The Quantization of Linear Gravitational Perturbations and the Hadamard Condition

David Stephen Hunt

A thesis submitted for the degree of PhD

University of York

Department of Mathematics

September 2012

Abstract

The quantum field theory describing linear gravitational perturbations is important from a cosmological viewpoint, in particular when formulated on de Sitter spacetime, which is used in inflationary models. There is currently an ongoing controversy pertaining to the existence of a de Sitter invariant vacuum state for free gravitons. This thesis is a mathematically rigorous study of the theory and all constructions are performed in as general a setting as is possible, which allows us to then specialise to a particular spacetime when required. In particular, to study the case of de Sitter spacetime with a view to resolving the aforementioned controversy. The main results include the full construction of the classical phase space of the linearized Einstein system on a background cosmological vacuum spacetime, which includes proving when various gauge choices can be made. In particular, we prove that within a normal neighbourhood of any Cauchy surface, in a globally hyperbolic spacetime, one may pass to the synchronous gauge. We also consider the transverse-traceless gauge but show that there is a topological obstruction to achieving this, which rules out its general use. In constructing the phase space it is necessary to obtain a weakly non-degenerate symplectic product. We prove that this can be achieved for the case that the background spacetime admits a compact Cauchy surface by using results from the Arnowitt-Deser-Misner (ADM) formalism, specifically the initial data splittings due to Moncrief. The system is quantized using Dirac's prescription, which permits the construction of an algebra of observables consisting of gauge-invariant smeared fields. It is shown that this algebra satisfies a time-slice condition. Finally, the states of the system are considered: we formulate the Hadamard condition and show that the Fock vacuum in Minkowski spacetime satisfies this definition.

Contents

List of Figures

Acknowledgements

First and foremost, I would like to sincerely thank my parents for their unwavering support, without which I would not have had the opportunity to undertake this degree. I would especially like to thank my supervisor, Dr C. J. Fewster, for his tireless help and support throughout my research. I would also like to thank the Department of Mathematics for accommodating me for the duration of my research, and thank all the staff for their accessibility whenever a problem arose. In particular, I would like to thank Dr S. P. Eveson, Dr A. Higuchi, Dr B. S. Kay and Dr. I. McIntosh. Finally, can I thank all my fellow PhD students and friends who have supported me throughout my studies. In particular, I would like to thank R. Friswell, B. Lang, L. Loveridge, T. Potts and J. Waldron for many extremely useful discussions throughout the duration of my research.

Declaration

Chapters 4 & 5 and section 6.1 of chapter 6 are based upon the paper "Quantization of linearized gravity in cosmological vacuum spacetimes" [34], which was cowritten by the author and his supervisor, Dr C. J. Fewster. The paper has been submitted for publication in the journal, Reviews of Mathematical Physics. The remainder of chapter six represents as yet unpublished research carried out in conjunction with Dr C. J. Fewster. The rest of this thesis, except where indicated, represents my own work.

Chapter 1

Introduction

The twentieth century heralded many great advances in our understanding of the physical world. Around 1915, Einstein put forward his general theory of relativity, which provided explanations for previously unexplained phenomena, such as the advance of the perihelion of Mercury, and also made new predictions, like the deflection of light rays by massive objects. It is currently the standard theory of gravity, against which all future expansions need to be judged. However, even before Einstein had proposed general relativity, various experiments were revealing that an entirely new framework, far removed from the 'classical' approaches to physics, was required to describe nature on the microscopic scale. This framework came to be known as quantum theory and since its introduction the philosophical implications of it have been a cause for great debate amongst the scientific community. However, there is no doubting its extraordinarily accurate experimental predictions and the focus since its formulation has been on incorporating various physical systems into its framework. In particular, a tremendous amount of research has been devoted to the attainment of a full quantum description of gravity. A number of candidate theories have been put forward, most notably string theory and loop quantum gravity, but as yet a quantum theory of gravity proves to be elusive.

A first approximation to combining gravity and quantum theory comes from the subject of quantum field theory in curved spacetimes. Here one treats the gravitational field classically within the framework of general relativity and studies the behaviour of quantum matter fields propagating on various spacetimes, with the goal of gaining an insight into the interaction between gravity and quantum theory. Currently the best descriptions for the behaviour of matter are conventional quantum field theories (formulated in Minkowski spacetime); therefore if we are to claim to have understood the various particle physics experiments that are carried out at colliders across the world within a lightly curved spacetime geometry and not the perfectly flat Minkowski spacetime, then it is vital that it be possible to formulate quantum field theories within curved spacetimes. Also, the early universe and other cosmological phenomena are all described by non-flat spacetime geometries, so if we wish to understand more clearly the behaviour of matter throughout the universe, then quantum field theory in curved spacetimes provides a very good method of approaching this. Indeed it is only when conditions become extreme, such as at the singularity of a black hole or the Big Bang singularity that the theory will break down and a full quantum description of gravity becomes necessary.

Many striking predictions have been made using quantum field theory in curved spacetimes, but the most celebrated of these are the Hawking effect [56], where a black hole is predicted to radiate particles at the Hawking temperature, and the Fulling-Unruh effect [46, 89], which describes how an accelerated observer in Minkowski spacetime will register the presence of particles in the vacuum state that is set by the inertial observers. The book of Wald [92] provides a full discussion of these topics.

Over the past fifty years, research into how to correctly formulate a quantum field theory on a curved spacetime has provided significant insights into the inner workings of quantum field theory itself. In particular, it has been found which structures and concepts are necessary to actually formulate the theory and which are merely useful for simplifying calculations. Most notably, when it comes to formulating rigorous results, it has led to the abandonment of the particle approach to quantum field theory, due primarily to the lack of Poincaré covariance and a preferred vacuum state on a general curved spacetime, and the use of the algebraic approach, first laid down for the case of Minkowski spacetime by Haag and Kastler [53] and fully generalised to arbitrary globally hyperbolic curved spacetimes by the local covariant approach of Brunetti, Fredenhagen and Verch [22]. We will give a more detailed discussion about this in section 3.1.

Unfortunately, quantum field theory in curved spacetimes still possesses the caveat that the gravitational field is treated entirely as a classical object. A first approximation to overcoming this is to utilise linear perturbation theory: by fixing a background spacetime and studying the behaviour of linear perturbations of the spacetime metric. Specifically, one quantizes these perturbations and treats them as another quantum field propagating on the chosen fixed classical background. This approach has found applications particularly in early universe cosmology, where one studies tensor fluctuations in the cosmic microwave background (CMB), see [94, Ch. 4 & 10] for a discussion of inflation and tensor fluctuations. A particular class of spacetimes used in the study of cosmology are the Friedmann-Robertson-Walker spacetimes, and there has been extensive research pertaining to the behaviour of quantized linear gravitational perturbations on these spacetimes, for further information see [43] and references therein.

Another important cosmological case is de Sitter spacetime, which becomes relevant for issues relating to inflation. In fact, its importance becomes even greater due to the persistence of a controversy concerning the existence of a suitable de Sitter invariant vacuum state for the free graviton field. By 'suitable' it is meant that the singular behaviour of the graviton two-point function in this state is of the Hadamard form, meaning that the state is a Hadamard state, and that there are no physical infrared divergences present. In [42] it was shown that if one constructs the graviton two-point function in the transverse-traceless and synchronous gauge, associated with the conformally flat coordinates defined on the Poincaré coordinate patch of de Sitter spacetime, then the expression for the two-point function is divergent due to an infrared divergence in the integral over modes. There are two distinct parties to this controversy, those [61] who maintain that there does exist a de Sitter invariant state and that the infrared divergence is a gauge artifact, and others [71] who argue to the non-existence of such a state. The main contention between the two sides is the validity of the use of Euclidean and analytic continuation methods, and the freedom to add gauge-breaking terms.

Therefore it seems appropriate to attempt to resolve these differences by means of another method, which is devoid of any of the previously mentioned techniques. The major theme of this thesis is to elaborate on an approach that was proposed jointly by the author with Dr C. J. Fewster in [34]. What is proposed is a rigorous framework for the consideration of the quantization of linear gravitational perturbations on general globally hyperbolic spacetimes, which obey the vacuum Einstein equation with cosmological constant,

$$G_{ab} + \Lambda g_{ab} = 0.$$

The reason for the restriction to these vacuum spacetimes is to ensure the gauge invariance of the linearized Einstein tensor with cosmological constant, which will be fully discussed in section 4.2. Such a formulation will then permit one to, when required, specialise down to a specific choice of background and also provides a setting for a rigorous investigation to be made into the Hadamard states using techniques from microlocal analysis.

Of course, there exist numerous treatments of the quantization of linear gravitational perturbations. In particular, the paper of [3] stands out for its treatment of the Hadamard condition. However, this quantization scheme was formulated by introducing gauge-breaking terms and ghost fields, and the Hadamard condition itself was not formulated using the now accepted and fully rigorous microlocal definition, which was originally introduced for the scalar field by Radzikowski [79]. The microlocal definition was expanded to include theories formulated in terms of vector bundles by Sahlmann and Verch [82]. As the motivation for formulating this approach was to circumvent the methodologies used by the parties in the controversy, we wish to avoid the introduction of gauge-breaking terms and auxiliary fields, and so will not discuss the results of [3] any further.

We adopt a minimalistic approach along the lines first laid down by Dimock for the case of the electromagnetic field [28], which like linearized gravity is a gauge theory. Indeed it is an interesting point that Dimock studied the quantization of the scalar field [26], Dirac field [27] and the electromagnetic field [28] on arbitrary globally hyperbolic spacetimes, but did not consider the graviton case. Therefore our approach also acts to fill a gap in the literature. In fact, not long after our paper [34] was submitted, the paper of Hack and Schenkel [55] appeared and showed how our framework fits nicely into their general approach to the quantization of linear theories with gauge invariance.

The essential content of Dimock's idea from the electromagnetic case is that the smeared fields, which are the basic observables of a quantum field theory, should be gauge invariant. There exist numerous treatments where this is not the case, and hence such approaches require further supplementary conditions (such as in the Gupta-Bleuler method), which just appear to over-complicate matters. This becomes notable by the desire to allow arbitrary smearings of the vector potential: if one permits arbitrary smearing tensors to be used, then the result is not a gauge invariant object; however, if one only works with a restricted class of smearing tensors, then it is possible to make the smeared vector potential observables into gauge invariant objects and, as we will see, resolve several issues that arise in the other methods. Dimock's papers each utilise the framework of the algebraic approach to quantum field theory and Dirac's prescription [29] for quantization. These two ideas entail that for our case, we construct a classical phase space and define a class of observables (functions) on this space. This system is then quantized by constructing an algebra of quantum observables, where Dirac's prescription is utilised to obtain the algebraic relations obeyed by the quantum observables.

We now briefly illustrate our approach. Given a spacetime $(M, \boldsymbol{g})$ that solves the vacuum Einstein equation with cosmological constant, the classical phase space for linearized gravity $\mathscr{P}_{\mathbb{C}}(M)$ consists of gauge equivalence classes of solutions to the linearized Einstein equation. Such equivalence classes are denoted by $[\boldsymbol{\gamma}]$. The observables we consider are the classical counterparts of the quantum smeared fields and are defined as follows. Given a smooth compactly supported rank $(0,2)$ tensor field $\boldsymbol{f}$, one considers the smeared field observable $F_{\boldsymbol{f}} : \mathscr{P}_{\mathbb{C}}(M) \to \mathbb{C}$ defined by

$$F_{\boldsymbol{f}}([\boldsymbol{\gamma}]) := \int_M \gamma_{ab} f^{ab} dvol_{\boldsymbol{g}}.$$

To ensure that this observable is gauge-invariant, that is, independent of the choice of representative used in the integral, one must restrict the smearing tensors to those which satisfy $\nabla^a f_{(ab)} = 0$.

One might be concerned as to whether this class of observables is 'large enough', in the sense that they are sufficient to distinguish different points of the phase space. In fact, as we will see in section 5.4, this issue is closely related to the weak non-degeneracy of the symplectic product, which we can show to be weakly non-degenerate for the case that the background spacetime admits a compact Cauchy surface. This result utilises the decompositions of initial data by Moncrief [72], which were derived in the Arnowitt-Deser-Misner (ADM) formulation of general relativity (we review this formulation in section 5.2). We prefer not to employ the ADM framework anywhere else as it emphasises a slicing of the spacetime, which might bring with it suspicions of dependence of our method on a particular slicing, coordinates or choices of linearized lapse and shift functions.

The remaining issue to be addressed, regarding the classical observables, is the calculation of the Poisson bracket of two of the observables. This is explicitly computed in section 5.4 and is found to agree with a result previously posited by Lichnerowicz [70]. With the phase space and observables established, the system can now be quantized by means of the algebraic framework and Dirac's prescription. The result is a $*$-algebra of observables that consists of polynomials of smeared quantum fields.

Having established this solid and rigorous framework, one is free to examine issues relating to Hadamard states using techniques from microlocal analysis. For the case of electromagnetism, this approach is described in [35], and due to the close links between the theories that we have already mentioned, we will adapt their approach to the linearized gravity case. A definition of Hadamard states is given in section 6.2.2, but it requires the introduction of two new concepts, namely the trace and trace-reversal of a bi-distribution that acts on smooth compactly supported rank $(0,2)$ tensor fields. We show that the vacuum state of the standard Minkowski Fock space construction of the graviton field is a

Hadamard state by our definition. However, we were not able to resolve this issue for the case of de Sitter spacetime, but we do provide discussions regarding its resolution within the conclusions made in chapter 7.

The overall strategy that has been put forward has numerous advantages. Firstly, it achieves the goal of avoiding the use of gauge-breaking terms or auxiliary fields, which are used by the parties to the controversy that was described earlier. Secondly, the approach is completely gauge-invariant, so there can be no arguments that any results depend upon a choice of gauge. Though, in particular instances, it will prove useful to make use of certain choices of gauge to simplify calculations involving gauge-invariant objects. Thirdly, the method can be implemented in arbitrary globally hyperbolic cosmological vacuum spacetimes, such as the important de Sitter case. Fourthly, it provides a nice clean separation between issues relating to observables (the algebra) and the issues pertaining to states, such as the Hadamard condition. This separation is one of the key points of the algebraic framework. Fifthly, the approach circumvents the known [18, 86] non-existence of a Wightman theory of linearized gravity on Minkowski spacetime that allows for arbitrary smearings. We will discuss the final point at length in section 6.1. Finally, a fully rigorous definition of Hadamard states is given, which has the potential to resolve the de Sitter controversy. There is however a disadvantage to our method, namely that the restriction on the available class of smearing tensors would prevent one from coupling this field to other fields. However, here we are purely interested in the free theory and so this issue is not of immediate concern.

This thesis is structured as follows. Chapter 2 outlines the various mathematical structures that will be used throughout the thesis. Particular attention is paid to the theory of differential operators and Green's operators, which will play a prominent role in the discussions of later chapters. We show how one can extend the action of the 'standard' Green's operators of [11], from smooth compactly supported tensor fields to smooth time-compact tensor fields, in particular, this leads to the proof of a generalisation of the usual exact sequence of [11, Thm 3.4.7]. This generalisation will be extremely useful when it comes to discussing the observables of the theory and isolating the algebraic relations which they satisfy. In chapter 2 we also provide a brief introduction to the theory of wavefront sets, which characterise the singular behaviour of a distribution. This concept will be important when it comes to addressing the issue of Hadamard states, where their very definition is given in terms of a wavefront set.

In chapter 3 we give an introduction to quantum field theory in curved spacetime. A brief historical overview of the development of the algebraic approach is given as well as the advantages of its use. We then develop the theory of the real scalar field along exactly the same lines as we will use for linearized gravity, so as to provide the reader with a familiar setting with which to compare our approach. When the phase space for the scalar field is constructed we address the frequently overlooked issue of how a smooth structure is placed upon it. The presence of a smooth structure is important as the standard definition of the Poisson bracket, see, for example, [1, pp. 566-568], is given in terms of the differential of the observables. We show how the use of a Frölicher space provides the answer to this issue. Finally in chapter 3 we give a brief historical overview of Hadamard states for the scalar field, noting the history of their development to becoming the primary attribute for a state in a free quantum field theory to be deemed physical. The Hadamard condition itself is also discussed within the framework of microlocal analysis, in preparation for considering the Hadamard condition for the free graviton quantum field.

Chapter 4 fully introduces the classical theory of linearized gravity. We introduce the theory using the geometrical approach of Stewart and Walker [84] and describe how one obtains the linearized Einstein equation that governs the behaviour of the perturbation. We will restrict attention to solutions obeying a certain type of boundary condition known as spacelike-compactness (see Definition 2.4.1). The motivations for this restriction are given in section 3.2.1. The restriction to spacelike-compactness forces one to consider two subspaces of pure gauge perturbations, which are in general not equal. As emphasised earlier, our approach is built upon gauge independence; however, particular choices of gauge do prove to be useful for technical purposes and so we describe three well-known choices of gauge and give particular attention to the circumstances under which they can be achieved. These gauges are the de Donder gauge, the transverse-traceless gauge and the synchronous gauge. The discussions of the de Donder gauge is largely standard. However, we are able to prove new results concerning the existence of the transverse-traceless gauge and the synchronous gauge. In particular, we find that there exists a topological obstruction to the attainment of the transverse-traceless gauge if the background spacetime solves the vacuum Einstein equation with vanishing cosmological constant. We also prove that on a normal neighbourhood of any Cauchy surface, the synchronous gauge condition can be satisfied by an arbitrary perturbation. The existence and uniqueness of solutions to the linearized Einstein equation is also reviewed in that chapter and we provide proofs for the case that the perturbation has spacelike-compact support. Similar results concerning existence and uniqueness are sketched in [41] for the case that the background spacetime admits a compact Cauchy surface. The final section of chapter 4 proves various results concerning the Green's operators used in linearized gravity. In particular, we show how any solution to the linearized Einstein equation is gauge equivalent to a perturbation that is equal to the action of the solution operator on a smearing tensor. We also show how the action of the Green's operators intertwines with the action of other operations, notably the trace, trace-reversal and the Lie derivative.

Chapter 5 considers the construction of the phase space and classical observables for linearized gravity. In particular, we prove that when the background spacetime admits a compact Cauchy surface, the space of degeneracies of the pre-symplectic product are precisely the pure gauge solutions. This is achieved using the splitting theorem's of Moncrief [72], which are proved within the ADM framework. We provide a full introduction to the ADM framework, as well as generalising the Moncrief splitting theorems from the case that the background spacetime solves the vacuum Einstein equation with vanishing cosmological constant to the case that the background solves the vacuum Einstein equation with a possibly non-vanishing cosmological constant. Unfortunately, a proof that, for the non-compact case, the space of degeneracies are just pure gauge solutions is not presently forthcoming and so we must conjecture that the result continues to hold in that instance. In chapter 5 we also define our classical observables and show how they are gauge invariant, and calculate their Poisson bracket. We also show how the commutator of our observables is equal to that originally conjectured by Lichnerowicz [70]. He used analogy with electromagnetism as well as previous results from Minkowski spacetime to motivate his commutator, whereas we show that the commutator arises from the Dirac quantization of our classical observables.

Chapter 6 deals with the quantum field theory. We discuss the algebra of observables for the graviton quantum field and show that this algebra respects a time-slice condition, meaning that the algebra of a slice of spacetime (containing a Cauchy surface) coincides with the algebra of the entire spacetime. A brief discussion is given into how our construction respects the axioms of local covariance [22], and we also discuss how our approach circumvents the issues pointed out by Strocchi regarding the non-existence of a Wightman formulation of the graviton quantum field on Minkowski spacetime. The final part of this chapter then deals with issues relating to states. We briefly comment on issues first raised by Moncrief [75] regarding linearization instabilities; this result states that if the background spacetime admits a compact Cauchy surface and global Killing vector fields, then for a state to be deemed physical it must be invariant under the action of the group of isometries (the Killing vector fields). However, our main focus in this chapter is on fully defining the Hadamard condition and checking whether or not our definition is respected by the standard Fock vacuum state from Minkowski spacetime. In fact, as we will show, the Minkowski Fock vacuum does respect our definition of Hadamard. Unfortunately, we are not able to report here on a resolution to the de Sitter controversy.

Finally, in chapter 7 we provide concluding remarks, where we summarise the main results of this thesis and provide an outlook for future research possibilities.

Chapter 2

Mathematical Structures

Motivated by the study of quantum fields on curved spacetimes, in this chapter we will discuss the various mathematical structures which underpin the theory and that will subsequently be used in later chapters. After firstly discussing the notations used here, we will begin by defining which mathematical restrictions we will place upon the spacetimes considered here. We assume that the reader has a knowledge of general relativity and differential geometry equivalent to part I and appendices A, B & C of [91].

2.1 Notation

We now briefly review the notation that will be used in the subsequent chapters, beginning with the function spaces. Our attention will be restricted to smooth tensor fields on a manifold M; as such we adopt the notation $C^\infty(M;\mathbb{K})$ for the space of smooth $\mathbb{K}$-scalar valued functions on M, and use $C^\infty(T^a_b(M;\mathbb{K}))$ to denote the space of smooth rank (a,b) $\mathbb{K}$-valued tensor fields on M. In the subsequent chapters we will always be clear as to our choice of $\mathbb{K}=\mathbb{R}$ or $\mathbb{K}=\mathbb{C}$. The support of a function or tensor field is defined to be the topological closure of the subset of points on which the field is non-zero. The subspaces consisting of elements with compact support are denoted by $C_0^\infty(M;\mathbb{K})$ and $C_0^\infty(T^a_b(M;\mathbb{K}))$ respectively. Two further subscripts on the function spaces will appear later, they are: SC denoting that a tensor has spacelike-compact support (see Definition 2.4.1), and TC denoting that a tensor has timelike-compact support (see Definition 2.4.4). The notation $S^2_0(M;\mathbb{K})$ (resp. $S^0_2(M;\mathbb{K})$) will be used to denote the symmetric elements of $T^2_0(M;\mathbb{K})$ (resp. $T^0_2(M;\mathbb{K})$).

Three conventions will be utilised to denote tensorial objects. Boldface type is used to indicate a tensor written with its indices suppressed. We also employ the abstract index notation, see [91, p. xi], as well as component expressions in a particular basis. It will be explicitly stated when coordinate expressions are being used, so one should assume otherwise that the abstract index convention is in use.

We adopt the standard convention that $\boldsymbol{w}^\flat$ denotes the covariant form of a vector field $\boldsymbol{w}$ and $\boldsymbol{v}^\sharp$ denotes the contravariant form of a covector field $\boldsymbol{v}$. In both instances the spacetime metric $\boldsymbol{g}$ is used to perform the transformation.

We take, as in [57, 91], the Riemann tensor to be defined by

$$R_{abc}{}^{d}\omega_d := (\nabla_a\nabla_b - \nabla_b\nabla_a)\omega_c \tag{2.1.1}$$

and the Ricci tensor by

$$R_{ac} := R_{abc}{}^{b}. \tag{2.1.2}$$

Finally, we will work in units for which $\hbar = c = 1$.

2.2 Spacetime

A *spacetime* is a pair $(M, \boldsymbol{g})$ consisting of a four-dimensional, smooth, real, connected, Hausdorff, orientable manifold without boundary M together with a smooth Lorentzian metric $\boldsymbol{g}$ of signature $(-+++)$, with respect to which M is time-orientable. All spacetimes will also be assumed to be globally hyperbolic, which is a condition that will be fully discussed in section 2.3. As our spacetimes are four-dimensional, real, connected, Hausdorff and admit a smooth Lorentzian metric, then by the Theorem in the appendix of [49], our spacetimes will be second countable, and hence [1, Prop. 5.5.5] paracompact. Note that our signature convention is not the standard one used when discussing quantum field theory in curved spacetimes, where the $(+---)$ convention is most prevalent. However, our $(-+++)$ convention is prevalent in the mathematical relativity literature, which motivates our choice.

The condition of time-orientability ensures that a consistent notion of future and past can be made across the entire manifold. Specifically, at each point $p \in M$ the tangent space T_pM can be divided into three classes: timelike vectors satisfy $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v}) < 0$, null vectors obey $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v}) = 0$ and spacelike vectors fulfil the condition $\boldsymbol{g}(\boldsymbol{v}, \boldsymbol{v}) > 0$, where $\boldsymbol{v} \in T_pM$. The timelike vectors form a double cone (vertices meeting at the origin) with the null vectors making up the boundary. However, there is no natural distinction between the two cones, that is, no natural future or past direction. A time-orientation is a continuous choice of one of these cones at each point of the manifold, and this is equivalent to choosing a continuous timelike vector field [77, Lem. 5.32].

On a time-orientable manifold one is able to discuss the causal structure of a spacetime. From this point forward we take causal curve to mean a curve whose tangent vector is always non-spacelike. A curve is said to be a future-directed timelike (resp. causal) curve if its tangent vector is always future-directed and timelike (resp. causal). Past-directed timelike curves and past-directed causal curves are defined similarly. The chronological (resp. causal) future of a set S is defined to be the collection of points that can be connected to at least one element of S by a smooth past-directed timelike (resp. causal) curve. The chronological and causal past of a set are defined similarly by replacing past-directed with future-directed. The chronological (resp. causal) future of a set S is denoted by $I^+(S)$ (resp. $J^+(S)$) and the chronological (resp. causal) past of that set is denoted by $I^-(S)$ (resp. $J^-(S)$).

2.3 Global hyperbolicity

We shall now discuss the causal restriction of globally hyperbolicity, which is a condition that we assume our spacetimes satisfy. The condition was first introduced by Jean Leray in his unpublished lecture notes on hyperbolic partial differential equations from 1952. He was motivated to obtain solutions to such differential equations on manifolds. The essential content is that global hyperbolicity ensures the well-posedness of certain partial differential equations and hence the uniqueness of certain Green's operators by ruling out various causal pathologies that a general spacetime, which may or may not solve the Einstein equation, can contain. Global hyperbolicity is therefore important from a physical standpoint where we want to be able to make predictions about the future and retrodictions of past events.

The definition of global hyperbolicity used here is the one given by Bernal and Sánchez, see item (i) from [16, p. 748].

Definition 2.3.1 *A spacetime $(M, \boldsymbol{g})$ is globally hyperbolic if it contains no closed causal curves and the set $J^+(p) \cap J^-(q)$ is compact for all $p, q \in M$.*

Previously it was thought that the condition of strong causality, which entails that there are no 'almost' closed causal curves (see [77, Dfn 14.11]), was required instead of just causality, but [16, Thm 3.2] showed that the condition of compactness of $J^+(p) \cap J^-(q)$ for all $p, q \in M$ means that causality implies strong causality.

This definition of global hyperbolicity is equivalent to the perhaps more intuitive definition that a spacetime admits a particular type of hypersurface known as a Cauchy surface. To explain what a Cauchy surface is, we will utilise the concept known as the domain of dependence of a set. Essentially, a point p will be in the domain of dependence of a set S if the state of, for instance, a physical field at that point is completely determined by its state on the surface S; this entails that no other 'information' can be propagated to p without first coming into contact with S. In relativity it is assumed that the speed of light is the speed limit at which information can be propagated and so the domain of dependence is defined accordingly.

Firstly we have to consider when a curve can be extended. Following [91, p. 193], a point p is said to be a future (resp. past) endpoint of a curve $\gamma : (a, b) \to M$ if the image of the curve converges to p for any increasing (resp. decreasing) sequence, where by increasing we mean that the curve parameter value increases as the sequence increases. Specifically, given any open neighbourhood O of p there exists a $T \in (a, b)$ such that for all $t > T$ (resp. $t < T$) we have $\gamma(t) \in O$. If a curve does not possess a future (resp. past) endpoint then it will be said to be future (resp. past) inextendible.

This allows one to define the domain of dependence of a set $S \subset M$. However, S will not be completely arbitrary. It will be assumed to be achronal, meaning that no two points in it can be joined by a timelike curve. This restriction is imposed because we are interested in applying the domain of dependence only to sets that will in a certain sense represent an instant of time. The future domain of dependence of an achronal set S is [77, Dfn 14.35] the set

$$D^+(S) := \{p \in M \mid \text{every past-inextendible smooth causal curve through } p \text{ hits } S\}.$$

It is important to only use inextendible curves because an extendible curve could freely terminate just prior to reaching S and hence this would entail, trivially, that $D^+(S) = S$. The past domain of dependence $D^-(S)$ is defined similarly but with future-inextendible replacing past-inextendible. The (total) domain of dependence of S is then defined to be

$$D(S) := D^+(S) \cup D^-(S).$$

A Cauchy surface is [91, p. 201] a closed, achronal set whose domain of dependence is the entire spacetime manifold. One could equivalently use the definition [77, Dfn 14.28] that a Cauchy surface is a subset met once by every inextendible timelike curve; equivalence of the two definitions comes from [77, Lem. 14.29]. Note that some authors [57, pp. 204-205] use acausal in their definition instead of achronal, and so exclude some possible Cauchy surfaces that would be considered by others. However, here we are only ever concerned with spacelike Cauchy surfaces and so in that case the difference between the two definitions is irrelevant. Now, by [77, Prop. 14.25], a Cauchy surface will be a topological hypersurface, and it serves as an initial surface where suitable data can be prescribed for the Cauchy problem. In fact, one can slice up an entire globally hyperbolic spacetime into a foliation of Cauchy surfaces of constant 'time'.

Originally Geroch [51, Thm 11] proved that global hyperbolicity is equivalent to the existence of a (topological) Cauchy surface and that the spacetime manifold is homeomorphic [51, Property 7] to the product manifold $\mathbb{R} \times \Sigma$, where Σ is the (topological) Cauchy surface. However, this is not very satisfactory, given that we predominantly work with smooth manifolds and attempt to solve differential equations smoothly. In a series of papers [13, 14, 15, 16], Bernal and Sánchez showed how Geroch's result can be generalised from the continuous case to the smooth case. Their main result [13, Thm 1] entails that if $(M, \boldsymbol{g})$ is globally hyperbolic then it is diffeomorphic to $\mathbb{R} \times \Sigma$, where Σ is a smooth spacelike Cauchy surface.

We end this section by stating some examples of globally hyperbolic and non-globally hyperbolic spacetimes. Minkowski spacetime, de Sitter spacetime and the exterior Schwarzschild spacetime are all globally hyperbolic, whereas anti de Sitter spacetime and the Gödel universe are not globally hyperbolic.

2.4 Partial differential equations on manifolds

The classical fields that we will consider all obey certain types of partial differential equations. It will therefore be necessary to appeal to various results concerning differential equations defined on manifolds and so those results are collated here.

2.4.1 Differential operators

We now summarise the main results of [11] that are relevant to our case. Everything will be explained in terms of smooth tensor fields on a globally hyperbolic spacetime $(M, \boldsymbol{g})$, rather than in terms smooth sections of general vector bundles as is done in [11]. Note that smooth tensor fields are an example of smooth sections of a vector bundle.

We will primarily be interested in certain types of second-order differential operators $P : C^\infty(T^a_b(M;\mathbb{K})) \to C^\infty(T^a_b(M;\mathbb{K}))$. Given a $\boldsymbol{u} \in C^\infty(T^a_b(M;\mathbb{K}))$, then in local coordinates, the action of a second-order linear differential operator takes the form

$$P(\boldsymbol{u})^A_B = m^{ijAC}_{\ \ \ BD}(x)\frac{\partial}{\partial x^i}\frac{\partial u^D_C}{\partial x^j} + n^{iAC}_{\ \ BD}(x)\frac{\partial u^D_C}{\partial x^i} + o^{AC}_{BD}(x)u^D_C,$$

where A and D are a shorthand notation for a contravariant spacetime indices, and B and C are a shorthand notation for b covariant spacetime indices. The $m^{ijAC}_{\ \ \ BD}(x)$, $n^{iAC}_{\ \ BD}(x)$ and $o^{AC}_{BD}(x)$ are smooth matrices on the spacetime $(M, \boldsymbol{g})$.

The principal symbol of this operator is obtained [24, p. 397] by considering only the leading order derivative terms, and replacing the partial derivatives ∂_j by $i\xi_j \in T^0_1(M;\mathbb{C})$ in those terms. Hence, at each point $x \in M$, the principal symbol of P is a linear map from rank (a, b) tensors at x to itself, given by

$$\sigma_x(P)(\xi) = -m^{ijAC}_{\ \ \ BD}(x)\xi_i\xi_j,$$

where $\xi \in T^*_x M$. The differential operator P will be said to have injective principal symbol if for each $\xi \neq 0$, the linear map $\sigma_x(P)(\xi)$ is injective [12, p. 383]. If $\sigma_x(P)(\xi) = -g^{ij}\xi_i\xi_j\delta^A_D\delta^C_B$, where $\boldsymbol{g}$ is the spacetime metric and $\delta^A_D = \delta^{\mu^1}_{\nu^1}\ldots\delta^{\mu^a}_{\nu^a}$, then P is said to be normally hyperbolic[1] [11, Ch. 1.5]. Note that if the metric was Riemannian as opposed to Lorentzian, then such an operator would be elliptic.

At this point we introduce the notion of spacelike-compactness [11, p. 90], which will be used extensively throughout the upcoming material.

Definition 2.4.1 *A tensor field is said to be spacelike-compact (SC) if its support is contained within $J(K)$ for some compact subset $K \subset M$.*

Of course, for a spacetime that admits a compact Cauchy surface, all tensor fields are spacelike-compact since $M \subset J(\Sigma)$, for any compact Cauchy surface Σ.

Normally hyperbolic differential operators arise very frequently in physics, for instance, the Klein-Gordon equation for the real scalar field and in the gauge theories of electromagnetism and linearized gravity, where the Maxwell equations in the Lorenz gauge and the linearized Einstein equations in the de Donder gauge both reduce to normally hyperbolic differential equations. The Cauchy problem is well-posed for normally hyperbolic differential operators. This means that given some smooth initial data compactly supported

[1] Note that [11, Ch. 1.5] use the convention that $\partial_i \to \xi_j$, but instead insert a minus sign by hand.

on a Cauchy surface (and a compactly supported source term) then there exists a unique smooth solution with spacelike-compact support [11, Thm 3.2.11] and the solution depends continuously on the initial data [11, Thm 3.2.12].

An object which will arise frequently throughout this thesis is a distribution. Given a finite-dimensional $\mathbb{K}$-vector space V, then, for us[2], a V-valued distribution[3] will be a continuous linear map, which takes elements of $C_0^\infty(T_b^a(M;\mathbb{K}))$ to V. One can endow the space $C_0^\infty(T_b^a(M;\mathbb{K}))$ with a suitable topology and so obtain a notion of convergence, see [11, p. 2] for details. Continuity of a distribution $\boldsymbol{u}$ is then expressed [11, Dfn 1.1.2] by the requirement that for all convergent sequences $(\boldsymbol{f})_n \in C_0^\infty(T_b^a(M;\mathbb{K}))$, where $\boldsymbol{f}_n \to \boldsymbol{f}$, we have $\boldsymbol{u}(\boldsymbol{f}_n) \to \boldsymbol{u}(\boldsymbol{f})$, where convergence here is with respect to the standard topology on a finite-dimensional vector space.

