn$ where f is a smooth positive function on $\mathscr{N}$. In particular, $K_{\mathscr{N}}$ is orientable as a vector bundle.* □

Remark. In fact, generalising the notion of null line bundle to null submanifolds of any codimension, so that the null line bundle consists of (the zero vectors and) null vectors both tangent and normal to the submanifold, it can easily be shown [Kup87] that Proposition 2.3.1 holds true *verbatim* if 'hypersurface' is replaced by 'embedded submanifold'.

Corollary 2.3.2 ([Kup87, Cor. 5]). *Any null hypersurface $\mathscr{N}$ in an orientable and time-orientable Lorentzian manifold (M, g) is orientable.*

Proof. Choose a global timelike vector field T on M. Let ω be a global nowhere-vanishing volume form on M. Since T is transverse to $\mathscr{N}$, the contraction $\iota_T\omega$ defines a nowhere-vanishing top-degree form on $\mathscr{N}$. □

Definition 2.3.3. Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold, and $\mathscr{N}$ be a null hypersurface with global future-directed tangent null vector field n as per Proposition 2.3.1. Any maximally extended integral curve of n is called a *null generator of $\mathscr{N}$*.

Lemma 2.3.4. *Let (M, g) be a Lorentzian manifold, and $\mathscr{N}$ be a null hypersurface. Then, if n is a (local) tangent null vector field on $\mathscr{N}$, $\nabla_X n \in T_p\mathscr{N}$ for any $X \in T_p\mathscr{N}$.*

Proof. Since n is also normal to $\mathscr{N}$, it suffices to show that $g(n, \nabla_X n) = 0$. Since $\nabla g = 0$,

$$g(n, \nabla_X n) = \frac{1}{2} X(g(n, n))$$

and, since n is null, the result follows. □

Proposition 2.3.5. *Let (M, g) be a Lorentzian manifold, and $\mathscr{N}$ be a null hypersurface with (local or global) tangent null vector field n. Then, on the subset of $\mathscr{N}$ on which n is defined, there exists a smooth real-valued function f such that $\nabla_n n = fn$. That is, the integral curves of n are null pregeodesics. In particular, the null generators of a null hypersurface in a time-oriented Lorentzian manifold are null pregeodesics and can be reparametrised to null geodesics.*

Proof. Since n is normal to $\mathscr{N}$ it suffices to show that, for any p in the domain of n, and any $X \in T_p\mathscr{N}$, $g(\nabla_n n, X) = 0$. Extend X to a local vector field $\hat{X}$ defined in an M-neighbourhood of p and tangent to $\mathscr{N}$. Similarly extend n to a local vector field $\hat{n}$ on M (defined w.l.o.g. on the same neighbourhood). Then the commutator vector field $[\hat{X}, \hat{n}]$ is also tangent to $\mathscr{N}$. We thus calculate (omitting evaluations at p for notational simplicity)

$$\begin{aligned}
g(\nabla_n n, X) &= g(\nabla_{\hat{n}}\hat{n}, \hat{X}) \\
&= \hat{n}(g(\hat{n}, \hat{X})) - g(\hat{n}, \nabla_{\hat{n}}\hat{X}) && (\nabla g = 0) \\
&= -g(\hat{n}, \nabla_{\hat{n}}\hat{X}) && (n \text{ is normal and tangent to } \mathscr{N}) \\
&= -g(\hat{n}, \nabla_{\hat{X}}\hat{n}) - g(\hat{n}, [\hat{n}, \hat{X}]) && (\nabla \text{ is torsion-free}) \\
&= 0 && (\text{Lemma 2.3.4, and } [\hat{n}, \hat{X}] \in T\mathscr{N}).
\end{aligned}$$

It is standard and a matter of simple verification that, if $c : I \to M$ is an integral curve of a vector field n satisfying $\nabla_n n \propto n$, c may be reparametrised to a curve $\tilde{c} : J \mapsto M$ satisfying $D_{\dot{\tilde{c}}}\dot{\tilde{c}} = 0$. □

We remark that something which Proposition 2.3.5 does *not* imply is that any global future-directed null vector field n tangent to an arbitrary null hypersurface can be globally rescaled to yield a *geodesic vector field* $\tilde{n}$, i.e. one satisfying $\nabla_{\tilde{n}}\tilde{n} = 0$. An obvious obstruction arises if one of the null generators is a closed curve, and any of (and therefore all) its null geodesic reparametrisations returns to the same point with a different velocity. So it is clear that, to ensure that no such obstructions occur, it is necessary to inject some additional causal assumptions. While in [Kup87, Sec. 4] a number of different sufficient conditions were derived, we will only focus on one of them as it is the only one relevant for our purposes.

Definition 2.3.6. Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold, $\mathscr{N}$ be a null hypersurface and $\mathscr{S}$ be an embedded submanifold of $\mathscr{N}$ which is spacelike and of codimension 1 in $\mathscr{N}$. Then we say that $\mathscr{S}$ is a *cross-section* of[14] $\mathscr{N}$ if there exists a diffeomorphism $\psi : \mathscr{N} \to \mathbb{R} \times \mathscr{S}$ such that the following hold:

- $\psi(\mathscr{S}) = \{0\} \times \mathscr{S}$;
- denoting by $\partial/\partial U$ the vector field on $\mathbb{R} \times \mathscr{S}$ induced from the standard $\mathrm{d}/\mathrm{d}t$ vector field on $\mathbb{R}$ by the identification $T(\mathbb{R} \times \mathscr{S}) \cong T\mathbb{R} \times T\mathscr{S}$, $\psi^*(\partial/\partial U)$ is a null, future-directed, vector field tangent to $\mathscr{N}$. Equivalently, for any $x \in \mathscr{S}$ the curve $\psi^{-1}(\cdot, x) : \mathbb{R} \to \mathscr{N}$ is a null generator of $\mathscr{N}$.

[14]Kupeli [Kup87, Def. 16] prefers to say that $\mathscr{N}$ is *causally separated by $\mathscr{S}$*.

We now show that a submanifold is a cross-section for a null hypersurface in the abstract sense of Definition 2.3.6 if and only if any null generator 'registers' on it once and never returns to it.

Lemma 2.3.7 ([Kup87, Lem. 17]). *Let $\mathscr{N}$ be a null hypersurface in a time-oriented Lorentzian manifold $(M, g, \mathfrak{t})$, and n be an arbitrary null, future-directed, vector field tangent to $\mathscr{N}$. Then a spacelike, codimension-1 embedded submanifold $\mathscr{S}$ of $\mathscr{N}$ is a cross-section of $\mathscr{N}$ according to Definition 2.3.6 if and only if any maximal integral curve of n intersects $\mathscr{S}$ at precisely one parameter value.*

Proof. We begin by noting that, since $\mathscr{S}$ is spacelike, n is transverse to $\mathscr{S}$.
($\Leftarrow$) We can assume that n is complete. Indeed, if not then, picking a complete Riemannian metric e on $\mathscr{N}$, $n/\|n\|_e$ is complete and still satisfies all our requirements. By the flowout theorem [Lee13, Thm. 9.20], the maximal flow of n, $\theta : \mathbb{R} \times \mathscr{N} \to \mathscr{N}$, restricts to a diffeomorphism $\Phi : \mathbb{R} \times \mathscr{S} \to \mathscr{N}$. Then, $\psi := \Phi^{-1}$ has the properties required by Definition 2.3.6.
($\Rightarrow$) This is obvious for the vector field $\psi^*(\partial/\partial U)$ in Definition 2.3.6 by combining the two defining properties of ψ there. But then it is also true for any positive rescaling of this field, such as n. □

The Lemma above allows to find a sufficient condition for the rescalability of a null vector field, tangent to a null hypersurface, to a geodesic one.

Proposition 2.3.8 ([Kup87, Thm. 18]). *Let $\mathscr{N}$ be a null hypersurface in a time-oriented Lorentzian manifold $(M, g, \mathfrak{t})$, and n be a global future-directed null vector field on $\mathscr{N}$. If $\mathscr{N}$ admits a cross-section $\mathscr{S}$ then n can be globally rescaled to yield a future-directed null vector field $\tilde{n}$ on $\mathscr{N}$ satisfying $\nabla_{\tilde{n}}\tilde{n} = 0$.*

Proof. Without loss of generality, assume that n is complete – since we can always rescale it as explained in the proof of Lemma 2.3.7 to obtain a complete vector field. By Proposition 2.3.5, there exists a smooth function $f : \mathscr{N} \to \mathbb{R}$ such that $\nabla_n n = fn$. Let Φ be the flow of n when restricted to $\mathscr{S}$, as in the proof of Lemma 2.3.7, let $f_\Phi := f \circ \Phi : \mathbb{R} \times \mathscr{S} \to \mathbb{R}$, and let $\pi_{\mathbb{R}}$, $\pi_{\mathscr{S}}$ denote the compositions of $\psi = \Phi^{-1} : \mathscr{N} \to \mathbb{R} \times \mathscr{S}$ with the projections onto the first and second factor in $\mathbb{R} \times \mathscr{S}$ (respectively). We verify that the (clearly, smooth) vector field

$$\tilde{n}|_p := \exp\left\{ -\int_0^{\pi_{\mathbb{R}}(p)} f_\Phi(U', \pi_{\mathscr{S}}(p))\, \mathrm{d}U' \right\} n|_p \tag{2.21}$$

is geodesic. Indeed, since Φ is the (restricted) flow of n, letting $P : \mathscr{N} \to \mathbb{R}$ denote minus the exponent in Equation (2.21), it follows from the Fundamental Theorem of

Calculus that $n(P) = f$, so that

$$\nabla_{\tilde{n}}\tilde{n} = e^{-P}\nabla_n(e^{-P}n) = e^{-2P}\nabla_n n + e^{-P}n(e^{-P})n = e^{-2P}fn - e^{-2P}n(P)n = 0.$$

Hence, $\tilde{n}$ is geodesic and its integral curves are null geodesics contained in $\mathscr{N}$. □

If $\mathscr{N}$ admits a geodesic, future-directed, null, global tangent vector field n then we call the maximally extended (in $\mathscr{N}$) integral curves of n the *null geodesic generators* of $\mathscr{N}$. Viewed as geodesics in the ambient manifold M, these may of course fail to be future or past inextendible.

Proposition 2.3.9. *Let $(M, g, \mathfrak{t})$ be globally hyperbolic. Then any null hypersurface $\mathscr{N}$ whose null generators, when reparametrised as null geodesics entirely contained in $\mathscr{N}$, are future and past inextendible as geodesics in (M, g), admits a cross-section.*

Proof. Let $\mathscr{C}$ be a smooth spacelike Cauchy surface for $(M, g, \mathfrak{t})$. Then $\mathscr{C}$ and $\mathscr{N}$ intersect transversely. Since $\mathscr{C}$ is a spacelike Cauchy surface, it is intersected precisely once by any future and past inextendible causal curve in M. In particular, by our hypothesis on the null generators of $\mathscr{N}$, it is intersected exactly once by each null generator of $\mathscr{N}$. Therefore, $\mathscr{S} := \mathscr{C} \cap \mathscr{N}$ is non-empty, and thus it is a spacelike, codimension-1 embedded submanifold of $\mathscr{N}$ intersected at precisely one parameter value by each null generator of $\mathscr{N}$. The result follows from Lemma 2.3.7. □

2.3.2 Gaussian null coordinates

We now wish to demonstrate the existence of coordinates adapted to a null hypersurface and defined in a way both natural and useful (Corollaries 2.3.13 and 2.3.14) for our applications to hyperbolic PDE theory and field theory. The proof of this result, i.e. Theorem 2.3.11 below, will require the following preliminary observation. We have already recalled the definition of the normal exponential map $\exp_S^\perp : NS \to M$ for a submanifold S of a Lorentzian (or even semi-Riemannian) manifold (M, g), and we have already recalled (Proposition 2.1.9 here, and Proposition 7.26 in [O'N83]) the important existence result for normal neighbourhoods of S in the case where the induced metric on S is non-degenerate. That normal neighbourhoods cannot exist in the sense of Proposition 2.1.9 if the submanifold is null is clear by dimensional considerations alone, since in that case the dimension of NS is strictly smaller than the dimension of M, and no open subsets of the two can be diffeomorphic (via $\exp_S^\perp$ or otherwise). However, by examining the proofs of Lemma 25 and Proposition 7.26 in [O'N83], it is clear that the only property of NS entering the arguments there is that *NS is a complementary vector subbundle to TS in $TM|_S$.* – i.e., that $T_qM = N_qS \oplus T_qS$ for all $q \in S$. Thus, Proposition 2.1.9 has the following generalisation.

Proposition 2.3.10. *Let (M, g) be a semi-Riemannian manifold and S be an embedded submanifold of codimension greater than or equal to 1 in M. Let $L \xrightarrow{\pi_L} S$ be a complementary vector subbundle to TS in $TM|_S$. Then there exists an open neighbourhood $\mathcal{Z}$ of the set of zero vectors in L such that $\exp\restriction_{\mathcal{Z}}$ is a diffeomorphism onto an open neighbourhood of S in M.* □

Remark. Restricting $\mathcal{Z}$ if necessary, we may in fact assume that each $\mathcal{Z}_q := \mathcal{Z} \cap \pi_L^{-1}\{q\}$ is star-shaped with respect to the zero vector.

While the first treatment of the coordinates described in the theorem below is often attributed to Moncrief and Isenberg [MI83] (see also [LR12, Sec. III]), they made earlier appearances in e.g. [Pen72, p. 60] and in the proof of Theorem 3.3.2 in [Fri75].

Theorem 2.3.11 (Gaussian null coordinates)**.** *Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold and $\mathscr{N}$ be a null hypersurface which admits a cross-section. Then, if $\Gamma \subseteq \mathscr{N}$ is the image of a null geodesic generator of $\mathscr{N}$, there exists a coordinate neighbourhood $\mathcal{U}$ for M, containing Γ, and coordinates (v, u, x^A) on $\mathcal{U}$, such that $\mathcal{U} \cap \mathscr{N} = \{v = 0\}$ and the metric in these coordinates takes the form*

$$g = v\varphi\, \mathrm{d}u^2 + 2\, \mathrm{d}u\, \mathrm{d}v + 2v\zeta_A\, \mathrm{d}x^A\, du + h_{AB}\, \mathrm{d}x^A\, \mathrm{d}x^B$$

for smooth functions φ, ζ_A, h_{AB}.

Proof. Let $\mathscr{S}$ be a cross-section for $\mathscr{N}$, and let n be a future-directed (geodesic or not) tangent null vector field on $\mathscr{N}$. Use $\mathscr{S}$ and the flow of n to construct a diffeomorphism $\Phi : \mathcal{D} \to \mathscr{N}$ where $\mathcal{D}$ is an open subset of $\mathbb{R} \times \mathscr{S}$ containing $\{0\} \times \mathscr{S}$ (unlike in the proof of Lemma 2.3.7, since n may be incomplete it is no longer guaranteed that $\mathcal{D} = \mathbb{R} \times \mathscr{S}$). Then, since $\mathscr{S}$ is a cross-section, there exists a $x \in \mathscr{S}$ such that, letting $\mathcal{D}_x := \mathcal{D} \cap (\mathbb{R} \times \{x\})$, Γ in the statement of the theorem is equal to $\Phi(\mathcal{D}_x)$. Now, let $(\mathcal{W}, (x^A))$ be a coordinate chart for $\mathscr{S}$ around x. Then, letting

$$\mathcal{V} := \Phi(\mathcal{D} \cap (\mathbb{R} \times \mathcal{W})) \quad \text{and} \quad \chi = (u, x^A) = [\mathrm{id}_{\mathbb{R}} \times (x^A)] \circ \Phi^{-1}\restriction_{\mathcal{V}},$$

$(\mathcal{V}, \chi)$ is a coordinate chart for $\mathscr{N}$ whose domain contains Γ together with the images of all other null geodesic generators of $\mathscr{N}$ which intersect $\mathscr{S}$ in $\mathcal{W}$. By construction, $\partial/\partial u = n\restriction_{\mathcal{V}}$, while the vectors $\partial/\partial x^A$ are spacelike and normal to $\partial/\partial u$, i.e.

$$g\left(\frac{\partial}{\partial u}, \frac{\partial}{\partial u}\right) = g\left(\frac{\partial}{\partial u}, \frac{\partial}{\partial x^A}\right) = 0.$$

Next, let l be a <u>null</u> vector field along $\mathcal{V}$, smooth as a section of the pullback bundle $TM|_{\mathcal{V}}$, and satisfying

$$g(l, n) = \alpha \quad \text{and} \quad g\left(l, \frac{\partial}{\partial x^A}\right) = 0, \tag{2.22}$$

for a strictly positive $\alpha \in C^\infty(\mathcal{V})$. That such an l exists (for any α), and is in fact uniquely determined by these requirements, follows by a reasoning analogous to the one used in the paragraph above Proposition 2.3.1. We will actually take for simplicity $\alpha \equiv 1$. It follows that l is (future-directed and) transverse to $\mathcal{V}$, and thus generates a complementary rank-1 vector subbundle L to the tangent bundle $T\mathcal{V}$ in $TM|_{\mathcal{V}}$. By Proposition 2.3.10, there is an open neighbourhood $\mathcal{Z}$ in L of the set of zero vectors, and an open neighbourhood $\mathcal{U}$ of $\mathcal{V}$ in M, such that $\exp\restriction_{\mathcal{Z}} : \mathcal{Z} \to \mathcal{U}$ is a diffeomorphism. Let $\phi : L \to \mathbb{R} \times \mathcal{V}$ denote the global trivialization of L induced by the global basis $\{l\}$. Then $\exp \circ \phi^{-1}\restriction_{\phi(\mathcal{Z})}$ is a diffeomorphism from an open set in $\mathbb{R} \times \mathcal{V}$ containing $\{0\} \times \mathcal{V}$ onto $\mathcal{U}$. Its inverse $\psi : \mathcal{U} \to \phi(\mathcal{Z})$ can be then composed with $\mathrm{id}_{\mathbb{R}} \times \chi$ to yield a chart

$$(\mathcal{U}, (v, \chi)) = (\mathcal{U}, (v, u, x^A)).$$

By construction, $\partial/\partial v$ is a geodesic vector field on $\mathcal{U}$ which is null (as it is equal to l) on $\mathcal{V} \subset \mathcal{U}$; it follows then that it is null everywhere on $\mathcal{U}$, i.e. that in this coordinate system

$$g_{vv} = g\left(\frac{\partial}{\partial v}, \frac{\partial}{\partial v}\right) \equiv 0.$$

On the other hand, we already knew that $g_{uu} = g_{uA} = g_{vA} = g_{uv} - 1 = 0$ on $\mathcal{V} \subseteq \mathscr{N}$. In particular, since $\mathcal{V}$ is given, in this coordinate chart, by $\{v = 0\}$, there exist functions $\varphi, \zeta_A \in C^\infty(\mathcal{U})$ such that

$$g_{uu} = v\varphi \quad \text{and} \quad g_{uA} = v\zeta_A.$$

Finally, let the index a indicate either a u-component or an x^A-component. We calculate (using $\nabla g = 0$ in the first step, and the torsion-free condition together with $\nabla_{\partial_v}\partial_v = 0$ in the second)

$$\begin{aligned}\partial_v(g_{va}) &= \partial_v\big(g(\partial_v, \partial_a)\big) = g(\nabla_{\partial_v}\partial_v, \partial_a) + g(\partial_v, \nabla_{\partial_v}\partial_a) = g(\partial_v, \nabla_{\partial_a}\partial_v)\\ &= \frac{1}{2}\partial_a\big(g(\partial_v, \partial_v)\big) = 0.\end{aligned}$$

Together with the initial condition $g_{vA} = g_{vu} - 1 = 0$ on $\mathcal{V}$, and by the remark following Proposition 2.3.10, we conclude that actually $g_{vA} = 0$ and $g_{vu} = 1$ throughout $\mathcal{U}$. The claimed form of the metric follows. □

Definition 2.3.12. We call the coordinate chart $(\mathcal{U}, (v, u, x^A))$ in Theorem 2.3.11 a *Gaussian null coordinate chart for* $\mathscr{N}$ *around* Γ.

The reason for us defining Gaussian null coordinates for null hypersurfaces lies in the fact that wave and wave-like equations in these coordinates take a special form, in which no repeated derivatives in directions transverse to the hypersurface occur at the location of the hypersurface. Namely, let $\Box : C^\infty(M) \to C^\infty(M)$ be the *d'Alembert operator* given in abstract index notation and in a choice of local coordinates (respectively) by

$$\Box\phi := \nabla_a \nabla^a \phi = \frac{1}{\sqrt{-|g|}} \partial_\mu \big(\sqrt{-|g|} g^{\mu\nu} \partial_\nu \phi\big). \tag{2.23}$$

Then, a straightforward calculation leads to the following important result.

Corollary 2.3.13. *Let (M, g) be a four-dimensional Lorentzian manifold. If there is a coordinate chart for M such that g takes the form in Equation (2.22), then $|g| = |h| := \det h_{AB}$ and there is a smooth function Ψ such that*

$$\begin{aligned}\Box =& v\left(v\frac{h_{22}[\zeta_1]^2 - 2h_{12}\zeta_1\zeta_2 + h_{11}[\zeta_2]^2)}{|h|} - \varphi\right)\frac{\partial^2}{\partial v^2} \\ &- 2v\frac{h_{22}\zeta_1 - h_{12}\zeta_2}{|h|}\frac{\partial^2}{\partial v\,\partial x^1} - 2v\frac{h_{11}\zeta_2 - h_{12}\zeta_1}{|h|}\frac{\partial^2}{\partial v\,\partial x^2} \\ &+ 2\frac{\partial^2}{\partial u\,\partial v} + \Psi\frac{\partial}{\partial v} \\ &+ \textit{terms in } \partial_{x^1},\, \partial_{x^2},\, \partial^2_{x^1},\, \partial^2_{x^2},\, \partial^2_{x^1x^2}.\end{aligned}$$

In the lower or higher dimensional cases, it still holds that there are no ∂_u, ∂^2_u, $\partial^2_{ux^A}$ terms, that the coefficients in front of ∂^2_v and $\partial^2_{vx^A}$ vanish when $v = 0$, that the coefficient in front of ∂^2_{uv} is equal to 2, and that the coefficient in front of ∂_v is a smooth function. $\Box$

Corollary 2.3.14. *Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold and $\mathscr{N}$ be a null hypersurface which admits a cross-section. Let $(\mathcal{U}, (v, u, x^A))$ be a Gaussian null coordinate chart for $\mathscr{N}$ as in Theorem 2.3.11. Then, for any vector field X on M and any two $q, F \in C^\infty(M)$, the equation $(\Box + X + q)\phi = F$, when evaluated on $\mathscr{N} \cap \mathcal{U} = \{v = 0\}$, takes the form*

$$2\frac{\partial}{\partial u}\left(\frac{\partial\phi}{\partial v}\right) + \big(\Psi + X^v\big)\frac{\partial\phi}{\partial v} = F + D^{(1)}_0\phi \tag{2.24}$$

where $\Psi \in C^\infty(\mathcal{U})$ is as in Corollary 2.3.13, and $D^{(1)}_0$ is a linear second-order differential operator involving u- and x^A–derivatives only, with coefficients independent of ϕ. In general, letting

$$\phi^{(n)}_v := \frac{\partial^n\phi}{\partial v^n}\restriction_{\mathscr{N}\cap\mathcal{U}},$$

the differential consequence obtained by taking the n-th order ($n \geq 1$) v-derivative of $(\Box + X + q)\phi = F$, evaluated on $\mathscr{N} \cap \mathcal{U}$, takes the form

$$2\frac{\partial}{\partial u}\left[\phi^{(n)}_v\right] + \Omega^{(n)}\phi^{(n)}_v = \frac{\partial^n F}{\partial v^n} + \sum_{i=0}^{n-1} D^{(n)}_i \phi^{(i)}_v \tag{2.25}$$

where $\Omega^{(n)} \in C^\infty(\mathcal{U})$ is independent of ϕ and each $D_i^{(n)}$ is a linear differential operator of order at most 2 involving u- and x^A-derivatives only, with coefficients independent of ϕ. $\square$

2.4 Generalities on Green hyperbolic and normally hyperbolic differential operators

This section will quickly review some key results in the theory of a special class of second order, linear, partial differential operators with smooth coefficients acting on section of vector bundles over Lorentzian manifolds. The *normally hyperbolic* operators – also sometimes called *wave operators*[15] – which comprise this class, when acting on sections of vector bundles over *globally hyperbolic* spacetimes, admit a well-posed *initial value formulation* on (smooth) spacelike Cauchy surfaces. This means that, in considering the equation

$$Pu = f, \tag{2.26}$$

where $P : \Gamma^\infty(F_1) \to \Gamma^\infty(F_2)$ is the differential operator in question and f is a given (sufficiently regular, e.g. smooth) section of $F_2 \to M$, given such a Cauchy surface $\mathscr{C}$ and a pair u_0, u_1 of (sufficiently regular, e.g. smooth) sections of $F_1|_{\mathscr{C}}$ – called *Cauchy data* – on such a Cauchy surface $\mathscr{C}$, there exists a unique solution u of Equation (2.26) whose restriction to $\mathscr{C}$ equals u_0, and whose first (covariant) derivative in the direction given by the global unit normal vector field along $\mathscr{C}$ equals u_1. Notice that since $\mathscr{C}$ is assumed spacelike, this vector is everywhere transverse to $\mathscr{C}$ – and indeed we could have chosen any other smooth vector field along $\mathscr{C}$ and transverse to it. The regularity of the solution u to this *Cauchy problem* depends on the regularity of u_1, u_2 and of the 'inhomogeneity' f, but (i) u is smooth if they all are, and (ii) it depends continuously on them if they are allowed to vary within appropriately topologized spaces of sections with given regularity.

2.4.1 Green hyperbolicity

Normally hyperbolic operators (to be defined formally below), which are always of second order, when acting on sections of vector bundles over globally hyperbolic spacetimes belong to the wider class of *Green hyperbolic* partial differential operators on Lorentzian manifolds, which were studied in detail in [BG12, Kha14, Bär15] and which may be of any order. Indeed, the fact that they are Green hyperbolic is often all that is needed to

[15] Not to be confused with the Hilbert space linear operators arising in scattering theory – the *Møller operators* – sometimes indicated by the same name.

construct well-behaved (particularly relative to the requirement of *Einstein causality*) *quantum* field theories based on the classical ones which they define. Such operators are axiomatically defined by the requirement that they (and their formal adjoints, see Definition 2.4.2) admit *advanced* and *retarded Green operators* – often referred to as advanced and retarded *propagators* of *Green's functions*, particularly in the physics literature. As we will see in the definition below, the latter are particular kinds of 'inverses' of the differential operator with special support properties relative to the causal structure defined by the underlying Lorentzian metric.

Definition 2.4.1. Let $\mathscr{M} = (M, g, \mathfrak{t})$ be a spacetime and $P : \Gamma^\infty(F_1) \to \Gamma^\infty(F_2)$ be a linear differential operator (with smooth coefficients). A *retarded* (+) or advanced (−) Green operator for P – relative to the causal structure on $\mathscr{M}$ – is a linear map $E^\pm : \Gamma_0^\infty(F_2) \to \Gamma^\infty(F_1)$ such that

(i) $E^\pm \circ P{\restriction}_{\Gamma_0^\infty(F_1)} = \mathrm{id}_{\Gamma_0^\infty(F_1)}$;

(ii) $P \circ E^\pm{\restriction}_{\Gamma_0^\infty(F_2)} = \mathrm{id}_{\Gamma_0^\infty(F_2)}$;

(iii) $\mathrm{supp}(E^\pm g) \subseteq J^\pm(\mathrm{supp}\, g)$ for all $g \in \Gamma_0^\infty(F_2)$.

Definition 2.4.2. Let (M, g) be a Lorentzian manifold and $\mathrm{d}\mu_g$ be the associated volume density. Let $P : \Gamma^\infty(F_1) \to \Gamma^\infty(F_2)$ be a differential operator between vector bundles over a common field $\mathbb{K}$. The *formal adjoint* or *formally dual operator* P^* of P (with respect to $\mathrm{d}\mu_g$) is the differential operator $P^* : \Gamma^\infty(F_2^*) \to \Gamma^\infty(F_1^*)$ uniquely defined by

$$\int_M \psi(P\varphi)\, \mathrm{d}\mu_g = \int_M [P^*\psi](\varphi)\, \mathrm{d}\mu_g$$

for all $\psi \in \Gamma^\infty(F_2^*)$ and $\varphi \in \Gamma^\infty(F_1^*)$ such that $\mathrm{supp}\,\psi \cap \mathrm{supp}\,\varphi$ is compact. If $F_1 = F_2 = F$ and we have a preferred vector bundle isomorphism $F \to F^*$ covering the identity, then P^* can be regarded as an operator $\Gamma^\infty(F) \to \Gamma^\infty(F)$, and P is *formally self-adjoint* relative to this identification if $P = P^*$. If the isomorphism $F \to F^*$ is realised by a smooth, fiberwise non-degenerate bilinear pairing $\langle \cdot, \cdot \rangle : F \times_M F \to \mathbb{K}$, then P is formally self-adjoint if and only if

$$\int_M \langle \psi, P\varphi \rangle\, \mathrm{d}\mu_g = \int_M \langle P\psi, \varphi \rangle\, \mathrm{d}\mu_g$$

for all $\psi, \varphi \in \Gamma^\infty(F)$ such that $\mathrm{supp}\,\psi \cap \mathrm{supp}\,\varphi$ is compact.

P^* allows one to linearly extend the action of P to distributional sections. Recall (see also Appendix C) that, in the presence of a preferred smooth volume density, $\Gamma^{-\infty}(F_i)$ (for $i = 1, 2$) can be defined as the continuous dual space of $\Gamma_0^\infty(F_i^*)$. Then, $P : \Gamma^{-\infty}(F_1) \to$

$\Gamma^{-\infty}(F_2)$ is defined by

$$[P\tau](\varphi) = \tau(P^*\varphi) \quad \forall\, \varphi \in \Gamma_0^\infty(F_2^*),$$

and is continuous in the standard (weak-*) topology on distributional sections. We can thus speak of *distributional solutions of* P, meaning elements in the kernel of this extension of P.

Definition 2.4.3 (Green hyperbolic operators)**.** If $(M, g, \mathfrak{t})$ is a spacetime and $P : \Gamma^\infty(F_1) \to \Gamma^\infty(F_2)$ is a linear differential operator (with smooth coefficients) such that both P and P^* have advanced and retarded Green operators, then P is said to be *Green hyperbolic*. In particular, P is Green hyperbolic if and only if P^* is.

If P is a Green hyperbolic operator on a globally hyperbolic spacetime, we may use its retarded Green operators to solve a 'Cauchy problem' of sorts in which, given an inhomogeneity $g \in \Gamma_0^\infty(F_2)$, we seek a solution u to $Pu = g$ which vanishes to infinite order on a spacelike Cauchy surface $\mathscr{C}^-$ such that $\operatorname{supp} g \subseteq J^+(\mathscr{C}^-)$. Clearly, $u := E^+g$ does the job. Similarly, E^-g solves the analogous 'homogeneous Cauchy problem' posed on a spacelike Cauchy surface $\mathscr{C}^+$ with $\operatorname{supp} g \subseteq J^-(\mathscr{C}^+)$.

Recall now the subspaces of (distributional) sections defined according to their support properties in Definition 2.1.40. Recall that a differential operator is always support non-increasing. The following theorem collects results proved in [Bär15], and we also refer to that paper for careful definitions of all topologies involved.

Theorem 2.4.4. *If* $P : \Gamma^\infty(F_1) \to \Gamma^\infty(F_2)$ *is a Green hyperbolic operator on a globally hyperbolic spacetime, then it is bijective as a map* $\Gamma^\infty_{\mathrm{pc}}(F_1) \to \Gamma^\infty_{\mathrm{pc}}(F_2)$ *and as a map* $\Gamma^\infty_{\mathrm{fc}}(F_1) \to \Gamma^\infty_{\mathrm{fc}}(F_2)$*. The resulting unique inverses are extensions, denoted by* $\overline{E^\pm}$*, of* $E^\pm$*, which still enjoy the support property* $\operatorname{supp}(\overline{E^\pm} g) \subseteq J^\pm(\operatorname{supp} g)\ \forall\ g \in \Gamma^\infty_{\mathrm{pc/fc}}(F_2)$*. In particular, the original Green operators of* P *as in Definition 2.4.1 are unique. By duality, similar statements hold when* P *acts on distributional sections. The original Green operators and all such extensions are continuous.* □

We denote the extended Green operators on distributional sections by $\widehat{E^+} : \Gamma^{-\infty}_{\mathrm{pc}}(F_2) \to \Gamma^{-\infty}_{\mathrm{pc}}(F_1)$ and $\widehat{E^-} : \Gamma^{-\infty}_{\mathrm{fc}}(F_2) \to \Gamma^{-\infty}_{\mathrm{fc}}(F_1)$.

If P is Green hyperbolic and the spacetime is globally hyperbolic then we can use its (unique by Theorem 2.4.4) advanced and retarded Green operators to define the important *causal Green operator* – also variously called *causal propagator* or (somewhat improperly) *causal Green's function*. If P is self-adjoint, the causal Green operator will allow us to define a symplectic structure on the space of *spatially compact solutions* $S := \ker P\restriction_{\Gamma^\infty_{\mathrm{sc}}(F)}$, and this is important for quantization.

Definition 2.4.5. Let $(M, g, \mathfrak{t})$ be a spacetime and $P : \Gamma^\infty(F_1) \to \Gamma^\infty(F_2)$ be a Green hyperbolic operator. The causal propagator $E : \Gamma_0^\infty(F_2) \to \Gamma^\infty(F_1)$ is defined by

$$E \coloneqq E^- - E^+$$

and clearly satisfies $E[\Gamma_0^\infty(F_2)] \subseteq \Gamma_{\mathrm{sc}}^\infty(F_1)$. If the spacetime is also globally hyperbolic, using the extended Green operators $\overline{E^\pm}$ we may also define

$$\overline{E} \coloneqq \overline{E^-} - \overline{E^+} : \Gamma_{\mathrm{tc}}^\infty(F_2) \to \Gamma^\infty(F_1),$$

which is an extension of E.

If P is a Green hyperbolic operator on a globally hyperbolic spacetime, the (extended) Green operators $\overline{E^\pm} : \Gamma_{\mathrm{pc/fc}}^\infty(F_2) \to \Gamma_{\mathrm{pc/fc}}^\infty(F_1)$ for P, and (extended) Green operators for P^* which we will denote by $\overline{G^\pm} : \Gamma_{\mathrm{pc/fc}}^\infty(F_1^*) \to \Gamma_{\mathrm{pc/fc}}^\infty(F_2^*)$, are in duality in the following sense. Note that the roles of 'advanced' and 'retarded' get interchanged.

Proposition 2.4.6 (Duality between Green operators [Bär15, Lem. 3.21] and [Kha14])**.** *Let $(M, g, \mathfrak{t})$ be globally hyperbolic and $P : \Gamma^\infty(F_1) \to \Gamma^\infty(F_2)$ be a linear Green hyperbolic operator. Then, with the notation in the paragraph above,*

$$\int_M \left[\overline{G^+}\alpha\right](\beta)\, \mathrm{d}\mu_g = \int_M \alpha\left(\overline{E^-}\beta\right) \mathrm{d}\mu_g$$

for all $\alpha \in \Gamma_{\mathrm{pc}}^\infty(F_1^*)$ and $\beta \in \Gamma_{\mathrm{fc}}^\infty(F_2)$ such that $\mathrm{supp}(\overline{G^+}\alpha) \cap \mathrm{supp}(\overline{E^-}\beta)$ is compact. Similarly,

$$\int_M \left[\overline{G^-}\alpha\right](\beta)\, \mathrm{d}\mu_g = \int_M \alpha\left(\overline{E^+}\beta\right) \mathrm{d}\mu_g$$

for all $\alpha \in \Gamma_{\mathrm{fc}}^\infty(F_1^*)$ and $\beta \in \Gamma_{\mathrm{pc}}^\infty(F_2)$ such that $\mathrm{supp}(\overline{G^-}\alpha) \cap \mathrm{supp}(\overline{E^+}\beta)$ is compact. In particular, if $F_1 = F_2 = F$ and P is formally self-adjoint with respect to a fiberwise non-degenerate bilinear pairing $\langle\cdot,\cdot\rangle : F \times_M F \to \mathbb{K}$, then $\overline{G^\pm} = \overline{E^\pm}$ and the above implies that

$$\int_M \left\langle \overline{E^\pm}\psi, \varphi \right\rangle \mathrm{d}\mu_g = \int_M \left\langle \psi, \overline{E^\mp}\varphi \right\rangle \mathrm{d}\mu_g$$

whenever $\psi \in \Gamma_{\mathrm{pc/fc}}^\infty(F)$ and $\varphi \in \Gamma_{\mathrm{fc/pc}}^\infty(F)$ are such that $\mathrm{supp}(\overline{E^\pm}\psi) \cap \mathrm{supp}(\overline{E^\mp}\varphi)$ is compact. In particular, in this latter case we have

$$\int_M \langle E\psi, \varphi\rangle\, \mathrm{d}\mu_g = -\int_M \langle \psi, E\varphi\rangle\, \mathrm{d}\mu_g$$

for all $\psi, \varphi \in \Gamma_0^\infty(F)$. □

It is evident that E maps test sections to spatially compact *solutions* of $Pu = 0$, and that $\mathrm{supp}(Eg) \subseteq J(\mathrm{supp}\, g)$ for all $g \in \Gamma_0^\infty(F_2)$ – justifying the terminology. Actually, in the presence of global hyperbolicity one can prove further statements about E which, together with the ones just made, can be conveniently summarised by means of an exact sequence as was done in [BGP07, Thm. 3.4.7] for normally hyperbolic operators, and generalised to Green operators and distributional sections in [Bär15, Thm. 3.22 & Thm. 4.3]. The last surjection in each sequence in Theorem 2.4.7 below was not explicitly included in [Bär15], or even in [BGP07] in the special normally hyperbolic case. In the case of normally hyperbolic operators, it is a result of Corollary 5 in [Gin09], while in the case of general Green operators the (simple) proof is in [Kha14, Lem. 2.1 & Prop. 2.1], but we will repeat it below in the interest of being self-contained.

Theorem 2.4.7. *If $(M, g, \mathfrak{t})$ is globally hyperbolic and P is a Green hyperbolic operator then the sequence of vector spaces*

$$\{0\} \longrightarrow \Gamma_0^\infty(F_1) \xrightarrow{P} \Gamma_0^\infty(F_2) \xrightarrow{E} \Gamma_{\mathrm{sc}}^\infty(F_1) \xrightarrow{P} \Gamma_{\mathrm{sc}}^\infty(F_2) \longrightarrow \{0\} \tag{2.27}$$

is exact, implying in particular that E is onto $S := \ker P\restriction_{\Gamma_{\mathrm{sc}}^\infty(F_1)}$, that $\ker E = P[\Gamma_0^\infty(F_1)]$ and therefore also that $S \cong \Gamma_0^\infty(F_2)/P[\Gamma_0^\infty(F_1)]$. With $\widehat{E} := \widehat{E^-} - \widehat{E^+} : \Gamma_0^{-\infty}(F_2) \to \Gamma_{\mathrm{sc}}^{-\infty}(F_1)$, the sequence

$$\{0\} \longrightarrow \Gamma_0^{-\infty}(F_1) \xrightarrow{P} \Gamma_0^{-\infty}(F_2) \xrightarrow{\widehat{E}} \Gamma_{\mathrm{sc}}^{-\infty}(F_1) \xrightarrow{P} \Gamma_{\mathrm{sc}}^{-\infty}(F_2) \longrightarrow \{0\} \tag{2.28}$$

is also exact. □

Proof. We prove the last surjection. The argument is identical in the smooth and distributional case, but we show it in the smooth case for definiteness. Pick two Cauchy surfaces $\mathscr{C}^-$ and $\mathscr{C}^+$ with $\mathscr{C}^\pm \subset I^\pm(\mathscr{C}^\mp)$. Then $\{I^+(\mathscr{C}^-), I^-(\mathscr{C}^+)\}$ is an open cover of M, and we can pick a partition of unity $\{\chi_+, \chi_-\}$ subordinate to this cover. Let $\phi \in \Gamma_{\mathrm{sc}}^\infty(F_2)$ so there is a compact set K such that $\operatorname{supp}\phi \subseteq J(K)$. It follows that

$$\begin{aligned}\operatorname{supp}(\chi_+\phi) \subseteq J(K) \cap J^+(\mathscr{C}^-) &= [J^+(K) \cap J^+(\mathscr{C}^-)] \cup [J^-(K) \cap J^+(\mathscr{C}^-)] \\ &\subseteq J^+(K) \cup J^+\big(J^-(K) \cap J^+(\mathscr{C}^-)\big) \\ &= J^+\big(K \cup [J^-(K) \cap J^+(\mathscr{C})^-]\big).\end{aligned}$$

Since $\mathscr{C}^-$ is a Cauchy surface and K is compact, the set in the last line is the causal future of a compact set, therefore $\chi_+\phi$ belongs to $\Gamma_{\mathrm{ret}}^\infty(F_2)$. In a similar way one sees that $\chi_-\phi \in \Gamma_{\mathrm{adv}}^\infty(F_2)$. Therefore, $\overline{E^\pm}(\chi_\pm\phi)$ also belongs to $\Gamma_{\mathrm{ret/adv}}^\infty(F_2)$, the sum $\overline{E^+}(\chi_+\phi) + \overline{E^-}(\chi_-\phi)$ is in $\Gamma_{\mathrm{sc}}^\infty(F_2)$, and the claim follows from

$$\phi = \chi_+\phi + \chi_-\phi = P\overline{E^+}(\chi_+\phi) + P\overline{E^-}(\chi_-\phi) = P[\overline{E^+}(\chi_+\phi) + \overline{E^-}(\chi_-\phi)]. \qquad \square$$

Finally, the results in Proposition 2.4.6 and the exactness of the sequence in Theorem 2.4.7 can be combined to equip the space of smooth, spatially compact solutions to a Green hyperbolic operator with a symplectic structure, in the formally self-adjoint case.

Theorem 2.4.8 (Symplectic structure on spatially compact solutions of formally self-adjoint operators, [BG12, Prop. 3.4])**.** *Let $(M, g, \mathfrak{t})$ be globally hyperbolic and let $F \to M$*

be a vector bundle over a field $\mathbb{K}$*, endowed with a non-degenerate fiberwise inner product* $\langle\cdot,\cdot\rangle$*. If* P *is formally self-adjoint relative to this inner product then the antisymmetric* $\mathbb{K}$*-bilinear map* $\Gamma_0^\infty(F)\times\Gamma_0^\infty(F)\to\mathbb{K}$ *given by*

$$(\psi,\varphi)\mapsto\int_M\langle\psi,E\varphi\rangle\,\mathrm{d}\mu_g$$

descends to a symplectic form on $S:=\ker P_{\Gamma_{\mathrm{sc}}^\infty(F)}\cong\Gamma_0^\infty(F)/P[\Gamma_0^\infty(F)]$*.*

It is well-known that, in the smaller class of normally hyperbolic operators (see the next subsection), the expression for the symplectic product between two spatially compact solutions can be expressed as an integral over any smooth Cauchy surface involving the value of the sections and of their (first, covariant relative to a 'P-compatible' connection) normal derivatives on that Cauchy surface, see for instance [BGP07, Lem. 3.22 & Lem. 4.7.7]. It is perhaps less well-known that a generalisation of this result still holds true in the Green hyperbolic case, but this time the 'differential data' of the two solutions to be integrated along the Cauchy surface may contain higher-order derivatives. This is because a general linear differential operator $P:\Gamma^\infty(F_1)\to\Gamma^\infty(F_2)$ and its formal adjoint $P^*:\Gamma^\infty(F_2^*)\to\Gamma^\infty(F_1^*)$ satisfy a relation which, assuming spacetime orientability for simplicity, takes the form

$$[\alpha(P\varphi)-(P^*\alpha)(\varphi)]\,\mathrm{vol}_g=\mathrm{d}\mathcal{G}[\alpha,\varphi], \tag{2.29}$$

where, with $\dim M=n+1$, $\mathcal{G}:\Gamma^\infty(F_2^*)\times\Gamma^\infty(F_1)\to\Omega^n(M;\mathbb{K})$ is a bilinear, bidifferential, form-valued operator. Equation (2.29) is the *Green–Vinogradov formula*, introduced in a local sense in [Vin84a, Vin84b], and see Theorem 6.2 in [AB02] for a proof of the global existence of such a $\mathcal{G}$. The pullback of $\mathcal{G}[\alpha,\varphi]$ to an arbitrary smooth Cauchy surface $\mathscr{C}$ can be integrated and the remarkable result is that, if the Cauchy surface is given the appropriate 'future orientation' and whenever α and φ are spatially compact solutions of P^* and P respectively, the resulting integral,

$$\int_{\mathscr{C}}\iota^*\mathcal{G}[\alpha,\varphi],\quad\text{is always equal to}\quad\int_M\beta(\varphi)\,\mathrm{vol}_g=\int_M\alpha(\psi)\,\mathrm{vol}_g$$

for any $\beta\in\Gamma_0^\infty(F_1^*)$ such that $G\beta=\alpha$ and any $\psi\in\Gamma_0^\infty(F_2)$ such that $E\psi=\varphi$. In particular, the integral in Theorem 2.4.8 defining the linear symplectic structure on S in the formally self-adjoint case may be always re-expressed as an integral over a smooth Cauchy surface involving only the values and derivatives of the solutions themselves – and no test functions. We refer to Section 2.5 in [Kha14], and specifically to Lemma 2.5 there, for further details, and to the rest of that paper for a study of the ramifications of this in the context of Lagrangian field theories.

2.4.2 Normal hyperbolicity

As we already mentioned above, normally hyperbolic operators on Lorentzian manifolds are particular kinds of second-order linear partial differential operators with smooth coefficients which, when acting on sections of vector bundles over globally hyperbolic spacetimes, are automatically Green hyperbolic. As a matter of fact, that this is the case is a deep result the proof of which was a major achievement starting with the seminal work of Leray [Ler53] (in which the notion of global hyperbolicity was first introduced), and continuing with Lichnerowicz' and Choquet-Bruhat's contributions [Lic61, CB67] and with Friedlander's book [Fri75] in which local results are achieved by a different method. The more recent textbooks [BGP07] and [Rin09] fill gaps in these earlier presentations and give complete accounts of these results in a modern language.

A normally hyperbolic operator is, roughly speaking, a second-order differential operator whose highest order coefficients in any choice of coordinates are the components of the inverse metric tensor $g^{-1} \in \Gamma^\infty(TM \otimes TM)$. A more formal definition requires the notion of the *principal symbol* of a general linear differential operator $L : \Gamma^\infty(F_1) \to \Gamma^\infty(F_2)$, which we now give. For further details, and coordinate-free definitions, see e.g. [Nic07, Ch. 10].

Definition 2.4.9 (Principal symbol)**.** Let M be a smooth manifold of dimension d, let $F_1 \to M$ and $F_2 \to M$ be vector bundles of rank N_1 and N_2 respectively and over the same field $\mathbb{K}$, and let L be a linear differential operator with smooth coefficients and of order $k \in \mathbb{N}_0$. The *principal symbol* of L is the map

$$\sigma_L : T^*M \to \mathrm{Hom}_{\mathbb{K}}(F_1, F_2)$$

whose action we will now describe in terms of a local coordinate chart $(\mathcal{U}, x = (x^\mu) : \mathcal{U} \to \mathcal{V})$ for M and local trivializations of F_1 and F_2 adapted to these coordinates. Given such choices, there exists a unique collection of smooth, $(N_2 \times N_1)$-matrix–valued functions A_α defined on $\mathcal{U}$ and such that the following holds: If s is an arbitrary local section of $F_1 \to M$ on $\mathcal{U}$ and we denote by $\tilde{s} : \mathcal{V} \to \mathcal{V} \times \mathbb{R}^{N_1}$ its expression in terms of the chosen coordinates and of the trivialization of F_1, and also denote by $\tilde{t} : \mathcal{V} \to \mathcal{V} \times \mathbb{R}^{N_2}$ the analogous expression of any local section of $F_2 \to M$ on $\mathcal{U}$, then

$$\widetilde{L\tilde{s}} = \sum_{|\alpha| \leq k} (A_\alpha \circ x^{-1}) \cdot \partial^\alpha \tilde{s}.$$

The principal symbol is then defined on any covector $\xi = \xi_\mu \mathrm{d}x^\mu \in T_p^*\mathcal{U}$ to be the element of $\mathrm{Hom}_{\mathbb{K}}(F_1, F_2)$ represented in the chosen trivializations by the matrix

$$\widetilde{\sigma_L}(\xi) = \sum_{|\alpha| = k} \xi^\alpha A_\alpha(p)$$

where $\xi^\alpha := \xi_1^{\alpha_1} \cdots \xi_d^{\alpha_d}$.

Remark. It follows immediately that the principal symbol of L is homogeneous of degree k, i.e. that $\sigma_L(\lambda\xi) = \lambda^k\sigma_L(\xi)$ for any covector ξ and any $\lambda \in \mathbb{K}$.

Definition 2.4.10 (Normally hyperbolic operators)**.** Let (M, g) be a Lorentzian manifold and $F \to M$ be a vector bundle. A second-order linear differential operator $P : \Gamma^\infty(F) \to \Gamma^\infty(F)$ is said to be *normally hyperbolic* if

$$\sigma_P(\xi) = g^{-1}(\xi, \xi)\,\mathrm{id}_F \quad \text{for all } \xi \in T^*M.$$

Although normally hyperbolic operators form a wide class, the only example which we will treat with any detail in this thesis is the most general normally hyperbolic operator acting on smooth real-valued functions, i.e. on sections of the trivial line bundle $F = M \times \mathbb{R}$. This takes the form

$$P = \Box + X + q \tag{2.30}$$

where $\Box$ is the d'Alembert operator defined in Equation (2.23), X is a smooth vector field, and q is a smooth real-valued function. When X vanishes and q is a constant, P is called a *Klein–Gordon* operator.

Lemma 2.4.11. *If (M, g) is a Lorentzian manifold and $P : C^\infty(M) \to C^\infty(M)$ is of the form in Equation (2.30), then P is formally self-adjoint relative to the obvious identification of $F := M \times \mathbb{R}$ with its dual bundle F^* if, and only if, the vector field X vanishes.*

Proof. It is well-known and a straightforward exercise using the divergence theorem that $\Box : C^\infty(M) \to C^\infty(M)$ is self-adjoint in the sense stated. It is obvious that the multiplication operator q too is formally self-adjoint. Since the operation of taking formal adjoints is linear, the 'if' direction is clear. For the 'only if' direction, by a similar reasoning it follows that we must have $X = X^*$. This means that

$$\int_M \psi X\varphi \,\mathrm{d}\mu_g = \int_M (X\psi)\varphi \,\mathrm{d}\mu_g$$

for all $\varphi, \psi \in C_0^\infty(M)$. But $\psi X\varphi = X(\psi\varphi) - (X\psi)\varphi$ by the Leibniz rule, and thus

$$\int_M \psi X\varphi \,\mathrm{d}\mu_g = -\int_M (X\psi)\varphi \,\mathrm{d}\mu_g$$

by integration by parts since the functions have compact support and the manifold is assumed without boundary. Thus, $X\varphi$ vanishes for all $\varphi \in C_0^\infty(M)$ in the sense of distributions and, since it is actually smooth, it must vanish in the sense of functions. Then X must be zero as a differential operator, and thus as a vector field. $\Box$

2.4.3 Characteristic initial value problems

Quite generically, the *characteristic set* of a linear partial differential operator P, denoted by $\operatorname{Char} P$, is the subset of $T^*M\backslash\mathbf{0}$ on which σ_P fails to be an injective linear map [Tre80, p. 80]. In particular, in the scalar case $P : C^\infty(M) \to C^\infty(M)$,

$$\operatorname{Char} P \coloneqq \{\xi \in T^*M \mid \xi \neq 0 \text{ and } \sigma_P(\xi) = 0\}\,.$$

A *characteristic hypersurface* for P is then a hypersurface $\mathscr{N}$ whose conormal bundle is entirely contained in $\operatorname{Char} P$.

Characteristic hypersurfaces are problematic for the following reason: Let P be of order k. Then, in local coordinates adapted to a characteristic hypersurface – i.e. such that the hypersurface is locally defined by the vanishing of one of the coordinate functions – we cannot rewrite the equation $Pu = f$ in such a way as to express the k-th order transverse derivatives of u in terms of the value of u and of its other derivatives up to order k. On the other hand, $Pu = f$, when evaluated on $\mathscr{N}$, reduces to compatibility conditions between $u\restriction_{\mathscr{N}}$, its derivatives along $\mathscr{N}$, and derivatives along $\mathscr{N}$ of the transverse derivatives of u up to order $k-1$. Thus we cannot freely specify the value of u and of its transverse derivatives up to order $k-1$ and expect to obtain a solution in a neighbourhood of a point in $\mathscr{N}$. This is completely unlike the situation in the standard Cauchy problem.

The *characteristic initial value problem*, also called *Goursat problem*, is the problem of finding solutions to a (linear) partial differential equation with prescribed values on a characteristic hypersurface:

$$Pu = f \quad \text{with} \quad u\restriction_{\mathscr{N}} = u_0$$

for some function u_0 defined on $\mathscr{N}$. Existence and/or uniqueness of solutions are not guaranteed, and one can only hope to establish them on a case-by-case basis. Our work in the next section will be an example of one such endeavour.

2.5 The global characteristic initial value problem on achronal hypersurfaces and the definition of S_A and S_B in [KW91]

The purpose of this section is to sketch the proof of existence and uniqueness results pertaining characteristic initial value problems for linear wave-like equations on globally hyperbolic Lorentzian manifolds, posed on null hypersurfaces with favourable causal properties. While the precise statements of these results are, to the best of our knowledge, novel, in essence they and their proofs are adaptations of ideas first presented by

A. D. Rendall in the much more general context of *quasilinear* wave equations in an influential 1990 paper [Ren90]. The nonlinear nature of such equations meant that Rendall's results were local in nature. Our main contribution will be to provide a geometric framework in which Rendall's arguments can be globalised in the linear case. A more superficial difference with Rendall's results is given by the fact that the characteristic initial value surfaces in [Ren90] consist of two intersecting smooth null hypersurfaces; morally speaking, the initial value surfaces which, in our work, play the role of Rendall's intersecting hypersurfaces are achronal boundaries of the form $\dot{I}^+(S)$ and $\dot{I}^-(S)$, for S a subset of a null hypersurface with the causal properties discussed in Subsections 2.5.3 and 2.5.4.

2.5.1 Whitney's Extension Theorem

At the heart of Rendall's argument is an application of a version of the classical *Whitney extension theorem* of differential topology [Whi34, Gla58, Mal66, Ste70, Tou72], see also [Hör90a, Thm. 2.3.6]. Given an open set $\Omega \subseteq \mathbb{R}^n$, the latter deals with the problem of extending, to a function in $C^m(\Omega)$ (with m finite or even infinite), a function originally defined, together with its 'derivative data', only on a closed subset X of Ω. The collection given by the function and its derivative data on X is required to satisfy certain compatibility conditions which are described completely in terms of X and arise naturally as necessary conditions that any suitable extension would have to satisfy by virtue of Taylor's formula. Then, the classical Whitney extension theorem states that smoothness on $X \subseteq \Omega$ in the sense of Whitney is equivalent to the existence of an extension to Ω that is C^m in the usual sense. To wit, call *Whitney data* on X a collection

$$\{f^\alpha : X \to \mathbb{R} \mid |\alpha| \leq m\} \tag{2.31}$$

of continuous functions f^α, where α denotes a multi-index and possibly $m = \infty$. If m is finite then this data is said to be C^m*-smooth in the sense of Whitney* – or to define a C^m*-Whitney field* – if, for each multi-index α with $|\alpha| \leq m$,

$$f^\alpha(x) - \sum_{|\beta| \leq m - |\alpha|} \frac{f^{\alpha+\beta}(y)}{\beta!}(x-y)^\beta \quad \text{is} \quad o\left(\|x-y\|^{m-|\alpha|}\right) \tag{2.32}$$

uniformly on compact subsets of X as $\|x - y\| \to 0$. If $m = \infty$ then the data is said to be C^∞*-smooth in the sense of Whitney* – or to define a C^∞*-Whitney field* – if, for any integer k, the subcollection $\{f^\alpha\}_{|\alpha| \leq k}$ is C^k-smooth in the sense of Whitney. It is clear that, for any $m \in \mathbb{N}_0 \cup \{\infty\}$ and any closed subset X, the Whitney data collection on X all of whose members equal the zero function on X defines a C^m-Whitney field on X.