In forthcoming results it will be necessary to see how one can extend the notions of support and action of a differential operator to distributions. We begin by addressing the question of the support of a distribution. Given a distribution $\boldsymbol{u} : C_0^\infty(T_b^a(M;\mathbb{K})) \to V$, then we denote by $S_{\boldsymbol{u}}$ the set of points $x \in M$ that each have an open neighbourhood $O_x \subset M$ for which $\boldsymbol{u}(\boldsymbol{f}) = 0$ for all test functions $\boldsymbol{f} \in C_0^\infty(T_b^a(O_x;\mathbb{K}))$. The support of $\boldsymbol{u}$, denoted $\mathrm{supp}(\boldsymbol{u})$, is then [11, Dfn 1.1.7] the complement of $S_{\boldsymbol{u}}$ in M. Note that this definition of support applied to functions and tensor fields is equivalent to the usual definition of the support of a function and a tensor field, which is the closure of the set of points on which the function or tensor field does not vanish.

To extend the action of a differential operator to distributions requires the introduction of the formal adjoint of a differential operator. Specifically, associated to each linear differential operator $P : C^\infty(T_b^a(M;\mathbb{K})) \to C^\infty(T_b^a(M;\mathbb{K}))$ is [11, p. 5] another uniquely determined linear differential operator $P^* : C^\infty(T_b^a(M;\mathbb{K})) \to C^\infty(T_b^a(M;\mathbb{K}))$ called the formal adjoint of P, which is essentially calculated via integration by parts as follows: for all $\boldsymbol{f}, \boldsymbol{f}' \in C_0^\infty(T_b^a(M;\mathbb{K}))$ we have

$$\int_M f_A^B P(f)_B^A dvol_{\boldsymbol{g}} = \int_M P^*(f')'^B_A f_B^A dvol_{\boldsymbol{g}},$$

where A is shorthand notation for a contravariant spacetime indices, and B is shorthand notation for b covariant spacetime indices. Note that [11] do not assume the presence of a

[2]If like [11, Dfn 1.1.2] we were considering general vector bundles, then the domain of the distribution would just be the smooth compactly supported sections of the vector bundle in question.

[3]In Hörmander's terminology, they are distribution densities, but due to the presence of a preferred density, namely the one associated with the spacetime metric, the distinction between distributions and distribution densities is irrelevant, see [63, pp. 144-145] for details.

spacetime metric and so define the adjoint in terms of the dual bundle. One is now able to extend the action of a linear differential operator from smooth sections of a vector bundle to distributions on the vector bundle by defining [11, p. 5], for a distribution $\boldsymbol{u}$,

$$(P\boldsymbol{u})(f) := \boldsymbol{u}(P^* f)$$

for all $f \in C_0^\infty(T^a_b(M; \mathbb{K}))$.

A differential operator P is said to be formally self-adjoint if $P = P^*$. The differential operators that we will be concerned with will all be formally self-adjoint.

2.4.2 Fundamental solutions and Green's operators

Here we will discuss the main aspects of Green's operators, which act in a certain sense as inverses to a differential operator. As the hyperbolic differential operators that we consider will be formally self-adjoint, then we will restrict attention to only that case in the forthcoming material. Note that the theory of Green's operators for non self-adjoint differential operators is fully treated in [11].

Associated to each point $x \in M$ and to the vector bundle under consideration (in our case, the tensor bundle) is a distribution δ_x called the Dirac-delta distribution. The action of this distribution is to take a test tensor field $\boldsymbol{f} \in C_0^\infty(T^a_b(M; \mathbb{K}))$ and evaluate it at $x \in M$, that is,

$$\delta_x(\boldsymbol{f}) = \boldsymbol{f}|_x.$$

The purpose of this distribution is to make the Dirac-delta function into a rigorous concept.

We now begin a full discussion of Green's operators beginning with the objects which they are constructed from, namely the fundamental solutions. A distribution which when acted on by a linear differential operator P gives the Dirac-delta distribution δ_x is known as a fundamental solution for P at x. Specifically it is a distribution $F_x : C_0^\infty(T^a_b(M; \mathbb{K})) \to C_0^\infty(T^a_b(M; \mathbb{K}))|_x$ satisfying

$$P(F_x) = \delta_x.$$

If $\text{supp}(F_x) \subset J^+(x)$ (resp. $\text{supp}(F_x) \subset J^-(x)$) then F_x is called the retarded[4] (resp. advanced) fundamental solution at x. Given a normally hyperbolic operator P on a spacetime $(M, \boldsymbol{g})$, then it is known, see [11, Thm 3.3.1], for example, that at each point $x \in M$, there

[4] Our naming convention agrees with the standard usage in electrodynamics, see [64], for example, but differs from the convention chosen in [11].

exists a unique advanced fundamental solution associated with P and a unique retarded fundamental solution associated with P. One should note that various works, for example, [44, 70] and indeed [11, Ch. 2], establish this and later results only for the local case, that is, only on an open neighbourhood[5] $O_x \subset M$ of x. The implication would be that for the local results, F_x would only be defined for $C_0^\infty(T^a_b(O_x;\mathbb{K}))$ as opposed to the full $C_0^\infty(T^a_b(M;\mathbb{K}))$ in the global case. Henceforth the retarded (resp. advanced) fundamental solution will be denoted by F_x^+ (resp. F_x^-).

We now consider linear maps $\boldsymbol{E}^\pm : C_0^\infty(T^a_b(M;\mathbb{K})) \to C^\infty(T^a_b(M;\mathbb{K}))$ that will be inverses to the differential operator P on the space $C_0^\infty(T^a_b(M;\mathbb{K}))$. Specifically, for any $\boldsymbol{f} \in C_0^\infty(T^a_b(M;\mathbb{K}))$ we have

$$P\boldsymbol{E}^\pm \boldsymbol{f} = \boldsymbol{f}, \tag{2.4.1}$$

and

$$\boldsymbol{E}^\pm P\boldsymbol{f} = \boldsymbol{f}. \tag{2.4.2}$$

They are known as the advanced $(-)$/retarded $(+)$ Green's operators[6] and are uniquely singled out by their support properties, namely that $\operatorname{supp} \boldsymbol{E}^\pm \boldsymbol{f} \subset J^\pm(\operatorname{supp} \boldsymbol{f})$. The existence and uniqueness of such operators, on globally hyperbolic spacetimes, is guaranteed by [11, Cor. 3.4.3], and [11, Prop. 3.4.2] entails that they take the form

$$(\boldsymbol{E}^\pm \boldsymbol{f})(x) = F_x^\mp(\boldsymbol{f}), \tag{2.4.3}$$

where $F_x^\mp$ are the advanced/retarded fundamental solutions for P at x.

A subset $S \subset M$ is said [11, p. 18] to be past (resp. future) compact if $J^-(p) \cap S$ (resp. $J^+(p) \cap S$) is compact for all $p \in M$. This definition can be applied to the support of a tensor field. A $\boldsymbol{T} \in C^\infty(T^a_b(M;\mathbb{K}))$ is said to have past/future compact support if $\operatorname{supp}(\boldsymbol{T})$ is compact to the past/future. In particular, since $J^\pm(\operatorname{supp} \boldsymbol{f})$ are past/future compact [11, Lem. A.5.7] and $\operatorname{supp} \boldsymbol{E}^\pm \boldsymbol{f} \subset J^\pm(\operatorname{supp} \boldsymbol{f})$, then $\boldsymbol{E}^\pm \boldsymbol{f}$ also have past/future compact support. This leads us to consider the following theorem.

Theorem 2.4.2 *Given a normally hyperbolic operator $P : C^\infty(T^a_b(M;\mathbb{K})) \to C^\infty(T^a_b(M;\mathbb{K}))$, then there exists unique solutions with past/future compact support to the equation $P(\boldsymbol{\Phi}) = \boldsymbol{f}$, where $\boldsymbol{f} \in C_0^\infty(T^a_b(M;\mathbb{K}))$. These solutions are $\boldsymbol{\Phi} = \boldsymbol{E}^\pm \boldsymbol{f}$ respectively, where $\boldsymbol{E}^\pm$ are*

[5]There are further restrictions imposed on this neighbourhood, which we do not discuss here, but instead refer the reader to [11, Ch. 2].

[6]Note that [11, Dfn 3.4.1] use the naming convention that $\boldsymbol{E}^+$ is the advanced and $\boldsymbol{E}^-$ is the retarded operator.

the retarded/advanced Green's operators for P.

Proof. Consider the $\boldsymbol{E}^+$ case, the $\boldsymbol{E}^-$ case follows analogously. We know from the preceding discussions that there exists a linear operator $\boldsymbol{E}^+$ such that $\boldsymbol{\Phi} = \boldsymbol{E}^+\boldsymbol{f}$ solves $P(\boldsymbol{\Phi}) = \boldsymbol{f}$ with past compact support. Now, assume that there exists another solution $\boldsymbol{\chi}$ with past compact support solving $P(\boldsymbol{\chi}) = \boldsymbol{f}$. Then the difference $\boldsymbol{\Phi} - \boldsymbol{\chi}$ solves $P(\boldsymbol{\Phi} - \boldsymbol{\chi}) = 0$ and has past compact support. Therefore, by [11, Thm 3.1.1], it vanishes and so $\boldsymbol{\Phi} = \boldsymbol{\chi}$. ■

We now define an operator that will be prevalent throughout future chapters. The advanced-minus-retarded solution operator is defined as $\boldsymbol{E} := \boldsymbol{E}^- - \boldsymbol{E}^+$.

In section 4.7 we will refer several times to a result of [11], which gives an exact sequence built from the operators P and $\boldsymbol{E}$. Recall that a sequence is exact if at each stage the image of a map equals the kernel of the subsequent map. As this result [11, Thm 3.4.7] is important, it is included here.

Theorem 2.4.3 *Given a spacetime and a normally hyperbolic operator P with associated advanced-minus-retarded solution operator $\boldsymbol{E}$, then the following exact sequence holds:*

$$0 \longrightarrow C_0^\infty(T^a_b(M;\mathbb{K})) \xrightarrow{P} C_0^\infty(T^a_b(M;\mathbb{K})) \xrightarrow{\boldsymbol{E}} C_{SC}^\infty(T^a_b(M;\mathbb{K})) \xrightarrow{P} C_{SC}^\infty(T^a_b(M;\mathbb{K})).$$

Proof. Our definition of a spacetime, see chapter 2.2, satisfies the requirements of [11, Thm 3.4.7] and so the result follows from that theorem. ■

We now consider a new support property, which will be known as time-compactness. Intuitively, a tensor field will be said to have time-compact support if its support is bounded between two Cauchy surfaces. This is made more precise in the following definition.

Definition 2.4.4 *A tensor field $\boldsymbol{T} \in C^\infty(T^a_b(M;\mathbb{K}))$ is said to have time-compact support if there exists two Cauchy surfaces Σ and Σ' with $\Sigma \subset I^+(\Sigma')$, such that* $\operatorname{supp}\boldsymbol{T} \subset (J^+(\Sigma')\setminus \Sigma') \cap (J^-(\Sigma) \setminus \Sigma)$.

The subspace of time-compact smooth rank (a, b) $\mathbb{K}$-valued tensor fields on M is denoted by $C_{TC}^\infty(T^a_b(M;\mathbb{K}))$.

It will be necessary in section 5.4 to consider how the action of the Green's operators can be extended from smooth compactly supported tensor fields to smooth time-compact tensor fields. We achieve this extension as follows. Given some $\boldsymbol{f} \in C_{TC}^\infty(T^a_b(M;\mathbb{K}))$, let $\boldsymbol{E}^+\boldsymbol{f}$ denote the unique solution to

$$P(\boldsymbol{\Phi}) = \boldsymbol{f} \tag{2.4.4}$$

with past compact support. To show that such a solution to (2.4.4) exists, select a smooth spacelike Cauchy surface Σ, with future-pointing unit normal vector $\boldsymbol{n}$, such that $\Sigma \cap J^+(\operatorname{supp} \boldsymbol{f}) = \emptyset$, and let the initial data $(\boldsymbol{\Phi}|_\Sigma, \nabla_{\boldsymbol{n}} \boldsymbol{\Phi}|_\Sigma)$ vanish. Then, using the arguments from Corollary 5[7] of [10, Ch. 3], there exists a solution, denoted by $\boldsymbol{E}^+\boldsymbol{f}$, whose support lies within $J(\operatorname{supp} \boldsymbol{f})$. However, $\boldsymbol{E}^+\boldsymbol{f}$ vanishes to the past of $\operatorname{supp} \boldsymbol{f}$ by the choice of initial data, meaning that it is supported within $J^+(\operatorname{supp} \boldsymbol{f})$. Hence, we have $\operatorname{supp} \boldsymbol{E}^+\boldsymbol{f} \subset J^+(\operatorname{supp} \boldsymbol{f}) \subset J^+(\Sigma)$ and therefore $\operatorname{supp} \boldsymbol{E}^+\boldsymbol{f} \cap J^-(p) \subset J^+(\Sigma) \cap J^-(p)$ for all $p \in M$. On a globally hyperbolic spacetime $(M, \boldsymbol{g})$, it holds [77, Lem. 14.22] that $J^\pm(p)$ are closed for all $p \in M$. Therefore $\operatorname{supp} \boldsymbol{E}^+\boldsymbol{f} \cap J^-(p)$ is a closed subset of the compact [77, Lem. 14.40] set $J^+(\Sigma) \cap J^-(p)$, and so $\operatorname{supp} \boldsymbol{E}^+\boldsymbol{f} \cap J^-(p)$ is also compact. The uniqueness of the solution then follows from [11, Thm 3.3.1].

Similarly let $\boldsymbol{E}^-\boldsymbol{f}$ be the unique solution to (2.4.4) with future compact support. We now define the extensions of the Green's operators to time-compact tensor fields to be the linear maps $\boldsymbol{E}^\pm : C^\infty_{TC}(T^a_b(M;\mathbb{K})) \to C^\infty(T^a_b(M;\mathbb{K}))$ given by $\boldsymbol{f} \mapsto \boldsymbol{E}^\pm\boldsymbol{f}$, where $\boldsymbol{E}^\pm\boldsymbol{f}$ is the unique solution to (2.4.4) with past/future compact support. The following lemma establishes that these operators satisfy the relations (2.4.1) and (2.4.2) of the standard Green's operators as well as having identical support properties.

Lemma 2.4.5 *The extended Green's operators $\boldsymbol{E}^\pm : C^\infty_{TC}(T^a_b(M;\mathbb{K})) \to C^\infty(T^a_b(M;\mathbb{K}))$ satisfy:*

1. $P(\boldsymbol{E}^\pm\boldsymbol{f}) = \boldsymbol{f}$;
2. $\boldsymbol{E}^\pm P\boldsymbol{f} = \boldsymbol{f}$;
3. $\operatorname{supp} \boldsymbol{E}^\pm\boldsymbol{f} \subset J^\pm(\operatorname{supp} \boldsymbol{f})$,

for all $\boldsymbol{f} \in C^\infty_{TC}(T^a_b(M;\mathbb{K}))$.

Proof. (i) By the definition of the extended Green's operators, we have $P(\boldsymbol{E}^\pm\boldsymbol{f}) = \boldsymbol{f}$ for all $\boldsymbol{f} \in C^\infty_{TC}(T^a_b(M;\mathbb{K}))$. (ii) From our definition, we know that $\boldsymbol{\Phi} = \boldsymbol{E}^\pm P\boldsymbol{f}$ are the unique

[7] This corollary shows how the result [11, Thm 3.2.11] may be generalised to include the case that both the initial data and source have no restrictions placed upon their supports. The result is achieved by constructing an increasing sequence of relatively compact, globally hyperbolic subsets of the spacetime that cover the spacetime manifold. One then introduces a sequence of smooth compactly supported functions, one for each of the globally hyperbolic regions, that are equal to unity on the respective region. Using these functions to make the initial data and the source compactly supported, one can solve the standard Cauchy problem [11, Thm 3.2.11]. The main solution to the non-compactly supported Cauchy problem is then defined to be equal to the solution one obtains on each of the globally hyperbolic regions that were constructed. It is also shown that this solution will have support contained within $J(N)$, where N is the union of the supports of the initial data with the support of the source.

solutions to $P(\mathbf{\Phi}) = P(\boldsymbol{f})$ with past/future compact support. Therefore $P(\boldsymbol{E}^{\pm}P\boldsymbol{f} - \boldsymbol{f}) = 0$ and as both $\boldsymbol{E}^{\pm}P\boldsymbol{f}$ and $\boldsymbol{f}$ have past/future compact support, then by [11, Thm 3.3.1] we have $\boldsymbol{E}^{\pm}P\boldsymbol{f} = \boldsymbol{f}$. (iii) The support properties were established in the construction of the solutions $\boldsymbol{E}^{\pm}\boldsymbol{f}$. ■

The motivation for constructing these extensions is to generalise the exact sequence of Theorem 2.4.3, which is achieved by the following theorem.

Theorem 2.4.6 *Given a spacetime and a normally hyperbolic operator P with associated advanced-minus-retarded solution operator $\boldsymbol{E}$, then the following sequence is exact:*

$$0 \longrightarrow C^{\infty}_{TC}(T^a_b(M;\mathbb{K})) \xrightarrow{P} C^{\infty}_{TC}(T^a_b(M;\mathbb{K})) \xrightarrow{\boldsymbol{E}} C^{\infty}(T^a_b(M;\mathbb{K})) \xrightarrow{P} C^{\infty}(T^a_b(M;\mathbb{K})).$$

Proof. We begin by showing that at each stage the image of a map is contained within the kernel of the subsequent map, that is, the sequence is a complex. The composition of the first two maps vanishes because P is linear. At the second element we have $\boldsymbol{E}P\boldsymbol{f} = 0$ for all $\boldsymbol{f} \in C^{\infty}_{TC}(T^a_b(M;\mathbb{K}))$ by part (*ii*) of Lemma 2.4.5. Finally, at the third element, we have by the definition of the extended Green's operators, $P\boldsymbol{E}\boldsymbol{f} = 0$ for all $\boldsymbol{f} \in C^{\infty}_{TC}(T^a_b(M;\mathbb{K}))$. Therefore this sequence forms a complex.

We now need to show the reverse inclusions, that the kernel of a map is contained within the image of its predecessor. By displaying this we will show that the sequence is exact. We again consider each stage of the sequence individually. Firstly, the kernel of $P : C^{\infty}_{TC}(T^a_b(M;\mathbb{K})) \to C^{\infty}_{TC}(T^a_b(M;\mathbb{K}))$ consists of those $\boldsymbol{f} \in C^{\infty}_{TC}(T^a_b(M;\mathbb{K}))$ such that $P(\boldsymbol{f}) = 0$. As $\operatorname{supp} \boldsymbol{f}$ does not expand under the action of P and because any time-compact tensor is both past and future compact, then we have, by [11, Thm 3.3.1], $\boldsymbol{f} = 0$. The image of the zero element is again the zero element and so we have exactness at the first stage.

At the second element, the kernel of $\boldsymbol{E} : C^{\infty}_{TC}(T^a_b(M;\mathbb{K})) \to C^{\infty}(T^a_b(M;\mathbb{K}))$ consists of those $\boldsymbol{f} \in C^{\infty}_{TC}(T^a_b(M;\mathbb{K}))$ such that $\boldsymbol{E}\boldsymbol{f} = 0$, meaning that $\boldsymbol{E}^{-}\boldsymbol{f} = \boldsymbol{E}^{+}\boldsymbol{f}$. In this case, $\operatorname{supp} \boldsymbol{E}^{+}\boldsymbol{f} \subset \operatorname{supp} \boldsymbol{E}^{+}\boldsymbol{f} \cap \operatorname{supp} \boldsymbol{E}^{-}\boldsymbol{f} \subset J^{+}(\operatorname{supp} \boldsymbol{f}) \cap J^{-}(\operatorname{supp} \boldsymbol{f})$, which is bounded between any two Cauchy surfaces that are used to show that $\boldsymbol{f}$ is time-compact, and therefore $\boldsymbol{E}^{+}\boldsymbol{f} = \boldsymbol{E}^{-}\boldsymbol{f} \in C^{\infty}_{TC}(T^a_b(M;\mathbb{K}))$. By definition, we have $P(\boldsymbol{E}^{+}\boldsymbol{f}) = \boldsymbol{f}$, and hence $\boldsymbol{f} \in P(C^{\infty}_{TC}(T^a_b(M;\mathbb{K})))$.

Exactness at the third stage is shown by utilising the methodology of [92, Lem. 3.2.1]. The kernel of $P : C^{\infty}(T^a_b(M;\mathbb{K})) \to C^{\infty}(T^a_b(M;\mathbb{K}))$ consists of those $\mathbf{\Phi} \in C^{\infty}(T^a_b(M;\mathbb{K}))$ such that $P(\mathbf{\Phi}) = 0$. Select two arbitrary Cauchy surfaces Σ_1 and Σ_2 such that $\Sigma_2 \subset$

$I^+(\Sigma_1)$, now select a further two Cauchy surfaces Σ_3 and Σ_4 such that $\Sigma_1 \subset I^+(\Sigma_3)$ and $\Sigma^2 \subset I^-(\Sigma_4)$. Let $\chi \in C^\infty(M;\mathbb{K})$ be such that $\chi = 0$ in $J^-(\Sigma_1)$ and $\chi = 1$ in $J^+(\Sigma_2)$, since $J^-(\Sigma_1)$ and $J^+(\Sigma_2)$ are closed and disjoint sets, then the existence of such a function is guaranteed by [1, Prop. 5.5.8]. Define $\boldsymbol{f} := -P(\chi\boldsymbol{\Phi})$, it is clear that $\operatorname{supp}\boldsymbol{f} \subset (J^+(\Sigma_3) \setminus \Sigma_3) \cap (J^-(\Sigma_4) \setminus \Sigma_4)$ and so $\boldsymbol{f} \in C^\infty_{TC}(T^a_b(M;\mathbb{C}))$ by Definition 2.4.4. Now, $\boldsymbol{\psi} = -\chi\boldsymbol{\Phi}$ solves $P(\boldsymbol{\psi}) = \boldsymbol{f}$ with past-compact support. Hence, by uniqueness, we have $-\chi\boldsymbol{\Phi} = \boldsymbol{E}^+\boldsymbol{f}$. Also, $\boldsymbol{\psi} = (1-\chi)\boldsymbol{\Phi}$ solves $P(\boldsymbol{\psi}) = \boldsymbol{f}$ with future compact support and therefore, by uniqueness of solutions, $(1-\chi)\boldsymbol{\Phi} = \boldsymbol{E}^-\boldsymbol{f}$. Combining these results we see that $\boldsymbol{E}\boldsymbol{f} = \boldsymbol{E}^-\boldsymbol{f} - \boldsymbol{E}^+\boldsymbol{f} = (1-\chi)\boldsymbol{\Phi} - (-\chi\boldsymbol{\Phi}) = \boldsymbol{\Phi}$, so $\boldsymbol{\Phi}$ is in the image of $\boldsymbol{E}$. Hence, the sequence is exact. ■

2.5 The theory of differential forms

The results of section 4.5.2 will make extensive use of the theory of differential forms. As such we will now state the various conventions used, which are consistent with those of [1, 35, 78].

The space of $\mathbb{K}$-valued p-form fields on M is denoted by $\Omega^p(M;\mathbb{K})$ with a subscript 0 being added to denote compactly supported p-form fields or a subscript SC being added to denote spacelike-compact p-form fields.

Given a p-form field $\boldsymbol{\alpha} \in \Omega^p(M;\mathbb{K})$ and a q-form field $\boldsymbol{\beta} \in \Omega^q(M;\mathbb{K})$, it is possible [1, Prop. 6.3.6] to construct a $p+q$-form field called the wedge product $\boldsymbol{\alpha}\wedge\boldsymbol{\beta} \in \Omega^{p+q}(M;\mathbb{K})$ of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. In terms of components, the wedge product is given by

$$(\alpha\wedge\beta)_{a_1\dots a_{p+q}} = \frac{(p+q)!}{p!q!}\alpha_{[a_1\dots a_p}\beta_{a_{p+1}\dots a_{p+q}]}.$$

This in turn allows for the definition of the Hodge star operator, which is a map $* : \Omega^p(M;\mathbb{K}) \to \Omega^{n-p}(M;\mathbb{K})$ uniquely defined [1, Prop. 6.2.12] pointwise by the condition that for all $\boldsymbol{\alpha},\boldsymbol{\beta} \in \Omega^p(M;\mathbb{K})$,

$$\boldsymbol{\alpha}\wedge *\boldsymbol{\beta} = (\boldsymbol{\alpha},\boldsymbol{\beta})_{\boldsymbol{g}}dvol_{\boldsymbol{g}},$$

where $(\boldsymbol{\alpha},\boldsymbol{\beta})_{\boldsymbol{g}}$ is the contraction of $\boldsymbol{\alpha}$ with $\boldsymbol{\beta}$ using the spacetime metric $\boldsymbol{g}$. The volume element $dvol_{\boldsymbol{g}}$ is an n–form field on the manifold. Explicitly it is [91, eq. B.2.17] given by $(dvol_{\boldsymbol{g}})_{a_1\dots a_n} = \epsilon_{a_1\dots a_n}$, where

$$\epsilon_{a_1\dots a_n} = \sqrt{|\det\boldsymbol{g}|}\tilde{\epsilon}_{a_1\dots a_n}$$

is the Levi-Civita tensor and $\tilde{\epsilon}_{a_1\ldots a_n}$ is the Levi-Cevita symbol, which takes the value of 1 when the indices are an even permutation of $(1\ldots n)$, the value -1 when they are an odd permutation and zero otherwise. Given an $\boldsymbol{\alpha} \in \Omega^p(M;\mathbb{K})$, then the Hodge star of $\boldsymbol{\alpha}$ is [91, p. 88] given by

$$(*\alpha)_{a_{p+1}\ldots a_n} = \frac{1}{p!}\alpha^{a_1\ldots a_p}\epsilon_{a_1\ldots a_p a_{p+1}\ldots a_n},$$

which agrees with the pointwise formula given in [1, p. 413].

The square of the Hodge star operator is pointwise [1, Prop. 6.2.13] a multiple of the identity operator given by

$$(*)^2 = (-1)^{p(n-p)+s},$$

where n is the dimension of the manifold and s is the index of the metric $\boldsymbol{g}$, that is, the number of negative entries present in the matrix representation of an orthogonal decomposition of $\boldsymbol{g}$. For the case of a four-dimensional manifold with a Lorentzian metric of either signature, the square of the Hodge star operator acting on a p-form simplifies to $(*)^2 = (-1)^{p+1}$, whilst on a spacelike Cauchy surface of the said spacetime, it becomes just the identity operator, $(*)^2 = 1$, in the $(-+++)$ conventions and minus the identity operator, $(*)^2 = -1$, in the $(+---)$ conventions.

The Hodge star operator is used to construct a pairing [1, p. 538] between p-forms on a manifold M; given a $\boldsymbol{\alpha} \in \Omega^p(M;\mathbb{K})$ and $\boldsymbol{\beta} \in \Omega^p_0(M;\mathbb{K})$ then one defines

$$\langle \boldsymbol{\alpha}, \boldsymbol{\beta}\rangle_M := \int_M \overline{\boldsymbol{\alpha}} \wedge *\boldsymbol{\beta},$$

where $\overline{\boldsymbol{\alpha}}$ denotes the complex conjugate of $\boldsymbol{\alpha}$. This definition makes sense because the above will just be the integral over M of the scalar function $(\overline{\boldsymbol{\alpha}}, \boldsymbol{\beta})_{\boldsymbol{g}}$ with respect to the volume element $dvol_{\boldsymbol{g}}$. If M were compact then the restriction that $\boldsymbol{\beta}$ be compactly supported is trivially satisfied.

The exterior derivative $d : \Omega^p(M;\mathbb{K}) \to \Omega^{p+1}(M;\mathbb{K})$ is the fundamental derivative operator used in differential forms. Its action, in component form, on a p-form $\boldsymbol{\alpha}$ is given by

$$(d\alpha)_{a_1\ldots a_{p+1}} = (p+1)\nabla_{[a_1}\alpha_{a_2\ldots a_{p+1}]}.$$

One can easily replace the covariant derivative with the partial derivative as the terms involving the Christoffel symbols vanish by virtue of their symmetry properties. Also the square of the exterior derivative vanishes identically, due to the equality of mixed partial derivatives for smooth fields.

By combining the exterior derivative with the Hodge star operator, it is possible to construct another derivative operator known as the codifferential δ. It is the map $\delta : \Omega^p(M;\mathbb{K}) \to \Omega^{p-1}(M;\mathbb{K})$ defined [1, Dfn 6.5.21] by

$$\delta = (-1)^{n(p-1)+s+1} * d* \tag{2.5.1}$$

and the condition that it always annihilates zero-forms. Using (2.5.1), that $d^2 \equiv 0$ and that $(*)^2$ is a multiple of the identity, then it is clear that $\delta^2 \equiv 0$. On a four-dimensional spacetime M with Lorentzian metric $\boldsymbol{g}$ of either signature, $\delta = *d*$; whilst on a spacelike Cauchy surface of such a spacetime, we have $\delta = (-1)^p * d*$ in the $(- + ++)$ convention and $\delta = (-1)^{p+1} * d*$ in the $(+ - --)$ convention.

Finally, Stokes' Theorem [1, Thm 7.2.8] states that, given an n-dimensional orientable compact manifold M with boundary ∂M and given an $\boldsymbol{\alpha} \in \Omega^{n-1}(M;\mathbb{K})$ then

$$\int_{intM} d\boldsymbol{\alpha} = \int_{\partial M} i^*\boldsymbol{\alpha},$$

where $i : \partial M \to M$ is the inclusion map. Stokes' Theorem can then be used to show that, when M is boundaryless, δ is the adjoint of d with respect to the product $\langle \cdot, \cdot \rangle_M$, that is,

$$\langle d\boldsymbol{\alpha}, \boldsymbol{\beta} \rangle_M = \langle \boldsymbol{\alpha}, \delta\boldsymbol{\beta} \rangle_M \tag{2.5.2}$$

for $\boldsymbol{\alpha} \in \Omega^{p-1}(M;\mathbb{K})$ and $\boldsymbol{\beta} \in \Omega^p(M;\mathbb{K})$, provided at least one of them is compactly supported when M is not compact.

2.6 Microlocal analysis

In this section we briefly describe some techniques of microlocal analysis which permit one to examine the singular (unsmooth) behaviour of distributions. We first state and discuss these results for the case of scalar distributions on $\mathbb{R}^n$ before discussing the generalisation to distributions on manifolds and to vector bundle distributions.

We begin by considering scalar functions and distributions on $\mathbb{R}^n$. Given an open subset $U \subseteq \mathbb{R}^n$, we adopt the standard notation of $\mathscr{D}(U)$ for $C_0^\infty(U;\mathbb{C})$ and $\mathscr{D}'(U)$ for the space of distributions on $\mathscr{D}(U)$, that is, the space of continuous linear functionals, or the dual space. Given a distribution $u \in \mathscr{D}'(U)$, then continuity is expressed, see Theorem 2 in [10, Sec. 4.2.1], by the requirement that for each compact subset $K \subset U$, there exist constants

$C \in \mathbb{R}$ and $k \in \mathbb{N}$, such that we have the following estimate:

$$|u(f)| < C \sum_{|\alpha|<k} \sup_{x\in K} |\partial^\alpha f|$$

for all $f \in \mathscr{D}(K)$, where α is a multi-index. Also, we adopt the standard notation of $\mathscr{E}(U)$ for $C^\infty(U;\mathbb{C})$ and $\mathscr{E}'(U)$ for the dual space of distributions. In this instance, by Theorem 1 of [10, Sec. 4.2.1], continuity is expressed in an exactly analogous manner. Given a $u \in \mathscr{E}'(U)$, then there exists a compact subset $K \subset U$, a $C \in \mathbb{R}$ and a $k \in \mathbb{N}$, such that we have the following estimate:

$$|u(f)| < C \sum_{|\alpha|<k} \sup_{x\in K} |\partial^\alpha f|$$

for all $f \in \mathscr{E}(U)$. The elements of $\mathscr{E}'(U)$ are in fact all compactly supported, see, for example, the argument given around the estimate in [10, eq. (4.3)] for a proof of this.

The operation which allows one to analyse the singular behaviour of a function is the Fourier transform. Therefore it is natural to implement this operation on distributions to investigate their behaviour. For a full exposition of these methods, see the famous text of Hörmander [63] or the more recent introduction to microlocal analysis given by Strohmaier in [10, Ch. 4].

To define the Fourier transform of a distribution, one begins with the compactly supported ones. Given a $u \in \mathscr{E}'(U)$, its Fourier transform is defined [63, Thm 7.1.14] to be the function $\hat{u}(k) := u(e^{ik\cdot x})$. As $u \in \mathscr{E}'(U)$ and $f_k(x) = e^{ik\cdot x}$ is a smooth function, then $u(f_k)$ is well-defined. Even if one is given an arbitrary distribution v, then one can still Fourier analyse it. To achieve this, one multiplies v by a $\chi \in C_0^\infty(\mathbb{R}^n;\mathbb{C})$; the resulting distribution χu is compactly supported and its action is given by $(\chi v)(f) = v(\chi f)$ for all $f \in C_0^\infty(\mathbb{R}^n;\mathbb{C})$. One can then Fourier analyse the behaviour of the compactly supported distribution χv, and vary χ to give further insights into the behaviour of v

To motivate the definition of smoothness used for distributions, we briefly consider the behaviour of the Fourier transform of a smooth compactly supported function on $\mathbb{R}^n$. It is a well-known result, see, for example, the beginning of [63, Sec. 8.1], that given a $f \in C_0^\infty(U;\mathbb{C})$, its Fourier transform will satisfy the following estimate:

$$|\hat{f}(k)| \leq \frac{C_N}{1+|k|^N}$$

for all $N \in \mathbb{N}$, where $C_N \in \mathbb{R}$ is constant for each N. Hence we see that smooth compactly supported functions have Fourier transforms which as $|k|$ becomes large, decay faster than any power of $|k|$. The standard terminology is that smooth compactly supported functions have 'rapidly decaying' Fourier transforms.

With this estimate in mind, and the fact that we can Fourier analyse an arbitrary distribution in $\mathscr{D}'(U)$, one can give [32, Dfn 3.1] a criterion for smoothness of a distribution. Given an open subset $U \subseteq \mathbb{R}^n$, one calls a point $(x, k) \in U \times (\mathbb{R}^n \setminus \{0\})$ a regular direction of a distribution $u : C_0^\infty(U; \mathbb{C}) \to \mathbb{C}$ if there exists a $\chi \in C_0^\infty(U; \mathbb{C})$ such that $\chi(x) \neq 0$ and an open conic neighbourhood[8] $\Gamma \subset \mathbb{R}^n \setminus \{0\}$ containing k, such that

$$|\widehat{(\chi u)}(k)| \leq \frac{C_N}{1 + |k|^N}$$

for all $k \in \Gamma$ and for all $N \in \mathbb{N}$, where $C_N \in \mathbb{R}$ is constant for each N.

The regular directions describe where and in which directions the distribution is smooth. The wavefront set of u, denoted $\mathrm{WF}(u)$ is [63, Dfn 8.1.2] the complement in $U \times (\mathbb{R}^n \setminus \{0\})$ of the set of regular directions of u, and so the wavefront set describes exactly how the distribution fails to be smooth. As will be discussed in section 3.3, in quantum field theory the class of physical states of a free theory are characterised by the singular behaviour of their two-point function. The wavefront set contains all of this information and so is used in the definition of the physical states.

Having considered distributions defined on $\mathbb{R}^n$, we now consider the generalisation of the wavefront set to distributions on manifolds. The local Euclidean nature of manifolds facilitates an easy transition to this case. The manifolds are assumed to obey all of the assumptions set out in section 2.2. Now, a manifold M admits local coordinate charts (U_i, ψ_i), where $U_i \subset M$ is open and $\psi_i : U_i \to \mathbb{R}^n$. Given a distribution $u \in \mathscr{D}'(M)$, then [63, Dfn 6.3.3] it has a representative distribution, denoted by ${\psi_i}^* u$, in each coordinate chart (U_i, ψ_i). In the chart (U_i, ψ_i), a point $(p, \xi) \in T^*U_i$ has the chart expression $(x, k) = (\psi_i(p), \psi_i^* \xi) \in \mathbb{R}^n \times (\mathbb{R}^n \setminus \{0\})$. This allows for the wavefront set of a distribution on a manifold to be defined [63, p. 265] by saying that $(p, \xi) \in \mathrm{WF}(u)$ if and only if $(x, k) \in \mathrm{WF}({\psi_i}^* u)$. The wavefront set is a subset of the cotangent bundle of M (with the zero-section excised). As described on [63, p. 265], the definition of $\mathrm{WF}(u)$ is independent of the choice of coordinates because under a change of coordinates, one has a diffeomorphism $\phi : \psi_i(U_i) \to \psi_j(U_j)$, where $\psi_i(U_i), \psi_j(U_j) \subset \mathbb{R}^n$ are open, and it holds that $\mathrm{WF}(\phi^* u) =$

[8] An open conic neighbourhood is an open set which is scale invariant, that is, if $p \in \Gamma$ then $\lambda p \in \Gamma$ for all positive $\lambda \in \mathbb{R}$.

$\phi^*\mathrm{WF}(u)$ for all $u \in \mathscr{D}'(\psi_i(U_i))$.

For the case of vector bundles [63, p. 265] and [82, Sec. 2.3], one uses local trivialisations, so that a smooth compactly supported section of a k-dimensional vector bundle corresponds locally to a k-tuple $(f_1, \ldots, f_k)$ of smooth compactly supported scalar functions. This correspondence is one-to-one and induces a one-to-one correspondence between vector bundle distributions and a k-tuple of scalar distributions. Under the correspondence between distributions, the wavefront set of the vector bundle distribution on an open set that trivialises the vector bundle, is equal to the union of the wavefront sets of the scalar distributions that represent it. One then says that the wavefront set of a general vector bundle distribution is defined as the collection of points such that their coordinate representation, in a local trivialisation, lies in the wavefront set of the representative distribution of that trivialisation.

Chapter 3

Quantum Field Theory in Curved Spacetimes

In this chapter we will discuss how one formulates a quantum field theory on a curved spacetime within the framework of the algebraic approach. The example of the free real scalar field will serve to fully illustrate this approach, which will be used for the case of linearized gravity in chapters 5 & 6. Finally, in preparation for their consideration in the free graviton case, we will give a review of the notion of Hadamard states for the scalar field.

3.1 Algebraic approach

The standard textbook approach to constructing a quantum field theory is to use mode expansions, see, for instance, the book of Birrell and Davies [17]. Such a treatment results in the construction of a Hilbert space, a Fock space, using the creation and annihilation operators associated to the modes and the selection of a vacuum vector that is annihilated by all annihilation operators. On this Hilbert space, observable quantities are represented by self-adjoint operators. However, the construction of the Fock space is dependent upon the choice of modes. For the case of Minkowski spacetime, when one is working with special relativity, a particularly special class of modes are picked out, namely the positive frequency plane waves. In that instance, the resulting Hilbert space is the unique Fock space whose vacuum state vector is invariant under the action of the Poincaré group [17, p. 47], which has the physical interpretation that all inertial observers will agree on what is to be regarded as the vacuum state. The fact that we have a preferred Fock space entails that it is meaningful to talk about particles, but for curved spacetimes this is no longer the case.

A general curved spacetime will possess no symmetries that can be used to pick out a preferred representation, through invariance of a vacuum state, and so the particle interpretation becomes blurred. Such issues were first pointed out and discussed by Fulling [46]. In fact, he described how the non-uniqueness of the particle description arises even if one is examining a subset of Minkowski spacetime, in that instance, Rindler spacetime. What he described became to be known as the Fulling-Unruh effect, where an accelerated observer in Minkowski spacetime sees the inertial vacuum state as containing particles; this is compared to the lack of particles registered by an inertial observer. For a full discussion, see [92, Ch. 5].

The algebraic approach is completely different in that one does not need a Hilbert space to formulate the theory. The focus is entirely on the algebraic relations between the observable quantities and not on representing them on a Hilbert space. One achieves this by constructing an abstract algebra that will contain all the observables of the theory. Once the algebra has been constructed, then one is free to discuss issues pertaining to states, and with that issues relating to measurements through expectation values.

The algebraic approach was initially championed by several authors throughout the 1950s, notably Araki, Haag and Segal. However, in those treatments, the focus was on operator algebras on Hilbert spaces. A true abstract treatment, devoid of any reference to Hilbert spaces, was first laid down by Haag and Kastler [53]. Unfortunately, their approach still had the caveat that it was restricted to the case of Minkowski spacetime.

The approach of Haag and Kastler is to lay down a collection of axioms which an algebra that describes the observables of a quantum field theory in Minkowski spacetime should obey. Since being first laid down, these axioms have undergone modifications. We now discuss the axioms as they appear in [52, p. 110].

One begins by considering the open subsets of Minkowski spacetime with compact closure. The reason for considering only the regions with compact closure is to ensure that one is only ever taking into account those observables that can be measured within a finite region such as a laboratory. This rules out global observables such as total energy and charge. For each such subset $\mathscr{O}$, a quantum field theory assigns an algebra $\mathcal{A}(\mathscr{O})$, which will contain all the local observables that can be measured within the region $\mathscr{O}$. Each $\mathcal{A}(\mathscr{O})$ is assumed to be a C^*-algebra with the $*$ operation being interpreted as taking the adjoint. The smallest C^*-algebra containing the union of all the algebras over all the regions with compact closure will be denoted by $\mathcal{A}$ and is known as the algebra of observables for the spacetime. When we discuss the scalar field and linearized gravity we will use $*$-algebras instead of C^*-algebras. The assignment $\mathscr{O} \mapsto \mathcal{A}(\mathscr{O})$ is frequently referred to as the selection of a net, but we caution that this differs from the usual usage of net [80, p. 96], which is a map from a directed system to a topological space, in that the target space is no longer a topological space.

To ensure that the net describes the observables of a quantum field theory, certain extra conditions need to be imposed on it. These conditions are the Haag-Kastler axioms and we now state them in their modern form:

1. *Isotony:* given two regions $\mathscr{O}_1$ and $\mathscr{O}_2$ such that $\mathscr{O}_1 \subset \mathscr{O}_2$ then $\mathcal{A}(\mathscr{O}_1) \subset \mathcal{A}(\mathscr{O}_2)$;

2. *Locality:* if two regions $\mathscr{O}_1$ and $\mathscr{O}_2$ are spacelike separated then the algebras $\mathcal{A}(\mathscr{O}_1)$ and $\mathcal{A}(\mathscr{O}_2)$ commute, in the sense of them being subalgebras of $\mathcal{A}$;

3. *Poincaré covariance:* the algebra $\mathcal{A}$ carries a representation of the Poincaré group $\mathcal{P}$ via automorphisms α_g, where $g \in \mathcal{P}$, and these automorphisms satisfy $\alpha_g(\mathcal{A}(\mathscr{O})) = \mathcal{A}(g(\mathscr{O}))$ for all regions $\mathscr{O}$ and for all $g \in \mathcal{P}$.

4. *Causality:* given a region $\mathscr{O}$, let $\hat{\mathscr{O}}$ denote its causal completion[1], then $\mathcal{A}(\mathscr{O}) = \mathcal{A}(\hat{\mathscr{O}})$.

The first condition ensures that if a measurement can be made in a region $\mathscr{O}_1$, which satisfies $\mathscr{O}_1 \subset \mathscr{O}_2$, then that measurement can also be performed in the second region $\mathscr{O}_2$. The second condition is the point where relativity enters, it says that measurements made in causally disconnected regions cannot influence each other. The third condition expresses the idea that a relativistic theory in Minkowski spacetime should be Poincaré covariant. The fourth and final condition displays the presence of a hyperbolic dynamical law. In particular, if one can extend this relation to include certain unbounded regions, which contain a Cauchy surface for the ambient spacetime, then a time-slice condition will hold. The time-slice condition was first discussed by Haag & Schroer [54] under the title of "primitive causality", and it entails that the algebra of a causally convex[2] neighbourhood of a Cauchy surface (a slice of spacetime) coincides with the algebra of the entire spacetime. As the fourth relation is only defined for relatively compact subsets of Minkowski spacetime,

[1] Following [52, p. 143], the causal complement of $\mathscr{O}$ is defined as the set $\mathscr{O}^{\perp} := M \setminus J(\mathscr{O})$. The causal completion of $\mathscr{O}$ is defined to be the set $\hat{\mathscr{O}} := (\mathscr{O}^{\perp})^{\perp}$. We caution that there do exist differing notions of causal complement, see [36, App. A.3] for a full discussion.

[2] A subset Ω is said to be causally convex if every smooth causal curve with endpoints in Ω is contained entirely within Ω [36, p. 8].

if one can extend it to include non-relatively compact causally convex subsets, then as the causal completion of a slice of spacetime is the whole spacetime, the algebra of a slice will equal the algebra of the spacetime. The existence of a time-slice condition is important for making physical predictions: one can determine the state of system by examining its expectation values on all elements of the algebra. However, if one had to do this for elements of the algebra that were localised anywhere within the spacetime, then it would be totally impractical to obtain the state. Even if the time-slice condition holds, it is still impractical to know all the expectation values in an entire slice, but the time-slice condition does show that, in principle, it is sufficient to just know the expectation values at a fixed time rather than having to know them at all times.

The generalisation of the Haag-Kastler axioms to globally hyperbolic curved spacetimes was achieved by Dimock [26]. The axioms remain on the whole unaltered except for the axiom concerning Poincaré covariance, which needs to be suitably modified. In the theory of general relativity, two spacetimes are physically equivalent if there exists an isometry between them. Therefore one expects that isometric spacetimes will have algebras that are isomorphic to one another. Specifically [26, p. 220], for any isometry $i : (M, \boldsymbol{g}) \to (\tilde{M}, \tilde{\boldsymbol{g}})$ there is an isomorphism $\alpha_i : \mathcal{A} \to \tilde{\mathcal{A}}$ such that $\alpha_i(\mathcal{A}(\mathscr{O})) = \tilde{\mathcal{A}}(i(\mathscr{O}))$ with $\alpha_{Id_M} = Id_{\mathcal{A}}$ and $\alpha_{i_1 \circ i_2} = \alpha_{i_1} \circ \alpha_{i_2}$. As a consistency check, this reduces to the usual Poincaré covariance in Minkowski spacetime, as the isometries there form the Poincaré group.

The preceding formulation has since been superseded by the methods of locally covariant quantum field theory due to Brunetti, Fredenhagen and Verch (BFV) [22]. One can easily recover the earlier approaches of Haag & Kastler and Dimock within this new formulation, see [22, Sec. 2.4]. One should note that when one recovers the Haag & Kastler approach, there is an additional and physically well-motivated constraint placed upon the subregions labelling the local algebras, namely that they must be causally convex. This means that when considered as regions (spacetimes) by themselves, the causal relationships are exactly the same as when they are included in the main spacetime, which entails that one cannot introduce new nor destroy existing causal relationships. The BFV approach is to construct a theory simultaneously on all physically admissible spacetimes, that is, globally hyperbolic spacetimes, using methods from category theory. We will now briefly outline this approach in a non-categorical language. One considers two collections of objects: the set of all four-dimensional globally hyperbolic spacetimes and the set of unital C^*-algebras. However each of these sets comes with some additional structure, namely a collection of maps between the elements of the set. For the set of spacetimes, these maps are isometric embeddings that preserve time and spatial orientations as well as ensuring that their images are causally convex. By composing the maps one can successively embed a spacetime into another spacetime and then into another one whilst preserving all the desired properties, and so obtain a map from the first spacetime into the final spacetime. Also, for each spacetime, there is an identity map. The maps associated with the algebras are injective (faithful) unit-preserving $*$-homomorphisms. These maps can also be combined using composition of maps to successively embed one algebra within two other algebras. There also exists an identity map for each algebra.

Just as for Haag & Kastler, where a quantum field theory is thought of as the assignment of regions to algebras, in this instance, a locally covariant quantum field theory is a map between the sets and their maps. Specifically it assigns to each spacetime an algebra, and associates to each isometric embedding an injective unit-preserving $*$-homomorphism between the algebras of the spacetimes involved in the embeddings. Note that an identity map of a spacetime corresponds to the identity map of the corresponding algebra, and compositions of embeddings is respected by the assignment of algebra embeddings. One then imposes a causality relation, meaning a local commutativity relation, as well as a time-slice condition on this assignment to complete the theory. This is a very brief description; for the full details we refer the reader to [22, Sec. 2.1]. Note that our approach to the case of linearized gravity will not focus on the categorical formulation but instead on the actual construction of the algebra of observables for a given spacetime, though we will, in chapter 6, describe briefly how our construction fits into the BFV framework.

Once an algebra containing all of the observables of the particular theory of interest has been constructed, then one needs to consider the issues of states. This then allows for a link up with measurements and hence experimental predictions. A state ω on an algebra $\mathcal{A}$ is a linear functional $\omega : \mathcal{A} \to \mathbb{C}$ satisfying two conditions. The first condition is positivity, $\omega(A^*A) \geq 0$ for all $A \in \mathcal{A}$. The second condition is the normalization condition, $\omega(\mathbb{1}) = 1$, where $\mathbb{1}$ is the unit of $\mathcal{A}$. The physical interpretations and connections to the standard Hilbert space treatments of quantum theory arise through the Gel'fand, Naimark and Segal (GNS) construction.

We now outline this construction for the case of a $*$-algebra. For a full description, including the case of a C^*-algebra, see [52, pp. 122-124]. One begins by choosing a state ω on the algebra $\mathcal{A}$. From this, the GNS construction yields a quadruple $(\mathscr{H}_\omega, \mathscr{D}_\omega, \pi_\omega, \Omega_\omega)$, where $\mathscr{H}_\omega$ is a Hilbert space, $\mathscr{D}_\omega \subset \mathscr{H}_\omega$ is a dense subspace, π_ω is a $*$-homomorphism from the algebra to unbounded operators on $\mathscr{H}_\omega$, and $\Omega_\omega \in \mathscr{H}_\omega$ is a cyclic vector, which means that the set $\{\pi_\omega(A)\Omega_\omega \mid A \in \mathcal{A}\}$ is dense in $\mathscr{H}_\omega$, in fact this set is $\mathscr{D}_\omega$. We include subscript ω's to highlight the dependence upon the choice of state used. The GNS quadruple is unique up to unitary equivalence [5, p.38] and the physical interpretations come from, as we will see, $\omega(A)$ being equal to the expectation value of a representation of the observable A in the Hilbert space $\mathscr{H}_\omega$.

The Hilbert space $\mathscr{H}_\omega$ is constructed directly from the algebra $\mathcal{A}$ by defining the product of two elements $A, B \in \mathcal{A}$ to be $\langle A, B\rangle := \omega(A^*B)$. Using the properties of states, one can show that such a pairing satisfies all but one of the requirements of being an inner product. What prevents it from being an inner product is the possible existence of non-trivial elements with zero-norm. This can be rectified by quotienting the algebra by the subspace consisting of elements with zero-norm. The resulting space $\mathscr{D}_\omega$ consists of equivalence classes of elements of the form $[A]_\omega$, where $[A]_\omega := \{B \in \mathcal{A} \mid B = A + C, \text{where}\, \omega(C^*C) = 0\}$. This space thus has an inner product given by $\langle [A]_\omega, [B]_\omega\rangle = \omega(A^*B)$. The completion of $\mathscr{D}_\omega$ with respect to the norm induced by the inner product gives the Hilbert space $\mathscr{H}_\omega$.

One now defines a map π_ω, which allows for the elements of the algebra (observables) to be represented as operators on $\mathscr{D}_\omega \subset \mathscr{H}_\omega$. By denoting an arbitrary element $[B]_\omega \in \mathscr{D}_\omega$ by ψ, then for any $A \in \mathcal{A}$ one defines the action of $\pi_\omega(A)$ to be $\pi_\omega(A)\psi := [AB]_\omega$. The map π_ω can easily be shown to be a $*$-homomorphism. By defining $\Omega_\omega = [\mathbb{1}]_\omega$, one sees immediately from the definition of π_ω that this vector is cyclic, and that $\omega(A) = \langle \Omega_\omega, \pi_\omega(A)\Omega_\omega\rangle$. This final point gives the interpretation of $\omega(A)$ as the expectation value of the observable A in the state ω.

As one can see from the preceding discussions, the algebraic formulation cleanly separates issues relating to observables, that is, relating to the algebra from issues relating to states, meaning linear functionals on the algebra. We will now highlight these advantages even more clearly by discussing the real scalar field.

3.2 Real scalar field

We now describe how one constructs the algebra of observables for the real scalar field on an arbitrary globally hyperbolic spacetime. This issue was first considered by Dimock [26] to serve as an example of a system which satisfied his generalisations of the Haag-Kastler axioms. Our approach differs from his in that his focus is more on dealing with initial data as opposed to solutions which we focus on. The construction that we now give will set the scene for our treatment of linearized gravity in later chapters.

We consider the case of the minimally-coupled[3] scalar field $\phi \in C^\infty(M;\mathbb{R})$ obeying the Klein-Gordon equation,

$$(\Box - m^2)\phi, \tag{3.2.1}$$

where $\Box := g^{ab}\nabla_a\nabla_b$. One may also obtain the Klein-Gordon equation from an action principle. For the real scalar field, such an action is

$$S(\phi) = \frac{1}{2}\int_M (g^{ab}\nabla_a\phi\nabla_b\phi + m^2\phi^2) dvol_{\boldsymbol{g}}. \tag{3.2.2}$$

The covariant conjugate momentum, $\pi^a = \frac{\delta S}{\delta \nabla_a \phi}$, to this action is

$$\pi^a = g^{ab}\nabla_b\phi,$$

and the Euler-Lagrange equation of (3.2.2) is precisely (3.2.1).

3.2.1 Phase space

The first issue is to construct a suitable phase space and class of observables for the real scalar field. This then facilitates the attainment of a quantum theory, which will be fully discussed after the phase space is constructed.

When considering field theories, the phase space is constructed from the space of solutions to the field equation, possibly with the solutions supplemented by certain boundary conditions as well. Such an approach may be unfamiliar to the student of classical mechanics, where one considers pairs of positions and momentums as points in the phase space, and a system evolves, according to Hamilton's equations, by following a curve of such points. However, even in that instance, one could also consider the space of solutions by identifying initial data with the corresponding solution (provided that the solution is unique). This is what is done here; we could equally well have considered initial data, just as was done in [26]. Note that for linear gauge theories, like linearized gravity, things are not quite so simple, where one has uniqueness of solutions only up to pure gauge solutions. The key point about a phase space is that it is supposed to describe the state of a system, and the space of solutions will certainly achieve that goal here.

When it comes to considering the quantum theory, we would like to be able to permit smearings against complex-valued test functions. As such, the phase space that we

[3]The non minimally-coupled case has a ξR term in the equation of motion, where ξ is a real constant, and R is the Ricci scalar.

construct will consist of complex-valued solutions to the Klein-Gordon equation[4]. This space is just the complexification of the space of real-valued solutions. Using the space of complex solutions will also permit the quantization to be achieved by the functorial methods discussed in [37, Sec. 5]. Of course, the phase space of real solutions is contained within the complex one, and indeed if one approached the quantum theory using the Weyl algebra, then one would use the real phase space to do this [11, Sec 4.2].

To construct the complex phase space, we begin by considering the space of smooth spacelike-compact complex-valued solutions to (3.2.1),

$$S(M;\mathbb{C}) := \{\phi \in C^{\infty}_{SC}(M;\mathbb{C}) \mid (\Box - m^2)\phi = 0\}.$$

The reasons for the additional boundary condition of spacelike-compactness are: (i) to ensure that certain integrals, such as the yet-to-be-defined symplectic product, are well defined, and (ii) if one specifies initial data of compact support then the solution to the Klein Gordon equation will be spacelike-compact [11, Thm 3.2.11].

One should note that a standard assumption prevailing throughout the literature is to use the boundary condition: compact support on Cauchy surfaces. However, such a condition is imprecise as it is unclear whether there is a dependence on the foliation of Cauchy surfaces chosen. An example that highlights this arises on two-dimensional Minkowski spacetime, where we work with the global inertial coordinates (t, x). Consider two foliations of spacelike Cauchy surfaces, one given by the surfaces $t = \frac{x}{2} + c$ and the other given by the surfaces $t = -\frac{x}{2} + d$, where c, d are real constants that label the surfaces in each foliation. If one now considers a field whose support lies within the spacetime slab built from the surfaces in the first foliation for which $c \in (-\epsilon, \epsilon)$, where $0 < \epsilon < k$ and $\epsilon, k \in \mathbb{R}$, then such a field will not necessarily have compact support on the Cauchy surfaces in the first foliation, but it will always have compact support on the Cauchy surfaces of the second foliation.

Even if one was able to make the definition of compact support on Cauchy surfaces precise, there is still the following quandary to resolve: all spacelike-compact fields have compact intersection with any Cauchy surface [11, Cor. A.5.4], but the converse, namely that any field with compact support on Cauchy surfaces will be spacelike-compact, is not true, as the following example from [32, p. 17] shows: consider four-dimensional Minkowski spacetime and let $f \in C^{\infty}(\mathbb{R}^4;\mathbb{C})$ have support of the form: $\operatorname{supp} f = \bigcup_{n\in\mathbb{N}} \left\{\frac{1}{n}\right\} \times B_n$, where B_n is the open ball of unit radius in $\mathbb{R}^3$ centred at $(4n, 0, 0) \in \mathbb{R}^3$. In this case

[4] We caution that they are not the solutions from the theory describing a complex scalar field.

$\operatorname{supp} f \not\subset J(K)$ for any compact $K \subset \mathbb{R}^4$ but does have compact intersection with the $t = 0$ hypersurface.

However, for the case that we are considering, both the imprecision of the definition and the possibility that it would not be spacelike-compact are nullified by the properties of the field equations and global hyperbolicity. If a solution has compact support on some foliation of Cauchy surfaces, then its initial data on one of these Cauchy surfaces will also be compactly supported. Hence, by property (ii) of spacelike-compactness from above, this solution will be spacelike-compact, and consequently will have compact support on all foliations by Cauchy surfaces.

All of the preceding discussions just serve to highlight why we prefer to work with the much cleaner and precise property of spacelike-compactness. Also, the restriction to SC coupled with the Klein-Gordon differential operator being normally hyperbolic entails that, by Theorem 2.4.3, any $\phi \in S(M, \mathbb{C})$ can be written as

$$\phi = Ef, \tag{3.2.3}$$

where $E = E^- - E^+$ is the advanced-minus-retarded solution operator associated with $(\Box - m^2)$ and $f \in C_0^\infty(M; \mathbb{C})$.

To make $S(M, \mathbb{C})$ into a phase space, one needs to endow it with a symplectic product. We have seen that the equation of motion for the real scalar field can arise from an action principle (3.2.2), where the integrand would be a Lagrangian for the scalar field. The presence of a Lagrangian means that the machinery of [68] is available for use. In fact, it entails that we can endow the space of smooth spacelike-compact real-valued solutions $S(M; \mathbb{R})$ with a symplectic product; one can then extend this to $S(M; \mathbb{C})$ by linearity. Recall, that, just as in the case of classical mechanics [1, Ch. 8.1], one defines a symplectic product ω_ϕ at each point $\phi \in S(M; \mathbb{C})$ by its action on the elements of the tangent space $T_\phi S(M; \mathbb{C})$ at that point. However, since in this case $S(M; \mathbb{C})$ is a vector space, we have the identification $S(M; \mathbb{C}) \cong T_\phi S(M; \mathbb{C})$ for any $\phi \in S(M; \mathbb{C})$, which means that ω_ϕ will act on elements from $S(M; \mathbb{C})$. By using the scalar field Lagrangian, $\mathscr{L} = \frac{1}{2}(g^{ab}\nabla_a\phi\nabla_b\phi + m^2\phi^2)$, in the expression [68, eqs (2.21) & (2.23)] for the symplectic product of a Lagrangian field theory whose Lagrangian depends only on the field and its first derivatives, and then extending the result by linearity to complex solutions, the symplectic product of any two solutions $\phi_1, \phi_2 \in S(M; \mathbb{C})$ is given by the standard expression,

$$\omega_\Sigma(\phi_1, \phi_2) := \int_\Sigma (\phi_1 \pi_2 - \phi_2 \pi_1) dvol_{\boldsymbol{h}}, \tag{3.2.4}$$

where

$$\pi = -n_a \pi^a = -\nabla_{\boldsymbol{n}} \phi, \tag{3.2.5}$$

and Σ is a spacelike Cauchy surface with future-pointing unit normal vector $\boldsymbol{n}$.

As (3.2.4) stands, there is the possibility for dependence on the Cauchy surface Σ where it is evaluated. Actually, as we will now show, the symplectic product is independent of choice of Cauchy surface, due to it being defined on the space of solutions. After proving this we will henceforth drop the subscript Σ from ω. To prove the Cauchy surface independence of ω_Σ, we define, for any $\phi_1, \phi_2 \in S(M; \mathbb{C})$, a current

$$j^a(\phi_1, \phi_2) := \phi_2 \pi_1^a - \phi_1 \pi_2^a,$$

whose divergence is given by

$$\nabla_a j^a(\phi_1, \phi_2) = \phi_2 \Box \phi_1 - \phi_1 \Box \phi_2 = \phi_2(\Box - m^2)\phi_1 - \phi_1(\Box - m^2)\phi_2. \tag{3.2.6}$$

One can see, using (3.2.5), that

$$\omega_\Sigma(\phi_1, \phi_2) := \int_\Sigma n_a j^a(\phi_1, \phi_2) dvol_{\boldsymbol{h}}.$$

This now allows us to prove the result.

Lemma 3.2.1 *On the space of solutions $S(M; \mathbb{C})$, the symplectic product is independent of the choice of spacelike Cauchy surface.*

Proof. Let Σ and Σ' be two arbitrary spacelike Cauchy surfaces. Without loss of generality let $\Sigma \subset I^+(\Sigma')$. Denote by V the region bounded by these two Cauchy surfaces. Applying Gauss' Theorem to the current $j^a(\phi_1, \phi_2)$ on this region one obtains

$$\int_V \nabla_a j^a(\phi_1, \phi_2) dvol_{\boldsymbol{g}} = -\int_\Sigma n_a j^a(\phi_1, \phi_2) dvol_{\boldsymbol{h}} + \int_{\Sigma'} n_a j^a(\phi_1, \phi_2) dvol_{\boldsymbol{h}}.$$

Assuming that both ϕ_1 and ϕ_2 are solutions to the Klein Gordon equation, then by (3.2.6) the divergence of the associated current vanishes and hence we have the desired result. ∎

We have been referring to (3.2.4) as a symplectic product, even though we have not yet verified that it is weakly non-degenerate, meaning that there are no non-trivial elements of $S(M;\mathbb{C})$ that have vanishing symplectic product with all elements of $S(M;\mathbb{C})$. Such elements would be referred to as degeneracies. We now prove that ω is weakly non-degenerate.

Lemma 3.2.2 *On the space $S(M;\mathbb{C})$, the only element of $S(M;\mathbb{C})$ that has vanishing symplectic product with all elements of $S(M;\mathbb{C})$ is the trivial element.*

Proof. Assume that $\phi \in S(M;\mathbb{C})$ is a degeneracy, so $\omega(\phi,\phi') = 0$ for all $\phi' \in S(M;\mathbb{C})$. Now select an arbitrary spacelike Cauchy surface Σ with future-pointing unit normal vector $\boldsymbol{n}$. Let ϕ' be the solution with initial data $(\phi'|_\Sigma, \nabla_{\boldsymbol{n}}\phi'|_\Sigma) = (0, \phi^*|_\Sigma)$. Calculating the symplectic product of ϕ and ϕ' at Σ we find that

$$0 = \omega(\phi,\phi') = -\int_\Sigma |\phi|^2 dvol_{\boldsymbol{h}}.$$

Hence, $\phi|_\Sigma = 0$. Now take ϕ' to be the solution with initial data $(\phi'|_\Sigma, \nabla_{\boldsymbol{n}}\phi'|_\Sigma) = (\nabla_{\boldsymbol{n}}\phi^*|_\Sigma, 0)$. This gives

$$0 = \omega(\phi,\phi') = -\int_\Sigma |\nabla_{\boldsymbol{n}}\phi|^2 dvol_{\boldsymbol{h}},$$

which means that $\nabla_{\boldsymbol{n}}\phi|_\Sigma = 0$. Therefore the initial data for ϕ vanish and so $\phi = 0$ globally [11, Cor. 3.2.4]. ■

For the case of linearized gravity the proof of non-degeneracy is not so simple because one has to take into account constraints on the initial data.

As we know from (3.2.3), any spacelike-compact solution may be written in terms of the advanced-minus retarded solution operator. If one uses this in the symplectic product, then one obtains the following standard (see [92, Lem 3.2.1], for example) and useful result.

Lemma 3.2.3 *Given any $\phi \in S(M;\mathbb{C})$ and any $f \in C_0^\infty(M;\mathbb{C})$, then*

$$\omega(Ef,\phi) = -\int_M \phi f dvol_{\boldsymbol{g}},$$

where E is the advanced-minus-retarded solution operator associated with the differential operator $(\Box - m^2)$.

Proof. Expanding out the symplectic product using (3.2.4), one finds that

$$\omega(Ef,\phi) = -\int_\Sigma (\phi\nabla_{\boldsymbol{n}}Ef - Ef\nabla_{\boldsymbol{n}}\phi) dvol_{\boldsymbol{h}}. \tag{3.2.7}$$

Define the covector field $v_a := \phi\nabla_a E^+ f - E^+ f\nabla_a\phi$, whose divergence is

$$\nabla^a v_a = \phi\Box E^+ f - E^+ f\Box\phi = \phi(\Box - m^2)E^+ f - E^+ f(\Box - m^2)\phi = \phi f, \tag{3.2.8}$$

where in the final equality we have used that $(\Box - m^2)E^+ f = f$ and $(\Box - m^2)\phi = 0$.

Now select two spacelike Cauchy surfaces Σ and Σ' satisfying $\Sigma \subset I^+(\operatorname{supp} f)$ and $\Sigma' \subset I^-(\operatorname{supp} f)$. Denote the region bounded by these two Cauchy surfaces by V. Applying Gauss' Theorem to the covector $\boldsymbol{v}$ on the region V we have

$$\int_V \nabla^a v_a dvol_{\boldsymbol{g}} = -\int_\Sigma n^a v_a dvol_{\boldsymbol{h}} + \int_{\Sigma'} n^a v_a dvol_{\boldsymbol{h}}.$$

Using the definition of $\boldsymbol{v}$, the result (3.2.8), and that $E^+ f$ and its derivative vanish at Σ' by the support properties of the $E^\pm f$'s, we have

$$\int_V \phi f dvol_{\boldsymbol{g}} = \int_\Sigma (E^+ f\nabla_{\boldsymbol{n}}\phi - \phi\nabla_{\boldsymbol{n}} E^+ f) dvol_{\boldsymbol{h}}.$$

As $E^- f$ and its first derivatives vanish at Σ, one may replace E^+ by $-E$ to give

$$\int_V \phi f dvol_{\boldsymbol{g}} = \int_\Sigma (\phi\nabla_{\boldsymbol{n}} E f - E f\nabla_{\boldsymbol{n}}\phi) dvol_{\boldsymbol{h}}.$$

Combining this with (3.2.7) and using that $\int_V \phi f dvol_{\boldsymbol{g}} = \int_M \phi f dvol_{\boldsymbol{g}}$, gives the result. ■

We now shift our focus towards considering the observables of the theory. We are attempting to describe a quantum theory of the scalar field, therefore the basic observables should be the smeared quantum fields [26]. Hence, we will define classical counterparts of the smeared quantum fields, and these will constitute a minimal collection of observables needed to formulate the theory. They are minimal in the sense that there are other observables, most notably the stress-energy tensor, that will not be elements of this collection. However, when it comes to considering the quantum theory, it is possible to obtain a reasonable result for the expectation value of the stress-energy tensor by restricting the states of the theory to those that are of the Hadamard class. These states will be discussed further in section 3.3.

The classical observables are defined by associating to each $f \in C_0^\infty(M;\mathbb{C})$ an observable $F_f : S(M;\mathbb{C}) \to \mathbb{C}$, whose action is given by

$$F_f(\phi) = \int_M \phi f dvol_{\boldsymbol{g}}. \tag{3.2.9}$$

One notes immediately from Lemma 3.2.3 that

$$F_f(\phi) = -\omega(Ef, \phi). \tag{3.2.10}$$

This collection of observables satisfy a number of relations. The simplest of these relations are stated in the next theorem, where the involution map $*$ is just complex conjugation. When we consider the quantum observables, $*$ will correspond to taking the adjoint.

Theorem 3.2.4 *The observables satisfy:*

(i) Complex linearity: $F_{\alpha f+\beta f'}(\phi) = \alpha F_f(\phi) + \beta F_{f'}(\phi)$ *for all* $\alpha, \beta \in \mathbb{C}$ *and all* $f, f' \in C_0^\infty(M; \mathbb{C})$*;*

(ii) Hermiticity: $F_f(\phi)^* = F_{f^*}(\phi^*)$ *for all* $f \in C_0^\infty(M; \mathbb{C})$*.*

Proof. (i) uses linearity of the integrand and linearity of integration, and (ii) uses the properties of complex-conjugation. ■

Note that when one only considers the subspace of real solutions, the hermiticity relation reduces to $F_f(\phi)^* = F_{f^*}(\phi)$ for all $f \in C_0^\infty(M; \mathbb{C})$, and it is this form of the relation that will carry over to the quantum theory of the real scalar field.