Theorem 2.5.1 (Classical Whitney Extension Theorem)**.** *Let $X \subseteq \Omega \subseteq \mathbb{R}^n$ be as above, and $\{f^\alpha\}_{|\alpha|\leq m}$ be Whitney data as in (2.31), with possibly $m = \infty$. There exists $f \in C^m(\Omega)$ such that $D^\alpha f\restriction_X = f^\alpha \ \forall\ |\alpha| \leq m$ if and only if the data is C^m-smooth in the sense of Whitney.* □

Jets and the global Whitney extension theorem

This deep result generalises easily to the case where the function is valued in a Euclidean space $\mathbb{R}^d$. It also generalises to the case where $\mathbb{R}^n$ is replaced by a smooth manifold M, $\Omega = M$ and $X \subseteq M$ is a closed subset; although this came as no surprise to experts in differential topology, to the best of our knowledge such a generalisation was first explicitly stated and proved in the published literature rather recently, namely in Appendix A to the paper [BCFT07]. The idea of the proof of this generalisation, which we will henceforth refer to as the *global Whitney extension theorem* and whose statement we will reproduce below, is a standard game in differential topology: suppose that, using the classical Whitney extension theorem on coordinate charts or otherwise, one is able to obtain local Whitney extensions on sufficiently small neighbourhoods in M; a partition of unity argument then yields a global extension by gluing together these local extensions. We will specialise our statement to the case of real-valued functions for simplicity.

The *jet bundle* language provides arguably the most concise and elegant formulation of the global Whitney extension theorem. The details of the jet bundle formalism are not essential here, and can be found in several references ([Mic80, Sau89, KMS93] to mention only a few). Suffice to say that, given two smooth manifolds M and N and $m \in \mathbb{N}_0$, we can define the set $J^m(M;N)$ of m-jets of maps from M to N as follows: an element of $J^m(M;N)$ is an equivalence classes of pairs (f,x), where $x \in M$ and f is a smooth map from an open set in M containing x to N, and two such pairs (f,x) and (f',x') are considered equivalent iff $x = x'$, $f(x) = f'(x')$, and f and f' have equal partial derivatives up to order m in one (and hence any other) coordinate chart for M and N around x and $f(x)$ respectively. That is, the local expressions for f and f' have equal Taylor developments up to order m. $J^m(M;N)$ can be given the structure of a finite-dimensional smooth manifold and even of a fiber bundle over $M \times N$, and one can define the *infinite jet space* $J^\infty(M;N)$ as a certain projective limit built from the $J^m(M;N)$ [Mic80, Ch. 4]. As a set, $J^\infty(M;N)$ may also be defined as a set of equivalence classes of (local) maps from M to N just as above, with m replaced by ∞. What's more, $J^\infty(M;N)$ can be given the structure of an infinite-dimensional manifold modelled on a locally convex topological vector space [Mic80, p. 88]. For any $m \in \mathbb{N}_0 \cup \{\infty\}$, any smooth map $f : M \to N$ gives rise to a smooth map

$$j^m f : M \to J^m(M;N), \quad \text{where } j^m f(x) \coloneqq \text{the equivalence class of } (f,x) \text{ in } J^m(M;N)$$

called the (m-jet) *prolongation* or *extension* of f. Now, specialising to the case $N = \mathbb{R}^N$, of course a smooth map $M \to \mathbb{R}^N$ may be viewed as a smooth section of the trivial vector bundle $(M \times \mathbb{R}^N, \mathrm{pr}_1, M, \mathbb{R}^N)$. More generally, given a vector bundle $\boldsymbol{\pi} \coloneqq (F, \pi, M, V)$ and $m \in \mathbb{N}_0$, we can define $J^m(\boldsymbol{\pi})$ to be the set of equivalence classes of smooth sections of $\boldsymbol{\pi}$ in just the same way as above. Therefore, under our notation, in general $J^m(\boldsymbol{\pi}) \subseteq J^m(M;F)$. $J^m(\boldsymbol{\pi})$ too is a finite-dimensional smooth manifold, and also the total space of a vector bundle over M, called the *m-jet bundle of* $\boldsymbol{\pi}$. Again by a projective limit procedure, we may define the *infinite jet bundle of* $\boldsymbol{\pi}$, $J^\infty(\boldsymbol{\pi})$, and equip it with the structure of an infinite-dimensional Fréchet manifold [Sau89, Ch. 7]. Prolongation of a smooth section ϕ of $\boldsymbol{\pi}$ gives rise to a smooth section $j^m\phi$ of $J^m(\boldsymbol{\pi}) \to M$, for any $m \in \mathbb{N}_0 \cup \{\infty\}$. We note that, if $\boldsymbol{\pi}$ is the trivial line bundle $(M \times \mathbb{R}, \mathrm{pr}_1, M, \mathbb{R})$ then there is a canonical vector bundle isomorphism (covering the identity map on M) between $J^1(\boldsymbol{\pi})$ and $(T^*M \times \mathbb{R}, \pi_{T^*M} \circ \mathrm{pr}_1, M, \mathbb{R}^{\dim M + 1})$.

Theorem 2.5.2 (Global Whitney Extension Theorem, Thm. A.1 in [BCFT07])**.** *Let M be a smooth manifold, $X \subseteq M$ be a closed set, and $m \in \mathbb{N}_0 \cup \{\infty\}$. Let $f : X \to J^m(M;\mathbb{R})$ be a section of $J^m(M;\mathbb{R}) \xrightarrow{\pi_m} M$ along X, i.e. $\pi_m \circ f = \mathrm{id}_X$. Assume that f has local C^m extensions, in the sense that for each $x \in X$ there exist:*

(1) open neighbourhoods V_x and U_x of x in M such that $V_x \subseteq \overline{V_x} \subseteq U_x$;

(2) a compact subset $K_x \subseteq X$ such that $X \cap V_x = K_x \cap V_x$ and $K_x \subseteq U_x$;

(3) a $F_x \in C^m(U_x;\mathbb{R})$ such that $[j^m F_x](y) = f(y)$ for each $y \in K_x$.

Then, there exists a $F \in C^m(M;\mathbb{R})$ such that $[j^m F](x) = f(x)$ for all $x \in X$. That is, f admits a global C^m extension. □

As pointed out in [BCFT07], the result actually extends in a straightforward way to mappings $M \to \mathbb{R}^N$ and thereby directly even to sections of arbitrary vector bundles over M.

Borel's Lemma

An important corollary of the Whitney extension theorem is a generalised version ([GG73, Lem. 2.5], [Hör90a, Thm. 1.2.6 & Cor. 1.3.4]), which we will now enunciate, of the classical *Borel Lemma* [Bor95].

Corollary 2.5.3 (n-dimensional Borel's Lemma)**.** *Let $\Lambda \subseteq \mathbb{R}^n$ be open and let $(f_0, f_1, \ldots)$ be a sequence of functions in $C^\infty(\Omega)$. Then, for any open interval $I \subseteq \mathbb{R}$ such that $0 \in I$, there exists an $F \in C^\infty(I \times \Lambda)$ such that $\dfrac{\partial^k F}{\partial t^k}(0, \cdot) = f_k$ for all $k \geq 0$.*

Proof. While this result can be easily proved without recourse to Whitney's extension theorem, it is of interest to see that this is what the latter reduces to in the case of the (closed) hyperplane $X := \{0\} \times \Lambda$ in the open subset $\Omega := I \times \Lambda$ of $\mathbb{R}^{n+1}$. We must prescribe complete Whitney data on X which extends the sequence $(f_k)_{k\geq 0}$ of already given 'transverse derivatives', and verify that just knowing that each f_k is smooth suffices to establish the estimates given by (2.32). To wit: Since we are looking for a smooth extension having f_k as the k-th order transverse derivative, for any multi-index ρ in $\mathbb{N}_0^n$ we are forced to prescribe the Whitney data functions as

$$f^{(k,\rho)}(0, \cdot) = \partial^\rho f_k \quad \forall\ k \geq 0. \tag{2.33}$$

Now consider two arbitrary points $x := (0, z)$ and $y := (0, w)$ belonging to X. If $\beta = (i, \gamma) \in \mathbb{N}_0^{n+1}$ is a multi-index then

$$(x - y)^\beta = (0 - 0)^i \cdot (z - w)^\gamma.$$

Therefore, this quantity vanishes unless $i = 0$. The difference in (2.32) then becomes

$$f^\alpha(x) - \sum_{|\gamma| \leq m - |\alpha|} \frac{f^{\alpha+(0,\gamma)}(0, w)}{\gamma!}(z - w)^\gamma.$$

If $\alpha = (k, \rho)$ with $k \in \mathbb{N}_0$ and $\rho \in \mathbb{N}_0^n$ then, using Equation (2.33), this reads

$$\begin{aligned}\partial^\rho f_k(z) - &\sum_{|\gamma| \leq m-k-|\rho|} \frac{f^{(k,\rho+\gamma)}(0, w)}{\gamma!}(z - w)^\gamma \\ &= \partial^\rho f_k(z) - \sum_{|\gamma| \leq m-k-|\rho|} \frac{\partial^\gamma[\partial^\rho f_k](w)}{\gamma!}(z - w)^\gamma.\end{aligned}$$

Since $\partial^\rho f_k$ is smooth on Λ, by Taylor's theorem this quantity is $o(||z - w||^{m-k-|\rho|})$ uniformly on compact subsets of Λ as $||z - w||^{m-k-|\rho|} \to 0$. The result follows. □

Remark. As is usual in differential topology, the above result is the essential local ingredient (via a partition of unity argument) in proving a yet more general version of Borel's Lemma. Namely, we can replace Ω with a smooth manifold of dimension $n + 1$, $\Lambda \subset \Omega$ with a closed codimension-1 embedded submanifold, and $\partial/\partial t$ with a smooth vector field on a neighbourhood of Λ which, on Λ, is everywhere (non-zero and) transverse to Λ. There exists then a $F \in C^\infty(\Omega)$ such that $T^k(F)\restriction_\Lambda$ is equal, for any $k \geq 0$, to a prescribed $f_k \in C^\infty(\Lambda)$.

'Gluing' Whitney fields and regularly separated sets

A natural problem often arises when trying to make use of Whitney's extension theorem in applications, and it will affect our discussion in the following subsections.[16] The problem – which has counterparts in the case of finite differentiability but which we will only formulate in the C^∞ case – is the following:

> **'Gluing problem' for C^∞-Whitney fields.** Let X, Y be two closed subsets of $\mathbb{R}^n$. Let F, G be arbitrary C^∞-Whitney fields on X and Y respectively, which agree on $X \cap Y$. *Under what conditions on X and Y does there always exist a C^∞-Whitney field on $X \cup Y$ whose restrictions to X and Y coincide with F and G respectively?*

This problem has a definite resolution, given as Theorem 5.5 in [Mal66] and recalled below as Theorem 2.5.5. The latter relies on a fundamental definition first given in a 1959 paper by Łojasiewicz [Łoj59]. The exact variant which we present below as Definition 2.5.4 is as in the paper [RSZ12, Prop. 4.3.6]. Recall that for a point $p \in \mathbb{R}^n$ and any $A \subseteq \mathbb{R}^n$, $d(p, A)$ denotes the standard metric space distance between p and A.

Definition 2.5.4 (After [Łoj59, pp. 91–92])**.** Two subsets X, Y of $\mathbb{R}^n$ are said to be *regularly separated* (or 'regularly situated' in [Mal66]) if either $X \cap Y = \emptyset$ or one of the following equivalent conditions is fulfilled:

i) for all $x_0 \in X \cap Y$, there exists a neighbourhood U of x_0 and constants $C > 0$, $\lambda > 0$, such that

$$d(x, X) + d(x, Y) \geq C d(x, X \cap Y)^\lambda \text{ for all } x \in U;$$

ii) for all $x_0 \in X \cap Y$, there exists a neighbourhood U of x_0 and constant, $C' > 0$, $\lambda > 0$, such that

$$d(x, Y) \geq C' d(x, X \cap Y)^\lambda \text{ for all } x \in X \cap U;$$

iii) condition ii) holds when the roles of X and Y are interchanged.

If, in either case, λ can be chosen to be equal to 1, then we say that X and Y are *simply separated.*

The resolution of the 'gluing problem' above can now be stated.

Theorem 2.5.5. *Let X, Y be two closed subsets of $\mathbb{R}^n$. Then the following are equivalent:*

- *the gluing problem for C^∞-Whitney fields can be answered in the affirmative for X and Y;*
- *X and Y are regularly separated.*

[16]That this problem requires attention when trying to solve the characteristic initial value problem for hyperbolic equations in Rendall's approach was already implicitly recognised in [Ren90].

In Subsection 2.5.4, we will require a more practical criterion to establish the regular separateness of two sets. We recall below one which will suit our purposes and which can be found in the existing literature, see e.g. Lemma 6 and its Corollary in [Paw02] (but the germ of the idea can already be found in [Łoj59, Lem. 1 & Lem. 2]). In the statement, if $f : X \to Y$ then $\Gamma(f) \subseteq X \times Y$ denotes the graph of f.

Lemma 2.5.6. *Let $\Omega \subseteq \mathbb{R}^k$ be non-empty and open, and $\theta : \Omega \to \mathbb{R}^m$ be a locally Lipschitz map. Then $\overline{\Gamma(\theta)}$ and any subset S of $\mathbb{R}^{k+m} \setminus (\Omega \times \mathbb{R}^m)$ are simply separated.*

Sketch of proof. We first prove the claim for $S = \mathbb{R}^{k+m} \setminus (\Omega \times \mathbb{R}^m)$. Without loss of generality, we may assume that θ is globally Lipschitz with Lipschitz constant L. Denote the continuous extension of θ to $\overline{\Omega}$ (which is Lipschitz with the same Lipschitz constant) by $\overline{\theta}$. Given $a \in \overline{\Omega}$, let $b \in \dot{\Omega}$ be such that $\|a - b\|_{\mathbb{R}^k} = d_{\mathbb{R}^k}(a, \dot{\Omega})$. Then

$$d_{\mathbb{R}^{k+m}}((a, \overline{\theta}(a)), S) = d_{\mathbb{R}^k}(a, \dot{\Omega}) = \|a - b\|_{\mathbb{R}^k}.$$

Now use

$$\left\|(a, \overline{\theta}(a)) - (b, \overline{\theta}(b))\right\|^2_{\mathbb{R}^{k+m}} = \|a - b\|^2_{\mathbb{R}^k} + \left\|\overline{\theta}(a) - \overline{\theta}(b)\right\|^2_{\mathbb{R}^m} \leq \left(1 + L^2\right) \|a - b\|^2_{\mathbb{R}^k}$$

to obtain the desired inequality.

$$\begin{aligned} d_{\mathbb{R}^{k+m}}((a, \overline{\theta}(a)), S) &\geq \frac{1}{\sqrt{L^2 + 1}} \left\|(a, \overline{\theta}(a)) - (b, \overline{\theta}(b))\right\|_{\mathbb{R}^{k+m}} \\ &\geq \frac{1}{\sqrt{L^2 + 1}} d_{\mathbb{R}^{k+m}}((a, \overline{\theta}(a)), \overline{\Gamma(\theta)} \cap S). \end{aligned}$$

The result then extends to any other subset (closed or not) of $\mathbb{R}^{k+m} \setminus (\Omega \times \mathbb{R}^m)$ by an application of Proposition 2 in [Paw02]. □

2.5.2 An illustrative example of Rendall's method

Let $M = \mathbb{R}^2$ and $q' \in \mathbb{R}$, and consider the homogeneous Klein–Gordon equation

$$\Box\phi(t, x) = \left(\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2}\right) \phi(t, x) = q'\phi(t, x).^{17}$$

In double null coordinates $u = t - x$, $v = t + x$, and denoting the expression for ϕ in these coordinates by ψ, this equation reads as

$$\frac{\partial^2 \psi}{\partial u \, \partial v}(u, v) = \frac{\partial}{\partial u}\left(\frac{\partial \psi}{\partial v}\right)(u, v) = q\psi(u, v) \tag{2.34}$$

[17] Normally, q' here would be $-m^2$ for some mass $m \geq 0$. Avoiding minus signs in the right-hand side of this equation will simplify our calculations.

with $q = q'/4$. The (u, v) coordinate system is in fact a very simple example of a Gaussian null coordinate system for the null hypersurface (in this case, the null line) $\mathscr{N} = \{(t, x) \in M \mid v(t, x) = t + x = 0\}$. Assume it be known that ψ is a solution of Equation (2.34) with given data on $\widetilde{\mathscr{N}} := \big\{(u, v) \in \mathbb{R}^2 \bigm| v = 0\big\}$ of compact support, i.e.

$$\psi(u, 0) = f(u), \quad f \in C_0^\infty(\mathbb{R}).^{18} \tag{2.35}$$

Then it follows directly from Equation (2.34), and by taking its v-derivatives, that

$$\frac{\partial}{\partial u}\left(\frac{\partial \psi}{\partial v}\right)(u, 0) = q\psi(u, 0) = qf(u) \quad \text{and} \quad \frac{\partial}{\partial u}\left(\frac{\partial^n \psi}{\partial v^n}\right)(u, 0) = q\frac{\partial^{n-1}\psi}{\partial v^{n-1}}(u, 0).$$

Defining $\psi_v^{(n)}(u) = \dfrac{\partial^n \psi}{\partial v^n}(u, 0)$ for all $n \geq 0$, these equations together with Equation (2.35) are equivalent to the iterative system of ODEs

$$\begin{cases} \psi_v^{(0)} = f \\ \frac{\mathrm{d}\psi_v^{(n)}}{\mathrm{d}u} = q\psi_v^{(n-1)} \quad \text{for } n \geq 1. \end{cases} \tag{2.36}$$

Of course, initial conditions are required to uniquely solve this system. We impose the following: let $u_{\min} = \inf \operatorname{supp} f$; then we supplement system (2.36) with the initial conditions $\psi_v^{(n)}(u_{\min}) = 0 \;\forall\; n \geq 0$, which are actually easily seen to be equivalent to $\psi_v^{(n)}(u) = 0 \;\forall\; u \leq u_{\min} \;\forall\; n \geq 0$. Geometrically, thinking of $M = \mathbb{R}^2$ as being equipped with the standard Minkowski metric $\mathrm{d}t^2 - \mathrm{d}x^2$, this is saying that the original function ϕ and all of its v-derivatives, restricted to $\mathscr{N}$, have support to the causal future of $\operatorname{supp} f$ (the latter being identified here with the obvious subset of $\mathscr{N}$). Then, the unique solution to system (2.36) is easily seen to be given by the sequence of smooth functions

$$\psi_v^{(n)} = q^n \mathcal{I}^n f \qquad \text{where} \qquad [\mathcal{I}g](u) := \int_{-\infty}^{u} g(u')\,\mathrm{d}u'. \tag{2.37}$$

Conversely, let $\psi_{\text{app}} \in C^\infty(\mathbb{R}^2)$ be such that

$$\frac{\partial^n \psi_{\text{app}}}{\partial v^n}(u, 0) = \psi_v^{(n)}(u) = q^n\,[\mathcal{I}^n f](u) \;\forall\; u \in \mathbb{R}.$$

Then ψ_{app} solves Equation (2.34), and all its differential consequences, on $\widetilde{\mathscr{N}}$ (i.e. it is an 'approximate' solution). That such a function exists is a result of the generalised Borel's Lemma, Corollary 2.5.3. Therefore, the Whitney data collection $\{\psi^\alpha\}_{\alpha \in \mathbb{N}_0^2}$, given by

$$\psi^{(i,j)}(u, 0) := \frac{\mathrm{d}^i \psi_v^{(j)}}{\mathrm{d}u^i}(u) = q^j \frac{\mathrm{d}^i}{\mathrm{d}u^i}\left[\mathcal{I}^j f\right](u) \quad \forall\; i, j \in \mathbb{N}_0, \tag{2.38}$$

[18] In what follows, we will occasionally abuse notation and freely identify f with the corresponding smooth function defined on the submanifold $\mathscr{N}$ of M.

defines a C^∞-Whitney field on any subset of $\widetilde{\mathscr{N}}$ which is closed in $\mathbb{R}^2$. In particular, the restriction $\Psi_1 := \{\psi^\alpha\restriction_{\widetilde{\mathscr{N}_1}}\}$ of this field to the closed set $\widetilde{\mathscr{N}_1} := [u_{\min}, +\infty) \times \{0\}$ is a C^∞-Whitney field on $\widetilde{\mathscr{N}_1}$ which vanishes at the point $(u_{\min}, 0)$.

Now, consider a second closed subset of $\mathbb{R}^2$, specifically $\widetilde{\mathscr{N}_2} := \{u_{\min}\} \times [0, +\infty)$. This choice is geometrically motivated since then, upon changing back to (t, x) coordinates, $\widetilde{\mathscr{N}_1} \cup \widetilde{\mathscr{N}_2}$ is mapped to the topological boundary of the causal future of the point $(t, x) \in M$ with $(u(t, x), v(t, x)) = (u_{\min}, 0)$ – that is, $p = (t, x) = (u_{\min}/2, -u_{\min}/2)$. The key observation at this point is that $\widetilde{\mathscr{N}_1}$ and $\widetilde{\mathscr{N}_2}$ are regularly separated – in fact, simply separated – in $\mathbb{R}^2$, see Definition 2.5.4. Indeed, their intersection $\widetilde{\mathscr{N}_1} \cap \widetilde{\mathscr{N}_2} = \{(u_{\min}, 0)\}$, and the claimed property follows here simply from the fact that the sets intersect orthogonally.[19] It follows from Theorem 2.5.5, and by the observation directly above the statement of Theorem 2.5.1, that the Whitney data $\{\Psi^\alpha\}_{\alpha\in\mathbb{N}_0^2}$ on $\widetilde{\mathscr{N}_1} \cup \widetilde{\mathscr{N}_2}$, defined by

$$\Psi^\alpha = \begin{cases} \Psi_1^\alpha = \psi^\alpha\restriction_{\widetilde{\mathscr{N}_1}} & \text{on } \widetilde{\mathscr{N}_1} \\ 0 & \text{on } \widetilde{\mathscr{N}_2}, \end{cases}$$

defines a C^∞-Whitney field on $\widetilde{\mathscr{N}_1} \cup \widetilde{\mathscr{N}_2}$.

2.5.2.1 Reduction to the Cauchy problem in this example

Returning to our original motivation coming from wanting to solve the Klein–Gordon equation [‘in (t, x)-coordinates’] with data prescribed on the null hypersurface $\mathscr{N}$, it follows from the Whitney extension theorem that there exists a $\phi_{\mathrm{app}} \in C^\infty(\mathbb{R}^2)$ which, by construction:

(a) solves the Klein–Gordon equation and all its differential consequences on $\dot{J}^+(p)$, with $p = (u_{\min}/2, -u_{\min}/2)$ as above;

(b) attains the zeroth-order values on $\mathscr{N}$ given by $f \in C_0^\infty(\mathscr{N})$;

(c) vanishes to infinite order on $\dot{J}^+(p) \setminus \mathscr{N}$.

Of course, ϕ_{app} may not yet solve the Klein–Gordon equation on *any* open set in M. We must therefore remove from it the ‘error’ preventing it from achieving this. This seems like a delicate procedure as we do not want to spoil the fact that ϕ_{app} *does*, at least, solve the initial conditions on $\mathscr{N}$. Thinking heuristically, assume such an error exists

[19] But notice that, by the cosine rule, the property would still hold if, say, $\widetilde{\mathscr{N}_2}$ were to be rotated around $\widetilde{\mathscr{N}_1} \cap \widetilde{\mathscr{N}_2}$ so as to intersect $\widetilde{\mathscr{N}_1}$ at any other angle between 0 and π.

and is smooth, and denote it by ϕ_{err}. Then it must satisfy at the very least

$$\begin{aligned} \phi_{\text{err}} &= 0 && \text{on } \mathscr{N} \\ P(\phi_{\text{app}} - \phi_{\text{err}}) = 0 \iff P\phi_{\text{err}} &= P\phi_{\text{app}} && \text{on some open set in } M, \end{aligned}$$

where $P = \Box - q'$ is the Klein–Gordon differential operator. Now, observe that it follows from (a) that $P\phi_{\text{app}}$ vanishes to infinite order on $\dot{J}^+(p)$. Therefore, letting

$$\chi_p^+ \coloneqq \chi_{J^+(p)} = \begin{cases} 1 & \text{on } J^+(p) \\ 0 & \text{on } M \setminus J^+(p), \end{cases}$$

the function

$$f^+ \coloneqq \chi_p^+ \cdot P\phi_{\text{app}}$$

is smooth on $\mathbb{R}^2$ and supported in $J^+(p)$. But we *can* solve

$$\phi_{\text{err}} = 0 \qquad \text{on } \mathscr{N} \tag{2.39}$$

$$P\phi_{\text{err}} = f^+ \qquad \text{on } M. \tag{2.40}$$

Indeed, f^+ is smooth and with retarded support; thus, as explained in Section 2.4, the unique smooth solution, with retarded support, to the second equation alone is obtained by acting on f^+ with the retarded Green operator E^+:

$$\phi_{\text{err}} = E^+ f^+ \qquad \text{solves Equation (2.40).}$$

It then follows at once that Equation (2.39) is automatically satisfied. We can now define

$$\phi^+ \coloneqq \phi_{\text{app}} - \phi_{\text{err}} = \phi_{\text{app}} - E^+ f^+ = \phi_{\text{app}} - E^+[\chi_p^+ \cdot P\phi_{\text{app}}],$$

and the construction is such that:

(i) $\phi^+ \in C^\infty(M)$;

(ii) $P\phi^+ = 0$ on $J^+(p)$ [which contains the open set $I^+(p)$];

(iii) $\phi^+\restriction_{\mathscr{N}} = \phi_{\text{app}}\restriction_{\mathscr{N}}$ is equal to the prescribed function $f \in C_0^\infty(\mathscr{N})$.

The procedure described in this section is the core of Rendall's argument in [Ren90], as applied to the linear Klein–Gordon equation on $\mathbb{R}^2$. As already pointed out, our conclusions are global, while Rendall's – being however applicable to vastly more general PDE systems – were local.

A completely analogous construction, which we will not repeat and amounts to a change in time orientation, would yield a $\phi^- \in C^\infty(M)$ also attaining the values prescribed by f on $\mathscr{N}$, but with $P\phi^- = 0$ on $J^-(p)$ instead.

2.5.3 Our general geometric setup

While we have just illustrated all the main ideas of our adaptation of Rendall's method, by working out in full detail the example of the Klein–Gordon equation on $M = \mathbb{R}^2$, we now wish to seek the most general setup to which the same ideas can carry over. The (1+1)-dimensional Minkowski spacetime, and the null hypersurface (which was just a line) considered in the example above, clearly enjoy several favourable local and global geometric properties. As we will illustrate in this and the next subsection, the most important ones which allow the arguments to survive generalisation are the global hyperbolicity of the spacetime and the achronality of $\mathscr{N}$.

Proposition 2.5.7. *Let $(M, g, \mathfrak{t})$ be a time-oriented Lorentzian manifold, $\mathscr{N}$ be an achronal null hypersurface in M, and S be a subset of $\mathscr{N}$. Define*[20]

$S^+ := \{p \in \mathscr{N} \mid \exists\, q \in S \text{ s.t. } p = q \text{ or } p \text{ comes after } q \text{ along the null generator through } q\}$

[resp. define S^- by replacing 'after' with 'before'].

(a) *It holds that $S^+ \subseteq \mathscr{N} \cap \dot{I}^+(S)$ [resp. $S^- \subseteq \mathscr{N} \cap \dot{I}^-(S)$].*

Suppose in addition that the null generators of $\mathscr{N}$, when reparametrised as null geodesics entirely contained in $\mathscr{N}$, are future [resp. past] inextendible as geodesics in $(M, g, \mathfrak{t})$.

(b) *$J^+(\mathscr{N}) \setminus I^+(\mathscr{N}) = \mathscr{N}$ [resp. $J^-(\mathscr{N}) \setminus I^-(\mathscr{N}) = \mathscr{N}$].*

(c) *In either case ('future' or 'past') $J^+(\mathscr{N}) \cap J^-(\mathscr{N}) = \mathscr{N}$ and therefore $J^+(A) \cap J^-(A) \subseteq \mathscr{N}$ for any $A \subseteq \mathscr{N}$.*

Finally, assume also, in addition to the above, that $(M, g, \mathfrak{t})$ is globally hyperbolic.

(d) *If S is future [resp. past] causally complete in M then $S^+ = \mathscr{N} \cap \dot{I}^+(S) = \mathscr{N} \cap \dot{J}^+(S) = \mathscr{N} \cap J^+(S)$ [resp. $S^- = \mathscr{N} \cap \dot{I}^-(S) = \mathscr{N} \cap \dot{J}^-(S) = \mathscr{N} \cap J^-(S)$].*

(e) *With S as in* **(c)***, if $\mathscr{N}$ is also closed then S^+ is closed in M [resp. S^- is closed in M].*

Proof. We illustrate the arguments for **(a)** in the case of S^+ and $\dot{I}^+(S)$, since the statements involving the corresponding objects with $+$ replaced by $-$ then follow simply by a change in time orientation. Similarly, we will prove **(b)**, **(c)**, **(d)** and **(e)** in the case where the assumption on $\mathscr{N}$ holds with the word 'future'.

[20] Recall that, according to our Definition 2.3.3, a null (geodesic) generator of a null hypersurface is always future-directed and maximally extended.

The generic inclusion $S^+ \subseteq \mathscr{N} \cap \dot{I}^+(S)$ in **(a)** follows from the fact that, on the one hand, $S^+ \subseteq J^+(S) \subseteq \overline{I^+}(S)$ by construction and, on the other hand, $S^+ \cap I^+(S) = \emptyset$ because $\mathscr{N}$ is achronal and S, $S^+ \subseteq \mathscr{N}$. Hence $S^+ \subseteq \mathscr{N} \cap [\overline{I^+}(S) \setminus I^+(S)] = \mathscr{N} \cap \dot{I}^+(S)$. Let us now prove **(b)**. Since $\mathscr{N}$ is achronal, the inclusion $\mathscr{N} \subseteq J^+(\mathscr{N}) \setminus I^+(\mathscr{N})$ is obvious and we only need to show that $J^+(\mathscr{N}) \setminus I^+(\mathscr{N}) \subseteq \mathscr{N}$. So let $p, x \in M$ be such that $p \in \mathscr{N}$ and: (i) there exists a future-directed causal curve γ_1 from p to x; (ii) there do *not* exist future-directed timelike curves from any point in $\mathscr{N}$ – in particular, from p – to x. Then we want to argue that $x \in \mathscr{N}$. By Corollary 2.1.21, γ_1 can be reparametrised to be a smooth null (and achronal) geodesic, so without loss of generality we assume that it already had that parametrisation. Denote by $\nu : I \to \mathscr{N}$ a reparametrisation, as a null geodesic with $\nu(0) = p$, of the generator of $\mathscr{N}$ through p. Concatenating $\nu\restriction_{I\cap(-\infty,0]}$ with γ_1, we obtain a causal curve γ connecting points in $\mathscr{N}$ to x. By construction, there seems to be the possibility of γ being a broken null geodesic. However, if γ were not actually everywhere smooth, again by Corollary 2.1.21 there would exist a timelike curve from any of the points in $\nu(I \cap (-\infty, 0)) \subseteq \mathscr{N}$ to p, contradicting our assumption that $x \notin I^+(\mathscr{N})$. The initial portion of γ is, by construction, a portion of the smooth null geodesic ν; hence, by geodesic uniqueness this must also be the case for the entirety of $\gamma\restriction_{[0,+\infty)}$, unless the latter is a proper extension of $\nu\restriction_{I\cap[0,+\infty)}$. But since $\nu\restriction_{I\cap[0,+\infty)}$ is future inextendible by hypothesis, this cannot be the case. In particular, x belongs to the image of a generator of $\mathscr{N}$ and is therefore in $\mathscr{N}$ as we wanted to show. **(c)** then follows easily by the achronality of $\mathscr{N}$ and again by the general fact that if $p \leq x \ll q$ then $p \ll q$:

$$J^+(\mathscr{N}) \cap J^-(\mathscr{N}) = \underbrace{\{I^+(\mathscr{N}) \cap J^-(\mathscr{N})\}}_{=\emptyset} \cup \{[J^+(\mathscr{N}) \setminus I^+(\mathscr{N})] \cap J^-(\mathscr{N})\} = \mathscr{N}.$$

Let us now turn to **(d)**. One inclusion was already proved in **(a)**, so one only needs to prove that $\mathscr{N} \cap \dot{I}^+(S) \subseteq S^+$ under the additional assumption of global hyperbolicity of $(M, g, \mathfrak{t})$ and of future causal completeness of S. By Lemma 2.1.38, in this case $J^+(S)$ is equal to the closure of $I^+(S)$, whence $\dot{I}^+(S) = \overline{I^+}(S) \setminus I^+(S) = J^+(S) \setminus I^+(S)$. It follows that the points in $\mathscr{N} \cap \dot{I}^+(S)$ are precisely those points on $\mathscr{N}$ which cannot be reached from S by following a future-directed timelike curve, but can be reached from S by following a future-directed causal curve. That any such causal curve must be a null geodesic entirely contained in $\mathscr{N}$, and thus that such points must belong to S^+, then follows from an argument analogous to the one used in the above proof of part **(b)**; we omit the details. That $\mathscr{N} \cap \dot{I}^+(S) = \mathscr{N} \cap \dot{J}^+(S)$ is a result of the fact that, for any subset V of an arbitrary time-oriented Lorentzian manifold, the interior of $J^\pm(V)$ equals $I^\pm(V)$. Finally, $\mathscr{N} \cap \dot{J}^+(S) = \mathscr{N} \cap J^+(S)$ in this case follows from the closedness of $J^+(S)$ together with the achronality of $\mathscr{N}$.

(e) then follows immediately since $\dot{I}^+(S)$ is closed. □

2.5.4 Main result on the global characteristic initial value problem; solutions 'falling entirely through' a null hypersurface; Kay and Wald's S_A and S_B

The material in this subsection is the culmination of much of the preparatory work done in this chapter. We will draw on the discussion and general results from this and the previous two sections in order to formulate general (and novel in the precise form we will give) existence and uniqueness results concerning hyperbolic PDEs on manifolds with initial values posed on suitable characteristic hypersurfaces. Finally, we will specialise these results to the setting of spacetimes with bifurcate Killing horizons described in Section 2.2. This will allow us to justify some claims, made without proof in [KW91] and in the literature referencing that paper, whose validity is essential to the analysis carried out there.

Let $(M, g, \mathfrak{t})$ be a globally hyperbolic Lorentzian manifold, and $\mathscr{N}$ be a null hypersurface whose null generators, when reparametrised as null geodesics entirely contained in $\mathscr{N}$, are future and past inextendible in (M, g). We proved in Proposition 2.3.9 that in this case $\mathscr{N}$ admits a cross-section $\mathscr{S}$. In turn, by Theorem 2.3.11 we can find a Gaussian null coordinate chart $(\mathcal{U}, (v, u, x^A))$ for $\mathscr{N}$ around the image of any of its null geodesic generators. Let P be a normally hyperbolic, scalar differential operator of the form in Equation (2.30). In Corollary 2.3.14, we obtained the form, in the chosen Gaussian null coordinate system, of the sequence of restrictions to $\mathscr{N} \cap \mathcal{U}$ of

$$P\phi = F \quad \text{for} \quad F \in C^\infty(M),$$

and of all its differential consequences obtained by taking derivatives of all orders with respect to the coordinate vector transverse to $\mathscr{N}$ – namely, $\partial/\partial v$ in the notation used there and also adopted in what follows.[21] Generalising what we did in the low-dimensional example covered in Subsection 2.5.2, we now notice that this sequence is actually equivalent to a sequence of *linear, first-order, inhomogeneous, ordinary* differential equations *smoothly depending on parameters*, with coefficients and inhomogeneities determined, for each $n \geq 1$, by X, q, F, $\phi_v^{(n-1)}$ and by smooth functions coming from the geometry of (M, g). More precisely:

[21] Albeit not written out explicitly, these are referred to as 'propagation equations' in [Ren90].

- for any given $n \geq 1$ and each fixed value of the tuple of x^A-coordinates in the chosen Gaussian null coordinate system, we have a linear, first-order ODE in the variable u, for the function $u \mapsto \phi_v^{(n)}(u, x^A)$;[22]
- the coefficients and inhomogeneities of the resulting families of ODEs, regarded as being parametrised by (x^A), depend smoothly on (x^A).

We would like to solve this infinite tower of ODEs depending on parameters recursively, augmenting it with and starting from the $n = 0$ case

$$\phi_v^{(0)} = f\restriction_{\mathcal{U}\cap\mathscr{N}},$$

for a given $f \in C^\infty(\mathscr{N})$. Again as in Subsection 2.5.2, a corresponding sequence of (x^A-parametrised) initial conditions is also needed to single out unique sequences of (x^A-parametrised) solutions. We will restrict attention to two types of initial conditions, which arise naturally and are related by a change in time orientation of the ambient Lorentzian manifold. To motivate their definitions, we first observe that, if $\operatorname{supp} f$ and $\operatorname{supp} F$ are both causally complete subsets of M as we assume from now on, then we can pick smooth spacelike Cauchy surfaces $\mathscr{C}^\mp$ such that $\operatorname{supp} f \cup \operatorname{supp} F \subset I^\pm(\mathscr{C}^\mp)$, in which case $J^\pm(\operatorname{supp} f \cup \operatorname{supp} F) \subset I^\pm(\mathscr{C}^\mp)$ too. By construction and by the simple recursive structure of our tower of equations, the following is then true of any sequence $\big(\phi_v^{(n)}\big)_{n=0}^\infty$ of x^A-parametrised solutions when seen as functions on $\mathscr{N} \cap \mathcal{U}$, if we use the notation introduced in Proposition 2.5.7 but denote $(\operatorname{supp} F \cap \mathscr{N})^\pm$ more compactly by $(\operatorname{supp} F)^\pm$:[23]

- $\phi_v^{(0)} = 0$ on $(\mathscr{N} \cap \mathcal{U}) \setminus \operatorname{supp} f \supseteq (\mathscr{N} \cap \mathcal{U}) \setminus [(\operatorname{supp} f)^\pm \cup (\operatorname{supp} F)^\pm]$;
- for any $n \geq 1$, if $\phi_v^{(n-1)} = 0$ on $(\mathscr{N} \cap \mathcal{U}) \setminus [(\operatorname{supp} f)^\pm \cup (\operatorname{supp} F)^\pm]$ then the same is true of $\phi_v^{(n)}$ if it is also the case that $\phi_v^{(n)} = 0$ on $\mathscr{C}^\mp \cap (\mathscr{N} \cap \mathcal{U})$.

We say that the sequence of x^A-parametrised initial conditions, given by $\phi_v^{(n)} = 0$ on $\mathscr{C}^- \cap (\mathscr{N} \cap \mathcal{U})$ for all $n \geq 1$, is *of type* $+$; similarly, the sequence of x^A-parametrised initial conditions, given by $\phi_v^{(n)} = 0$ on $\mathscr{C}^+ \cap (\mathscr{N} \cap \mathcal{U})$ for all $n \geq 1$, will be referred to as being *of type* $-$. It is clear that any two distinct Cauchy surfaces $\mathscr{C}_1^-$, $\mathscr{C}_2^-$ with the properties above yield the same sequence of x^A-parametrised solutions if initial conditions of type $+$ are imposed in both cases; similarly for two distinct $\mathscr{C}_1^+$, $\mathscr{C}_2^+$ and initial conditions of type $-$. Therefore, the notion just introduced is actually independent of the choice of Cauchy surface(s); indeed, by virtue of what was noticed above we may characterise the two types in more succinct and geometrical terms as follows:

[22] Here we are, of course, using the same notation for the coordinate-invariant object $\phi_v^{(n)} \in C^\infty(\mathscr{N} \cap \mathcal{U})$ and for its coordinate expression in (u, x^A)-coordinates.

[23] Note: with this convention, in general $(\operatorname{supp} F)^\pm$ is larger than $(\operatorname{supp}[F\restriction_{\mathscr{N}}])^\pm$.

(+) $\phi_v^{(n)} = 0$ on $(\mathscr{N} \cap \mathcal{U}) \setminus [(\operatorname{supp} f)^+ \cup (\operatorname{supp} F)^+]$ for all $n \geq 1$;

(−) $\phi_v^{(n)} = 0$ on $(\mathscr{N} \cap \mathcal{U}) \setminus [(\operatorname{supp} f)^- \cup (\operatorname{supp} F)^-]$ for all $n \geq 1$.

Trivially, initial conditions belonging to either type depend smoothly on (x^A). Standard results on the smooth dependence of solutions of ODEs on parameters and initial conditions (see e.g. [CL55, Sec. 1.7])[24] then guarantee that, if $\phi_v^{(n-1)}$ is jointly smooth in u and the x^A-parameters, so is $\phi_v^{(n)}$ as determined by the n-th x^A-parametrised family of ODEs together with initial conditions of either type (+) or type (−). Therefore, corresponding to either type of initial conditions, we obtain:

(+) a sequence $\big(\phi_{v,+}^{(n)}\big)_{n=0}^{\infty}$ of functions in $C^\infty(\mathscr{N} \cap \mathcal{U})$ with $\operatorname{supp} \phi_{v,+}^{(n)}$ contained in the closure in $\mathscr{N} \cap \mathcal{U}$ of $\mathcal{U} \cap [(\operatorname{supp} f)^+ \cup (\operatorname{supp} F)^+]$;

(−) a sequence $\big(\phi_{v,-}^{(n)}\big)_{n=0}^{\infty}$ of functions in $C^\infty(\mathscr{N} \cap \mathcal{U})$ with $\operatorname{supp} \phi_{v,-}^{(n)}$ contained in the closure in $\mathscr{N} \cap \mathcal{U}$ of $\mathcal{U} \cap [(\operatorname{supp} f)^- \cup (\operatorname{supp} F)^-]$.

Since either sequence is uniquely determined by the functions F and f (provided their supports are causally complete), we call it the *(F, f)-sequence of transverse derivatives of type + [resp. −]* with respect to the chosen Gaussian null coordinate neighbourhood $\mathcal{U}$. By Borel's Lemma, since $\mathscr{N} \cap \mathcal{U} = \{v = 0\}$ is closed in $\mathcal{U}$, we can find a function $\phi_\pm^{\mathcal{U}} \in C^\infty(\mathcal{U})$ whose sequence of v-derivatives restricted to $\mathscr{N} \cap \mathcal{U}$ equals the the (F, f)-sequence of transverse derivatives of type $\pm$. The infinite-order jet $j^\infty \phi_\pm^{\mathcal{U}}$ of any such function is uniquely determined, at any point in $\mathscr{N} \cap \mathcal{U}$, by the pair (f, F) – and is obtained by differentiating the elements of $\big(\phi_{v,\pm}^{(n)}\big)_{n=0}^{\infty}$ in directions tangent to $\mathscr{N}$, generalising what was done in (2.38) in Subsection 2.5.2. In particular, we obtain a continuous section of $J^\infty(M, \mathbb{R}) \to M$ along $\mathscr{N} \cap \mathcal{U}$, which admits an extension to a function in $C^\infty(\mathcal{U})$.

The procedure just described can be repeated for all members of an open cover $\{\mathcal{U}_i\}$ of $\mathscr{N}$ by Gaussian null coordinate charts, yielding for each i a continuous section $\Phi_{i,\pm} : \mathscr{N} \cap \mathcal{U}_i \to J^\infty(M; \mathbb{R})$ which admits an extension to a function in $C^\infty(\mathscr{N} \cap \mathcal{U}_i)$. The differential operator P, the and the notion of initial condition of type $\pm$, are defined in coordinate-independent and geometric terms. It follows that, if $i \neq j$,

$$\Phi_{i,\pm} = \Phi_{j,\pm} \quad \text{on } \mathscr{N} \cap \mathcal{U}_i \cap \mathcal{U}_j.$$

Therefore, we can patch together all such sections to obtain a global continuous section

$$\Phi_\pm : \mathscr{N} \to J^\infty(M; \mathbb{R}) \tag{2.41}$$

[24]These results are only local in the case of non-linear ODEs, but become global if the ODEs are linear.

which admits local C^∞ extensions. If we now also assume that $\mathscr{N}$ is achronal and closed, by Proposition 2.5.7 part **(e)**, and the global Whitney extension theorem, Theorem 2.5.2, we can obtain a *global* C^∞ extension of $\Phi_\pm$ to an element of $C^\infty(M)$, i.e. $\Phi_\pm$ is a C^∞-Whitney field on $\mathscr{N}$. By construction then, under all the assumptions made so far:

i) $\operatorname{supp}\Phi_\pm \subseteq (\operatorname{supp} f)^\pm \cup (\operatorname{supp} F)^\pm$;

ii) the 'zeroth-order part' of $\Phi_\pm$ equals f;

iii) any smooth extension ϕ_{app} of $\Phi_\pm$ has $P\phi_{\text{app}} - F$ vanishing to infinite order on $\mathscr{N}$.

Now, by part **(d)** of Proposition 2.5.7, under all the assumptions made so far

$$J^\pm(\operatorname{supp} f \cup [\operatorname{supp} F \cap \mathscr{N}]) \cap \mathscr{N} = (\operatorname{supp} f \cup [\operatorname{supp} F \cap \mathscr{N}])^\pm = (\operatorname{supp} f)^\pm \cup (\operatorname{supp} F)^\pm.$$

Unburdening the notation somewhat by denoting the first of the two sets on the leftmost side by $J^\pm_{(f,F)}$, we can define the following two subsets of M:

$$\begin{aligned} X_\pm &:= J^\pm_{(f,F)} \cap \mathscr{N} = (\operatorname{supp} f)^\pm \cup (\operatorname{supp} F)^\pm, \\ Y_\pm &:= \overline{J^\pm_{(f,F)} \setminus X_\pm} = \overline{J^\pm_{(f,F)} \setminus [(\operatorname{supp} f)^\pm \cup (\operatorname{supp} F)^\pm]}. \end{aligned}$$

$X_\pm$ and $Y_\pm$ are closed and, since $J^\pm_{(f,F)}$ is closed, $X_\pm \cup Y_\pm = J^\pm_{(f,F)}$ and we may regard the closures in the definition of $Y_\pm$ as being taken in $J^\pm_{(f,F)}$ with the relative topology. Then, denoting the set-theoretic operator of closure in a subset A of M equipped with the relative topology as cl_A, and similarly the topological boundary operator in A as bd_A,

$$\begin{aligned} X_\pm \cap Y_\pm &= X_\pm \cap \operatorname{cl}_{J^\pm_{(f,F)}} \left[J^\pm_{(f,F)} \setminus X_\pm \right] \\ &= \operatorname{cl}_{J^\pm_{(f,F)}} X_\pm \cap \operatorname{cl}_{J^\pm_{(f,F)}} \left[J^\pm_{(f,F)} \setminus X_\pm \right] \\ &= \operatorname{bd}_{J^\pm_{(f,F)}} X_\pm \end{aligned} \tag{2.42}$$

(the fact that $\mathscr{N}$ is closed was used in obtaining the second equality). We now put to use the fact that both $\mathscr{N}$ and $J^\pm_{(f,F)}$ are embedded topological submanifolds. This is also a good time to recapitulate the assumptions made in the course of the discussion so far.

Lemma 2.5.8. *Let $(M, g, \mathfrak{t})$ be a globally hyperbolic Lorentzian manifold, and $\mathscr{N}$ be a closed, achronal hypersurface whose null generators, when reparametrised as null geodesics entirely contained in $\mathscr{N}$, are future [resp. past] inextendible as geodesics in (M, g). Further, let $f \in C^\infty(\mathscr{N})$ and $F \in C^\infty(M)$ be such that* supp f *and* supp F *are* *causally complete. Then, with $X_\pm$, $Y_\pm$ defined as above,*

$$X_\pm \cap Y_\pm \subseteq \mathrm{cl}_{\mathscr{N}} \left[\mathscr{N} \setminus X_\pm\right].$$

Proof. Let $\dim M = n+1$. By Equation (2.42), we need to show that $\mathrm{bd}_{\dot{J}^\pm_{(f,F)}} X_\pm \subseteq \mathrm{cl}_{\mathscr{N}} \left[\mathscr{N} \setminus X_\pm\right]$. We proceed by contradiction, assuming instead that we can find $p \notin \mathrm{cl}_{\mathscr{N}} \left[\mathscr{N} \setminus X_\pm\right]$ such that $p \in \mathrm{bd}_{\dot{J}^\pm_{(f,F)}} X_\pm$. Since $p \in \mathscr{N} \setminus \mathrm{cl}_{\mathscr{N}} \left[\mathscr{N} \setminus X_\pm\right]$ which equals the interior of $X_\pm$ in $\mathscr{N}$, p belongs to a subset U, open in $\mathscr{N}$, which is entirely contained in $X_\pm$ and thus also in $\dot{J}^\pm_{(f,F)}$. Since $\mathscr{N}$ has the subspace topology, there exists an open set $\mathscr{U}$ in M such that $U = \mathscr{U} \cap \mathscr{N}$. Since $\mathscr{N}$ is an embedded hypersurface, without loss of generality we may assume that there is a homeomorphism

$$\varphi : U \to \varphi(U) \text{ open in } \mathbb{R}^n,$$

where U has the subspace topology inherited from, equivalently, M, $\mathscr{N}$ or $\dot{J}^\pm_{(f,F)}$. But, by well-known results [O'N83, Prop. 14.25], $\dot{J}^\pm_{(f,F)}$ is too an embedded topological (indeed, actually locally Lipschitz) hypersurface of M. It follows that there is an open set $\mathscr{V}$ in M, containing p, and a homeomorphism

$$\psi : \mathscr{V} \cap \dot{J}^\pm_{(f,F)} =: V \to \psi(V) \text{ open in } \mathbb{R}^n,$$

where V has the subspace topology inherited from, equivalently, M or $\dot{J}^\pm_{(f,F)}$. Now, since $U \subseteq \dot{J}^\pm_{(f,F)}$,

$$U \cap V = U \cap \mathscr{V}$$

and is therefore open in U. It follows that $\varphi(U \cap V)$ is open in $\mathbb{R}^n$, and the composition

$$\mathbb{R}^n \supseteq \varphi(U \cap V) \xrightarrow{\varphi^{-1}\restriction_{\varphi(U\cap V)}} U \cap V \xrightarrow{\psi\restriction_{U\cap V}} \psi(U \cap V) \subseteq \mathbb{R}^n,$$

where the topology of the set in the middle is the subspace topology inherited from $\dot{J}^\pm_{(f,F)}$, is an injective continuous map. By invariance of domain, $\psi(U \cap V)$ is open in $\mathbb{R}^n$ and this composition is a homeomorphism. This proves that $U \cap V$ is an open subset of $\dot{J}^\pm_{(f,F)}$, entirely contained in $X_\pm$. Therefore it cannot belong to the boundary $\mathrm{bd}_{\dot{J}^\pm_{(f,F)}} X_\pm$, and the claim is proved. □

The reason why Lemma 2.5.8 is relevant to the discussion that preceded it is that it allows to formulate a natural 'gluing problem' as follows. By Property **i)** on p. 86 and by the definition of $X_\pm$, the continuous section $\Phi_\pm$ defined in (2.41) has support contained in $X_\pm$. Therefore, it vanishes on $\mathscr{N} \setminus X_\pm$ and, by continuity, it must actually vanish on $\mathrm{cl}_{\mathscr{N}}\,[\mathscr{N} \setminus X_\pm]$. By Lemma 2.5.8 then,

$$\Phi_\pm = 0 \quad \text{on} \quad X_\pm \cap Y_\pm. \tag{2.43}$$

We have a natural gluing problem for C^∞-Whitney fields on manifolds, generalising the situation we already encountered in Subsection 2.5.2: $\Phi_\pm$, and the zero section $\mathbf{0}_{Y_\pm}$ of $J^\infty(M,\mathbb{R}) \to M$ along $Y_\pm$, are C^∞-Whitney fields on the closed sets $X_\pm$ and $Y_\pm$ respectively, and by Equation (2.43) they agree on $X_\pm \cap Y_\pm$. Now, around any point in $\dot{J}^\pm_{(f,F)}$ we can find a diffeomorphism from an open set $\mathcal{W}$ in M – which has dimension $n+1$ – to an open set in $\Omega' \times \mathbb{R}$ where $\Omega' \subseteq \mathbb{R}^n$ is open, such that the image of $\mathcal{W} \cap \dot{J}^\pm_{(f,F)}$ under this diffeomorphism is the graph $\Gamma(\theta')$ of a Lipschitz function $\theta' : \Omega' \to \mathbb{R}$. Then, Lemma 2.5.6 may be applied around any point in $X_\pm \cap Y_\pm$ with the following assignments: Ω is the graph projection of $\Gamma(\theta') \setminus \widetilde{X_\pm}$, where $\widetilde{X_\pm}$ denotes the diffeomorphic image of $\mathcal{W} \cap X_\pm$; $\theta : \Omega \to \mathbb{R}$ is $\theta' \restriction_\Omega$; the set S is $\Gamma(\theta' \restriction_{\Omega' \setminus \Omega}) = \widetilde{X_\pm}$. Globalising, we have proven:

Proposition 2.5.9. *Under the same assumptions as in Lemma 2.5.8, $X_\pm$ and $Y_\pm$ are simply separated. There exists a C^∞-Whitney field $\Phi^\pm_{(f,F)}$ on $X_\pm \cup Y_\pm = \dot{J}^\pm_{(f,F)}$ which equals $\Phi_\pm$ on $X_\pm$ and zero on $Y_\pm$, and therefore there exists a function $\phi^\pm_{\mathrm{app}} \in C^\infty(M)$ whose Whitney data on $\dot{J}^\pm_{(f,F)}$ is the one given by $\Phi^\pm_{(f,F)}$.* □

We have terminated all the hard work needed to generalise the procedure described in 2.5.2.1. The remainder of the construction carries through in the same way, and as in 2.5.2.1 we specialise to the homogeneous case $F \equiv 0$, so that $\dot{J}^\pm_{(f,F)} = \dot{J}^\pm(\operatorname{supp} f)$. Namely, let $\phi^\pm_{\mathrm{app}}$ be as in Proposition 2.5.9. Then:

- $P\phi^\pm_{\mathrm{app}}$ vanishes to infinite order on $\dot{J}^\pm(\operatorname{supp} f)$ by construction, so $\chi_{J^\pm(\operatorname{supp} f)} \cdot P\phi^\pm_{\mathrm{app}}$ is smooth and has support in $J^\pm(\operatorname{supp} f)$ which is past/future compact.
- $E^\pm[\chi_{J^\pm(\operatorname{supp} f)} \cdot P\phi^\pm_{\mathrm{app}}]$ is smooth and has support still in $J^\pm(\operatorname{supp} f)$. In particular, it vanishes to infinite order on $\dot{J}^\pm(\operatorname{supp} f)$.
- $\phi^\pm_f \coloneqq \phi^\pm_{\mathrm{app}} - E^\pm[\chi_{J^\pm(\operatorname{supp} f)} \cdot P\phi^\pm_{\mathrm{app}}]$ is smooth on M, equals f on $\mathscr{N}$, vanishes to infinite order on $\dot{J}^\pm(\operatorname{supp} f) \setminus \mathscr{N}$, and

$$P\phi^\pm_f = P\phi^\pm_{\mathrm{app}} - \chi_{J^\pm(\operatorname{supp} f)} \cdot P\phi^\pm_{\mathrm{app}} = 0 \quad \text{on } J^\pm(\operatorname{supp} f).$$

Now consider the function $\phi_f : M \to \mathbb{R}$ defined as follows:

$$\phi_f(x) = \begin{cases} \phi^+_f(x) & \text{if } x \in J^+(\operatorname{supp} f) \\ \phi^-_f(x) & \text{if } x \in J^-(\operatorname{supp} f) \\ 0 & \text{if } x \notin J(\operatorname{supp} f) = J^+(\operatorname{supp} f) \cup J^-(\operatorname{supp} f). \end{cases} \tag{2.44}$$

Part **(c)** of Proposition 2.5.7 guarantees that this is well-defined, since both ϕ_f^+ and ϕ_f^- equal f on $\mathscr{N}$. Together with the fact that $\phi_f^\pm$ vanishes (to infinite order) on $\dot{J}^\pm(\operatorname{supp} f) \setminus \mathscr{N}$, this guarantees that ϕ_f is globally continuous. However, it may fail to be globally C^2 and therefore it may not be a classical solution in general.

There remain at least three natural questions about our construction so far:

Q1 Is ϕ_f thus constructed at least a solution *in a weak sense* of the equation $P\phi_f = 0$?

Q2 Given f, is ϕ_f unique in some natural sense? Clearly, the auxiliary functions $\phi^\pm_{\mathrm{app}}$ used in constructing ϕ_f were not uniquely determined.

Q3 If **Q2** can be answered in the affirmative, is the resulting (well-defined) assignment $f \mapsto \phi_f$ linear?

Q4 Is the assignment $f \mapsto \phi_f$ continuous in some natural choice of function spaces and topologies?

We will not touch on **Q4** in this work, but we now proceed to address the other questions.

Answering Q1: Distributional solution property

Recall (see also Appendix C) that we can view (real-valued) functions on M which are locally integrable with respect to the measure defined by $\mathrm{d}\mu_g$ as distributions in a natural way: if ϕ is such a function then the assignment

$$C_0^\infty(M) \ni \psi \mapsto \int_M \phi\psi \, \mathrm{d}\mu_g$$

defines a (real) distribution.