The third relation obeyed by these observables is interpreted as stating that the Klein Gordon equation holds weakly. It arises by considering under what circumstances an observable F_f will just be the trivial observable $F_f(\phi) = 0$ for all $\phi \in S(M; \mathbb{C})$.

Theorem 3.2.5 *Given any* $f \in C_0^\infty(M; \mathbb{C})$*, then*

$$F_f(\phi) = 0$$

for all $\phi \in S(M; \mathbb{C})$ *if and only if* $f = (\Box - m^2)g$ *for some* $g \in C_0^\infty(M; \mathbb{C})$*.*

Proof. Consider the linear map F, which assigns an observable F_f to each compactly supported smooth function f. By examining the kernel of this map, we will discover which f's give trivial observables. Using (3.2.10), one can see that $F_f(\phi) = 0$ for all $\phi \in S(M; \mathbb{C})$ if and only if $\omega(Ef, \phi) = 0$ for all $\phi \in S(M; \mathbb{C})$. The non-degeneracy of ω entails that this is true if and only if $Ef = 0$. Using the exact sequence of Theorem 2.4.3, then $Ef = 0$ if and only if $f = (\Box - m^2)g$ for some $g \in C_0^\infty(M; \mathbb{C})$. ■

The final aspect of classical theory that will be examined is the one remaining relation, namely the Poisson bracket of two of the observables. We will follow the definition given in [1, p. 568]. However, this definition requires one to use the differential of an observable, but as it stands, the space $S(M;\mathbb{C})$ does not come equipped with a smooth structure that will allow for operations from calculus to be performed. This issue is not discussed in the existing literature, but there is a very elegant method due to Frölicher that allows one to endow a smooth structure to an infinite-dimensional manifold. The resulting space is called a Frölicher space, see [66, Ch. 23] for a full discussion. To make $S(M;\mathbb{C})$ into a Frölicher space, we follow the methodology set out in [66, p. 239], where one chooses a set of scalar functions on $S(M;\mathbb{C})$ to serve as a generating set and then defines a curve to be smooth if its composition with each of the generating functions is a smooth map. One then obtains the space of smooth functions by saying that a function is smooth if its composition with a smooth curve is a smooth map. For the case of $S(M;\mathbb{C})$, the symplectic product ω will be used to construct the functions that will generate the smooth structure: a curve $c : \mathbb{R} \to S(M;\mathbb{C})$ will be said to be smooth if the map $t \mapsto \omega(\phi, c(t))$ is smooth for all $\phi \in S(M;\mathbb{C})$. A function $G : S(M;\mathbb{C}) \to \mathbb{C}$ is then deemed to be smooth if $G \circ c : \mathbb{R} \to \mathbb{C}$ is a smooth function for every smooth curve c. Here the functions $\omega(\phi, \cdot)$, or equally $\omega(\cdot, \phi)$, for some $\phi \in S(M;\mathbb{C})$ are the generating set of the Frölicher space. Note that the $\omega(\phi, \cdot)$ are by definition contained within the set of smooth functions, and hence the symplectic product is smooth in both of its arguments. We know from (3.2.10) that the functions F_f can be expressed in terms of the symplectic product and so all of the F_f's that we consider here are smooth functions.

The differential of a scalar function on the Frölicher space can therefore be defined in an exactly analogous manner to the finite-dimensional case, see [83, eq. (1.3.1)], for example. Given an $F \in C^\infty(S(M;\mathbb{C}))$, one defines its differential dF pointwise by the condition $dF_\phi(V_\phi) = V_\phi(F)$ for all $V_\phi \in T_\phi S(M;\mathbb{C})$. Recall that tangent vectors are defined to be equivalence classes of curves, where two smooth curves c_1 and c_2 are deemed to be equivalent at $\phi \in S(M;\mathbb{C})$ if $\frac{d}{dt}(F \circ c_1)|_{t=0} = \frac{d}{dt}(F \circ c_2)|_{t=0}$ for all smooth functions F and the parameterisations are chosen so that $c_1(0) = c_2(0) = \phi$. As $V_\phi(F) = \frac{d}{dt}(F \circ c)|_{t=0}$, then we see that the differential of a smooth function can be defined on a Frölicher space.

With the preceding structures defined, we are now able to discuss the Poisson bracket of two observables. Following [1, pp. 566-568], we take the Poisson bracket of two smooth functions $F, G \in C^\infty(S(M;\mathbb{C}))$ to be given in terms of their exterior derivatives by

$$\{F, G\}([\boldsymbol{\gamma}]) = dF((dG)^{\sharp_\omega})|_{[\boldsymbol{\gamma}]}, \tag{3.2.11}$$

where the Hamiltionian vector field $(dG)^{\sharp_\omega}$ induced by G ($\sharp_\omega$ denotes the vector field generated using ω) satisfies

$$\omega_{[\gamma]}((dG)^{\sharp_\omega}|_{[\gamma]}, \phi) = dG|_{[\gamma]}(\phi) \tag{3.2.12}$$

for all $\phi \in T_{[\gamma]}S(M;\mathbb{C})$. We will show in the proof of Theorem 3.2.6 that, for our case, $(dG)^{\sharp_\omega}|_{[\gamma]}$ is uniquely defined by the condition (3.2.12). Here $\omega_{[\gamma]} : T_{[\gamma]}S(M;\mathbb{C}) \times T_{[\gamma]}S(M;\mathbb{C}) \to \mathbb{C}$ is the symplectic form at $[\gamma] \in S(M;\mathbb{C})$. Under the identification $T_{[\gamma]}S(M;\mathbb{C}) \cong S(M;\mathbb{C})$, which was discussed earlier when the symplectic product was defined, we can replace $\omega_{[\gamma]}$ by $\omega : S(M;\mathbb{C}) \times S(M;\mathbb{C}) \to \mathbb{C}$. This now allows for the Poisson bracket of our observables to be calculated.

Theorem 3.2.6 *The Poisson bracket of two observables $F_f, F_{f'}$, where $f, f' \in C_0^\infty(M;\mathbb{C})$, is given by*

$$\{F_f, F_{f'}\} = -E(f, f'),$$

where the bi-distribution E is defined for all $f, f' \in C_0^\infty(M;\mathbb{C})$ by

$$E(f, f') := \int_M f E f' dvol_{\boldsymbol{g}} = F_f(Ef') = -\omega(Ef, Ef'). \tag{3.2.13}$$

Proof. Using the definition of the Poisson bracket (3.2.11) & (3.2.12), we see that

$$\{F_f, F_{f'}\} = dF_f((dF_{f'})^{\sharp_\omega}), \tag{3.2.14}$$

and

$$\omega((dF_{f'})^{\sharp_\omega}, \phi) = dF_{f'}(\phi).$$

As F_f is a linear map, we have $dF_f = F_f$. Hence,

$$\omega((dF_{f'})^{\sharp_\omega}, \phi) = F_{f'}(\phi) = -\omega(Ef', \phi),$$

where we have used (3.2.10). By non-degeneracy of ω, we see that $(dF_{f'})^{\sharp_\omega} = -Ef'$. Substituting this result into (3.2.14) and again using that $dF_f = F_f$, we see that

$$\{F_f, F_{f'}\} = dF_f((dF_{f'})^{\sharp_\omega}) = -F_f(Ef').$$

Using the definition of the observables (3.2.9) and the definition of the bi-distribution (3.2.13), the right-hand side of this equation is minus $E(f, f')$. The final equality in (3.2.13) follows directly from (3.2.10). ■

3.2.2 Quantization

We now move to quantize the classical theory and so obtain a quantum description of the real-scalar field. This will be achieved using the algebraic approach discussed in section 3.1. Specifically, for each globally hyperbolic spacetime $(M, \boldsymbol{g})$ we will seek an associated algebra of observables $\mathcal{A}(M, \boldsymbol{g})$ made up of quantum analogues of the classical observables (the F_f's). A prescription for obtaining quantum observables from classical ones and with that the algebraic relations between them was provided by Dirac. His approach is outlined in his famous text [29, Ch. 22]. The basic premise is that given a collection of classical observables, then one seeks quantum observables represented by operators on a Hilbert space whose commutators are equal to i times their corresponding classical Poisson brackets. However, we will not actually seek a Hilbert space nor any operators acting on that Hilbert space. Instead we will use Dirac's commutator identity as one of several relations that will be enforced on an abstract algebra to give the final algebra of observables $\mathcal{A}(M, \boldsymbol{g})$. The other relations will be the direct analogous of their classical counterparts. We will denote the quantum observables by $\phi(f)$, where ϕ is not to be confused with a classical solution, it merely labels that this object is describing the real scalar quantum field. The complete list of relations obeyed by our quantum observables are:

(i) *Complex-linearity*: $\phi(\alpha f + \beta f') = \alpha\phi(f) + \beta\phi(f')$ for all $f, f' \in C_0^\infty(M; \mathbb{C})$ and for all $\alpha, \beta \in \mathbb{C}$;

(ii) *Hermiticity*: $\phi(f)^* = \phi(f^*)$ for all $f \in C_0^\infty(M; \mathbb{C})$;

(iii) *Field equation*: $\phi((\Box - m^2)f) = 0$ for all $f \in C_0^\infty(M; \mathbb{C})$;

(iv) *Commutator*: $[\phi(f), \phi(f')] = -iE(f, f')$ for all $f, f' \in C_0^\infty(M; \mathbb{C})$).

In the second relation, the involution $*$ is interpreted as taking the adjoint.

One might be concerned about the validity of Dirac's prescription, in particular the potential of operator ordering ambiguities arising (see [95, p. vi], for example) and also the possibility of issues pertaining to the domains of unbounded operators and with that the definition of commutators, see [80, Ch. 8] for a full discussion of issues relating to unbounded operators. Now, the classical observables that we consider are linear and do not suffer from ordering ambiguities, so that poses no obstruction, and as we are not in this instance considering representations of the observables as operators on a Hilbert space, there are no issues regarding the domain of the unbounded operators representing the observables, nor issues regarding the definition of the commutator of unbounded operators. Therefore these two potential obstructions do not cause any issues with our use of Dirac's prescription.

The construction of the algebra of observables proceeds as follows [33, Ch. 5]. We first construct the free unital $*$-algebra $\mathcal{A}$ generated by all the $\phi(f)$'s. This means that $\mathcal{A}$ consists of finite linear combinations of finite products of the $\phi(f)$'s, their adjoints $\phi(f)^*$ and the unit $\mathbb{1}$. However, as it stands, this algebra is too large, in that it does not take into account the relations (i)-(iv) obeyed by the quantum fields. Therefore there will exist elements of $\mathcal{A}$ which can be manipulated into the same form using the relations, meaning that they are actually the same observable. To remove this issue one has to enforce the relations on the algebra $\mathcal{A}$. This is achieved as follows: consider the subset $\mathcal{P} \subset \mathcal{A}$ consisting of all finite linear combinations of elements of the form ABC such that $A, C \in \mathcal{A}$ and B is one of the following

$$\begin{aligned}
&\phi(\alpha f + \beta f') - \alpha\phi(f) - \beta\phi(f') \\
&\phi(f)^* - \phi(f^*) \\
&\phi((\Box - m^2)f) \\
&\phi(f)\phi(f') - \phi(f')\phi(f) + iE(f, f')
\end{aligned}$$

for all $\alpha, \beta \in \mathbb{C}$ and $f, f' \in C_0^\infty(M; \mathbb{C})$. This subset will thus consist of elements that are to be regarded as zero, meaning that under the enforcement of the relations, they could be shown to be the same observable. The subset $\mathcal{P}$ is a linear subspace of $\mathcal{A}$. It is also invariant under the adjoint operator as one can see from the definition of $\mathcal{P}$ and the properties of the adjoint, and $\mathcal{P}$ is also, by definition, invariant under products from the left and right by elements of $\mathcal{A}$. All of these properties entail that $\mathcal{P}$ is a $*$-ideal.

The algebra of observables $\mathcal{A}(M, \boldsymbol{g})$ for the real scalar field on the globally hyperbolic spacetime $(M, \boldsymbol{g})$ is obtained by quotienting the algebra $\mathcal{A}$ by this $*$-ideal $\mathcal{P}$, so

$$\mathcal{A}(M, \boldsymbol{g}) := \mathcal{A}/\mathcal{P}.$$

For the purpose of aesthetics, we do not include equivalence class parentheses on the elements $\phi(f) \in \mathcal{A}(M, \boldsymbol{g})$. It is assumed that they are implicit. The algebraic operation of taking products carries straight over to the quotient space from $\mathcal{A}$.

We have constructed the algebra of observables for the entire spacetime manifold with elements being labelled by the test functions. One recovers the local algebras by considering the subalgebra of elements for which the support of smearing tensors labelling them are contained within a region $\mathscr{O}$ that has compact closure.

Finally, to show that $\mathcal{A}(M, \boldsymbol{g})$ obeys the time-slice condition, we will prove that any test function on the spacetime can be decomposed into a test function supported purely within a connected causally convex neighbourhood of a Cauchy surface and a function that is in the image of the Klein-Gordon operator acting on smearing functions. Combining this decomposition with the third relation of the observables, namely that $\phi((\Box - m^2)f) = 0$ for all $f \in C_0^\infty(M; \mathbb{C})$, it is clear that if one knows the algebra on the neighbourhood, then one knows it on the entire spacetime, and hence the time-slice property holds. We now prove the decomposition.

Theorem 3.2.7 *Given an arbitrary Cauchy surface Σ and a connected causally convex neighbourhood $\mathscr{N}$ of Σ, then each $f \in C_0^\infty(M; \mathbb{C})$ may be decomposed as*

$$f = \tilde{f} + (\Box - m^2)h,$$

for some $\tilde{f} \in C_0^\infty(\mathscr{N}; \mathbb{C})$ and $h \in C_0^\infty(M; \mathbb{C})$.

Proof. For the first part of this proof, namely the construction of a suitable χ^+, we utilise the ideas of [37, Lem 3.1]. Since $\mathscr{N}$ is causally convex, then it will be a globally hyperbolic subset of M. Therefore one can select two Cauchy surfaces $\Sigma^\pm$ lying to the future/past of Σ, which are both still contained within $\mathscr{N}$. Take two scalar functions $\chi^\pm \in C^\infty(M; \mathbb{C})$, which satisfy $\chi^+ = 1$ in $J^+(\Sigma^+)$, $\chi^+ = 0$ in $J^-(\Sigma^-)$, and $\chi^+ + \chi^- = 1$ globally. (Since $J^+(\Sigma^+)$ and $J^-(\Sigma^-)$ are closed and disjoint sets, then the existence of χ^+ is guaranteed by [1, Prop. 5.5.8]. One then defines χ^- by the condition $\chi^+ + \chi^- = 1$.)

Now, define

$$\tilde{f} := -P(\chi^+ E f).$$

One sees immediately from this that $\tilde{f} = P(\chi^- E f)$. As $-\chi^+ E f$ and $\chi^- E f$ solve $P(\phi) = \tilde{f}$ with past and future compact support respectively, we conclude, by Theorem 2.4.2, that $-\chi^+ E f = E^+ \tilde{f}$ and $\chi^- E f = E^- \tilde{f}$. Combining these results gives $E\tilde{f} = Ef$. Now by the exact sequence of Theorem 2.4.3, we have $E(\tilde{f} - f) = 0$ if and only if $\tilde{f} - f = P(h)$ for some $h \in C_0^\infty(M; \mathbb{C})$. ■

3.3 Hadamard states for the scalar field

As discussed earlier, a state on $\mathcal{A}(M, \boldsymbol{g})$ is a linear functional $\omega : \mathcal{A}(M, \boldsymbol{g}) \to \mathbb{C}$ satisfying $\omega(A^* A) \geq 0$ for all $A \in \mathcal{A}(M, \boldsymbol{g})$ and $\omega(\mathbb{1}) = 1$. If one were to take the entire collection of states on the algebra $\mathcal{A}(M, \boldsymbol{g})$ as being physically admissible, then one could run into serious difficulties. Notably, one might not be able to define a suitable expression for the expectation value of the stress-energy for the quantum field. The possession of such an object is crucial if one wants to explore issues relating to backreaction, that is, how the presence of the quantum field influences the background spacetime geometry. To remedy this particular situation, one can impose a requirement on the states, which ensures that physically reasonable results can be calculated. The Hadamard condition is the requirement that proves sufficient to these purposes and it has become the basic property required of a state (from a free quantum field theory) to be deemed physical.

The Hadamard condition fixes the singular behaviour of the two point function of the quantum field in a particular class of states. We will henceforth restrict attention to quasi-free states [21, p. 640], that is, states whose n-point functions vanish if n is odd, and if n is even, are completely determined by the two-point function, meaning they are made up of products of the two-point function. This restriction entails that the focus can be entirely upon the two-point function rather than being concerned about the other n-point functions.

The Hadamard condition came to prominence through attempts at defining an expectation value for stress-energy tensor through the point-splitting technique, which was first proposed by De Witt [25]. Here, instead of considering the ill-defined Wick square $\omega(\phi(x)\phi(x))$, one considers the two-point function $\omega(\phi(x)\phi(x'))$, where $x \neq x'$. The expression $\omega(\phi(x)\phi(x'))$ is a well-defined bi-distribution that is singular in the limit $x' \to x$. To resolve this and obtain something smooth, one subtracts off another bi-distribution that has been locally constructed from the geometry and the Klein-Gordon operator. Such a construction uses techniques formulated by Hadamard and hence his name is attributed to these states [92, Sec 4.6]. If one was working in Minkowski spacetime then the bi-distribution that is subtracted off corresponds to the expression $\omega(\phi(x)\phi(x'))$ in the vacuum state and so gives the standard normal ordering prescription available there. However, for many years this procedure did not have a precise and rigorous formulation. This situation was rectified by the seminal paper of Kay and Wald [65], where a fully precise and rigorous definition was supplied. Returning briefly to the discussion of the stress-energy tensor, if the two point function has this desired Hadamard behaviour then an expectation value of the stress-energy tensor in this state may be constructed [90, Sec. IV].

Soon after the definition of the Hadamard form was laid down by Kay and Wald, it was reduced into an elegant form by Radzikowski [79] who brought to fruition the pioneering work of Duistermaat and Hörmander [31]. He showed how the Hadamard condition of Kay and Wald [65, Sec. 3.3] could be characterised in terms of the two-point function's wavefront set. Recall that the wavefront set highlights the singular behaviour of a distribution and so the use of this object is well-motivated in this instance. Radzikowski's definition is now the modern and well-established definition of Hadamard states and it is the one that will be used here. The result states that a bi-distribution $W : C_0^\infty(M;\mathbb{C}) \times C_0^\infty(M;\mathbb{C}) \to \mathbb{C}$, in this instance the two-point function, which is a bi-solution to $(\Box - m^2)$ modulo smooth function is said to have Hadamard form if [79, Thm 5.1] its wavefront set is

$$\mathrm{WF}(W) = \{(x, \boldsymbol{k}; x', -\boldsymbol{k}') \in \dot{T}^*(M \times M) \mid (x, \boldsymbol{k}) \sim (x', \boldsymbol{k}') \text{ and } \boldsymbol{k} \in \overline{V}_x^+\},$$

and its antisymmetric part is given by

$$W(f, \tilde{f}) - W(\tilde{f}, f) = -iE(f, \tilde{f}),$$

for all $f, \tilde{f} \in C_0^\infty(M;\mathbb{C})$, where E was defined in (3.2.13). In the wavefront set condition, $\overline{V}_x^+$ denotes the closed future lightcone in T_x^*M, and the equivalence relation is defined by $(x, \boldsymbol{k}) \sim (x', \boldsymbol{k}')$ if and only if $\boldsymbol{k}'$ is the parallel transport of $\boldsymbol{k}$ along a null geodesic connecting x and x'. For the special case that $x = x'$, the condition reduces to $\boldsymbol{k} = \boldsymbol{k}'$ is null.

With these definitions established, a state ω is said to be a Hadamard state if and only if its two-point function is a Hadamard form bi-distribution. Though having defined what a Hadamard state is, one is still left with the question of existence of Hadamard states on general globally hyperbolic curved spacetimes. A proof of existence on such spacetimes was achieved by [47] who used a deformation argument to show that the problem reduces to finding Hadamard states on ultrastatic spacetimes, which they duly established.

Chapter 4

Linearized Gravity

As its name suggests, linearized gravity is a linear approximation to the classical theory of gravity described by general relativity. It entails fixing some known background spacetime $(M, \boldsymbol{g}_0)$ and then approximating the behaviour of other spacetimes that are in a certain sense, to be defined, close to the background. This is achieved by using quantities defined purely on the background spacetime.

By utilising an approximation to a full theory, one must always bear in mind the circumstances under which the approximation will be valid. Indeed, the domain of applicability of linearized gravity includes, for example [91, Sec. 4.4], deriving the Newtonian approximation from full general relativity and the description of gravitational waves.

The full setup of the theory will be now described. Apart from section 4.1, this chapter constitutes an expansion of chapters two & three and appendix A.2 of the paper [34] cowritten by the author with Dr C. J. Fewster.

4.1 Perturbations of spacetimes

In this section we review the geometrical setup proposed by Stewart & Walker [84] for the consideration of perturbations of spacetimes. They were partly motivated by the work of Sachs [81, pp. 556-557] who considered the issue of formulating perturbation theory in a coordinate invariant manner. The Stewart & Walker framework utilises the work of Geroch [50], which concerns how to take limits of spacetimes.

The assumptions that Stewart & Walker make [84, p. 53] are that a spacetime $(M, \boldsymbol{g})$ consists of a smooth, real, connected, Hausdorff four-dimensional manifold M together with a smooth Lorentzian metric $\boldsymbol{g}$. These assumptions coincide with ours, see section 2.2, except that we further impose that our spacetimes be orientable, time-orientable and globally hyperbolic. Note that Stewart & Walker choose the opposite signature convention to ours for their spacetime metric, but this difference has no influence on any results.

As mentioned in the preamble of this chapter, one begins with a known background spacetime $(M_0, \boldsymbol{g}_0)$, which is where all the comparisons between physical fields will be made. One then postulates that there exists a (perturbed) spacetime $(M_1, \boldsymbol{g}_1)$ and that these two spacetimes are members of a continuous (to be defined) one-parameter family of spacetimes $(M_\lambda, \boldsymbol{g}_\lambda)$ labelled by λ. To describe this situation more satisfactorily, [84] utilise the idea of [50] and introduce a five-dimensional, smooth, Hausdorff manifold N, which contains the M_λ's as smooth, properly embedded and non-intersecting, four-dimensional submanifolds. The parameter λ is a continuous real-valued function on this five manifold, and its level surfaces are the M_λ's. Having given meaning to the continuous one-parameter family of spacetimes, the issue now moves to considering how to compare the physical quantities on different spacetimes.

Although it would simplify matters greatly, one cannot just freely compare tensors on different manifolds any more than one cannot freely compare tensors from different points on the same manifold. Comparisons can only be made at one spacetime point. This is the reason behind our earlier statement that all comparisons will be made on the background spacetime. To do this requires the introduction of a suitably smooth map between the manifolds from which we can then use the pullback map to move the required tensor fields to the background for comparison. By selecting a map between the manifolds we will thus be making a decision as to when a point $p_\lambda \in M_\lambda$ is to be identified with a point $p_0 \in M_0$. However, by making such a choice, we should not be prejudicing the physics, meaning that any results should not depend upon the choice of map used. The freedom in this choice of map thus constitutes a gauge freedom.

A suitable map may be constructed by using the local flow along the integral curves of a vector field $\boldsymbol{V} \in C^\infty(T^1_0(N; \mathbb{R}))$, which is nowhere tangent to any of the M_λ's; such a vector field will be said to be transverse [84]. In order to ensure consistency, $\boldsymbol{V}$'s normalisation is chosen so that the parameter along its integral curves coincides exactly with λ. Therefore $\boldsymbol{V}$'s flow, ϕ_λ, would map points in M_0 to points in M_λ.

If we are interested in some physical quantity described by a tensor field $\boldsymbol{T}_\lambda$ on each M_λ, then it can be compared against the background value $\boldsymbol{T}_0$ as follows. Given a suitably normalised transverse vector field, using the one-parameter family of pullbacks associated with the local flow, one can define a tensor field $\phi^*_\lambda \boldsymbol{T}_\lambda$ on M_0. The pullback quantity $\phi^*_\lambda \boldsymbol{T}_\lambda$

depends smoothly on λ and so can be expanded using Taylor's Theorem as:

$$\phi_\lambda^*(\boldsymbol{T}_\lambda) = \boldsymbol{T}_0 + \lambda \pounds_{\boldsymbol{V}} \boldsymbol{T}_\lambda|_{\lambda=0} + O(\lambda^2),$$

where $\boldsymbol{V}$ is the vector field on N generating the identification map ϕ_λ. The order λ term, $\pounds_{\boldsymbol{V}} \boldsymbol{T}_\lambda|_{\lambda=0}$, is called the linear perturbation of $\boldsymbol{T}$.

One can highlight the gauge freedom that we mentioned earlier by considering two choices of transverse vector field $\boldsymbol{V}, \boldsymbol{U}$, and examining the difference between the two associated linear perturbations:

$$\pounds_{\boldsymbol{V}} \boldsymbol{T}_\lambda|_{\lambda=0} - \pounds_{\boldsymbol{U}} \boldsymbol{T}_\lambda|_{\lambda=0} = \pounds_{\boldsymbol{w}} \boldsymbol{T}_\lambda|_{\lambda=0},$$

where we have used the properties of the Lie derivative and defined, $\boldsymbol{w} := (\boldsymbol{V} - \boldsymbol{U})|_{\lambda=0} \in C^\infty(T_0^1(M_0; \mathbb{R}))$. This shows that the gauge freedom in the linear perturbations is precisely characterised by elements of the form $\pounds_{\boldsymbol{w}} \boldsymbol{T}_\lambda|_{\lambda=0}$, where $\boldsymbol{w}$ is a smooth vector field on M_0. For the case of perturbations of the spacetime metric, these pure gauge terms become $\pounds_{\boldsymbol{w}} \boldsymbol{g}$.

The spacetime metric is the dynamical object that describes the gravitational field in general relativity. By applying the above procedure to it, we can obtain a linear approximation of the perturbed spacetime metric in terms of the background spacetime metric and an object, henceforth referred to as the perturbation, defined on the background spacetime. Specifically,

$$\phi_\lambda^*(\boldsymbol{g}_\lambda) = \boldsymbol{g}_0 + \lambda \pounds_{\boldsymbol{V}} \boldsymbol{g}_\lambda|_{\lambda=0} + O(\lambda^2). \tag{4.1.1}$$

We also henceforth use a shortened notation for the perturbation, $\boldsymbol{\gamma} := \pounds_{\boldsymbol{V}} \boldsymbol{g}_\lambda|_{\lambda=0}$. Coupling this with a standard abuse of notation, namely using just $\boldsymbol{g}_\lambda$ for the pulled-back spacetime metric $\phi_\lambda^*(\boldsymbol{g}_\lambda)$, yields the well-known expansion of the spacetime metric,

$$\boldsymbol{g}_\lambda = \boldsymbol{g}_0 + \lambda \boldsymbol{\gamma} + O(\lambda^2). \tag{4.1.2}$$

The perturbation $\boldsymbol{\gamma}$ is the primary object of linearized gravity and it will appear throughout the remainder of this thesis.

For a spacetime to be deemed physically admissible in general relativity it must solve the Einstein equations. If the background and perturbed spacetime are assumed to solve the Einstein equation, then we are thus now led to consider what equation the perturbation $\boldsymbol{\gamma}$ has to obey if (4.1.2) is going to be said to approximate the metric of the perturbed spacetime. This question is answered in the next section of this chapter.

4.2 Linearized Einstein equation

As we have just discussed, it will be assumed that the spacetimes in our one-parameter family all obey the Einstein equation,

$$G_{ab} + \Lambda g_{ab} = \kappa T_{ab}, \tag{4.2.1}$$

where $G_{ab} = R_{ab} - \frac{1}{2}R g_{ab}$ is the Einstein tensor, Λ is the cosmological constant and T_{ab} is the stress tensor. We now wish to study the behaviour of the perturbed spacetime metric by using its decomposition (4.1.2) in terms of the background metric and a linear perturbation.

If one pulls back the Einstein equation from the perturbed spacetime and then expands it as a Taylor series, one obtains

$$G^0_{ab} + \Lambda g^0_{ab} + \lambda L_{ab}(\boldsymbol{\gamma}) + O(\lambda^2) = \kappa T^0_{ab} + \lambda\kappa T^{lin}_{ab} + O(\lambda^2), \tag{4.2.2}$$

where a superscript zero refers to the background quantity, $L_{ab}(\boldsymbol{\gamma})$, henceforth referred to as the linearized Einstein tensor with cosmological constant, is the linearization of the term $G_{ab} + \Lambda g_{ab}$, and T^{lin}_{ab} is the linearized stress-energy tensor. As the background spacetime solves the Einstein equation, $G^0_{ab} + \Lambda g^0_{ab} = \kappa T^0_{ab}$, then up to order λ^2 terms, we are left with the linearized Einstein equations,

$$L_{ab}(\boldsymbol{\gamma}) = \kappa T^{lin}_{ab}, \tag{4.2.3}$$

to govern the behaviour of the perturbation $\boldsymbol{\gamma}$. We will set $T^{lin}_{ab} \equiv 0$.

Therefore we are left with the linearized Einstein equation,

$$L_{ab}(\boldsymbol{\gamma}) = 0. \tag{4.2.4}$$

To calculate the left-hand side of this equation, one follows the standard method of linearizing a differential equation: postulate the existence of a one parameter family of exact solutions $\boldsymbol{g}(\lambda)$, and calculate $\frac{d}{d\lambda}\left(G_{ab}(\lambda) + \Lambda g_{ab}(\lambda)\right)|_{\lambda=0}$, which gives:

$$\begin{aligned} L_{ab}(\boldsymbol{\gamma}) = -\frac{1}{2}\nabla_a\nabla_b\gamma - \frac{1}{2}\Box\gamma_{ab} + g^{cd}\nabla_c\nabla_{(a}\gamma_{b)d} \\ - \frac{1}{2}g_{ab}\left(\nabla^c\nabla^d\gamma_{dc} - \Box\gamma - \gamma_{cd}R^{cd}\right) + \left(\Lambda - \frac{1}{2}R\right)\gamma_{ab}, \end{aligned} \tag{4.2.5}$$

where we have used that $\boldsymbol{\gamma} = \frac{d}{d\lambda}\boldsymbol{g}(\lambda)|_{\lambda=0}$. Note that just as in the previous chapter, $\Box := g^{ab}\nabla_a\nabla_b$.

We now need to consider the issue of gauge invariance of L, that is, if a perturbation $\boldsymbol{\gamma}$ solves this equation, then so should any other perturbation related to $\boldsymbol{\gamma}$ by a pure gauge term, as they would be physically equivalent. Therefore we seek to enforce the condition $L_{ab}(\pounds_{\boldsymbol{w}}\boldsymbol{g}) = 0$ for all $\boldsymbol{w} \in C^\infty(T^1_0(M;\mathbb{R})$, which will result in L being gauge invariant. One may calculate, using (4.2.5), that

$$L_{ab}(\pounds_{\boldsymbol{w}}\boldsymbol{g}) = \pounds_{\boldsymbol{w}}(G_{ab} + \Lambda g_{ab}). \tag{4.2.6}$$

The right-hand side of this vanishes for all $\boldsymbol{w}$ if and only if $G_{ab} + \Lambda g_{ab} = 0$ [84, Lem. 2.2]. Therefore we henceforth restrict attention to cosmological vacuum background spacetimes, that is, background spacetimes whose metric obeys the cosmological vacuum Einstein equation, $G_{ab} + \Lambda g_{ab} = 0$. This restriction entails that the linearized Einstein tensor with cosmological constant (4.2.5) simplifies to

$$L_{ab}(\boldsymbol{\gamma}) = -\frac{1}{2}\nabla_a\nabla_b\gamma - \frac{1}{2}\Box\gamma_{ab} + g^{cd}\nabla_c\nabla_{(a}\gamma_{b)d} - \frac{1}{2}g_{ab}\left(\nabla^c\nabla^d\gamma_{dc} - \Box\gamma - \Lambda\gamma\right) - \Lambda\gamma_{ab}. \tag{4.2.7}$$

Later it will be important, when considering the algebraic relations obeyed by the observables, to consider how L's action can be extended to include general smooth rank $(0,2)$ tensor fields. In fact, since any rank $(0,2)$ tensor field decomposes into a symmetric and an antisymmetric part, by extending L's action to antisymmetric tensors, then its action will be extended to all rank $(0,2)$ tensor fields. The extension we choose entails that when L acts on an antisymmetric tensor field then it vanishes identically and reduces back to (4.2.7) when acting on symmetric tensor fields. Explicitly, for any $\boldsymbol{f} \in C^\infty(T^0_2(M;\mathbb{R}))$, L is defined to be

$$L_{ab}(\boldsymbol{f}) = -\frac{1}{2}g_{ab}(\nabla^c\nabla^d f_{(cd)} - \Box f - \Lambda f) - \Lambda f_{(ab)} - \Box f_{(ab)} - \frac{1}{2}\nabla_a\nabla_b f + \frac{3}{2}\nabla^c\nabla_{(a}f_{bc)}. \tag{4.2.8}$$

Unfortunately, the linearized Einstein tensor with cosmological constant is a non-hyperbolic differential operator. This presents difficulties when it comes to dealing with issues pertaining to the existence of solutions to the linearized Einstein equation (4.2.4). However, in this case, it is not so troublesome because there exists a decomposition of $L_{ab}(\boldsymbol{\gamma})$ into the action of a hyperbolic differential operator acting on $\boldsymbol{\gamma}$ and the trace-reversal (to be defined shortly) of a pure gauge perturbation. As we will see in section 4.6.1, this hyperbolic differential operator will be utilised to prove the existence of solutions to the linearized Einstein equation. The decomposition requires the introduction of two operators. The first is the normally hyperbolic partial differential operator

$$P_{ab}{}^{cd} := \Box\delta^c{}_a\delta^d{}_b - 2R^c{}_{ab}{}^d. \tag{4.2.9}$$

Note that if one takes the trace of the action of P on some $\boldsymbol{f} \in C^\infty(T^a_b(M;\mathbb{R}))$, then one obtains

$$g^{ab}P_{ab}{}^{cd}f_{cd} = \Box f + 2R^{ab}f_{ab}, \tag{4.2.10}$$

where $f = f^a_a$ is the trace of $\boldsymbol{f}$.

Differential operators of the form of (4.2.9) were previously considered by Lichnerowicz [70], see equation (10.4) of that reference. He considered the construction of propagators obeying such equations, but his constructions were purely local in nature, unlike the results that we will present later. In fact the Laplacian he defined, which we denote by $\Box_L$, is related to our P via

$$P_{ab}{}^{cd}f_{cd} = -\Box_L f_{ab} + R_a{}^c f_{cb} + R_b{}^c f_{ac}.$$

The second operation that is required is the trace-reversal of a perturbation. Given a $\boldsymbol{\gamma} \in C^\infty(T^0_2(M;\mathbb{R}))$, the trace-reversal of $\boldsymbol{\gamma}$, denoted by $\overline{\boldsymbol{\gamma}}$, is defined as

$$\overline{\gamma}_{ab} := \gamma_{ab} - \frac{1}{2}g_{ab}\gamma,$$

and it satisfies the conditions $\overline{\overline{\gamma}}_{ab} = \gamma_{ab}$ and $\overline{\gamma} = -\gamma$.