Definition 2.5.10. Let (M, g, τ) be a time-oriented Lorentzian manifold, and P be a normally hyperbolic differential operator with smooth coefficients acting on real-valued functions, i.e. an operator of the form given in Equation (2.30). Given a closed null hypersurface $\mathscr{N}$ and $f \in C^\infty(\mathscr{N})$, we will refer to a continuous function τ on M with the following properties:

(a) τ is a distributional solution of P,

(b) $\tau\restriction_{\mathscr{N}} = f$,

(c) $\operatorname{supp} \tau \subseteq \overline{I}(\operatorname{supp} f) := \overline{I^+}(\operatorname{supp} f) \cup \overline{I^-}(\operatorname{supp} f)$,

as a *solution which falls entirely through $\mathscr{N}$ with data f.*

With these definitions in place, **Q1** is the question of whether or not ϕ_f as defined in Equation (2.44) is a solution falling entirely through $\mathscr{N}$ with data f. The following simple resolution was suggested to us by A. Strohmaier. It is based on a well-known divergence identity, proved e.g. in [BGP07, Lem. 3.2.1] in greater generality, and which we will also use in answering **Q2**.

Proposition 2.5.11. *Let (M,g) be a Lorentzian manifold, and P be an operator of the form given in Equation (2.30). Then, for every $u, v \in C^2(M)$,*

$$uPv - [P^*u]v = -\operatorname{div}(j[u,v]) \tag{2.45}$$

where $j[u,v]$ is a C^1 vector field uniquely defined by

$$g(j[u,v], Z) = [\nabla_Z u]v - u[\nabla_Z v] \quad \forall\, Z \in \mathfrak{X}(M),$$

i.e., in abstract index notation, by $j[u,v]_a = [\nabla_a u]v - u[\nabla_a v]$. □

We now need to couple Proposition 2.5.11 with a version of the divergence theorem [also known as the *Gauss–Green(–Ostrogradsky)* theorem] of differential calculus which is sufficiently general for our purposes. A well-known elementary version in the context of oriented Riemannian manifolds with smooth boundary is usually obtained from *Stokes Theorem*, which states that

$$\int_M \mathrm{d}\omega = \int_{\partial M} \omega \tag{2.46}$$

if ω is a differential form of degree equal to $\dim M - 1$, sufficiently regular (say, C^1) and with compact support on M. In the case of Lorentzian manifolds with smooth boundary it is often explained how to use Equation (2.46) to again obtain a simple version of the divergence theorem which applies whenever ∂M is everywhere non-degenerate for the metric g – that is, whenever ∂M is everywhere either spacelike or timelike.

We will however require a generalisation of the divergence theorem which accommodates for both the possibility that the boundary of the integration region may be rougher than C^∞ or even than C^1, and for the fact that portions of it may be null. Not requiring orientability of the manifold would also be a bonus. We now briefly discuss how such a generalisation can be extracted from the existing literature.

The discipline of *geometric measure theory* [Whi57, Fed69, Har93] allows to extend tools from the classical differential geometry of smooth manifolds to a larger class of geometric objects that are not necessarily smooth, in such a way that integration on these objects retains a geometric significance. In particular, it allows for far-reaching and natural generalisations of precisely the sort of theorems of differential calculus which we are now discussing – such as Stokes' Theorem and the divergence theorem. We will in fact not need anywhere near the full power of the results in modern geometric measure theory here, but it is good to bear in mind that they are there, should one desire to further sharpen the procedure we are outlining. We first recall how the divergence theorem can be obtained independently of Stokes' theorem, since it is really a statement involving densities and not forms. In particular, no orientability assumptions are needed. We follow [LS90, Secs. 10.5 & 10.6].

Namely, let M be a smooth manifold, with or without boundary, ρ be a smooth section of the bundle $|\Lambda M| \to M$ of $(n{+}1)$ densities on M, and Υ be a C^1 vector field on M. Then we may define the *divergence of* Υ *relative to* ρ to be the density

$$\mathrm{div}_\rho \Upsilon = \pounds_\Upsilon \rho$$

where $\pounds$ denotes the Lie derivative. Now, in the case in which $\partial M = \emptyset$ we wish to define a special class of domains in M.

Definition 2.5.12. If M is a smooth manifold of dimension $n+1$ without boundary, an open set $D \subseteq M$ is called a *domain with regular boundary in* M if, for every $p \in M$, there is a chart (U, φ) about p such that one of the following three possibilities holds:

i) $U \cap D = \emptyset$;

ii) $U \subseteq D$;

iii) $\varphi(U \cap D) = \varphi(U) \cap \left\{(y^0, \ldots, y^n) \in \mathbb{R}^{n+1} \;\middle|\; y^n > 0\right\}$.

If D is a domain with regular boundary, then as discussed in [LS90, p. 420], we may give the *topological* boundary $\dot{D}$ of D the structure of a smooth manifold of dimension n which is in particular an embedded hypersurface in the ambient manifold M. Now, given any point $p \in \dot{D}$, a vector $\xi \in T_pM \setminus T_p\dot{D}$ has the property that any differentiable curve $c : I \to M$ with $\dot{c}(0) = \xi$ is either entirely contained in D or entirely contained in $M \setminus \overline{D}$ for sufficiently small positive t. In the first [resp. second] case, we say that ξ is *inward* [resp. *outward*] *pointing with respect to* D, or that it points towards [resp. away from] D.

The final ingredient before we are able to state a preliminary version of the divergence theorem comes from the following observation: let i be the embedding $\dot{D} \to M$. Then the pointwise *interior product* of Υ with ρ at each point $p \in \dot{D}$, defined by

$$\iota_\Upsilon \rho(X_0, \ldots, X_n) = \rho(T_pi(X_0), \ldots, T_pi(X_n), \Upsilon(p)) \quad \forall\ X_0, \ldots, X_n \in T_p\dot{D}, \tag{2.47}$$

just as in the case of differential forms, defines a smooth density on $\dot{D}$. We are now ready to state the classical divergence theorem for densities, Theorem 6.1 in [LS90].

Theorem 2.5.13. *Let M be a smooth manifold (without boundary), Υ be a C^1 vector field with compact support on M and D be a domain with regular boundary in M. Defining the function ϵ_Υ on $\dot{D}$ by*

$$\epsilon_\Upsilon(x) = \begin{cases} 1 & \textit{if } \Upsilon(x) \textit{ points away from } D \\ -1 & \textit{if } \Upsilon(x) \textit{ points away from } D \\ 0 & \textit{if } \Upsilon(x) \textit{ is tangent to } \dot{D} \end{cases}$$

Then

$$\int_{\overline{D}} \mathrm{div}_\rho \Upsilon = \int_D \mathrm{div}_\rho \Upsilon = \int_{\dot{D}} \epsilon_\Upsilon \cdot \iota_\Upsilon \rho. \qquad \square$$

Now, if (M, g) is a semi-Riemannian manifold, ∇ is the Levi-Civita connection associated to g, and $\mathrm{d}\mu_g$ is the volume density arising from g, it is standard that

$$\mathrm{div}_{\mathrm{d}\mu_g} \Upsilon = \mathrm{div}\, \Upsilon \, \mathrm{d}\mu_g = \mathrm{Tr}(\nabla \Upsilon) \, \mathrm{d}\mu_g = (\nabla_a \Upsilon^a) \, \mathrm{d}\mu_g,$$

so that by Theorem 2.5.13 we have

$$\int_D \mathrm{div}\, \Upsilon \, \mathrm{d}\mu_g = \int_{\dot{D}} \epsilon_\Upsilon \cdot \iota_\Upsilon \mathrm{d}\mu_g. \tag{2.48}$$

The important observation at this point is that although we stated Theorem 2.5.13 for domains with *regular* boundary in the sense of Definition 2.5.12, this is an unnecessary restriction. Indeed, the fact that Theorem 2.5.13 extends in a natural way to domains with much rougher boundaries – and to vector fields of regularity lower than C^1 – is one of the great achievements of geometric measure theory. We will not attempt to lower the regularity of our vector fields here.[25] However, as shown in Appendix I in [Tay06] (see also [EG15, Thm. 5.16]), Theorem 2.5.13 holds almost *verbatim* if D is merely an open domain whose topological boundary can be locally represented as the graph of a Lipschitz function – in other words, if $\dot{D}$ is an embedded locally Lipschitz hypersurface in M. In that case, the integrand $\epsilon_\Upsilon \cdot \iota_\Upsilon \rho$ is not defined everywhere on $\dot{D}$, but is still defined *almost everywhere* on $\dot{D}$ since, by Rademacher's theorem on Lipschitz functions (see p. 29), $\dot{D}$ has a tangent space at almost all of its points.

We now make some further pointwise considerations on the integrand $\epsilon_\Upsilon \cdot \iota_\Upsilon \mathrm{d}\mu_g$ in Equation (2.48). Locally around any point $p \in \dot{D}$ (at which $\dot{D}$ is differentiable), M is orientable and the metric volume density $\mathrm{d}\mu_g$ and metric volume *form* vol_g – both

[25] In [BW15], for instance, a version of the divergence theorem involving vector fields which were merely in the Sobolev space H^1 – and for boundaries which were locally Lipschitz – was needed. Then, the integrand in the integral over $\dot{D}$ is intended in the sense of the trace theorem for Sobolev spaces (see e.g. [Alt16, A8.8] or [Tie15, Thm. 1.7.2]).

restricted to a suitable open set around p – are related at p by $\mathrm{d}\mu_g = |\mathrm{vol}_g|$, no matter what local choice of orientation is made. It follows that, at any $p \in \dot{D}$, $\iota_\Upsilon \mathrm{d}\mu_g$ defined in Equation (2.47), and the more customary interior product $\iota_\Upsilon \mathrm{vol}_g$, are related by

$$\iota_\Upsilon \mathrm{d}\mu_g = |i^*[\iota_\Upsilon \mathrm{vol}_g]|. \tag{2.49}$$

Now, let $0 \neq m \in T_p^* M$ be a (co)normal to $\dot{D}$ at p, i.e. a covector such that $m(T_p \dot{D}) = \{0\}$.[26] The relation

$$m \wedge \sigma = \mathrm{vol}_g \tag{2.50}$$

can be inverted to obtain a uniquely defined n-covector σ at p (where $n = \dim M - 1$). We calculate, using the fact that the interior product for forms is a derivation, and the fact that $i^* m = 0$ for any conormal:

$$\begin{aligned} i^*[\iota_\Upsilon \mathrm{vol}_g] &= i^*[\iota_\Upsilon (m \wedge \sigma)] = i^*[(\iota_\Upsilon m) \wedge \sigma + (-1)^{\deg m}\, m \wedge (\iota_\Upsilon \sigma)] \\ &= m(\Upsilon) i^* \sigma - i^* m \wedge i^*(\iota_\Upsilon \sigma) \\ &= m(\Upsilon) i^* \sigma \quad \text{at } p. \end{aligned} \tag{2.51}$$

Now suppose that, at a differentiability point p of $T_p \dot{D}$, the tangent space $T_p \dot{D}$ is null with respect to the Lorentzian metric g. This means that, for any choice of conormal m at p, the vector $m^\sharp \in T_p M$ obtained by index raising is null and belongs to $T_p \dot{D}$. Let Υ be the vector field $j[u, v]$ defined in Proposition 2.5.11. Then

$$m(\Upsilon) = [\nabla_{m^\sharp} u] v - u [\nabla_{m^\sharp} v] \tag{2.52}$$

and, by Equations (2.49) and (2.51),

$$\epsilon_\Upsilon \cdot \iota_\Upsilon \mathrm{d}\mu_g = \epsilon_\Upsilon \cdot \big| [\nabla_{m^\sharp} u] v - u [\nabla_{m^\sharp} v] \big| \cdot |i^* \sigma| \tag{2.53}$$

for *any* conormal m at p if σ is given by Equation (2.50). We are now ready to give our resolution to **Q1**.

Theorem 2.5.14. *Let ϕ_f be defined as in Equation (2.44). Then $P\phi_f = 0$ distributionally, and thus ϕ_f is a solution which falls entirely through $\mathscr{N}$.*

Proof. We need to check that $\phi_f[P^*\psi] = 0$ for all $\psi \in C_0^\infty(M)$. To unburden the notation, we denote the smooth functions ϕ_f^+ and ϕ_f^- in Equation (2.44) simply by ϕ^+

[26] We may say that m is inward pointing if $m(Y) > 0$ for all vectors $Y \in T_p M \setminus T_p \dot{D}$ which are inward pointing with respect to D, and outward pointing if $m(Y) < 0$ for all such vectors. But this distinction will not be needed here.

and ϕ^- (respectively). Since $\operatorname{supp}\phi_f \subseteq J^+(\operatorname{supp} f) \cup J^-(\operatorname{supp} f)$, we have

$$\begin{aligned}\phi_f[P^*\psi] = \int_M \phi_f P^*\psi \, \mathrm{d}\mu_g &= \int_{J^+(\operatorname{supp} f)} \phi_f P^*\psi \, \mathrm{d}\mu_g + \int_{J^-(\operatorname{supp} f)} \phi_f P^*\psi \, \mathrm{d}\mu_g \\ &= \int_{\overline{I^+}(\operatorname{supp} f)} \phi^+ P^*\psi \, \mathrm{d}\mu_g + \int_{\overline{I^-}(\operatorname{supp} f)} \phi^- P^*\psi \, \mathrm{d}\mu_g.\end{aligned}$$

We may use Proposition 2.5.11 to convert the integrands in the last line to

$$\left\{[P\phi^\pm]\psi + \operatorname{div}(j[\psi, \phi^\pm])\right\} \, \mathrm{d}\mu_g.$$

But on $\overline{I^\pm}(\operatorname{supp} f)$, $P\phi^\pm = 0$. We are left with

$$\phi_f[P^*\psi] = \sum_\pm \int_{\overline{I^\pm}(\operatorname{supp} f)} \operatorname{div}(j[\psi, \phi^\pm]) \, \mathrm{d}\mu_g.$$

Since $\psi \in C_0^\infty(M)$ and $\phi^\pm \in C^\infty(M)$, the vector field $\Upsilon := j[\psi, \phi^\pm]$ is actually smooth and compactly supported on M. Furthermore, $I^\pm(\operatorname{supp} f)$ is open in M and, as we already discussed, $\dot{I}^\pm(\operatorname{supp} f) = \dot{J}^\pm(\operatorname{supp} f)$ is an embedded locally Lipschitz hypersurface in M. We may thus apply the divergence theorem to each integral and obtain

$$\phi_f[P^*\psi] = \sum_\pm \int_{\dot{J}^\pm(\operatorname{supp} f)} \epsilon^\pm_{j[\psi,\phi^\pm]} \cdot \iota_{j[\psi,\phi^\pm]} \mathrm{d}\mu_g.$$

But since $\phi^\pm$ vanishes to infinite order on $\dot{J}^\pm(\operatorname{supp} f) \setminus \mathscr{N}$, we have in fact

$$\phi_f[P^*\psi] = \sum_\pm \int_{\dot{J}^\pm(\operatorname{supp} f) \cap \mathscr{N}} \epsilon^\pm_{j[\psi,\phi^\pm]} \cdot \iota_{j[\psi,\phi^\pm]} \mathrm{d}\mu_g = \sum_\pm \int_{(\operatorname{supp} f)^\pm} \epsilon^\pm_{j[\psi,\phi^\pm]} \cdot \iota_{j[\psi,\phi^\pm]} \mathrm{d}\mu_g. \tag{2.54}$$

The first crucial observation is now that, whenever p belongs to $(\operatorname{supp} f)^+$ [resp. $(\operatorname{supp} f)^-$] and is a point at which $\dot{J}^+(\operatorname{supp} f)$ [resp. $\dot{J}^-(\operatorname{supp} f)$] is differentiable, then $T_p\dot{J}^+(\operatorname{supp} f) = T_p\mathscr{N}$ [resp. $T_p\dot{J}^-(\operatorname{supp} f) = T_p\mathscr{N}$] when both are regarded as subspaces of T_pM. Therefore, by the discussion preceding the statement of this theorem, for any covector $m \in T_p^*M$ which is (co)normal to $\dot{J}^+(\operatorname{supp} f)$ [resp. $\dot{J}^-(\operatorname{supp} f)$] at p, $m^\sharp \in T_p\mathscr{N}$ and

$$m\big(j[\psi, \phi^\pm]\big) = [\nabla_{m^\sharp}\psi]\phi^\pm - \psi[\nabla_{m^\sharp}\phi^\pm] = [\nabla_{m^\sharp}\psi]f - \psi\, m^\sharp(f), \tag{2.55}$$

the last equality following from the fact that both ϕ^+ and ϕ^- are identically equal to f on $\mathscr{N}$, by construction. In particular, this quantity vanishes outside $\operatorname{supp} f$ and Equation (2.54) becomes

$$\phi_f[P^*\psi] = \sum_\pm \int_{\operatorname{supp} f} \epsilon^\pm_{j[\psi,\phi^\pm]} \cdot \iota_{j[\psi,\phi^\pm]} \mathrm{d}\mu_g,$$

where the integrand on the right-hand side is defined almost everywhere. The second observation is that, whenever p belongs to $(\operatorname{supp} f)^+ \cap (\operatorname{supp} f)^- \supseteq \operatorname{supp} f$, then *any vector in T_pM which is outward pointing with respect to $I^+(\operatorname{supp} f)$ is inward pointing with respect to $I^-(\operatorname{supp} f)$ (and vice-versa).* Let $\boldsymbol{m}$ now be a field of covectors defined at each point of $\operatorname{supp} f$ and everywhere normal to $\mathscr{N}$ – but otherwise unrestricted and in particular not necessarily continuous. Define the corresponding field of n-covectors $\boldsymbol{\sigma}$ by demanding that Equation (2.50) be satisfied at each point, i.e. that

$$\boldsymbol{m}(p) \wedge \boldsymbol{\sigma}(p) = \mathrm{vol}_g(p) \quad \forall\, p \in \operatorname{supp} f.$$

Then, by Equation (2.53), if $i^\pm$ is the embedding of $\dot{J}^\pm(\operatorname{supp} f)$ into M,

$$\phi_f[P^*\psi] = \sum_{\pm} \int_{\operatorname{supp} f} \epsilon^{\pm}_{j[\psi,\phi^\pm]} \cdot \big|[\nabla_{\boldsymbol{m}^\sharp} u]v - u[\nabla_{\boldsymbol{m}^\sharp} v]\big| \cdot |(i^\pm)^*\boldsymbol{\sigma}|.$$

But clearly, when viewed as forms on tangent spaces to $\mathscr{N}$, $(i^+)^*\boldsymbol{\sigma}$ and $(i^-)^*\boldsymbol{\sigma}$ are equal. Therefore, we can prove that $\phi_f[P^*\psi] = P\phi_f[\psi] = 0$ if we can argue that $\epsilon^+_{j[\psi,\phi^+]} = -\epsilon^-_{j[\psi,\phi^-]}$ on $\operatorname{supp} f$. But this follows from Equation (2.55), which implies that

$$\boldsymbol{m}(p)(j[\psi,\phi^+]) = \boldsymbol{m}(p)(j[\psi,\phi^-]);$$

this, in turn, readily implies that if $j[\psi,\phi^+]$ is tangent to $T_p\dot{J}^+(\operatorname{supp} f)$ then $j[\psi,\phi^-]$ is tangent to $T_p\dot{J}^-(\operatorname{supp} f)$, and that if $j[\psi,\phi^+]$ is inward pointing relative to $I^+(\operatorname{supp} f)$ then $j[\psi,\phi^-]$ is outward pointing relative to $I^-(\operatorname{supp} f)$ (and vice-versa). But this is precisely the statement that $\epsilon^+_{j[\psi,\phi^+]} = -\epsilon^-_{j[\psi,\phi^-]}$. This concludes the proof of the theorem. □

Answering Q2 & Q3: Uniqueness and Linear Dependence

Equation (2.55) is also at the heart of the argument for uniqueness, which is very simple, provided we are more precise as to what exactly we mean by 'uniqueness'. The following Lemma will show us the way. In fact, a stronger version of the Lemma below was proved as Theorem 23 in [BW15] for functions with lower regularity than we will prescribe, but we do not go into the details of such a strengthening in the interest of keeping the discussion concise.

Lemma 2.5.15. *Let $(M, g, \mathfrak{t})$ be a time-oriented and globally hyperbolic Lorentzian manifold. Let $\mathscr{N}$ be a closed, achronal and null hypersurface whose null generators, when reparametrised as null geodesics entirely contained in $\mathscr{N}$, are future and past inextendible in (M, g). Let P be a differential operator of the form given in Equation (2.30). Let S be a future [resp. past] causally complete subset of $\mathscr{N}$, and ζ be a continuous real-valued function with the following properties:*

- *ζ is C^2 on $I^+(S)$ [resp. on $I^-(S)$];*
- *$\zeta\restriction_{I^+(S)}$ [resp. $\zeta\restriction_{I^-(S)}$] can be extended to a function ζ^+ [resp. ζ^-] defined and C^2 in a neighbourhood of $J^+(S)$ [resp. $J^-(S)$];*
- *ζ^+ [resp. ζ^-] vanishes to first order on $\dot{J}^+(S) \setminus \mathscr{N}$ [resp. $\dot{J}^-(S) \setminus \mathscr{N}$];*
- *$P\zeta = 0$ on $I^+(S)$ [resp. on $I^-(S)$];*
- *$\zeta\restriction_{\dot{J}^+(S)} = 0$ [resp. $\zeta\restriction_{\dot{J}^-(S)} = 0$].*

Then, $\zeta = 0$ on $J^+(S)$ [resp. on $J^-(S)$].

Proof. Viewing ζ as a distribution, and considering the case in which S is future causally complete, we need to show that $\int_M \zeta\psi \, \mathrm{d}\mu_g = 0$ for all test functions $\psi \in C_0^\infty(M)$ with $\operatorname{supp}\psi \subseteq J^+(S)$. For such test functions, actually

$$\int_M \zeta\psi \, \mathrm{d}\mu_g = \int_{J^+(S)} \zeta^+\psi \, \mathrm{d}\mu_g$$

where ζ^+ is the extension assumed to exist in the statement of the Lemma. Now, the adjoint operator P^* is also normally hyperbolic on M, and thus admits causal Green operators which we will denote by $G^\pm$, while $E^\pm$ will denote the causal Green operators for P. Therefore, for such a test function ψ, and using Proposition 2.5.11,

$$\begin{aligned}\int_M \zeta^+\psi \, \mathrm{d}\mu_g = \int_{J^+(S)} \zeta^+ P^* G^- \psi \, \mathrm{d}\mu_g &= \int_{J^+(S)} \left\{ [G^-\psi] P\zeta^+ + \operatorname{div}(j[G^-\psi, \zeta^+]) \right\} \mathrm{d}\mu_g \\ &= \int_{J^+(S)} \operatorname{div}(j[G^-\psi, \zeta^+]) \, \mathrm{d}\mu_g.\end{aligned}$$

As was done in the proof of Theorem 2.5.14, we use the divergence theorem – which can be done since $G^-\psi$ and ζ^+ may be ‘cut off’ outside $J^+(S)$ in such a way that (using the same letters to denote the result of this cutting off) $\operatorname{supp} j[G^-\psi, \zeta^+]$ is compact – and re-express the integrand of the resulting integral over $\dot{J}^+(S)$ pointwise using an analog of Equation (2.55). Since ζ^+ vanishes to first order on $\dot{J}^+(S) \setminus \mathscr{N}$ and it is C^2, it does so also on the closure of this set and we need only worry about the contribution coming from $\dot{J}^+(S) \setminus \overline{\dot{J}^+(S) \setminus \mathscr{N}}$, which is an open subset of $\dot{J}^+(S)$ contained in $\mathscr{N}$. But, in this portion, the tangent spaces to $\dot{J}^+(S)$ are null, and thus contain their normal vectors, almost everywhere. Since ζ^+ vanishes identically on $\dot{J}^+(S)$, so too do its directional derivatives in directions tangent to $\dot{J}^+(S)$ (when the latter are available), in particular on this portion. It follows that the boundary integrand vanishes pointwise on this portion, and we have proved that ζ vanishes as a distribution on $J^+(S)$. □

Remark. There is a sense that we could have asked less of $\zeta^\pm$ in the hypotheses of Lemma 2.5.15, while still being able to draw the same conclusion. In particular, although we have not attempted to prove this, one might expect that, with $S \subseteq \mathscr{N}$ future [resp. past] causally complete, $\dot{J}^+(S)$ [resp. $\dot{J}^-(S)$] should have *null* tangent spaces whenever they exist – i.e. it should be a characteristic locally Lipschitz hypersurface for P. If that were the case, then assuming that ζ^+ [resp. ζ^-] vanishes everywhere on $\dot{J}^+(S)$ [resp. on $\dot{J}^-(S)$] would be sufficient without the additional assumption, which we have made in the statement of the lemma to keep ourselves on the side of care, that it does so also to first order on $\dot{J}^+(S) \setminus \mathscr{N}$ [resp. on $\dot{J}^-(S) \setminus \mathscr{N}$]. It is often stated without proof that (specialising w.l.o.g to the 'future' case), in the case of a singleton $S = \{p\}$, $\dot{J}^+(S)$ does have the property we seek, but we are not aware of a location in the literature where this is proven carefully. While it is obviously true that through any point in $J^+(p) \setminus \{p\}$ there is a null geodesic $\gamma : I \to M$ with $\gamma(\mathring{I}) \subseteq \dot{J}^+(p) \setminus \{p\}$, this in itself is not sufficient to prove that any tangent space to $\dot{J}^+(p) \setminus \{p\}$ must be null – indeed, one may regard several *timelike* hypersurfaces as being 'ruled' by null geodesics in a similar fashion! The other assumption which must be used is the achronality of $\dot{J}^+(p) \setminus \{p\}$. If we *knew* that $\dot{J}^+(p) \setminus \{p\}$ is C^1, we could (as is done in [Kup87, Thm. 1]) prove that any tangent space to it cannot contain timelike vectors – and thus really must be null – by arguing that any C^1 curve contained in $\dot{J}^+(p) \setminus \{p\}$, with an initial timelike tangent vector, would have derivatives which, by continuity, would have to remain initially timelike. And this would contradict the achronality of $\dot{J}^+(p) \setminus \{p\}$. But if we only knew that $\dot{J}^+(p) \setminus \{p\}$ is locally Lipschitz then we would run into difficulties. We could easily construct a locally Lipschitz curve starting at a point of differentiability of $\dot{J}^+(p) \setminus \{p\}$, entirely contained in $\dot{J}^+(p) \setminus \{p\}$ and with initially timelike tangent vector in the sense of Rademacher's theorem. But it would not be immediately guaranteed that the derivatives would vary continuously on their dense subset of definition (in the subspace topology inherited from the interval of definition of the curve) at the initial point. It seems to us that one has to dig into the finer differentiability properties of $\dot{J}^+(p)$. We now sketch our proposed resolution for the interested reader: $\dot{J}^+(p) \setminus \{p\}$ is a locally achronal, closed, *past* null geodesically ruled topological hypersurface in M.[27] Now, Chruŝciel et al. have shown ([CDGH01, Thm. 2.2], but we refer also to [Min15a]) that any such subsets are better than just locally Lipschitz – they are *semi-convex*. This property implies that they are almost everywhere *twice*-differentiable in the 'Alexandrov' sense (see [CDGH01, Prop. 2.1]) – and they are so on the same set of full measure on which they are once-differentiable. It seems to us that this would yield the desired continuity property of the almost-everywhere–defined derivatives of our curve above, and the same argument as in the C^1 case would then carry through. If this strategy works in the case of a point, we believe that it will also work for a more general future causally complete subset $S \subseteq \mathscr{N}$. We will leave the careful verification of these claims to future work. For a related result, see also [Chr98].

[27] We are grateful to E. Minguzzi for clarifying a point to this effect made tacitly in [CGM16].

We are now ready for our uniqueness theorem, which follows immediately from the previous lemma and which could also be easily strengthened using Sobolev spaces as is done in [BW15].

Theorem 2.5.16. *Let $(M, g, \mathfrak{t})$ be a time-oriented and globally hyperbolic Lorentzian manifold. Let $\mathscr{N}$ be a closed, achronal and null hypersurface whose null generators, when reparametrised as null geodesics entirely contained in $\mathscr{N}$, are future and past inextendible in (M, g). Let P be a differential operator of the form given in Equation (2.30). Finally, let $f \in C^\infty(\mathscr{N})$ have causally complete support. Then, the corresponding solution ϕ_f falling entirely through $\mathscr{N}$ with data f, given by Equation (2.44), is unique among all real-valued functions ϕ on M with the following properties:*

- *ϕ is C^2 on $I^+(\operatorname{supp} f) \cup I^-(\operatorname{supp} f)$;*
- *$\phi\restriction_{I^\pm(\operatorname{supp} f)}$ can be extended to a function defined and C^2 in a neighbourhood of $J^\pm(\operatorname{supp} f)$;*
- *$P\phi = 0$ on $I^+(\operatorname{supp} f) \cup I^-(\operatorname{supp} f)$;*
- *$\phi\restriction_{\mathscr{N}} = f$;*
- $\operatorname{supp} \phi \subseteq J(\operatorname{supp} f)$. □

With the hypotheses and notation of Theorem 2.5.16, let

$$S_{\mathscr{N}} := \{\phi_f \mid f \in C^\infty(\mathscr{N}) \text{ and } \operatorname{supp} f \text{ is causally complete}\} \subseteq C^0(M) \cap \ker P. \quad (2.56)$$

We can use our uniqueness result, which shows that the assignment

$$C^\infty(\mathscr{N}) \to C^0(M) \cap \ker P, \quad \text{given by} \quad f \mapsto \phi_f,$$

is well-defined (i.e. independent of all choices made in defining ϕ_f), to also argue that this assignment is linear, and hence that $S_{\mathscr{N}}$ is a linear subspace of $C^0(M) \cap \ker P$. This will answer **Q3** in the affirmative.

Proposition 2.5.17. *Under the hypotheses of Theorem 2.5.16, for any $\alpha, \beta \in \mathbb{R}$ and any two $f_1, f_2 \in C^\infty(\mathscr{N})$ with causally compact support,*

$$\phi_{\alpha f_1 + \beta f_2} = \alpha \phi_{f_1} + \beta \phi_{f_2}.$$

Proof. With $f := \alpha\phi_{f_1} + \beta\phi_{f_2}$, clearly both ϕ_f and $\alpha\phi_{f_1} + \beta\phi_{f_2}$ restrict to f on $\mathscr{N}$. $\alpha\phi_{f_1} + \beta\phi_{f_2}$ has support contained in $J(\operatorname{supp} f_1) \cup J(\operatorname{supp} f_2) = J(\operatorname{supp} f_1 \cup \operatorname{supp} f_2)$. On the other hand, $\operatorname{supp} f \subseteq \operatorname{supp} f_1 \cup \operatorname{supp} f_2$ and thus $\operatorname{supp} \phi_f \subseteq J(\operatorname{supp} f_1 \cup \operatorname{supp} f_2)$.

Therefore, ϕ_f and $\alpha\phi_{f_1} + \beta\phi_{f_2}$ must be equal by uniqueness in the sense of Theorem 2.5.16. □

2.5.5 Bifurcate Killing horizons and spaces of solutions S_A and S_B

Our original motivation for the work done in Section 2.3 and in this section, which eventually culminated in the general existence and uniqueness theorems just proved for characteristic initial value problems on globally hyperbolic manifolds, was to put on more rigorous grounds some claims and constructions made without proof by the authors of [KW91].

In the presence of a bifurcate Killing horizon structure, we saw in Section 2.2 that two null, closed and achronal hypersurfaces naturally arise, namely the two 'horizons' denoted by $\mathcal{H}_A$ and $\mathcal{H}_B$ there and in [KW91]. We also showed in detail how the assumed completeness of the Killing field translated into the property that both $\mathcal{H}_A$ and $\mathcal{H}_B$ contain all their maximally extended null generators. It follows that, if $\mathscr{M} = (M, g, \mathfrak{t}, \mathfrak{o}, \xi)$ is a spacetime with a bifurcate Killing horizon in the sense of Definition 2.2.11, then $(M, g, \mathfrak{t})$, together with either $\mathcal{H}_A$ or $\mathcal{H}_B$, satisfy all of the hypotheses of our existence and uniqueness theorems concerning characteristic initial value problems for (scalar) normally hyperbolic differential operators, i.e. all of the hypotheses listed in Theorem 2.5.16. Therefore, we may define spaces

$$S_{\mathcal{H}_A}, S_{\mathcal{H}_B} \subseteq C^0(M) \cap \ker P$$

of distributional solutions, falling entirely through $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, according to Equation (2.56). If f in Equation (2.56) is further restricted to have compact support, then we obtain further linear subspaces which we denote by S_A and S_B. It is precisely the existence of these subspaces (and the fact that they indeed consist of distributional solutions!) that the authors of [KW91] claimed without proof. But our analysis in this section has remedied the problem. Therefore, the discussion on this point in [KW91] does not require any modification. This is unlike some other problematic aspects of the analysis in [KW91] which we will analyse in detail in Chapter 4.

2.6 Summary

The first four sections in this chapter introduced mathematical background on, broadly speaking, the following topics: general Lorentzian geometry; the theory of spacetimes with a bifurcate Killing horizon; the geometry of null hypersurfaces and the existence of useful coordinate systems adapted to them (the 'Gaussian null coordinate systems' of Theorem 2.3.11); the well-established theory of wave-like equations on globally hyperbolic Lorentzian manifolds. Sections 2.1 and 2.4 contained no original work. On the other hand, some technical gaps were filled in the available treatments of the topic dealt with in Section 2.2. In particular, we faced head-on the problem of determining under what conditions the two 'horizons', naturally arising if a Killing horizon is present which vanishes on a spacelike submanifold of a given spacetime, are well-behaved hypersurfaces of the spacetime. To the best of our knowledge, detailed proofs as some of the ones given in that section were not previously available. As an aside, we presented a new proof of the constancy of the 'surface gravity' of a bifurcate Killing horizon (Theorem 2.2.7). Section 2.3 was heavily based, in the first instance, on work of Kupeli [Kup87], with a minor addition being given by Proposition 2.3.9. We then reviewed the Gaussian null coordinate systems which are vastly used when dealing with null hypersurfaces. In doing so, we aimed to obtain an existence result which were as sharp as possible. This was because such a result would then be of crucial importance in the arguments given in Section 2.5.

Section 2.5 contains mostly original work, which was however heavily inspired by work carried out by Rendall in [Ren90]. We expect that our final global existence and uniqueness results, summarised in Theorem 2.5.16, can in fact also be obtained by different methods – for example, by globalising some of the results in [Fri75], or by using approaches closer in spirit to the ones adopted in [MzH90] or in [Hör90b]. From the point of view of the interplay with the rest of the work in this thesis, the general results in Section 2.5, when specialised to the context of spacetimes with a bifurcate Killing horizon, allowed us to fully vindicate some important claims made without proof in [KW91]. Since these claims were among the essential ingredients of the entire analysis carried out there, and since our work in Chapters 3 and 4 will build on parts of that analysis, it was important to reassure ourselves that these specific claims held true.

Chapter 3

Quantum no-go results

3.1 Introduction

Thanks to a number of results obtained in the 1990's, it is known that (leaving aside some technicalities) if one quantizes a linear scalar field on a globally hyperbolic spacetime with a one-parameter group of isometries possessing a bifurcate Killing horizon, then there is at most one[1] state which is invariant under those isometries and which is (locally) Hadamard.[2] Furthermore, for some notable cases, such as Kerr and Schwarzschild–de Sitter, it was proved in [KW91] that there is *no* such state.[3] For Kerr, this was a consequence of superradiance; for Schwarzschild–de Sitter, one argument for the no-go result was based on the fact that, should such a state exist, the Hawking temperatures associated with the black hole horizon and the cosmological horizon would be different. Another argument relied on what, in quantum information theory, is now known as monogamy (although this notion had not yet been coined at the time).

[1]In fact, such a uniqueness result was proven in [KW91] under the restriction that the state in question be quasi-free (with vanishing one-point function) [KW91, Haa96, BR97] and with the local Hadamard condition replaced by a certain global Hadamard condition (see next footnote). However, in [Kay93] a general result was obtained which enabled one to drop the quasi-free restriction while, as conjectured in [Kay88, GK89] and proved in [RV96, Rad96, Rad92] on any globally hyperbolic spacetime, locally Hadamard states are necessarily globally Hadamard. See also Footnote 19 and Chapter 4.

[2]A (locally or globally) Hadamard state for a linear quantum field theory is a state whose two-point function has the (local or global) Hadamard property – local Hadamard meaning roughly that its short distance singularity should be the appropriate generalisation to a curved spacetime of the short-distance singularity of the two-point function of the vacuum state and of other physically relevant states in Minkowski space, while the global Hadamard condition on a globally hyperbolic spacetime also rules out the possibility of singularities for spacelike separated pairs of points. For full definitions, see e.g. [KW91] or the recent review [KM14]. See also the important microlocal reformulation of the global Hadamard condition in [Rad96] and see [Mor03] for spacetime dimensions other than $1+3$.

[3]We remark that, as pointed out in [KW91], to prove such a no-go result, it suffices to prove that there is no such quasi-free state, since if there was such a state at all, the quasi-free state with the same two-point function (and zero one-point function) – i.e. the 'liberation' in the sense of [Kay93] – would also be such a state.

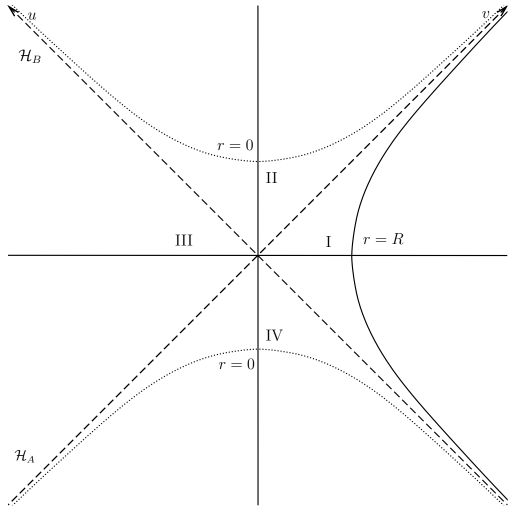

FIGURE 3.1: This is a dual purpose figure. In one interpretation, it represents the Kruskal spacetime bounded by a single box in the right wedge (region I) at $r = R$ (with r the Schwarzschild coordinate and each point representing a two-sphere). In another interpretation, it represents (1+n)-dimensional Minkowski space to the left of a hypersurface (referred to in the text as a 'mirror') at some constant Rindler spatial coordinate r in the right Rindler wedge (in this case each point represents an (n-1)-plane). The dotted lines are only relevant to the Kruskal interpretation, in which case they portray the future and past singularities at $r = 0$.

In this chapter we conjecture, and give heuristic arguments for, a further such non-existence result which concerns a massless or massive linear scalar field on a spacetime which one might think would represent a spherically symmetric maximally extended black hole in equilibrium in a spherical box. Namely, the region of the Kruskal spacetime to the left of a stationary hypersurface at some fixed value R of the Schwarzschild radial coordinate r, represented by the hyperbola in Figure 3.1 (where, as usual, each point represents a two-sphere).[4] I.e. we argue that, completing the specification of the system by imposing (say) Dirichlet boundary conditions at the box, there is no Schwarzschild-isometry invariant Hadamard state on this spacetime (when the notion of 'Hadamard', usually applied to globally-hyperbolic spacetimes, is suitably adapted to the presence of a timelike boundary).

The basic plausible expectations about the space of classical solutions, from which we will argue for this no-go conjecture in the next section, are that, on the one hand,

(a) the reflection at the box in the right wedge will cause solutions which 'fall entirely through' (see Section 3.2) the right A-horizon ($\mathcal{H}_A^R$ in the Penrose diagram, Figure 3.2) to coincide with solutions which 'fall entirely through' the right B-horizon ($\mathcal{H}_B^R$ in Figure 3.2).

[4]Our no-go conjecture for Kruskal in a box applies equally to the part of the globally-hyperbolic region of non-extremal Reissner-Nordström spacetime to the left of a similar stationary hypersurface at fixed Schwarzschild radial coordinate r but, for simplicity we shall only refer to the Kruskal case in the main text.

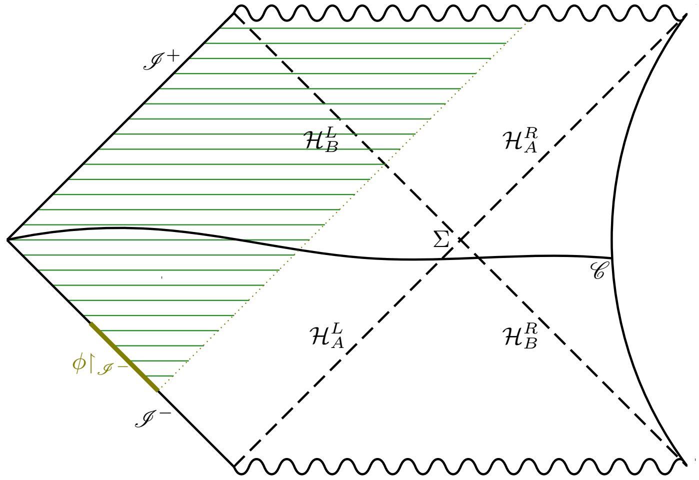

Figure 3.2: Penrose diagram for the region of the Kruskal spacetime bounded by a single box, cf. Figure 3.1. $\mathscr{C}$ is an initial-value surface on which the Cauchy-Dirichlet problem for the Klein–Gordon equation is well-posed. The shaded area represents the support of the solution ϕ discussed in Sections 3.1 and 3.2.

On the other hand,

(b) there exist solutions (one such suffices for our argument) which are non-vanishing on the left B-horizon but which vanish on the entire A-horizon.

The plausibility of Property (b) is particularly easy to see for the massless case since, in fact, any solution, ϕ, with non-zero Cauchy data on $\mathscr{I}^-$ (see the Penrose diagram, Figure 3.2) and zero Cauchy data on $\mathcal{H}_A$ would be expected to have a non-zero value on $\mathcal{H}_B$ expressing the fact that not all the solution would be reflected back out to infinity, but rather, some of it will fall through $\mathcal{H}_B^L$ into the black hole. (Whether or not this property holds obviously doesn't depend on whether or not the spacetime is cut off at a box-wall in the right wedge.) For massless and massive fields, one can rely, instead, e.g. on the existence of wave operators, $\Omega_0^\pm$ and $\Omega_1^\pm$[5] for the scattering theory on exterior Schwarzschild demonstrated in [DK87, DK86] together with the expectation that the S-matrix component $(\Omega_1^+)^*\Omega_0^-$ will not be zero. In fact this is now rigorously established in the massless case in Theorem 10 of [DRSR14].[6]

We remark that if there is also an image box in the left wedge (located at the wedge-reflected set of spacetime points to those occupied by the right-wedge box – below we shall refer to this as the case of two boxes) we expect that there *will* exist an isometry-invariant Hadamard state on the region between the two boxes. Indeed, we expect the

[5] $\Omega_0^\pm$ maps solutions of the Klein–Gordon equation on Minkowski space into solutions on exterior Schwarzschild (identified here with our Kruskal left wedge) which resemble them at late/early times and $\Omega_1^\pm$ maps solutions of the massless 'wave equation' in 1+1 Minkowski space times the bifurcation 2-sphere into solutions on exterior Schwarzschild and (as explained in [DK87]) effectively solves the characteristic initial-value problem for data on the future/past horizon.

[6] We thank Mihalis Dafermos for drawing this to our attention.

latter to be a counterpart to the Hartle–Hawking–Israel state [HH76, Isr76, San15] in maximally extended Kruskal. Thus our no-go conjecture is reliant upon there being just one box rather than two.

Geometrically, this setup appears analogous to Minkowski spacetime (of any dimension) to the left of a hypersurface at some constant Rindler spatial coordinate in the right wedge (see Figure 3.1), i.e. to the left of a uniformly accelerating mirror (assumed to be 'planar' and infinitely extended in the spatial dimensions suppressed in Figure 3.1). Here, Schwarzschild-isometry invariance is replaced by boost invariance. One might therefore think that a similar non-existence result would hold for boost-invariant Hadamard states for Klein–Gordon fields on such spacetimes. And, in the absence of a rigorous proof of our conjecture for Kruskal, it would obviously be of interest if one could more easily give a rigorous proof of the non-existence of boost-invariant Hadamard states for some such Minkowskian system. However, Property (b) above only holds for scalar fields in Minkowski space when those fields are massless and the Minkowski space is 1+1 dimensional. This is because, except in this special case, a solution to the Klein–Gordon equation in Minkowski space (say with compact support on spacelike Cauchy surfaces) which vanishes on a single null plane, vanishes everywhere. This is proven for the case of massless fields and spacetime dimension greater than 2 in e.g. [Wal94, pp. 109–110], whereas in the case of massive fields in any spacetime dimension Ullrich [Ull07, Thm. 3] proved an even more general result using Fourier analytic methods.[7]

In view of the above, and aside from making our above conjecture for the Kruskal case, the main purpose of the present thesis is to prove a rigorous version of such a non-existence result for this latter 1+1 massless system with Dirichlet boundary conditions. Even for this much simpler problem, it will turn out that we have to deal with a number of complications which arise from the well-known special infra-red pathology [Sch63, Wig67, SW70, Kay85, FR87, DM06] of the 1+1 massless Klein–Gordon field as well as with complications due to the presence of a boundary. In fact, even in the absence of boundaries, because of that special infra-red pathology, there are several inequivalent mathematical notions which could be regarded as making the phrase 'boost-invariant Hadamard state' precise for the massless scalar field in 1+1 Minkowski space. What we succeed in doing (with Theorem 3.4.7 in Section 3.4.3) is to prove that, with a particular such notion, when suitably adapted to the presence of a single mirror – namely what we call the 'strongly boost-invariant globally-Hadamard' property of Definition 3.4.6 in Section 3.4.3 – then (in the presence of a single mirror) there is no state which has this property.

[7]However, we point out that we believe that a more geometrically-flavoured argument can also be produced by exploiting the fact that massive fields in Minkowski time exhibit dispersion, i.e. they decay in time according to some well-known estimates.

We believe this no-go theorem deserves to be regarded as a suitable counterpart to the no-go result we conjecture for Kruskal because, as we will also point out in Section 3.4.3, there *does* exist a strongly boost-invariant globally-Hadamard state both in full 1+1 Minkowski space and in the case where there is a second mirror located at the wedge-reflected set of spacetime points to those occupied by the right-wedge mirror (which we shall call the case of two mirrors) – the boost-invariant Hadamard state in the absence of mirrors being a suitably defined version of the usual Minkowski vacuum state, while the state for two mirrors was constructed in [Kay15]. Also, we think that the method of proof of this result should provide useful lessons towards a proof of our conjecture about the Kruskal case. Note that our notion of 'strongly boost-invariant globally-Hadamard' makes precise the notion of 'boost-invariant *global* Hadamard state' since, for reasons we will explain in Section 3.4.2, it is not obvious that a local-to-global result (see Footnote 2) applies in the 1+1 massless case.

Our conjecture in the Kruskal case has an obvious application to understanding the nature of the idealised black holes in boxes which play a basic role in black hole thermodynamics [Haw76, GH93]. A natural question is whether a black hole in equilibrium in a box[8] has a semiclassical description in terms of a fixed Lorentzian classical spacetime together with a Hadamard state of a quantum field defined on it – where both the classical spacetime and the Hadamard state are isometry-invariant. Amongst the various possibilities one can imagine for the background spacetime, and ignoring back reaction, one might consider the following three: (A) the region of Kruskal to the left of a single box as in Figures 3.1 and 3.2; (B) the region of Kruskal between two boxes as in Figure 3.3; (C) the region of exterior Schwarzschild alone to the left of a single box (i.e. the right wedge of any of the figures 3.1, 3.2 or 3.3). An earlier paper [Kay15] argued that both (A) and (B) should be ruled out due to the existence of classical and/or quantum small perturbations such that, as a consequence of reflection at the box, their (renormalized) stress-energy grows arbitrarily large near the future horizon(s) and/or near the bifurcation surface and argued in favour of (C) with the proviso that the region near the horizon be considered to be essentially quantum-gravitational and non-classically describable rather as envisaged in 't Hooft's 'brick wall' model [tH85]. However the arguments against (A) in [Kay15] were less strong than the arguments against (B). Our conjectured no-go theorem, if true, tells us that, on the background (A), no isometry-invariant Hadamard state is possible and this reinforces our reasons for rejecting (A).

It is also of interest to compare our no-go result for the massless scalar field in 1+1 Minkowski with claims made in the literature (see e.g. [FD76, DF77, BD84]) concerning

[8]Here we leave aside the issue that a Schwarzschild black hole in equilibrium in a box is believed to be thermodynamically unstable [Haw76].

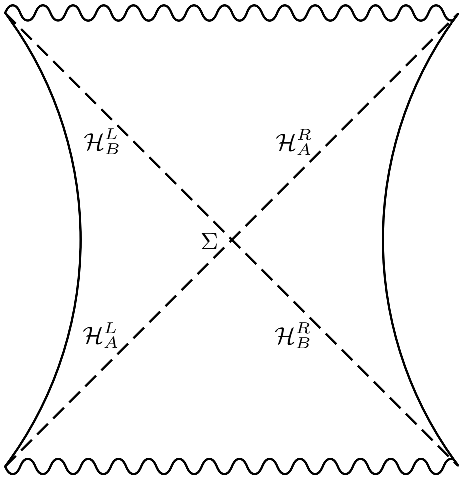

FIGURE 3.3: Penrose diagram for the region of the Kruskal spacetime bounded by two boxes, possibility (B) in Section 3.1. We conjecture that a 'Hartle–Hawking–Israel–like' state exists for the Klein–Gordon field on this spacetime when Dirichlet conditions are imposed at the boundary.

radiation by accelerating mirrors in 1+1 dimensions. As pointed out in that work, a mirror which starts out inertial – with the state of the field the initial vacuum state – and later undergoes uniform acceleration doesn't radiate during the period of uniform acceleration. This might seem to suggest that there would be a quantum state of the field such that an eternally accelerating mirror wouldn't radiate at all and that might, in its turn, seem to suggest that there would exist a boost-invariant Hadamard quantum state. And one might think that there would in fact exist a strongly boost-invariant globally-Hadamard state in the sense of the present thesis. But we prove that there isn't one; for there to be such a boost-invariant Hadamard state, it would seem to be required for there to be a symmetrically placed uniformly decelerating image mirror in the left wedge.

3.2 Basic idea of our argument for the no-go theorem

We next wish to explain the basic idea behind both our no-go conjecture for (massive or massless) Klein–Gordon on Kruskal and our proof of our analogous no-go result for the massless 1+1 Minkowski one-mirror system. In Kruskal we take our equation to be

$$P\phi = (\Box_g + m^2)\phi = 0 \tag{3.1}$$

where m is a non-negative mass. (Note that the often considered ξR term will anyway vanish in Kruskal.) In our 1+1 Minkowskian theorem we insist that m be zero.

In both cases, we rely on the well-posedness of the Cauchy problem for (3.1) when supplemented by Dirichlet boundary conditions at the box/mirror. Of course, neither the region of Kruskal to the left of our box, nor the region of 1+1 Minkowski space to the left of our mirror are globally hyperbolic and thus neither have Cauchy surfaces in the strict sense. However, with our boundary conditions on the box/mirror, one expects the Cauchy problem to be well posed, at least in the sense of uniqueness, for data on initial-value surfaces which are the restrictions, to the region to the left of the box/mirror, of Cauchy surfaces for the whole of Kruskal/Minkowski. Indeed, this can easily be verified in the 1+1 Minkowski case; for the Kruskal case we expect a suitable extension of known results on the mixed Cauchy–Dirichlet problem (see e.g. Theorem 24.1.1 in [Hör94] or the monograph [GV96]) to apply. And it will still to be possible to talk, in each case, about the space S of smooth (real-valued) solutions of this mixed Cauchy–Dirichlet problem whose restriction to all such initial-value surfaces[9] has compact support, along the lines of that discussed in [KW91]. And this space will be equipped with a manifestly antisymmetric bilinear form σ defined, in terms of an arbitrary (possibly partially null) smooth initial-value surface $\mathscr{C}$, by

$$\sigma(\phi_1, \phi_2) := \int_{\mathscr{C}} n_a j^a[\phi_1, \phi_2] \, \mathrm{vol}_{\mathscr{C}}, \tag{3.2}$$

where $j^a[\phi_1, \phi_2] := \phi_1 \nabla^a \phi_2 - \phi_2 \nabla^a \phi_1$, $\mathscr{C}$ is given the induced orientation as the boundary of $J^-(\mathscr{C})$,[10] and the forms n and $\mathrm{vol}_{\mathscr{C}}$ are such that, on $\mathscr{C}$, $n \wedge \mathrm{vol}_{\mathscr{C}}$ equals the volume form vol_g induced by the spacetime metric. The independence of the right-hand side of Equation (3.2) from the initial-value surface $\mathscr{C}$ is a consequence, using Gauss' theorem, of the fact that $\nabla_a j^a[\phi_1, \phi_2] = 0$ whenever ϕ_1 and ϕ_2 are solutions to Equation (3.1), together with the fact that, due to the Dirichlet boundary conditions, no boundary terms arise from integrating along the spacetime boundary. One expects that, once a full characterisation for the allowed initial data for solutions in S is available, it will be possible to show that σ is in fact non-degenerate on S, and therefore a symplectic form.

Similarly to in [KW91] – and proceeding, in the Kruskal-like variant, under the same fiction explained in the Note Added in Proof at the end of [KW91] (see the discussion at the end of this section) – an important role will be played by 'subspaces', S_A and S_B, of S which we assume are nontrivial in the case of Kruskal, and which in both cases consist of solutions of Equation (3.1) satisfying the Dirichlet boundary condition on ∂M and which 'fall entirely through' the A- and B-horizons $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, in the sense of our Definition 2.5.10 when adapted to the context of a spacetime with a smooth

[9] These initial-value surfaces should be understood to contain the relevant boundary points and therefore not as being entirely contained in the interior of the spacetime.

[10] I.e. the boundary orientation for which Stokes' Theorem applies.

boundary (see Section 5.1) and to the fact that we are imposing (Dirichlet) boundary conditions. More precisely:

Definition 3.2.1. In both the Kruskal-like and Minkowski-like spacetimes-with-boundaries which we are considering, ϕ will belong to S_A if it is at least continuous, it is a distributional solution of Equation (3.1) in $\operatorname{Int} M$ and even a C^2 one on an open neighbourhood of ∂M, it vanishes identically on ∂M, $\phi\restriction_{\mathcal{H}_A}$ is a smooth function on $\mathcal{H}_A$ whose support $K := \operatorname{supp}(\phi\restriction_{\mathcal{H}_A})$ is compact, and $\operatorname{supp}\phi \subseteq J(K) = J^+(K) \cup J^-(K)$. We define S_B analogously.

For a massless scalar field in 1+1 Minkowski space without any mirrors, S_B would consist of right-moving solutions and S_A of left-moving solutions. When we have our mirror in the right wedge, we can also fully characterise both spaces: S_B consists of solutions which are right-moving to the causal past of the B-horizon , and S_A consists of solutions which are left-moving to the causal future of the A-horizon. This is explained in more detail in Section 3.4.1. On the other hand, we do not yet have a full characterisation of S_A and S_B, or even a rigorous proof of their non-triviality, in the Kruskal-like setting for any value of the mass. Nonetheless, we proceed under the assumption that, in the Kruskal-like case, S_A and S_B are large enough for our purposes. We will also need to define appropriate subspaces of S_A and S_B in both settings, and again proceed under the assumption that they are large enough for our purposes in the Kruskal-like setting.

Definition 3.2.2. In both cases, S_A^R is to denote the subspace of S_A consisting of solutions whose restrictions to $\mathcal{H}_A$ are compactly supported to the causal future of, and strictly away from, the bifurcation surface. The 'R' superscript stands for 'right' as this is the portion of $\mathcal{H}_A$ to the right of the bifurcation surface in Figure 3.1. We define S_A^L, S_B^R and S_B^L similarly with obvious changes.

In Appendix A, we will recall the general theory of the quantization of linear Bose systems via the so-called Weyl-algebra approach. In particular, we will review the standard definitions for the notions of *state*, *quasifree state* and *one-particle structures*. In Section 3.3, we will recall how this theory is applied to the case of Klein–Gordon fields on general globally hyperbolic spacetimes, where the class of Hadamard states plays a special role, and we will sketch a strategy for adapting this theory to situations with timelike boundaries so as to properly define the notion of 'Hadamard state' and, thereby, to be able to formulate in a precise way our conjecture that there is no isometry invariant Hadamard state on Kruskal in a box. Then Section 3.4 will show how to implement this strategy for massless fields on 1+1 Minkowski with a mirror in a way which also copes with the special infra-red pathology, thereby enabling us both to properly formulate and prove our no-go theorem. For us to explain the basic idea behind our conjecture and theorem in the present section, however, all that we shall rely on are the following two facts:

- First, just as in the globally hyperbolic case mentioned in Footnote 3, to show that there is no isometry-invariant Hadamard state, it suffices to show there is no isometry-invariant *quasi-free* Hadamard state (with zero one-point function), see Appendix A.
- Second, as explained in Appendix A, to every quasi-free state of the theory there corresponds a *one-particle structure*, $(K, \mathscr{H})$. That is a, Hilbert space (the *one-particle Hilbert space*), $\mathscr{H}$, and a real-linear map, $K : S \to \mathscr{H}$, such that $KS+iKS$ is dense in $\mathscr{H}$, which is *symplectic* in the sense that

$$2\mathrm{Im}\,\langle K\phi_1 \mid K\phi_2\rangle = \sigma(\phi_1, \phi_2) \tag{3.3}$$

  for all pairs of classical solutions, $\phi_1, \phi_2 \in S$.

Furthermore, and similarly to Kruskal without a box or (1+3)-dimensional Minkowski without a mirror, we expect that the existence of an isometry-invariant Hadamard state for Kruskal with our box implies, by similar arguments to those given in [KW91] the following explicit formula for $\left\langle K\phi_B^1 \mid K\phi_B^2\right\rangle$ for any $\phi_B^1, \phi_B^2 \in S_B$:

$$\left\langle K\phi_B^1 \mid K\phi_B^2\right\rangle = -\frac{1}{\pi}\lim_{\varepsilon\to 0^+}\int \frac{f_1(u_1,s)f_2(u_2,s)}{(u_1-u_2-i\varepsilon)^2}\,\mathrm{d}u_1\mathrm{d}u_2\mathrm{d}^2s, \tag{3.4}$$

where f_1 is the restriction of ϕ_B^1 and f_2 the restriction of ϕ_B^2 to the B-horizon, and this is coordinatized in the usual way by affine parameter, u,[11] and the usual set of 'angular' variables, denoted by s, and the integration can be thought of as over two copies of the real line and one copy of the bifurcation sphere.

For our massless scalar field in 1+1 Minkowski with a mirror, it turns out that the existence of an isometry-invariant state which is Hadamard in the precise sense we will define (i.e. the 'strongly boost-invariant globally-Hadamard' property of Definition 3.4.6 in Section 3.4.3) entails a similar formula, with the dependence on s and the integration over s removed. And of course there will be a similar formula, for ϕ_A^1 and ϕ_A^2 and the A-horizon.

As discussed in [KW91] (cf. Equation (1.1) there; we refer also to Observation 6.1 and Proposition 7.2 in [DK87]), Equation (3.4) tells us that the restriction of the two-point function for the u derivative of the field to the B-horizon can be identified (up to a trivial dependence on s) with the restriction of the two-point function for the u derivative of a

[11] Aside from having the opposite signature convention to [KW91], we differ from [KW91] by denoting affine parameter on our horizons by u and v, rather than U and V.

free massless real scalar field in 1+1 Minkowski space (without a mirror) to the null line $t = -x$, where u is now identified with $t - x$, and where t and x are the usual Minkowski coordinates. In view of this (or directly from the formula) one can conclude (see again [KW91]) the following crucial facts[12]

(A) KS_A and KS_B are dense in complex-linear subspaces $\mathscr{H}_A$ and $\mathscr{H}_B$ of $\mathscr{H}$ (respectively). As explained in Appendix A of [KW91], and reproduced in Appendix A to the present thesis as Proposition A.0.4, this is equivalent to the fact that the state restricted to fields 'symplectically smeared' with solutions in either S_A or S_B is a pure state. In the special case of 1+1 Minkowski (without a mirror) it corresponds to the fact that the Minkowski vacuum is a pure state when restricted to either the left or right-moving sector.

(B) $KS_A^R + iKS_A^R$ is dense in $\mathscr{H}_A$ and $KS_B^R + iKS_B^R$ is dense in $\mathscr{H}_B$. This corresponds to the fact that the (massless) 1+1 Minkowski vacuum, restricted to sums of products of (derivatives of) fields restricted to a single null line has the Reeh–Schlieder property [SW00] for fields localised on a half null-line. Cf. Proposition A.0.5 in Appendix A.

We are now in a position to explain the basic idea behind both our hoped-for proof of our no-go conjecture for Kruskal in a box and our proof of our no-go theorem for our massless field in 1+1 Minkowski with a mirror.

First we point out that, for the 1+1 Minkowski case, the 'basic plausible expectations about the space of classical solutions' discussed in Section 3.1 may be reformulated in terms of our subspaces of solutions as follows:

(a) $S_A^R = S_B^R$;

(b) There exists a $\phi \in S$ such that $\sigma(\phi, \phi_B^L) \neq 0$ for some $\phi_B^L \in S_B^L$, but for which $\sigma(\phi, \phi_A) = 0$ for all $\phi_A \in S_A$.

Combining the (purely classical) statements in (a) and (b) with (A) and (B) above quickly leads to a contradiction, as we will now explain. By the first part of (b) and Equation (3.3), $K\phi$ cannot be orthogonal (with orthogonality here and throughout this this paragraph meant in the sense of the Hilbert space $\mathscr{H}$) to KS_B^L and hence, *a fortiori* it cannot be orthogonal to KS_B – so, by (A), it cannot be orthogonal to $\mathscr{H}_B$. On the other hand, Equation (3.3) and the last part of (b), together with (A), imply that $K\phi$ is orthogonal to $\mathscr{H}_A$. To see this, we will use the following general observation: If

[12] Actually in our proof of our no-go theorem, i.e. of Theorem 3.4.7 in Section 3.4.3, facts (A) and (B) about the one-particle structure $(K, \mathscr{H})$ are arrived at by directly relating it to the one-particle structure $(K_{\mathbb{M}}, \mathscr{H}_{\mathbb{M}})$ associated to the vacuum state, $\omega_{\mathbb{M}}$, on the 'physical' Weyl algebra for the massless wave equation in (1+1)-Minkowski space by a somewhat different version of the argument which doesn't (need to) refer to the formula (3.4).

$\mathscr{H}$ is a complex Hilbert space, and $\mathscr{K} \subseteq \mathscr{H}$ is a real-linear subspace whose closure is *complex*-linear, then, for any $\Phi \in \mathscr{H}$, $\langle \Phi \mid \mathscr{K} \rangle = \{0\}$ if and only if $\mathrm{Im}\,\langle \Phi \mid \mathscr{K} \rangle = \{0\}$ if and only if $\mathrm{Re}\,\langle \Phi \mid \mathscr{K} \rangle = \{0\}$. [Proof of first 'if': Suppose that $\mathrm{Im}\,\langle \Phi \mid \mathscr{K} \rangle = \{0\}$. Note that $\mathrm{Re}\,\langle \Phi \mid \mathscr{K} \rangle = \mathrm{Im}\,\langle \Phi \mid i\mathscr{K} \rangle$. Under the assumptions on $\mathscr{K}$, $i\mathscr{K} \subseteq \overline{\mathscr{K}}$, whereupon a simple limit argument shows that $\mathrm{Im}\,\langle \Phi \mid \mathscr{K} \rangle = \{0\} \implies \mathrm{Im}\,\langle \Phi \mid i\mathscr{K} \rangle = \{0\}$ and we are done. The proof of the second 'if' is analogous.] By (B), to say that $K\phi \perp KS_A$ is tantamount to saying that it is orthogonal to $KS_A^R + iKS_A^R$. But, by (a), this is the same thing as saying that it is orthogonal to $KS_B^R + iKS_B^R$, which, by (B), has the same closure as KS_B, namely $\mathscr{H}_B$. Thus, on the assumption that there exists a stationary Hadamard state, $K\phi$ is both not orthogonal to $\mathscr{H}_B$ and orthogonal to $\mathscr{H}_B$ – a contradiction.

For Kruskal in a box, Property (a) above cannot strictly hold since we would expect a solution which falls entirely through the right B-horizon to have a restriction to the right A-horizon which fails to be supported away from the bifurcation point and moreover we would expect it to fail to be compactly supported, but rather to have a tail at large v. However, we conjecture that the closure in $\mathscr{H}$ of KS_A^R will equal the closure in $\mathscr{H}$ of KS_B^R (or rather an appropriate substitute for this statement will hold when one removes the fiction we referred to above and discuss further below). It is easy to see that this 'closure conjecture' would immediately lead to the same contradiction.

The fiction we referred to above concerns an error in the original version of [KW91] which we have also (knowingly) made above. As was pointed out in the Note Added in Proof in that paper, the notion of 'C^∞ solutions which fall entirely through one of the horizons', as in the apparent 'definitions' of S_A etc. in that paper and above in the Kruskal case, is problematic since a solution which actually falls entirely through one of the horizons in the sense explained above cannot be C^∞ – smoothness failing when one crosses from one side of the horizon to the other. The Note Added in Proof of [KW91] showed how one can repair this error while maintaining the spirit of the basic arguments there by working with a certain class of solutions (which are everywhere C^2) and end up with rigorous results with essentially the same physical content as those originally announced. In particular the no-go results in that paper continue to hold with thus-corrected arguments. A new, improved way to deal with some of the technical issues in the Note Added in Proof in [KW91] will be described in Chapter 4 of this thesis.

Clearly, in the case of Kruskal, what we have written above, while we find it highly plausible, falls considerably short of being a rigorously stated theorem and proof. To have a rigorously stated theorem one would need to show that the expectations mentioned in Section 3.3.2 below hold so that the strategy we sketch there for defining what is meant by a Hadamard state can be implemented. And then to turn the above-explained idea for a proof into a rigorous proof one would need to remove the above fiction, presumably with similar methods to those introduced in the Note Added in Proof in [KW91], prove the above 'closure conjecture' or some effective replacement for it, and justify in detail the various statements made above which were described as 'expectations'. As we anticipated in the Introduction, in the absence of all that, what we can and do provide, in Section 3.4, is a rigorous formulation and proof of our no-go result for a massless field in 1+1 Minkowski with a mirror.

3.3 Quantization of Klein–Gordon quantum fields

3.3.1 Globally hyperbolic case

Let $(M, g, \mathfrak{t})$ be a time-oriented, globally hyperbolic spacetime of dimension $1 + n$. We adopt the 'mostly minus' signature convention for the metric, as we did throughout Chapter 2. In Section 2.4, we recalled that the Klein–Gordon operator in Equation (3.1) is Green hyperbolic (actually, even normally hyperbolic) and self-adjoint. We will denote, as we did there, by S the space of 'regular' real-valued classical solutions to the Klein–Gordon equation, i.e. of smooth and spatially compact functions in $\ker P$.

As we also saw in greater generality in Section 2.4, S is naturally equipped with a linear symplectic structure. Explicitly, the symplectic product of any two $\phi_1, \phi_2 \in S$ is known to be given by Equation (3.2), where $\mathscr{C}$ is any smooth Cauchy surface. Equivalently, it is given by the 'covariant' expression

$$\sigma(\phi_1, \phi_2) = \int_M F_1 \phi_2 \, \mathrm{d}\mu_g = \int_M F_1 (E F_2) \, \mathrm{d}\mu_g =: E(F_1, F_2), \tag{3.5}$$

where $E \coloneqq E^- - E^+ : C_0^\infty(M) \to S$ is the causal propagator of P, and $F_1, F_2 \in C_0^\infty(M)$ are such that $EF_1 = \phi_1$ and $EF_2 = \phi_2$.

The *Weyl algebra* recipe for quantization of general linear systems outlined in Appendix A can now be straightforwardly applied to (S, σ), thus yielding a Weyl algebra of canonical commutation relations $\mathscr{A} = \mathscr{W}(S, \sigma)$. In view of the existence of the causal propagator E relating test functions to solutions, if ω is a C^2 state on $\mathscr{A}$, then its two-point function λ_2 (see Appendix A) induces a bidistribution[13] on M defined for all test functions F_1, F_2 by

$$\Lambda(F_1, F_2) = \lambda_2(EF_1, EF_2). \tag{3.6}$$

[13]Henceforth, for a manifold (without boundary) N, we use the word 'bidistribution on N' to simply indicate a bilinear functional $C_0^\infty(N) \times C_0^\infty(N) \to \mathbb{C}$, without any continuity requirements.

We will henceforth refer to λ_2 as the 'symplectically smeared two-point function' and to Λ as the 'spacetime smeared two-point function'. In view of the general properties of C^2 states listed in Appendix A, of the sequence (2.27) and of Equation (3.5), Λ will satisfy for all $F_1, F_2 \in C_0^\infty(M)$:

1. (*Commutation relations*) $2i\,\mathrm{Im}[\Lambda(F_1,F_2)] = \Lambda(F_1,F_2) - \Lambda(F_2,F_1)] = iE(F_1,F_2)$;
2. (*Positivity*) ReΛ has analogous symmetry and positivity properties to (i)–(ii) in Appendix A (with σ, Φ_1, Φ_2 replaced by E, F_1, F_2 respectively);
3. (*Distributional bisolution property*) $\Lambda(PF_1, F_2) = \Lambda(F_1, PF_2) = 0$.

For a state on $\mathscr{A}$ to be physically relevant, of course, not only must its spacetime smeared two-point function, Equation (3.6), exist, but it must also satisfy the (local or global) Hadamard condition. For general globally hyperbolic spacetimes, we refer to the discussion and references in Footnote 2. In the present thesis, the only case we will discuss in detail is the (1+1)-dimensional massless case, the correct formulation of which will, in fact, be the focus of the next section.

3.3.2 Case of spacetimes with timelike boundaries

We would next like to sketch how we expect the quantization procedure for Klein–Gordon fields outlined above could be adapted to the case of 'spacetimes with boundary' (M, g), where M is now a manifold with boundary whose boundary is timelike and $(\mathrm{Int}\,M, g\restriction_{\mathrm{Int}\,M})$ – where $\mathrm{Int}\,M$ denotes the interior of M – is extendible to a globally hyperbolic spacetime. This class of course includes our Kruskal-in-a-box or Minkowski-with-a-mirror spacetimes.

First, we expect that methods akin to those in [Hör94, GV96] will show that, with the addition of suitable homogeneous boundary conditions on the timelike boundary, the Cauchy problem is well-posed for suitable initial data on suitable initial-value surfaces, as already discussed at the start of Section 3.2 for the case of Dirichlet boundary conditions. In particular, such suitable initial data, when smooth and of compact support (where it is to be understood that the support could include points on the timelike boundary), should be in one-to-one correspondence with smooth spatially compact[14] solutions to this mixed problem, and (once the class of 'suitable' initial-value surfaces has been precisely identified) these should in turn be equivalently characterised as being the smooth solutions whose restriction to all suitable initial-value surfaces has compact support. Defining S as the space of spatially compact smooth solutions to this mixed

[14]Just as in the globally hyperbolic case, a spatially compact function ϕ on M is one such that $\mathrm{supp}\,\phi \subseteq J(K)$ for a compact set K, however in this case we allow K to contain points on the timelike boundary.

problem, we then expect, as discussed in Section 3.2, that Equation (3.2) will define a symplectic form σ on S.

Furthermore, we expect that one will be able to construct retarded and advanced Green's operators $E^{\pm}$ which, in addition to satisfying the same requirements as the analogous objects in the globally hyperbolic case – listed as (i)–(ii) in the previous section – are such that $E^{\pm}F\restriction_{\partial M}$ satisfies the given boundary conditions. The domain of $E^{\pm}$ here should at least include $F \in C_0^{\infty}(\mathrm{Int}\, M)$. In the next section we will explicitly construct such objects in the case of the massless wave equation in the region of (1+1)-dimensional Minkowski spacetime to the left of a uniformly accelerating mirror. As we will observe in that case, in general the analogous sequence to (2.27) will no longer be exact since the kernel of $E = E^{-} - E^{+}$ will be strictly larger than the image of P. Furthermore, both in that case and in the general case one doesn't expect that E will be onto S.[15]

Assuming that the expectations in the previous paragraphs are fulfilled, we propose that a state on the Weyl algebra $\mathscr{W}(S, \sigma)$ be called Hadamard if its symplectically smeared two-point function exists and if its spacetime smeared two-point function, defined at least on $C_0^{\infty}(\mathrm{Int}\, M) \times C_0^{\infty}(\mathrm{Int}\, M)$ by Equation (3.6), satisfies the following condition:

Definition 3.3.1. A bidistribution on $\mathrm{Int}\, M$ will be said to be globally Hadamard if, for any causally convex open subset $\mathcal{O}$ of $\mathrm{Int}\, M$ which, when equipped with the restriction of the metric to $\mathrm{Int}\, M$, is a globally hyperbolic spacetime in its own right, the restriction of Λ to smearings with test functions supported inside $\mathcal{O}$ is globally Hadamard in the standard sense.

Here we recall that a subset U of a spacetime (N, g) is called *causally convex* if, whenever two points $x, y \in U$ can be connected by a causal curve γ in N, then the portion of γ between x and y is entirely contained in U. Notice that, if $\mathcal{O}$ is a causally convex globally hyperbolic subset of $\mathrm{Int}\, M$, then denoting by $E^{\pm}_{\mathcal{O}} : C_0^{\infty}(\mathcal{O}) \to C^{\infty}(\mathcal{O})$ the unique retarded/advanced Green operators for the Klein–Gordon equation on $\mathcal{O}$, it is easy to verify that, for all $F \in C_0^{\infty}(\mathcal{O})$, we will have

$$[E^{\pm}F]\restriction_{\mathcal{O}} = E^{\pm}_{\mathcal{O}}F. \tag{3.7}$$

Indeed, that this will be the case follows since, as it is easy to check, $E^{\pm}$ followed by restriction to $\mathcal{O}$ will have, as an operator on $C_0^{\infty}(\mathcal{O})$, the support properties and left/right inverse properties which uniquely determine the retarded/advanced Green operators on $\mathcal{O}$.

[15] It is an interesting open question (as far as we know) – again both in the general case and in the (1+1)-dimensional example we will study – whether the domains of $E^{\pm}$ can be suitably extended in such a way that the resulting advanced-minus-retarded propagator is onto S.

The above proposal would seem to fit nicely with the paradigm of locally covariant (quantum) field theory proposed by Brunetti, Fredenhagen and Verch [BFV03] and indeed allow an extension of that paradigm to include spacetimes with (timelike) boundaries. Physically, since a spacetime boundary can only be detected by sending a signal to it and receiving one in return, our requirement corresponds to saying that, if we localize the quantum state by only performing measurements within globally hyperbolic regions $\mathcal{O}$ which do not 'causally intercommunicate' with the boundary – i.e. such that there are no future-directed piecewise smooth causal curves which begin in $\mathcal{O}$, hit the boundary and then return to $\mathcal{O}$ – we should not be able to tell whether our universe possesses a real boundary, or whether we are witnessing an 'unusual' state on a different, unbounded spacetime. A similar ideology was already contained in [Kay79], where it was pointed out that such a view is necessary in order to clarify the conceptual issues underlying the Casimir effect. It also appeared in [FOP07] in the context of the investigation of quantum energy conditions for spacetimes with boundaries.

3.4 No-go result for massless fields in 1+1-dimensions with a mirror

3.4.1 Classical theory

In this section we consider in detail the classical theory of a massless real scalar field on the spacetime with boundary, (M, η), consisting of the portion M of (1+1)-dimensional Minkowski spacetime 'to the left of' (and including) the worldline of a point-like mirror on a timelike trajectory of uniform and eternal acceleration. Without loss of generality we assume that the Minkowskian pseudo-norm of the 2-acceleration is always equal to -1 (clearly our no-go result does not depend on the numerical value of this quantity). Picking a global inertial frame (t, x) such that, when the proper time τ along the mirror's worldline equals 0, the mirror is located at $(t = 0, x = 1)$ and $\mathrm{d}t/\mathrm{d}\tau|_{\tau=0} = 1$, we represent (M, η) by $M = \mathbb{R}^2 \setminus \{(t, x) \mid x^2 - t^2 > 1,\ x > 0\}$ and $\eta = \mathrm{d}t^2 - \mathrm{d}x^2$. The manifold M is depicted in Figure 3.1, with ($R = 1$ and) the vertical (respectively horizontal) axis representing the t-axis (respectively x-axis).

As already pointed out, this spacetime is not globally hyperbolic due to the presence of the timelike boundary given by the mirror's trajectory. It possesses a one-parameter group β_τ of isometries given by the flow of the Killing vector field $k = x\partial/\partial t + t\partial/\partial x$[16]

[16]Explicitly, in global inertial coordinates, $\beta_\tau(t, x) = (\cosh(\tau)t + \sinh(\tau)x, \sinh(\tau)t + \cosh(\tau)x)$ or, in terms of the null coordinates (u, v) introduced below, $\beta_\tau(u, v) = (e^{-\tau}u, e^{\tau}v)$

describing homogeneous Lorentz boosts in the x-direction. k has a bifurcate Killing horizon given by $\mathcal{H}_A \cup \mathcal{H}_B$, where $\mathcal{H}_A = \{(t,x) \mid t = x\}$ and $\mathcal{H}_B = \{(t,x) \mid t = -x\}$.

We immediately note that any real-valued, smooth solution ϕ on M to

$$\Box\phi = 0, \qquad \phi\restriction_{\partial M} = 0, \tag{3.8}$$

can be written globally as a sum $\phi(t,x) = f(t-x) + g(t+x)$ for two smooth functions f and g with $g(v) = -f(-1/v)$ for all $v > 0$. This can be checked e.g. by writing the above equation in the null coordinates $u(t,x) = t - x$ and $v(t,x) = t + x$. It is also easy to check that for any such solution ϕ which, in addition, has spatially compact support (see Section 3.3), the functions f and g must have the additional property that there exist u_0 and v_0 such that, for some $a \in \mathbb{R}$, $f(u) = a \; \forall \; u \geq u_0$ and $g(v) = -a \; \forall \; v \leq v_0$. Thus we have complete knowledge of the vector space S of spatially compact, smooth (and real-valued) solutions discussed in Section 3.3.2. And, again as envisaged in that section and in Section 3.2, Equation (3.2) defines a manifestly antisymmetric bilinear form $\sigma : S \times S \to \mathbb{R}$, independent of the initial-value surface $\mathscr{C}$ as explained in Section 3.2. Since it is easy to check that the Cauchy-Dirichlet problem is well-posed (in the sense of both existence and uniqueness) for initial data of compact support in the interior of the particular initial-value surface $\mathscr{C} = \{(t,x) \mid t = 0\} \cap M$, one could prove the non-degeneracy of σ directly by picking, for any $\phi_1 \in S$, which will have some initial data $(\varphi_1, \pi_1) \in C_0^\infty(\mathscr{C}) \oplus C_0^\infty(\mathscr{C})$,[17] ϕ_2 to be the solution with initial data $(\varphi_2, \pi_2) \in C_0^\infty(\mathrm{Int}\,\mathscr{C}) \oplus C_0^\infty(\mathrm{Int}\,\mathscr{C})$ where (φ_2, π_2) approximate $(-\pi_1, \varphi_1)$ (respectively) 'sufficiently well' for $\sigma(\phi_1, \phi_2)$ to be greater than 0. This can always be done by picking $\varphi_2 = -\psi\pi_1$ and $\pi_2 = \psi\varphi_1$ where $\psi \in C_0^\infty(\mathrm{Int}\,\mathscr{C}) \subset C_0^\infty(\mathscr{C})$ is such that $0 < \psi < 1$ and $\psi = 1$ everywhere but on a small enough neighbourhood of the boundary point $(t = 0, x = 1)$ of $\mathscr{C}$. Indeed, we expect a generalisation of this strategy to apply to the more general setup described in Section 3.3.2. We will also provide another, independent, proof of the non-degeneracy of σ later in this section.

Thus we have endowed S with the structure of a symplectic vector space (S, σ). A simple calculation, which e.g. starts with the expression for σ in terms of the $t = 0$ initial-value surface mentioned above and then involves a change of variables, shows that, for any

[17] Note that, since $\mathscr{C}$ is a manifold with boundary, functions in $C_0^\infty(\mathscr{C})$ – which are by definition smooth functions with compact support on $\mathscr{C}$ – need not be supported away from the boundary; indeed, they needn't even vanish *at* the boundary (although for this specific choice of $\mathscr{C}$, both pieces of Cauchy data will have to vanish at the boundary because of the Dirichlet boundary condition).

$\phi_1, \phi_2 \in S$,

$$\sigma(\phi_1, \phi_2) = 2\int_{-\infty}^{+\infty} f_1(u) f_2'(u)\,\mathrm{d}u + 2\int_{-\infty}^{0} g_1(v) g_2'(v)\,\mathrm{d}v \tag{3.9}$$

$$= 2\int_{0}^{+\infty} f_1(u) f_2'(u)\,\mathrm{d}u + 2\int_{-\infty}^{+\infty} g_1(v) g_2'(v)\,\mathrm{d}v, \tag{3.10}$$

where, f_1, g_1, f_2, g_2 are any smooth functions such that $\phi_1(t,x) = f_1(t-x) + g_1(t+x)$ and $\phi_2(t,x) = f_2(t-x) + g_2(t+x)$. These explicit expressions will be important in the next paragraph.

Let S_A and S_B denote the linear subspaces of S consisting of those solutions which 'fall entirely through' $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively. A geometric definition of these was already given in the third paragraph of Section 3.2. However, a more explicit characterisation is also available here: $\phi \in S_B$ (respectively $\phi \in S_A$) if and only if $\phi(t,x) = f(t-x)+g(t+x)$ with the 'right mover' f belonging to $C_0^\infty(\mathbb{R})$ and the 'left mover', $g(v)$, being equal to zero for all $v \leq 0$, and to $-f(-1/v)$ for all $v > 0$ (respectively the 'left mover' g belonging to $C_0^\infty(\mathbb{R})$ and the 'right mover', $f(u)$, being equal to zero for all $u \geq 0$, and to $-g(-1/u)$ for all $u < 0$). Thus, solutions in S_B (respectively S_A) are uniquely determined by their restriction to $\mathcal{H}_B$ (respectively $\mathcal{H}_A$). And indeed, the initial value problem is well-posed on Cauchy surfaces which include portions of $\mathcal{H}_B$ (respectively $\mathcal{H}_A$), for data supported on those portions. For any pair ϕ_1, ϕ_2 of S_B-solutions (respectively S_A-solutions), the second (respectively first) summand on the right-hand side of Equation (3.9) (respectively Equation (3.10)) vanishes, and thus $\sigma(\phi_1, \phi_2)$ can be interpreted as twice the integral along $\mathcal{H}_B$ (respectively $\mathcal{H}_A$) of $\phi_1 \partial_u \phi_2$ (respectively $\phi_1 \partial_v \phi_2$). Moreover, let $(S_{\mathbb{M}}, \sigma_{\mathbb{M}})$ denote the symplectic vector space of spatially compact, smooth, real-valued solutions to the massless wave equation on $(\mathbb{R}^2, \eta)$, and let $S_{\text{r-mov}}$ and $S_{\text{l-mov}}$ denote the vector subspaces of $S_{\mathbb{M}}$ consisting of right-moving and left moving (respectively) solutions. Then, as is well known (or easy to show), $(S_{\text{r-mov}}, \sigma_{\mathbb{M}})$ and $(S_{\text{l-mov}}, \sigma_{\mathbb{M}})$[18] are symplectic vector spaces in their own right and one has the following important result, whose proof is immediate.

Proposition 3.4.1. *The map $T_B : S_B \to S_{\text{r-mov}}$, defined by sending $\phi \in S_B$ to the unique Minkowski-space right-moving solution with the same data as ϕ on $\mathcal{H}_B$, is a presymplectic isomorphism between (S_B, σ) and $(S_{\text{r-mov}}, \sigma_{\mathbb{M}})$. Thus in particular (S_B, σ) is a symplectic space and the map is a symplectic isomorphism. (And similarly, with B replaced by A and* r-mov *replaced by* l-mov*.)*

[18] Throughout the text we adopt the convention that, if (S, σ) is a symplectic vector space and T is a vector subspace of S, then the presymplectic vector space $(T, \sigma\restriction_{T\times T})$ is written simply as (T, σ).

We can now also define a proper linear subspace S_0 of S by $S_0 := S_A + S_B$, and subspaces $S_A^R, S_A^L \subset S_A$, $S_B^R, S_B^L \subset S_B$ just as explained in Section 3.2, that is e.g.

$$S_A^{L/R} := \left\{ \phi \in S_A \;\middle|\; \mathrm{supp}(\phi\restriction_{\mathcal{H}_A}) \subset \mathcal{H}_A^{L/R} \right\},$$

with $\mathcal{H}_A^L$ and $\mathcal{H}_A^R$ the 'left' and 'right' portions of the A-horizon, i.e. $\mathcal{H}_A^L := \mathcal{H}_A \cap \{(t,x) \mid x < 0\}$ and $\mathcal{H}_A^R := \mathcal{H}_A \cap \{(t,x) \mid x > 0\}$ (and similarly with $S_B^{L/R}$ and $\mathcal{H}_B^{L/R}$). It is clear that (S_A, σ), (S_B, σ), $(S_A^{L/R}, \sigma)$, $(S_B^{L/R}, \sigma)$ are all symplectic spaces (indeed, for (S_A, σ), (S_B, σ) this was already established in Proposition 3.4.1). It is also clear that T_B restricts to a symplectic isomorphism between $(S_B^{L/R}, \sigma)$ and $(S_{\text{r-mov}}^{L/R}, \sigma_{\mathbb{M}})$, where $(S_{\text{r-mov}}^{L/R}, \sigma_{\mathbb{M}})$ is the symplectic subspace of $(S_{\text{r-mov}}, \sigma_{\mathbb{M}})$ consisting of purely right-moving solutions in $S_{\mathbb{M}}$ whose data on $\mathcal{H}_B$ is supported strictly to the left/right of the origin (and similarly, with B replaced by A and r-mov replaced by l-mov).

We wish next to show that the presymplectic space (S_0, σ) is also actually a symplectic space.[19] In fact we will prove a stronger result. Note first that the formula on the right-hand side of Equation (3.2) is still well-defined and antisymmetric when only one of the solutions is spatially compact, and Equations (3.9)–(3.10) are still valid in that case.

Proposition 3.4.2. *Suppose ϕ is any (not necessarily spatially compact) smooth solution to (3.8) on M which is symplectically orthogonal to both S_A and S_B, i.e. $\sigma(\phi_A, \phi) = 0 = \sigma(\phi_B, \phi)$ for all $\phi_A \in S_A$ and $\phi_B \in S_B$. Then $\phi = 0$.*

Proof. Let f, g be smooth functions such that $\phi(t,x) = f(t-x) + g(t+x)$. Solutions in S_B have the form $\phi_B(t,x) = h(t-x) + k(t+x)$ where h is any function in $C_0^\infty(\mathbb{R})$ and $k(v) = -\vartheta(v)h(-1/v)$. Therefore, if ϕ is symplectically orthogonal to S_B then Equation (3.9) implies that

$$\int_{-\infty}^{+\infty} h(u) f'(u)\, \mathrm{d}u = 0$$

for *all* $h \in C_0^\infty(\mathbb{R})$. This implies that f' is identically zero and thus that f equals a constant. A similar argument shows that g equals a constant. Thus ϕ is also constant. But then it must be zero since it is assumed to vanish on ∂M. □

As already anticipated in the Introduction, two further important observations for the purposes of this thesis are that, with the above definitions and using Equations (3.9)–(3.10), it is clearly the case that

[19] While we were proving Proposition 3.4.2 we noticed that there seems to be a gap in the arguments on a corresponding issue in [KW91]: While it was clear that the (S_A, σ) and (S_B, σ) of that paper are symplectic spaces (and the same is also true for the spaces called $(\tilde{S}_A, \hat{\sigma})$ and $(\tilde{S}_B, \hat{\sigma})$, as we show in Chapter 4) it was also tacitly assumed that (with our fiction) the space called (S_0, σ) and (without our fiction) the space called $(\tilde{S}_0, \hat{\sigma})$ are symplectic spaces. However this was never established there. The entirety of Chapter 4 is dedicated to tackling this problem and describing possible ways of filling the gap in some cases of physical interest.

- $S_A^R = S_B^R$,
- S_B^L is symplectically orthogonal to S_A. Similarly, S_A^L is symplectically orthogonal to S_B.

As final 'classical' ingredients necessary to formulate and then to prove our no-go result in the remainder of this section, we need to construct retarded/advanced Green operators $E^{\pm}$ appropriate to our Cauchy-Dirichlet problem on M, as discussed in Section 3.3.2. Namely, $E^{\pm}$ should be such that, for all $F \in C_0^{\infty}(\operatorname{Int} M)$,

$$\Box E^{\pm} F = E^{\pm} \Box F = F, \tag{3.11}$$

$$E^{\pm} F \restriction_{\partial M} = 0, \tag{3.12}$$

$$\operatorname{supp}(E^{\pm} F) \subseteq J^{\pm}(\operatorname{supp} F). \tag{3.13}$$

The resulting causal propagator $E = E^{-} - E^{+} : C_0^{\infty}(\operatorname{Int} M) \to C^{\infty}(M)$ will then clearly map to S.

We will now argue that $E^{\pm}$ with the above properties can indeed be constructed. In what follows, for each $p \in M$ we denote by $m_{+}(p)$ [resp. $m_{-}(p)$] the set of all future [resp. past] endpoints on ∂M of (smoothly) inextendible null geodesics passing through p. Equivalently, $m_{\pm}(p)$ is the intersection between ∂M and the topological boundary (in M) of $J^{\pm}(p) = J^{\pm}(\{p\})$. In particular, $m_{\pm}(p)$ is either empty or a singleton, and $m_{\pm}(p) = \{p\}$ if $p \in \partial M$. See Figure 3.4.

It is well-known and easy to verify that the unique advanced and retarded Green operators for the scalar wave equation in full (1+1)-dimensional Minkowski spacetime, which we denote by $E_{\mathbb{M}}^{\pm}$, are given by

$$[E_{\mathbb{M}}^{\pm} F](p) = \frac{1}{2} \int_{J_{\mathbb{M}}^{\mp}(p)} F \, \mathrm{d}\mu_{\eta},$$

where $p \in \mathbb{R}^2$, $F \in C_0^{\infty}(\mathbb{R}^2)$, $J_{\mathbb{M}}^{\mp}(p)$ is the causal past/future of p in the full Minkowski space, and $\mathrm{d}\mu_{\eta}$ denotes the metric volume element. Consequently, the causal propagator $E_{\mathbb{M}}$ is given by

$$[E_{\mathbb{M}} F](p) = \frac{1}{2} \left\{ \int_{J_{\mathbb{M}}^{+}(p)} - \int_{J_{\mathbb{M}}^{-}(p)} \right\} F \, \mathrm{d}\mu_{\eta} = \frac{1}{2} \left\{ \int_{V(p)} - \int_{U(p)} \right\} F \, \mathrm{d}\mu_{\eta} \tag{3.14}$$

where we have defined the sets $V(p) := \{p' : v(p') \geq v(p)\}$, $U(p) := \{p' : u(p') \leq u(p)\}$, with u and v the global null coordinates defined above. The first term in the rightmost expression is a function of the v-coordinate of p only, while the second is a function of the u-coordinate only. Thus one retrieves the expression of the solution as a sum of a left mover and a right mover, which we denote by $g_{\mathbb{M}}(v)$ and $f_{\mathbb{M}}(u)$ respectively.

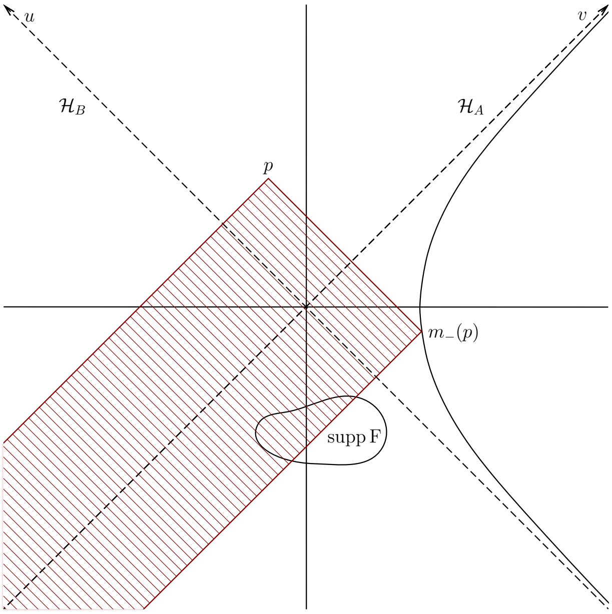

FIGURE 3.4: Illustration of the definition of the retarded propagator E^+ given in Theorem 3.4.4. Integrating one half times the source F over the shaded region gives $[E^+F](p)$. The definition of E^- can, of course, be illustrated similarly.

We next make a definition before finally being able to state the result on existence of advanced and retarded Green operators in the presence of our mirror.

Definition 3.4.3. For any open subset $X \subset \mathbb{R}^2$, we denote the space of compactly supported smooth functions on X with vanishing integral with respect to the Minkowski metric measure by $C^\infty_{00}(X)$. That is,

$$C^\infty_{00}(X) := \left\{ F \in C^\infty_0(X) \,\middle|\, \int_X F \,\mathrm{d}\mu_\eta = 0 \right\}.$$

Note that in what follows we will sometimes identify test functions defined on an open subset X with test functions on the whole of Minkowski space (by extending them to be zero outside of X). It is easy to see from Equation (3.14) that, in the full Minkowski space theory, $E_\mathbb{M}[C^\infty_{00}(\mathbb{R}^2)]$ consists of all solutions (to the massless wave equation) of the form $f(t-x)+g(t+x)$ with $f, g \in C^\infty_0(\mathbb{R})$. That is, defining the subspaces $S_{A,\mathbb{M}}$, $S_{B,\mathbb{M}}$ and $S_{0,\mathbb{M}} := S_{A,\mathbb{M}} + S_{B,\mathbb{M}}$ of $S_\mathbb{M}$, in a manner analogous to the way we defined S_A, S_B and $S_0 = S_A + S_B$, one has $S_{0,\mathbb{M}} = E_\mathbb{M}[C^\infty_{00}(\mathbb{R}^2)]$.

Theorem 3.4.4. *The linear operators* $E^{\pm} : C_0^\infty(\operatorname{Int} M) \to C^\infty(M)$ *defined, for all* $F \in C_0^\infty(\operatorname{Int} M)$ *and* $p \in M$*, by*

$$[E^{\pm}F](p) = \frac{1}{2}\left\{\int_{J^{\mp}(p)} - \int_{J^{\mp}(m_{\mp}(p))}\right\} F \,\mathrm{d}\mu_\eta$$
$$= [E_{\mathbb{M}}^{\pm}F](p) - \frac{1}{2}\int_{J^{\mp}(m_{\mp}(p))} F \,\mathrm{d}\mu_\eta, \tag{3.15}$$

(see Figure 3.4) satisfy Equations (3.11)–(3.13). Furthermore, $S_0 = E[C_{00}^\infty(\operatorname{Int} M)]$.

Remark. The second summand on the right-hand side of Equation (3.15) equals zero (for any test function) at any point p for which $m_{\mp}(p) = \emptyset$ [whereupon the integration domain $J^{\mp}(m_{\mp}(p))$ is also empty]. When $m_{\mp}(p)$ consists of the point $p_{\mp}$, it equals $[E_{\mathbb{M}}^{\pm}F](p_{\mp})$.

Proof. That each $E^{\pm}F$ is smooth is obvious since our test functions have compact support. The boundary condition, Equation (3.12), and the support property, Equation (3.13), also hold trivially.

We now turn to the equations in (3.11), i.e. to the two-sided inverse property, on the domain $C_0^\infty(\operatorname{Int} M)$, of $E^{\pm}$ with respect to the d'Alembert operator $\Box$. We carry out the proof explicitly in the case of E^+; the arguments for E^- are analogous. In view of the fact that the corresponding object $E_{\mathbb{M}}^+$ on full Minkowski space is already known to satisfy the analogous two-sided inverse property for all test functions (and thus in particular for those supported in $\operatorname{Int} M$), we need to check that the operator $D^+ = E^+ - E_{\mathbb{M}}^+$, whose action is defined by the second summand on the right-hand side of Equation (3.15), is such that $D^+\Box F = 0 = \Box D^+ F$ whenever $F \in C_0^\infty(\operatorname{Int} M)$. Using the remark above and, again, the left-inverse property for $E_{\mathbb{M}}^+$, it is easy to see that the first of these identities holds because any $F \in C_0^\infty(\operatorname{Int} M)$ vanishes on ∂M. To verify the second identity, we first express D^+F in terms of the null coordinates u and v. For any $p \in M$, $m_-(p)$ is empty if $v(p) \le 0$, and contains only the point with null coordinates $u_- = -1/v(p)$ and $v_- = v(p)$ if $v(p) > 0$. Therefore, one has

$$\widetilde{[D^+F]}(u,v) = -\frac{\vartheta(v)}{4}\int_{u' \le -1/v} \tilde{F}(u',v')\,\mathrm{d}u'\,\mathrm{d}v', \tag{3.16}$$

where the tilde indicates that one is dealing with the coordinate expression of a function in the (u,v) coordinate system, and ϑ denotes the Heaviside step function. The right-hand side of Equation (3.16) is clearly annihilated by $\partial/\partial u$, and thus in particular by $\Box = 4\partial^2/\partial u \partial v$. This completes the proof of the right-inverse property for E^+.

In order to prove the second statement in the theorem, we first point out that it is straightforward to check that, for any $F \in C_0^\infty(\operatorname{Int} M)$,

$$\widetilde{[EF]}(u,v) = f_{\mathbb{M}}(u) + g_{\mathbb{M}}(v) - \vartheta(-u)g_{\mathbb{M}}(-1/u) - \vartheta(v)f_{\mathbb{M}}(-1/v), \tag{3.17}$$

where $f_{\mathbb{M}}$ and $g_{\mathbb{M}}$ denote the right- and left-moving parts of $E_{\mathbb{M}}F$ obtained in the manner described in the discussion under Equation (3.14). That is,

$$f_{\mathbb{M}}(u) = -\frac{1}{4}\int_{u' \leq u} \tilde{F}(u', v')\,\mathrm{d}u'\,\mathrm{d}v' \quad \text{and} \quad g_{\mathbb{M}}(v) = \frac{1}{4}\int_{v' \geq v} \tilde{F}(u', v')\,\mathrm{d}u'\,\mathrm{d}v'. \tag{3.18}$$

Since $f_{\mathbb{M}}$ and $g_{\mathbb{M}}$ have compact support when $F \in C_{00}^{\infty}(\operatorname{Int} M)$, it follows that $E[C_{00}^{\infty}(\operatorname{Int} M)] \subseteq S_0$. To prove the reverse inclusion, it clearly suffices to show that S_A and S_B are individually contained in $E[C_{00}^{\infty}(\operatorname{Int} M)]$. We give the argument for S_B, the one for S_A being entirely analogous. If $\phi \in S_B$ then $\tilde{\phi}(u,v) = h(u) + k(v)$ where $h \in C_0^{\infty}(\mathbb{R})$ and $k(v) = -\vartheta(v)h(-1/v)$. In view of Equation (3.17), it therefore suffices to find an $F \in C_{00}^{\infty}(\operatorname{Int} M)$ such that $f_{\mathbb{M}}$ and $g_{\mathbb{M}}$ in Equation (3.18) equal h and 0 respectively (i.e. F needs to integrate to zero, be supported in $\operatorname{Int} M$ and generate the pure right-mover – in the full Minkowski space theory – described by h). This can be done as follows: Pick any $\chi \in C_0^{\infty}(\mathbb{R})$ with the properties that $\operatorname{supp}\chi \subset (-\infty, 0)$ and $\int_{\mathbb{R}} \chi(x)\,\mathrm{d}x = 1$. Then, the function F defined by

$$\tilde{F}(u,v) = -4h'(u)\chi(v) \tag{3.19}$$

clearly fulfills the required properties. □

To make contact with the general discussion in Section 3.3.2, we remark that we have *not* proved that $E : C_0^{\infty}(\operatorname{Int} M) \to C^{\infty}(M)$ is onto S. Indeed, as pointed out there, we don't expect this to be the case. Nor, as also anticipated there, is the kernel of the causal propagator constructed in Theorem 3.4.4 equal to $\Box[C_0^{\infty}(\operatorname{Int} M)]$, as one can see from Equation (3.17). Indeed, one need only pick a test function $F \in C_0^{\infty}(\operatorname{Int} M)$ which, on the entire Minkowski space, would propagate to a non-zero solution with right- and left-moving parts $f_{\mathbb{M}}$ and $g_{\mathbb{M}}$ respectively (obtained again in the manner described in the discussion under Equation (3.14)), which are such that $f_{\mathbb{M}}(u) = \vartheta(-u)g_{\mathbb{M}}(-1/u)$ and $g_{\mathbb{M}}(v) = \vartheta(v)f_{\mathbb{M}}(-1/v)$ for all $u, v \in \mathbb{R}$. Then $EF = 0$ but F cannot equal $\Box G$ for any $G \in C_0^{\infty}(\operatorname{Int} M)$ since $E_{\mathbb{M}}F \neq 0$ in full Minkowski space. See Figure 3.5. In Section 5.2.1, we will tackle the general problem of deriving exact sequences analogous to (2.27) for hyperbolic boundary-value problems. Those results suggest that to obtain an exact sequence one should seek enlargements of the advanced and retarded Green operators to include test functions which do not necessarily vanish at the boundary of M, and also seek smooth functions with the properties enjoyed by χ_- and χ_- in the statement of Theorem 5.2.3. While we believe this to be possible (see the discussion following the proof of Theorem 5.2.3), we will not pursue it in detail in this section.

As another side remark, we note that, equipped with the above results, one can straightforwardly imitate an argument which is standard in the globally hyperbolic setup (see

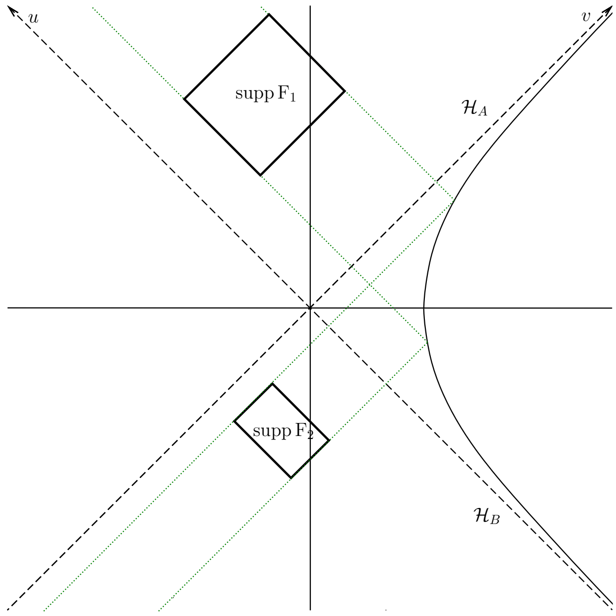

FIGURE 3.5: Illustration of the failure of $\ker E$ to be equal to $\Box[C_0^\infty(\operatorname{Int} M)]$. The F in the discussion in the main text (which in this illustration has disconnected support) is to be identified with $F_1 - F_2$. Then $EF = 0$ but F cannot equal $\Box G$ for any $G \in C_0^\infty(\operatorname{Int} M)$ since $E_{\mathbb{M}}F \neq 0$ in full Minkowski space.

e.g. [BGP07, Lemma 3.2.2]) to show that, for any $F \in C_0^\infty(\operatorname{Int} M)$ and $\phi \in S$,

$$\int_M F\phi \, \mathrm{d}\mu_\eta = \sigma(EF, \phi). \tag{3.20}$$

Equation (3.20) provides the alternative way, promised above, to show the non-degeneracy of σ. Indeed, for any given $\phi \in S$ it is clearly possible to find a test function $F \in C_0^\infty(\operatorname{Int} M)$, not in the kernel of E (i.e. not generating the zero solution) and such that $\int_M F\phi \, \mathrm{d}\mu_\eta \neq 0$ – any F which is everywhere non-zero and is sufficiently localised around a point where ϕ attains a non-zero value will do.

We conclude this section by briefly discussing the action of Lorentz boost isometries on elements of S. The one-parameter group $(\beta_\tau)_{\tau\in\mathbb{R}}$ of Lorentz-boost isometries yields a one-parameter abelian group of linear symplectomorphisms $\mathscr{T}_\tau : S \to S$ via pullback by the inverse maps, i.e. $\mathscr{T}_\tau\phi := \phi \circ \beta_{-\tau}$. Explicitly, if $\phi(t,x) = f(t-x) + g(t+x)$ then $[\mathscr{T}_\tau\phi](t,x) = f_\tau(t-x) + g_\tau(t+x)$ where $f_\tau(u) = f(e^{a\tau}u)$ and $g_\tau(v) = g(e^{-a\tau}v)$.

3.4.2 The infra-red pathology and the Hadamard notion

We now wish to discuss the prospects for identifying an appropriate framework for the quantization of the massless field on (M, η). We first recall some of the issues arising in the quantization of massless fields in *full* (1+1)-dimensional Minkowski spacetime.

As we mentioned in the Introduction and in Section 3.2, in attempting to define a ground state representation there, one is faced with an infra-red pathology which has been extensively discussed in the literature, starting from the foundational work in [Sch63] and in [Wig67, pp. 204–215] (further aspects were discussed in e.g. [SW70, Kay85, FR87, DM06]). To recall the issue: One might attempt to define the quantum field as a genuine operator-valued distribution[20] by proceeding in the usual way involving creation and annihilation operators on the standard bosonic Fock space $\mathscr{F} = \bigoplus_{n=0}^{\infty} L^2(\mathbb{R})^{\odot n}$. One would then demand that the Fock vacuum vector Ω belong to a common invariant (and dense) domain for all thus defined field operators. However, in general the resulting one-particle vectors $\hat{\phi}(F)\Omega$ – generated by acting on the vacuum with the candidate quantum field smeared with an arbitrary test function F on spacetime – might not be square integrable. In fact, if $\tilde{F}(k)$ is the Fourier transform, $(1/\sqrt{2\pi}) \int_{\mathbb{R}^2} F(x) e^{-ik\cdot x}\, \mathrm{d}^2 x$ of F, the vacuum belongs to the domain of $\hat{\phi}(F)$ if and only if $\tilde{F}(0) = 0$. This problem starkly manifests itself at the level of the tentative 'two-point function', which is formally given by

$$\left\langle \Omega \,\middle|\, \hat{\phi}(F)\hat{\phi}(G)\Omega \right\rangle = \pi \int_{-\infty}^{+\infty} \tilde{F}(|p|, -p)\tilde{G}(-|p|, p)\, \frac{\mathrm{d}p}{|p|}. \tag{3.21}$$

Indeed, the above clearly diverges (logarithmically) unless one of $\tilde{F}(0)$ or $\tilde{G}(0)$ equals zero.

Thus the usual quantization procedure fails to produce, via Equation (3.21), a bidistribution, Λ, on $\mathbb{R}^2$ representing two-point correlators, because one can't allow for generic test functions. If, however, one restricts to smearings with elements of the linear subspace $C_{00}^{\infty}(\mathbb{R}^2)$ of Definition 3.4.3, then both this 'two-point functional' exists and (by construction via creation and annihilation operators) satisfies the positivity properties $\Lambda(F, F) \geq 0$, $E_{\mathbb{M}}(F, G)^2 \leq 4\Lambda(F, F)\Lambda(G, G)$[21] required for a probabilistic interpretation.

In the Weyl-algebraic approach to quantization which we adopt in this thesis (see Appendix A), what is problematic is the attempt to define a ground state with respect to time translations on the Weyl algebra $\mathscr{A}_{\mathbb{M}} = \mathscr{W}(S_{\mathbb{M}}, \sigma_{\mathbb{M}})$ generated by the symplectic space $(S_{\mathbb{M}}, \sigma_{\mathbb{M}})$ defined in Section 3.4.1. But we observe that, if we restrict to the Weyl subalgebra $\mathscr{A}_{0,\mathbb{M}} = \mathscr{W}(S_{0,\mathbb{M}} = E_{\mathbb{M}}[C_{00}^{\infty}(\mathbb{R}^2)], \sigma_{\mathbb{M}})$ then there *is* an unproblematic ground state with respect to time translations, namely the state whose spacetime smeared two-point function is precisely the 'two-point functional' of the previous paragraph. In Section 3.4.3 we will refer to this state on $\mathscr{A}_{0,\mathbb{M}}$ – which, we remark in passing, is a quasi-free state – as $\omega_{\mathbb{M}}$, and to its symplectically smeared two-point function as $\lambda_{\mathbb{M}}$. In

[20]It is irrelevant to this discussion whether the quantum field is to be smeared with test functions in $C_0^{\infty}(\mathbb{R}^2)$ or, say, test functions in Schwartz space $\mathscr{S}(\mathbb{R}^2; \mathbb{R})$. But we will work with the former space because it's technically more appropriate for our needs in this section.

[21]If we let $\mathscr{D}_0(\mathbb{R}^2)$ denote the complexification of $C_{00}^{\infty}(\mathbb{R}^2)$ then these positivity conditions can be succinctly expressed as $\Lambda^{\mathbb{C}}(\bar{F}, F) \geq 0 \ \forall\ F \in \mathscr{D}_0(\mathbb{R}^2)$, where $\Lambda^{\mathbb{C}}$ denotes the extension by complex bilinearity of Λ to a bilinear form on $\mathscr{D}_0(\mathbb{R}^2)$.

view of this, from now on we adopt the view (essentially what in [FR87] is termed the 'liberal' approach to dealing with the infra-red pathology) that our 'physical algebra' is this Weyl subalgebra $\mathscr{A}_{0,\mathbb{M}}$ and 'physical states' are to be sought amongst positive linear functionals on $\mathscr{A}_{0,\mathbb{M}}$.

A price to pay for working in this framework is that the spacetime smeared two-point functions of our thus-defined physical states are only defined as bilinear functionals $C_{00}^{\infty}(\mathbb{R}^2) \times C_{00}^{\infty}(\mathbb{R}^2) \to \mathbb{C}$, and therefore do not define true bidistributions on $\mathbb{R}^2$. As a result, what one might mean by a globally (or even locally!) 'Hadamard' state becomes problematic. We propose to overcome this by declaring that a state on $\mathscr{A}_{0,\mathbb{M}}$ be called globally Hadamard if its spacetime smeared two-point function $\Lambda : C_{00}^{\infty}(\mathbb{R}^2) \times C_{00}^{\infty}(\mathbb{R}^2) \to \mathbb{C}$ (exists and) admits an extension $\Lambda^{\text{ext}} : C_0^{\infty}(\mathbb{R}^2) \times C_0^{\infty}(\mathbb{R}^2) \to \mathbb{C}$ which is globally Hadamard (on $\mathbb{R}^2$). Note that this extension need not satisfy any positivity property beyond positivity (in the above sense) when restricted to smearings in $C_{00}^{\infty}(\mathbb{R}^2)$. The (1+1)-dimensional version of the global Hadamard condition for bidistributions was written down in [Mor03] (along with versions appropriate to all other spacetime dimensions). For a massless theory in any globally hyperbolic open subset $\mathcal{O}$ of (1+1)-dimensional Minkowski space, it simply amounts to the following.

Definition 3.4.5 (*Global Hadamard condition on $\mathcal{O}$, massless case*)**.** A bidistribution Λ on $\mathcal{O}$ satisfies the global Hadamard condition if there exists a Cauchy surface $\mathscr{C}$ for $(\mathcal{O}, \eta)$, a causal normal neighbourhood $\mathcal{N} \subseteq \mathcal{O}$ of $\mathscr{C}$, a 'smoothing function' $\chi \in C^{\infty}(\mathcal{N} \times \mathcal{N})$, a global temporal function T on $\mathcal{O}$ increasing towards the future,[22] and a *smooth* function $H_{\mathcal{N}}$ on $\mathcal{N} \times \mathcal{N}$ such that, for all $F, G \in C_0^{\infty}(\mathcal{N})$,

$$\Lambda(F, G) = \lim_{\varepsilon \to 0^+} \int_{\mathcal{N} \times \mathcal{N}} \left(-\frac{\chi(x,y)}{4\pi} \ln \frac{-s_{\varepsilon,T}(x,y)}{\lambda^2} + H_{\mathcal{N}}(x,y) \right) F(x) G(y) \, \mathrm{d}\mu_{\eta}(x) \, \mathrm{d}\mu_{\eta}(y).$$

In the above, for all $\varepsilon > 0$,

$$s_{\varepsilon,T}(x,y) := s(x,y) - 2i\varepsilon(T(x) - T(y)) - \varepsilon^2,$$

with $s(x,y) = (x-y)^2$ and the branch-cut of the logarithm chosen to lie on the negative real axis. Finally, λ is a length scale introduced for dimensional reasons, but clearly the property being defined does not depend on it.

Clearly, the ground state on the physical algebra $\mathscr{A}_{0,\mathbb{M}}$ is a globally Hadamard state in this sense. To prepare the ground for our discussion on the case of the spacetime (M, η) we're interested in, where the Lorentz boosts are the only continuous isometries, we notice that actually more is true about this state on $\mathscr{A}_{0,\mathbb{M}}$, namely that one can find an extension of its spacetime smeared two-point function which, on its larger domain

[22] We refer to [KW91, Rad96] for complete definitions of $\mathcal{N}$, χ and T.

$C_0^\infty(\mathbb{R}^2) \times C_0^\infty(\mathbb{R}^2)$, is still boost-invariant, a weak bisolution of the wave equation, and satisfies the canonical commutation relations. Indeed, $\Lambda_{\mathbb{M}}$ defined by

$$\Lambda_{\mathbb{M}}(F,G) = -\frac{1}{4\pi} \lim_{\varepsilon \to 0^+} \int \log[-(x-y)^2 + i\varepsilon(x^0 - y^0)] F(x) G(y) \, \mathrm{d}\mu_\eta(x) \, \mathrm{d}\mu_\eta(y) \quad (3.22)$$

gives such an extension (and so does '$\Lambda_{\mathbb{M}}(x,y) + c$' for any $c \in \mathbb{C}$). It can be seen that, indeed, no such extension can satisfy the necessary positivity conditions for all test functions.