On the background spacetimes that we restrict to, these two operators commute as the next lemma shows.

Lemma 4.2.1 *On a cosmological vacuum background spacetime, P commutes with trace reversal. In particular, $P(\overline{\boldsymbol{\gamma}}) = 0$ if and only if $P(\boldsymbol{\gamma}) = 0$.*

Proof. Given a $\boldsymbol{f} \in C^\infty(T^0_2(M;\mathbb{R}))$, we now compute what $\overline{P(\boldsymbol{f})}$ equals:

$$\begin{aligned}\overline{P_{ab}{}^{cd}f_{cd}} = \overline{(\Box\delta^c{}_a\delta^d_b - 2R^c{}_{ab}{}^d)f_{cd}} &= (\Box\delta^c{}_a\delta^d_b - 2R^c{}_{ab}{}^d)f_{cd} - \frac{1}{2}g_{ab}(\Box f + 2R^{cd}f_{cd}) \\ &= \Box(f_{ab} - \frac{1}{2}g_{ab}f) - 2R^c{}_{ab}{}^d f_{cd} - g_{ab}\Lambda f,\end{aligned}$$

where we have used (4.2.10). Now, $g_{cd}R^{c}{}_{ab}{}^{d}f = -R_{ab}f = -\Lambda g_{ab}f$ and therefore

$$\begin{aligned}\overline{P_{ab}{}^{cd}f_{cd}} &= \Box(f_{ab} - \frac{1}{2}g_{ab}f) - 2R^{c}{}_{ab}{}^{d}f_{cd} + g_{cd}R^{c}{}_{ab}{}^{d}f \\ &= \Box\overline{f}_{ab} - 2R^{c}{}_{ab}{}^{d}(f_{cd} - \frac{1}{2}g_{cd}f) = \Box\overline{f}_{ab} - 2R^{c}{}_{ab}{}^{d}\overline{f}_{cd} = P_{ab}{}^{cd}\overline{f}_{cd}.\end{aligned}$$

■

We now move to consider the decomposition of L, which is analogous to the well-known result from electromagnetism, where the Maxwell equation for the vector potential is equal to a hyperbolic differential equation plus a pure gauge term. Specifically, using the differential forms notation of section 2.5, if $\boldsymbol{A}$ is the vector potential then the source-free Maxwell equation takes the form, $-\delta d\boldsymbol{A} = 0$. The Laplace-Beltrami operator $\tilde{\Box}\boldsymbol{A} = -(\delta d + d\delta)\boldsymbol{A}$ is a normally hyperbolic differential operator and it is immediately clear from its definition that

$$-\delta d\boldsymbol{A} = \tilde{\Box}\boldsymbol{A} + d\delta\boldsymbol{A}, \tag{4.2.11}$$

which shows how the Maxwell equation is equal to the action of a hyperbolic differential operator on the vector potential, and a pure gauge term. The version for the linearized Einstein tensor with cosmological constant is now proven. One should note the prominence of the trace-reversal in this decomposition. In fact, it will have a prevalent role throughout our discussions. Both the Maxwell and linearized gravity decompositions have been subsequently discussed by Hack and Schenkel [55].

Theorem 4.2.2 *For any $\boldsymbol{\gamma} \in C^\infty(S^0_2(M;\mathbb{R}))$, on a cosmological vacuum background space-time,*

$$2L_{ab}(\boldsymbol{\gamma}) = -P_{ab}{}^{cd}\overline{\gamma}_{cd} + (\overline{\pounds_{(\nabla\cdot\overline{\gamma})^\sharp}\boldsymbol{g}})_{ab} \tag{4.2.12}$$

or equivalently

$$2\overline{L_{ab}(\boldsymbol{\gamma})} = -P_{ab}{}^{cd}\gamma_{ab} + (\pounds_{(\nabla\cdot\overline{\gamma})^\sharp}\boldsymbol{g})_{ab}. \tag{4.2.13}$$

Proof. The Lie derivative term is

$$(\overline{\pounds_{(\nabla\cdot\overline{\gamma})^\sharp}\boldsymbol{g}})_{ab} = \nabla_a\nabla^c\overline{\gamma}_{cb} + \nabla_b\nabla^c\overline{\gamma}_{ca} - g_{ab}\nabla^c\nabla^d\overline{\gamma}_{dc},$$

which upon expanding $\overline{\boldsymbol{\gamma}}$ becomes

$$
\begin{aligned}
(\overline{\pounds_{(\nabla\cdot\overline{\gamma})^\sharp}\boldsymbol{g}})_{ab} &= \nabla_a\nabla^c\gamma_{cb} - \frac{1}{2}\nabla_a\nabla_b\gamma + \nabla_b\nabla^c\gamma_{ca} - \frac{1}{2}\nabla_b\nabla_a\gamma - g_{ab}\nabla^c\nabla^d\gamma_{dc} + \frac{1}{2}g_{ab}g_{dc}\nabla^c\nabla^d\gamma \\
&= \nabla_a\nabla^c\gamma_{cb} + \nabla_b\nabla^c\gamma_{ca} - \nabla_a\nabla_b\gamma - g_{ab}(\nabla^d\nabla^c\gamma_{cd} - \frac{1}{2}\Box\gamma), \qquad (4.2.14)
\end{aligned}
$$

where we have used the result for a torsion-free connection that covariant derivatives commute when acting on scalar functions, that is, $\nabla_a\nabla_b\gamma = \nabla_b\nabla_a\gamma$. The first and second terms on the right-hand side of (4.2.14) may be rearranged as follows. (We show this for the first term only as the second term is identical except for interchange of the indices a and b.) Utilising the Riemann tensor identity (2.1.1) and that $R_{ab} = \Lambda g_{ab}$ we have

$$
\begin{aligned}
\nabla_a\nabla^c\gamma_{bc} = g^{cd}\nabla_a\nabla_d\gamma_{bc} &= g^{cd}(\nabla_d\nabla_a\gamma_{bc} + R_{adb}{}^e\gamma_{ec} + R_{adc}{}^e\gamma_{be}) \\
&= \nabla^c\nabla_a\gamma_{bc} + R_a{}^c{}_b{}^e\gamma_{ec} - R_a{}^e\gamma_{be} \\
&= \nabla^c\nabla_a\gamma_{bc} - R^c{}_{ab}{}^e\gamma_{ec} - \Lambda\gamma_{ba}.
\end{aligned}
$$

Therefore

$$
(\overline{\pounds_{(\nabla\cdot\overline{\gamma})^\sharp}\boldsymbol{g}})_{ab} = \nabla^c\nabla_a\gamma_{bc} + \nabla^c\nabla_b\gamma_{ac} - 2\Lambda\gamma_{ab} - 2R^c{}_{ab}{}^d\gamma_{cd} - \nabla_a\nabla_b\gamma - g_{ab}(\nabla^d\nabla^c\gamma_{cd} - \frac{1}{2}\Box\gamma). \qquad (4.2.15)
$$

The $P(\overline{\boldsymbol{\gamma}})$ term is

$$
-P_{ab}{}^{cd}\overline{\gamma}_{cd} = -\Box\gamma_{ab} + \frac{1}{2}g_{ab}\Box\gamma + 2R^c{}_{ab}{}^d\gamma_{cd} + \Lambda g_{ab}\gamma. \qquad (4.2.16)
$$

The first identity (4.2.12) is obtained by combining (4.2.15) and (4.2.16), and then comparing this with (4.2.7). The second identity (4.2.13) follows from the first by using Lemma 4.2.1 and that trace-reversal is an involution. ■

We now discuss two identities that will both be utilised in section 4.7 when we consider Green's operators. The first identity shows the result of taking the divergence of $P(\boldsymbol{\gamma})$.

Lemma 4.2.3 *For any $\boldsymbol{\gamma} \in C^\infty(T^0_2(M;\mathbb{R}))$, on a cosmological vacuum background spacetime,*

$$
\nabla^a(P_{ab}{}^{cd}\gamma_{cd}) = (\Box + \Lambda)\nabla^a\gamma_{ab}. \qquad (4.2.17)
$$

Proof. Expanding out the left-hand side of (4.2.17) gives

$$
\nabla^a(P_{ab}{}^{cd}\gamma_{cd}) = \nabla^a\Box\gamma_{ab} - 2(\nabla^a R^c{}_{ab}{}^d)\gamma_{cd} - 2R^c{}_{ab}{}^d\nabla^a\gamma_{cd}. \qquad (4.2.18)
$$

On cosmological vacuum spacetimes we have that $R_{ab} = \Lambda g_{ab}$ and so, in this case, the contracted Bianchi identity, $\nabla_a R_{bcd}{}^a + \nabla_b R_{cd} - \nabla_c R_{bd} = 0$, reduces to $\nabla_a R_{bcd}{}^a = 0$. Hence (4.2.18) becomes

$$\nabla^a (P_{ab}{}^{cd}\gamma_{cd}) = \nabla^a \Box\gamma_{ab} - 2R^c{}_{ab}{}^d \nabla^a \gamma_{cd}. \tag{4.2.19}$$

We now re-express the first term on the right-hand side of this equation,

$$\begin{aligned}\nabla^a \Box\gamma_{ab} &= g^{ac} g^{de} \nabla_c \nabla_d \nabla_e \gamma_{ab} \\ &= g^{ac} g^{de} (\nabla_d \nabla_c \nabla_e \gamma_{ab} + R_{cde}{}^f \nabla_f \gamma_{ab} + R_{cda}{}^f \nabla_e \gamma_{fb} + R_{cdb}{}^f \nabla_e \gamma_{af}),\end{aligned} \tag{4.2.20}$$

where in the second line we have used the Riemann tensor identity (2.1.1). The first term on the right-hand side of (4.2.20) can again be manipulated using the Riemann tensor to give

$$\begin{aligned}\nabla_d (\nabla_c \nabla_e \gamma_{ab}) &= \nabla_d (\nabla_e \nabla_c \gamma_{ab} + R_{cea}{}^h \gamma_{hb} + R_{ceb}{}^h \gamma_{ah}) \\ &= \nabla_d \nabla_e \nabla_c \gamma_{ab} + (\nabla_d R_{cea}{}^h)\gamma_{hb} + R_{cea}{}^h \nabla_d \gamma_{hb} + (\nabla_d R_{ceb}{}^h)\gamma_{ah} + R_{ceb}{}^h \nabla_d \gamma_{ah}.\end{aligned} \tag{4.2.21}$$

Substituting (4.2.21) back into (4.2.20) and performing the contractions over the indices a & c and d & e one finds that

$$\begin{aligned}\nabla^a \Box\gamma_{ab} = \Box\nabla^a \gamma_{ab} &+ (\nabla^e R_e{}^h)\gamma_{hb} + R^{dh} \nabla_d \gamma_{hb} + (\nabla^e R^a{}_{eb}{}^h)\gamma_{ah} \\ &+ R^{ad}{}_b{}^h \nabla_d \gamma_{ah} - R^{af} \nabla_f \gamma_{ab} + R^{ef} \nabla_e \gamma_{fb} + R^{ae}{}_b{}^f \nabla_e \gamma_{af}.\end{aligned}$$

The second and fourth terms on the right-hand side vanish whilst the third and sixth terms cancel each other. Finally, substituting $R_{ab} = \Lambda g_{ab}$ and relabelling indices gives

$$\nabla^a \Box\gamma_{ab} = \Box\nabla^a \gamma_{ab} + \Lambda \nabla^a \gamma_{ab} + 2R^{ad}{}_b{}^c \nabla_d \gamma_{ac}. \tag{4.2.22}$$

Combining (4.2.19) and (4.2.22) yields the desired result. ■

The second identity considers the action of P on pure gauge perturbations.

Lemma 4.2.4 *Given a* $\boldsymbol{w} \in C^\infty(T^1_0(M;\mathbb{R}))$ *on a cosmological vacuum background spacetime, then*

$$P(\pounds_{\boldsymbol{w}} \boldsymbol{g}) = \pounds_{(\Box+\Lambda)\boldsymbol{w}} \boldsymbol{g}. \tag{4.2.23}$$

Proof. Expanding out the right-hand side of (4.2.23) gives

$$(\pounds_{(\Box+\Lambda)\boldsymbol{w}}\boldsymbol{g})_{ab} = \nabla_a(\Box + \Lambda)w_b + \nabla_b(\Box + \Lambda)w_a = 2\nabla_{(a}\Box w_{b)} + 2\Lambda\nabla_{(a}w_{b)}. \tag{4.2.24}$$

Using the Riemann tensor identity (2.1.1) we have

$$\begin{aligned}\nabla_a\Box w_b &= g^{cd}\nabla_a\nabla_c\nabla_d w_b = g^{cd}(\nabla_c\nabla_a\nabla_d w_b + R_{acd}{}^e\nabla_e w_b + R_{acb}{}^e\nabla_d w_e)\\ &= g^{cd}(\nabla_c\nabla_d\nabla_a w_b + (\nabla_c R_{adb}{}^f)w_f + R_{adb}{}^f\nabla_c w_f + R_{acd}{}^e\nabla_e w_b + R_{acb}{}^e\nabla_d w_e),\end{aligned}$$

whereupon substituting $R_{ab} = \Lambda g_{ab}$ gives

$$\nabla_a\Box w_b = \Box(\nabla_a w_b) - \Lambda\nabla_a w_b + 2R_{a\ b}^{\ c\ d}\nabla_c w_d + w_d\nabla^c R_{acb}{}^d.$$

We know from the proof of Lemma 4.2.3 that $\nabla_a R_{bcd}{}^a = 0$ on cosmological vacuum background spacetimes, hence

$$\nabla_a\Box w_b = \Box(\nabla_a w_b) - \Lambda\nabla_a w_b + 2R_{a\ b}^{\ c\ d}\nabla_c w_d.$$

Substituting this and the interchanged indices version into (4.2.24) gives the result (4.2.23). ■

One may also obtain the linearized Einstein equation (4.2.4) by performing a second order expansion, in terms of the metric perturbation $\boldsymbol{\gamma}$, of the Einstein-Hilbert action

$$S = \int_M (R - 2\Lambda)dvol_{\boldsymbol{g}}, \tag{4.2.25}$$

for a cosmological vacuum background spacetime. The first-order terms in such an expansion contribute a total divergence and the zeroth-order terms make up the Lagrangian of the background spacetime. It is the quadratic part of the Lagrangian that we seek and this is given by

$$\mathscr{L} = T^{abcdef}\nabla_a\gamma_{bc}\nabla_d\gamma_{ef} + S^{abcd}\gamma_{ab}\gamma_{cd} \tag{4.2.26}$$

with

$$T^{abcdef} = \frac{1}{4}(g^{ad}g^{bc}g^{ef} + g^{af}g^{d(b}g^{c)e} + g^{d(b}g^{c)f}g^{ae} - g^{ad}g^{e(b}g^{c)f} - g^{a(e}g^{f)d}g^{bc} - g^{d(b}g^{c)a}g^{ef})$$

and

$$S^{abcd} = \frac{\Lambda}{4} g^{ac} g^{bd} + \frac{\Lambda}{4} g^{bc} g^{ad} - \frac{\Lambda}{4} g^{ab} g^{cd}.$$

These two tensors possess the following symmetries. T^{abcdef} is symmetric on interchange of the indices b with c, e with f, and abc with def. S^{abcd} is symmetric on interchange of the indices a with b, c with d, and ab with cd.

To obtain the Euler-Lagrange equation, one varies the Lagrangian (the action) with respect to γ_{ab} and $\nabla_c \gamma_{ab}$. The variation with respect to $\nabla_c \gamma_{ab}$ will be called the covariant conjugate momentum $\mathbf{\Pi}$ and explicitly it is given by

$$\Pi^{cab} = -\frac{1}{2} \nabla^c \gamma^{ab} + \frac{1}{2} g^{ab} \nabla^c \gamma - \frac{1}{2} g^{ab} \nabla_d \gamma^{cd} - \frac{1}{4} g^{cb} \nabla^a \gamma - \frac{1}{4} g^{ca} \nabla^b \gamma + \frac{1}{2} \nabla^a \gamma^{cb} + \frac{1}{2} \nabla^b \gamma^{ca}. \quad (4.2.27)$$

The Euler-Lagrange equations then give

$$\nabla_c \Pi^{cab} - 2 S^{abcd} \gamma_{cd} = 0. \quad (4.2.28)$$

The next lemma establishes that the left-hand side of this equation is always equal to the linearized Einstein tensor with cosmological constant, and so (4.2.28) is just the linearized Einstein equation.

Lemma 4.2.5 *On a cosmological vacuum background spacetime, for $\boldsymbol{\gamma} \in C^\infty(S^0_2(M;\mathbb{R}))$ the following equality holds*

$$L^{ab}(\boldsymbol{\gamma}) = \nabla_c \Pi^{cab} - 2 S^{abcd} \gamma_{cd}.$$

Proof. Explicitly, the expressions on the right-hand side of this equation are

$$\nabla_c \Pi^{cab} = -\frac{1}{2} \Box \gamma^{ab} + \frac{1}{2} g_{ab} \Box \gamma - \frac{1}{2} g^{ab} \nabla_c \nabla_d \gamma^{cd} - \frac{1}{2} \nabla^a \nabla^b \gamma + \frac{1}{2} \nabla_c \nabla^a \gamma^{cb} + \frac{1}{2} \nabla_c \nabla^b \gamma^{ca} \quad (4.2.29)$$

and

$$\begin{aligned} 2 S^{abcd} \gamma_{cd} &= \frac{\Lambda}{2} \gamma^{ab} + \frac{\Lambda}{2} \gamma^{ba} - \frac{\Lambda}{2} g^{ab} \gamma \\ &= \Lambda \gamma^{ab} - \frac{\Lambda}{2} g^{ab} \gamma. \end{aligned} \quad (4.2.30)$$

Combining (4.2.29) and (4.2.30) in $\nabla_c \Pi^{cab} - 2 S^{abcd} \gamma_{cd}$, and then comparing this to (4.2.7) gives the result. ■

Henceforth we will restrict attention to perturbations with spacelike-compact support, that is, perturbations whose support is contained within $J(K)$ for some compact subset $K \subset M$. The reasons for doing this are discussed extensively, for the case of the scalar field, in section 3.2.1. As we discussed there, spacelike-compactness ensures that certain integrals will be well-defined. Also, as solutions to a homogeneous hyperbolic differential equation with compactly supported initial data will have spacelike compact support [11, Thm 3.2.11], then the choice is well-motivated for the scalar field case. Of course, for linearized gravity, the linearized Einstein equation is not hyperbolic, but we will frequently exploit its relationship to the hyperbolic differential operator P, and so this adds further reason to consider spacelike-compact perturbations. Note that the condition of compact support on Cauchy surfaces is used in the literature, see [9], for example.

We also henceforth make the following notational choices: the space of spacelike-compact symmetric rank $(0, 2)$ tensors are denoted by

$$\mathscr{T}(M;\mathbb{R}) = C^\infty_{SC}(S^0_2(M;\mathbb{R})),$$

whilst the subspace of this space that consists of solutions to the linearized Einstein equation is denoted by

$$\mathscr{S}(M;\mathbb{R}) = \{\boldsymbol{\gamma} \in \mathscr{T}(M;\mathbb{R}) \mid L_{ab}(\boldsymbol{\gamma}) = 0\}.$$

4.3 Linearization instabilities

The way in which we have set up the linearized Einstein system, one might be led to believe that once one has solved the linearized Einstein equation then one has a good approximation to a solution to the full Einstein equation. However, as was first pointed out by [20] for the case of a spacetime with topology $\mathbb{R} \times T^3$, there exist spurious solutions to the linearized Einstein equation that will not be tangent to a curve of exact solutions to the Einstein equation. Hence, these spurious solutions should not be deemed physically admissible. A spacetime is said to be linearization stable if and only if there are no spurious solutions to the linearized Einstein equations [73, p. 493]. It was shown by Moncrief [73] that spacetimes admitting a compact Cauchy surface and solving the vacuum Einstein equation with no cosmological constant are linearization stable if and only if they contain no global Killing vector fields. He subsequently showed in [74] that in a spacetime solving the vacuum Einstein equation, possessing a compact Cauchy surface and non-trivial global Killing vector fields, a necessary condition for a solution to the linearized Einstein equation to approximate a curve of exact solutions is that the Taub conserved quantity associated to each Killing vector field must vanish. The Taub conserved quantities are defined relative to a spacelike Cauchy hypersurface by [74, eq. (3,19)]:

$$T_{\boldsymbol{X}}(\boldsymbol{\gamma}) := \int_{\Sigma} (X^a n^b L^2_{ab}(\boldsymbol{\gamma}, \boldsymbol{\gamma})) dvol_{\boldsymbol{h}},$$

where $\boldsymbol{X}$ is the Killing vector field, $\boldsymbol{n}$ is the future-pointing unit normal vector to Σ and $L^2_{ab}(\boldsymbol{\gamma}, \boldsymbol{\gamma})$ consists of the terms quadratic in $\boldsymbol{\gamma}$ from $\frac{d^2}{d\lambda^2}(G_{ab}(\lambda))|_{\lambda=0}$. The Taub quantities $T_{\boldsymbol{X}}(\boldsymbol{\gamma})$ are independent of the choice of Cauchy surface [74, Sec. 3] and are also gauge invariant [74, Sec. 5].

To prove all of the preceding results, Moncrief used methods from the Arnowitt-Deser-Misner formulation of general relativity. We discuss this formulation at length in section 5.2, as related results due to Moncrief are required to establish the weak non-degeneracy of the symplectic product, which we define in (5.1.1).

The preceding analysis still holds true if one includes a non-zero cosmological constant in the Einstein equations, that is, if one works, as we do, with cosmological vacuum spacetimes. One can prove this by following Moncrief's arguments from the papers [73, 74], but using the modified forms (for the $\Lambda \neq 0$ case) of the various quantities used there. The relevant quantities are available in section 5.2. Therefore linearization instability analysis holds true for all cosmological vacuum spacetimes, which admit a compact Cauchy surface.

Having established the necessary condition for a perturbation to be deemed physical, it is interesting to discover whether this condition is also sufficient. It turns out, for the case of a spacetime admitting a compact Cauchy surface and solving the vacuum Einstein equation with vanishing cosmological constant, that the vanishing of the Taub quantities is sufficient to ensure that a solution to the linearized Einstein equation approximates a curve of exact solutions. However, this was proven [6] under the restriction that one of the compact Cauchy surfaces has constant mean curvature, and at this point we are not aware of any generalisations of this result.

For the case that the spacetime only admits non-compact Cauchy surfaces then it is not so clear whether such issues also exist. As is discussed in [74, Sec. 6], non-compactness introduces certain boundary terms which entail that the correspondence between $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$ and its adjoint (see equation (5.2.6)), which is used in [74, Sec. 2], is not available in this instance. One might think that due to our restriction of spacelike-compactness, any such boundary terms would be removed, which would thus ensure that the correspondence (5.2.6) holds in the non-compact case too, and hence the issues concerning Killing vectors are also relevant for this case too. However, this is debatable and we have not yet managed to resolve it.

The linearization stability issues for spacetimes admitting a compact Cauchy surface carry over to the quantum theory [75]. However, they do not manifest themselves in the algebra of observables but rather in states of the theory. Specifically, they determine whether or not a state is physically admissible. Our focus for the forthcoming chapters is the construction of the algebra of observables and so we postpone any further discussion of these issues until we consider states in section 6.2.

4.4 Pure gauge subspaces

We will now discuss the spaces of pure gauge perturbations. Due to the restriction to spacelike-compact perturbations we are led to consider two spaces of pure gauge perturbations. Primarily we will be concerned with the space

$$\mathscr{G}(M;\mathbb{R}) := \{\pounds_{\boldsymbol{w}}\boldsymbol{g} \mid \boldsymbol{w} \in C^{\infty}_{SC}(T^1_0(M;\mathbb{R}))\},$$

which consists of pure gauge perturbations that are sourced by spacelike-compact vector fields. However, it will also be necessary, when the background spacetime does not admit a compact Cauchy surface, to consider the expanded space

$$\hat{\mathscr{G}}(M;\mathbb{R}) := \{\pounds_{\boldsymbol{w}}\boldsymbol{g} \in C^{\infty}_{SC}(S^0_2(M;\mathbb{R})) \mid \boldsymbol{w} \in C^{\infty}(T^1_0(M;\mathbb{R}))\}.$$

This consists of spacelike-compact pure gauge perturbations that are sourced by purely smooth vector fields. It is clear from the definitions that $\mathscr{G}(M;\mathbb{R}) \subseteq \hat{\mathscr{G}}(M;\mathbb{R})$, and that the two sets will be equal whenever the background spacetime admits a compact Cauchy surface. In fact, in this case, all tensor fields will be spacelike-compact as $M \subseteq J(\Sigma)$, where Σ is a compact Cauchy surface. The implications of when the two sets are not equal become most apparent when we consider the observables in section 5.4.

It is important to understand under what circumstances these two spaces are not equal. From the definitions one can see that for the spaces to differ it is necessary that the background spacetime admit vector fields which satisfy Killing's equation in regions of the form $M \setminus J(K)$ for some compact subset K. However, whilst this condition is necessary, it is not sufficient for the spaces to differ. This will be illustrated using the case that the background spacetime is Minkowski, for which the two pure gauge spaces are equal.

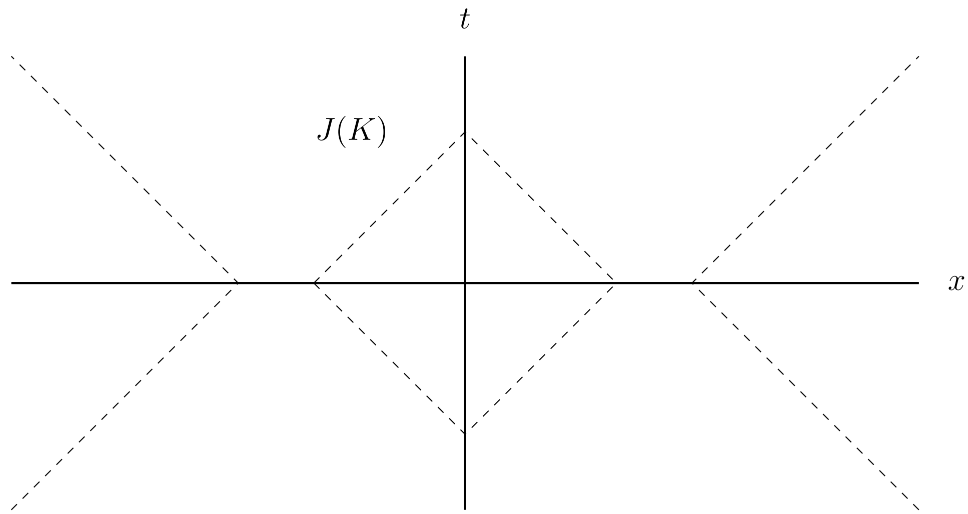

Figure 4.1: Minkowski spacetime with y, z directions supressed and $J(K)$ highlighted, where K is an annulus in the $t = 0$ plane.

To see this, let $\pounds_{\boldsymbol{w}}\boldsymbol{\eta} \in \hat{\mathscr{G}}(\mathbb{R}^4;\mathbb{R})$, hence $\boldsymbol{w} \in C^\infty(T^1_0(\mathbb{R}^4;\mathbb{R}))$ obeys Killing's equation outside of $J(K)$ for some compact $K \subset \mathbb{R}^4$. We will show that $\pounds_{\boldsymbol{w}}\boldsymbol{\eta} = \pounds_{\boldsymbol{v}}\boldsymbol{\eta}$, where $\boldsymbol{v}$ is smooth and spacelike-compact.

It will be necessary to assume that $M \setminus J(K)$ is connected. However, there do exist examples where this is not true. For example, take K as being a closed annulus in the $t = 0$ hyperplane with inner radius r_1 and outer radius r_2. The complement of $J(K)$, illustrated in figure 4.1, will consist of two disconnected regions, one that is relatively compact and the other which is not. However, by expanding K to being the closed disc of radius r_2 in the $t = 0$ hyperplane, one eliminates the relatively compact disconnected component and obtains a connected causal complement. In fact, for any compact K in Minkowski spacetime, one can find another compact set $\tilde{K}$, a closed ball of some fixed radius, such that $K \subset \tilde{K}$ and $M \setminus J(\tilde{K})$ is connected. So without loss of generality it can be assumed that $M \setminus J(K)$ is connected in Minkowski spacetime.

On a connected n-dimensional Lorentzian manifold, the collection of Killing vector fields form [77, Lem. 9.28] a finite-dimensional vector space (a Lie algebra) whose dimension is bounded by $\frac{n(n+1)}{2}$. Maximally symmetric spacetimes are those for which the dimension of the space of Killing vector fields equals the bound. Minkowski spacetime is an example of such a spacetime and the 'usual' Killing vectors (that generate the translations, rotations and boosts) form a basis of the space of Killing vector fields.

As we can assume that the region $M \setminus J(K)$ is connected, the space of Killing vector fields on it form a vector space that is at most ten-dimensional. The restriction of the global Minkowski Killing vectors to this region will thus exhaust the bound on the dimension and so the restricted global Killing vectors form a basis for the space of Killing vector fields on the region $M \setminus J(K)$. Now, by assumption, $\boldsymbol{w}|_{M\setminus J(K)}$ is a Killing vector and so can be expressed as a linear combination of the restricted global Minkowski Killing vectors. In fact, the same linear combination of global Killing vectors will again be a global Killing vector field, which will be denoted by $\boldsymbol{\xi}$, and we therefore have $\boldsymbol{\xi}|_{M\setminus J(K)} = \boldsymbol{w}|_{M\setminus J(K)}$. By subtracting the Lie-derivative of the spacetime metric with respect to $\boldsymbol{\xi}$ from $\pounds_{\boldsymbol{w}}\boldsymbol{\eta}$ we still obtain $\pounds_{\boldsymbol{w}}\boldsymbol{\eta}$ because $\boldsymbol{\xi}$ is a Killing vector, so

$$\pounds_{\boldsymbol{w}}\boldsymbol{\eta} = \pounds_{\boldsymbol{w}}\boldsymbol{\eta} - \pounds_{\boldsymbol{\xi}}\boldsymbol{\eta} = \pounds_{\boldsymbol{w}-\boldsymbol{\xi}}\boldsymbol{\eta}.$$

The vector field $\boldsymbol{w} - \boldsymbol{\xi}$ is spacelike-compact and hence $\pounds_{\boldsymbol{w}}\boldsymbol{\eta} \in \mathscr{G}(\mathbb{R}^4; \mathbb{R})$, which shows that $\mathscr{G}(M; \mathbb{R})$ and $\hat{\mathscr{G}}(M; \mathbb{R})$ agree on Minkowski spacetime.

To see how the sets $\mathscr{G}(M; \mathbb{R})$ and $\hat{\mathscr{G}}(M; \mathbb{R})$ can differ, consider Minkowski spacetime with the causal future and past of the origin removed; this is still a globally hyperbolic spacetime and it inherits all of the Killing vector fields of Minkowski spacetime. Let K be the set of points in the $t = 0$ hyperplane whose radial coordinate lies within $[R, 2R]$ for some $R > 0$. The spacetime together with K and $J(K)$ are shown in figure 4.2, where spherical polars are used for the spatial coordinates and the θ, ϕ directions have been suppressed. Note that neither of the disconnected regions of $M \setminus J(K)$ are relatively compact.

Now select an arbitrary Killing vector field $\boldsymbol{\xi}$ and let $\boldsymbol{w} = f\boldsymbol{\xi}$, where f is a real scalar function that is constant in the region $M \setminus J(K)$, where $f(t, r, \theta, \phi) = \alpha$ for $r < R$, $f(t, r, \theta, \phi) = \beta$ for $r > 2R$ and $\alpha \neq \beta$ are both constant. This gives $\pounds_{\boldsymbol{w}}\boldsymbol{\eta} = 0$ outside of $J(K)$ but now, due to the differing constants, we cannot subtract off a global Killing vector field that will kill off $\boldsymbol{w}$ in both regions simultaneously. Therefore $\pounds_{\boldsymbol{w}}\boldsymbol{\eta} \notin \mathscr{G}(M; \mathbb{R})$ and hence $\mathscr{G}(M; \mathbb{R}) \neq \hat{\mathscr{G}}(M; \mathbb{R})$.

4.5 Gauge choices

Given a perturbation $\boldsymbol{\gamma}$, its gauge equivalence class $[\boldsymbol{\gamma}]$ is defined to be

$$[\boldsymbol{\gamma}] := \{\boldsymbol{\gamma} + \pounds_{\boldsymbol{w}}\boldsymbol{g} \mid \pounds_{\boldsymbol{w}}\boldsymbol{g} \in \hat{\mathscr{G}}(M; \mathbb{R})\},$$

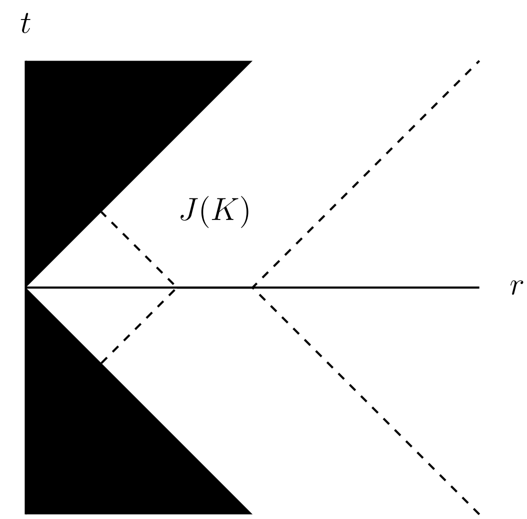

Figure 4.2: Minkowski spacetime with the causal future and past of the origin removed. Here K is again an annulus in the $t = 0$ plane and spherical polars are used as spatial coordinates.

that is, it consists of the collection of all perturbations which are physically equivalent to $\boldsymbol{\gamma}$.

By making a choice of gauge one picks out a subset of each equivalence class whose elements all obey certain desired properties. This will prove useful when considering gauge invariant objects as particular gauge choices can be used to simplify calculations whilst maintaining the physics.

The gauge will be said to be totally fixed if within each equivalence class the subset of elements obeying the desired properties consists of a single element. This means that further gauge transformations can not be made to move around in the equivalence class whilst maintaining these conditions. The gauge will be not totally fixed if the subset consists of more than one element, that is, there exists a residual gauge freedom.

In this subsection we shall be concerned with identifying when particular choices of gauge can be made, beginning with our primary choice, the de Donder gauge. All results are proven with respect to the smaller pure gauge subspace $\mathscr{G}(M;\mathbb{R})$ which entails that the results are the strongest that they can be for spacelike-compact perturbations.