The arguments we made above in favour of taking the 'physical algebra' to be $\mathscr{A}_{0,\mathbb{M}}$ privileged the role of the usual Minkowski ground state (i.e. the Poincaré invariant vacuum). One might nevertheless still want to explore what could be said about (globally) Hadamard states on the 'full' Weyl algebra $\mathscr{A}_{\mathbb{M}}$. Within the (technically inequivalent) approach to quantization based on the 'full' Borchers–Uhlmann algebra, Schubert [Sch13] has recently shown that there are no time-translation–invariant Hadamard states; it seems reasonable to expect that a similar result will hold within our Weyl-algebra framework. And, as we prove in Appendix B, there is no C^2 state on $\mathscr{A}_{\mathbb{M}}$ which is globally Hadamard and boost-invariant, either. This is another reason to take the view that the 'physical algebra' is $\mathscr{A}_{0,\mathbb{M}}$.

3.4.3 The non-existence theorem

Having carefully set up the classical theory for massless fields on our one-mirror spacetime (M, η), and having clarified our perspective on both the appropriate strategy to deal with spacetimes with boundaries (in Section 3.3.2), and the status of the infrared pathology for massless fields on full (1+1)-dimensional Minkowski spacetime, we now turn to the theory obtained by quantizing the classical system analysed in Section 3.4.1. For this theory, we are now in a position to rigorously define an appropriate class of quantum states for which we are able to prove a non-existence theorem (Theorem 3.4.7) which, arguably (see however Footnote 24) is analogous to the non-existence result which we conjecture for Kruskal. Namely, the class of 'strongly boost-invariant globally-Hadamard' states of Definition 3.4.6 below. Indeed, we will show that once our definitions are in place, the strategy outlined in Section 3.2 becomes a rigorous proof of this theorem once Equation (3.4) is established.

In the previous section we have argued that the 'physical algebra' for massless fields on full (1+1)-dimensional Minkowski space is the Weyl subalgebra $\mathscr{A}_{0,\mathbb{M}}$ of $\mathscr{A}_{\mathbb{M}}$ generated by Minkowski-space solutions in $S_{0,\mathbb{M}}$. Similarly, here we regard the 'physical algebra' for massless fields on (M, η), satisfying Dirichlet boundary conditions on ∂M, to be *not*

$\mathscr{A} := \mathscr{W}(S,\sigma)$, but rather its subalgebra $\mathscr{A}_0 := \mathscr{W}(S_0,\sigma)$ generated by solutions in S_0 [cf. Section 3.4.1 for definitions of the symplectic vector spaces (S,σ) and (S_0,σ)].

Definition 3.4.6. A *strongly boost-invariant globally-Hadamard* state on $\mathscr{A}_0$ is a boost-invariant state on $\mathscr{A}_0$ whose spacetime smeared two-point function Λ exists and admits an extension Λ^{ext} to a bidistribution on $\operatorname{Int} M$ which is (i) globally Hadamard in the sense of Definition 3.3.1, (ii) boost-invariant and (iii) a weak bisolution of the wave equation.[23]

We remark that one could contemplate replacing the word 'global' in this definition by the word 'local' and thereby define a notion of 'strongly boost-invariant *locally*-Hadamard'. However, in view of the fact that no assumption of positivity is made for the extension of the spacetime smeared two-point function, the local-to-global theorem of Radzikowski [RV96, Rad92] will presumably not be available to conclude that the two notions are equivalent and it is not clear whether we would be able to prove that there is no state satisfying the local version of the definition.

We point out that, with $\operatorname{Int} M$ replaced by $\mathbb{R}^2$ and $\mathscr{A}_0$ replaced by $\mathscr{A}_{0,\mathbb{M}}$ in the above definition, there obviously *is* a strongly boost-invariant globally-Hadamard state on $\mathscr{A}_{0,\mathbb{M}}$ – namely $\omega_{\mathbb{M}}$ as we in fact pointed out at the end of the previous section. And most importantly, with the obvious replacements, in the case with two mirrors (see the Introduction) there *is* a strongly boost-invariant globally-Hadamard state, namely the 'Hartle-Hawking-Israel-like state' constructed in [Kay15] with two-point function given by Equation (5) in that paper – as one may verify by inspection of that formula.

In contrast, however…

Theorem 3.4.7. *There is no strongly boost-invariant globally-Hadamard state on* $\mathscr{A}_0$.[24]

We first record and prove a preliminary lemma.

Lemma 3.4.8. *For any two solutions* ϕ_1, ϕ_2 *in* S_B *one can find a causally convex and globally hyperbolic open subregion* $\mathcal{O}$ *of* $\operatorname{Int} M$*, a pair of test functions* $F_1, F_2 \in C_{00}^\infty(\mathcal{O})$ *and a Cauchy surface* $\mathscr{C}$ *for* $\mathcal{O}$ *containing a portion of* $\mathcal{H}_B$*, such that*

- *the Cauchy data for* ϕ_1 *and* ϕ_2 *on* $\mathscr{C}$ *vanish outside* $\mathscr{C} \cap \mathcal{H}_B$*;*

[23] It is not assumed that this extension still satisfies the canonical commutation relations for all test functions, i.e. that $\Lambda^{\text{ext}}(F,G) - \Lambda^{\text{ext}}(G,F) = iE(F,G)$ for all $F, G \in C_0^\infty(\operatorname{Int} M)$ [of course these are satisfied for pairs of test functions belonging to the subspace $C_{00}^\infty(\operatorname{Int} M)$].

[24] This theorem of course implies that there are no boost-invariant states on the 'full' Weyl algebra $\mathscr{A}$ with globally Hadamard spacetime smeared two-point function (in the sense of Definitions 3.3.1 and 3.4.5), since the restriction to $\mathscr{A}_0$ of any such state would obviously be a strongly boost-invariant globally-Hadamard state on $\mathscr{A}_0$. However, it does *not* imply that there is no state on $\mathscr{A}_0$ which is boost-invariant and whose spacetime smeared two-point (exists and) admits an extension to a globally Hadamard bidistribution on $\operatorname{Int} M$, i.e. one satisfying (i) but not (ii) and/or (iii) in Definition 3.4.6.

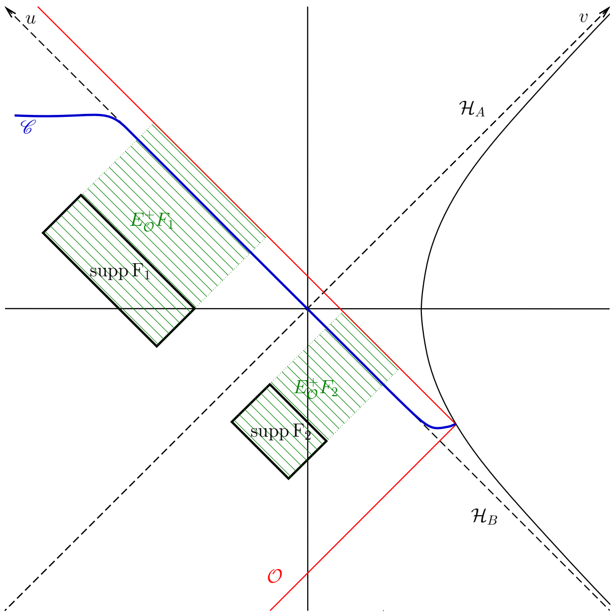

FIGURE 3.6: Illustration of the set $\mathcal{O}$, Cauchy surface $\mathscr{C}$ and test functions F_1, F_2 constructed in Lemma 3.4.8 and used in the proof of Theorem 3.4.7.

- *F_1 and F_2 have support in $I^-_{\mathcal{O}}(\mathcal{H}_B \cap \mathcal{O}) \cap I^-_{\mathcal{O}}(\mathscr{C})$, $EF_1 = \phi_1$ and $EF_2 = \phi_2$ (here $I^{\pm}_{\mathcal{O}}(S)$ denotes the chronological future/past of a subset S in $\mathcal{O}$);*
- *$E_{\mathbb{M}}F_1$ is the full Minkowski space solution which is purely right-moving and with restriction to $\mathcal{H}_B$ equal to $\phi_1\restriction_{\mathcal{H}_B}$, i.e. $E_{\mathbb{M}}F_1 = T_B\phi_1$ where T_B is the linear symplectomorphism of Proposition 3.4.1 (and a similar statement with $F_1 \leftrightarrow F_2$, and $\phi_1 \leftrightarrow \phi_2$).*

Moreover, $\mathcal{O}$ can be taken to be geodesically convex, and therefore a causal normal neighbourhood of any of its Cauchy surfaces. All the above holds equally with $\mathcal{H}_B \leftrightarrow \mathcal{H}_A$ and $T_B \leftrightarrow T_A$.

Proof. Since $\phi_i \in S_B$ $(i = 1, 2)$, there exists a unique function $f_i \in C_0^\infty(\mathbb{R})$ such that $\tilde{\phi}_i(u, v) = f_i(u) - \vartheta(v) f_i(-1/v)\ \forall\, u, v$. Pick $u_{\mathrm{m}} < 0$ with $\operatorname{supp} f_1 \cup \operatorname{supp} f_2 \subset (u_{\mathrm{m}}, +\infty)$. Then $\mathcal{O} := \{(t, x) \mid u(t, x) > u_{\mathrm{m}},\ v(t, x) < -1/u_{\mathrm{m}}\}$ is clearly a causally and geodesically convex, globally hyperbolic open subregion of $\operatorname{Int} M$, and for $|u_{\mathrm{m}}|$ sufficiently large it is clear that a Cauchy surface $\mathscr{C}$ for $\mathcal{O}$ can be found satisfying the requirements in the statement of the Lemma, see Figure 3.6.

In order to prove the statements about F_1, F_2 one proceeds just as in the proof of Theorem 3.4.4 (cf. in particular Equations (3.17)–(3.18) and the discussion following these), namely picking any $\chi \in C_0^\infty(\mathbb{R})$ such that $\operatorname{supp} \chi \subset (-\infty, 0)$ and $\int_{\mathbb{R}} \chi(x)\, \mathrm{d}x = 1$, and then defining $\tilde{F}_i(u, v) = -4\chi(v) f'(u)$.[25] □

[25]Note that, defining $\psi(s) = \int_{-\infty}^{s} \chi(s')\, \mathrm{d}s'$ and $\xi_i(t, x) = -\psi(v(t, x))\phi_i(t, x)$, this amounts to setting $F_i = \Box \xi_i$.

Proof of Theorem 3.4.7. As already outlined in the Introduction and in Section 3.2, one need only prove that there are no *quasi-free* strongly boost-invariant globally-Hadamard states. Thus, suppose such a quasi-free state ω exists with spacetime smeared two-point function $\Lambda : C_{00}^\infty(\mathrm{Int}\, M) \times C_{00}^\infty(\mathrm{Int}\, M) \to \mathbb{C}$, and let Λ^{ext} be an extension of Λ satisfying (i), (ii) and (iii) in Definition 3.4.6. Let $\phi_1, \phi_2 \in S_B$ and pick a causally and geodesically convex, open, globally hyperbolic subset $\mathcal{O}$ of $\mathrm{Int}\, M$, a Cauchy surface $\mathscr{C}$ for $\mathcal{O}$ and a pair of test functions $F_1, F_2 \in C_{00}^\infty(\mathcal{O})$, as in the statement and proof of Lemma 3.4.8. Since Λ^{ext} is a globally Hadamard bidistribution on $\mathcal{O}$, results on the *propagation of the global Hadamard form* contained in [FSW78, KW91] guarantee that we are free to choose $\mathcal{N} = \mathcal{O}$, $\chi \equiv 1$ and $T(t, x) = t$ as the causal normal neighbourhood, 'smoothing function' and 'global time function' in Definition 3.4.5. In terms of these, the global Hadamard condition for Λ^{ext} simply reduces to the existence of a function $H_{\mathcal{O}} \in C^\infty(\mathcal{O} \times \mathcal{O})$ such that

$$\Lambda^{\mathrm{ext}}(F, G) - \Lambda_{\mathbb{M}}(F, G) = \int_{\mathcal{O}\times\mathcal{O}} H_{\mathcal{O}}(x, y) F(x) G(x) \,\mathrm{d}\mu_\eta(x) \,\mathrm{d}\mu_\eta(y) \tag{3.23}$$

for all $F, G \in C_0^\infty(\mathcal{O})$, where $\Lambda_{\mathbb{M}}$ is as defined in Equation (3.22). We remark that, since Λ^{ext} and $\Lambda_{\mathbb{M}}$ are both weak bisolutions of the wave equation, then $H_{\mathcal{O}}$ is a (smooth) bisolution of the wave equation. Also, since both $\Lambda^{\mathrm{ext}}\restriction_{C_0^\infty(\mathcal{O})\times C_0^\infty(\mathcal{O})}$ (by assumption) and $\Lambda_{\mathbb{M}}\restriction_{C_0^\infty(\mathcal{O})\times C_0^\infty(\mathcal{O})}$ are invariant under the (local) one-parameter group of Lorentz boosts applied to the two copies of $C_0^\infty(\mathcal{O})$ simultaneously, it follows that $H_{\mathcal{O}}$ is annihilated by the formal adjoint X^* of the infinitesimal generator $X = X_1 \oplus X_2 = (x_1 \partial/\partial t_1 + t_1 \partial/\partial x_1) \oplus (x_2 \partial/\partial t_2 + t_2 \partial/\partial x_2)$ (where, for $i = 1, 2$, t_i and x_i are inertial coordinates on the i-th copy of $\mathcal{O}$). Since $X^* = -X$, it follows that $H_{\mathcal{O}}$ is constant on the integral curves of X on $\mathcal{O}\times\mathcal{O}$. Together with global smoothness (and in particular smoothness at the point $(0, 0; 0, 0)$), this clearly implies that $H_{\mathcal{O}}$ is constant on the portion of $\mathcal{H}_B \times \mathcal{H}_B$ contained within $\mathcal{O} \times \mathcal{O}$.

Now recall that the test functions F_1 and F_2 were chosen to both have support in $I_{\mathcal{O}}^-(\mathcal{H}_B \cap \mathcal{O}) \cap I_{\mathcal{O}}^-(\mathscr{C})$ (see again Figure 3.6). Let $\alpha \in C^\infty(\mathcal{O})$ and F be any test function supported in $I_{\mathcal{O}}^-(\mathscr{C})$. Then, proceeding similarly to Equations (B.12)–(B.13) in Appendix B of [KW91], and noting that $\Box = \nabla^a \nabla_a$,

$$\begin{aligned}
\int_{\mathcal{O}} \alpha F \,\mathrm{d}\mu_\eta &= \int_{I_{\mathcal{O}}^-(\mathscr{C})} \alpha F \,\mathrm{d}\mu_\eta \\
&= \int_{I_{\mathcal{O}}^-(\mathscr{C})} \alpha \Box E_{\mathcal{O}}^+ F \,\mathrm{d}\mu_\eta \\
&= \int_{I_{\mathcal{O}}^-(\mathscr{C})} [\Box \alpha] E_{\mathcal{O}}^+ F \,\mathrm{d}\mu_\eta + \int_{I_{\mathcal{O}}^-(\mathscr{C})} \nabla^a [\alpha \overleftrightarrow{\nabla}_a E_{\mathcal{O}}^+ F] \,\mathrm{d}\mu_\eta
\end{aligned}$$

$$= \int\limits_{I^-_{\mathcal{O}}(\mathscr{C})} [\Box\alpha] E^+_{\mathcal{O}} F \, \mathrm{d}\mu_\eta + \int\limits_{\mathscr{C}} n_a [\alpha \overleftrightarrow{\nabla}^a E^+_{\mathcal{O}} F] \, \mathrm{d}\mu_{\mathscr{C}}$$

$$= \int\limits_{I^-_{\mathcal{O}}(\mathscr{C})} [\Box\alpha] E^+ F \, \mathrm{d}\mu_\eta - \int\limits_{\mathscr{C}} n_a [\alpha \overleftrightarrow{\nabla}^a E F] \, \mathrm{d}\mu_{\mathscr{C}}, \tag{3.24}$$

where $E^\pm_{\mathcal{O}}$ denotes the retarded/advanced Green operator for $\Box$ on $\mathcal{O}$, in the fourth step Gauss' law has been applied, and in the final step we used the fact that $E^- F$ vanishes on a neighbourhood of $\mathscr{C}$, together with Equation (3.7).
Recalling the fact that $H_{\mathcal{O}} \in C^\infty(\mathcal{O} \times \mathcal{O})$ is a bisolution of the wave equation, and applying Equation (3.24) twice, with first α interpreted as $\int_{\mathcal{O}} H_{\mathcal{O}}(\cdot, x_2) F_2(x_2) \, \mathrm{d}\mu_\eta(x_2)$ and F interpreted as F_1, and then with α interpreted as $H_{\mathcal{O}}(x_1, \cdot)$ for arbitrary fixed $x_1 \in \mathcal{O}$ and F interpreted as F_2, yields

$$\int\limits_{\mathcal{O}\times\mathcal{O}} H_{\mathcal{O}}(x_1, x_2) F_1(x_1) F_2(x_2) \, \mathrm{d}\mu_\eta(x_1) \, \mathrm{d}\mu_\eta(x_2)$$

$$= \int\limits_{\mathscr{C}\times\mathscr{C}} H_{\mathcal{O}}(x_1, x_2) \overleftrightarrow{\nabla}^{1a} \overleftrightarrow{\nabla}^{2b} \phi_1(x_1) \phi_2(x_2) \, n_a(x_1) n_b(x_2) \, \mathrm{d}\mu_{\mathscr{C}}(x_1) \mathrm{d}\mu_{\mathscr{C}}(x_2)$$

$$= 4 \int\limits_{(\mathcal{H}_B \times \mathcal{H}_B)\cap(\mathcal{O}\times\mathcal{O})} [\nabla^{1a} \nabla^{2b} H_{\mathcal{O}}(x_1, x_2)] \phi_1(x_1) \phi_2(x_2) \, n_a(x_1) n_b(x_2) \, \mathrm{d}\mu_{\mathcal{H}_B}(x_1) \mathrm{d}\mu_{\mathcal{H}_B}(x_2)$$

$$= 0.$$

In the second step, we have used the fact that the Cauchy data for ϕ_1 and ϕ_2 are supported in $\mathscr{C} \cap \mathcal{H}_B$ and performed two integrations by parts. The final equality is a consequence of the constancy of $H_{\mathcal{O}}$ on $(\mathcal{H}_B \times \mathcal{H}_B) \cap (\mathcal{O} \times \mathcal{O})$. This proves that $\Lambda(F_1, F_2) = \Lambda_{\mathbb{M}}(F_1, F_2)$. In terms of the symplectically smeared two-point function λ_2 of our state ω, this means that

$$\lambda_2(\phi_1, \phi_2) = \lambda_{\mathbb{M}}(E_{\mathbb{M}} F_1, E_{\mathbb{M}} F_2),$$

where we recall that $\lambda_{\mathbb{M}}$ denotes the symplectically smeared two-point function of the (1+1)-dimensional Minkowski vacuum state $\omega_{\mathbb{M}}$ on $\mathscr{A}_{0,\mathbb{M}}$ discussed in Section 3.4.2. But since F_1 and F_2 were chosen so that $E_{\mathbb{M}} F_1 = T_B \phi_1$ and $E_{\mathbb{M}} F_2 = T_B \phi_2$, and since $\phi_1, \phi_2 \in S_B$ are arbitrary, we conclude that in fact

$$\lambda_2(\phi_1, \phi_2) = \lambda_{\mathbb{M}}(T_B \phi_1, T_B \phi_2) \tag{3.25}$$

for *all* $\phi_1, \phi_2 \in S_B$. Next, let $(K, \mathscr{H})$ be the one-particle structure associated to ω, and let $(K_{\mathbb{M}}, \mathscr{H}_{\mathbb{M}})$ be the one-particle structure associated to $\omega_{\mathbb{M}}$ (see Proposition A.0.2 in

Appendix A), then Equation (3.25) implies that

$$\langle K\phi_1 \mid K\phi_2\rangle_{\mathscr{H}} = \langle K_{\mathbb{M}}T_B\phi_1 \mid K_{\mathbb{M}}T_B\phi_2\rangle_{\mathscr{H}_{\mathbb{M}}} \tag{3.26}$$

(and similarly for $\phi_1, \phi_2 \in S_A$ and T_A). Now it is known (cf. pages 89–90 in [KW91]) that

($A_{\mathbb{M}}$) $K_{\mathbb{M}}S_{\text{r-mov}}$ and $K_{\mathbb{M}}S_{\text{l-mov}}$ are dense in complex-linear subspaces $\mathscr{H}_{\text{r-mov}}$ and $\mathscr{H}_{\text{l-mov}}$ of $\mathscr{H}_{\mathbb{M}}$ (respectively);

($B_{\mathbb{M}}$) $K_{\mathbb{M}}S^R_{\text{r-mov}} + iK_{\mathbb{M}}S^R_{\text{r-mov}}$ is dense in $\mathscr{H}_{\text{r-mov}}$ and $K_{\mathbb{M}}S^R_{\text{l-mov}} + iK_{\mathbb{M}}S^R_{\text{l-mov}}$ is dense in $\mathscr{H}_{\text{l-mov}}$.

But Equation (3.26) implies that the obvious corresponding properties, i.e. (A) and (B) of Section 3.2, are inherited by $(K, \mathscr{H})$. In detail: define the complex Hilbert spaces $\mathscr{H}_1$ and $\mathscr{H}_2$ by $\mathscr{H}_1 := \overline{K_{\mathbb{M}}S_{\text{r-mov}} + iK_{\mathbb{M}}S_{\text{r-mov}}}$ with the restriction of the inner product in $\mathscr{H}_{\mathbb{M}}$, and $\mathscr{H}_2 := \overline{KS_B + iKS_B}$ with the restriction of the inner product in $\mathscr{H}$; define also the real-linear subspaces $M_1 := K_{\mathbb{M}}S_{\text{r-mov}} \subseteq \mathscr{H}_1$ and $M_2 := KS_B \subseteq \mathscr{H}_2$. Then, clearly, $M_1 + iM_1$ is dense in $\mathscr{H}_1$ and $M_2 + iM_2$ is dense in $\mathscr{H}_2$. We recall that both $K : S_0 \to \mathscr{H}$ and $K_{\mathbb{M}} : S_{0,\mathbb{M}} \to \mathscr{H}_{\mathbb{M}}$ are symplectic maps of real symplectic vector spaces (the symplectic forms on the respective codomains being given by twice the imaginary part of the inner product), and therefore are injective. It follows that $(K_{\mathbb{M}}\restriction_{S_{\text{r-mov}}})^{-1} : M_1 \to S_{\text{r-mov}}$ exists, and that the map

$$m : M_1 \to M_2 \quad \text{defined by} \quad m = K \circ T_B \circ (K_{\mathbb{M}}\restriction_{S_{\text{r-mov}}})^{-1}$$

is real-linear and bijective. In addition, Equation (3.26) entails precisely that m preserves inner products. By Lemma A.0.3, m then extends uniquely to a complex-linear isomorphism, which we call U_B, from $\mathscr{H}_1$ to $\mathscr{H}_2$. That is, we have a unique Hilbert space isomorphism

$$U_B : \overline{K_{\mathbb{M}}S_{\text{r-mov}} + iK_{\mathbb{M}}S_{\text{r-mov}}} \to \overline{KS_B + iKS_B}.$$

restricting to $m : K_{\mathbb{M}}S_{\text{r-mov}} \to KS_B$. Therefore, any subset of the Hilbert space on the right-hand side which corresponds, under U_B^{-1}, to a dense subset of the Hilbert space on the left-hand side, must itself be dense. Since $U_B^{-1}(KS_B) = m^{-1}(KS_B) = K_{\mathbb{M}}S_{\text{r-mov}}$ and we know [see ($A_{\mathbb{M}}$) above] that $K_{\mathbb{M}}S_{\text{r-mov}}$ is dense in $\overline{K_{\mathbb{M}}S_{\text{r-mov}} + iK_{\mathbb{M}}S_{\text{r-mov}}}$, it follows that KS_B is dense, i.e. we have proved property (A) of Section 3.2. The proof with $A \leftrightarrow B$, and the proof of property (B), are completely analogous.

The arguments leading to the final contradiction were already given in Section 3.2. □

We remark that the connection between the above proof and the heuristic discussion in Section 3.2 is made clearer if we note that, for any pair $\phi_1, \phi_2 \in S_{\text{r-mov}}$,

$$\langle K_{\mathbb{M}}\phi_1 \mid K_{\mathbb{M}}\phi_2\rangle_{\mathscr{H}_{\mathbb{M}}} = -\frac{1}{\pi} \lim_{\varepsilon \to 0^+} \int \frac{f_1(u_1) f_2(u_2)}{(u_1 - u_2 - i\varepsilon)^2} \, \mathrm{d}u_1 \mathrm{d}u_2,$$

where $\phi_1(t,x) = f_1(t-x)$ and $\phi_2(t,x) = f_2(t-x)$ (see e.g. Observation 6.1 and Proposition 7.2 in [DK87]). Equivalently, f_1, f_2 can be thought of geometrically as the restrictions of ϕ_1 and ϕ_2 (respectively) to $\mathcal{H}_B$.

3.5 Conclusions and future directions

In this chapter, we proved that there is no 'strongly boost-invariant globally-Hadamard' state – a notion which we have introduced in Definition 3.4.6 – for the massless wave equation to the left of an eternally uniformly accelerating mirror (with vanishing boundary conditions on the mirror) in 1+1 Minkowski spacetime. This, we argued, lends support to our conjecture that there is no isometry-invariant Hadamard state (in the sense which we have introduced in Section 3.3.2) for the Klein–Gordon equation defined on the region of Kruskal to the left of a surface of constant Schwarzschild radial coordinate in the right wedge (with vanishing boundary conditions on that box).

As was already suggested by Kay in [Kay15], the conjecture, if true, would suggest that there may be fundamental difficulties in attempting to provide a semi-classical description of a black hole confined to a spherical static box – a scenario which, as we recalled in Chapter 1, is of basic importance in discussions of black hole thermodynamics – by basing it on a portion of the Kruskal spacetime which includes the Killing horizons. Indeed, the paper [Kay15] pointed out a number of senses in which the right wedge horizons become (both classically and quantum mechanically) unstable for the our 1+1 model system with an accelerating mirror, and argued for a similar problem for our Klein–Gordon Kruskal system confined to a box. The tentative conclusion there was that any semi-classical description in the right wedge must break down at the right-wedge horizons – and it was suggested that it makes no sense to consider the spacetime as continuing to have any existence beyond these horizons.

One possible way around such a conclusion might be if there were one or more *non-stationary* Hadamard states on the region of Kruskal to the left of the box which are nevertheless stationary when restricted to the region of the right wedge to the left of the box. If this could be shown to also be impossible it would strengthen the above conclusion further, whereas if it would turn out to be possible it would perhaps undermine it. But it would still be strange if an equilibrium state of a black hole in a box were to be modelled mathematically by a state whose domain of definition includes the left wedge but which is not stationary when restricted to that left wedge. It would obviously be of interest to try to settle this question – or the obvious counterpart question for the wave equation in 1+1 Minkowski space to the left of an eternally uniformly accelerating mirror.

Converting our arguments in favour of the Kruskal conjecture into a rigorous proof would be of considerable interest and we now comment on the main challenges which lie ahead. First, one would need to gain full control over the relevant PDE theory in the presence of Dirichlet (and possibly of even more general) boundary conditions – this will be more thoroughly discussed in Section 5.4. Then, the validity of Property (b) on p. 103 and of its 'symplectic translation' on p. 110 would need to be rigorously assessed. Finally, the 'closure conjecture' on p. 111 would also need to be more precisely formulated and proved.

On a more technical level, it would be of interest to determine whether the notion of 'strongly boost-invariant globally-Hadamard' state can be relaxed while maintaining intact the conclusion of Theorem 3.4.7, particularly by dropping either of the requirements (ii) and (iii) in Definition 3.4.6 – see our comments in Footnote 24.

A topic which, although we have not highlighted this so far, is related to our investigations in this chapter, is that of deciding on a more stringent notion of what it might mean for a state to be (globally) Hadamard in quantised systems with boundary conditions. Definition 3.3.1 avoided dealing with the timelike boundary and did not prescribe how exactly the singularities of the two-point function ought to behave upon reaching the boundary. To find a Hadamard notion which, given a boundary condition, appropriately includes such a prescription, is an interesting and open problem. An investigation into this issue was recently initiated in [DF16].

Chapter 4

Filling a gap in [KW91]

4.1 Introducing the problem

This chapter is dedicated to precisely pointing out, and attempting to fill, a technical gap in some of the arguments of the paper [KW91]. We should stress that, while it was consideration of Theorem 3.4.2 which naturally led us to notice this gap (see also the Author's Declaration and Footnote 19), the discussion in this chapter is logically separate from the rest of the thesis – although it may well be that the methods used here will turn out to be useful in the attempt to rigorously prove our no-go conjecture of Sections 3.1 and 3.2 for Kruskal-in-a-box.

Our work in Chapter 2 (see in particular Section 2.5.5) has demonstrated that the spaces called S_A and S_B by Kay and Wald are quite large [indeed, as large as $C_0^\infty(\mathcal{H}_A)$ and $C_0^\infty(\mathcal{H}_B)$ respectively] and their elements do indeed have the properties claimed in that paper. In particular, it follows that linear symplectic forms σ_A and σ_B may be defined on S_A and on S_B (respectively), by using Equation (3.2) with $\mathscr{C}$ replaced by $\mathcal{H}_A$ and $\mathcal{H}_B$ respectively, cf. Equation (4.4) in [KW91]. Given this, one would also like for it to be the case that Equation (3.2), with $\mathscr{C}$ an *arbitrary* smooth Cauchy surface, defines an extension σ_0 of both σ_A and σ_B to the space $S_0 = S_A + S_B$. This extension would be bilinear and antisymmetric.

Assuming for a moment that this can be done, the gap to be filled in [KW91] is that the authors of that paper do not prove that the *a priori pre*-symplectic space $(S_0 = S_A + S_B, \sigma_0)$ is actually symplectic, i.e. it is not proven there that σ_0 (assumed defined on S_0) is non-degenerate. This gap needs to be filled, in particular, since the proof of Theorem 4.2 in [KW91] (which, we recall, establishes certain important uniqueness and KMS properties) relied on the presumed non-degeneracy of σ_0 on S_0, in the presence of which Lemma 4.1 in the same paper could be applied to yield the desired results.

As a matter of fact, Kay and Wald remarked, in the Note Added in Proof in [KW91] (cf. the discussion in Section 3.2), that while (S_A, σ_A) and (S_B, σ_B) are indeed, individually, well-defined linear symplectic spaces, elements of S_0 may not be regular enough for the extension σ_0 above to be easily definable by Equation (3.2). Indeed, we saw in great detail in Section 2.5.4 of this thesis that, while the elements of S_A and of S_B (respectively) are continuous, solve the Klein–Gordon equation in a weak sense, and are spatially compact, they are not guaranteed to be globally once differentiable as they can exhibit jump discontinuities in their derivatives in directions transverse to $\mathcal{H}_A$ and $\mathcal{H}_B$ (respectively). This creates awkwardness in defining σ_0 on S_0 by integration over Cauchy surfaces and, even more crucially, means that S_A, S_B and S_0 are not subspaces of the space S of smooth spatially compact solutions. As explained in that Note Added in Proof, the way to address this issue and its implications on the analysis in the main body of [KW91] – which incorrectly assumed the inclusions $S_A, S_B, S_0 \subseteq S$ – is to embed (S, σ) into a 'suitably larger' linear symplectic space $(\hat{S}, \hat{\sigma})$, while simultaneously restricting attention to better behaved, but 'large' enough, proper subspaces $\tilde{S}_A$, $\tilde{S}_B$ of S_A and S_B (respectively) such that $\tilde{S}_A$, $\tilde{S}_B$ and $\tilde{S}_0 \coloneqq \tilde{S}_A + \tilde{S}_B$ are contained in $\hat{S}$.

In view of these observations, the gap that's *really* to be filled is an analogous one to the one mentioned above: Is $(\tilde{S}_0, \hat{\sigma})$ a symplectic subspace of $(\hat{S}, \hat{\sigma})$? Actually, Kay and Wald do not give explicit arguments that even $(\tilde{S}_A, \hat{\sigma})$ and $(\tilde{S}_B, \hat{\sigma})$ are symplectic (sub)spaces – but once all definitions are in place this is not difficult to establish and, in Section 4.2, we will give a simple argument, which holds on the entire class of spacetimes considered in [KW91], that $(\tilde{S}_A, \hat{\sigma})$ and $(\tilde{S}_B, \hat{\sigma})$ are indeed symplectic.

Concerning the ramifications onto the (linear, spin-0) *quantum* theory considered in [KW91], filling this gap in full generality would allow to complete the proof of a certain modification of Theorem 4.2 given in the Note Added in Proof in [KW91]. As argued there, this modified theorem retains the physical significance of the (mathematically incorrect) original statement, while still allowing to carry out proofs of the (unchanged) uniqueness and non-existence theorems which are among the main achievements of that work (see p. 17 in this thesis). The modified theorem in question involves the 'natural extension' to the Weyl algebra (see Appendix A) $\hat{\mathscr{A}}$ over $(\hat{S}, \hat{\sigma})$, of a quasifree, isometry-invariant Hadamard state on the Weyl algebra $\mathscr{A}$ over (S, σ). Indeed, the existence of such a natural extension procedure is a requirement on any candidate choice of symplectic extension of $\hat{S}$, as we will illustrate in this chapter.

We will give two different lines of argument (the first of which applies to the massless Klein–Gordon equation, the second to more general Klein–Gordon equations with isometry-invariant potentials) each of which establishes that $(\tilde{S}_0, \hat{\sigma})$ is symplectic for certain spacetimes with bifurcate Killing horizons, including the notable cases of the

Minkowski and Kruskal spacetime. As explained in the next paragraph but one, both lines of arguments rely in particular – but not only! – on the existence of isometry-invariant Hadamard states for the Klein–Gordon field and spacetime under consideration.

As we mentioned in the introduction to this thesis, and recalled two paragraphs above the current one, it is argued in [KW91, Ch. 6] that, on the globally hyperbolic patches of Schwarzschild–de Sitter (with non-zero Schwarzschild mass) and of sub-extremal Kerr, there can be no isometry-invariant Hadamard state. The non-existence arguments there use the 'modified Theorem 4.2' mentioned above and thus assume that the relevant $(\tilde{S}_0, \hat{\sigma})$ are symplectic spaces. The situation is curious: on the one hand, these non-existence proofs have gaps because it was not shown that $(\tilde{S}_0, \hat{\sigma})$ is symplectic in those spacetimes; on the other hand, the two strategies known to us for establishing this rely on the existence of precisely the kind of quantum states which we are trying (with [KW91]) to prove do *not* exist! It is of course still possible to repair the non-existence proofs in [KW91], which were *reductiones ad absurdum*, under these circumstances: Assuming instead that isometry-invariant Hadamard states *did* exist for those spacetimes, if their existence can be shown to imply that $(\tilde{S}_0, \hat{\sigma})$ must be symplectic, then one would arrive at the same contradictions as Kay and Wald did. For the Klein–Gordon equation on both Kerr and of Schwarzschild–de Sitter (even allowing for isometry-invariant, but sufficiently regular, potential terms in place of a constant m^2 term), we will succeed in proving that the implication

A : "There exists an isometry-invariant Hadamard state" $\Longrightarrow$ **B** : "$(\tilde{S}_0, \hat{\sigma})$ is symplectic"

is true, and as just explained this leads to establishing that statement **A** is, in fact, false, but we will not be able to establish the truth value of statement **B**. When we refer, below, to 'filling the gap' in the case of Kerr and Schwarzschild–de Sitter, it needs to be borne in mind that this is the sense we intend.

The common starting point for both lines of argument is that, as we will show in Theorem 4.2.5 in Section 4.2, if

(i) there exists an isometry-invariant Hadamard state on $\mathscr{A}$, and
(ii) the entire spacetime coincides with the 'domain of C^{k-3}-determinacy' (with integer $k \geq 5$) of the bifurcate Killing horizon $\mathcal{H}_A \cup \mathcal{H}_B$ (this notion will be introduced in Definition 4.2.3),

then degenerate elements[1] of $(\tilde{S}_0, \hat{\sigma})$ are necessarily 'zero modes', i.e. are invariant under the isometries. Once this is established, it immediately follows that $(\tilde{S}_0, \hat{\sigma})$ is symplectic

[1] That is, elements whose pre-symplectic product with all other elements is zero.

for all those choices of spacetime (with bifurcate Killing horizon) and of Klein–Gordon operator such that (a) Conditions (i) and (ii) above are satisfied, and (b) there do not exist non-zero isometry-invariant solutions in the resulting $\tilde{S}_0$.[2] Notice that, as will also be explained in Section 4.2, our definitions of the spaces $\hat{S}$, $\tilde{S}_A$ and $\tilde{S}_B$ (and therefore also $\tilde{S}_0$) will be slightly different from (and, morally speaking, more general than) the ones originally presented in the Note Added in Proof in [KW91].

Sections 4.3 and 4.4 will present our two different lines of argument which allow to establish the absence of 'zero modes' in the cases of interest listed above. We remark that (a) our methods actually allow to prove a stronger statement, namely the absence of zero modes amongst solutions of sufficient regularity and not just amongst solutions in $\tilde{S}_0$,[3] and that (b) since neither of our lines of argument will require that Conditions (i) and (ii) hold, the 'cases of interest' *include* Schwarzschild–de Sitter and Kerr. However, we have not succeeded in ascertaining whether or not there are zero modes in $\tilde{S}_0$ in the case of the de Sitter spacetime.

4.2 Preliminaries and the common starting point

We begin by reviewing the definitions of the 'enlarged' symplectic space $(\hat{S}, \hat{\sigma})$ and of its subspaces $\tilde{S}_A, \tilde{S}_B \subset \hat{S}$. Actually, the definition of $(\hat{S}, \hat{\sigma})$ given in [KW91] is not entirely satisfactory: $\hat{S}$ is defined there to be the set of real-valued solutions to the Klein–Gordon equation with C_0^5 data on a Cauchy surface, $\mathscr{C}$, which contains the entire bifurcation surface Σ. It seems not totally clear whether, in this definition, $\mathscr{C}$ is a fixed Cauchy surface, chosen once and for all, or whether it is allowed to depend on the solution. Either way there would appear to be serious difficulties: If the Cauchy surface is allowed to depend on the solution, then there is no reason why $\hat{S}$ should be a vector space. If it is assumed to be fixed once and for all, then at least all solutions in the resulting $\hat{S}$ are C^2 on spacetime and have spatially compact support, so that a symplectic form $\hat{\sigma}$ on $\hat{S}$ can be defined in the obvious way – extending the symplectic form σ on S. However – a

[2]If one were to adopt the fiction explained in Section 3.2 that S_A, S_B and therefore $S_0 = S_A + S_B$ are subspaces of S then there is a simple (though of course false) argument showing that solutions ϕ in S which are symplectically orthogonal to the whole of S_0 are isometry invariant – and therefore, apparently, also that (S_0, σ) is a symplectic space if there do not exist non-zero isometry-invariant solutions in S_0. This argument does not need to appeal to the existence of any particular quantum state, and therefore Condition (ii) above is not needed. Namely, as explained on page 91 of [KW91] in the paragraph preceding Lemma 4.1, and under Condition (i) above, in a first step one easily shows that such a ϕ must be constant on each null generator of each horizon (we note that, in that passage of [KW91], it is stated erroneously that such a solution must be constant on each horizon, but presumably what was intended is what we wrote above). Then, in virtue of the fact that the isometries map solutions to solutions, and by the very definition of the domain of determinacy, one can conclude that the solution will be isometry-invariant.

[3]The term 'regularity' is here used informally to indicate conditions on both the differentiability and the asymptotic behaviour of the solution. We will not attempt to precisely identify 'minimal' regularity assumptions which are sufficient for ruling out zero modes.

statement to the contrary in [KW91, p. 133] notwithstanding – there is no reason why the action of the isometries on (S, σ) will extend to an action on $(\hat{S}, \hat{\sigma})$. This is because the pullback by the isometries of a solution in the thus defined $\hat{S}$ may fail to have C^5 Cauchy data on the chosen Cauchy surface $\mathscr{C}$.

In order to overcome these difficulties we now give what we consider to be an improved definition for $(\hat{S}, \hat{\sigma})$. It will also turn out that the resulting framework will allow us to tackle the main problem dealt with in this chapter in a clean and precise fashion. We begin by pointing out that the authors of [KW91] already suggested that an enlarged symplectic space of solutions $\hat{S}$ could alternatively be defined by using Cauchy data on $\mathscr{C}$ belonging to appropriate Sobolev spaces. In order to turn this idea into a rigorous recipe we will draw upon constructions and results as presented in a recent paper [BW15] by Bär and Wafo concerning the Cauchy and characteristic initial value problems for an arbitrary second-order normally hyperbolic operator P acting on distributional sections of a vector bundle over a globally hyperbolic spacetime. We refer to Section 2.4 of this thesis for a review of the notion of normal hyperbolicity, and to Appendix C for a review of Sobolev spaces on manifolds.

To wit, for any choice of foliation of the spacetime by smooth spacelike Cauchy surfaces, Bär and Wafo define spaces of spatially compact solutions to the homogeneous 'wave equation' which have 'finite k-energy' ($k \in \mathbb{R}$) along the foliation, and then show that these spaces do not actually depend on the choice of foliation. More precisely, given a choice of (smooth) Cauchy temporal function $t : M \to \mathbb{R}$ for the spacetime M, one can first define, for each $k \in \mathbb{R}$, spaces $C^\ell(t(M), H^k_{\mathrm{loc}}(\mathscr{C}_\bullet))$ of ℓ-times continuously differentiable sections of the bundle $\{H^k_{\mathrm{loc}}(\mathscr{C}_s)\}_{s \in t(M)}$, where $H^k_{\mathrm{loc}}(\mathscr{C}_s)$ is the space of locally Sobolev sections of the restriction of the original vector bundle to the Cauchy surface $\mathscr{C}_s = t^{-1}\{s\}$. As explained in [BW15], these spaces can then be straightforwardly embedded as subspaces of distributional sections of the original vector bundle over M. It is therefore legitimate to further restrict attention to elements of $C^\ell(t(M), H^k_{\mathrm{loc}}(\mathscr{C}_\bullet))$ which correspond to distributional sections with spatially compact support on M; this way, one obtains the spaces denoted by $C^\ell_{sc}(t(M), H^k(\mathscr{C}_\bullet))$ in [BW15]. The space of *finite k-energy sections* (with respect to t) is then defined by

$$\mathscr{F}\mathscr{E}^k_{sc}(t) = C^0_{sc}(t(M), H^k(\mathscr{C}_\bullet)) \cap C^1_{sc}(t(M), H^{k-1}(\mathscr{C}_\bullet))$$

(this is Definition 1 in [BW15], though we have suppressed some of the notation there). The main result (which is Corollary 18 in [BW15]) for the purposes of the present chapter is the fact that, for any two Cauchy temporal functions t, t',

$$\mathscr{F}\mathscr{E}^k_{sc}(t) \cap \ker P = \mathscr{F}\mathscr{E}^k_{sc}(t') \cap \ker P \tag{4.1}$$

(having omitted an embedding into the space of distributional sections from both sides in the interest of notational clarity). One can thus unambiguously speak of a space $\mathscr{FE}^k_{sc}(\ker P)$ of finite k-energy *solutions* of the 'homogeneous wave equation' for P – with the property that $\mathscr{FE}^k_{sc}(\ker P) = \mathscr{FE}^k_{sc}(t) \cap \ker P$ for all Cauchy temporal functions t. Topologizing $\mathscr{FE}^k_{sc}(\ker P)$ in the manner discussed in Section 2.7.6 of [BW15], one has that the spatially compact *smooth* solutions of $Pu = 0$ form a dense subset of $\mathscr{FE}^k_{sc}(\ker P)$. Furthermore, in a four-dimensional spacetime, by Corollary 20 in [BW15] and the Sobolev embedding theorem, if $\mathbb{N} \ni k \geq 5$ then $\mathscr{FE}^k_{sc}(\ker P) \subset C^{k-3}(M) \subset C^2(M)$.

In view of the above (and returning to the specific framework of [KW91]) we give our alternative definition of the space $\hat{S}$ as one of the spaces $\hat{S}^k = \mathscr{FE}^k_{sc}(\ker P)$, with $\mathbb{N} \ni k \geq 5$ to be determined later. It is then to be understood that, unless stated otherwise, any statement involving '$\hat{S}^k$' (and the later defined '$\tilde{S}^k_A$', '$\tilde{S}^k_B$' and '$\tilde{S}^k_0$') in the remainder of this chapter will hold for any choice of $\mathbb{N} \ni k \geq 5$. In the next Proposition we prove that the 'obvious' antisymmetric bilinear form $\hat{\sigma}$ (which we refrain from denoting instead by the more cumbersome '$\hat{\sigma}^k$') on $\hat{S}^k$ is indeed non-degenerate and thus a symplectic form.[4]

Proposition 4.2.1. *For any $\mathbb{N} \ni k \geq 5$, the antisymmetric bilinear form $\sigma : \hat{S}^k \times \hat{S}^k \to \mathbb{R}$ defined, for any smooth spacelike Cauchy surface $\mathscr{C}$ with future-directed unit normal vector field ν, by*

$$\hat{\sigma}(\phi, \psi) = \int_{\mathscr{C}} (\phi \nabla_\nu \psi - \psi \nabla_\nu \phi) \, \mathrm{d}\mu_{\mathscr{C}} \tag{4.2}$$

(where $\mathrm{d}\mu_{\mathscr{C}}$ denotes the volume density on $\mathscr{C}$ coming from the pullback of the metric g to $\mathscr{C}$) is non-degenerate. That is, $(\hat{S}^k, \hat{\sigma})$ is a (real) symplectic vector space.

Proof. We will actually prove the stronger statement that, for any $\mathbb{N} \ni k \geq 5$, every $\phi \in \hat{S}^k$ such that $\hat{\sigma}(\phi, \psi) = 0$ for all ψ in $S = C^\infty_{\mathrm{sc}}(M) \cap \ker P \subset \hat{S}^k$, must be identically zero. Namely, suppose such a ϕ exists. Let t be a smooth Cauchy temporal function adapted to $\mathscr{C}$ in the sense that $\mathscr{C}$ is one of its level sets, say $\mathscr{C} = t^{-1}\{\tau\}$ for some $\tau \in \mathbb{R}$. We recall the isomorphism of topological vector spaces

$$\hat{S}^k \cong H^k_c(\mathscr{C}) \oplus H^{k-1}_c(\mathscr{C}) \tag{4.3}$$

given by $u \mapsto (u\restriction_{\mathscr{C}}, (\nabla_\nu u)\restriction_{\mathscr{C}})$ and proven as Corollary 14 in [BW15]. Under this isomorphism, the dense subspace S of $\hat{S}^k$ is identified with the dense subspace $C^\infty_0(\mathscr{C}) \oplus C^\infty_0(\mathscr{C})$ of $H^k_c(\mathscr{C}) \oplus H^{k-1}_c(\mathscr{C})$, and the form defined by Equation (4.2) can be written as

$$\hat{\sigma}(\phi, \psi) = \int_{\mathscr{C}} (fp' - pf') \, \mathrm{d}\mu_{\mathscr{C}} \tag{4.4}$$

[4]We remark that, presumably, a similar argument to the one we will make was implicitly assumed in [KW91].

where (f, p) and (f', p') are the images of ϕ and ψ (respectively) under the isomorphism (4.3). In turn, we may (and will) view the right-hand side of Equation (4.4) as

$$\langle f \mid p' \rangle_{\mathscr{C}} - \langle p \mid f' \rangle_{\mathscr{C}} \tag{4.5}$$

where $\langle \cdot \mid \cdot \rangle_{\mathscr{C}}$ denotes the standard inner product on $L^2_{\mathbb{C}}(\mathscr{C}, \mathrm{d}\mu_{\mathscr{C}})$ (though we could equally well use, instead, the inner product on the corresponding L^2 space of *real-valued* functions). The original assumption about ϕ corresponds to the assumption that the object in (4.5) vanishes for all f', $p' \in C_0^\infty(\mathscr{C})$. Since the latter space is dense in $L^2_{\mathbb{C}}(\mathscr{C}, \mathrm{d}\mu_{\mathscr{C}})$, there exist sequences $(f_n)_n$ and $(p_n)_n$ which tend to $-p$ and f in the L^2 sense (respectively), and such that each f_n, $p_n \in C_0^\infty(\mathscr{C})$. Denoting by ψ_n the element of S corresponding to the pair (f_n, p_n), we have

$$\hat{\sigma}(\phi, \psi_n) = 0 \quad \forall\, n \in \mathbb{N}$$

and at the same time

$$\hat{\sigma}(\phi, \psi_n) = \langle f \mid p_n \rangle_{\mathscr{C}} - \langle p \mid f_n \rangle_{\mathscr{C}} \longrightarrow \|f\|^2_{\mathscr{C}} + \|p\|^2_{\mathscr{C}} \quad \text{as } n \to \infty.$$

Together, these two facts clearly imply that f and p vanish identically, and therefore that so does ϕ. □

The 'awkwardness' mentioned above in naturally defining an action on $\hat{S}$ by the group of isometries disappears when we use the thus defined $\hat{S}^k$: Indeed, this follows from Equation (4.1) together with the fact that the composition of a Cauchy temporal function with an isometry preserving the time orientation yields another Cauchy temporal function. Finally a quasi-free Hadamard state on the Weyl algebra $\mathscr{A}$ over (S, σ) will possess a natural quasi-free extension to the Weyl algebra $\hat{\mathscr{A}}^k$ over $(\hat{S}^k, \hat{\sigma})$ by the same reasoning as in [KW91]. Namely, one has the following result.

Proposition 4.2.2. *Let ω be a quasi-free, globally*[5] *Hadamard state on $\mathscr{A} = \mathscr{W}(S, \sigma)$.*

[5]The reader might wonder why we insist on adding the word 'global' at this point, in view of Radzikowski's resolution of Kay's conjecture. This is because Radzikowski's local-to-global theorem for Hadamard stated of the Klein–Gordon field, Theorem 9.2 in [RV96], resolves Kay's conjecture whenever it is also known that the spacetime smeared two-point function of the state defines a continuous (bi)distribution (the adjective 'continuous' here and elsewhere is perhaps a pleonasm when referred to a (bi)distribution, but we keep it for emphasis). However, we are working in the original framework of [KW91] where no such continuity assumption was made. We are at present unaware of any explicit result in the literature proving that the two-point functions of a globally Hadamard state, with 'globally Hadamard' taken in the strict sense of [KW91], always defines a continuous bidistribution. Equation (4.6) in the main text suggests a strategy for arriving at such a result: Namely, the two-point function of a globally Hadamard state is given by a composition

$$C_0^\infty(M) \times C_0^\infty(M) \to S \times S \to \mathbb{C}$$

where the first arrow is given by applying the causal propagator E to each slot and the second arrow is defined by the right-hand side of Equation (4.6). By virtue of the Schwartz kernel theorem (see e.g.

Let $(K, \mathscr{H})$ be its associated one-particle structure. Then, for any sequence $(\phi_n)_n$ in S which tends to $\hat{\phi} \in \hat{S}^k$ (for $k \geq 5$) in the latter's topology, the sequence $(K\phi_n)_n$ has a limit which depends only on $\hat{\phi}$. We denote the resulting map

$$\hat{S}^k \to \mathscr{H}, \quad \hat{\phi} \mapsto \lim_{n\to\infty} K\phi_n$$

by $\hat{K}$, so that the statement above is simply that $\hat{K}$ is sequentially continuous. Finally, $(\hat{K}, \mathscr{H})$ is the one-particle structure associated to a quasi-free state $\hat{\omega}$ on $\hat{\mathscr{A}}^k = \mathscr{W}(\hat{S}^k, \hat{\sigma})$, which (following [KW91]) we call the natural extension *of ω from $\mathscr{A}$ to $\hat{\mathscr{A}}^k$.*

Proof. (Sketch) It was proved in [BW15] (Corollary 20) that there is a continuous embedding

$$\hat{S}^k \to H^k_{\text{loc}}(M)$$

into the Fréchet space of locally Sobolev functions on M. Therefore, any sequence $(\phi_n)_n$ in S which tends to $\hat{\phi} \in \hat{S}^k$ in the topology of $\hat{S}^k$ does so also in the topology of $H^k_{\text{loc}}(M)$. Using the isomorphism (4.3) for any smooth spacelike Cauchy surface $\mathscr{C}$, and since the spaces $H^j_c(\mathscr{C})$ are defined as strict inductive limits of spaces of Sobolev functions supported in compact sets of $\mathscr{C}$, it follows[6] that the Cauchy data for the ϕ_n have support contained in a common compact set $K \subseteq \mathscr{C}$ and therefore, by finite speed of propagation for the Cauchy problem, that $\operatorname{supp}\phi_n \subseteq J(K)$ for all n. Therefore, the intersections

$$\operatorname{supp}(\phi_n) \cap \mathscr{C}_1 \quad \text{and} \quad \operatorname{supp}(\phi_n) \cap \mathscr{C}_2$$

are all contained in compact sets K_1, K_2 of $\mathscr{C}_1$ and $\mathscr{C}_2$ respectively. We now pick two relatively compact open neighbourhoods A_1, A_2 (in M) of K_1 and K_2 (respectively), and two smooth, compactly supported bump functions χ_1, χ_2 with $\chi_1 \equiv 1$ on $\overline{A_1}$ and

Theorem 5.2 in [Hör90a]), the two-point function (i.e, this composition) is representable by a (unique) continuous bidistribution if and only if it is a separately continuous bilinear form. In turn (see e.g. Theorem 2.1.4 in [Hör90a]), one need only check *sequential* continuity at zero of the maps given, for fixed $\phi \in S$, by the composition

$$C^\infty_0(M) \ni F \mapsto (EF, \phi) \mapsto \langle KEF \mid K\phi \rangle_{\mathscr{H}} \in \mathbb{C}$$

(notice that by the canonical commutation relations one does not need to also check continuity 'in the other slot'). By Definition 3.4.6 and Proposition 3.4.8 in [BGP07], if a sequence $(F_n)_n$ of test functions converges to zero in the standard topology of $C^\infty_0(M)$, then the sequence $(EF_n)_n$ converges to the zero function in the standard Frechét space topology of $C^\infty(M)$ and there exists a compact set $K \subset M$ such that each $\operatorname{supp}(EF_n) \subseteq J(K)$. Thus if one can prove that, for any compact set K and sequence $(\varphi_n)_n$ in $C^\infty_{\text{sc}}(M)$ such that $\operatorname{supp}(\varphi_n) \subseteq J(K)$ and $(\varphi_n)_n \to 0$ in $C^\infty(M)$, the right-hand side of Equation (4.6), with φ_n instead of ϕ_1, tends to zero for any $\phi_2 \in S$, then indeed one has the result that there are no globally Hadamard distributions in the Kay–Wald sense which do not define continuous bidistributions. However, we shall not attempt to turn this strategy into a rigorous proof in this thesis.

[6]If X is the strict inductive limit of an sequence of spaces $\{X_n\}$ with X_n closed in X_{n+1} for each n, then a sequence $(x_i)_{i\in\mathbb{N}} \subset X$ converges in X if and only there is an n so that $(x_i)_{i\in\mathbb{N}} \subset X_n$ and $(x_i)_{i\in\mathbb{N}}$ converges in X_n.

$\chi_2 \equiv 1$ on $\overline{A_2}$. Then the convergence of $(\phi_n)_n$ in $H^k_{\rm loc}(M)$ implies that, for $i = 1, 2$,

$$(\chi_i \phi_n)_n \to \chi_i \hat{\phi} \quad \text{in } H^k(\operatorname{supp}(\chi_i)).$$

By using a doubling procedure as described for instance in Section 2.6.2 of [BW15] (or otherwise), one can satisfy the hypotheses of (a version of) the Sobolev embedding theorem. Indeed, the doubling produces a compact Riemannian manifold without boundary $\mathscr{M}_i = (M_i, e_i)$ and a continuous embedding of $H^k(\operatorname{supp}(\chi_i))$ into the standard Sobolev space $H^k(\mathscr{M}_i)$. Since we work in four spacetime dimensions and we are taking $k \geq 5$, it follows that the sequence $(\chi_i \phi_n)_n$ converges in the Banach space $C^2(\mathscr{M}_i)$. Thus in particular ϕ_n and its derivatives up to order 2 converge uniformly – to $\hat{\phi}$ and its corresponding derivatives – on the compact sets $K_1 \subseteq \mathscr{C}_1$ and $K_2 \subseteq \mathscr{C}_2$.