4.5.1 de Donder gauge

A perturbation $\boldsymbol{\gamma}$ will be said to obey the de Donder condition if $\nabla^a \overline{\gamma}_{ab} = 0$. This condition is the analogue of the Lorenz gauge from electromagnetism.

Before proving that the de Donder gauge can be reached we first require the following result.

Lemma 4.5.1 *For any* $\boldsymbol{w} \in C^\infty(T^1_0(M;\mathbb{R}))$*, on a cosmological vacuum background spacetime, we have*

$$\nabla \cdot \overline{\pounds_{\boldsymbol{w}}\boldsymbol{g}} = (\Box + \Lambda)(\boldsymbol{w})^\flat.$$

Proof. By the definition of trace-reversal, $(\overline{\pounds_{\boldsymbol{w}}\boldsymbol{g}})_{ab} = \nabla_a w_b + \nabla_b w_a - g_{ab}\nabla_c w^c$. Taking the divergence of this gives

$$\nabla^a(\overline{\pounds_{\boldsymbol{w}}\boldsymbol{g}})_{ab} = \nabla^a\nabla_a w_b + \nabla^a\nabla_b w_a - \nabla_b\nabla^a w_a. \tag{4.5.1}$$

The second term on the right-hand side may be rearranged as

$$\nabla^a\nabla_b w_a = g^{ac}\nabla_c\nabla_b w_a = g^{ac}(\nabla_b\nabla_c w_a + R_{cba}{}^d w_d) = \nabla_b\nabla^a w_a + R_b{}^d w_d = \nabla_b\nabla^a w_a + \Lambda w_b, \tag{4.5.2}$$

where we have used the Riemann tensor identity (2.1.1) and that $R_{ab} = \Lambda g_{ab}$. Combining equations (4.5.1) and (4.5.2) gives the desired result. ■

We now prove that any spacelike-compact symmetric perturbation is gauge equivalent, by an element of $\mathscr{G}(M;\mathbb{R})$, to a spacelike-compact de Donder perturbation. Such a result was proved for the case of a spacetime that admits a compact[1] Cauchy surface (and hence just smooth tensor fields) in [41, Prop. 4.3]. It should also be noted that symmetry of the perturbation is not actually required in the proof and hence such a result also holds for general spacelike-compact perturbations.

Theorem 4.5.2 *The space* $\mathscr{T}(M;\mathbb{R})$ *may be decomposed as*

$$\mathscr{T}(M;\mathbb{R}) = \mathscr{T}^{dD}(M;\mathbb{R}) + \mathscr{G}(M;\mathbb{R}),$$

where $\mathscr{T}^{dD}(M;\mathbb{R}) = \{\boldsymbol{\gamma} \in \mathscr{T}(M;\mathbb{R}) \mid \nabla^a\overline{\gamma}_{ab} = 0\}$*. The intersection* $\mathscr{G}^{dD}(M;\mathbb{R}) = \mathscr{T}^{dD}(M;\mathbb{R}) \cap \mathscr{G}(M;\mathbb{R})$ *is given by*

$$\mathscr{G}^{dD}(M;\mathbb{R}) = \{\pounds_{\boldsymbol{w}}\boldsymbol{g} \mid \boldsymbol{w} \in C^\infty_{SC}(T^1_0(M;\mathbb{R})),\ (\Box + \Lambda)\boldsymbol{w} = 0\},$$

which specifies the residual gauge freedom that the de Donder gauge possesses.

[1] The assumption of compactness is made at the beginning of [41, Sec. 2].

Proof. Given $\boldsymbol{\gamma} \in \mathscr{T}(M;\mathbb{R})$, let $\boldsymbol{\gamma}' = \boldsymbol{\gamma} + \pounds_{\boldsymbol{w}}\boldsymbol{g}$ for an arbitrary $\boldsymbol{w} \in C^{\infty}_{SC}(T^1_0(M;\mathbb{R}))$. Taking the divergence of the trace-reversal of this gives

$$\nabla^a \overline{\gamma}'_{ab} = \nabla^a \overline{\gamma}_{ab} + \nabla^a \overline{\pounds_{\boldsymbol{w}}\boldsymbol{g}}_{ab},$$

which upon using Lemma 4.5.1 becomes

$$\nabla^a \overline{\gamma}'_{ab} = \nabla^a \overline{\gamma}_{ab} + (\Box + \Lambda) w_b.$$

Therefore $\boldsymbol{\gamma}' \in \mathscr{T}^{dD}(M;\mathbb{R})$ if and only if $\boldsymbol{w}$ obeys

$$(\Box + \Lambda) w^b = -\nabla^a \overline{\gamma}_a{}^b. \tag{4.5.3}$$

This is a hyperbolic differential equation but the source is not compactly supported on spacetime and therefore it is not possible to use [11, Thm 3.2.11] to prove that there exists a unique solution which is spacelike-compact. However, as we discussed when defining the extensions of the Green's operators to time-compact tensor fields in section 2.4.2, Corollary 5 of [10, Ch. 3] shows how the result [11, Thm 3.2.11] may be generalised to the case that both the initial data and source have no restrictions placed upon their supports. It is also shown that the solution to such a system will have support contained within $J(N)$, where N is the union of the supports of the initial data with the support of the source. For our case, the initial data for $\boldsymbol{w}$ will have compact support on Σ, in fact we assume that they vanish, and the source has spacelike-compact support within some $J(K)$ for a compact subset K. Therefore the unique solution will have spacelike-compact support since $J(J(K)) = J(K)$. Hence $\boldsymbol{\gamma}' - \boldsymbol{\gamma} \in \mathscr{G}(M;\mathbb{R})$ and so the splitting is proved.

The space $\mathscr{G}^{dD}(M;\mathbb{R})$ consists of pure gauge solutions that satisfy the de Donder condition, therefore by Lemma 4.5.1 this consists of spacelike-compact vector fields $\boldsymbol{w}$ which satisfy $(\Box + \Lambda)\boldsymbol{w} = 0$. That such pure gauge solutions constitute the residual gauge freedom in the de Donder gauge follows immediately from (4.5.3) with the assumption that $\boldsymbol{\gamma}$ is de Donder, meaning that the source vanishes identically. ■

There is a natural corollary to the previous theorem, namely that the space of solutions to the linearized Eisntein equation splits as above.

Corollary 4.5.3 *On a cosmological vacuum background spacetime, the space* $\mathscr{S}(M;\mathbb{R})$ *decomposes as*

$$\mathscr{S}(M;\mathbb{R}) = \mathscr{S}^{dD}(M;\mathbb{R}) + \mathscr{G}(M;\mathbb{R}), \tag{4.5.4}$$

where $\mathscr{S}^{dD}(M;\mathbb{R}) = \{\boldsymbol{\gamma} \in \mathscr{T}^{dD}(M;\mathbb{R}) \mid L_{ab}(\boldsymbol{\gamma}) = 0\}$ *is the space of de Donder gauge solutions. Moreover,* $\mathscr{S}^{dD}(M;\mathbb{R}) \cap \mathscr{G}(M;\mathbb{R}) = \mathscr{G}^{dD}(M;\mathbb{R})$.

Proof. By the splitting of Theorem 4.5.2, linearity of the linearized Einstein equation and that $L_{ab}(\pounds_{\boldsymbol{w}}\boldsymbol{g}) = 0$ for all pure gauge solutions, the desired splitting is achieved. ■

On inspection of Theorem 4.2.2 it is clear that for linearized gravity solutions obeying the de Donder condition $\nabla \cdot \overline{\boldsymbol{\gamma}} = 0$, on a cosmological vacuum spacetime, the linearized Einstein equation reduces to the hyperbolic differential equation

$$P_{ab}{}^{cd}\overline{\gamma}_{cd} = \Box\overline{\gamma}_{ab} - 2R^{c}{}_{ab}{}^{d}\overline{\gamma}_{cd} = 0, \tag{4.5.5}$$

or equivalently, by Lemma 4.2.1,

$$\Box\gamma_{ab} - 2R^{c}{}_{ab}{}^{d}\gamma_{cd} = 0. \tag{4.5.6}$$

In addition, we see directly from this equation and (4.2.10) that, on a cosmological vacuum spacetime, the trace of $\boldsymbol{\gamma}$ obeys

$$(\Box + 2\Lambda)\gamma = 0. \tag{4.5.7}$$

4.5.2 Transverse-traceless gauge

The transverse-traceless gauge is mentioned frequently throughout the literature and it is defined to be those de Donder perturbations whose trace vanishes, that is, a transverse-traceless perturbation $\boldsymbol{\gamma}$ obeys $\nabla^a\gamma_{ab} = 0$ and $\gamma = 0$. Unfortunately, as we will show, it is not possible to put spacelike-compact perturbations into the transverse-traceless gauge on an arbitrary cosmological vacuum background spacetime. In fact, for vacuum spacetimes with a non-vanishing cosmological constant, such as de Sitter spacetime, it turns out that the transverse-traceless gauge can be achieved for spacelike-compact perturbations, at least when the perturbation obeys the linearized Einstein equation. The obstruction to attaining the transverse-traceless gauge arises for vacuum spacetimes with a vanishing cosmological constant, in which case a de Donder solution $\boldsymbol{\gamma}$ can be put into the transverse-traceless gauge if and only if

$$\int_{\Sigma} (\nabla_{\boldsymbol{n}}\gamma) dvol_{\boldsymbol{h}} = 0$$

on some, and hence any, Cauchy surface Σ, where $\boldsymbol{n}$ is the future-pointing unit normal vector to Σ.

Having to place such a further restriction on the collection of background spacetimes available for use is undesirable and therefore we will not seek to utilise the transverse-traceless gauge in deriving various results in later chapters. In fact, as it turns out, the de Donder gauge will prove sufficient for our purposes. However, for completeness we will include a full treatment of the transverse-traceless gauge.

As we have seen in the previous section, it is always possible to gauge transform an arbitrary spacelike-compact perturbation to the de Donder gauge. Therefore, to achieve the transverse-traceless gauge, we are left with the goal of generating a gauge transformation which keeps us in the de Donder gauge and also eliminates the trace. Respectfully, these conditions are given by the equations

$$(\Box + \Lambda)\boldsymbol{w}^\flat = 0 \tag{4.5.8}$$

and

$$\nabla_a w^a = \frac{1}{2}\gamma. \tag{4.5.9}$$

This system involves a yet to be found vector $\boldsymbol{w}$ and an already specified scalar γ. Therefore, to solve this system, we can utilise the framework of differential forms and appeal to the methodology of [78, Prop. II.6] who considers solving an almost identical system for the case of electromagnetism. They prove that the existence of a solution to $\Box\boldsymbol{A} = 0$ that also obeys the Lorenz gauge $\nabla^a A_a = 0$ globally is equivalent to the existence of Cauchy data that satisfy certain constraints. For the case of the system (4.5.8) and (4.5.9) we will prove that similar constraints are required and it is these constraints which cause the obstruction to achieving the transverse-traceless gauge.

Following [78], we write the Cauchy data for $\boldsymbol{w}$ in the language of differential forms. Begin with a Cauchy surface Σ and let $i : \Sigma \to M$ be the inclusion map. Then, given any $\boldsymbol{w} \in \Omega^1(M; \mathbb{R})$, define the following forms on the Cauchy surface Σ,

$$\boldsymbol{w}_{\boldsymbol{(0)}} := i^*\boldsymbol{w} \in \Omega^1(\Sigma; \mathbb{R}) \tag{4.5.10}$$

$$\boldsymbol{w}_{\boldsymbol{(d)}} := - *_\Sigma\, i^*\, *_M\, d\boldsymbol{w} \in \Omega^1(\Sigma; \mathbb{R}) \tag{4.5.11}$$

$$w_{(\delta)} := i^*\delta\boldsymbol{w} \in \Omega^0(\Sigma; \mathbb{R}) \tag{4.5.12}$$

$$w_{(n)} := - *_\Sigma\, i^*\, *_M\, \boldsymbol{w} \in \Omega^0(\Sigma; \mathbb{R}). \tag{4.5.13}$$

The $\boldsymbol{w}_{\boldsymbol{(0)}}$ and $w_{(n)}$ together constitute the value of $\boldsymbol{w}$ on Σ, whereas $\boldsymbol{w}_{\boldsymbol{(d)}}$ and $w_{(\delta)}$ make up the value of the forward normal derivative on Σ, that is, $\nabla_{\boldsymbol{n}}\boldsymbol{w}|_\Sigma$. Hence the quantities

(4.5.10), (4.5.11), (4.5.12) and (4.5.13) form the Cauchy data for $\boldsymbol{w}$.

Similarly, for scalar functions, given any $\gamma \in \Omega^0(M;\mathbb{R})$ we define

$$\gamma_{(0)} := i^*\gamma \in \Omega^0(\Sigma;\mathbb{R}) \tag{4.5.14}$$

$$\gamma_{(d)} := - *_\Sigma i^* *_M d\gamma \in \Omega^0(\Sigma;\mathbb{R}). \tag{4.5.15}$$

Note that there is no $\gamma_{(\delta)}$ nor a $\gamma_{(n)}$ because γ is a zero form, and so $\delta\gamma \equiv 0$, and as $*_M\gamma$ is a four-form then its pullback to Σ is a four-form, so it is compelled to vanish because Σ is three-dimensional. Equation (4.5.14) is just the restriction of the scalar function γ to the surface Σ and so constitutes the initial value. The initial time-derivative is given by (4.5.15). Therefore, just as for $\boldsymbol{w}$, the zero-forms $\gamma_{(0)}$ and $\gamma_{(d)}$ constitute the Cauchy data for γ.

In forms notation, the equation of motion for the vector field sourcing a gauge transformation that keeps one in the de Donder gauge (4.5.8) is written as

$$-(\delta d + d\delta)\boldsymbol{w} + 2\Lambda\boldsymbol{w} = 0 \tag{4.5.16}$$

and the constraint to eliminate the trace (4.5.9) becomes

$$\delta\boldsymbol{w} = \frac{1}{2}\gamma. \tag{4.5.17}$$

It will be necessary to consider Green's identities for zero-forms and one-forms which obey a differential equation of the form of (4.5.16). Specifically, if $\boldsymbol{w} \in \Omega^1_{SC}(M;\mathbb{R})$ solves (4.5.16), then

$$\langle \boldsymbol{w}, \boldsymbol{f}\rangle_M = \langle \boldsymbol{w_{(0)}}, (\boldsymbol{Ef})_{\boldsymbol{(d)}}\rangle_\Sigma + \langle w_{(\delta)}, (\boldsymbol{Ef})_{(n)}\rangle_\Sigma - \langle \boldsymbol{w_{(d)}}, (\boldsymbol{Ef})_{\boldsymbol{(0)}}\rangle_\Sigma - \langle w_{(n)}, (\boldsymbol{Ef})_{(\delta)}\rangle_\Sigma, \tag{4.5.18}$$

where $\boldsymbol{f} \in \Omega^1_0(M;\mathbb{R})$ and $\boldsymbol{E}$ is the advanced-minus-retarded solution operator for the differential operator $-(\delta d + d\delta) + 2\Lambda$ acting on 1-forms (see section 2.4.2 for further details about Green's operators). The result (4.5.18) follows from the formula obtained in [78, Sec. 2.4] by a change of sign convention, though there is a sign error in equation (2.21) of [78]. Alternatively it can be obtained from the similar result in [48, Appx A] by noting that [48] uses the retarded-minus-advanced propagator, whereas we use the advanced-minus-retarded. The scalar case identity is stated in (4.5.25) below.

This puts us in a position to state and prove a theorem which shows what conditions on the Cauchy data for the vector field (covector field here) sourcing the gauge transformation are equivalent to being able to reach the transverse-traceless gauge globally on $(M, \boldsymbol{g})$.

Theorem 4.5.4 *Suppose* $\boldsymbol{w} \in \Omega^1_{SC}(M;\mathbb{R})$ *solves* $(-(\delta d + d\delta) + 2\Lambda)\boldsymbol{w} = 0$ *and* $\boldsymbol{\gamma}$ *is a de Donder solution, then* $\delta \boldsymbol{w} = \frac{1}{2}\gamma$ *if and only if* $w_{(\delta)} = \frac{1}{2}\gamma_{(0)}$ *and* $\delta \boldsymbol{w_{(d)}} + 2\Lambda w_{(n)} = \frac{1}{2}\gamma_{(d)}$.

Proof. ($\Rightarrow$) The pullback of $\delta\boldsymbol{w} = \frac{1}{2}\gamma$ to the Cauchy surface Σ gives the first constraint $w_{(\delta)} = \frac{1}{2}\gamma_{(0)}$. The second constraint is found as follows: apply $- *_\Sigma i^* *_M d$ to $\delta\boldsymbol{w} = \frac{1}{2}\gamma$ and then use (4.5.15) to obtain

$$- *_\Sigma i^* *_M d\delta \boldsymbol{w} = \frac{1}{2}\gamma_{(d)}. \tag{4.5.19}$$

Utilising the equation of motion (4.5.16) means that (4.5.19) rearranges to

$$*_\Sigma i^* *_M \delta d\boldsymbol{w} - 2\Lambda *_\Sigma i^* *_M \boldsymbol{w} = \frac{1}{2}\gamma_{(d)}. \tag{4.5.20}$$

By definition of $w_{(n)}$, see (4.5.13), the second term on the left-hand side is just $2\Lambda w_{(n)}$. The first term on the left-hand side of (4.5.20) can be re-expressed as

$$*_\Sigma i^* *_M \delta d\boldsymbol{w} = *_\Sigma i^* *_M *_M d *_M d\boldsymbol{w} = *_\Sigma i^* d *_M d\boldsymbol{w} = *_\Sigma d i^* *_M d\boldsymbol{w}, \tag{4.5.21}$$

where the first equality uses that $\delta = *d*$ on spacetime, the second equality uses that, when acting on p-forms, $(*_M)^2 = (-1)^{p+1}$, and the third equality uses that the exterior derivative d commutes with the pullback i^*. In section 2.5 we found that $(*_\Sigma)^2 = 1$ and $\delta = (-1)^p *_\Sigma d *_\Sigma$ on Σ, therefore (4.5.21) becomes

$$*_\Sigma i^* *_M \delta d\boldsymbol{w} = *_\Sigma d *_\Sigma *_\Sigma i^* *_M d\boldsymbol{w} = -\delta *_\Sigma i^* *_M d\boldsymbol{w} = \delta \boldsymbol{w}_{(d)}. \tag{4.5.22}$$

Hence (4.5.20) is $\delta\boldsymbol{w}_{(d)} + 2\Lambda w_{(n)} = \frac{1}{2}\gamma_{(d)}$.

($\Leftarrow$) To prove that a $\boldsymbol{w}$ with Cauchy data obeying the restrictions set out will in fact obey the global constraint $\delta\boldsymbol{w} = \frac{1}{2}\gamma$ on $(M, \boldsymbol{g})$, we first take an arbitrary $f \in \Omega^0_0(M;\mathbb{R})$ and compute

$$\begin{aligned}\langle \delta\boldsymbol{w}, f\rangle_M = \langle \boldsymbol{w}, df\rangle_M = \langle \boldsymbol{w_{(0)}}, (\boldsymbol{E}df)_{\boldsymbol{(d)}}\rangle_\Sigma + \langle w_{(\delta)}, (\boldsymbol{E}df)_{(n)}\rangle_\Sigma \\ - \langle \boldsymbol{w_{(d)}}, (\boldsymbol{E}df)_{\boldsymbol{(0)}}\rangle_\Sigma - \langle w_{(n)}, (\boldsymbol{E}df)_{(\delta)}\rangle_\Sigma,\end{aligned} \tag{4.5.23}$$

where the first equality uses (2.5.2) and the second equality uses the Green's identity for one-forms (4.5.18). The advanced-minus-retarded solution operator $\boldsymbol{E}$ commutes with the exterior derivative [78, Prop. 2.1], that is, $\boldsymbol{E}d = dE$, where E is the advanced-minus-retarded solution operator associated with the differential operator $-(\delta d + d\delta) + 2\Lambda$ acting on scalar functions. It is also true that $(du)_{(d)} \equiv 0$ for any scalar function $u \in \Omega^0(M;\mathbb{C})$, since $d^2 \equiv 0$. Therefore (4.5.23) reduces to

$$\langle \delta \boldsymbol{w}, f\rangle_M = \langle w_{(\delta)}, (dEf)_{(n)}\rangle_\Sigma - \langle \boldsymbol{w_{(d)}}, (dEf)_{\boldsymbol{(0)}}\rangle_\Sigma - \langle w_{(n)}, (dEf)_{(\delta)}\rangle_\Sigma.$$

For the second term on the right-hand side we can use that the pullback and the exterior derivative commute [1, Thm 6.4.4], and then (2.5.2) to obtain

$$\langle \delta \boldsymbol{w}, f\rangle_M = \langle w_{(\delta)}, (dEf)_{(n)}\rangle_\Sigma - \langle \boldsymbol{\delta w_{(d)}}, (Ef)_{(0)}\rangle_\Sigma - \langle w_{(n)}, (dEf)_{(\delta)}\rangle_\Sigma.$$

Now substitute the Cauchy data constraints to obtain

$$\langle \delta \boldsymbol{w}, f\rangle_M = \frac{1}{2}\langle \gamma_{(0)}, (dEf)_{(n)}\rangle_\Sigma - \frac{1}{2}\langle \gamma_{(d)}, (Ef)_{(0)}\rangle_\Sigma + 2\Lambda\langle w_{(n)}, (Ef)_{(0)}\rangle - \langle w_{(n)}, (dEf)_{(\delta)}\rangle_\Sigma.$$

The final term on the right-hand side can be re-expressed as follows. First, $(dEf)_{(\delta)} = (\delta dEf)_{(0)}$, and as the scalar function Ef solves $(-\delta d + 2\Lambda)Ef = 0$ we therefore have $(dEf)_{(\delta)} = (\delta dEf)_{(0)} = 2\Lambda(Ef)_{(0)}$. Hence,

$$\langle \delta \boldsymbol{w}, f\rangle_M = \frac{1}{2}\langle \gamma_{(0)}, (Ef)_{(d)}\rangle_\Sigma - \frac{1}{2}\langle \gamma_{(d)}, (Ef)_{(0)}\rangle_\Sigma, \tag{4.5.24}$$

where we have used that, in this case, $u_{(d)} = (du)_{(n)}$ for any scalar function $u \in \Omega^0(M;\mathbb{C})$.

The trace of a de Donder solution $\boldsymbol{\gamma}$ satisfies the scalar wave equation (4.5.7), which in forms notation is $(-\delta d + 2\Lambda)\gamma = 0$. The scalar Green's identity for γ is

$$\langle \gamma, f\rangle_M = \langle \gamma_{(0)}, (Ef)_{(d)}\rangle_\Sigma - \langle \gamma_{(d)}, (Ef)_{(0)}\rangle_\Sigma \tag{4.5.25}$$

for any $f \in \Omega^0_0(M;\mathbb{R})$. Comparing this with (4.5.24) gives $\langle \delta \boldsymbol{w}, f\rangle_M = \frac{1}{2}\langle \gamma, f\rangle_M$ for all $f \in \Omega^0_0(M;\mathbb{R})$, and hence $\delta \boldsymbol{w} = \frac{1}{2}\gamma$. One can prove this last statement by noting that we have $\langle (\delta \boldsymbol{w} - \frac{1}{2}\gamma), f\rangle_M = 0$ for all $f \in \Omega^0_0(M;\mathbb{R})$. Hence, if one selects an arbitrary point $p \in M$ and works in a local coordinate neighbourhood of that point, then using a bump function χ centred on p and supported within the chart, one can generate a smooth, positive, compactly supported function $\chi(\delta\boldsymbol{w} - \frac{1}{2}\gamma)$. This function will satisfy

$$\int_M (\delta\boldsymbol{w} - \frac{1}{2}\gamma)^2 \chi dvol_{\boldsymbol{g}} = 0.$$

Since the integrand is positive, it will vanish globally, and therefore $\delta\boldsymbol{w} - \frac{1}{2}\gamma = 0$ wherever $\chi \neq 0$, in particular at p itself. Since p was arbitrary, one can do this at every point of M, and so $\delta\boldsymbol{w} - \frac{1}{2}\gamma = 0$ globally on M. ■

This puts us in a position to prove the applicability of the transverse-traceless gauge. Note that this decomposition is proven only for the case of solutions, unlike the de Donder gauge decomposition in Theorem 4.5.2, due to Theorem 4.5.4 relying on the assumption that the perturbation be a de Donder linearized gravity solution.

Theorem 4.5.5 *For cosmological vacuum spacetimes with $\Lambda \neq 0$, one may perform the following decomposition of the space of spacelike-compact solutions:*

$$\mathscr{S}(M;\mathbb{R}) = \mathscr{S}^{TT}(M;\mathbb{R}) + \mathscr{G}(M;\mathbb{R}).$$

Proof of Theorem 4.5.5. We know from Corollary 4.5.3 that $\mathscr{S}(M;\mathbb{R}) = \mathscr{S}^{dD}(M;\mathbb{R}) + \mathscr{G}(M;\mathbb{R})$, therefore if we can decompose the space of de Donder solutions as $\mathscr{S}^{dD}(M;\mathbb{R}) = \mathscr{S}^{TT}(M;\mathbb{R}) + \mathscr{G}(M;\mathbb{R}) \cap \mathscr{S}^{dD}(M;\mathbb{R})$, then we can achieve the splitting. As was shown in Theorem 4.5.4, this decomposition can be made so long as the Cauchy data for the gauge vector field satsify certain constraints. Therefore the goal is to obtain such suitable Cauchy data, which as we will now see, can be done for the $\Lambda \neq 0$ background case. Given a perturbation $\boldsymbol{\gamma} \in \mathscr{S}^{dD}(M;\mathbb{R})$ on a cosmological vacuum spacetime $(M, \boldsymbol{g})$ with $\Lambda \neq 0$, the constraints of Theorem 4.5.4 are satisfied by the Cauchy data: $w_{(0)} = \frac{1}{4\Lambda} d\gamma_{(0)}$, $w_{(d)} = 0$, $w_{(n)} = \frac{1}{4\Lambda}\gamma_{(d)}$ and $w_{(\delta)} = \frac{1}{2}\gamma_{(0)}$. In fact, the solution with this data is $\boldsymbol{w} = \frac{1}{4\Lambda} d\gamma$, which corresponds to the choice made in [59, eq. (9)] for the case that the background is de Sitter spacetime. Therefore appropriate Cauchy data exist and one may gauge transform from the de Donder gauge to the transverse-traceless gauge. ■

For the $\Lambda = 0$ case, there are topological obstructions. Specifically, the second constraint of Theorem 4.5.4 reduces to $\delta\boldsymbol{w}_{(\boldsymbol{d})} = \frac{1}{2}\gamma_{(d)}$, and whether such an equation can be satisfied becomes a cohomological question: the scalar $\gamma_{(d)}$ is co-closed, $\delta\gamma_{(d)} = 0$, but is it co-exact? This means that can a suitable one-form, in this case $2\boldsymbol{w}_{(\boldsymbol{d})}$, be found whose codifferential is $\gamma_{(d)}$. Instead of working with codifferentials we can equally work with the ordinary exterior derivative operator. If one expands out $\delta\boldsymbol{w}_{(\boldsymbol{d})} = \frac{1}{2}\gamma_{(d)}$ and then applies a

Hodge star operation to both sides, then the result is

$$d(*_\Sigma \boldsymbol{w_{(d)}}) = -\frac{1}{2} *_\Sigma \gamma_{(d)},$$

where $*_\Sigma\gamma_{(d)}$ is a 3-form on Σ and hence is necessarily closed. In solving this equation there are two cases to consider, which depend upon whether or not $\boldsymbol{w}$ has compact support on Cauchy surfaces. As we wish to consider elements of $\mathscr{G}(M;\mathbb{R})$ then we require the case that $\boldsymbol{w}$ has compact support on Cauchy surfaces. This means that, from [1, Thm 7.5.19(i)], $*_\Sigma\gamma_{(d)}$ is exact if and only if

$$\int_\Sigma *_\Sigma\gamma_{(d)} = 0.$$

Now, $*_\Sigma\gamma_{(d)} = i^* *_M d\gamma$ and expanding this out we have

$$(*_M d\gamma)_{bcd} = \epsilon_{abcd}\nabla^a\gamma = -\epsilon_{abcd}n^a n^e \nabla_e\gamma + \epsilon_{abcd}h^{ae}\nabla^e\gamma, \tag{4.5.26}$$

where we have used the metric decomposition $g_{ab} = -n_a n_b + h_{ab}$ with $\boldsymbol{n}$ the future-pointing unit normal vector to Σ. To pullback (4.5.26) to Σ requires that we project down the free-indices using $h_a{}^b$, see [57, Ch. 2.7], which gives

$$h_b{}^f h_c{}^i h_d{}^j (*_M d\gamma)_{fij} = -h_b{}^f h_c{}^i h_d{}^j \epsilon_{afij} n^a n^e \nabla_e\gamma + h_b{}^f h_c{}^i h_d{}^j \epsilon_{afij} h^{ae}\nabla^e\gamma. \tag{4.5.27}$$

The second term on the right-hand side is compelled to vanish as $h_b{}^f h_c{}^i h_d{}^j \epsilon_{afij} h^{ae}$ is a four-form acting at each point purely on vectors in the three-dimensional subspace tangent to Σ. Hence the pullback is identified with the first-term, $-h_b{}^f h_c{}^i h_d{}^j \epsilon_{afij} n^a \nabla_{\boldsymbol{n}}\gamma = -dvol_{\boldsymbol{h}}\nabla_{\boldsymbol{n}}\gamma$, where the equality comes from [69, Prop. 13.24].

For the case that $\boldsymbol{w}$ is not restricted to have compact support on Cauchy surfaces and the Cauchy surface is non-compact, then [1, Thm 7.5.19(iii)] gives $H^3(\Sigma) = 0$ and so $*_\Sigma\gamma_{(d)}$ will always be exact. Hence the transverse-traceless gauge can always be attained in that instance.

4.5.3 Synchronous gauge

Given a timelike vector field $\boldsymbol{t}$, a perturbation $\boldsymbol{\gamma}$ will be said to be synchronous if $t^a\gamma_{ab} = 0$. In some treatments, for example, the case of Minkowski spacetime considered in [91, p. 80], the definition of the synchronous gauge is made in a coordinate dependent way, that is, $\boldsymbol{\gamma}$ is said to be synchronous if $\gamma_{0\mu} = 0$. One can immediately see that this is included in our definition by selecting local coordinates (t, x^i) where $\boldsymbol{t} = \left(\frac{\partial}{\partial t}\right)$, which thus gives $t^\mu = \delta^\mu{}_0$ and so $t^\mu \gamma_{\mu\nu} = \delta^\mu{}_0 \gamma_{\mu\nu} = \gamma_{0\nu}$. An example of the use of the definition of synchronous that we have adopted here is given by [59, p. 4].

In order to be able to gauge transform a perturbation $\boldsymbol{\gamma}$ to the synchronous gauge, one must solve

$$t^a \nabla_a w_b + t^a \nabla_b w_a = -t^a \gamma_{ab} \tag{4.5.28}$$

for $\boldsymbol{w}$. To solve this equation, even locally, will require us to make a specific choice of timelike vector field $\boldsymbol{t}$, which is associated with a particular type of neighbourhood of a Cauchy surface. In fact, we will find that given any Cauchy surface we can reach the synchronous gauge on such a neighbourhood of that surface. The neighbourhood is constructed using the exponential map, which [77, pp. 70-71] diffeomorphically maps an open subset O (including the zero-vector) of the tangent space at a point $p \in M$ to an open subset (containing p) of the manifold. It is achieved by mapping a vector $v \in O \subseteq T_p M$ to the point a unit-parameter distance along the geodesic with initial data p and v.

If one considers an embedded submanifold $S \subset M$ of codimension greater than or equal to one, then this surface will possess a non-trivial normal bundle consisting [77, p. 198] of vectors which are orthogonal, with respect to the spacetime metric, to all vectors tangent to the surface. By restricting the exponential map to act only on the normal bundle of S, one obtains [77, p. 199] a map called the normal exponential map, which is denoted $\exp^\perp$. For this map, all the geodesics meet S orthogonally.

A normal neighbourhood of S is [77, p. 199] a neighbourhood of S that is diffeomorphic under $\exp^\perp$ to a connected neighbourhood of the zero section in the normal bundle of S.

We now restrict attention to the case where S is a smooth spacelike Cauchy surface for $(M, \boldsymbol{g})$, and denote this Cauchy surface by Σ. Since Σ is an embedded submanifold of M, then by [77, Prop. 7.26], it will possess a normal neighbourhood. On this normal neighbourhood $\mathscr{O}$ we can construct a timelike vector field which will allow one to transform to the synchronous gauge. Specifically, as any point in $\mathscr{O}$ lies on a normal geodesic emanating from Σ, we can define a vector field on $\mathscr{O}$ by stating that its value at each point is equal to the tangent vector of the normal geodesic passing through that point. Such a vector field will be smooth, future-pointing, geodesic, hypersurface-orthogonal and will have unit-magnitude. We henceforth call this vector field the normal field of Σ in $\mathscr{O}$ and denote it by $\tilde{\boldsymbol{n}}$. Note that $\tilde{\boldsymbol{n}}|_\Sigma = \boldsymbol{n}$, where $\boldsymbol{n}$ is the future-pointing unit normal vector to Σ.

We are now in a position to state and prove the theorem that allows one to reach the synchronous gauge. However, before we do that, we will briefly consider how this result links up with the findings present in the current literature. Our theorem includes the result [9, Lem. 1.1], which shows that, given a Cauchy surface Σ with future-pointing unit normal vector $\boldsymbol{n}$, then one can make a gauge transformation to achieve the condition $n^a\gamma_{ab}|_\Sigma = 0$. Also the proof of our theorem treats in detail the solution of the equations arising in [9, Lem. 1.1].

Theorem 4.5.6 *Let Σ be a smooth spacelike Cauchy surface with future-pointing unit normal vector $\boldsymbol{n}$. Let $\mathscr{O}$ be any open normal neighbourhood of Σ, whose closure is contained in another normal neighbourhood of Σ. Then*

$$\mathscr{T}(M;\mathbb{R}) = \mathscr{T}^{synch}_{\Sigma,\mathscr{O}}(M;\mathbb{R}) + \mathscr{G}(M;\mathbb{R}),$$

where $\mathscr{T}^{synch}_{\Sigma,\mathscr{O}}(M;\mathbb{R}) = \{\boldsymbol{\gamma} \in \mathscr{T}(M;\mathbb{R}) \mid \tilde{n}^a\gamma_{ab} = 0 \text{ on } \mathscr{O}\}$ and $\tilde{\boldsymbol{n}}$ is the normal field of Σ in $\mathscr{O}$. All elements of $\mathscr{T}^{synch}_{\Sigma,\mathscr{O}}(M;\mathbb{R})$ satisfy $n^a\gamma_{ab}|_\Sigma = 0$.

Remark. Given any normal neighbourhood of Σ, we may restrict to a smaller normal neighbourhood whose closure is contained in the original. Therefore the existence of $\mathscr{O}$ in the hypothesis is not restrictive and is made to ensure that there is sufficient room to let $\boldsymbol{w}$ decay smoothly to zero outside of $\mathscr{O}$.

Proof of Theorem 4.5.6. Let $\boldsymbol{\gamma} \in \mathscr{T}(M;\mathbb{R})$ be arbitrary and let $\pounds_{\boldsymbol{w}}\boldsymbol{g} \in \mathscr{G}(M;\mathbb{R})$. Then the condition $\boldsymbol{\gamma} + \pounds_{\boldsymbol{w}}\boldsymbol{g} \in \mathscr{T}^{synch}_{\Sigma,\mathscr{O}}(M;\mathbb{R})$ requires that

$$\tilde{n}^a\nabla_a w_b + \tilde{n}^a\nabla_b w_a = -\tilde{n}^a\gamma_{ab} \tag{4.5.29}$$

hold on the normal neighbourhood $\mathscr{O}$. In order to solve (4.5.29) we decompose $\boldsymbol{w}$ using the vector field $\tilde{\boldsymbol{n}}$, obtaining $\boldsymbol{w} = -W_0\tilde{\boldsymbol{n}} + \boldsymbol{w}_\parallel$, where $W_0 = \tilde{n}_a w^a$. Now contract both sides of (4.5.29) with $\tilde{\boldsymbol{n}}$, which gives

$$2\tilde{n}^a\tilde{n}^b\nabla_a w_b = -\tilde{n}^a\tilde{n}^b\gamma_{ab}. \tag{4.5.30}$$

By the definition of $\tilde{\boldsymbol{n}}$ we have that $\tilde{n}^a\nabla_a\tilde{n}^b = 0$ on $\mathscr{O}$ and therefore (4.5.30) simplifies to

$$\nabla_{\tilde{\boldsymbol{n}}} W_0 = -\frac{1}{2}\tilde{n}^a\tilde{n}^b\gamma_{ab}. \tag{4.5.31}$$

Thus the 'time' component, W_0, of the gauge vector field must obey the simple first-order differential equation (4.5.31). The remaining components of $\boldsymbol{w}$ must also satisfy a certain first-order differential equation. By utilising that $\boldsymbol{w}_{\|} = \boldsymbol{w} + W_0\tilde{\boldsymbol{n}}$, then (4.5.29) becomes

$$-(\tilde{n}^a\nabla_a W_0)\tilde{n}_b + \tilde{n}^a\nabla_a w^{\|}_b + \nabla_b W_0 + \tilde{n}^a\nabla_b w^{\|}_a = -\tilde{n}^a\gamma_{ab}, \tag{4.5.32}$$

where we have used that $\tilde{n}^a\nabla_a\tilde{n}^b = 0$, $\tilde{n}^a\tilde{n}_a = -1$ and $\tilde{n}^a\nabla_b\tilde{n}_a = 0$. Upon substituting (4.5.31) and then rearranging one finds that $\boldsymbol{w}$ obeys

$$(\nabla_{\tilde{\boldsymbol{n}}} w_{\|})_b - w_{\|a}\nabla_b\tilde{n}^a = -\tilde{n}^a\gamma_{ab} - \nabla_b W_0 - \frac{1}{2}\tilde{n}^a\tilde{n}^c\gamma_{ac}\tilde{n}_b, \tag{4.5.33}$$

which is the first-order differential equation for $\boldsymbol{w}_{\|}$ alluded to earlier.

Having obtained the appropriate differential equations we must now prove that a solution $\boldsymbol{w}$ exists. In both cases we will choose vanishing initial data. We proceed by first obtaining a solution W_0 to (4.5.31) on $\mathscr{O}$. By definition of the normal neighbourhood $\mathscr{O}$, through each point $p \in \Sigma$ there is a unit speed geodesic $\lambda_p : I \to \mathscr{O}$, which is normal to Σ at p, with $0 \in I \subset \mathbb{R}$ and $\lambda_p(0) = p$. Along each such geodesic λ_p equation (4.5.31) is just a simple first-order differential equation and can therefore be integrated along the geodesic to give a solution

$$(W_0 \circ \lambda_p)(t) = -\frac{1}{2}\int_0^t (\tilde{n}^a\tilde{n}^b\gamma_{ab} \circ \lambda_p)(s)ds. \tag{4.5.34}$$

We now define the scalar function W_0 at any $q \in \mathscr{O}$ by

$$W_0(q) := (W_0 \circ \lambda_{p_q})(t_q), \tag{4.5.35}$$

where p_q is the point on Σ through which the unique unit speed normal geodesic λ_{p_q} passing through q intercepts Σ and t_q is the parameter distance along λ_{p_q} that q is from p_q. The scalar function (4.5.35) satisfies (4.5.31) by definition and it is smooth on $\mathscr{O}$ because whenever q is varied the values of t_q and p_q change smoothly due to the exponential map being a diffeomorphism, and $\tilde{n}^a\tilde{n}^b\gamma_{ab}$ is smooth.

To obtain the remaining part of $\boldsymbol{w}$, that is, the component $\boldsymbol{w}_{\|}$ on $\mathscr{O}$, we solve (4.5.33) locally within a neighbourhood of each normal geodesic and then patch together the results using a partition of unity.

The geodesic neighbourhoods used are constructed as follows: for each $q \in \Sigma$, let $\mathscr{N}_q \subset \Sigma$ be an open normal, with respect to the induced Riemannian metric on Σ, neighbourhood of q. Hence on $\mathscr{N}_q$ there are [57, p. 34] well-defined Riemannian normal coordinates x^i ($i = 1, 2, 3$) based at q, and basis vector fields $\boldsymbol{e}_i$ associated with the coordinates x^i. For each $q \in \Sigma$, let $\mathscr{M}_q$ be the open set of points in $\mathscr{O}$ connected to Σ by geodesics emanating normally from $\mathscr{N}_q$. As the $\mathscr{N}_q$'s cover Σ, then by the definition of $\mathscr{O}$ as a normal neighbourhood of Σ, the $\mathscr{M}_q$'s form an open cover for $\mathscr{O}$. Within each $\mathscr{M}_q$ we can introduce Gaussian normal coordinates given by: the proper time t along the geodesics, with $t = 0$ on Σ, and the Riemannian normal coordinates x^i mentioned previously. In such coordinates, (4.5.33) becomes

$$\frac{d(w_{\|})_i(t,x)}{dt} - 2\Gamma^j{}_{i0}(t,x)(w_{\|})_j(t,x) = -\gamma_{0i} - \frac{\partial W_0(t,x)}{\partial x^i}. \tag{4.5.36}$$

This system is just a matrix differential equation on some open subset of $\mathbb{R}^4$ and so can be solved using techniques from [88, Sec. 1.6]. As $(w_{\|})_0 = 0$ in these coordinates, we have thus found $\boldsymbol{w}_{\|}$ on $\mathscr{M}_q$. This process is repeated on each $\mathscr{M}_q$ for all $q \in \Sigma$. Wherever two of these neighbourhoods intersect one can compare the solution from each neighbourhood with each other. Since on the intersection of the neighbourhoods they will both solve the same inhomogeneous differential equation with vanishing initial data, their difference will solve the homogeneous version with vanishing data and so will vanish. Therefore their difference will vanish on the intersection and so they are the same solution. This ensures consistency between patches.

We now need to stitch together these results to obtain a smooth solution $\boldsymbol{w}_{\|}$ on $\mathscr{O}$. However, just any partition of unity will not suffice, it needs to be specially constructed so that the resulting $\boldsymbol{w}_{\|}$ obeys (4.5.33). As Σ is an embedded submanifold of M, it inherits the topological properties of M; in particular, it will be second-countable and Hausdorff. This means that, by [93, Thm 1.11], the open cover $\{\mathscr{N}_q \mid q \in \Sigma\}$ of Σ by normal neighbourhoods will admit a countable partition of unity $\{\chi_\lambda \mid \lambda \in I\}$ subordinate to the cover with $\operatorname{supp} \chi_\lambda$ compact for each $\lambda \in I$. Hence, for each $\lambda \in I$ there exists a $q \in \Sigma$ such that $\operatorname{supp} \chi_\lambda \subset \mathscr{N}_q$. To obtain a suitable partition of unity $\{\tilde{\chi}_\lambda \mid \lambda \in I\}$ on the $\mathscr{M}_q$'s, we propagate the partition of unity $\{\chi_\lambda \mid \lambda \in I\}$ off Σ and onto $\mathscr{O}$. This is achieved by solving $\nabla_{\tilde{\boldsymbol{n}}} \tilde{\chi}_\lambda = 0$ with $\tilde{\chi}_\lambda|_\Sigma = \chi_\lambda$, for each λ. As we have already seen for the case of W_0, this can be done by integrating the equation along the integral curves of $\tilde{\boldsymbol{n}}$.

The desired $\boldsymbol{w}_{\|}$ on $\mathscr{O}$ is obtained by stitching together, using the partition of unity $\{\tilde{\chi}_\lambda \mid \lambda \in I\}$, the various solutions $\boldsymbol{w}_{\|}^\lambda$ to (4.5.36) from each set $\mathscr{M}_q$ containing $\tilde{\chi}_\lambda$. Explicitly, $\boldsymbol{w}_{\|}$ is given by

$$\boldsymbol{w}_{\|} = \sum_\lambda \tilde{\chi}_\lambda \boldsymbol{w}_{\|}^\lambda. \tag{4.5.37}$$

This obeys (4.5.33) on $\mathscr{O}$ as $\nabla_{\tilde{\boldsymbol{n}}} \tilde{\chi}_\lambda = 0$ there. Together with W_0 this yields the appropriate $\boldsymbol{w}$ to transform to the synchronous gauge on $\mathscr{O}$. Outside of $\mathscr{O}$ we let $\boldsymbol{w}$ smoothly decay to zero.

To verify that $\boldsymbol{w}$ has spacelike-compact support, first consider the support of W_0. On the region exterior to $\operatorname{supp}\boldsymbol{\gamma}$, which is spacelike compact, (4.5.31) reduces to $\nabla_{\boldsymbol{n}} W_0 = 0$ and so $W_0 = \textit{constant}$ along each normal geodesic emanating from Σ, as long as the geodesic does not enter $\operatorname{supp}\boldsymbol{\gamma}$. Choosing $W_0|_\Sigma = 0$ yields $W_0 = 0$ along every geodesic that does not intersect $\operatorname{supp}\boldsymbol{\gamma}$; hence $W_0|_{\mathscr{O}}$ is spacelike-compact. Using this means that outside $\operatorname{supp}\boldsymbol{\gamma}$ equation (4.5.33) reduces to $(\nabla_{\tilde{\boldsymbol{n}}} w_\parallel)_b - w_{\parallel a}\nabla_b \tilde{n}^a = 0$; in Gaussian normal coordinates, the right-hand side of (4.5.36) vanishes. Thus with $\boldsymbol{w}_\parallel|_\Sigma = 0$, the solution $w_{\parallel i}$ vanishes in every $\mathscr{M}_q$ that does not intersect $\operatorname{supp}\boldsymbol{\gamma}$, so $\boldsymbol{w}|_{\mathscr{O}}$ is spacelike-compact. Outside $\mathscr{O}$ we let $\boldsymbol{w}$ smoothly decay to zero. Hence $\boldsymbol{w}$ may be chosen to be compactly supported and therefore $\pounds_{\boldsymbol{w}}\boldsymbol{g} \in \mathscr{G}(M;\mathbb{R})$. ■

For the case of Minkowski spacetime, the exponential map is globally defined [77, Ex. 3.34]. Therefore we can obtain a global normal neighbourhood for an ordinary $t = \textit{constant}$ Cauchy surface and hence obtain the synchronous gauge globally.

4.6 Existence and uniqueness of solutions to the linearized Einstein equation

We will now explicitly prove the existence and uniqueness, up to gauge, of spacelike-compact solutions to the linearized Einstein equation. Until the appearance of the author's joint paper [34], which is the primary work on which this thesis is based, such a treatment of the existence and uniqueness of solutions had proved to be elusive within the existing literature.

4.6.1 Existence

We begin by selecting an initial surface, which will be a smooth spacelike Cauchy surface Σ with future-pointing unit normal vector $\boldsymbol{n}$. This surface is where the initial data will be specified. The data consists of two pieces, the initial value of the perturbation $\boldsymbol{\gamma}|_\Sigma$ and the forward normal derivative $\nabla_{\boldsymbol{n}}\boldsymbol{\gamma}|_\Sigma$. In order to simplify notation and improve the aesthetics, we introduce the initial data map, $\mathrm{Data}_\Sigma : C^\infty(T^0_2(M;\mathbb{R})) \to C^\infty((T^0_2(M;\mathbb{R})|_\Sigma) \oplus C^\infty((T^0_2(M;\mathbb{R})|_\Sigma)$ defined by

$$\mathrm{Data}_\Sigma(\boldsymbol{\gamma}) := (\boldsymbol{\gamma}|_\Sigma, \nabla_{\boldsymbol{n}}\boldsymbol{\gamma}|_\Sigma). \tag{4.6.1}$$

Note that the notation $C^\infty((T^0_2(M;\mathbb{R})|_\Sigma)$ means the restriction to points in Σ of tensor fields in $C^\infty(T^0_2(M;\mathbb{R}))$.

The initial data for the linearized Einstein equation cannot however be freely specified. The data must satisfy certain constraints, which come from particular components of the linearized Einstein tensor with cosmological constant. Specifically, the components $L_{ab}(\boldsymbol{\gamma})n^b$, as they do not include any second order time-derivatives and so will be completely fixed once the initial data has been selected. For identical reasons as for the case of the initial data map $\mathrm{Data}_\Sigma(\cdot)$, we now introduce the constraint map, $\boldsymbol{C}^\Sigma : C^\infty((T^0_2(M;\mathbb{R})|_\Sigma) \oplus C^\infty((T^0_2(M;\mathbb{R})|_\Sigma) \to C^\infty(T^0_1(M;\mathbb{R})|_\Sigma)$, which satisfies

$$\boldsymbol{C}^\Sigma(\mathrm{Data}_\Sigma(\boldsymbol{\gamma})) = \boldsymbol{C}^\Sigma(\boldsymbol{\gamma}|_\Sigma, \nabla_{\boldsymbol{n}}\boldsymbol{\gamma}|_\Sigma) := n^a L_{ab}(\boldsymbol{\gamma})|_\Sigma. \tag{4.6.2}$$

Note that the gauge invariance of $L_{ab}(\boldsymbol{\gamma})$ (on cosmological vacuum spacetimes) entails that $\boldsymbol{C}^\Sigma \circ \mathrm{Data}_\Sigma(\boldsymbol{\gamma})$ is also gauge invariant. Also, the precise form of $\boldsymbol{C}^\Sigma$ in terms of $\boldsymbol{\gamma}|_\Sigma$ and $\nabla_{\boldsymbol{n}}\boldsymbol{\gamma}|_\Sigma$ is not required here and is therefore not included. We now see that for initial data to be admissible, then they must lie in the kernel of $\boldsymbol{C}^\Sigma$.

Theorem 4.6.1 *Let Σ be a smooth spacelike Cauchy surface with future-pointing unit normal vector $\boldsymbol{n}$. For any initial data $\boldsymbol{\zeta}$, $\boldsymbol{\xi} \in C^\infty_0(S^0_2(M;\mathbb{R})|_\Sigma)$ satisfying the initial value constraint $\boldsymbol{C}^\Sigma(\boldsymbol{\zeta},\boldsymbol{\xi}) = 0$ there exists a solution $\boldsymbol{\gamma} \in \mathscr{T}(M;\mathbb{R})$ to $L_{ab}(\boldsymbol{\gamma}) = 0$ such that* $\mathrm{Data}_\Sigma(\boldsymbol{\gamma}) = (\boldsymbol{\zeta},\boldsymbol{\xi})$.

Proof. It will be necessary to extend the specified initial data $(\boldsymbol{\zeta},\boldsymbol{\xi})$ off the surface Σ to obtain smooth tensor fields on M whose restriction to Σ is the data. The extension will be made in a completely arbitrary way and the reason for doing it is to obtain objects which can then be manipulated in a smooth manner. Firstly we extend $\boldsymbol{\zeta}$. This is achieved by considering a normal neighbourhood of Σ, see section 4.5.3 for details on such neighbourhoods, on which we use parallel transport along the normal geodesics to obtain a tensor field $\tilde{\boldsymbol{\zeta}}$. Specifically, we solve $\nabla_{\tilde{\boldsymbol{n}}}\tilde{\boldsymbol{\zeta}} = 0$ with $\tilde{\boldsymbol{\zeta}}|_\Sigma = \boldsymbol{\zeta}$, where $\tilde{\boldsymbol{n}}$ is the normal vector field. The explicit form of the solution to this particular system would be given by (4.5.35). By our choice of initial data, the solution $\tilde{\boldsymbol{\zeta}}$ satisfies $\mathrm{Data}_\Sigma(\tilde{\boldsymbol{\zeta}}) = (\boldsymbol{\zeta}, 0)$. Now take an arbitrary extension of $\boldsymbol{\xi}$, calling the result $\tilde{\boldsymbol{\xi}}$, such that $\tilde{\boldsymbol{\xi}}|_\Sigma = \boldsymbol{\xi}$ and define $\boldsymbol{\chi} = \tilde{\boldsymbol{\zeta}} + s\tilde{\boldsymbol{\xi}}$, where for any point p in the normal neighbourhood, s is the parameter distance along the unique normal geodesic connnecting p to Σ. As $s = 0$ on Σ, we have

$$\boldsymbol{\chi}|_\Sigma = \tilde{\boldsymbol{\zeta}}|_\Sigma = \boldsymbol{\zeta}.$$

Also,

$$\nabla_{\tilde{\boldsymbol{n}}}\boldsymbol{\chi}|_\Sigma = \nabla_{\tilde{\boldsymbol{n}}}\tilde{\boldsymbol{\zeta}}|_\Sigma + (\nabla_{\tilde{\boldsymbol{n}}}s)|_\Sigma\tilde{\boldsymbol{\xi}}|_\Sigma + s|_\Sigma\nabla_{\tilde{\boldsymbol{n}}}\tilde{\boldsymbol{\xi}}|_\Sigma,$$

so again as $s = 0$ on Σ, and $\nabla_{\tilde{\boldsymbol{n}}}\tilde{\boldsymbol{\zeta}}|_\Sigma = 0$, the above simplifies to

$$\nabla_{\tilde{\boldsymbol{n}}}\boldsymbol{\chi}|_\Sigma = (\nabla_{\tilde{\boldsymbol{n}}}s)|_\Sigma\tilde{\boldsymbol{\xi}}|_\Sigma = \tilde{\boldsymbol{\xi}}|_\Sigma = \boldsymbol{\xi},$$

where the second equality comes from $\tilde{\boldsymbol{n}}$ being unit-parameterised. Finally, as $\tilde{\boldsymbol{n}}|_\Sigma = \boldsymbol{n}$ we have $\nabla_{\tilde{\boldsymbol{n}}}\boldsymbol{\chi}|_\Sigma = \nabla_{\boldsymbol{n}}\boldsymbol{\chi}|_\Sigma$ and so $\mathrm{Data}_\Sigma(\boldsymbol{\chi}) = (\boldsymbol{\zeta}, \boldsymbol{\xi})$.

We thus have a suitably smooth object on which we can perform a de Donder gauge transformation. By Theorem 4.5.2 there exists a $\boldsymbol{w} \in C^\infty_{SC}(T^1_0(M;\mathbb{R}))$ such that $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\chi} + \pounds_{\boldsymbol{w}}\boldsymbol{g}$ obeys the de Donder condition $\nabla^a\overline{\tilde{\gamma}}_{ab} = 0$ and

$$\begin{aligned}(\tilde{\gamma}_{ab} - 2\nabla_{(a}w_{b)})|_\Sigma &= \chi_{ab}|_\Sigma = \zeta_{ab}\\ n^c\nabla_c(\tilde{\gamma}_{ab} - 2\nabla_{(a}w_{b)})|_\Sigma &= n^c\nabla_c\chi_{ab}|_\Sigma = \xi_{ab}.\end{aligned} \tag{4.6.3}$$

We now use $\tilde{\boldsymbol{\gamma}}$'s initial data to obtain a unique solution to the hyperbolic differential equation $P(\boldsymbol{\gamma}) = 0$, see (4.2.9) for the definition of P and Theorem 4.2.2 for its relationship to the linearized Einstein tensor with cosmological constant. Specifically, let $\hat{\boldsymbol{\gamma}}$ be the unique solution to $P_{ab}{}^{cd}\hat{\gamma}_{cd} = 0$ whose initial data satisfies $\mathrm{Data}_\Sigma(\hat{\boldsymbol{\gamma}}) = \mathrm{Data}_\Sigma(\tilde{\boldsymbol{\gamma}})$. The existence and uniqueness of $\hat{\boldsymbol{\gamma}}$ is guaranteed by [11, Thm 3.2.11]. The aim now is to show that $\hat{\boldsymbol{\gamma}}$ obeys the de Donder condition. By Lemma 4.2.3, $\nabla^a\overline{\hat{\gamma}}_{ab}$ obeys the hyperbolic equation $(\Box + \Lambda)(\nabla^a\hat{\gamma}_{ab}) = 0$; it also vanishes on Σ because $\nabla^a\overline{\hat{\gamma}}_{ab}|_\Sigma = \nabla^a\overline{\tilde{\gamma}}_{ab}|_\Sigma = 0$. Therefore if it can be shown that $\nabla_{\boldsymbol{n}}\nabla^a\overline{\hat{\gamma}}_{ab}|_\Sigma = 0$, then $\hat{\boldsymbol{\gamma}}$ will be de Donder. To prove this requires the following.

Lemma 4.6.2 *On a cosmological vacuum background spacetime, for any solution $\hat{\boldsymbol{\gamma}}$ to $P(\hat{\boldsymbol{\gamma}}) = 0$, it holds that $n^c\nabla_c(\nabla^a\overline{\hat{\gamma}}_{ab})|_\Sigma = 2L_{ab}(\hat{\boldsymbol{\gamma}})n^a|_\Sigma$.*

Proof. Combining the hypothesis Lemma 4.2.1 and Theorem 4.2.2 gives

$$2L_{ab}(\hat{\boldsymbol{\gamma}}) = (\overline{\pounds_{(\nabla\cdot\overline{\hat{\gamma}})^\sharp}\boldsymbol{g}})_{ab}.$$

Contracting both sides of this equation with $\boldsymbol{n}$ and then expanding out the right-hand side explicitly gives

$$2L_{ab}(\hat{\boldsymbol{\gamma}})n^a|_\Sigma = n^a\nabla_a\nabla^c\overline{\hat{\gamma}}_{cb}|_\Sigma + n^a\nabla_b\nabla^c\overline{\hat{\gamma}}_{ca}|_\Sigma - n_b\nabla^d\nabla^c\overline{\hat{\gamma}}_{cd}|_\Sigma. \tag{4.6.4}$$

The spacetime metric $\boldsymbol{g}$ can be decomposed, see [57, Ch. 2.7], in terms of the normal vector $\boldsymbol{n}$ and a projection operator ${q_a}^b$ so that $g_{ab} = -n_a n_b + q_{ab}$. This allows for tensors to be split into their components normal and tangential to Σ. Applying such a procedure to (4.6.4) gives

$$2L_{ab}(\hat{\boldsymbol{\gamma}})n^a|_\Sigma = n^a\nabla_a\nabla^c\bar{\hat{\gamma}}_{cb}|_\Sigma - n^a n_b n^d \nabla_d \nabla^c \bar{\hat{\gamma}}_{ca}|_\Sigma + n^a {q_b}^d \nabla_d \nabla^c \bar{\hat{\gamma}}_{ca}|_\Sigma + n_b n^d n^e \nabla_e \nabla^c \bar{\hat{\gamma}}_{cd}|_\Sigma - n_b q^{de} \nabla_e \nabla^c \bar{\hat{\gamma}}_{ca}|_\Sigma. \quad (4.6.5)$$

The second and fourth terms on the right-hand side of this equation cancel each other. Finally, as $\nabla^a\bar{\hat{\gamma}}_{ab}|_\Sigma = 0$ the third and fifth terms on the right-hand side of (4.6.5) vanish because they take the derivative of $\nabla^a\bar{\hat{\gamma}}_{ab}$ tangential to Σ. Hence we obtain

$$2L_{ab}(\hat{\boldsymbol{\gamma}})n^a|_\Sigma = n^c\nabla_c(\nabla^a\bar{\hat{\gamma}}_{ab})|_\Sigma.$$

■

As the constraints are gauge invariant, the following chain of equalities holds,

$$L_{ab}(\hat{\boldsymbol{\gamma}})n^a|_\Sigma = \boldsymbol{C}^\Sigma(\mathrm{Data}_\Sigma(\hat{\boldsymbol{\gamma}})) = \boldsymbol{C}^\Sigma(\mathrm{Data}_\Sigma(\tilde{\boldsymbol{\gamma}})) = \boldsymbol{C}^\Sigma(\mathrm{Data}_\Sigma(\boldsymbol{\chi})) = 0$$

and thus upon using Lemma 4.6.2 we see that $n^c\nabla_c(\nabla^a\bar{\hat{\gamma}}_{ab})|_\Sigma = 0$. Therefore we have shown that $\mathrm{Data}_\Sigma(\nabla\cdot\bar{\hat{\boldsymbol{\gamma}}}) = 0$ and so $\nabla\cdot\bar{\hat{\boldsymbol{\gamma}}} = 0$ by [11, Cor. 3.2.4]. Hence, $\hat{\boldsymbol{\gamma}}$ satisfies $P(\hat{\boldsymbol{\gamma}}) = 0$ and $\nabla\cdot\bar{\hat{\boldsymbol{\gamma}}} = 0$ and is therefore a solution to the linearized Einstein equation, $L(\hat{\boldsymbol{\gamma}}) = 0$.

Finally, by undoing the original gauge transformation, used to put $\boldsymbol{\chi}$ into the de Donder gauge, we obtain a solution $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}} - \pounds_{\boldsymbol{w}}\boldsymbol{g}$ to $L(\boldsymbol{\gamma}) = 0$, which satisfies $\mathrm{Data}_\Sigma(\boldsymbol{\gamma}) = (\boldsymbol{\zeta}, \boldsymbol{\xi})$. To see that this final statement is true, use $\mathrm{Data}_\Sigma(\hat{\boldsymbol{\gamma}}) = \mathrm{Data}_\Sigma(\tilde{\boldsymbol{\gamma}})$ and the expression for $\mathrm{Data}_\Sigma(\tilde{\boldsymbol{\gamma}})$ given in (4.6.3). ■

4.6.2 Uniqueness

Having addressed, in Theorem 4.6.1, the issue of the existence of a solution to the linearized Einstein equation with initial data satisfying the constraints, we now move to find out whether the solution will be uniquely specified by its initial data. It turns out that it will be unique up to gauge equivalence.

Theorem 4.6.3 *Suppose* $\boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \mathscr{S}(M;\mathbb{R})$ *with* $\mathrm{Data}_\Sigma(\boldsymbol{\gamma}) = \mathrm{Data}_\Sigma(\boldsymbol{\gamma}')$ *on some spacelike Cauchy surface* Σ*. Then* $\boldsymbol{\gamma} = \boldsymbol{\gamma}' + \pounds_{\boldsymbol{w}}\boldsymbol{g}$*, also written as* $\boldsymbol{\gamma} \sim \boldsymbol{\gamma}'$*, where* $\pounds_{\boldsymbol{w}}\boldsymbol{g} \in \mathscr{G}(M;\mathbb{R})$*. If, additionally,* $\boldsymbol{\gamma}, \boldsymbol{\gamma}' \in \mathscr{S}^{dD}(M;\mathbb{R})$ *then gauge equivalence is replaced by equality.*

Proof. Let $\boldsymbol{\xi} = \boldsymbol{\gamma} - \boldsymbol{\gamma}'$, which satisfies $\mathrm{Data}_\Sigma(\boldsymbol{\xi}) = 0$ by definition. Now gauge transform $\boldsymbol{\xi}$ to the de Donder gauge. Theorem 4.5.2 entails that this de Donder gauge transformation can be made with a spacelike-compact vector field $\boldsymbol{w}$ whose initial data vanishes. Therefore we have a de Donder perturbation $\tilde{\boldsymbol{\gamma}} = \boldsymbol{\xi} + \pounds_{\boldsymbol{w}}\boldsymbol{g}$ whose initial data is

$$\begin{aligned} \tilde{\gamma}_{ab}|_\Sigma &= (\nabla_a w_b + \nabla_b w_a)|_\Sigma = 0 \\ n^c \nabla_c \tilde{\gamma}_{ab}|_\Sigma &= n^c \nabla_c (\nabla_a w_b + \nabla_b w_a)|_\Sigma. \end{aligned} \tag{4.6.6}$$

We now prove that the time-derivative (4.6.6) is compelled to vanish. This is achieved by considering how the first term on the right-hand side of (4.6.6) may be re-expressed. Note that the second term is just the first term with indices reversed and so we need only consider the first term. Upon utilising the Riemann tensor identity (2.1.1) we have

$$n^c \nabla_c \nabla_a w_b|_\Sigma = n^c \nabla_a \nabla_c w_b|_\Sigma + n^c R_{cab}{}^d w_d|_\Sigma,$$

which can then be rearranged, using the Leibniz rule, to

$$n^c \nabla_c \nabla_a w_b|_\Sigma = \nabla_a (n^c \nabla_c w_b)|_\Sigma - (\nabla_a n^c) \nabla_c w_b|_\Sigma + n^c R_{cab}{}^d w_d|_\Sigma.$$

Now, by the choice of data for $\boldsymbol{w}$ we know that $w_b|_\Sigma = 0$ and $\nabla_a w_b|_\Sigma = 0$. Hence,

$$n^c \nabla_c \nabla_a w_b|_\Sigma = \nabla_a (n^c \nabla_c w_b)|_\Sigma.$$

Using the decomposed metric tensor this can be re-expressed as

$$n^c \nabla_c \nabla_a w_b|_\Sigma = -n_a n^d \nabla_d (n^c \nabla_c w_b)|_\Sigma + q_a{}^d \nabla_d (n^c \nabla_c w_b)|_\Sigma.$$

The second term on the right-hand side vanishes because $(n^c \nabla_c w_b)|_\Sigma = 0$. We will now show that the first term on the right-hand side also vanishes. By evaluating the de Donder equation of motion for $\boldsymbol{w}$ on the Cauchy surface and using that $\mathrm{Data}_\Sigma(\boldsymbol{\xi}) = 0$, then it holds that $(\Box + \Lambda)\boldsymbol{w}|_\Sigma = -\nabla \cdot \overline{\boldsymbol{\xi}}|_\Sigma = 0$. Expanding out the left-hand side of this equation gives

$$(\Box + \Lambda) w_c|_\Sigma = -n^a n^b \nabla_a \nabla_b w_c|_\Sigma + q^{ab} \nabla_a \nabla_b w_c|_\Sigma + \Lambda w_c|_\Sigma$$
$$= -n^a \nabla_a n^b \nabla_b w_c|_\Sigma + (n^a \nabla_a n^b) \nabla_b w_c|_\Sigma + q^{ab} \nabla_a \nabla_b w_c|_\Sigma + \Lambda w_c|_\Sigma. \quad (4.6.7)$$

As the initial data for $\boldsymbol{w}$ vanish, the final three terms on the right-hand side will vanish and therefore $0 = (\Box + \Lambda) w_c|_\Sigma = -n^a \nabla_a n^b \nabla_b w_c|_\Sigma = n^c \nabla_c \nabla_a w_b|_\Sigma$. Therefore by (4.6.6), $n^c \nabla_c \tilde{\gamma}_{ab}|_\Sigma = 0$. This entails that $\tilde{\boldsymbol{\gamma}}$ satisfies $\mathrm{Data}_\Sigma(\tilde{\boldsymbol{\gamma}}) = 0$ and we know, by assumption, that it is a solution to the linearized Einstein equation obeying the de Donder condition and so it satisfies $P(\tilde{\boldsymbol{\gamma}}) = 0$. This equation in conjunction with the vanishing initial data entails [11, Cor. 3.2.4] that $\tilde{\boldsymbol{\gamma}} = 0$. Therefore $\boldsymbol{\xi}$ is gauge equivalent to the trivial solution and so $\boldsymbol{\gamma}$ is gauge equivalent to $\boldsymbol{\gamma}'$.

If both the solutions $\boldsymbol{\gamma}$ and $\boldsymbol{\gamma}'$ obey the de Donder equation, then they will both solve the hyperbolic equation $P(\boldsymbol{\gamma}) = P(\boldsymbol{\gamma}') = 0$ with identical initial data and will therefore be the same solution [11, Cor. 3.2.4]. ■

4.7 Green's operators and their intertwinings

The general theory of Green's operators was discussed at length in section 2.4.2. Here we wish to apply that theory to the specific case of the differential operators arising in linearized gravity. We will show how each gauge equivalence class of linearized gravity solutions admits elements that may be written in terms of the advanced-minus-retarded solution operator $\boldsymbol{E}$ and we will show how the actions of the various Green's operators intertwine with the action of other operations, such as the trace, the trace-reversal and the Lie-derivative. These results will be important in later chapters, where, for example, the advanced-minus-retarded solution operator appears in the Poisson bracket of the basic classical observables and consequently in the commutator of their quantum counterparts.

To begin, in section 4.2 it was noted that P is a normally hyperbolic operator and so we have the following lemma.

Lemma 4.7.1 *Any $\boldsymbol{\gamma} \in \mathscr{T}(M;\mathbb{R})$ solving $P(\boldsymbol{\gamma}) = 0$ may be written as $\boldsymbol{\gamma} = \boldsymbol{E}\boldsymbol{f}$ with $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{R}))$, where $\boldsymbol{E}$ is the advanced-minus-retarded solution operator associated with P.*

Proof. By the exact sequence of Theorem 2.4.3, any spacelike-compact element of the kernel of P is equal to an element in the image of $\boldsymbol{E}$. ■

This shows that any solution to $P(\boldsymbol{\gamma}) = 0$ may be written as $\boldsymbol{Ef}$. However, for it to be a linearized gravity solution then it must also satisfy the de Donder condition. In order to obtain the circumstances under which this occurs, we first need to show that the trace and the trace-reversal operations commute with the Green's operators.