The remainder of this proof (which will be more sketchy) is borrowed from [KW91], cf. the last full paragraph in p. 133 there. Namely, as it was proved there (see p. 85 in [KW91]), for $\omega, K, \mathscr{H}$ as in the statement of the Theorem,

$$\langle K\phi_1 \mid K\phi_2 \rangle_{\mathscr{H}} = \lim_{\varepsilon \to 0^+} \int_{\mathscr{C}_1 \times \mathscr{C}_2} \Lambda^{T,n}_\varepsilon(x_1, x_2) \overleftrightarrow{\nabla}_{1a} \overleftrightarrow{\nabla}_{2b} \phi_1(x_1) \phi_2(x_2) \, \mathrm{d}S^a_1 \, \mathrm{d}S^b_2 \tag{4.6}$$

where ϕ_1, $\phi_2 \in S$, $\mathscr{C}_1$, $\mathscr{C}_2$ are smooth spacelike Cauchy surfaces one of which lies entirely in the chronological future of the other, and the kernel $\Lambda^{T,n}_\varepsilon(x_1, x_2)$ (with T a time function and $n \in \mathbb{N}$) is defined on a causal normal neighbourhood $\mathscr{N}$ of $\mathscr{C}_1$ such that $\mathscr{C}_2 \subset \mathcal{N}$. By an integration by parts argument presented in the first full paragraph under Equation (B.11) in [KW91], the right-hand side of Equation (4.6) can be rewritten as an integral over $\mathscr{C}_1 \times \mathscr{C}_2$ of a sum of terms each involving a product of a locally integrable function of x_1 and x_2 times ϕ_1 and ϕ_2 or their derivatives of order at most 2. By using $\phi_m - \phi_n$ instead of ϕ_1 and ϕ_2, we see that the fact that $(\phi_n)_n$ is Cauchy in C^2 norm, as proved above, readily implies that

$$\|K(\phi_m - \phi_n)\|^2_{\mathscr{H}} \leq C \, \|\phi_m - \phi_n\|_{C^2(K_1)} \, \|\phi_m - \phi_n\|_{C^2(K_2)} \tag{4.7}$$

for some $C \geq 0$. Therefore, $(K\phi_n)_n$ is Cauchy and thus has a limit in $\mathscr{H}$, as we wished to prove. Independence of the limit from the choice of sequence $(\phi_n)_n$ is easily seen by consideration of the fact that if $(\phi'_n)_n$ is another sequence which tends to $\hat{\phi}$ then, if K'_1 and K'_2 are the compact sets of $\mathscr{C}_1$ and $\mathscr{C}_2$ corresponding to K_1 and K_2 above, then we have $\|\phi_n - \phi'_n\|_{C^2(K_i \cup K'_i)} \to 0$ as $n \to \infty$ for $i = 1, 2$, which by a similar inequality to (4.7) shows that $K(\phi_n - \phi'_n) \to 0$ as $n \to \infty$.

It is a very easy verification that $(\hat{K}, \mathscr{H})$ satisfies the requirements of a one-particle structure over $(\hat{S}^k, \hat{\sigma})$. Finally, the natural extension $\hat{\omega}$ is defined by $\hat{\omega}[W(\hat{\phi})] = \exp(-\frac{1}{2}\|\hat{K}\hat{\phi}\|^2_{\mathscr{H}})$. □

Just as in [KW91], in the case of a Klein–Gordon equation with isometry-invariant potential, spaces of solutions $\tilde{S}_A^k$ and $\tilde{S}_B^k$ can now be defined in such a way that they are at the same time 'large' subspaces of S_A and S_B (respectively) and suitable subspaces of $\hat{S}^k$. Notice that our $\tilde{S}_A^5$ and $\tilde{S}_B^5$ coincide with the $\tilde{S}_A$ and $\tilde{S}_B$ in [KW91] (respectively). The key observation is that any function in S_A whose restriction to the A-horizon is of the form $\partial^k(U^k g)/\partial U^k$, for some compactly supported and smooth function g on the A-horizon, has C_0^k data on any Cauchy surface $\mathscr{C}$ containing the bifurcation surface. A similar statement (with U replaced by V) holds for functions in S_B. Denoting the linear spaces of such solutions by $\tilde{S}_A^k$ and $\tilde{S}_B^k$, this means that, for $k \geq 5$, $\tilde{S}_A^k, \tilde{S}_B^k \subset \hat{S}^k$ as desired. We also let $\tilde{S}_0^k = \tilde{S}_A^k + \tilde{S}_B^k$.

To show that the restriction of $\hat{\sigma}$ to $\tilde{S}_A^k$ is non-degenerate, recall that any solution in $\tilde{S}_A^k$ is smooth and of compact support when restricted to the A-horizon. Suppose ϕ is a degenerate element in $(\tilde{S}_A^k, \hat{\sigma})$, i.e. suppose that $\hat{\sigma}(\phi, \psi) = 0$ for all $\psi \in \tilde{S}_A^k$. Denote by f the restriction of ϕ to the A-horizon. Starting from Equation (4.4) in [KW91] and integrating by parts $k+1$ times one readily sees that $U^k \frac{\partial^{k+1} f}{\partial U^{k+1}} = 0$. Therefore, since f is smooth, $\frac{\partial^{k+1} f}{\partial U^{k+1}} = 0$ everywhere on the horizon. So f is a polynomial of degree at most k in the affine parameter U, whose coefficients are (compactly supported, smooth) functions of the coordinates on the bifurcation surface. But no such polynomial can have compact support on the A-horizon unless it's zero. Thus there are no non-zero degenerate elements and $(\tilde{S}_A^k, \hat{\sigma})$ is a symplectic space. $(\tilde{S}_B^k, \hat{\sigma})$ is also a symplectic space by a similar argument.

We now turn to what we already called the 'common starting point' for both our strategies: That is, we aim to show that, under Conditions (i)–(ii) above, any degenerate element of $(\tilde{S}_0^k, \hat{\sigma})$ is necessarily isometry-invariant. Before giving a proof of this fact, we must introduce the notion of *domain of C^n-determinacy* (with respect to the Klein–Gordon operator) of a subset $U \subseteq M$, with $n \in \mathbb{N} \cup \{\infty\} \cup \{\omega\}$, which appears in the formulation of our Condition (ii).

Definition 4.2.3. The 'domain of C^n-determinacy' $\mathscr{D}^{(n)}[U]$ (with respect to the Klein–Gordon operator) of $U \subseteq M$ is the set of points $p \in M$ such that every C^n solution which vanishes on U must vanish at p.

Remark. Kay and Wald's 'domain of determinacy' (cf. pages 64–65 in [KW91]) coincides with what we would call the 'domain of C^∞-determinacy.' It is also clear that the inclusions $\mathscr{D}^{(l)}[U] \subseteq \mathscr{D}^{(m)}[U]$ hold for $l \leq m$.

The following Lemma will be used in the proof of the 'common starting point', Theorem 4.2.5 below.

Lemma 4.2.4. *Let $k \in \mathbb{N}$ with $k \geq 5$, and let ω be a quasi-free Hadamard state on $\mathscr{A}$, with associated one-particle structure $(K, \mathscr{H})$. Let $\hat{K} : \hat{S}^k \to \mathscr{H}$ be the 'natural' extension* *of $K : S \to \mathscr{H}$ according to Proposition 4.2.2. Then the one-parameter unitary group $U(t)$ on the one-particle Hilbert space $\mathscr{H}$ for ω which implements the 'time translations' $\mathscr{T}(t) : S \to S$ also implements the 'time translations' $\hat{\mathscr{T}}(t) : \hat{S}^k \to \hat{S}^k$, i.e.*

$$U(t)\hat{K} = \hat{K}\hat{\mathscr{T}}(t). \tag{4.8}$$

Proof. For any $\hat{\psi} \in \hat{S}^k$, by definition

$$\hat{K}\hat{\psi} = \lim_{n\to\infty} K\psi_n \tag{4.9}$$

where $(\psi_n)_{n\in\mathbb{N}}$ is a sequence of solutions in S which converges to $\hat{\psi}$ in the topology for $\hat{S}^k = \mathscr{F}\mathscr{E}^k_{sc}(\ker P)$ given in [BW15] [and any such sequence yields the same limit on the right-hand-side of Equation (4.9)]. Since $U(t)$ is bounded,

$$\begin{aligned} U(t)\hat{K}\hat{\psi} &= U(t)\left[\lim_{n\to\infty} K\psi_n\right] \\ &= \lim_{n\to\infty} U(t)[K\psi_n] \\ &= \lim_{n\to\infty} K[\mathscr{T}(t)\psi_n]. \end{aligned}$$

The claim then follows since it is clear that $(\mathscr{T}(t)\psi_n)_{n\in\mathbb{N}}$ is a sequence in S which tends to $\hat{\mathscr{T}}(t)\hat{\psi}$ in the topology of $\hat{S}^k$. □

We can now give the statement and proof of the 'common starting point' mentioned on p. 137.[7]

Theorem 4.2.5. *Let $k \in \mathbb{N}$ with $k \geq 5$, and supposed the following conditions hold:*

(i) there exists an isometry-invariant Hadamard state on $\mathscr{A}$, and

(ii) the entire spacetime coincides with the domain of C^{k-3}-determinacy of the bifurcate Killing horizon $\mathcal{H}_A \cup \mathcal{H}_B$ according to Definition 4.2.3.

Then any solution in $\tilde{S}^k_0$ which is symplectically orthogonal to $\tilde{S}^k_0$ is isometry-invariant.

Proof. A proof was given in [KW91, p. 135] (under the unnecessary extra assumption that $k = 5$) that if Condition (i) above holds and Condition (ii) is replaced by

(ii') the entire spacetime coincides with the domain of C^∞-determinacy of the bifurcate Killing horizon $\mathcal{H}_A \cup \mathcal{H}_B$,

then any solution ϕ in S with the property that $\hat{\sigma}(\phi, \phi_0) = 0 \ \forall \ \phi_0 \in \tilde{S}_0$ must be isometry-invariant on the entire spacetime. We now describe how those arguments can

[7] Note that it was perhaps suggested in [KW91] that an even stronger result than Theorem 4.2.5 should hold, namely that (under the same hypotheses) any solution in $\hat{S}$ (rather than just $\tilde{S}_0$) which is symplectically orthogonal to $\tilde{S}_0$ is isometry-invariant. However, the integration by parts argument used in the proof of Theorem 4.2.5 does not straightforwardly adapt in that case, due to the fact that the restrictions of elements in $\hat{S}^k$ to either horizon are in general only in C^{k-3}.

be adapted for our purposes. Let $\psi_0 \in \tilde{S}_0^k$ be such that $\hat{\sigma}(\psi_0, \phi_0) = 0 \ \forall \ \phi_0 \in \tilde{S}_0^k$. Then, in particular, ψ_0 is symplectically orthogonal to the whole of $\tilde{S}_A^k$ and to the whole of $\tilde{S}_B^k$. We would like to apply an integration by parts argument similar to the one used above in the proof that $(\tilde{S}_A^k, \hat{\sigma})$ and $(\tilde{S}_B^k, \hat{\sigma})$ are symplectic to conclude that the restrictions of ψ_0 to $\mathcal{H}_A$ and $\mathcal{H}_B$ are polynomials of degree at most k in U and V respectively, whose coefficients in both cases are functions on the bifurcation surface. However, the restriction of ψ_0 to either horizon, while in C^k, may fail to be C^{k+1} at the bifurcation surface. To overcome this difficulty one can apply our integration by parts argument separately, first to symplectic products of ψ_0 with solutions in $\tilde{S}_A^{L,k}$ and then to symplectic products of ψ_0 with solutions in $\tilde{S}_A^{R,k}$, where

$$\tilde{S}_A^{L/R,k} = \left\{ \phi \in \tilde{S}_A^k \,\middle|\, \phi\text{'s data on } \mathcal{H}_A \text{ is of the form } \frac{\partial^k (U^k g)}{\partial U^k} \text{with } g \in C_0^\infty\left(\mathcal{H}_A^{L/R}\right) \right\}$$

and we also define the spaces $\tilde{S}_B^{L/R,k}$ in a similar fashion. Since the restrictions of ψ_0 to $\mathcal{H}_A^L$ and to $\mathcal{H}_A^R$ are smooth, it indeed follows that each of them is a polynomial in U of degree at most k whose coefficients are smooth functions on the bifurcation surface. (The fact that ψ_0 is C^k across the bifurcation surface will imply that the first k of these coefficients agree.) Analogous results clearly hold with A replaced by B and U replaced by V. Now let $\tilde{\mathscr{T}}(t)$ denote the time translation operator on $\tilde{S}_0^k$. We define a generalised version of the operator $Q(t)$ in Equation (N.4) in [KW91], namely

$${}^kQ(t) = \prod_{l=-k}^{k} [\tilde{\mathscr{T}}(t) - e^{l\kappa t}] : \tilde{S}_0^k \to \tilde{S}_0^k.$$

Just as in [KW91] one sees that since, for any j with $0 \le j \le k$, U^j is annihilated by $[\tilde{\mathscr{T}}(t) - e^{j\kappa t}]$, ${}^kQ(t)\psi_0$ vanishes on $\mathcal{H}_A$. Similarly, for any j with $0 \le j \le k$, V^j is annihilated by $[\tilde{\mathscr{T}}(t) - e^{-j\kappa t}]$, which implies that ${}^kQ(t)\psi_0$ vanishes on $\mathcal{H}_B$. Therefore ${}^kQ(t)\psi_0 = 0$ on $\mathcal{H}_A \cup \mathcal{H}_B$. Now, if ψ_0 were smooth – as is ϕ in the corresponding arguments in [KW91] – the very definition of the domain of (C^∞-)determinacy of a set would immediately imply that, under condition (ii') above, ${}^kQ(t)\psi_0 = 0$ throughout the spacetime. However, while ψ_0 is certainly everywhere C^{k-3}, it could fail to be everywhere smooth. Thus one cannot conclude that ${}^kQ(t)\psi_0 = 0$ if Condition (ii') alone holds. However, under the stronger Condition (ii) – namely under the assumption that the entire spacetime coincides with the domain of C^{k-3} determinacy of the bifurcate Killing horizon – the vanishing of ${}^kQ(t)\psi_0$ on $\mathcal{H}_A \cup \mathcal{H}_B$ does imply that ${}^kQ(t)\psi_0 = 0$ on the entire spacetime.

At this point, again just as in [KW91], we invoke Condition (i), i.e. the existence of an isometry-invariant Hadamard state on the Weyl algebra $\mathscr{A}$ over (S, σ). Without loss of generality, we can assume this state to be quasi-free and denote its associated one-particle Hilbert space structure by $(K, \mathscr{H})$. Let $\hat{K} : \hat{S}^k \to \mathscr{H}$ be the 'natural extension' of $K : S \to \mathscr{H}$. By Lemma 4.2.4, an equation analogous to Equation (N.6) in [KW91] holds. Namely:

$$ {}^k P(t) \hat{K} \psi_0 = 0 \tag{4.10} $$

where

$$ {}^k P(t) = \prod_{l=-k}^{k} [U(t) - e^{l\kappa t}]. $$

The desired result that ψ_0 is isometry-invariant then follows by straightforwardly adapting the arguments given in the first paragraph on page 136 in [KW91] – in particular, using in the final step the fact that $\hat{K} : \hat{S}^k \to \mathscr{H}$ is injective, which in turn follows from the property $2\,\mathrm{Im}\langle \hat{K}\hat{\psi} | \hat{K}\hat{\phi} \rangle = \hat{\sigma}(\hat{\psi}, \hat{\phi})\ \forall\ \hat{\psi}, \hat{\phi} \in \hat{S}^k$ and the nondegeneracy of $\hat{\sigma}$ established in Proposition 4.2.1. □

Corollary 4.2.6. *For any $\mathbb{N} \ni k \geq 5$, $(\tilde{S}^k_0, \hat{\sigma})$ is a symplectic space if Conditions (i)–(ii) are satisfied and there are no non-zero isometry-invariant solutions in $\tilde{S}^k_0$.* □

We end this section by discussing for which cases of physical interest our Conditions (i) and (ii) are known to hold. First of all, it is not hard to see that, for any Klein–Gordon equation with isometry-invariant potential, there is no difficulty in adapting the arguments given on pages 64-65 of [KW91] – which are based on the characteristic initial value formulations for the sets $J^{\pm}(\Sigma)$ and on an application of Holmgren's uniqueness theorem – to our $\mathscr{D}^{(n)}[\mathcal{H}_A \cup \mathcal{H}_B]$ for any $n \geq 2$ instead of Kay and Wald's $\mathscr{D}[\mathcal{H}_A \cup \mathcal{H}_B]$. It follows that, for any $k \geq 5$, Condition (ii) holds, for example, on Minkowski spacetime, on the Kruskal spacetime, on de Sitter spacetime, and on the globally hyperbolic patches of Kerr and Schwarzschild–de Sitter considered in [KW91]. As for Condition (i), it is known that isometry-invariant Hadamard states exist for the massive or massless Klein–Gordon field on both Minkowski spacetime and [San15] Kruskal spacetime, and for the massive or massless conformally coupled Klein–Gordon field on de Sitter spacetime [CT68, BD78].[8] On the other hand, the paper [KW91] contains proofs that no such

[8]The case of the massless *minimally coupled* Klein–Gordon field on de Sitter seems more subtle. While it was proved in [All85] that no fully de Sitter invariant state (Hadamard or not) exists, Hadamard states do exist [All85, AF87] which are invariant under the subgroups E(3) and O(4) of the de Sitter group (and it is presumed [AF87] that O(1, 3)-invariant Hadamard states also exist). However, none of these subgroups contain the 'de Sitter boost' isometries to which our analysis applies and we conjecture that there is no boost-invariant Hadamard state. Our grounds for this conjecture are that, were there to exist such a state, then it is plausible that its restriction to the 'right-wedge' (which is of course a static spacetime when the time evolution is taken to be the restriction of the de Sitter boost isometries) would be a KMS state. But it is known [Pol90] that (for reasons of bad infra-red behaviour) on the right-wedge, no ground state exists for this time evolution. Also by Lemma 6.2 in [KW91], we know quite generally that if a stationary linear Bose dynamical system admits a KMS state then it also admits a ground state, and thus there would be a contradiction. There are a number of obstacles, however, to making this argument rigorous: Even under the fiction explained in Section 3.2 we would only be able to rely on Theorem 4.2 of [KW91] to prove the KMS property on the subalgebra of the Weyl algebra for the right

states can exist in Kerr or Schwarzschild–de Sitter, although, as we explained in the fourth paragraph of this Appendix, these proofs have a gap that needs filling and that we will fill below by showing that $(\tilde{S}_0, \hat{\sigma})$ is symplectic under the (for these spacetimes, counter-factual) assumption that an isometry invariant Hadamard state exists.

4.3 'Decay along the horizons' strategy

Let us now present our first line of argument for showing the non-existence of isometry-invariant solutions in $\tilde{S}_0^k$. The idea is as follows: suppose (M, g) is a globally hyperbolic spacetime with a bifurcate Killing horizon $\mathcal{H}_A \cup \mathcal{H}_B$ and bifurcation surface Σ, and suppose that there exists a Cauchy surface $\mathscr{C}$ for M which contains Σ and such that $\mathscr{C} = \Sigma \cup (\mathscr{C} \cap \mathscr{L}) \cup (\mathscr{C} \cap \mathscr{R})$, where $\mathscr{L}$ and $\mathscr{R}$ are the left and right wedge regions (respectively) defined in Section 2 of [KW91]. Then, clearly (recall that the Killing field is assumed to be complete), an isometry-invariant solution $\phi \in \tilde{S}_0^k$ is identically zero on M if and only if, for all $p \in \mathscr{L} \cup \mathscr{R}$, $\phi(\tau_t(p)) \to 0$ as $t \to +\infty$. Thus, in the presence of appropriate 'pointwise decay' results for (sufficiently regular) solutions of the Klein–Gordon equation in question, the result follows.

In the case of the *massless* wave equation, recent papers by Dafermos, Rodnianski and Shlapentokh-Rothman [DR09, DRSR16, DR07] contain pointwise decay results which are sufficient for our purposes in the case of Kruskal and of the globally hyperbolic patches of Kerr and Schwarzschild–de Sitter considered in [KW91], *provided* that we pick $k \geq 5$ in the definition of $\tilde{S}_0^k$ large enough for the 'higher order weighted energies' defined in those papers to be finite. That this can always be done can be seen by inspection of the relevant formulae in those papers: in the case of Kruskal, we refer to the estimate (1.8) in Theorem 1.1 in [DR09]; for Kerr, we refer to the pointwise estimates in Corollary 3.1 in [DRSR16]; for Schwarzschild–de Sitter, see Theorem 1.1 in [DR07].[9]

wedge corresponding to classical solutions in the subspace of solutions $S_0^R = S_A^R + S_B^R$ and, of course, we don't even know if that theorem is applicable since we don't know if our symplectic form on S restricts to a symplectic form on this subspace. We also mention, in passing, that since the massless minimally coupled Klein–Gordon field on de Sitter has a classical zero mode (namely the constant solution) the strengthened uniqueness theorem, Theorem 5.1 in Chapter 5 of [KW91], is also inapplicable for the reasons explained in the introductory remarks in that Chapter. We are grateful to Atsushi Higuchi for helpful conversation on the topic of this footnote.

[9] We also notice that, in the somewhat analogous case of our Proposition 3.4.2, it is the Dirichlet boundary condition which provides the relevant 'decay' for our purposes there.

4.4 Strategy based on analytic elliptic regularity

An alternative approach to showing the non-existence of 'zero modes' in $\tilde{S}_0^k$ in a number of important cases, which requires less heavy machinery and is also more easily generalised to the case where suitable potentials (including e.g. a mass term) are included, is based on an application of *analytic elliptic regularity* [Joh55].[10] Therefore, we must assume the spacetime manifold and metric to be analytic in what follows.

First, we look at the case where the following two conditions hold:

(a) the restriction of the spacetime (M, g) and of the one-parameter group of isometries to either the left or the right wedge is analytically isometric to a (globally hyperbolic) *standard static* spacetime (see Section 3.2 of [San13] and references therein) of form $(\mathbb{R} \times C, \alpha \mathrm{d}t^2 - {}^3g)$ where α (the lapse function) is a positive function on C and ${}^3g$ is a Riemannian metric on the connected manifold C (with C, α and ${}^3g$ analytic);

(b) for any compact set $K \subset M$, the open set $M \setminus J(K)$ has non-empty intersection both with the right and with the left wedge.

It is easy to see that the following spacetimes satisfy the above conditions: Minkowski spacetime with Lorentz boosts as isometries; the Kruskal spacetime with standard 'time translation isometries'; suitable globally hyperbolic patches of the subextremal Reissner–Nordström spacetime and of the Schwarzschild–de Sitter spacetime (with non-zero black hole mass), again with their respective standard 'time translation isometries'. Importantly, the case of *de Sitter* spacetime will not be covered by this strategy due to the failure of condition (b) above.

Under condition (a), on (say) the right wedge, the Klein–Gordon equation with an analytic potential term V will take the form $\left(\alpha^{-1}\frac{\partial^2}{\partial t^2} - \mathcal{D} + V\right)\phi = 0$ where $\mathcal{D}$ is the Laplace–Beltrami operator for ${}^3g$. For an isometry-invariant solution, $\frac{\partial \phi}{\partial t}$ will be identically zero, and therefore so will be $\frac{\partial^2 \phi}{\partial t^2}$ and ϕ will satisfy the manifestly elliptic equation with analytic coefficients

$$\left(-\alpha^{-1}\frac{\partial^2}{\partial t^2} - \mathcal{D} + V\right)\phi = 0$$

(and we notice that the operator $-\alpha^{-1}\frac{\partial^2}{\partial t^2} + \mathcal{D}$ is of course nothing but minus the Laplace-Beltrami operator for the Riemannian metric $\alpha \mathrm{d}t + {}^3g$). Therefore, by analytic elliptic regularity, ϕ must be an analytic function on the right wedge. But, since $\phi \in \tilde{S}_0 \subset \hat{S}$ has Cauchy data – on a Cauchy surface $\mathscr{C}$ for the full spacetime which contains the bifurcation 2-sphere Σ, see [KW91] – of compact support, by finite propagation speed

[10] We would like to thank Robert Wald for suggesting this approach to us and for providing some guidance on how to deal with the case of Kerr, see below.

results it must vanish on $M \setminus J(K)$ where $K = \mathrm{supp}(\phi\restriction_{\mathscr{C}}) \cup \mathrm{supp}(\nabla_n \phi\restriction_{\mathscr{C}})$ and n denotes the vector field of unit normals to $\mathscr{C}$. Under condition (b) above, it must then vanish in an open subset of the right wedge. By analyticity and connectedness, it must vanish identically on the entire right wedge. A similar argument shows that it must vanish identically on the left wedge. Finally, ϕ must vanish on the entire spacetime by continuity at Σ.

An obvious local-to-global version of this argument also shows that the same conclusion holds if we only require, in condition (a) above, that the spacetime in the left and right wedges be simply static (with respect to the one-parameter group of isometries) rather than 'standard' static. However, outside these circumstances the argument won't straightforwardly apply. Nonetheless, under some mild restrictions on the possible potential terms which we shall state, one can also fill the gap for the case of the globally hyperbolic patch of (subextremal, maximally extended) Kerr defined on page 66 of [KW91] and denoted by $\mathscr{M}$ there, with Killing vector field $\xi_+ = \partial/\partial t + \Omega_+ \partial/\partial\varphi$ in Boyer-Lindquist coordinates (t, r, θ, φ). Here, denoting the black hole's angular momentum by a and its mass by M, $\Omega_+ = a/(r_+^2 + a^2)$ is the angular velocity of the black hole/Killing horizon situated at $r = r_+ = M + \sqrt{M^2 - a^2}$ and we recall that there is a cosmological horizon 'at' $r = r_- = M - \sqrt{M^2 - a^2}$. In the right wedge where the Boyer-Lindquist coordinates are regular, the Laplace-Beltrami operator associated with the Kerr metric is

$$\Box = \left[a^2 \sin^2\theta - \frac{(a^2+r^2)^2}{\Delta(r)}\right]\frac{\partial^2}{\partial t^2} - \frac{a^2}{\Delta(r)}\frac{\partial^2}{\partial\varphi^2} - \frac{2a[r^2+a^2-\Delta(r)]}{\Delta(r)}\frac{\partial^2}{\partial\varphi\partial t} + \frac{\partial}{\partial r}\left[\Delta(r)\frac{\partial}{\partial r}\right] + \not{\Delta}_{\mathbb{S}^2},$$

where $\Delta(r) = (r - r_+)(r - r_-)$ (so that $\Delta(r) > 0$ everywhere in the right wedge) and $\not{\Delta}_{\mathbb{S}^2} = \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\frac{\partial}{\partial\theta}\right) + \frac{1}{\sin^2\theta}\frac{\partial^2}{\partial\varphi^2}$ is the Laplacian on the two-dimensional unit sphere. Now, let u be a C^2 function on $\mathscr{M}$ which is invariant under the isometries generated by ξ_+. Then, everywhere in the right wedge,

$$\frac{\partial u}{\partial\varphi} = -\Omega_+^{-1}\frac{\partial u}{\partial t} \quad \text{and} \quad \frac{\partial^2 u}{\partial\varphi^2} = \Omega_+^{-2}\frac{\partial^2 u}{\partial t^2}. \tag{4.11}$$

Thus, if u is an isometry-invariant solution to $\Box u = 0$ on $\mathscr{M}$, belonging to $\tilde{S}_0$, then we can use the equations in (4.11) to 'trade' φ-derivatives for t-derivatives and obtain

$$\left\{F(r,\theta)\frac{\partial^2}{\partial t^2} + \frac{\partial}{\partial r}\left[\Delta(r)\frac{\partial}{\partial r}\right] + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\left[\sin\theta\frac{\partial}{\partial\theta}\right]\right\} u = 0 \tag{4.12}$$

where $F(r,\theta)$ is an analytic function for $(r,\theta) \in (r_+,\infty) \times (0,\pi)$. Clearly, the same equation will be satisfied by the Fourier coefficients

$$\hat{u}_m(t,r,\theta) := \int_0^{2\pi} u(t,r,\theta,\varphi)e^{-im\varphi}\,\mathrm{d}\varphi, \quad m \in \mathbb{Z}.$$

However, a simple calculation shows that, in virtue of the first equation in (4.11),

$$\frac{\partial \hat{u}_m}{\partial t} = im\Omega_+\hat{u}_m \quad \text{and, consequently,} \quad \frac{\partial^2 \hat{u}_m}{\partial t^2} = -m^2\Omega_+^2\hat{u}_m$$

for all $m \in \mathbb{Z}$. Pick a positive constant K and set $G(r,\theta) = -m^2\Omega_+^2[F(r,\theta) - K]$; then $\hat{u}_m$ solves $P\hat{u}_m = 0$ where

$$P = K\frac{\partial^2}{\partial t^2} + \Delta(r)\frac{\partial^2}{\partial r^2} + \frac{\partial^2}{\partial \theta^2} + \frac{\mathrm{d}\Delta}{\mathrm{d}r}(r)\frac{\partial}{\partial r} + \cot\theta\frac{\partial}{\partial\theta} + G(r,\theta).$$

P is a differential operator with analytic coefficients. An inspection of the highest order terms shows that it is elliptic on $\mathbb{R} \times (r_+,\infty) \times (0,\pi)$. Therefore, by analytic elliptic regularity, $\hat{u}_m$ is analytic. But $\hat{u}_m$ must vanish in an open set because of the support properties of $u \in \tilde{S}_0$.[11] Therefore $\hat{u}_m = 0$ for all $m \in \mathbb{Z}$. By the Fourier inversion formula, this in turn implies that $u = 0$ in the right wedge. Similar reasoning shows that u must vanish in the left wedge. Again, by continuity at the bifurcation surface this means that u must vanish on $\mathscr{M}$. For ease of presentation, we only showed the proof explicitly in the case of the massless wave equation. However, it is clear that an analytic potential term can be added with no change in the arguments, provided it is independent of the coordinate φ – as would of course be the case for a constant mass term or for a constant multiple of the Ricci scalar.

4.5 Conclusions and future directions

To conclude, the two lines of argument presented in this chapter have enabled us to fill the gap in [KW91] in many cases of interest (however, see our discussion in the introductory section of this Appendix for the meaning of 'filling the gap' in the cases of Schwarzschild–de Sitter and Kerr).

[11]To quickly see this, the reader may wish to consider a *projection diagram*, in the sense of [COS12] (see also Chapter 3 of [Chr15]), for the region of Kerr under consideration and denoted by $\mathscr{M}$ above. The projection diagram in Fig. 3 of [COS12] appears to closely resemble the more commonly seen conformal diagram for the submanifold corresponding to the axis of symmetry ($\theta = 0$ or $\theta = \pi$) of the Kerr solution. However, unlike the latter, the former captures causal properties of the entire spacetime (in a precise way discussed in Section 3 of [COS12]). In particular, since u above has spacelike compact support on $\mathscr{M}$, it follows that the projection of its support onto the (1+1)-dimensional diagram is spacelike compact with respect to the (1+1)-dimensional Minkowski metric. The claimed result then easily follows upon observing that the projection diagram is obtained by projecting out the Boyer-Lindquist coordinates θ and φ.

In the case of de Sitter spacetime, it is not obvious to us that there can be no non-zero solutions in $\tilde{S}_0$ which are invariant under the one-parameter group of isometries generating the bifurcate Killing horizon. Clearly, for massless minimally coupled fields, there *are* non-zero solutions in $\hat{S} \supset \tilde{S}_0$ which are invariant: namely, the constant (non-zero) solutions. Therefore, in particular, one would need to show that no non-zero constant solution can lie in $\tilde{S}_0$. But this would still not suffice to fill the gap.

Another direction for extending our results is to decide whether or not $(\tilde{S}_0, \hat{\sigma})$ is symplectic in the cases of Schwarzschild–de Sitter and Kerr. While the answer to this question would have no impact on the quantum theory, any proof would presumably require considerably different arguments to the ones, based on the existence of invariant Hadamard states, which we have been using. And one may learn valuable lessons in trying to find an alternative way to decide on the 'symplecticity' of $(\tilde{S}_0, \hat{\sigma})$ in general spacetimes with bifurcate Killing horizons.

Chapter 5

Notes towards a theory of spacetimes with timelike boundaries and boundary value problems

5.1 Elements of causal theory in the presence of timelike boundaries

Motivated by our study of the quantum theory of fields on curved spacetimes possessing boundaries, we would like to put on firmer ground the causal theory of such spacetimes. We assume that the reader has familiarised herself with the contents of Section 2.1 in this thesis, where general Lorentzian geometry and some elements of causal theory are reviewed in the context of ordinary open spacetimes.

5.1.1 STT∂s

Definition 5.1.1. Let M be a (Hausdorff, second countable) non-empty smooth manifold with (possibly empty) boundary and with $\dim M = n + 1 \geq 2$. If M is endowed with a metric g of Lorentzian signature and with a time orientation $\mathfrak{t}$ determined by a global smooth timelike vector field, then $\mathscr{M} := (M, g, \mathfrak{t})$ will be called a *spacetime-with-boundary.* Letting $g^{-1} \in \Gamma(TM \otimes TM)$ denote the inverse metric to g, we will say that $\mathscr{M}$ is a *spacetime-with-timelike-boundary* (STT∂) if the restriction of g^{-1} to the conormal bundle of ∂M is negative-definite; equivalently, if the pullback of g under the inclusion $\iota : \partial M \to M$ defines a metric of Lorentzian signature on ∂M. In this case, $\mathfrak{t}$ induces a time orientation on $(\partial M, \iota^* g)$, thus yielding a spacetime (without boundary) which we denote by $\partial\mathscr{M}$.

For clarity, we fix some nomenclature. A continuous, piecewise-smooth regular (i.e. its velocity vectors never vanish, including at break points) curve $\gamma : I \to M$ with finitely many break points is said to be *causal* (resp. *timelike*) if its velocity vectors are timelike or null (resp. timelike) and either all future-pointing or all past-pointing. Notice that, with this definition, the constant curves are not causal. We write $p \leq q$ (resp. $p \ll q$) in $\mathscr{M}$ if either $p = q$ or there exists a future-directed causal (resp. timelike) curve in $\mathscr{M}$ from p to q. Replacing 'future-directed' with 'past-directed' in the previous sentence similarly defines the relation $\geq$ (resp. $\gg$). The *causal future* $(+)$ and *past* $(-)$ of a point p in $\mathscr{M}$ is

$$J^{\pm}_{\mathscr{M}} = \{q \mid p \leq q\}$$

and the *chronological future* $(+)$ and *past* $(-)$ are

$$I^{\pm}_{\mathscr{M}} = \{q \mid p \ll q\}\,.$$

Finally, for an arbitrary subset $S \subset M$, we define $J^{\pm}_{\mathscr{M}}(S) = \bigcup_{p \in S} J^{\pm}_{\mathscr{M}}(p)$, $I^{\pm}_{\mathscr{M}}(S) = \bigcup_{p \in S} I^{\pm}_{\mathscr{M}}(p)$ and $J_{\mathscr{M}}(S) = J^{+}_{\mathscr{M}}(S) \cup J^{-}_{\mathscr{M}}(S)$. In the context of spacetimes-with-timelike-boundaries, the theorem below was proved as Proposition 3.5 in [Sol06]. We give an alternative proof which is valid on any spacetime-with-boundary (and, in particular, on one whose boundary is not everywhere timelike, or on one with empty boundary!).

Proposition 5.1.2. *In a spacetime-with-boundary $\mathscr{M}$, the chronological relation $\ll$ is open. That is, if $p \ll q$ in $\mathscr{M}$ then there are neighbourhoods X and Y of p and q respectively, such that $p' \ll q'$ for all $p' \in X$ and $q' \in Y$. A similar statement holds for the chronological relation $\gg$.*

Proof. Let $\gamma : [0,1] \to M$ be a timelike curve from p to q. Pick a chart $(U, \varphi : U \to V)$ around p, with $V \subseteq \mathbb{R}^{d+1}$ or $V \subseteq \mathbb{H}^{d+1}$ depending on whether p is an interior or boundary point (respectively).[1] Without loss of generality we can assume that $V = \varphi(U)$ is a convex subset of $\mathbb{R}^{d+1}$ or of $\mathbb{H}^{d+1}$. Denote $\varphi(p) \in V$ by x. Since γ is timelike and piecewise smooth, it is a smooth immersion when restricted to the interval $[0, a]$, where $a > 0$ is the location of the first break point of γ. In particular, it is an embedding locally around 0 [Lee13, Thm. 4.25], and, for small enough $\varepsilon > 0$, it is injective on $I_0 \coloneqq [0, \varepsilon)$ and $\gamma(I_0) \subset U$. We now consider the curve $\tilde{\gamma} = \varphi \circ \gamma\restriction_{I_0} : I_0 \to V$. Equipping V with the pushforward metric $\tilde{g} = \varphi_* g$ and pushforward time orientation $\tilde{\mathfrak{t}} = \varphi_* \mathfrak{t}$, so that $(U, g\restriction_U, \mathfrak{t})$ and $(V, \tilde{g}, \tilde{\mathfrak{t}})$ are time-oriented isometric, we will argue in the next paragraph that there exists a $\tau \in I_0 \setminus \{0\}$ for which a $\delta > 0$ can be found such that (a) the open ball (in either $\mathbb{R}^{d+1}$ or $\mathbb{H}^{d+1}$) $\mathcal{B}_\delta(x)$ is entirely contained in V, and (b) straight line segments in V between any $y \in \mathcal{B}_\delta(x)$ and $\tilde{\gamma}(\tau)$ are everywhere timelike and future-directed with respect to the geometry of $(V, \tilde{g})$. A completely analogous result will also hold, *mutatis*

[1] Here, $\mathbb{H}^{d+1}$ denotes the standard closed half space in $\mathbb{R}^{d+1}$.

mutandis, for the final endpoint q. Finally, the claim in the statement of the theorem clearly follows by pulling back these results to (M, g).

Suppose, instead, that no such τ can be found. Let $(\tau_n)_n$ be a sequence in $I_0 \setminus \{0\}$ with $\tau_n \to 0$ as $n \to \infty$. Denoting $x_n = \tilde{\gamma}(\tau_n)$, for each n we can find a $\delta_n > 0$ such that $\delta_n < \|x_n - x\|$ and such that, for all $y \in \mathcal{B}_{\delta_n}(x)$, the unit norm vectors

$$l_n(y) = \frac{x_n - y}{\|x_n - y\|}$$

are at Euclidean distance less than $1/n$ from the unit norm vector

$$l_n = \frac{x_n - x}{\|x_n - x\|}.$$

[This just follows from the continuity of $y \mapsto l_n(y)$ on the complement of the point x_n.] We then automatically also have $\delta_n \to 0$ as $n \to \infty$. Since we are rejecting the claim, for each n we obtain a pair (y_n, z_n) with $y_n \in \mathcal{B}_{\delta_n}(x)$, z_n belonging to the straight line segment between y_n and x_n, and the vector $t_n = (z_n, l_n(y_n))$ – which is of course tangent to the straight line segment between y_n and x_n – not both timelike and future-directed. It is clear that $z_n \to x$ as $n \to \infty$. We show below that

$$l_n(y_n) \to \frac{\tilde{\gamma}'(0)}{\|\tilde{\gamma}'(0)\|} \text{ as } n \to \infty \tag{5.1}$$

so that the sequence $(t_n)_n$ in the tangent bundle of V (trivialized as $V \times \mathbb{R}^{d+1}$) tends to a tangent vector at x which is timelike and future directed by assumption. But since the subset C of TV consisting of timelike and future directed tangent vectors is open, any limit point of $TV \setminus C$ must belong to $TV \setminus C$, and we get a contradiction.

The following steps are a proof of (5.1):

$$\begin{aligned}\left\| l_n(y_n) - \frac{\tilde{\gamma}'(0)}{\|\tilde{\gamma}'(0)\|} \right\| &\le \|l_n(y_n) - l_n\| + \left\| l_n - \frac{\tilde{\gamma}'(0)}{\|\tilde{\gamma}'(0)\|} \right\| \\ &\le \frac{1}{n} + \left\| \frac{x_n - x}{\|x_n - x\|} - \frac{\tilde{\gamma}'(0)}{\|\tilde{\gamma}'(0)\|} \right\|;\end{aligned}$$

as is elementary to check, the right-hand side of the last expression tends to zero as $n \to \infty$, and the claim is established. □

Consequently, for any $S \subset M$, the sets $I^{\pm}_{\mathscr{M}}(S)$ are always open. Note that, while in a general spacetime-with-boundary $\mathscr{M}$ the sets $I^{\pm}_{\mathscr{M}}(p)$ are always non-empty when p is an interior point, this might not be the case if p is a boundary point. On a spacetime-with-*timelike*-boundary, on the other hand, if $p \in \partial M$ we have $I^{\pm}_{\mathscr{M}}(p) \supset I^{\pm}_{\partial\mathscr{M}}(p) \neq \emptyset$. It is also obvious that, for any point p, $I^{\pm}_{\mathscr{M}}(p) \cap U \neq \emptyset$ for any neighbourhood U of p.

The following key results were established by Solis (Proposition 3.6 and Proposition 3.7 in [Sol06]).

Proposition 5.1.3. *On any STT∂ $\mathscr{M} = (M, g, \mathfrak{t})$:*

- **(i)** *if $p \ll q \leq r$ or $p \leq q \ll r$ then $p \ll r$;*
- **(ii)** *$J^{\pm}_{\mathscr{M}}(S) \subseteq \overline{I}^{\pm}_{\mathscr{M}}(S)$ for any $S \subseteq M$ (where the overline indicates the topological closure of the corresponding set).* □

We supplement them with an easy corollary which will be useful later.

Corollary 5.1.4. *On a STT∂ $\mathscr{M} = (M, g, \mathfrak{t})$, and for any $S \subseteq M$, it holds that (a) $\overline{J}^{\pm}_{\mathscr{M}}(S) = \overline{I}^{\pm}_{\mathscr{M}}(S)$, (b) $\dot{J}^{\pm}_{\mathscr{M}}(S) = \dot{I}^{\pm}_{\mathscr{M}}(S)$, and (c) $\mathring{J}^{\pm}_{\mathscr{M}}(S) = I^{\pm}_{\mathscr{M}}(S)$ (where the dots and circles indicate topological boundaries and interiors of the corresponding sets, respectively).*

Proof. Equality (a) follows from item **(ii)** of Proposition 5.1.3 and the fact that $J^{\pm}_{\mathscr{M}}(S) \supseteq I^{\pm}_{\mathscr{M}}(S)$. We can prove the left-to-right inclusion in Equality (b) in the following way:

$$\dot{J}^{\pm}_{\mathscr{M}}(S) = \overline{J}^{\pm}_{\mathscr{M}}(S) \setminus \mathring{J}^{\pm}_{\mathscr{M}}(S) = \overline{I}^{\pm}_{\mathscr{M}}(S) \setminus \mathring{J}^{\pm}_{\mathscr{M}}(S) \subseteq \overline{I}^{\pm}_{\mathscr{M}}(S) \setminus \mathring{I}^{\pm}_{\mathscr{M}}(S) = \dot{I}^{\pm}_{\mathscr{M}}(S),$$

having used equality (a) and $\mathring{J}^{\pm}_{\mathscr{M}}(S) \supseteq \mathring{I}^{\pm}_{\mathscr{M}}(S)$. We now prove the other inclusion in (b), considering the '+' case for simplicity and without loss of generality. Consider an arbitrary point $p \in \dot{I}^{+}_{\mathscr{M}}(S)$. Then, since $I^{+}_{\mathscr{M}}(S)$ is open, $p \in M \setminus I^{+}_{\mathscr{M}}(S)$. Since p is automatically in $\overline{J}^{+}_{\mathscr{M}}(S)$, in order to conclude that $p \in \dot{J}^{+}_{\mathscr{M}}(S)$ it suffices to show that $p \notin \mathring{J}^{+}_{\mathscr{M}}(S)$. Suppose, instead, that p is an interior point of $J^{+}_{\mathscr{M}}(S)$. Then there is an open neighbourhood $U \ni p$ entirely contained in $J^{+}_{\mathscr{M}}(S)$. Picking $q \in I^{-}_{\mathscr{M}}(p) \cap U$, we then have $z \leq q \ll p$ for some $z \in S$, and thus $p \in I^{+}_{\mathscr{M}}(z) \subseteq I^{+}_{\mathscr{M}}(S)$ by item **(i)** in Proposition 5.1.3. Since this contradicts our hypothesis, we conclude that $\dot{I}^{+}_{\mathscr{M}}(S) \subseteq \dot{J}^{+}_{\mathscr{M}}(S)$. Finally, equality (c) straightforwardly follows from the previous two. □

A simple application of Proposition 5.1.3(i) yields the additional results below where, just as in the case of spacetimes without boundary, an open subset X is defined to be *causally convex* if any causal curve between any two of its points is entirely contained in X.

Corollary 5.1.5. *Let $A, B \subseteq M$ be arbitrary subsets in a STT∂ $\mathscr{M} = (M, g, \mathfrak{t})$. Then $I^{\pm}_{\mathscr{M}}(A)$ and $I^{+}_{\mathscr{M}}(A) \cap I^{-}_{\mathscr{M}}(B)$ are causally convex subsets. Whenever open, so are the subsets $M \setminus J^{\pm}_{\mathscr{M}}(A)$ and $M \setminus J_{\mathscr{M}}(A)$.* □

The theory of *continuous* causal and timelike curves was also studied in detail by Solis (Chapter 3.1.3 in [Sol06]). One defines *continuous* timelike or causal curves, and their future and past-inextendibility, in the same way as for spacetimes with empty boundary.

Using continuous timelike or causal curves instead of piecewise smooth ones does not enlarge the chronological ($\ll$) or causal ($\leq$) relations.

5.1.2 Global hyperbolicity and Cauchy surfaces

Definition 5.1.6. Let $\mathscr{M} = (M, g, \mathfrak{t})$ be a STT∂. Then:

(a) A topological hypersurface $\mathscr{C}$ will be said to be a *Cauchy surface* for $\mathscr{M}$ if every future and past inextendible (continuous) timelike curve meets $\mathscr{C}$ precisely once.

(b) A C^1 function $t : M \to \mathbb{R}$ is said to be a *temporal function* if dt is timelike and future-directed (with respect to g^{-1}), and a *Cauchy* temporal function if in addition all its level sets are Cauchy surfaces (which are then necessarily C^1-embedded spacelike submanifolds).

(c) $\mathscr{M}$ will be said to be *globally hyperbolic* if it is strongly causal and if for all $p, q \in M$ the sets $J^+_{\mathscr{M}}(p) \cap J^-_{\mathscr{M}}(q)$ are compact.

Note that, if $\mathscr{M} = (M, g, \mathfrak{t})$ is a STT∂ with Cauchy surface $\mathscr{C}$, then just as in the case of spacetimes without boundary, we have $M = I^+_{\mathscr{M}}(\mathscr{C}) \uplus \mathscr{C} \uplus I^-_{\mathscr{M}}(\mathscr{C})$, implying in particular that $\mathscr{C}$ is closed. Making use of Proposition 5.1.3(i), one similarly sees that $M = J^{\pm}_{\mathscr{M}}(\mathscr{C}) \uplus I^{\mp}_{\mathscr{M}}(\mathscr{C})$. Thus $J^{\pm}_{\mathscr{M}}(\mathscr{C})$ is closed and its boundary [which, by virtue of Corollary 5.1.4, is also the boundary of its interior $I^{\pm}_{\mathscr{M}}(\mathscr{C})$] is $\mathscr{C}$.

The results in the next Proposition are due to Solis (Proposition 3.17 and Proposition 3.18 in [Sol06], see also Section 2.2 in [CGS09]).

Proposition 5.1.7. *Let $\mathscr{M} = (M, g, \mathfrak{t})$ be a globally hyperbolic STT∂. Let $A, B \subseteq M$ be compact. Then:*

(i) $J^{\pm}_{\mathscr{M}}(A)$ is closed, and thus also $J^{\pm}_{\mathscr{M}}(A) = \overline{I}^{\pm}_{\mathscr{M}}(A)$ (in particular, any globally hyperbolic STT∂ is causally simple);

(ii) $J^+_{\mathscr{M}}(A) \cap J^-_{\mathscr{M}}(B)$ is compact;

(iii) if $\mathscr{C}$ is a Cauchy surface for $\mathscr{M}$ then $J^{\pm}_{\mathscr{M}}(A) \cap J^{\mp}_{\mathscr{M}}(\mathscr{C})$ and $J^{\pm}_{\mathscr{M}}(A) \cap \mathscr{C}$ are compact. □

Just as in the case of spacetimes without boundaries, it is clear that any causally convex subset of a globally hyperbolic STT∂ inherits the structure of a STT∂ which is itself globally hyperbolic. The following result then follows immediately from Corollary 5.1.5.

Corollary 5.1.8. *Let $A, B \subseteq M$ be arbitrary subsets in a globally hyperbolic STT∂ $\mathscr{M} = (M, g, \mathfrak{t})$. Then $I^{\pm}_{\mathscr{M}}(A)$ and $I^+_{\mathscr{M}}(A) \cap I^-_{\mathscr{M}}(B)$ always yield causally convex globally*

hyperbolic sub-spacetimes of $\mathscr{M}$. If A is compact, then so do $M \setminus J^{\pm}_{\mathscr{M}}(A)$ and $M \setminus J_{\mathscr{M}}(A)$.
□

We conclude this section with a technical lemma which will be useful in the next section.

Lemma 5.1.9. *Let $\mathscr{M} = (M, g, \mathfrak{t})$ be a globally hyperbolic STT∂ with Cauchy surface $\mathscr{C}$, and A be a compact set (e.g. a one-point set). Then the interior of the closed set $D = J^-_{\mathscr{M}}(A) \cap J^+_{\mathscr{M}}(\mathscr{C})$ is $I^-_{\mathscr{M}}(A) \cap I^+_{\mathscr{M}}(\mathscr{C})$. Furthermore, we have*

$$D = \overline{\mathring{D}} \cup [\dot{J}^-_{\mathscr{M}}(A) \cap \mathscr{C}] = \overline{\mathring{D} \cup [\dot{J}^-_{\mathscr{M}}(A) \cap \mathscr{C}]}. \tag{5.2}$$

Proof. The interior of the intersection of finitely many subsets of a topological space always equals the intersection of the interiors. Together with the results collected so far about the sets $J^-_{\mathscr{M}}(A)$ (when A is compact) and $J^+_{\mathscr{M}}(\mathscr{C})$, this immediately gives the first part. For the second part notice that since D is closed and contains the closed subset $\dot{J}^-_{\mathscr{M}}(A) \cap \mathscr{C}$, the only non-trivial inclusion is $D \subseteq \overline{\mathring{D}} \cup [\dot{J}^-_{\mathscr{M}}(A) \cap \mathscr{C}]$. We have

$$\begin{aligned} D &= J^-_{\mathscr{M}}(A) \cap J^+_{\mathscr{M}}(\mathscr{C}) \\ &= [I^-_{\mathscr{M}}(A) \cup \dot{J}^-_{\mathscr{M}}(A)] \cap [I^+_{\mathscr{M}}(\mathscr{C}) \cup \mathscr{C}] \\ &= [I^-_{\mathscr{M}}(A) \cap I^+_{\mathscr{M}}(\mathscr{C})] \cup [\dot{J}^-_{\mathscr{M}}(A) \cap I^+_{\mathscr{M}}(\mathscr{C})] \cup [I^-_{\mathscr{M}}(A) \cap \mathscr{C}] \cup [\dot{J}^-_{\mathscr{M}}(A) \cap \mathscr{C}]. \end{aligned}$$

The result then follows since, as a moment's reflection reveals, each of the first three sets is contained in $\overline{\mathring{D}}$. Notice that the second equality in Equation (5.2) is simply an application of the fact that the union of the closures of finitely many sets always equals the closure of the union of those sets. □

5.2 Miscellaneous results on Green hyperbolic boundary value problems

5.2.1 A 'black-box' exact sequence for general Green hyperbolic boundary value problems

In this section we consider special kinds of linear differential operators acting on sections of *field bundles* (i.e. vector bundles) over a manifold with boundary M. Namely, we look at differential operators $P : \Gamma^\infty(F) \to \Gamma^\infty(\tilde{F})$ where $\tilde{F} \to M$ is the *densitized dual bundle* to $F \to M$, i.e. $\tilde{F} := F^* \otimes_M \bigwedge^n(T^*M)$ with the obvious bundle projection. We do so with a view to variational problems in which such operators arise naturally, as explained e.g. in [Kha14]. What we have in mind is the following: we would like to generalise the notion of Green hyperbolicity given in Section 2.4.1 to the context of globally hyperbolic

STT∂s, accommodating for the possibility that boundary conditions may be imposed at the boundary and under the assumption that the structural theorems on the existence of smooth spacelike Cauchy surfaces and Cauchy temporal functions, which are well-established in the boundaryless context, hold also in this category. A (homogeneous) boundary condition will be tantamount to the selection of a particular linear subspace of Γ – the subspace of sections satisfying the so far unspecified boundary condition. Then, assuming that an operator is Green hyperbolic compatibly with the chosen boundary condition in some sense, we would like to investigate whether or not a sequence of vector spaces akin to that of Theorem 2.4.7 can be exact.

Let $\mathscr{M}$ be a STT∂. Abbreviating $\Gamma^\infty(F)$ and $\Gamma_0^\infty(F)$ by Γ and Γ_0 respectively [and similarly abbreviating $\Gamma^\infty(\tilde{F})$ and $\Gamma_0^\infty(\tilde{F})$ by $\tilde{\Gamma}$ and $\tilde{\Gamma}_0$ respectively], we note that the causal structure coming from the metric and time orientation of $\mathscr{M}$ allows one to introduce some subspaces of Γ distinguished by their support properties just as was done (in the case of spacetimes without boundaries) in Definition 2.1.40. Namely:

$$\Gamma_\pm \coloneqq \left\{\phi \in \Gamma \;\middle|\; \operatorname{supp}\phi \subseteq J^\pm(K) \text{ for some compact } K\right\},$$

$$\Gamma_{\mathrm{sc}} \coloneqq \{\phi \in \Gamma \mid \operatorname{supp}\phi \subseteq J(K) \text{ for some compact } K\},$$

and corresponding subspaces $\tilde{\Gamma}_\pm$ and $\tilde{\Gamma}_{\mathrm{sc}}$ of $\tilde{\Gamma}$ are defined similarly. Notice that $\Gamma_\pm$ corresponds to what would have been denoted by $\Gamma^\infty_{\mathrm{ret/adv}}(F)$ in Definition 2.1.40.

Now let $\Gamma_\partial \subseteq \Gamma$ be a linear subspace of Γ, and define

$$\Gamma_{0,\partial} = \Gamma_0 \cap \Gamma_\partial, \qquad \Gamma_{\pm,\partial} = \Gamma_\pm \cap \Gamma_\partial, \qquad \Gamma_{\mathrm{sc},\partial} = \Gamma_{\mathrm{sc}} \cap \Gamma_\partial.$$

Definition 5.2.1. The linear differential operator $P : \Gamma \to \tilde{\Gamma}$ is a *Green hyperbolic operator with respect to the causal structure of* $\mathscr{M}$ *and the domain* $\Gamma_\partial \subseteq \Gamma$ if:

(a) there exists a unique $\phi_\pm \in \Gamma_{\pm,\partial}$ solving $P\phi_\pm = \tilde{\alpha}_\pm$ for any $\tilde{\alpha}_\pm \in \tilde{\Gamma}_\pm$;

(b) this solution satisfies $\operatorname{supp}\phi_\pm \subseteq J^\pm(\operatorname{supp}\tilde{\alpha}_\pm)$.

More succinctly, in that case we will speak of a differential operator which is Green hyperbolic with respect to $(\mathscr{M}, \Gamma_\partial)$. The resulting linear maps $\mathrm{G}_\pm : \tilde{\Gamma}_\pm \to \Gamma_{\pm,\partial}$ (given by sending $\tilde{\alpha}_\pm$ to $\phi_\pm$ above) are called the retarded (+) and advanced (−) Green operators for the triple $(\mathscr{M}, P, \Gamma_\partial)$, and their difference $\mathrm{G} = \mathrm{G}_+ - \mathrm{G}_- : \tilde{\Gamma}_0 \to \Gamma_{\mathrm{sc},\partial}$ is termed the causal propagator [for the triple $(\mathscr{M}, P, \Gamma_\partial)$].

Remark. Notice that we are not imposing that similar properties hold for the adjoint operator to P, contrary to what was done in Section 2.4. This is because the existence of retarded/advanced Green operators for the adjoint operator will not be needed in the proofs of the statements in this subsection.

The following result follows easily from Definition 5.2.1.

Proposition 5.2.2. *If P is Green hyperbolic with respect to $(\mathscr{M}, \Gamma_\partial)$, then $\mathrm{G}_\pm : \tilde{\Gamma}_\pm \to \Gamma_{\pm,\partial}$ is an inverse to $P\restriction_{\Gamma_{\pm,\partial}} : \Gamma_{\pm,\partial} \to \tilde{\Gamma}_\pm$, i.e.*

$$P \circ \mathrm{G}_\pm = \mathrm{id}_{\tilde{\Gamma}_\pm}, \tag{5.3}$$

$$\mathrm{G}_\pm \circ P\restriction_{\Gamma_{\pm,\partial}} = \mathrm{id}_{\Gamma_{\pm,\partial}}. \tag{5.4}$$

Consequently,

$$P \circ \mathrm{G} = 0, \tag{5.5}$$

$$\mathrm{G} \circ P\restriction_{\Gamma_{0,\partial}} = 0. \tag{5.6}$$

□

Now consider the sequence

$$\{0\} \longrightarrow \Gamma_{0,\partial} \xrightarrow{P} \tilde{\Gamma}_0 \xrightarrow{\mathrm{G}} \Gamma_{\mathrm{sc},\partial} \xrightarrow{P} \tilde{\Gamma}_{\mathrm{sc}} \longrightarrow \{0\}. \tag{5.7}$$

By Proposition 5.2.2, this sequence is a complex. We thus turn to the following question: what are sufficient conditions for (5.7) to be *exact*? A possible answer is given in the following theorem.

Theorem 5.2.3. *Suppose there exists a pair $\chi = \{\chi_+, \chi_-\}$ of functions in $C^\infty(M)$ with the following properties:*

- $\chi_+ + \chi_- \equiv 1$*;*
- *multiplication by $\chi_\pm$ maps $\tilde{\Gamma}_{\mathrm{sc}} \to \tilde{\Gamma}_\pm$, $\Gamma_{\mathrm{sc}} \to \Gamma_\pm$, $\Gamma_{\mathrm{sc},\partial} \to \Gamma_{\pm,\partial}$ and $\Gamma_{\mp,\partial} \to \Gamma_{0,\partial}$;*
- $P[\chi_+\psi] - \chi_+ P[\psi] \in \tilde{\Gamma}_0$ *whenever $\psi \in \Gamma_{\mathrm{sc},\partial}$.*

Then (5.7) is an exact sequence with witnessing contracting homotopy

$$\begin{array}{ccccccccccc}
\{0\} & \longrightarrow & \Gamma_{0,\partial} & \xrightarrow{P} & \tilde{\Gamma}_0 & \xrightarrow{\mathrm{G}} & \Gamma_{\mathrm{sc},\partial} & \xrightarrow{P} & \tilde{\Gamma}_{\mathrm{sc}} & \longrightarrow & \{0\} \\
 & & \downarrow{\scriptstyle \mathrm{id}} & \swarrow{\scriptstyle {}_\chi\mathrm{G}} & \downarrow{\scriptstyle \mathrm{id}} & \swarrow{\scriptstyle P_\chi} & \downarrow{\scriptstyle \mathrm{id}} & \swarrow{\scriptstyle \mathrm{G}_\chi} & \downarrow{\scriptstyle \mathrm{id}} & & \\
\{0\} & \longrightarrow & \Gamma_{0,\partial} & \xrightarrow[P]{} & \tilde{\Gamma}_0 & \xrightarrow[\mathrm{G}]{} & \Gamma_{\mathrm{sc},\partial} & \xrightarrow[P]{} & \tilde{\Gamma}_{\mathrm{sc}} & \longrightarrow & \{0\}
\end{array} \tag{5.8}$$

The cochain homotopy maps ${}_\chi\mathrm{G}$, P_χ and G_χ, are defined in terms of $\chi = \{\chi_+, \chi_-\}$ by

$${}_\chi\mathrm{G}[\alpha] = \chi_+ \mathrm{G}_-[\alpha] + \chi_- \mathrm{G}_+[\alpha] \tag{5.9}$$

$$P_\chi[\psi] = P[\chi_+\psi] - \chi_+ P[\psi] = -P[\chi_-\psi] + \chi_- P[\psi] \tag{5.10}$$

$$\mathrm{G}_\chi[\alpha] = \mathrm{G}_+[\chi_+\alpha] + \mathrm{G}_-[\chi_-\alpha] \tag{5.11}$$

Proof. It is clear that the requirements on $\chi = \{\chi_+, \chi_-\}$ ensure that the cochain homotopy maps are well-defined. The contracting homotopy identities are:

$$\begin{aligned} \forall\ \psi \in \Gamma_{0,\partial} \quad ({}_\chi\mathrm{G} \circ P)[\psi] &= \chi_+\mathrm{G}_-[P\psi] + \chi_-\mathrm{G}_+[P\psi] \\ &= (\chi_+ + \chi_-)\psi \\ &= \psi, \end{aligned}$$

which holds by virtue of Equation (5.4);

$$\begin{aligned} \forall\ \alpha \in \tilde{\Gamma}_0 \quad (P \circ {}_\chi\mathrm{G} + P_\chi \circ \mathrm{G})[\alpha] &= P[\chi_+\mathrm{G}_-\alpha] + P[\chi_-\mathrm{G}_+\alpha] \\ &\quad - P[\chi_-\mathrm{G}\alpha] + \chi_- P[\mathrm{G}\alpha] \\ &= P[\chi_+\mathrm{G}_-\alpha] + P[\chi_-\mathrm{G}_+\alpha] \\ &\quad - P[\chi_-\mathrm{G}_+\alpha] + P[\chi_-\mathrm{G}_-\alpha] \\ &= P[(\chi_+ + \chi_-)\mathrm{G}_-\alpha] \\ &= P[\mathrm{G}_-\alpha] \\ &= \alpha, \end{aligned}$$

where Equation (5.5) was used in the second step, and Equation (5.3) in the final step;

$$\begin{aligned} \forall\ \psi \in \Gamma_{\mathrm{sc},\partial} \quad (\mathrm{G} \circ P_\chi + \mathrm{G}_\chi \circ P)[\psi] &= \mathrm{G}_+[P_\chi\psi] - \mathrm{G}_-[P_\chi\psi] \\ &\quad + \mathrm{G}_+[\chi_+P\psi] + \mathrm{G}_-[\chi_-P\psi] \\ &= \mathrm{G}_+[P[\chi_+\psi] - \chi_+P\psi] - \mathrm{G}_-[-P[\chi_-\psi] + \chi_-P\psi] \\ &\quad + \mathrm{G}_+[\chi_+P\psi] + \mathrm{G}_-[\chi_-P\psi] \\ &= \mathrm{G}_+[P[\chi_+\psi]] + \mathrm{G}_-[P[\chi_-\psi]] \\ &= \chi_+\psi + \chi_-\psi \\ &= \psi, \end{aligned}$$

where the assumptions made on $\chi = \{\chi_+, \chi_-\}$ were used at various points and Equation (5.4) was used in the second to last step;

$$\begin{aligned} \forall\ \alpha \in \tilde{\Gamma}_{\mathrm{sc}} \quad (P \circ \mathrm{G}_\chi)[\alpha] &= P[\mathrm{G}_+[\chi_+\alpha]] + P[\mathrm{G}_-[\chi_-\alpha]] \\ &= (\chi_+ + \chi_-)\alpha \\ &= \alpha, \end{aligned}$$

which again holds simply in virtue of Equation (5.3) together with the assumption that $\chi_\pm$ sends $\tilde{\Gamma}_{\mathrm{sc}} \to \tilde{\Gamma}_\pm$. □

The significance of Theorem 5.2.3 is as follows: although we have not explicitly described, so far, the nature of the choice of domain Γ_∂, in practice what we have in mind here is that this be determined by a choice of *boundary conditions* to be satisfied by the field variables on the (non-empty) boundary of M. Now assume that $\mathscr{M}$ is a globally hyperbolic STT∂ and that it admits two smooth spacelike Cauchy surfaces $\mathscr{C}_1$, $\mathscr{C}_2$ with $\mathscr{C}_2 \subset I^+(\mathscr{C}_1)$. Then let $\chi_+ \in C^\infty(M)$ be equal to 1 on $J^+(\mathscr{C}_2)$ and equal to 0 on $J^-(\mathscr{C}_1)$. With $\chi_- \coloneqq 1 - \chi_+$, the pair $\chi = \{\chi_+, \chi_-\}$ will automatically satisfy all the hypotheses of Theorem 5.2.3, save possibly for the requirements that $\Gamma_{\mathrm{sc},\partial} \to \Gamma_{\pm,\partial}$ and $\Gamma_{\mp,\partial} \to \Gamma_{0,\partial}$ under multiplication by $\chi_\pm$. In the case of a Dirichlet boundary condition for a scalar field, i.e. $\Gamma_\partial = \{\phi \in C^\infty(M) \mid \phi\restriction_{\partial M} = 0\}$, then actually the latter requirements are also automatically satisfied. However, the situation is different already in the still very simple case of the *Neumann* boundary condition for a scalar field, i.e. $\Gamma_\partial = \{\phi \in C^\infty(M) \mid (\nabla_\nu\phi)\restriction_{\partial M} = 0\}$ where ν is the unit normal vector field along ∂M. There, we see that χ_+ needs to satisfy the additional requirement $(\nabla_\nu\chi_+)\restriction_{\partial M} = 0$. Proving the existence of a χ_+ satisfying this requirement while also being identically equal to 1 on $J^+(\mathscr{C}_2)$ and to 0 on $\mathscr{C}_1$ seems to us to be a typical Whitney-type extension problem (see Section 2.5.1). We leave its definitive resolution, and also the consideration of more general boundary conditions, to future investigation.

5.2.2 A few words on adjoints

Definition 5.2.4. Two differential operators $P, P^\dagger : \Gamma \to \tilde{\Gamma}$ are *mutually adjoint* if there exists a bilinear bidifferential operator $\mathcal{G} : \Gamma \times \Gamma \to \Omega^n(M)$, which we call a *Green form*, such that

$$[P\phi]\psi - [P^\dagger\psi]\phi = \mathrm{d}\mathcal{G}(\phi, \psi) \qquad \forall\ \phi, \psi \in \Gamma. \tag{5.12}$$

It can be shown that for any P there is a unique $P^\dagger$ such that P and $P^\dagger$ are mutually adjoint. This is the Green–Vinogradov formula, as explained at the end of Section 2.4.1. However, the form-valued operator $\mathcal{G}$ is not uniquely determined: it is only determined up to the addition of a locally exact bilinear form.

Notice there is a bilinear pairing $\langle\cdot,\cdot\rangle_\pm$ between $\tilde{\Gamma}_\pm$ and $\Gamma_{\mp,\partial}$, defined by

$$\langle\alpha, \phi\rangle_\pm = \int_M \alpha\phi, \qquad \alpha \in \tilde{\Gamma}_\pm,\ \phi \in \Gamma_{\mp,\partial}.$$

If the pairing is weakly non-degenerate then we can use it to embed $\tilde{\Gamma}_\pm$ and $\Gamma_{\mp,\partial}$ into the algebraic dual spaces $(\Gamma_{\mp,\partial})^*$ and $(\tilde{\Gamma}_\pm)^*$ respectively.

Now let P and $P^\dagger$ be mutually adjoint differential operators, and assume both of them to be Green hyperbolic with respect to (the same causal structure and) the same domain Γ_∂. Denote the advanced $(-)$ and retarded $(+)$ Green operators of P and $P^\dagger$ by $\mathrm{G}_\pm$

and $\mathrm{G}^\dagger_\pm$ respectively. Consider $\mathrm{G}_\pm : \tilde{\Gamma}_\pm \to \Gamma_{\pm,\partial}$. Its transpose $(G_\pm)^*$ is naturally a map $(\Gamma_{\pm,\partial})^* \to (\tilde{\Gamma}_\pm)^*$. For $\alpha \in \tilde{\Gamma}_\mp$ and $\beta \in \tilde{\Gamma}_\pm$, one has

$$\begin{aligned}
[(\mathrm{G}_\pm)^* \langle \alpha, \cdot \rangle_\mp]\beta = \langle \alpha, \mathrm{G}_\pm \beta \rangle_\mp &= \int_M \alpha \mathrm{G}_\pm \beta \\
&= \int_M [P^\dagger \mathrm{G}^\dagger_\mp \alpha] \mathrm{G}_\pm \beta \\
&= \int_M [P \mathrm{G}_\pm \beta] \mathrm{G}^\dagger_\mp \alpha - \int_M \mathrm{d}\mathcal{G}(\mathrm{G}_\pm \beta, \mathrm{G}^\dagger_\mp \alpha) \\
&= \int_M \beta \mathrm{G}^\dagger_\mp \alpha - \int_{\partial M} \mathcal{G}(\mathrm{G}_\pm \beta, \mathrm{G}^\dagger_\mp \alpha) \\
&= \langle \beta, \mathrm{G}^\dagger_\mp \alpha \rangle_\pm - \int_{\partial M} \mathcal{G}(\mathrm{G}_\pm \beta, \mathrm{G}^\dagger_\mp \alpha) \qquad (5.13)
\end{aligned}$$

where, in the second to last step, both Stokes' Theorem and Equation (5.12) – with $\phi = \mathrm{G}_\pm \beta$ and $\psi = \mathrm{G}^\dagger_\mp \alpha$ – were used. A similar result relates $(\mathrm{G}^\dagger_\pm)^*$ with $\mathrm{G}_\mp$ and contributions coming from boundary integrals.

5.3 Initial value problems for wave equations on globally hyperbolic STT∂s with Dirichlet boundary conditions

In this section, $P : C^\infty(M) \to C^\infty(M)$ is a second order normally hyperbolic (see Section 2.4.2) partial differential operator with smooth coefficients on the STT∂ $\mathscr{M} = (M, g, \mathfrak{t})$. That is,

$$Pu = \Box_g u + X[u] + \mathcal{V} u$$

where $\Box_g = g^{ab} \nabla_a \nabla_b$ is the d'Alembert operator associated with the metric, X is a smooth vector field, and $\mathcal{V}$ is a smooth function.

Lemma 5.3.1. *Let N be a smooth manifold with or without boundary and $F \xrightarrow{\pi} N$ be a real or complex vector bundle. Suppose f and h are continuous real or complex-valued functions on F which restrict to homogeneous functions of degree 2 (for example, but not necessarily, to quadratic forms) on each fiber of $F \xrightarrow{\pi} N$. Suppose h is positive-definite on each fiber and let $X \subseteq N$ be compact. Then there exists a constant $C' > 0$ such that*

$$|f(Z)| \leq C' h(Z)$$

for all $Z \in \pi^{-1} X$. If f is also positive-definite then there exists another constant $C > 0$ such that

$$f(Z) \geq C h(Z)$$

for all $Z \in \pi^{-1} X$.

Proof. This is a standard type of argument, notably used to prove that any two Riemannian metrics on a compact manifold are Lipschitz equivalent. Pick a smooth fiberwise Riemannian or Hermitian metric g on F and let

$$UF = \{Z \in F : g(Z, Z) = 1\}.$$

Consider the function $k : UF \to \mathbb{R}$ defined by $k(Z) = |f(Z)|/h(Z)$. Restricted to the compact subspace $\pi^{-1}X \cap UF$, k is a continuous function (as h never vanishes there), strictly positive if f is positive-definite. Therefore, it is bounded from above. If f is positive-definite then k is also bounded from below by a positive constant. The result follows. □

The following theorem is essentially an adaptation of Lemma 12.28 in [Rin09], aided by considerations (specifically, the choice of vector field in the final part the proof) made in [Hör94, p. 419]. The experienced PDE theorist would immediately recognise that, aside from the (very important) Lorentzian-geometric aspects of the proof, the core of the argument relies on obtaining an *energy identity* which, if coupled with a judicious choice of vector field, can be used to obtain useful estimates on domains 'swept out' appropriately by spacelike hypersurfaces. The use of such *energy estimates* in solving both linear and nonlinear hyperbolic PDEs has a very long and illustrious history, and we refer to [Går98] for a beautifully written review. The theorem we will prove is a local uniqueness result for solutions of the mixed Cauchy–Dirichlet problem, which relies on the existence of temporal functions with specific properties locally around a Cauchy surface in a globally hyperbolic STT∂. More precisely: assuming the existence of such temporal functions, the theorem establishes a precise *domain of dependence* property which establishes the *precise finite speed of propagation* of solutions satisfying Dirichlet boundary conditions. By 'precise' here, following [JMR05, Rau05], we mean that the speed of propagation of the initial data is not only finite, but is further constrained by the Lorentzian causal structure. This is of course well-known in the case of normally hyperbolic operators on globally hyperbolic spacetimes without boundaries, and it is the fundamental ingredient in proving that such operators are always Green hyperbolic. Our initial motivation for seeking a result such as Theorem 5.3.2 came from wanting to prove that normally hyperbolic differential operators are still Green hyperbolic – in the sense of Section 5.2.1 – when a timelike boundary is present, provided that the resulting spacetime is a globally hyperbolic STT∂ and that a variety of suitable boundary conditions are imposed. Theorem 5.3.2 below is the only preliminary result in this direction which we will prove here, and considers the case of Dirichlet boundary conditions. We show that the latter conditions are (unsurprisingly!) 'suitable' for Green hyperbolicity in the large, at least in principle.

Unfortunately, we will not be able to fully complete the program just described, even in the case of Dirichlet conditions. In the concluding section of this chapter, we will illustrate what we believe to be the main (and, in fact, probably also the only) details still to be worked out. In particular, we have not constructed temporal functions of the kind we will assume exist in the statement of Theorem 5.3.2, or even abstractly proved their existence, despite finding the latter very plausible.

Theorem 5.3.2 (Local uniqueness for wave equations on globally hyperbolic STT∂s with Dirichlet boundary conditions)**.** *Let $\mathscr{M} = (M, g, \mathfrak{t})$ be an $(n{+}1)$-dimensional globally hyperbolic STT∂ with a smooth spacelike Cauchy hypersurface $\mathscr{C}$. Let $\widetilde{\mathscr{M}} = (\widetilde{M}, \tilde{g}, \tilde{\mathfrak{t}})$ be an extension of $\mathscr{M}$ to a spacetime without boundary, so that there is an isometric time-orientation–preserving embedding $i : \mathscr{M} \to \widetilde{\mathscr{M}}$ such that $i(M)$ is closed in $\widetilde{M}$. Assume that $p \in I^+_{\mathscr{M}}(\mathscr{C})$ is such that there is a real-valued function t defined on an open neighbourhood W of $I^-_{\mathscr{M}}(p) \cap J^+_{\mathscr{M}}(\mathscr{C})$ and satisfying the following properties:*

1. *t is smooth and temporal on W;*
2. *letting $D = J^-_{\mathscr{M}}(p) \cap J^+_{\mathscr{M}}(\mathscr{C})$ so that $\mathring{D} = I^-_{\mathscr{M}}(p) \cap I^+_{\mathscr{M}}(\mathscr{C})$, if $t(\mathring{D}) = (t_1, t_2)$*[2] *(where we allow $t_2 = \infty$) then, for each $\tau \in (t_1, t_2)$ the level sets*
$$\mathscr{S}_\tau := t^{-1}\{\tau\}$$
are contained in $I^-_{\mathscr{M}}(p)$;
3. *there exists a $\tilde{t} \in (t_1, t_2)$ such that, for all $\tau \in (\tilde{t}, t_2)$, the sets*
$$L_\tau = I^-_{\mathscr{M}}(\mathscr{S}_\tau) \cap I^+_{\mathscr{M}}(\mathscr{C}) \subseteq \mathring{D},$$
viewed as embedded submanifolds with boundary of the extended spacetime $\widetilde{M}$, have
    - *(compact) closure entirely contained in $I^-_{\mathscr{M}}(p) \cap J^+_{\mathscr{M}}(\mathscr{C})$;*
    - *locally Lipschitz topological boundary which is the union of a set of zero measure with the following disjoint embedded hypersurfaces in $\widetilde{M}$: (a) a portion $\mathscr{B}_\tau$ of $\partial M \subset \widetilde{M}$; (b) a portion $\mathscr{I}_\tau$ of $J^-_{\mathscr{M}}(p) \cap \mathscr{C} \cap \operatorname{Int} M$; (c) a portion $\mathscr{S}'_\tau$ of $\mathscr{S}_\tau \cap I^+_{\mathscr{M}}(\mathscr{C}) \cap \operatorname{Int} M$.*

Furthermore let $V \supseteq J^-_{\mathscr{M}}(p) \cap J^+_{\mathscr{M}}(\mathscr{C})$ be open. If $u \in C^2(V)$ satisfies $Pu = 0$ on $I^-_{\mathscr{M}}(p) \cap I^+_{\mathscr{M}}(\mathscr{C})$ with $u = 0$ on $\partial M \cap V$, then u and $\mathrm{d}u$ *vanish on $J^-_{\mathscr{M}}(p) \cap J^+_{\mathscr{M}}(\mathscr{C})$ iff they vanish on $J^-_{\mathscr{M}}(p) \cap \mathscr{C}$.*

Proof. Of course the 'only if' statement is automatic, and we proceed to prove the other implication. In virtue of Lemma 5.1.9, and since u and $\mathrm{d}u$ are continuous on $V \supseteq D$

[2] Notice that, under the previous assumptions, $t(\mathring{D}) \subset \mathbb{R}$ is necessarily an open interval bounded from below.

and vanish on $J^-_{\mathscr{M}}(p) \cap \mathscr{C} \supseteq \dot{J}^-_{\mathscr{M}}(p) \cap \mathscr{C}$ by hypothesis, we only need to prove that u and $\mathrm{d}u$ vanish in the interior $\mathring{D} = I^-_{\mathscr{M}}(p) \cap I^+_{\mathscr{M}}(\mathscr{C})$.
As preparation we prove that, under the hypotheses of the theorem, for any $q \in \mathring{D}$ and any future-directed timelike curve $\gamma : [0, 1] \to M$ from q to p, we have

$$(t \circ \gamma)[0, 1) = [t(q), t_2). \tag{5.14}$$

Since t is temporal on $\mathring{D}$ and $\gamma[0, 1)$ is entirely contained in $\mathring{D}$, it follows that $t \circ \gamma\restriction_{[0,1)}$ is non-decreasing. Now let $t(q) < \tau < t_2$ and pick a point $r \in \mathring{D}$ such that $t(r) = \tau$. Then it is possible to find a neighbourhood U of p such that for any point z in $U \cap I^-_{\mathscr{M}}(p)$ there is a future-directed timelike curve ρ from r to z. In particular, we can pick z to be $\gamma(\lambda)$ for λ sufficiently close to 1. Since t is non-decreasing along ρ, we conclude that $\tau = t(r) \leq t(z) = t(\gamma(\lambda))$. Thus we have found that $(t \circ \gamma)[0, 1) \supseteq [t(q), \tau] \; \forall \, t(q) \leq \tau < t_2$, which readily implies Equation (5.14).
We now argue, using Equation (5.14), that

$$\mathring{D} = \bigcup_n L_{\tau_n} \tag{5.15}$$

where $(\tau_n)_n$ is any sequence such that $t_1 < \tau_n < t_2$ and $\tau_n \to t_2$ as $n \to \infty$. The inclusion $\supseteq$ is automatic. So let $q \in \mathring{D}$. Then $t_1 < t(q) < t_2$ so there exists an integer n such that $t(q) < \tau_n$; but by Equation (5.14), there exists a future-directed timelike curve from q to a point z_n with $t(z_n) = \tau_n$, i.e. to a point $z_n \in \mathscr{S}_{\tau_n}$. So $q \in L_{\tau_n}$, which completes the proof of Equation (5.15).
In particular, we can without loss of generality assume that the sequence $(\tau_n)_n$ in Equation (5.15) has the property that $\tilde{t} < \tau_n < t_2 \; \forall \; n$, with $\tilde{t}$ as in the hypothesis of the theorem. Then to prove the theorem it suffices to show that, for any τ with $\tilde{t} < \tau < t_2$, u and $\mathrm{d}u$ vanish on L_τ. The idea behind the proof of this fact is to construct a suitable vector field out of u and $\mathrm{d}u$, to integrate its divergence over L_τ, and to use the divergence theorem together with the assumptions made on the structure of the topological boundary of L_τ as a subset of the extended manifold (without boundary) $\widetilde{M}$. Define the first-order, tensor-valued bilinear bidifferential operator

$$T[u, v] = \mathcal{T}(j^1 u, j^1 v) = \mathrm{d}u \otimes \mathrm{d}v - \frac{1}{2} g \, g^{-1}(\mathrm{d}u, \mathrm{d}v).$$

[We refer to Section 2.5.1 for notions of jet bundles and jet extensions of sections; in particular, we are using here the isomorphism $J^1(M \times \mathbb{R}) \cong T^*M \times \mathbb{R}$ mentioned there.]
It holds that, for any point p and any two future-directed timelike vectors X, Y at p, the quadratic form

$$T^*_p M \ni \omega \mapsto \mathcal{T}(\omega, \omega)_{ab} X^a Y^b = \omega(X)\omega(Y) - \frac{1}{2} g(X, Y) g^{-1}(\omega, \omega) \tag{5.16}$$

is positive-definite. I.e. we are defining $\mathcal{T}_{ab}$ to be the stress-energy tensor of the massless scalar field, and the positive-definiteness of (5.16) is simply a statement of the well-known *dominant energy condition* which is enjoyed by this quantity, even 'off-shell'. A simple calculation shows that

$$\nabla^a T[u,u]_{ab} = (\Box_g u)(\mathrm{d}u)_b$$

so that, if Y is a vector field, then

$$\nabla^a (T[u,u]_{ab} Y^b) = Y(u)\Box_g u + T[u,u]_{ab}\nabla^a Y^b.$$

Now let $\xi[u]^a = e^{-\lambda t} T[u,u]^a{}_b Y^b$ where $\lambda > 0$ is a constant. Then

$$\operatorname{div}\xi[u] = e^{-\lambda t}\left\{Y[u]\Box_g u + T[u,u]_{ab}\nabla^a Y^b - \lambda T[u,u]_{ab}\nabla^a t\, Y^b\right\}.$$

Letting also $\eta[u] = e^{-\lambda t} u^2 Y$ we have

$$\operatorname{div}\eta[u] = e^{-\lambda t}\left\{-\lambda Y[t]u^2 + 2uY[u] + \operatorname{div} Y u^2\right\}$$

from which we obtain

$$\begin{aligned}\operatorname{div}(\xi[u] + \eta[u]) &+ \lambda e^{-\lambda t}\left\{T[u,u]_{ab}\nabla^a t\, Y^b + Y[t]u^2\right\}\\ &= e^{-\lambda t} Y[u]\Box_g u + e^{-\lambda t}\left\{T[u,u]\nabla^a Y^b + (\operatorname{div} Y)u^2 + 2uY[u]\right\}\end{aligned} \quad (5.17)$$

and, in turn,

$$\begin{aligned}&\operatorname{div}(\xi[u] + \eta[u]) + \lambda e^{-\lambda t}\{\overbrace{T[u,u]_{ab}\nabla^a t\, Y^b + Y[t]u^2}^{(I)}\}\\ &= -e^{-\lambda t} Y[u]Pu + e^{-\lambda t}\{\underbrace{T[u,u]\nabla^a Y^b + (\operatorname{div} Y)u^2 + Y[u](X[u] + \mathcal{V}u + 2u)}_{(II)}\}.\end{aligned} \quad (5.18)$$

The exact forms of the terms (I) and (II) above are not essential. What matters for the analysis which will follow is that:

- If Y and ∇t are defined, continuous and future-directed timelike everywhere on an open set $U \subseteq V$, and if additionally $Y[t] = g_{ab}Y^a\nabla^b t$ is strictly positive on U, then (I) on U defines a pointwise positive-definite quadratic form in u and its first derivatives, i.e. it is a pointwise positive-definite quadratic form on the first-order jet bundle $J^1E \to U$ of the trivial line bundle $E = U \times \mathbb{R} \to U$.
- (II) is too a (not necessarily positive definite or even semi-definite) quadratic form on $J^1E \to U$.

We now wish to use Lemma 5.3.1, with the vector bundle $J^1E \to U$ playing the role of $F \xrightarrow{\pi} N$, the quadratic forms defined by either of the differential expressions (I) and (II) playing the role of f, h being any auxiliary fiberwise positive-definite quadratic form on J^1E, and the closure of L_τ in U playing the role of the subset X. If U is an open neighbourhood of $I^-_{\mathscr{M}}(p) \cap J^+_{\mathscr{M}}(\mathscr{C})$ then we know (by hypothesis) that this closure is indeed contained in U. In particular, we may choose U to be the neighbourhood $W \subset I^-_{\mathscr{M}}(p) \cap J^+_{\mathscr{M}}(\mathscr{C})$ on which t is known to be temporal. As far as the application of Lemma 5.3.1 is concerned, it suffices to choose the vector field Y to be any other future-directed timelike vector field on W – for then $Y[t] = g_{ab}Y^a\nabla^b t$ is strictly positive on W. However, the rest of the proof our theorem will dictate a more stringent choice for Y, which will arise naturally below. For now, we conclude that there exist constants C_1, $C_2 > 0$ such that, on $\overline{L_\tau}$,

$$(I) \geq C_1 h(j^1u) \quad \text{and} \quad (II) \leq C_2 h(j^1u) \tag{5.19}$$

and thus Equation (5.18) leads to the inequality

$$\operatorname{div}(\xi[u] + \eta[u]) \leq -e^{-\lambda t} Y[u] P u - [\lambda C_1 - C_2] e^{-\lambda t} h(j^1 u). \tag{5.20}$$

on $\overline{L_\tau}$ and for all $\lambda > 0$. In particular, we may pick λ large enough so that $\lambda C_1 - C_2 = C_3 > 0$. Now if $Pu = 0$ throughout $\overline{L_\tau}$, as we are assuming, then this further reduces to

$$\operatorname{div}(\xi[u] + \eta[u]) \leq -C_3 e^{-\lambda t} h(j^1 u). \tag{5.21}$$

The function on the right-hand side is continuous on $\overline{L_\tau}$, is either negative or zero, and vanishes only at those points at which u and $\mathrm{d}u$ are equal to zero, because h was chosen positive-definite. Therefore,

$$\int_{\overline{L_\tau}} \operatorname{div}(\xi[u] + \eta[u])\, \mathrm{d}\mu_g \leq 0 \qquad \text{with equality iff } u \text{ and } \mathrm{d}u \text{ are identically zero on } \overline{L_\tau}.$$

We now show, and it is at this point that we use the hypothesis that u and $\mathrm{d}u$ vanish on $J^-_{\mathscr{M}}(p) \cap \mathscr{C}$ and that u vanishes on ∂M,[3] that if Y is chosen appropriately then we must also have

$$\int_{\overline{L_\tau}} \operatorname{div}(\xi[u] + \eta[u])\, \mathrm{d}\mu_g \geq 0. \tag{5.22}$$

Namely, let Y be a vector field with the property that it is everywhere future-directed and timelike, and that additionally, on ∂M, it is either tangent to ∂M or points towards M when the latter is regarded as a submanifold of $\widetilde{M}$. Below, we will discuss how one such vector field may be globally constructed, but for now we assume one is given.

[3] But not the fact that $Pu = 0$!

Using the assumptions on the regularity of the boundary made in the statement of the theorem, by the generalisation of Theorem 2.5.13 involving locally Lipschitz boundaries we see that if it is the case that the vector field

$$\xi[u] + \eta[u] = e^{-\lambda t}\big(T[u,u]^a{}_b Y^b + u^2 Y^a\big)$$

is either outward pointing or tangent to each smooth portion $\mathscr{B}_\tau$, $\mathscr{I}_\tau$ and $\mathscr{S}'_\tau$, then Inequality (5.22) follows.

($\mathscr{B}_\tau$) On $\partial M \cap V \supseteq \mathscr{B}_\tau$, u vanishes. Therefore, on $\mathscr{B}_\tau$, $\mathrm{d}u = \varphi n$ where φ is real-valued and n is an outward-directed conormal field along $\mathscr{B}_\tau$. There, we have as a result

$$\xi[u] + \eta[u] = e^{-\lambda t}\varphi^2 \left\{ n(Y)n^\sharp - \frac{1}{2} g^{-1}(n,n) Y \right\}, \tag{5.23}$$

and, since $n(n^\sharp) = g^{-1}(n,n)$, we see that

$$n(\xi[u] + \eta[u]) = \frac{e^{-\lambda t}\varphi^2}{2} g^{-1}(n,n) n(Y).$$

Now, $n(Y) \le 0$ because n was assumed outward directed and Y tangent or inward pointing, and $g^{-1}(n,n) < 0$ because n is conormal to a timelike hypersurface. It follows then that $n(\xi[u] + \eta[u]) \ge 0$, i.e. that $\xi[u] + \eta[u]$ is outward directed along $\mathscr{B}_\tau$, or tangent to $\mathscr{B}_\tau$.

($\mathscr{I}_\tau$) This is clear, since u and $\mathrm{d}u$ are both assumed to vanish on $J^-_{\mathscr{M}}(p) \cap \mathscr{C} \supseteq \mathscr{I}_\tau$.

($\mathscr{S}'_\tau$) Since $\mathscr{S}'_\tau$ is part of a level set of t and since $\mathrm{d}t$ never vanishes, $(\mathrm{d}t)\restriction_{\mathscr{S}'_\tau}$ is conormal to $\mathscr{S}'_\tau$. Since $\operatorname{grad} t$ is future-directed timelike and t increases strictly along future-directed causal curves, it is actually an outward conormal. Then,

$$\mathrm{d}t(\xi[u] + \eta[u]) = e^{-\lambda t}\big\{T[u,u]_{ab}(\nabla^a t)Y^b + u^2 g(\operatorname{grad} t, Y)\big\}$$

and all terms are positive due to the properties of T and to the fact that both $\operatorname{grad} t$ and Y are future-directed and timelike.

We conclude by proving the global existence of Y with the properties claimed above. Let Θ be a global, timelike, future-directed vector field on M – which exists since M is time-orientable. Let $f : M \to \mathbb{R}$ be a boundary-defining function for M [Lee13, p. 118], assumed w.l.o.g. such that $\mathrm{d}f$ is inward pointing on ∂M, and consider the open (in M) set

$$\mathcal{O} := \left\{ p \in M \;\middle|\; g^{-1}(\mathrm{d}f(p), \mathrm{d}f(p)) < 0 \right\}.$$

Then $\partial M \subseteq \mathcal{O}$ since ∂M is timelike. Now, consider the vector field Ξ defined on $\mathcal{O}$ by

$$\Xi^a = \Theta^a - \frac{\Theta^c \nabla_c f}{(\nabla^d f)(\nabla_d f)} \nabla^a f.$$

We calculate

$$\Xi^a \Xi_a = \Theta^a \Theta_a - 2\frac{(\Theta^a \nabla_a f)^2}{(\nabla^d f)(\nabla_d f)} + \frac{(\Theta^c \nabla_c f)^2}{[(\nabla^d f)(\nabla_d f)]^2}(\nabla^a f)(\nabla_a f) = \Theta^a \Theta_a - \frac{(\Theta^a \nabla_a f)^2}{(\nabla^d f)(\nabla_d f)} > 0,$$

so Ξ is also timelike. Furthermore,

$$\Xi^a \Theta_a = \Theta^a \Theta_a - \frac{(\Theta^a \nabla_a f)^2}{(\nabla^d f)(\nabla_d f)} = \Xi^a \Xi_a > 0,$$

so Ξ is future-directed. Ξ is tangent to ∂M since

$$\mathrm{d}f(\Xi) = \Xi^a \nabla_a f = \Theta^a \nabla_a f - \frac{\Theta^c \nabla_c f}{(\nabla^d f)(\nabla_d f)} (\nabla^a f)(\nabla_a f) = 0.$$

Consider the open cover of M given by $\{\operatorname{Int} M, \mathcal{O}\}$. Pick a (non-negative) partition of unity $\{\chi_1, \chi_2\}$ subordinate to this cover, and let

$$Y \coloneqq \chi_1 \Theta + \chi_2 \Xi \in \mathfrak{X}(M).$$

Then Y is future-directed timelike since it is the sum of two future-directed timelike vector fields, and it is tangent to ∂M since it equals Ξ there. We can even tweak this construction to obtain a future-directed timelike vector field which is inward pointing on ∂M, by defining Y instead as

$$Y = \chi_1 \Theta + \chi_2 \Xi'$$

with χ_1, χ_2, Θ as above, and

$$\Xi'^a \coloneqq \Xi^a - \delta \nabla^a f$$

for any $\delta \in C^\infty(\mathcal{O})$ positive and such that

$$\delta^2 < -\frac{\Xi^a \Xi_a}{(\nabla^a f)(\nabla_a f)}.$$

With these definitions, it is easy to check that Ξ'^a is future-directed and timelike on $\mathcal{O}$, and it is additionally inward pointing on ∂M. The claim on Y follows. $\square$

5.4 Discussion and outlook

Suppose we could prove that any globally hyperbolic STT∂ $\mathscr{M} = (M, g, \mathfrak{t})$ admits a (surjective, smooth) Cauchy temporal function $t : M \to \mathbb{R}$, in which case let $\mathscr{C}_\tau := t^{-1}\{\tau\}$. What's more, suppose that we could prove that, given any smooth spacelike Cauchy surface $\mathscr{C}$, there exists a Cauchy temporal function for which $\mathscr{C}$ is a level set. Consider the notation of Section 5.2.1 and assume that Γ_∂ is determined by homogeneous boundary conditions, so that for any open subset O of M we can consider the smooth sections of the field bundle restricted to O and satisfying the same boundary conditions, and denote them by Γ_∂^O – and the zero section of $F|_O$ always belongs to Γ_∂^O. Finally, suppose that we could prove following PDE result (to compare with Theorem 24.4.1 in [Hör94]):

'Theorema demonstrandum'. For any $f \in \tilde{\Gamma}_0$, there exists a unique solution $u_\pm \in \Gamma_\partial$ of $Pu = f$ such that u_+ [resp. u_-] vanishes on $t^{-1}(-\infty, \tau]$ [resp. on $t^{-1}[\tau, +\infty)$] whenever $\tau \in \mathbb{R}$ is such that $\operatorname{supp} f \subset I^+(\mathscr{C}_\tau)$ [resp. $\operatorname{supp} f \subset I^-(\mathscr{C}_\tau)$].

Remark. Notice that, under the other assumptions made in the first paragraph of this section, it would then automatically follow that $\operatorname{supp} u_\pm \subseteq J^\pm_{\mathscr{M}}(\operatorname{supp} f)$ by the following argument:

$$\mathscr{M}_\pm := (M_\pm = M \setminus J^\pm_{\mathscr{M}}(\operatorname{supp} f), g\restriction_{M_\pm}, \mathfrak{t}\restriction_{M_\pm})$$

is also a globally hyperbolic STT∂ (by Corollary 5.1.8), and $\mathscr{C}_\tau$ is entirely contained in M_+ [resp. in M_-] whenever $\tau \in \mathbb{R}$ is such that $\operatorname{supp} f \subset I^+_{\mathscr{M}}(\mathscr{C}_\tau)$ [resp. $\operatorname{supp} f \subset I^-_{\mathscr{M}}(\mathscr{C}_\tau)$] – since this implies that $J^+_{\mathscr{M}}(\operatorname{supp} f) \subseteq J^+_{\mathscr{M}}\big(I^+_{\mathscr{M}}(\mathscr{C}_\tau)\big) = I^+_{\mathscr{M}}(\mathscr{C}_\tau)$. Hence, $\mathscr{C}_\tau$ is a Cauchy surface for $\mathscr{M}_+$ [resp. for $\mathscr{M}_-$]. Under the assumptions in the first paragraph of this section, we can then find a Cauchy temporal function t_+ [resp. t_-] for $\mathscr{M}_+$ [resp. for $\mathscr{M}_-$] such that $\mathscr{C}_\tau = t_+^{-1}\{\tau\}$ [resp. such that $\mathscr{C}_\tau = t_-^{-1}\{\tau\}$]. The zero section of the field bundle restricted to M_+ [resp. to M_-] clearly solves $Pu = 0$ on M_+ [resp. on M_-], belongs to $\Gamma_\partial^{M_+}$ [resp. to $\Gamma_\partial^{M_-}$], and vanishes on $t_+^{-1}(-\infty, \tau]$ [resp. on $t_-^{-1}[\tau, +\infty)$], and it is the unique such section. But $u_\pm\restriction_{M_\pm}$ also satisfies those requirements (since $f\restriction_{M_\pm} = 0$). This immediately yields the claim on $\operatorname{supp} u_\pm$.

Assuming that our 'Theorema demonstrandum' holds as stated, i.e. for inhomogeneities f with compact support, then we expect that actually, owing to global hyperbolicity, one may extend it, and correspondingly also the statement on supports made in the Remark above, to inhomogeneities with more general causally restricted supports. E.g. it should be possible to adapt the proof of Theorem 3.8 in [Bär15] (see also the end of the proof of Theorem 12.17 in [Rin09], or the proof of Corollary 5 in [Gin09]) to show that f can be allowed to have retarded/advanced support, or even past/future compact support. So at the very least, one could define retarded/advanced Green operators $\mathrm{G}_\pm : \tilde{\Gamma}_\pm \to \Gamma_{\pm,\partial}$

in the sense of Definition 5.2.1, by setting

$$\mathrm{G}_{\pm} f = u_{\pm} \quad \text{for all } f \in \tilde{\Gamma}_{\pm}.$$

These will be the inverses to $P : \Gamma_{\pm,\partial} \to \tilde{\Gamma}_{\pm}$ by Proposition 5.2.2. This in turn demonstrates that using a different Cauchy temporal function in the original statement of our 'Theorema demonstrandum' must yield the same solution $u_{\pm}$.

With $\mathrm{G}_{\pm}$ as just constructed, one can then establish the exactness of sequence (5.7) as was done in the proof of Theorem 5.2.3, provided the boundary conditions are such that cutoff functions $\chi_{\pm}$ can be found as required by the statement of that theorem – see the discussion on this point at the end of Section 5.2.1. In turn, this would prepare the ground for the rigorous quantization of linear systems with boundary conditions.

In view of these considerations, given a linear differential operator P on a globally hyperbolic STT∂ and a choice of homogeneous boundary conditions, Green hyperbolicity relative to the boundary conditions can be established if:

1.) Lorentzian-geometric results which are well-known in the boundaryless scenario can be extended to globally hyperbolic STT∂s: at least, the existence of (smooth) Cauchy temporal functions and additionally the existence of such functions adapted to any smooth spacelike Cauchy surface;

2.) a 'Theorema demonstrandum' can be proved for the given choice of boundary condition.

While we have separated the two issues for convenience, there is overlap as we now wish to explain. Let us now specialise to the case of normally hyperbolic operators. One approach for proving existence and uniqueness in the Cauchy problem in that case is constructive, and essentially dates back to the work of M. Riesz [Rie49] (whose arguments in turn built on Hadamard's pioneering work [Had23]) in which explicit expressions were given for the advanced and retarded fundamental solutions of the wave equation in Minkowski spacetime. This approach was then extended to local regions in a curved spacetime by Friedlander [Fri75] using geodesic normal coordinates. No energy identities or inequalities need to be used at any step in this approach. Finally, global advanced and retarded fundamental solutions on globally hyperbolic manifolds can be found by patching up local ones (regardless of how the latter were found in the first place), as was already sketched in [CB67] and was proved in full detail in [BGP07].

It seems to us that attempting to adapt Friedlander's local constructions to the case with timelike boundaries and boundary conditions would meet with important difficulties related to the unwieldy interplay between geodesics and boundaries. On the other hand, the approach to solving the Cauchy problem based on energy methods (see p. 164) seems more promising. It would also, presumably, extend to other kinds of 'hyperbolic' PDEs, and in particular to *symmetric hyperbolic systems* [Ger96]. In the boundaryless globally hyperbolic case, and for normally hyperbolic operators, Chapter 12 of Ringström's book [Rin09] presents a complete and rigorous account of how a global existence and uniqueness result, which includes our 'Theorema demonstrandum' as a special case, can be proved using energy methods. A rough outline of how this is achieved is as follows:

(a) Prove that, around any point p on a smooth spacelike Cauchy surface $\mathscr{C}$, there is an open neighbourhood U in M such that for any point in $U \cap J^+(\mathscr{C})$ the hypotheses of the equivalent of our local uniqueness/sharp finite speed of propagation result, Theorem 5.3.2, are satisfied. In particular, the required temporal functions exist [Rin09, Lem. 12.7 & Lem. 12.8].

(b) Use a Cauchy temporal function adapted to $\mathscr{C}$ and a connectedness argument to promote the local finite speed of propagation result in (a) to a global result stating that if the Cauchy data for the homogeneous problem has support in $K \subseteq \mathscr{C}$, then the support of any solution to the future of $\mathscr{C}$ is supported in $J^+(K)$ – dually, if the Cauchy data vanishes on $\Omega \subseteq \mathscr{C}$ then any solution to the homogeneous problem vanishes on the future Cauchy development of Ω [Rin09, Cor. 12.12].

(c) Prove a global existence result for linear wave equations on $\mathbb{R}^{n+1}$ with initial data on on $\{0\} \times \mathbb{R}^n$ [Rin09, Thm. 8.6].

(d) Prove global existence to the future of $\mathscr{C}$ by constructing local solutions in small neighbourhoods of each point, using (b) and compactness arguments to argue that the local solutions agree on overlaps and that any solution defined up to a given finite Cauchy time can be extended to one defined up to a strictly later Cauchy time [Rin09, Thm. 12.17].

(e) Similarly prove global existence to the past of $\mathscr{C}$.

Our view is that what will require the most attention in generalising the above steps to the case with boundary conditions are, in addition to point *1.)* above, items (a) and (d).

Actually, as we saw in the proof of Theorem 5.3.2, the Dirichlet boundary condition is such that a vector field Y may be chosen to yield the necessary estimates when the 'energy form' $T[u, u]$ defined there is employed. Any other boundary condition to be deemed suitable must share the same property, although possibly with respect to a different choice of energy form. To the best of our understanding, this point is not always tackled explicitly in the literature on hyperbolic mixed problems. Exceptions which we are aware of are the claimed characterisation of suitable oblique derivative conditions in [Går77] (see also [MS82, Sec. 6]), and the discussion on domain of dependence properties in [Sak82, Sec. 3.6] for the class of boundary conditions considered there (see below). We would therefore deem it a worthwhile future endeavour to revisit this literature and to translate their results into appropriate analogs of Theorem 5.3.2.

Concerning (d), there already is, at this stage, a large literature on hyperbolic mixed problems from which it should be possible to draw adequate local existence results for a variety of boundary conditions. For example: [Hör94, Lem. 24.1.6] for the case of Dirichlet boundary conditions, the aforementioned work in [Går77] for oblique derivative problems, [Kre75] for first-order systems, and [Sak82, GV96] for hyperbolic problems satisfying the *uniform Lopatinsky condition.*

It remains therefore to discuss the purely Lorentzian-geometric issues, namely *1.)* on p. 172 and (a) above. Concerning the latter, in [Rin09, Lem. 12.8] it is geodesic normal coordinates which provide the desired temporal function on a neighbourhood of $I^-(q) \cap J^+(\mathscr{C})$ whenever q is sufficiently close to $\mathscr{C}$. The required temporal function there is the Lorentz norm of the displacement vector from q, which is defined on $I^-(q) \cap V$ where V is a geodesic normal coordinate neighbourhood such that $J^-(q) \cap J^+(\mathscr{C}) \subseteq V$. If a timelike boundary is present, this strategy would appear to be inadequate. While we have not yet found an adequate substitute, a possibility may be as follows:

> Since $I^-(q)$ is a globally hyperbolic sub-STT∂ of $\mathscr{M}$, if *1.)* can be solved in full generality then a *Cauchy* temporal function for $I^-(q)$ may do the job of satisfying the hypotheses of Theorem 5.3.2, at least when q is close enough to the Cauchy surface $\mathscr{C}$.

Since this strategy is strictly speaking not yet known to work even in the boundaryless case, it would be of interest first to verify it there. At any rate, our reason for claiming that there is overlap between the resolution of issues *1.)* and *2.)* on p. 172 should now be clear: in the absence of convenient constructions such as geodesic normal coordinates (at least when the point of interest lies on the boundary), we need more general ways of 'sweeping out' domains of the form $I^-(p) \cap J^+(\mathscr{C})$, and tentatively propose that Cauchy temporal functions may provide one such way.

It then appears to be of paramount importance, if one is to make progress in this research project, to succeed in extending the purely Lorentz-geometric results on the existence of Cauchy temporal functions (adapted to Cauchy surfaces) to the case of globally hyperbolic STT∂s. Unfortunately, many of the available proofs, for instance the ones in [BS05] or in [CGM16], seem to again rely on the fine properties of the exponential map on Lorentzian manifolds without boundary! What we believe is needed is a family of arguments based on more widely applicable methods of analysis or differential topology.

Recently, Fathi and Siconolfi [FS12] (see also [Fat15]) have provided a proof of the existence of Cauchy temporal functions for general globally hyperbolic *cone structures* on manifolds, using techniques imported from weak Kolmogorov–Arnold–Moser (KAM) theory. We believe that, in principle, their proof has great potential to be adapted with minimal modifications to the case of globally hyperbolic STT∂s. There are, however, some subtleties: In the special case (which is what we're interested in) of cone structures arising from the field of future-directed causal (or zero) vectors on a time-oriented Lorentzian manifold, the definition of global hyperbolicity in [FS12] consists of stable causality together with the requirement that the images of all causal curves connecting two given points lie in a compact (possibly empty) subset. Additionally, those authors define causal curves using the 'locally Lipschitz' view, i.e. according to an analog of Definition 2.1.17. As we saw in Section 2.1.2 (see also [HE73, Prop. 6.6.2] or [FS12, Lem. 4.4]), in the case of spacetimes without boundaries this definition of global hyperbolicity is no different from the one using (strong) causality together with compactness of all causal diamonds, and similarly it makes no difference to adopt continuous or even piecewise-smooth curves to define causal pasts and futures instead. But in the case of STT∂s, all of this remains to be rigorously established. In particular, as we saw in Section 5.1.1, Solis [Sol06] – and we with him – set up the causal theory of STT∂s using continuous causal curves and not locally Lipschitz ones. This may not eventually be a serious restriction to prove the existence of Cauchy temporal functions, but it would still be of interest to provide an explicit proof that including all locally Lipschitz causal curves does not in fact change causal futures and pasts of points, i.e. to prove an analog of Theorem 2.1.18.[4]

Perhaps more seriously, stable causality plays a crucial role in the methods used in [FS12], so it would need to be shown that Solis' and our notion of global hyperbolicity of STT∂s implies their stable causality. Should this *not* be the case, there may be an argument that global hyperbolicity of STT∂ in the sense introduced by Solis is not the appropriate notion, after all!

[4]Note that proving an analog of Theorem 2.1.18 when the spacetime has no boundary but the metric is only assumed to be continuous is to this date an open problem in Lorentzian geometry, see [Säm16, Sec. 7].

Appendix A

Weyl quantization of linear systems, quasi-free states and one-particle structures

We give here a brief overview of the standard *Weyl-algebra* approach to the quantization of (real, bosonic) linear systems [Seg63, BR97]. The starting point is the realization that the phase space of the classical theory is a (real) symplectic vector space (S, σ). The first step is to construct the *Weyl algebra* [Sla72] *over* (S, σ), denoted here by $\mathscr{W}(S, \sigma)$. This is the C*-algebra generated by a unit element $\mathbb{1}$ and by *Weyl operators* $W(\Phi)$ (for all $\Phi \in S$) satisfying the relations

$$W(\Phi_1)W(\Phi_2) = e^{-i\sigma(\Phi_1,\Phi_2)/2}W(\Phi_1 + \Phi_2), \qquad W(\Phi)^* = W(-\Phi),$$

which are to be regarded as exponentiated versions of the standard canonical commutation relations (and in particular imply that each $W(\Phi)$ is unitary and that $W(\mathbf{0}) = \mathbb{1}$).

The Weyl algebra construction is functorial in the sense that for any two linear symplectic spaces (S_1, σ_1) and (S_2, σ_2) and for any linear symplectic map $T : S_1 \to S_2$, one defines in a natural way a *-homomorphism $\alpha : \mathscr{W}(S_1, \sigma_1) \to \mathscr{W}(S_2, \sigma_2)$ between the corresponding Weyl algebras by setting

$$\alpha(W_1(\Phi)) = W_2(T\Phi) \ \forall \ \Phi \in S_1 \tag{A.1}$$

(and extending by linearity and continuity). If a one-parameter subgroup $(\mathscr{T}_\tau)_{\tau\in\mathbb{R}}$ of linear symplectomorphisms of (S, σ) is available, then, from the 'linear dynamical system' $(S, \sigma, \mathscr{T}_\tau)$, one obtains, via Weyl algebra quantization, the 'C^* dynamical system' $(\mathscr{A}, \alpha_\tau)$ where $\mathscr{A} = \mathscr{W}(S,\sigma)$ and $(\alpha_\tau)_{\tau\in\mathbb{R}}$ is the one-parameter group of *-automorphisms of $\mathscr{A}$ induced from $(\mathscr{T}_\tau)_{\tau\in\mathbb{R}}$ in the manner described by Equation (A.1).

We recall that a *state* on the Weyl algebra $\mathscr{A}$ is a positive linear functional ω such that $\omega(\mathbb{1}) = 1$. It is called *pure* if it cannot be expressed as a convex combination of any other two states, and *mixed* otherwise. Finally, ω is said to be *stationary* or *invariant* with respect to a one-parameter group $(\alpha_\tau)_{\tau\in\mathbb{R}}$ of *-automorphisms of $\mathscr{A}$ if, for all $\tau \in \mathbb{R}$, $\omega \circ \alpha_\tau = \omega$.

Correlation functions can be defined for sufficiently regular states; that is, one may define the one- and two-point functions

$$\lambda_1(\Phi) = \frac{\mathrm{d}}{\mathrm{d}t}\omega[W(t\Phi)]\bigg|_{t=0} \tag{A.2}$$

$$\lambda_2(\Phi_1,\Phi_2) = -\frac{\partial^2}{\partial s\partial t}\omega[W(s\Phi_1 + t\Phi_2)]e^{-ist\sigma(\Phi_1,\Phi_2)/2}\bigg|_{s,t=0}, \tag{A.3}$$

and similarly define higher n-point correlation functions λ_n, if the state is regular enough for the relevant derivatives to exist. Note that all correlation functions are multilinear in their arguments.

Two-point functions play a special role in quantum field theory. For now, note that if a state is C^2 (see e.g. [Kay93] for a definition), so that the one- and two-point functions exist, one may verify that λ_2 automatically satisfies the following properties for all $\Phi_1, \Phi_2 \in S$:

(i) $\mathrm{Im}[\lambda_2(\Phi_1,\Phi_2)] = \sigma(\Phi_1,\Phi_2)/2$;

(ii) $\mathrm{Re}\lambda_2 =: \mu$ is a symmetric, real-bilinear form on S satisfying

$$\mu(\Phi_1,\Phi_1) \geq 0, \qquad \sigma(\Phi_1,\Phi_2)^2 \leq 4\mu(\Phi_1,\Phi_1)\mu(\Phi_2,\Phi_2). \tag{A.4}$$

Condition (i) encodes the canonical commutation relations, and Condition (ii) results from positivity of the state.

The set of $\lambda_2 : S \times S \to \mathbb{C}$ satisfying Conditions (i) and (ii) is in one-to-one correspondence with the set of equivalence classes of *one-particle structures* over (S,σ), whose definition appeared already in Section 3.2, but which we repeat here for convenience.

Definition A.0.1 (One-particle structures)**.** These are pairs $(K, \mathscr{H})$, with $\mathscr{H}$ a complex Hilbert space and $K : S \to \mathscr{H}$ a *real-linear* map, such that for all $\Phi_1, \Phi_2 \in S$,

1. $KS + iKS$ is dense in $\mathscr{H}$;

2. $\operatorname{Im} \langle K\Phi_1 \mid K\Phi_2 \rangle_{\mathscr{H}} = \sigma(\Phi_1, \Phi_2)/2$.

Any two such pairs $(K, \mathscr{H})$ and $(K', \mathscr{H}')$ are said to be *equivalent* if there exists an isomorphism $U \colon \mathscr{H} \to \mathscr{H}'$ of Hilbert spaces such that $UK = K'$.

The correspondence works as follows. On the one hand, any one-particle structure $(K, \mathscr{H})$ over (S, σ) clearly yields a λ_2 satisfying Conditions (i) and (ii), namely $\lambda_2(\Phi_1, \Phi_2) = \langle K\Phi_1 \mid K\Phi_2 \rangle_{\mathscr{H}}$. Somewhat less trivially, the converse also holds.

Proposition A.0.2. *Given a $\lambda_2 : S \times S \to \mathbb{C}$ satisfying Conditions (i) and (ii), there exists a one-particle structure $(K, \mathscr{H})$ which is associated to λ_2 in the sense that $\langle K\Phi_1 \mid K\Phi_2 \rangle_{\mathscr{H}} = \lambda_2(\Phi_1, \Phi_2)$ for all $\Phi_1, \Phi_2 \in S$. Furthermore, any two such one-particle structures are equivalent in the sense of Definition A.0.1.*

The result above is proved in Appendix A of [KW91]. There, and in the discussion following Proposition 3.1 in Section 3.2, it was also pointed out that one may use this result to prove that, for any $\lambda_2 : S \times S \to \mathbb{C}$ satisfying Conditions (i) and (ii) above, the prescription

$$\omega[W(\Phi)] = \exp[-\lambda_2(\Phi, \Phi)/2] \quad \forall\ \Phi \in S \tag{A.5}$$

(and extension by linearity and continuity) defines a state on $\mathscr{A}$. Indeed, one may realize the right-hand side of Equation (A.5) as the expectation value in the Fock space vacuum, of the operator $W^{\mathscr{F}}(K\Phi) = \exp[\overline{a^\dagger(K\Phi) - (a^\dagger(K\Phi))^*}]$ on the Fock space over $\mathscr{H}$. Since $W(\Phi) \mapsto W^{\mathscr{F}}(K\Phi)$ defines a *-representation of the Weyl algebra, the result follows. One may then easily verify that ω has a two-point function and that this equals λ_2. Indeed, ω also has the following additional properties: (a) it is *analytic* (see e.g. [BR97], p. 38) so that, in particular, it is C^m for all m and all correlation functions exist; (b) the one-point function vanishes; (c) the 'truncated' n-point functions (see e.g. [Haa96, BR97]) vanish for $n > 2$ (in particular, all odd correlation functions vanish). Throughout the present thesis, and just as in [KW91], we will refer to states having Properties (a)–(c) as 'quasi-free', but remark that more properly they should be referred to as 'quasi-free states with vanishing one-point function'. Since analytic states with the same collections of n-point functions are identical, this also proves that any quasi-free state on the Weyl algebra is in the form of Equation (A.5), for some λ_2 satisfying Conditions (i) and (ii).

We recall a technical lemma of Hilbert space analysis which is used in the proof of the statement of uniqueness up to equivalence in Proposition A.0.2, and which we also invoke in the proof of our no-go Theorem 3.4.7.

Lemma A.0.3. *Let $\mathscr{H}_1$, $\mathscr{H}_2$ be complex Hilbert spaces and let $M_1 \subseteq \mathscr{H}_1$, $M_2 \subseteq \mathscr{H}_2$ be real-linear subspaces such that $M_1 + iM_1$ is dense in $\mathscr{H}_1$ and $M_2 + iM_2$ is dense in $\mathscr{H}_2$.*

Let $m : M_1 \to M_2$ *be a bijective real-linear map such that* $\langle x \mid y \rangle_{\mathscr{H}_1} = \langle mx \mid my \rangle_{\mathscr{H}_2}$. *Then* m *extends uniquely to a complex-linear isomorphism from* $\mathscr{H}_1$ *to* $\mathscr{H}_2$. □

So one concludes that quasi-free states over the Weyl algebra $\mathscr{A}$ are also in one-to-one correspondence with equivalence classes of one-particle structures over (S, σ), and thus we can freely speak of the (equivalence class of) one-particle structure(s) 'associated with' a given quasi-free state. What's more, a number of important properties which could be possessed by a quasi-free state have a 'translation' at the level of the corresponding one-particle structure(s). These 'one-particle versions' are often technically convenient to work with, and indeed are what allowed us to conjecture/prove the results in Chapter 3. We record below two such translations (for proofs, see Appendix A of [KW91] and [Kay85]), which are invoked in Section 3.2.

Proposition A.0.4. *A state* ω *is pure if and only if its associated one-particle structure* $(K_\omega, \mathscr{H}_\omega)$ *is such that* $K_\omega S$ *alone is dense in* $\mathscr{H}_\omega$.

Proposition A.0.5. *Let* $\tilde{\mathscr{A}}$ *denote the Weyl algebra over the symplectic vector space* $(\tilde{S}, \tilde{\sigma})$ *and* ω *be a state on* $\tilde{\mathscr{A}}$ *with associated one-particle structure* $(K_\omega, \mathscr{H}_\omega)$. *Then the* C^*-*subalgebra* $\tilde{\mathscr{A}}_R$ *of* $\tilde{\mathscr{A}}$ *generated by the subspace* R *of* $\tilde{S}$ *has the Reeh-Schlieder property*[1] *for* $(\tilde{\mathscr{A}}, \omega)$ *iff* $K_\omega R + iK_\omega R$ *is dense in* $\mathscr{H}_\omega$.

[1] Let ω be a state on a C^*-algebra $\mathscr{A}$ with GNS-triple [BR87] (ρ, H, Ω). Then the C^*-subalgebra $\mathscr{B}$ of $\mathscr{A}$ is said to have the *Reeh-Schlieder property* for $(\mathscr{A}, \omega)$ if $\rho(\mathscr{B})\Omega$ is dense in H.

Appendix B

More on the infrared pathology of massless fields in 1+1 dimensions

The purpose of this appendix is to show how the infrared pathology afflicting massless fields in (1+1)-dimensional Minkowski spacetime, which was discussed in detail in Section 3.4.2, manifests itself when considering the interplay between the global Hadamard condition for such fields and invariance under Lorentz boosts. Namely, we will prove the precise result stated below (where we recall that the notation $C_0^\infty(M)$, for M an arbitrary manifold, refers for us to the space of *real-valued* test functions on M). It is worth noting that our proof of this result will use several techniques pioneered in [KW91] – and indeed that the result itself may be morally regarded as complementing the list of non-existence theorems proved in that paper (but recall that the analysis [KW91] was restricted to models in four dimensions).

Theorem B.0.1. *Let $(S_{\mathbb{M}}, \sigma_{\mathbb{M}})$ be the real symplectic vector space of spatially compact solutions to the massless wave equation $\Box\phi = 0$ on $\mathbb{R}^2$. Let $\mathscr{A}_{\mathbb{M}} = \mathscr{W}(S_{\mathbb{M}}, \sigma_{\mathbb{M}})$ be the associated Weyl algebra of canonical commutation relations. Then there is no C^2 state ω on $\mathscr{A}_{\mathbb{M}}$ which is invariant under the automorphisms of $\mathscr{A}_{\mathbb{M}}$ induced by the one-parameter group of Lorentz boost isometries, and whose spacetime smeared two-point function $\Lambda : C_0^\infty(\mathbb{R}^2) \times C_0^\infty(\mathbb{R}^2) \to \mathbb{C}$ has the 'global Hadamard form' in the sense that*

$$H = \Lambda - \Lambda_{\mathbb{M}} \tag{B.1}$$

is representable by a function $h \in C^\infty(\mathbb{R}^2 \times \mathbb{R}^2)$.

Proof. Let us collect some easy facts about the bilinear functional $H : C_0^\infty(\mathbb{R}^2) \times C_0^\infty(\mathbb{R}^2) \to \mathbb{C}$ defined by Equation (B.1). Aside from the representability by a smooth function h, i.e.

$$H(F,G) = \int_{\mathbb{R}^2} h(x,y)F(x)G(y)\,\mathrm{d}^2x\,\mathrm{d}^2y \quad \forall\, F,G \in C_0^\infty(\mathbb{R}^2),$$

we have that

(a) $H(\mathscr{T}_\tau F, \mathscr{T}_\tau G) = H(F,G)\ \forall\, F,G \in C_0^\infty(\mathbb{R}^2)$, where $(\mathscr{T}_\tau)_{\tau\in\mathbb{R}}$ is the one-parameter group of linear transformations on $C_0^\infty(\mathbb{R}^2)$ given by pullback by (the inverse of) the Lorentz boost isometries;

(b) h is a classical bisolution to the wave equation, i.e. $[\Box_x h](x,y) = [\Box_y h](x,y) = 0\ \forall\, x,y \in \mathbb{R}^2$;

(c) since Λ and $\Lambda_{\mathbb{M}}$ individually satisfy the canonical commutation relations in the sense that

$$2\,\mathrm{Im}\,\Lambda_{(\mathbb{M})}(F,G) = \frac{1}{i}\left[\Lambda_{(\mathbb{M})}(F,G) - \Lambda_{(\mathbb{M})}(G,F)\right] = E_{\mathbb{M}}(F,G) \quad \forall\, F,G \in C_0^\infty(\mathbb{R}^2),$$

it follows that

$$\begin{aligned} H(F,G) - H(G,F) &= [\Lambda(F,G) - \Lambda_{\mathbb{M}}(F,G)] - [\Lambda(G,F) - \Lambda_{\mathbb{M}}(G,F)] \\ &= [\Lambda(F,G) - \Lambda(G,F)] - [\Lambda_{\mathbb{M}}(F,G) - \Lambda_{\mathbb{M}}(G,F)] \\ &= 0 \quad \forall\, F,G \in C_0^\infty(\mathbb{R}^2), \end{aligned}$$

i.e. $h(x,y) = h(y,x)\ \forall\, x,y \in \mathbb{R}^2$.

Without loss of generality, we may assume from now on that the state ω is quasi-free, since the existence of a merely C^2 state on $\mathscr{A}_{\mathbb{M}}$ which is Lorentz-boost–invariant and whose two-point function satisfies the hypotheses of the Theorem immediately implies the existence of a quasi-free one with the same properties – i.e. the unique quasi-free state (with vanishing one-point function) whose two-point function is the same as the two-point function of the original state.

In a first step, we argue that ω must coincide with the 'usual' vacuum state $\omega_{\mathbb{M}}$ on the Weyl subalgebra $\mathscr{A}_{0,\mathbb{M}}$ of $\mathscr{A}_{\mathbb{M}}$ generated by the subspace $S_{0,\mathbb{M}} = S_{\text{r-mov}} + S_{\text{l-mov}}$ consisting of sums of compactly supported left movers and compactly supported right movers. This fact is proved by the following chain of reasoning: (1) by methods completely analogous (and actually considerably simpler thanks to the absence of boundaries) to the methods used by us in the proof of Theorem 3.4.7 (which are in turn based on methods presented in Appendix B of [KW91]), one sees that the 'symplectically smeared' two-point function $\lambda : S_{\mathbb{M}} \times S_{\mathbb{M}} \to \mathbb{C}$ of ω, restricted to either $S_{\text{r-mov}} \times S_{\text{r-mov}}$ or $S_{\text{l-mov}} \times S_{\text{l-mov}}$, equals the restriction to the same sets of the 'symplectically smeared' two-point function $\lambda_{\mathbb{M}}$ of the state $\omega_{\mathbb{M}}$; (b) this implies in particular that the restriction of ω to the Weyl subalgebras $\mathscr{A}_{\text{r-mov}}$ and $\mathscr{A}_{\text{l-mov}}$ generated by $S_{\text{r-mov}}$ and $S_{\text{l-mov}}$ (respectively) equals the restriction to the same subalgebras of the 'usual' vacuum state $\omega_{\mathbb{M}}$; (c) since these restrictions are pure and quasifree, by Lemma 4.1 in [KW91] it follows that the $\omega\restriction_{\mathscr{A}_{0,\mathbb{M}}}$ is pure and uniquely determined; (d) finally, since $\omega_{\mathbb{M}}$ is pure and quasi-free, this implies the sought-after result that $\omega\restriction_{\mathscr{A}_{0,\mathbb{M}}} = \omega_{\mathbb{M}}$.

Since $S_{0,\mathbb{M}} = E_{\mathbb{M}}[C^\infty_{00}(\mathbb{R}^2)]$, the result in the above paragraph implies that

$$H(F, G) = 0 \quad \forall\; F,\, G \in C^\infty_{00}(\mathbb{R}^2) \tag{B.2}$$

(we recall that $C^\infty_{00}(\mathbb{R}^2) = \big\{\psi \in C^\infty_0(\mathbb{R}^2) \;\big|\; \int_{\mathbb{R}^2} \psi(x)\,\mathrm{d}^2x = 0\big\}$). So we now pick a $\chi \in C^\infty_0(\mathbb{R}^2)$ with $\int_{\mathbb{R}^2} \chi(x)\,\mathrm{d}^2x = 1$. Then for any $F \in C^\infty_0(\mathbb{R}^2)$ one has $F - (\int F)\chi \in C^\infty_{00}(\mathbb{R}^2)$ (where we have abbreviated $\int_{\mathbb{R}^2} F(x)\,\mathrm{d}^2x$ with $\int F$ since no confusion should arise from doing this), and therefore

$$\begin{aligned} 0 &= H\left(F - \left(\int F\right)\chi, G - \left(\int G\right)\chi\right) \\ &= H(F, G) - \left(\int F\right) H(\chi, G) - \left(\int G\right) H(F, \chi) + \left(\int F\right)\left(\int G\right) H(\chi, \chi) \end{aligned} \tag{B.3}$$

for all $F,\, G \in C^\infty_0(\mathbb{R}^2)$. Now consider the last term in Equation (B.3) as a bilinear map $C^\infty_0(\mathbb{R}^2) \times C^\infty_0(\mathbb{R}^2) \to \mathbb{C}$: this is simply the bidistribution defined by the constant kernel $c_1 = H(\chi, \chi)$, which is obviously a bisolution of the wave equation and is boost-invariant since Lorentz boosts are measure-preserving. Let us define $\tau \in C^\infty(\mathbb{R}^2; \mathbb{C})$ by $\tau(y) = \int_{\mathbb{R}^2} h(x, y)\chi(x)\,\mathrm{d}^2x$, so that τ is the smooth kernel of the distribution $\Psi \mapsto H(\chi, \Psi)$. Clearly τ is a solution to the wave equation. By the symmetry of h together with Equation (B.3) we have

$$h(x, y) + c_1 = \tau(x) + \tau(y) \quad \forall\; x,\, y \in \mathbb{R}^2. \tag{B.4}$$

Since the left-hand-side of Equation (B.4) is boost-invariant, so must be the right-hand side (seen as a function of both x and y). But then, by setting $x = y$, we see that this implies that τ is boost-invariant. Since it is also a solution of the wave equation (and is smooth), it must be equal to a constant c_2. Thus we conclude that

$$h(x, y) = 2c_2 - c_1 =: c, \tag{B.5}$$

i.e. h equals a complex constant. This in turn implies that $\Lambda = \Lambda_{\mathbb{M}} + c$. Since, as proved in [Wig67, p. 204], for no choice of constant c can this quantity satisfy the positivity properties of the (spacetime smeared) two-point function of a state, we reach a contradiction. This completes the proof. □

Appendix C

Sobolev Spaces on Manifolds

The purpose of this appendix is to introduce Sobolev spaces on smooth manifolds, and in particular spaces of locally Sobolev sections of vector bundles. We follow [Hör90a, Sec. 7.9], [Hör94, App. B], [Rin09, Ch. 5] and [BW15].

We begin by considering the case of the Euclidean space $\mathbb{R}^n$ and scalar functions/distributions. Throughout, $\mathcal{S} = \mathcal{S}(\mathbb{R}^n)$ denotes the space of complex-valued Schwartz functions on $\mathbb{R}^n$, and $\mathcal{S}' = \mathcal{S}'(\mathbb{R}^n)$ denotes its continuous dual, the space of Schwartz distributions. Recall that a Lebesgue measurable, complex-valued function u is said to be *locally* L^1, or to belong to $L^1_{\mathrm{loc}}(\mathbb{R}^n)$, if $u \cdot \chi_K \in L^1(\mathbb{R}^n)$ for any compact set $K \subset \mathbb{R}^n$. The space $L^1_{\mathrm{loc}}(\Omega)$, of locally L^1 functions on an open set $\Omega \subseteq \mathbb{R}^n$, is defined analogously. Locally L^1 functions with values in $\mathbb{R}$ and in any other finite-dimensional real or complex vector space are defined by demanding that all components in a basis (and therefore in any other basis) be locally L^1.

Definition C.0.1. Let $k \in \mathbb{N}_0$ and $\Omega \subseteq \mathbb{R}^n$ be open. Then a function $u \in L^1_{\mathrm{loc}}(\Omega)$ is said to be k *times weakly differentiable* if for every multi-index α with $|\alpha| \leq k$ there is a locally L^1 function u_α on Ω such that the following equation holds for any $\psi \in C_0^\infty(\Omega)$:

$$\int_\Omega u \partial^\alpha \psi \, \mathrm{d}x = (-1)^{|\alpha|} \int_\Omega u_\alpha \psi \, \mathrm{d}x.$$

If that is the case, the u_α are referred to as the *weak derivatives* of u. A locally integrable vector-valued function is said to be k times weakly differentiable if its components are.

Definition C.0.2. Let $k \in \mathbb{N}_0$, $1 \leq p < \infty$, $\Omega \subseteq \mathbb{R}^n$ be open, and V be a finite-dimensional real or complex vector space with norm and inner product denoted by $|\cdot|$ and $(\cdot, \cdot)$ respectively. Then $\mathcal{W}^{k,p}(\Omega; V)$ denotes the set of k times weakly differentiable functions (valued in V) such that all the weak derivatives are also in $L^p(\Omega; V)$. Let now

$$W^{k,p}(\Omega; V) \coloneqq \mathcal{W}^{k,p}(\Omega; V)/\sim \quad \text{where} \quad u_1 \sim u_2 \Leftrightarrow u_1 = u_2 \text{ a.e.}$$

(with respect to the Lebesgue measure). Then $W^{k,p}(\Omega;V)$ is a Banach space with the following (well-defined) norm

$$\|[u]\|_{W^{k,p}} := \left[\sum_{|\alpha|\le k}\int_\Omega |\partial^\alpha u|^p \,\mathrm{d}x\right]^{1/p} \qquad \forall\ u \in \mathcal{W}^{k,p}(\Omega;V).$$

The special case $p = 2$ yields a Hilbert space denoted by $H^k(\Omega;V) = \mathcal{W}^{k,2}(\Omega;V)$, with inner product

$$\langle [u_1] \mid [u_2]\rangle_{H^k} := \sum_{|\alpha|\le k}\int_\Omega (\partial^\alpha u_1, \partial^\alpha u_2)\,\mathrm{d}x \quad \forall\ u_1,u_2 \in \mathcal{W}^{k,2}(\Omega;V).$$

Clearly, $W^{0,p}(\Omega;V) = L^p(\Omega;V)$ so that in particular $H^0(\Omega;V) = L^2(\Omega;V)$. The special cases with $V = \mathbb{C}$ are denoted simply by $W^{k,p}(\Omega)$ and $H^k(\Omega)$.

Definition C.0.3. Let $s \in \mathbb{R}$. Then $H_{(s)}(\mathbb{R}^n)$ denotes the space of all $u \in \mathcal{S}'$ such that the Fourier transform $\hat{u} \in \mathcal{S}'$ is (representable by) a measurable function and $\hat{u}(\xi)(1+\|\xi\|^2)^{s/2}$ is square-integrable. $H_{(s)}(\mathbb{R}^n)$ is a Hilbert space with the inner product

$$\langle u_1 \mid u_2\rangle_{(s)} := \frac{1}{(2\pi)^n}\int \overline{\hat{u}_1(\xi)}\hat{u}_2(\xi)\left(1+\|\xi\|^2\right)^s \mathrm{d}\xi$$

and corresponding norm denoted by $\|\cdot\|_{(s)}$. We can define $H_{(s)}(\mathbb{R}^n;V)$ for any finite-dimensional real or complex vector space V by demanding that components be in $H_{(s)}(\mathbb{R}^n)$.

Proposition C.0.4. *Under the standard embedding $L^1_{\mathrm{loc}}(\mathbb{R}^n) \hookrightarrow \mathcal{S}'$, we have $H_{(k)}(\mathbb{R}^n) = H^k(\mathbb{R}^n)$ for all $k \in \mathbb{N}_0$. The norms on $H_{(k)}(\mathbb{R}^n)$ and on $H^k(\mathbb{R}^n)$ are equivalent.* □

Recall that, for any open subset $\Omega \subseteq \mathbb{R}^n$, the space $\mathscr{D}(\Omega)$ is the locally convex topological vector space of complex-valued functions which are defined, smooth, and with compact support on Ω. In other words, $\mathscr{D}(\Omega) = C_0^\infty(\Omega;\mathbb{C})$ as a set; its standard inductive limit (more precisely, *LF-space*) topology is defined e.g. in [Tre67, Ch. 13], and can be equivalently characterised by using the seminorms in Theorem 2.1.5 in [Hör90a]. Its continuous dual space is the space of distributions $\mathscr{D}'(\Omega)$. We now define local versions of the spaces in Definition C.0.3 (local versions of the spaces in Definition C.0.2 may also be defined, but we will not do so here). This will allow us to further extend the definition to arbitrary smooth manifolds.

Definition C.0.5. Let $s \in \mathbb{R}$ and $\Omega \subseteq \mathbb{R}^n$ be open. Then $H^{\mathrm{loc}}_{(s)}(\Omega)$ denotes the space of all $u \in \mathscr{D}'(\Omega)$ such that $\psi u \in H_{(s)}(\mathbb{R}^n)$ for every $\psi \in C_0^\infty(\Omega)$. $H^{\mathrm{loc}}_{(s)}(\Omega;V)$ can similarly be defined when V is a real or complex vector space.

If M is a smooth manifold, while there is only one possible definition of the space of (scalar) test functions, i.e. $\mathscr{D}(M) := C_0^\infty(M;\mathbb{C})$, there are at least two natural ways in which one might wish to define a space of distributions $\mathscr{D}'(M)$. A choice between the two is determined to a large extent by whether what is cherished is the presence of a canonically defined embedding from functions on M [say, from $C^0(M;\mathbb{C})$] to $\mathscr{D}'(M)$, or the convenience of defining spaces directly by duality. Namely, in the first approach (used e.g. by Hörmander), one defines $\mathscr{D}'(M)$ as

$$\mathscr{D}'(M) = [\Gamma_0^\infty(|\Lambda M| \otimes_{\mathbb{R}} C)]',$$

where C denotes the bundle $M \times \mathbb{C} \to M$ and $|\Lambda M|$ denotes the bundle of densities on M, so that $|\Lambda M| \otimes_{\mathbb{R}} C$ may be viewed as the bundle of *complex-valued* densities on M. A function in $C^0(M;\mathbb{C})$ [indeed, more generally one in $L^1_{\text{loc}}(M)$] gives rise to a linear form on $\Gamma_0^\infty(|\Lambda M| \otimes_{\mathbb{R}} C)$ defined by

$$\lambda \mapsto \int_M f\lambda \in \mathbb{C},$$

since $f\lambda$ is then a continuous [or locally L^1] complex-valued density on M with compact support, which can be integrated.

In the second approach, one directly defines

$$\mathscr{D}'(M) = [\mathscr{D}(M)]',$$

i.e. as the continuous dual of $\mathscr{D}(M)$. If a smooth density $\mathrm{d}\mu \in \Gamma^\infty(|\Lambda M|)$ is given, then functions on M can still be linearly embedded in the thus defined space of distributions, since the assignment

$$\mathscr{D}(M) \ni \psi \mapsto \int_M f\psi \,\mathrm{d}\mu \in \mathbb{C}$$

belongs to $[\mathscr{D}(M)]'$.

In other words: every $\mathrm{d}\mu \in \Gamma^\infty(|\Lambda M|)$ defines a linear isomorphism (indeed, a topological vector space isomorphism) between the two versions of $\mathscr{D}'(M)$ we have just defined. We adopt the second approach in most of this thesis, since we will always be dealing with smooth manifolds with a semi-Riemannian metric g and the latter yields a smooth volume density $\mathrm{d}\mu_g$. However, for the purposes of this Appendix, and in particular of the following definition, it is in fact more convenient to temporarily use the first definition of $\mathscr{D}'(M)$ we gave here.

Definition C.0.6 (Spaces of locally Sobolev functions on manifolds)**.** Let M be a smooth n-dimensional manifold and $s \in \mathbb{R}$. Then $H^{\text{loc}}_{(s)}(M)$, denoted also by $H^s_{\text{loc}}(M)$, is defined to be the vector space of all $u \in \mathscr{D}'(M)$ such that, for every smooth chart $(U, \kappa : U \to V)$

for M,

$$u_\kappa := (\kappa^{-1})^* u \in H^{\mathrm{loc}}_{(s)}(V)$$

where $(\kappa^{-1})^* u \in \mathscr{D}'(V)$ is defined by

$$[(\kappa^{-1})^* u](\varphi) = u[(\kappa^{-1})_*(\varphi \, \mathrm{d}\lambda)] \quad \forall \, \varphi \in \mathscr{D}(V).$$

In turn, $\mathrm{d}\lambda$ above denotes the restriction of the standard smooth Lebesgue density on $\mathbb{R}^n$ to V, and the pushforward $(\kappa^{-1})_*(\mathrm{d}\tilde{\lambda})$ of a compactly supported smooth density $\mathrm{d}\tilde{\lambda}$ on V is the compactly supported smooth density on M defined by

$$(\kappa^{-1})_*(\mathrm{d}\tilde{\lambda}) = \begin{cases} \kappa^*(\mathrm{d}\tilde{\lambda}) & \text{on } \kappa^{-1}(\operatorname{supp} \mathrm{d}\tilde{\lambda}) \\ 0 & \text{otherwise.} \end{cases}$$

Definition C.0.7 (Topology on $H^s_{\mathrm{loc}}(M)$)**.** $H^s_{\mathrm{loc}}(M)$ is equipped with the locally convex topology defined by the semi-norms

$$\|u\|_{\kappa,\varphi} := \|\varphi u_\kappa\|_{(s)}$$

where $\kappa : U \to V$ is an arbitrary smooth coordinate system on M, and φ is an arbitrary element of $C_0^\infty(V)$. In fact [Hör94, p. 475], the same topology is generated by any *countable* subfamily $\{\|\cdot\|_{\kappa_n,\varphi_n}\}_{n\in\mathbb{N}}$ with the property that $\{\kappa_n : U_n \to V_n\}$ is an atlas for M, $\varphi_n \in C_0^\infty(V_n)$, and $\bigcup_{n\in\mathbb{N}} \{x \in V_n \mid \varphi_n(x) \neq 0\} = M$. Therefore, $H^s_{\mathrm{loc}}(M)$ is a Fréchet space.

Recall that $\mathscr{E}'(M)$ denotes the subspace of $\mathscr{D}'(M)$ consisting of distributions with compact support, while for a closed subset K of M, $\mathscr{E}'(K) := \{u \in \mathscr{D}'(M) \mid \operatorname{supp} u \subseteq K\}$.

Definition C.0.8 (Spaces of locally Sobolev functions with fixed or arbitrary compact support)**.** Let M be a smooth manifold and $H^s_{\mathrm{loc}}(M)$ be defined as above. Then, for any compact subset K of M, we define the space of locally Sobolev functions supported on K, $H^s_K(M)$, by

$$H^s_K(M) := H^s_{\mathrm{loc}}(M) \cap \mathscr{E}'(K).$$

Finally, we define the space of locally Sobolev functions with compact support, $H^s_c(M)$, by

$$H^s_c(M) := H^s_{\mathrm{loc}}(M) \cap \mathscr{E}'(M).$$

Both spaces are equipped with the subspace topology inherited from $H^s_{\mathrm{loc}}(M)$.

In giving definitions C.0.6, C.0.7 and C.0.8, we followed the route given in [Hör90a, Sec. 7.9] and [Hör94, App. B]. Bär and Wafo [BW15] take a different but equivalent path to those definitions (generalised to sections of arbitrary vector bundles). Their approach has the virtue of coordinate independence, but is perhaps less elementary. Since their work plays an important role in Chapter 4, let us make some contact with their viewpoint these definitions, when specialised to the scalar case. Essentially, we first gave the definition of the largest of the spaces above, i.e. of $H^s_{\mathrm{loc}}(M)$, in a concrete way by using the existence of the very large space of distributions on M; then, we regarded the other spaces as subspaces equipped with the relative topology. On the other hand, Bär and Wafo begin by defining the *smallest* space, $H^s_K(M)$ for a compact K, as a Banach space by an abstract metric space completion procedure from $C^\infty_K(M;\mathbb{C}) = \{\phi \in C^\infty(M;\mathbb{C}) \mid \operatorname{supp}\phi \subseteq K\}$. Then, $H^s_c(M)$ is defined as the direct limit

$$H^s_c(M) \coloneqq \bigcup_{\substack{K \subseteq M \\ K \text{ compact}}} H^s_K(M)$$

of the direct system given by $\{H^s_K(M)\}_K$ and the inclusion maps, equipped with the strict inductive limit topology. Finally, they notice that the thus abstractly defined spaces $H^s_K(M)$ and $H^s_c(M)$ embed into $\mathscr{D}'(M)$. This allows them to define $H^s_{\mathrm{loc}}(M)$ by

$$H^s_{\mathrm{loc}}(M) \coloneqq \left\{ u \in \mathscr{D}'(M) \;\middle|\; \chi u \in H^s_c(M)\ \forall\ \chi \in C^\infty_0(M) \right\}$$

and equip it with a Fréchet space topology which is given by a countable family of seminorms defined in terms of the Banach space norms on $H^s_K(M)$ mentioned above.

Bibliography

[AB02] R. J. Alonso-Blanco, *On the Green–Vinogradov Formula*, Acta Appl. Math. **72**(1), 19–32 (2002).

[AF87] B. Allen and A. Folacci, *Massless minimally coupled scalar field in de Sitter space*, Phys. Rev. D **35**, 3771–3778 (1987).

[All85] B. Allen, *Vacuum states in de Sitter space*, Phys. Rev. D **32**, 3136–3149 (1985).

[Alt16] H. W. Alt, *Linear Functional Analysis: An Application-Oriented Introduction*, Universitext, Springer-Verlag, 2016.

[Bär15] C. Bär, *Green-Hyperbolic Operators on Globally Hyperbolic Spacetimes*, Comm. Math. Phys. **333**(3), 1585–1615 (2015), 1310.0738.

[BCFT07] H. Broer, R. Cushman, F. Fassò and F. Takens, *Geometry of KAM tori for nearly integrable Hamiltonian systems*, Erg. Th. Dyn. Syst. **27**(3), 725–741 (2007).

[BD78] T. S. Bunch and P. C. W. Davies, *Quantum field theory in de Sitter space: renormalization by point-splitting*, Proc. R. Soc. London A **360**(1700), 117–134 (1978).

[BD84] N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 1984.

[BEE96] J. K. Beem, P. Ehrlich and K. Easley, *Global Lorentzian Geometry, Second Edition*, volume 67 of *Pure and Applied Mathematics*, Marcel Dekker, New York, second edition, 1996.

[Bek73] J. D. Bekenstein, *Black Holes and Entropy*, Phys. Rev. D **7**(8), 2333–2346 (1973).

[BFV03] R. Brunetti, K. Fredenhagen and R. Verch, *The Generally Covariant Locality Principle – A New Paradigm for Local Quantum Field Theory*, Comm. Math. Phys. **237**(1-2), 31–68 (2003), math-ph/0112041.

[BG12] C. Bär and N. Ginoux, Classical and Quantum Fields on Lorentzian Manifolds, in *Global Differential Geometry*, edited by C. Bär, J. Lohkamp and M. Schwarz, volume 17 of *Springer Proceedings in Mathematics*, pages 359–400, Springer-Verlag, 2012, 1104.1158.

[BGP07] C. Bär, N. Ginoux and F. Pfäffle, *Wave Equations on Lorentzian Manifolds and Quantization*, ESI Lectures in Mathematics and Physics, European Mathematical Society, 2007, math/0806.1036.

[Bor95] E. Borel, *Sur quelques points de la théorie des fonctions*, Ann. Sci. École Norm. Sup. **12**(3), 9–55 (1895).

[Boy69] R. H. Boyer, *Geodesic Killing Orbits and Bifurcate Killing Horizons*, Proc. R. Soc. Lond. A **311**(1505), 245–252 (1969).

[BR87] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics 1*, Theoretical and Mathematical Physics, Springer-Verlag, second edition, 1987.

[BR97] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics 2*, Theoretical and Mathematical Physics, Springer-Verlag, second edition, 1997.

[BS05] A. N. Bernal and M. Sánchez, *Smoothness of Time Functions and the Metric Splitting of Globally Hyperbolic Spacetimes*, Comm. Math. Phys. **257**(1), 43–50 (2005), gr-qc/0512095.

[BS07] A. N. Bernal and M. Sánchez, *Globally hyperbolic spacetimes can be defined as 'causal' instead of 'strongly causal'*, Class. Quant. Grav. **24**(3), 745 (2007), gr-qc/0611138.

[BW15] C. Bär and R. T. Wafo, *Initial Value Problems for Wave Equations on Manifolds*, Math. Phys. Anal. Geom. **18**(1, Article 7), 1–29 (2015), 1408.4995.

[Cas48] H. B. G. Casimir, *On the Attraction Between Two Perfectly Conducting Plates*, Proc. Kon. Ned. Akad. Wetenschap **B51**, 793–795 (1948).

[CB67] Y. Choquet-Bruhat, Hyperbolic partial differential equations on a manifold, in *Battelle Rencontres: 1967 Lectures in Mathematics and Physics*, edited by C. DeWitt-Morette and J. A. Wheeler, pages 84–106, Benjamin, 1967.

[CDGH01] P. T. Chruściel, E. Delay, G. J. Galloway and R. Howard, *Regularity of horizons and the Area Theorem*, Ann. Inst. H. Poincaré **2**(1), 109–178 (2001), gr-qc/0001003.

[CFS08] A. M. Candela, J. L. Flores and M. Sánchez, *Global hyperbolicity and Palais–Smale condition for action functionals in stationary spacetimes*, Adv. Math. **218**(2), 515–536 (2008), math/0610175.

[CGM16] P. T. Chruściel, J. D. E. Grant and E. Minguzzi, *On Differentiability of Volume Time Functions*, Ann. H. Poincaré **17**(10), 2801–2824 (2016), 1301.2909.

[CGS09] P. T. Chruściel, G. J. Galloway and D. A. Solis, *Topological censorship for Kaluza-Klein space-times*, Ann. Inst. H. Poincaré **10**(5), 893–912 (2009), 0808.3233.

[Chr98] P. T. Chruściel, *A remark on differentiability of Cauchy horizons*, Class. Quant. Grav. **15**(12), 3845 (1998).

[Chr11] P. T. Chruściel, *Elements of causality theory*, (2011), 1110.6706.

[Chr15] P. T. Chruściel, *The Geometry of Black Holes*, preprint lecture notes, University of Vienna (2015), available at `http://tinyurl.com/p5omjwv` (retrieved 20 Sept 2016).

[CL55] A. Coddington and N. Levinson, *Theory of Ordinary Differential Equations*, International series in pure and applied mathematics, McGraw-Hill, 1955.

[COS12] P. T. Chruściel, C. R. Ölz and S. J. Szybka, *Space-time diagrammatics*, Phys. Rev. D **86**(12), 124041 (2012), 1211.1718.

[CT68] N. A. Chernikov and E. A. Tagirov, *Quantum theory of scalar field in de Sitter space-time*, Annales de l'I.H.P. Physique théorique **9**(2), 109–141 (1968).

[Dav75] P. C. W. Davies, *Scalar production in Schwarzschild and Rindler metrics*, J. Phys. A **8**(4), 609–616 (1975).

[DF77] P. C. W. Davies and S. A. Fulling, *Radiation from Moving Mirrors and from Black Holes*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **356**(1685), 237–257 (1977).

[DF16] C. Dappiaggi and H. R. C. Ferreira, *Hadamard states for a scalar field in anti-de Sitter spacetime with arbitrary boundary conditions*, (2016), 1610.01049.

[DK86] J. Dimock and B. S. Kay, *Classical and quantum scattering theory for linear scalar fields on the Schwarzschild metric. II*, J. Math. Phys. **27**(10), 2520–2525 (1986).

[DK87] J. Dimock and B. S. Kay, *Classical and quantum scattering theory for linear scalar fields on the Schwarzschild metric I*, Ann. Phys. **175**(2), 366–426 (1987).

[DM06] J. Dereziński and K. A. Meissner, Quantum Massless Field in 1+1 Dimensions, in *Mathematical Physics of Quantum Mechanics*, edited by J. Asch and A. Joye, volume 690 of *Lecture Notes in Physics*, pages 107–127, Springer-Verlag, 2006.

[DR07] M. Dafermos and I. Rodnianski, *The wave equation on Schwarzschild-de Sitter spacetimes*, (2007), 0709.2766.

[DR09] M. Dafermos and I. Rodnianski, *The red-shift effect and radiation decay on black hole spacetimes*, Comm. Pure Applied Math. **62**(7), 859–919 (2009), gr-qc/0512119.

[DRSR14] M. Dafermos, I. Rodnianski and Y. Shlapentokh-Rothman, *A scattering theory for the wave equation on Kerr black hole exteriors*, (2014), 1412.8379.

[DRSR16] M. Dafermos, I. Rodnianski and Y. Shlapentokh-Rothman, *Decay for solutions of the wave equation on Kerr exterior spacetimes III: The full subextremal case* $|a| < M$, Ann. Math. **183**(3), 787–913 (2016), 1402.7034.

[EG15] L. C. Evans and R. F. Gariepy, *Measure Theory and Fine Properties of Functions*, Textbooks in Mathematics, CRC Press, second revised edition, 2015.

[EMM12] G. F. R. Ellis, R. Maartens and M. A. H. MacCallum, *Relativistic Cosmology*, Cambridge University Press, 2012.

[Fat15] A. Fathi, *Time functions revisited*, Int. J. Geom. Methods Mod. Phys. **12**(8), 1560027 (2015).

[FD76] S. A. Fulling and P. C. W. Davies, *Radiation from a Moving Mirror in Two Dimensional Space-Time: Conformal Anomaly*, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences **348**(1654), 393–414 (1976).

[Fed69] H. Federer, *Geometric Measure Theory*, volume 153 of *Grundlehren der mathematischen Wissenschaften*, Springer–Verlag, 1969.

[FOP07] C. J. Fewster, K. D. Olum and M. J. Pfenning, *Averaged null energy condition in spacetimes with boundaries*, Phys. Rev. D **75**(2), 025007 (2007), gr-qc/0609007.

[FR87] S. A. Fulling and S. N. M. Ruijsenaars, *Temperature, periodicity and horizons*, Phys. Rep. **152**(3), 135–176 (1987).

[Fri75] F. G. Friedlander, *The wave equation on a curved space-time*, Cambridge Monographs in Mathematical Physics, Cambridge University Press, 1975.

[FS12] A. Fathi and A. Siconolfi, *On smooth time functions*, Math. Proc. Camb. Phil. Soc. **152**(2), 303–339 (2012).

[FSW78] S. A. Fulling, M. Sweeny and R. M. Wald, *Singularity structure of the two-point function quantum field theory in curved spacetime*, Comm. Math. Phys. **63**(3), 257–264 (1978).

[Ful73] S. A. Fulling, *Nonuniqueness of Canonical Field Quantization in Riemannian Space-Time*, Phys. Rev. D **7**(10), 2850–2862 (1973).

[Gal86] G. J. Galloway, *Curvature, causality and completeness in space-times with causally complete spacelike slices*, Math. Proc. Camb. Phil. Soc. **99**(2), 367–375 (1986).

[Går77] L. Gårding, *Le problème de la dérivée oblique pour l'équation des ondes*, C. R. Acad. Sci. Paris Série A **285**, 773–775 (1977), *Rectification* **285**, p. 1199. (1978).

[Går98] L. Gårding, Hyperbolic equations in the twentieth century, in *Séminaire et Congrès vol. 3: Matériaux pour l'histoire des mathématiques au XXe siècle (Nice, 1996)*, pages 37–68, Société Mathématique de France, 1998, `http://smf4.emath.fr/Publications/SeminairesCongres/1998/3/html/smf_sem-cong_3_37-68.php`.

[Ger68] R. Geroch, *Spinor structure of space-times in General Relativity. I*, J. Math. Phys. **9**(11), 1739–1744 (1968).

[Ger96] R. Geroch, Partial Differential Equations of Physics, in *General Relativity. Proceedings of the Forty-Sixth Scottish Summer School in Physics*, edited by G. S. Hall, J. R. Pulham and P. Osborne, pages 19–60, The Scottish Universities Summer School in Physics, 1996.

[GG73] M. Golubitsky and V. Guillemin, *Stable Mappings and Their Singularities*, Springer-Verlag, 1973.

[GH77] G. W. Gibbons and S. W. Hawking, *Cosmological event horizons, thermodynamics, and particle creation*, Phys. Rev. D **15**(10), 2738–2751 (1977).

[GH93] G. W. Gibbons and S. W. Hawking, *Euclidean Quantum Gravity*, World Scientific, 1993.

[Gin09] N. Ginoux, Linear Wave Equations, in *Quantum Field Theory on Curved Spacetimes: Concepts and Mathematical Foundations*, edited by C. Bär and K. Fredenhagen, volume 786 of *Lecture Notes in Physics*, pages 59–84, Springer-Verlag, 2009.

[GK89] G. Gonnella and B. S. Kay, *Can locally Hadamard quantum states have non-local singularities?*, Class. Quantum Grav. **6**(10), 1445 (1989).

[Gla58] G. Glaeser, *Étude de Quelques Algèbres Tayloriennes*, Journal d'Analyse Mathématique **6**(1), 1–124 (1958).

[GP76] G. W. Gibbons and M. J. Perry, *Black Holes in Thermal Equilibrium*, Phys. Rev. Lett. **36**(16), 985–987 (1976).

[GS86] M. H. Goroff and A. Sagnotti, *The ultraviolet behavior of Einstein gravity*, Nucl. Phys. B **266**(3), 709–736 (1986).

[GV96] L. R. Gindikin and S. Volevich, *Mixed problem for partial differential equations with quasihomogeneous principal part*, American Mathematical Society, 1996.

[GV14] G. J. Galloway and C. Vega, *Achronal Limits, Lorentzian Spheres, and Splitting*, Ann. Inst. H. Poincaré **15**(11), 2241–2279 (2014), 1211.2460v2.

[Haa96] R. Haag, *Local Quantum Physics: Fields, Particles, Algebras*, Texts and Monographs in Physics, Springer-Verlag, second edition, 1996.

[Had23] J. Hadamard, *Lectures on Cauchy's problem in linear partial differential equations*, Yale University Press, 1923.

[Hal90] G. S. Hall, *Conformal symmetries and fixed points in space-time*, J. Math. Phys. **31**(5), 1198–1207 (1990).

[Hal04] G. S. Hall, *Symmetries and Curvature Structure in General Relativity*, Lecture Notes in Physics, World Scientific, 2004.

[Har93] J. Harrison, *Stokes' theorem for nonsmooth chains*, Bull. Amer. Math. Soc. **29**(2), 235–242 (1993).

[Haw75] S. W. Hawking, *Particle creation by black holes*, Commun. Math. Phys. **43**(3), 199–220 (1975).

[Haw76] S. W. Hawking, *Black holes and thermodynamics*, Phys. Rev. D **13**(2), 191–197 (1976).

[HE73] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time*, Cambridge Monographs on Mathematical Physics, Cambridge University Press, 1973.

[Heu96] M. Heusler, *Black Hole Uniqueness Theorems*, Cambridge Lecture Notes in Physics, Cambridge University Press, 1996.

[HH76] J. B. Hartle and S. W. Hawking, *Path-integral derivation of black-hole radiance*, Phys. Rev. D **13**(8), 2188–2203 (1976).

[Hir76] M. W. Hirsch, *Differential Topology*, volume 33 of *Graduate Texts in Mathematics*, Springer-Verlag, 1976.

[Hör90a] L. Hörmander, *The Analysis of Linear Partial Differential Operators I: Distribution Theory and Fourier Analysis*, volume 256 of *Grundlehren der mathematischen Wissenschaften*, Springer-Verlag, second edition, 1990.

[Hör90b] L. Hörmander, *A remark on the characteristic Cauchy problem*, J. Funct. Anal. **93**(2), 270–277 (1990).

[Hör94] L. Hörmander, *The Analysis of Linear Partial Differential Operators III: Pseudo-Differential Operators*, volume 274 of *Grundlehren der mathematischen Wissenschaften*, Springer-Verlag, second edition, 1994.

[HP70] S. W. Hawking and R. Penrose, *The Singularities of Gravitational Collapse and Cosmology*, Proc. R. Soc. Lond. A **314**(1519), 529–548 (1970).

[Isr76] W. Israel, *Thermo-field dynamics of black holes*, Phys. Lett. A **57**(2), 107–110 (1976).

[JMR05] J.-L. Joly, G. Métivier and J. Rauch, *Hyperbolic domains of determinacy and Hamilton-Jacobi equations*, J. Hyperbolic Differ. Equ. **02**(03), 713–744 (2005).

[Joh55] F. John, *Plane Waves and Spherical Means Applied to Partial Differential Equations*, Interscience Publishers, 1955.

[Kay79] B. S. Kay, *Casimir effect in quantum field theory*, Phys. Rev. D **20**(12), 3052–3062 (1979), (originally appeared as a preprint under the title *The Casimir effect without magic*).

[Kay85] B. S. Kay, *The double-wedge algebra for quantum fields on Schwarzschild and Minkowski spacetimes*, Comm. Math. Phys. **100**(1), 57–81 (1985).

[Kay88] B. S. Kay, Quantum Field Theory in Curved Spacetime, in *Proceedings of the XVI International Conference on Differential Geometrical Methods in Theoretical Physics (Como, Italy, 1987)*, edited by K. Bleuler and M. Werner, pages 373–393, Kluwer Academic Publishers, 1988.

[Kay93] B. S. Kay, *Sufficient conditions for quasifree states and an improved uniqueness theorem for quantum fields on space–times with horizons*, J. Math. Phys. **34**(10), 4519–4539 (1993).

[Kay15] B. S. Kay, *Instability of enclosed horizons*, Gen. Rel. Grav. **47**(31), 1–27 (2015), 1310.7395.

[Kha14] I. Khavkine, *Covariant phase space, constraints, gauge and the Peierls formula*, Int. J. Mod. Phys. **29**(5) (2014), 1402.1282.

[KL16] B. S. Kay and U. Lupo, *Non-existence of isometry-invariant Hadamard states for a Kruskal black hole in a box and for massless fields on 1+1 Minkowski spacetime with a uniformly accelerating mirror*, Class. Quant. Grav. **33**(21), 215001 (2016), 1502.06582.

[KM14] I. Khavkine and V. Moretti, *Algebraic QFT in Curved Spacetime and quasifree Hadamard states: an introduction*, (2014), 1412.5945.

[KMS93] I. Kolář, P. W. Michor and J. Slovák, *Natural Operations in Differential Geometry*, Springer-Verlag, 1993, available at `https://www.mat.univie.ac.at/~michor/kmsbookh.pdf`.

[KN63] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry Volume 1*, Interscience Publishers, 1963.

[KN69] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry Volume 2*, Interscience Publishers, 1969.

[Kob58] S. Kobayashi, *Fixed points of isometries*, Nagoya Math. J. **13**, 63–68 (1958).

[Kre75] H.-O. Kreiss, Initial boundary value problems for hyperbolic partial differential equations, in *Proceedings of the Fourth International Conference on Numerical Methods in Fluid Dynamics, Vancouver (1974)*, pages 22–33, Springer, 1975.

[KSSV14] M. Kunzinger, R. Steinbauer, M. Stojković and J. A. Vickers, *A regularisation approach to causality theory for $C^{1,1}$-Lorentzian metrics*, Gen Rel. Grav. **46**(8), 1738 (2014), 1310.4404v3.

[Kup87] D. N. Kupeli, *On null submanifolds in spacetimes*, Geom. Dedicata **23**(1), 33–51 (1987).

[KW91] B. S. Kay and R. M. Wald, *Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on spacetimes with a bifurcate Killing horizon*, Phys. Rep. **207**(2), 49–136 (1991).

[Lee97] J. M. Lee, *Riemannian Manifolds: An Introduction to Curvature*, Graduate Texts in Mathematics, Springer-Verlag, 1997.

[Lee13] J. M. Lee, *Introduction to Smooth Manifolds*, Graduate Texts in Mathematics, Springer-Verlag, second edition, 2013.

[Ler53] J. Leray, *Hyperbolic Differential Equations*, Princeton lecture notes, Institute for Advanced Study, Princeton University (mimeographed), 1953.

[Lic61] A. Lichnerowicz, *Propagateurs et commutateurs en relativité générale*, Publ. IHÉS **10**, 5–56 (1961).

[Łoj59] S. Łojasiewicz, *Sur le problème de la division*, Studia Math. **18**(1), 87–136 (1959).

[LR12] J. Lucietti and H. S. Reall, *Gravitational instability of an extreme Kerr black hole*, Phys. Rev. D **86**(10), 104030 (2012), 1208.1437v3.

[LS90] L. H. Loomis and S. Sternberg, *Advanced Calculus*, Jones and Bartlett Publishers, second edition, 1990.

[Mal66] B. Malgrange, *Ideals of differentiable functions*, Oxford University Press, 1966.

[Mar72] K. B. Marathe, *A condition for paracompactness of a manifold*, **7**(3-4), 571–573 (1972).

[MI83] V. Moncrief and J. Isenberg, *Symmetries of cosmological Cauchy horizons*, Comm. Math. Phys. **89**(3), 387–413 (1983).

[Mic80] P. W. Michor, *Manifolds of differentiable mappings*, volume 3 of *Shiva mathematics series*, Shiva Publishing Ltd., 1980, available at `https://www.mat.univie.ac.at/~michor/manifolds_of_differentiable_mappings.pdf`.

[Min08a] E. Minguzzi, *Limit curve theorems in Lorentzian geometry*, J. Math. Phys. **49**(9), 092501 (2008), 0712.3942v2.

[Min08b] E. Minguzzi, *Non-imprisonment conditions on spacetime*, J. Math. Phys. **49**(6), 062503 (2008), 0712.3949v2.

[Min15a] E. Minguzzi, *Area Theorem and Smoothness of Compact Cauchy Horizons*, Comm. Math. Phys. **339**(1), 57–98 (2015), 1406.5919.

[Min15b] E. Minguzzi, *Convex neighborhoods for Lipschitz connections and sprays*, Monatshefte für Mathematik **177**(4), 569–625 (2015), 1308.6675.

[Mor03] V. Moretti, *Comments on the Stress-Energy Tensor Operator in Curved Spacetime*, Comm. Math. Phys. **232**(2), 189–221 (2003), gr-qc/0109048.

[MS82] R. B. Melrose and J. Sjöstrand, *Singularities of boundary value problems. II*, Comm. Pure Appl. Math. **35**(2), 129–168 (1982).

[MS08] E. Minguzzi and M. Sanchez, The Causal hierarchy of spacetimes, in *Recent Developments in pseudo-Riemannian geometry*, edited by H. Baum and D. Alekseevsky, ESI Lectures in Mathematics and Physics, pages 299–358, EMS, Zürich, 2008, gr-qc/0609119.

[MzH90] H. Müller zum Hagen, *Characteristic initial value problem for hyperbolic systems of second order differential equations*, Ann. Inst. H. Poincaré – Phys. Théor. **53**(2), 159–216 (1990).

[Nic07] L. I. Nicolaescu, *Lectures on the Geometry of Manifolds*, World Scientific, second edition, 2007, available at `https://www3.nd.edu/~lnicolae/Lectures.pdf`.

[O'N83] B. O'Neill, *Semi-Riemannian Geometry With Applications to Relativity*, Pure and Applied Mathematics, Academic Press, 1983.

[Par68] L. Parker, *Particle Creation in Expanding Universes*, Phys. Rev. Lett. **21**(8), 562–564 (1968).

[Paw02] W. Pawłucki, *A decomposition of a set definable in an o-minimal structure into perfectly situated sets*, Ann. Polon. Math. **79**(2), 171–184 (2002).

[Pen72] R. Penrose, *Techniques in Differential Topology in Relativity*, volume 7 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, SIAM, 1972.

[Pla16] Planck Collaboration, *Planck 2015 results*, A&A **594**, A13 (2016).

[Pol90] D. Polarski, *Minimally coupled scalar field on the static de Sitter space*, Phys. Rev. D **41**, 442–450 (1990).

[Rad92] M. J. Radzikowski, *The Hadamard condition and Kay's conjecture in (axiomatic) quantum field theory on curved space-time*, PhD thesis, Princeton University, 1992, available at `http://tinyurl.com/nnz769w` (retrieved 20 Sept 2016).

[Rad96] M. J. Radzikowski, *Micro-local approach to the Hadamard condition in quantum field theory on curved space-time*, Comm. Math. Phys. **179**(3), 529–553 (1996).

[Rau05] J. Rauch, *Precise finite speed with bare hands*, Methods Appl. Anal. **12**(3), 267–278 (09 2005).

[Ren90] A. D. Rendall, *Reduction of the Characteristic Initial Value Problem to the Cauchy Problem and Its Applications to the Einstein Equations*, Proc. R. Soc. Lond. A **427**(1872), 221–239 (1990).

[Rie49] M. Riesz, *L'intégrale de Riemann-Liouville et le problème de Cauchy*, Acta mathematica **81**(1), 1–222 (1949).

[Rin09] H. Ringström, *The Cauchy Problem in General Relativity*, ESI lectures in mathematics and physics, European Mathematical Society, 2009.

[RSZ12] J.-P. Ramis, J. Sauloy and C. Zhang, *Local analytic classification of q-difference equations*, (2012), 0903.0853v3.

[RV96] M. J. Radzikowski and R. Verch, *A local-to-global singularity theorem for quantum field theory on curved space-time*, Comm. Math. Phys. **180**(1), 1–22 (1996).

[Sak82] R. Sakamoto, *Hyperbolic Boundary Value Problems*, Cambridge University Press, 1982.

[Säm16] C. Sämann, *Global Hyperbolicity for Spacetimes with Continuous Metrics*, Ann. H. Poincaré **17**(6), 1429–1455 (2016), 1412.2408.

[San13] K. Sanders, *Thermal Equilibrium States of a Linear Scalar Quantum Field in Stationary Space-times*, Int. J. Mod. Phys. A **28**(10), 1330010 (2013), 1209.6068.

[San15] K. Sanders, *On the Construction of Hartle-Hawking-Israel States Across a Static Bifurcate Killing Horizon*, Lett. Math. Phys. **105**(4), 575–640 (2015), 1310.5537.

[Sau89] D. J. Saunders, *The Geometry of Jet Bundles*, volume 142 of *London Mathematics Society Lecture Note Series*, Cambridge University Press, 1989.

[Sch63] B. Schroer, *Infrateilchen in der Quantenfeldtheorie*, Fortschr. Physik **11**, 1–31 (1963).

[Sch13] S. Schubert, Über die Charakterisierung von Zuständen hinsichtlich der Erwartungswerte quadratischer Operatoren, Diploma thesis, University of Hamburg, 2013, available at `http://www.desy.de/uni-th/theses/Dipl_Schubert.pdf` (retrieved 20 Sept 2015).

[Seg63] I. E. Segal, *Mathematical Problems of Relativistic Physics*, Lectures in Applied Mathematics, American Mathematical Society, 1963.

[Sew82] G. Sewell, *Quantum fields on manifolds: PCT and gravitationally induced thermal states*, Ann. Phys. **141**(2), 201–224 (1982).

[Sla72] J. Slawny, *On factor representations and the C*-algebra of canonical commutation relations*, Comm. Math. Phys. **24**(2), 151–170 (1972).

[Sol06] D. A. Solis, *Global Properties of Asymptotically de Sitter and Anti de Sitter Spacetimes*, PhD thesis, University of Miami, 2006.

[Ste70] E. M. Stein, *Singular Integrals and Differentiability Properties of Functions*, volume 30 of *Princeton Mathematical Series*, Princeton University Press, 1970.

[SW70] R. F. Streater and I. F. Wilde, *Fermion states of a boson field*, Nucl. Phys. B **24**(3), 561–575 (1970).

[SW00] R. Streater and A. Wightman, *PCT, Spin and Statistics, and All That*, Landmarks in Physics, Princeton University Press, 2000.

[Tay06] M. E. Taylor, *Measure Theory and Integration*, Graduate studies in mathematics, American Mathematical Society, 2006.

[tH85] G. 't Hooft, *On the quantum structure of a black hole*, Nucl. Phys. B **256**(0), 727–745 (1985).

[Tie15] C. Tiee, *Computation and Visualization of Geometric Partial Differential Equations*, PhD thesis, University of California, San Diego, 2015, available at `https://ccom.ucsd.edu/~ctiee/ctiee-thesis-extended.pdf` (retrieved 20 Sept 2016).

[Tou72] J. C. Tougeron, *Idéaux de fonctions différentiables*, Erg. der Math., Springer-Verlag, 1972.

[Tre67] F. Treves, *Topological Vector Spaces, Distributions and Kernels*, Academic Press, 1967.

[Tre80] J.-F. Treves, *Introduction to Pseudodifferential and Fourier Integral Operators Volume 1: Pseudodifferential Operators*, University Series in Mathematics, Springer-Verlag, 1980.

[Tre11] J. H. Treude, Ricci curvature comparison in Riemannian and Lorentzian geometry, Diploma thesis, Albert-Ludwigs-Universität Freiburg, 2011, available at `http://freidok.uni-freiburg.de/volltexte/8405/pdf/diplomarbeit_-treude.pdf` (retrieved 1 Sept 2016).

[TU57] Y. Takahashi and H. Umezawa, *A general treatment of expanding systems*, Nuovo Cim. **6**(6), 1324–1334 (1957).

[Ull07] P. Ullrich, *Uniqueness in the characteristic Cauchy problem of the Klein–Gordon equation and tame restrictions of generalized functions*, Mathematische Zeitschrift **255**(4), 827–846 (2007), math-ph/0408022v1.

[Unr76] W. G. Unruh, *Notes on black hole evaporation*, Phys. Rev. **D14**, 870–892 (1976).

[Vin84a] A. M. Vinogradov, *Local symmetries and conservation laws*, Acta Appl. Math. **2**(1), 21–78 (1984).

[Vin84b] A. M. Vinogradov, *The $\mathscr{C}$-spectral sequence, Lagrangian formalism, and conservation laws. I. The linear theory. II. The nonlinear theory*, J. Math. Anal. Appl. **100**(1), 1–129 (1984).

[Wal94] R. M. Wald, *Quantum Field Theory in Curved Spacetime and Black Hole Thermodynamics*, Chicago Lectures in Physics, University of Chicago Press, 1994.

[Whi34] H. Whitney, *Analytic extensions of differentiable functions defined in closed sets*, Trans. Am. Math. Soc. **36**(1), 63–89 (1934).

[Whi57] H. Whitney, *Geometric Integration Theory*, Princeton University Press, 1957.

[Wig67] A. S. Wightman, Introduction to some aspects of the relativistic dynamics of quantized fields, in *Cargèse Lectures in Theoretical Physics: High energy electromagnetic interactions and field theory*, pages 171–291, Gordon and Breach Science Publishers, 1967.

[Zel72] Y. B. Zel'dovich, *Amplification of Cylindrical Electromagnetic Waves Reflected from a Rotating Body*, Zh. Eksp. Teor. Fiz. **62**, 2076–2081 (1972), English translation JETP **35**, 1085–1087 (1972).