Lemma 4.7.2 *Given any* $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{R}))$ *then* $g^{ab}(\boldsymbol{Ef})_{ab} = Ef$*, where* $f = f_a^a$ *is the trace of* $\boldsymbol{f}$*, and* E *is the scalar advanced-minus-retarded solution operator associated with the differential operator* $\Box + 2\Lambda$*.*

Proof. $\boldsymbol{E}^\pm \boldsymbol{f}$ are the unique solutions to $P(\boldsymbol{\gamma}) = \boldsymbol{f}$ with past/future compact support and $E^\pm f$ are the unique solutions to $(\Box + 2\Lambda)\gamma = f$ with past/future compact support. By taking the trace of $P(\boldsymbol{\gamma}) = \boldsymbol{f}$ and using (4.2.10), we see that $g^{ab}(\boldsymbol{E}^\pm \boldsymbol{f})_{ab}$ solve $(\Box + 2\Lambda)\gamma = f$ with past/future compact support. Therefore by the uniqueness of solutions with such support properties we have $g^{ab}(\boldsymbol{E}^\pm \boldsymbol{f})_{ab} = E^\pm f$. Since taking the trace is a linear operation, this yields $g^{ab}(\boldsymbol{Ef})_{ab} = Ef$. ■

Lemma 4.7.3 *For all* $\boldsymbol{f} \in C^\infty(T_2^0(M;\mathbb{R}))$*, we have* $\overline{\boldsymbol{Ef}} = \boldsymbol{E}\overline{\boldsymbol{f}}$*.*

Proof. By Theorem 2.4.2, $\tilde{\boldsymbol{\gamma}}^\pm = \boldsymbol{E}^\pm \overline{\boldsymbol{f}}$ are the unique solutions to $P(\tilde{\boldsymbol{\gamma}}^\pm) = \overline{\boldsymbol{f}}$ with past/future compact support and likewise $\boldsymbol{\gamma}^\pm = \boldsymbol{E}^\pm \boldsymbol{f}$ are the unique solutions to $P(\boldsymbol{\gamma}^\pm) = \boldsymbol{f}$ with past/future compact support. Since, by Lemma 4.2.1, trace-reversal commutes with P we have

$$P(\boldsymbol{E}^\pm \overline{\boldsymbol{f}}) = \overline{\boldsymbol{f}} = \overline{P(\boldsymbol{E}^\pm \boldsymbol{f})} = P(\overline{\boldsymbol{E}^\pm \boldsymbol{f}}).$$

Therefore by uniqueness (from the support properties) $\boldsymbol{E}^\pm \overline{\boldsymbol{f}} = \overline{\boldsymbol{E}^\pm \boldsymbol{f}}$ and so $\boldsymbol{E}\overline{\boldsymbol{f}} = \boldsymbol{E}^- \overline{\boldsymbol{f}} - \boldsymbol{E}^+ \overline{\boldsymbol{f}} = \overline{\boldsymbol{E}^- \boldsymbol{f}} - \overline{\boldsymbol{E}^+ \boldsymbol{f}} = \overline{\boldsymbol{Ef}}$, as trace-reversal is a linear operator. ■

Now we are in a position to show under what circumstances a solution $\boldsymbol{\gamma} = \boldsymbol{Ef}$ to $P(\boldsymbol{\gamma}) = 0$ will obey the de Donder condition.

Theorem 4.7.4 *For any* $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{R}))$*, we have* $\boldsymbol{Ef} \in \mathscr{S}^{dD}(M;\mathbb{R})$ *if and only if* $\nabla \cdot \overline{\boldsymbol{f}} \in (\Box + \Lambda) C_0^\infty(T_1^0(M;\mathbb{R}))$*.*

Proof. By the definition of the de Donder gauge, $\boldsymbol{Ef} \in \mathscr{S}^{dD}(M;\mathbb{R})$ if and only if $\nabla \cdot \overline{\boldsymbol{Ef}} \equiv 0$ or equivalently, using Lemma 4.7.3, $\nabla \cdot \boldsymbol{E}\overline{\boldsymbol{f}} = 0$.

Taking the divergence of $P(\boldsymbol{E}^\pm \overline{\boldsymbol{f}}) = \overline{\boldsymbol{f}}$ and utilising Lemma 4.2.3, we find that

$$(\Box + \Lambda)(\nabla \cdot \boldsymbol{E}^\pm \overline{\boldsymbol{f}}) = \nabla \cdot \overline{\boldsymbol{f}}.$$

Now, by Theorem 2.4.2, $\boldsymbol{w}^{\pm} = \hat{\boldsymbol{E}}^{\pm}\nabla\cdot\overline{\boldsymbol{f}}$ are the unique solutions to $(\Box+\Lambda)\boldsymbol{w}^{\pm} = \nabla\cdot\overline{\boldsymbol{f}}$ with past/future compact support, where $\hat{\boldsymbol{E}}^{\pm}$ are the retarded/advanced Green's operators for $(\Box+\Lambda)$ on *covector* fields. However, $\boldsymbol{v}^{\pm} = \nabla\cdot\boldsymbol{E}^{\pm}\overline{\boldsymbol{f}}$ also solve $(\Box+\Lambda)\boldsymbol{v}^{\pm} = \nabla\cdot\overline{\boldsymbol{f}}$ and have past/future compact support. Therefore we deduce, by uniqueness of the solutions, that $\nabla\cdot\boldsymbol{E}^{\pm}\overline{\boldsymbol{f}} = \hat{\boldsymbol{E}}^{\pm}\nabla\cdot\overline{\boldsymbol{f}}$ and hence

$$\hat{\boldsymbol{E}}\nabla\cdot\overline{\boldsymbol{f}} = \hat{\boldsymbol{E}}^{-}\nabla\cdot\overline{\boldsymbol{f}} - \hat{\boldsymbol{E}}^{+}\nabla\cdot\overline{\boldsymbol{f}} = \nabla\cdot\boldsymbol{E}^{-}\overline{\boldsymbol{f}} - \nabla\cdot\boldsymbol{E}^{+}\overline{\boldsymbol{f}} = \nabla\cdot(\boldsymbol{E}^{-}-\boldsymbol{E}^{+})\overline{\boldsymbol{f}} = \nabla\cdot\boldsymbol{E}\overline{\boldsymbol{f}}.$$

By assumption the right-hand side vanishes meaning that $\hat{\boldsymbol{E}}\nabla\cdot\overline{\boldsymbol{f}} = 0$, which occurs, by the exact sequence of Theorem 2.4.3, if and only if $\nabla\cdot\overline{\boldsymbol{f}} \in (\Box+\Lambda)C_0^{\infty}(T_1^0(M;\mathbb{R}))$. ■

In fact, as we will show, one is able to select an $\boldsymbol{f} \in C_0^{\infty}(S_2^0(M;\mathbb{R}))$ such that $\nabla\cdot\overline{\boldsymbol{f}} = 0$ by exploiting the residual gauge freedom in the de Donder gauge. To do this we will need to know how the action of the Green's operators intertwines with the action of the Lie derivative and this is what the next lemma shows.

Lemma 4.7.5 *Given a $\boldsymbol{v} \in C_0^{\infty}(T_0^1(M;\mathbb{R}))$ on a cosmological vacuum background spacetime, then*

$$\pounds_{\tilde{\boldsymbol{E}}\boldsymbol{v}}\boldsymbol{g} = \boldsymbol{E}(\pounds_{\boldsymbol{v}}\boldsymbol{g}), \tag{4.7.1}$$

where $\tilde{\boldsymbol{E}}$ is the advanced-minus-retarded solution operator for $(\Box+\Lambda)$ on vector *fields and $\boldsymbol{E}$ is the advanced-minus-retarded solution operator for P.*

Proof. We know that $(\Box+\Lambda)\tilde{\boldsymbol{E}}^{\pm}\boldsymbol{v} = \boldsymbol{v}$, so using this and Lemma 4.2.4 gives

$$P(\pounds_{\tilde{\boldsymbol{E}}^{\pm}\boldsymbol{v}}\boldsymbol{g}) = \pounds_{(\Box+\Lambda)\tilde{\boldsymbol{E}}^{\pm}\boldsymbol{v}}\boldsymbol{g} = \pounds_{\boldsymbol{v}}\boldsymbol{g}.$$

The unique solutions to this equation with past/future compact support are, by Theorem 2.4.2, $\boldsymbol{E}^{\pm}\pounds_{\boldsymbol{v}}\boldsymbol{g}$ and therefore $\pounds_{\tilde{\boldsymbol{E}}^{\pm}\boldsymbol{v}}\boldsymbol{g} = \boldsymbol{E}^{\pm}\pounds_{\boldsymbol{v}}\boldsymbol{g}$ by uniqueness. Combining these results as follows

$$\boldsymbol{E}\pounds_{\boldsymbol{v}}\boldsymbol{g} = \boldsymbol{E}^{-}\pounds_{\boldsymbol{v}}\boldsymbol{g} - \boldsymbol{E}^{+}\pounds_{\boldsymbol{v}}\boldsymbol{g} = \pounds_{\tilde{\boldsymbol{E}}^{-}\boldsymbol{v}}\boldsymbol{g} - \pounds_{\tilde{\boldsymbol{E}}^{+}\boldsymbol{v}}\boldsymbol{g} = \pounds_{\tilde{\boldsymbol{E}}^{-}\boldsymbol{v}-\tilde{\boldsymbol{E}}^{+}\boldsymbol{v}}\boldsymbol{g} = \pounds_{\tilde{\boldsymbol{E}}\boldsymbol{v}}\boldsymbol{g}$$

achieves the result (4.7.1). ■

This now allows us to prove that in each gauge equivalence class there is a de Donder solution that is equal to $\boldsymbol{E}\boldsymbol{f}$ with $\boldsymbol{f}$ obeying $\nabla\cdot\overline{\boldsymbol{f}} = 0$. This will be important in section 5.4 when we consider the fundamental observables of the theory, the smeared fields, which will have the restriction $\nabla\cdot\boldsymbol{f} = 0$ placed upon the smearing tensors.

Theorem 4.7.6 *Any $\boldsymbol{\gamma} \in \mathscr{S}(M;\mathbb{R})$ is gauge equivalent to an $\boldsymbol{Ef}$ for some $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{R}))$ satisfying $\nabla \cdot \overline{\boldsymbol{f}} = 0$.*

Proof. By Corollary 4.5.3 we have $\boldsymbol{\gamma} \sim \boldsymbol{\gamma}^{dD}$ for some de Donder representative $\boldsymbol{\gamma}^{dD}$. Lemma 4.7.1 and Theorem 4.7.4 entail that $\boldsymbol{\gamma}^{dD} = \boldsymbol{E}\tilde{\boldsymbol{f}}$ with $\nabla \cdot \overline{\tilde{\boldsymbol{f}}} = (\Box + \Lambda)\boldsymbol{v}^\flat$ for some $\boldsymbol{v} \in C_0^\infty(T_0^1(M;\mathbb{R}))$. We now perform a further de Donder gauge transformation on $\boldsymbol{E}\tilde{\boldsymbol{f}}$, meaning that the result is just a different de Donder representative of $[\boldsymbol{\gamma}]$. The gauge transformation is selected to be $\pounds_{\tilde{\boldsymbol{E}}\boldsymbol{v}}\boldsymbol{g} \in \mathscr{G}^{dD}(M;\mathbb{R})$. Thus we have

$$\boldsymbol{\gamma} \sim \boldsymbol{E}\tilde{\boldsymbol{f}} - \pounds_{\tilde{\boldsymbol{E}}\boldsymbol{v}}\boldsymbol{g} = \boldsymbol{E}(\tilde{\boldsymbol{f}} - \pounds_{\boldsymbol{v}}\boldsymbol{g}), \tag{4.7.2}$$

where the equality uses Lemma 4.7.5. Now define $\boldsymbol{f} := \tilde{\boldsymbol{f}} - \pounds_{\boldsymbol{v}}\boldsymbol{g}$, which is smooth and compactly supported on M. Calculating the divergence of the trace-reversal of $\boldsymbol{f}$ gives

$$\nabla \cdot \overline{\boldsymbol{f}} = \nabla \cdot \overline{\tilde{\boldsymbol{f}}} - \nabla \cdot (\overline{\pounds_{\boldsymbol{v}}\boldsymbol{g}}) = (\Box + \Lambda)\boldsymbol{v}^\flat - (\Box + \Lambda)\boldsymbol{v}^\flat = 0,$$

where Lemma 4.5.1 has been used. Finally, noting from (4.7.2) that $\boldsymbol{\gamma} \sim \boldsymbol{Ef}$ completes the proof. ■

The remaining two lemmas of this subsection will be required later to prove that the time-slice condition holds for the algebra of observables that we construct. First we show that in the gauge equivalence class of a solution to the inhomogeneous (non-hyperbolic) linearized Einstein equation having past/future compact support, there exists a representative with the same support properties and this representative is, in fact, just the retarded/advanced solution to the inhomogeneous hyperbolic equation associated with the inhomogeneous linearized Einstein equation.

Lemma 4.7.7 *On a cosmological vacuum background spacetime, given $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{R}))$, if $\boldsymbol{\gamma}^\pm \in \mathscr{T}(M;\mathbb{R})$ solves $L_{ab}(\boldsymbol{\gamma}^\pm) = f_{ab}$ with* supp $\boldsymbol{\gamma}^\pm$ *compact to the past/future then $\boldsymbol{\gamma}^\pm \sim -2\boldsymbol{E}^\pm\overline{\boldsymbol{f}}$.*

Proof. Theorem 4.5.2 shows that one may gauge transform a perturbation to the de Donder gauge. However, here we wish to preserve the additional support properties of $\boldsymbol{\gamma}^\pm$ too. Therefore it is necessary that the vector field generating the gauge transformation also have past/future compact support. We know that this vector field obeys the hyperbolic differential equation (4.5.3) and that the unique solutions to this equation with past/future compact support are, by Theorem 2.4.2, $\boldsymbol{w}^\pm = -\hat{\boldsymbol{E}}^\pm\nabla \cdot \overline{\gamma}$. Therefore $\boldsymbol{\gamma}' = \boldsymbol{\gamma}^\pm + \pounds_{\boldsymbol{w}^\pm}\boldsymbol{g}$ obeys $\nabla \cdot \overline{\gamma}' = 0$ with past/future compact support.

On a cosmological vacuum background spacetime, $L_{ab}(\pounds_{\boldsymbol{w}} \boldsymbol{g}) = 0$ and so $L_{ab}(\boldsymbol{\gamma}') = L_{ab}(\boldsymbol{\gamma}^{\pm}) = f_{ab}$, which simplifies, on account of the de Donder condition, see Theorem 4.2.2, to

$$P(\overline{\boldsymbol{\gamma}}') = -2\boldsymbol{f}.$$

The solutions to this inhomogeneous equation with past/future compact support are $\overline{\boldsymbol{\gamma}}' = -2\boldsymbol{E}^{\pm}\boldsymbol{f}$, see Theorem 2.4.2. Lemma 4.7.3 entails that $\boldsymbol{\gamma}' = -2\boldsymbol{E}^{\pm}\overline{\boldsymbol{f}}$ and hence the result. ■

As we have shown in Theorem 4.7.6, if we have a solution $\boldsymbol{E}\boldsymbol{f}$ with $\nabla \cdot \overline{\boldsymbol{f}} = 0$, then it is a de Donder solution. However, there exist pure gauge de Donder solutions, which are elements of the space $\mathscr{G}^{dD}(M;\mathbb{R})$ defined in Theorem 4.5.2[2]. Therefore we need to also consider under what circumstances such a solution is just pure gauge, meaning that it is only moving us about within the de Donder gauge of some fixed gauge equivalence class. This situation is what the next lemma considers.

Lemma 4.7.8 *Given a $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{R}))$ satisfying $\nabla \cdot \boldsymbol{f} = 0$, suppose that $\boldsymbol{E}\overline{\boldsymbol{f}} = \boldsymbol{E}\pounds_{\boldsymbol{v}}\boldsymbol{g}$ for some $\boldsymbol{v} \in C_0^\infty(T_0^1(M;\mathbb{R}))$. Then there exists a $\boldsymbol{h} \in C_0^\infty(S_2^0(M;\mathbb{R}))$ such that*

$$\boldsymbol{f} = -2L(\boldsymbol{h}). \tag{4.7.3}$$

Proof. $\boldsymbol{E}(\overline{\boldsymbol{f}} - \pounds_{\boldsymbol{v}}\boldsymbol{g}) = 0$ and so by the exact sequence of Theorem 2.4.3,

$$\overline{\boldsymbol{f}} = \pounds_{\boldsymbol{v}}\boldsymbol{g} + P(\boldsymbol{h}) \tag{4.7.4}$$

for some $\boldsymbol{h} \in C_0^\infty(S_2^0(M;\mathbb{R}))$. Taking the trace-reversal and then the divergence of (4.7.4) gives $\nabla \cdot \boldsymbol{f} = \nabla \cdot \overline{\pounds_{\boldsymbol{v}}\boldsymbol{g}} + \nabla \cdot \overline{P(\boldsymbol{h})}$. Lemmas 4.2.1 and 4.5.1 entail that this simplifies to $\nabla \cdot \boldsymbol{f} = (\Box + \Lambda)\boldsymbol{v}^\flat + \nabla \cdot P(\overline{\boldsymbol{h}})$ and then Lemma 4.2.3 means that it reduces to

$$\nabla \cdot \boldsymbol{f} = (\Box + \Lambda)(\boldsymbol{v}^\flat + \nabla \cdot \overline{\boldsymbol{h}}).$$

By assumption, $\nabla \cdot \boldsymbol{f} = 0$ and so as $\boldsymbol{v}$ and $\boldsymbol{h}$ are compactly supported, in particular, they have past/future compact support, then by [11, Thm 3.1.1], $\boldsymbol{v}^\flat = -\nabla \cdot \overline{\boldsymbol{h}}$. Inserting this into (4.7.4), trace-reversing and then applying Theorem 4.2.2 gives the result. ■

[2] On cosmological vacuum spacetimes, all pure gauge perturbations are linearized Einstein solutions.

Chapter 5

Phase space

Having thoroughly discussed all of the preliminary material dealing with the classical solutions of linearized gravity, we are now in a position to construct the phase space and define the observables of the theory. This will allow us, in the subsequent chapter, to quantize the theory using Dirac's prescription.

As highlighted in section 3.2.1, the phase space for linear field theories consists of the vector space of smooth solutions to the equation of motion, with, if required, certain support restrictions imposed. This space is also endowed with a symplectic product. However, matters are complicated here, as they are for the case of electromagnetism, by the presence of gauge symmetries.

This chapter is based upon section four and appendices B & C of the paper [34] cowritten by the author with Dr C. J. Fewster.

5.1 Pre-symplectic space

For exactly the same reasons as those given in section 3.2.1, we begin by considering the vector space of smooth spacelike-compact complex-valued perturbations that solve the linearized Einstein equation, that is, the space $\mathscr{S}(M;\mathbb{C})$. This space is the complexification of $\mathscr{S}(M;\mathbb{R})$. All of the results from chapter 4 in no way required the tensor fields to be real-valued and so they all continue to hold in the complex case too.

Following the arguments given for the case of the real scalar field in section 3.2.1, which concern how to equip the complexified space of solutions with a symplectic product, we again use the techniques of [68] and now the Lagrangian (4.2.26), to endow $\mathscr{S}(M;\mathbb{C})$ with a pre-symplectic product. Given a smooth spacelike Cauchy surface Σ with future-pointing unit normal vector $\boldsymbol{n}$, then the pre-symplectic product of solutions is defined to be

$$\omega_\Sigma(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) := \int_\Sigma (\gamma^1_{ab}\pi^{ab}_2 - \gamma^2_{ab}\pi^{ab}_1) dvol_{\boldsymbol{h}}, \tag{5.1.1}$$

where $\boldsymbol{\pi}$ is defined in terms of $\boldsymbol{\Pi}$, see (4.2.27), by

$$\pi^{ab} := -n_c \Pi^{cab}. \tag{5.1.2}$$

The presence of the minus sign reflects our choice of sign convention for the spacetime metric.

Just as for the scalar field, this product acts on elements of the underlying solution space rather than on elements of the tangent space; this is due to $\mathscr{S}(M;\mathbb{C})$ being a vector space and so one may identify $\mathscr{S}(M;\mathbb{C})$ with the tangent space $T_{\boldsymbol{\gamma}}\mathscr{S}(M;\mathbb{C})$ at any point $\boldsymbol{\gamma} \in \mathscr{S}(M;\mathbb{C})$. However, unlike for the scalar field there exist degeneracies here. Recall that a degeneracy is a non-trivial element whose product with every other element vanishes. The space of degeneracies is called the radical of ω_Σ. We will now be concerned with finding out exactly what the radical of ω_Σ is made up of. However, before doing that, we briefly consider whether the pre-symplectic product ω_Σ is independent of the choice of Cauchy surface. This is in fact the case when it acts on solutions, as the next lemma shows, and so henceforth, after proving this lemma, the subscript Σ will be dropped from ω when only solutions are considered.

Lemma 5.1.1 *Given $\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2 \in \mathscr{S}(M;\mathbb{C})$ and two spacelike Cauchy surfaces Σ and Σ', then $\omega_\Sigma(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = \omega_{\Sigma'}(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)$.*

Proof. By defining the current $j^c(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) := \gamma^2_{ab}\Pi^{cab}_1 - \gamma^1_{ab}\Pi^{cab}_2$, one immediately sees from (5.1.1) and (5.1.2) that

$$\omega_\Sigma(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = \int_\Sigma n_c j^c(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) dvol_{\boldsymbol{h}}. \tag{5.1.3}$$

The divergence of the current $j^c(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2)$ is given by

$$\nabla_c j^c(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = (\nabla_c \gamma^2_{ab})\Pi^{cab}_1 + \gamma^2_{ab}\nabla_c \Pi^{cab}_1 - (\nabla_c \gamma^1_{ab})\Pi^{cab}_2 - \gamma^1_{ab}\nabla_c \Pi^{cab}_2,$$

whereupon using the identity from Lemma 4.2.5 and (4.2.27) this becomes

$$\nabla_c j^c(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = \nabla_c \gamma^2_{ab} T^{cabdef} \nabla_d \gamma^1_{ef} + \gamma^2_{ab}(L^{ab}(\boldsymbol{\gamma}^1) - 2S^{abde}\gamma^1_{de}) \\ - \nabla_c \gamma^1_{ab} T^{cabdef} \nabla_d \gamma^2_{ef} - \gamma^1_{ab}(L^{ab}(\boldsymbol{\gamma}^2) - 2S^{abde}\gamma^2_{de}).$$

Using the symmetry properties of T^{abcdef} and S^{abcd}, all terms containing them cancel one another and so we are left with

$$\nabla_c j^c(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = \gamma^2_{ab} L^{ab}(\boldsymbol{\gamma}^1) - \gamma^1_{ab} L^{ab}(\boldsymbol{\gamma}^2),$$

which will vanish if $\boldsymbol{\gamma}^1$ and $\boldsymbol{\gamma}^2$ are solutions to the linearized Einstein equation, as we have assumed.

Now, without loss of generality we assume that $\Sigma \subset I^+(\Sigma')$, and denote by U the region bounded between these two Cauchy surfaces. The future-pointing unit normal vector on Σ is denoted by $\boldsymbol{n}$ and on Σ' is denoted by $\boldsymbol{n'}$. Applying Gauss' Theorem to the vector field j^a on the region U, it is clear that

$$\int_{\Sigma'} n'_a j^a dvol_{\boldsymbol{h}} - \int_{\Sigma} n_a j^a dvol_{\boldsymbol{h}} = \int_U \nabla_a j^c(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) dvol_{\boldsymbol{g}}.$$

As $\boldsymbol{\gamma}^1$ and $\boldsymbol{\gamma}^2$ are solutions, the right-hand side vanishes and so by using (5.1.3) we have

$$\omega_{\Sigma'}(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) - \omega_{\Sigma}(\boldsymbol{\gamma}^1, \boldsymbol{\gamma}^2) = 0$$

and hence the result. ■

We will now discuss the degeneracies of the product (5.1.1). Firstly, we will show that pure gauge perturbations are degeneracies. This entails that the product ω is a gauge invariant object. Here we work with the larger pure gauge subspace $\hat{\mathscr{G}}(M;\mathbb{C})$. In order to prove that the elements of $\hat{\mathscr{G}}(M;\mathbb{C})$ are degeneracies, we require the following identity that connects the initial value constraint (4.6.2) to the symplectic product. In this instance, we consider the symplectic product to be defined on all elements of $\mathscr{T}(M;\mathbb{C})$ and not just the subspace of solutions $\mathscr{S}(M;\mathbb{C})$. As such, for this theorem only and the lemmas contained within the proof of it, we have to include a subscript Σ to indicate the dependence of the pre-symplectic product on the choice of spacelike Cauchy surface Σ.

Theorem 5.1.2 *For any* $\boldsymbol{\gamma} \in \mathscr{T}(M;\mathbb{C})$, $\boldsymbol{w} \in C^\infty(T^1_0(M;\mathbb{C}))$ *and given any smooth space-*

like Cauchy surface Σ *with future-pointing unit normal vector* $\boldsymbol{n}$*, then we have*

$$\omega_\Sigma(\boldsymbol{\gamma}, \pounds_{\boldsymbol{w}}\boldsymbol{g}) = 2\int_\Sigma w^a L_{ab}(\boldsymbol{\gamma}) n^b dvol_{\boldsymbol{h}} = 2\int_\Sigma w^b C^\Sigma_b(\mathrm{Data}_\Sigma(\boldsymbol{\gamma})) dvol_{\boldsymbol{h}}. \tag{5.1.4}$$

Proof. The second equality follows directly from the definition of $\boldsymbol{C}^\Sigma$. To prove that the first equality holds, we will prove two further identities as lemmas within the main proof of the theorem. The proofs of these identities will utilise a vector field $\boldsymbol{v} \in C^\infty(T^1_0(M;\mathbb{C}))$, which has the properties that it agrees with the vector field $\boldsymbol{w}$ in a neighbourhood of Σ and vanishes to the far past of Σ. As far as we know, the origin of the use of such a vector field is [45], see the paragraph preceding equation (79) in that reference. The first identity is as follows.

Lemma 5.1.3 *With* $\boldsymbol{\gamma}$*,* $\boldsymbol{w}$ *and* $\boldsymbol{v}$ *as above, then*

$$\int_\Sigma w^a L_{ab}(\boldsymbol{\gamma}) n^b dvol_{\boldsymbol{h}} = -\int_{M^-} \nabla_{(a}v_{b)} L^{ab}(\boldsymbol{\gamma}) dvol_{\boldsymbol{g}}, \tag{5.1.5}$$

where $M^- = I^-(\Sigma)$ *is the region to the past of the Cauchy surface* Σ*.*

Proof. As $\boldsymbol{w} = \boldsymbol{v}$ on a neighbourhood of Σ, then

$$\int_\Sigma w^a L_{ab}(\boldsymbol{\gamma}) n^b dvol_{\boldsymbol{h}} = \int_\Sigma v^a L_{ab}(\boldsymbol{\gamma}) n^b dvol_{\boldsymbol{h}}.$$

Using the Gauss Theorem on the region M^- with the covector field $v^a L_{ab}(\boldsymbol{\gamma})$ then we have

$$\int_\Sigma v^a L_{ab}(\boldsymbol{\gamma}) n^b dvol_{\boldsymbol{h}} = -\int_{M^-} \nabla^b(v^a L_{ab}(\boldsymbol{\gamma})) dvol_{\boldsymbol{g}},$$

where we have used that $\boldsymbol{v}$ vanishes to the far past of Σ. The right-hand side can be expanded using the Leibniz rule, and then using the linearized Bianchi identity, $\nabla^a L_{ab}(\boldsymbol{\gamma}) = 0$, this becomes

$$\int_\Sigma v^a L_{ab}(\boldsymbol{\gamma}) n^b dvol_{\boldsymbol{h}} = -\int_{M^-} (\nabla^b v^a) L_{ab}(\boldsymbol{\gamma}) dvol_{\boldsymbol{g}}. \tag{5.1.6}$$

Finally, since L is symmetric then $L_{ab}(\boldsymbol{\gamma}) = L_{(ab)}(\boldsymbol{\gamma})$, and so because $\nabla^b v^a L_{(ab)}(\boldsymbol{\gamma}) = \nabla^{(b}v^{a)} L_{ab}(\boldsymbol{\gamma})$, (5.1.6) becomes

$$\int_\Sigma v^a L_{ab}(\boldsymbol{\gamma}) n^b dvol_{\boldsymbol{h}} = -\int_{M^-} \nabla^{(a}v^{b)} L_{ab}(\boldsymbol{\gamma}) dvol_{\boldsymbol{g}},$$

where we have relabelled indices and again used symmetry of L. ■

Now we use the pre-symplectic product (5.1.1) to re-express the right-hand side of (5.1.5).

Lemma 5.1.4 *With* $\boldsymbol{\gamma}$*,* $\boldsymbol{w}$ *and* $\boldsymbol{v}$ *as above,*

$$\omega_\Sigma(\boldsymbol{\gamma}, \pounds_{\boldsymbol{w}}\boldsymbol{g}) = -\int_{M^-} 2\nabla_{(a}v_{b)}L^{ab}(\gamma)dvol_{\boldsymbol{g}}.$$

Proof. Expanding the left-hand side using (5.1.1) gives

$$\omega_\Sigma(\boldsymbol{\gamma}, \pounds_{\boldsymbol{w}}\boldsymbol{g}) = \int_\Sigma n_a[2\nabla_{(b}w_{c)}\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\Pi^{abc}(\pounds_{\boldsymbol{w}}\boldsymbol{g})]dvol_{\boldsymbol{h}}.$$

As $\boldsymbol{w} = \boldsymbol{v}$ on a neighbourhood of Σ, then it is clear that

$$\int_\Sigma n_a[2\nabla_{(b}w_{c)}\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\Pi^{abc}(\pounds_{\boldsymbol{w}}\boldsymbol{g})]dvol_{\boldsymbol{h}} = \int_\Sigma n_a[2\nabla_{(b}v_{c)}\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g})]dvol_{\boldsymbol{h}}.$$

Now applying the Gauss Theorem on the region M^- with the vector field $2\nabla_{(b}v_{c)}\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g})$ gives

$$\begin{aligned}\int_\Sigma n_a[2\nabla_{(b}v_{c)}\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g})]dvol_{\boldsymbol{h}}& \\ = -\int_{M^-} \nabla_a[2\nabla_{(b}v_{c)}\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g})]dvol_{\boldsymbol{g}},& \quad (5.1.7)\end{aligned}$$

where we have again used that $\boldsymbol{v}$ vanishes to the far past of Σ. The integrand on the right-hand side of (5.1.7) is

$$\begin{aligned}\nabla_a(2\nabla_{(b}v_{c)}\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g})) &= 2\nabla_a\nabla_{(b}v_{c)}\Pi^{abc}(\boldsymbol{\gamma}) \\ &+ 2\nabla_{(b}v_{c)}\nabla_a\Pi^{abc}(\boldsymbol{\gamma}) - \nabla_a\gamma_{bc}\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g}) - \gamma_{bc}\nabla_a\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g}).\end{aligned}$$

Using (4.2.27) and the symmetries of T^{abcdef}, the first and third terms of this cancel. The remaining two terms are

$$\begin{aligned}2\nabla_{(b}v_{c)}\nabla_a\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\nabla_a\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g})& \\ = 2\nabla_{(b}v_{c)}(L^{bc}(\boldsymbol{\gamma}) - 2S^{bcde}\gamma_{de}) - \gamma_{bc}(L^{bc}(\pounds_{\boldsymbol{v}}\boldsymbol{g}) - 2S^{bcde}(\pounds_{\boldsymbol{v}}\boldsymbol{g})_{de},&\end{aligned}$$

where we have used the identity from Lemma 4.2.5. Finally, as $L^{bc}(\pounds_{\boldsymbol{v}}\boldsymbol{g}) = 0$ and using the symmetries of S^{bcde}, then

$$2\nabla_{(b}v_{c)}\nabla_a\Pi^{abc}(\boldsymbol{\gamma}) - \gamma_{bc}\nabla_a\Pi^{abc}(\pounds_{\boldsymbol{v}}\boldsymbol{g}) = 2\nabla_{(b}v_{c)}L^{bc}(\boldsymbol{\gamma}).$$

The combination of these results proves the lemma. ■

The proof of Theorem 5.1.2 is completed by combining Lemma 5.1.3 and Lemma 5.1.4. ■

This result puts us in a position to prove that pure gauge perturbations are degeneracies of the product (5.1.1) on the space of solutions to the linearized Einstein equation.

Lemma 5.1.5 $\hat{\mathscr{G}}(M;\mathbb{C})$ *is contained in the radical of* ω*.*

Proof. Suppose $\boldsymbol{w} \in C^\infty(T^1_0(M;\mathbb{C}))$ and $\boldsymbol{\gamma} \in \mathscr{S}(M;\mathbb{C})$, and let Σ be a smooth spacelike Cauchy surface. From Theorem 5.1.2 we have the identity

$$\omega(\boldsymbol{\gamma}, \pounds_{\boldsymbol{w}}\boldsymbol{g}) = 2\int_\Sigma w^b C^\Sigma_b(\mathrm{Data}_\Sigma(\boldsymbol{\gamma}))dvol_{\boldsymbol{h}}$$

and the right-hand side vanishes because $\boldsymbol{C}^\Sigma(\mathrm{Data}_\Sigma(\boldsymbol{\gamma})) = 0$. ■

The problem which now arises is to prove that $\hat{\mathscr{G}}(M;\mathbb{C})$ exhausts the space of degeneracies. For the case that $(M, \boldsymbol{g})$ admits a compact Cauchy surface we will show how this can be achieved, but for the case that $(M, \boldsymbol{g})$ does not admit a compact Cauchy surface, a proof that $\hat{\mathscr{G}}(M;\mathbb{C})$ exhausts the radical of ω has yet to be obtained. It is claimed in the footnote of [9, p. 59] that weak non-degeneracy holds provided the background spacetime does not admit Killing fields supported near spatial infinity, but no justification nor any references are ever given.

The proof, for the case that $(M, \boldsymbol{g})$ admits a compact Cauchy surface, that $\hat{\mathscr{G}}(M;\mathbb{C})$ exhausts the radical of ω is directly analogous to the case of electromagentism [28]. For that theory, one proves that all degeneracies are pure gauge by working with the Cauchy data for the vector potential $\boldsymbol{A}$, which are differential forms on the Cauchy surface. As the Cauchy surface is assumed compact, one can apply the Hodge decomposition to split the initial data into various pieces to prove [28, Prop. 5] that the degeneracies are only pure gauge.

However, recently, a new method [67] has been proposed which makes use of cohomology theory to prove that the degeneracies of the symplectic product of electromagnetism are still just pure gauge for the case of non-compact Cauchy surfaces, but with the initial data on the Cauchy surfaces still being compactly supported. Unfortunately, since there does not currently exist an analogue of cohomology theory for symmetric tensor fields, we will not be able to utilise this approach, and as such we will not discuss it any further here.

Therefore we are left with the same issues as [28], namely the existence of a suitable decomposition theorem. In fact, such a decomposition has been obtained by Moncrief [72], but it is computed within the Arnowitt-Deser-Misner (ADM) formalism [7] and is only valid for background spacetimes that solve the vacuum Einstein equation with vanishing cosmological constant. Therefore the task is to both reconcile the ADM approach with ours and generalise Moncrief's decomposition to the non-vanishing cosmological constant case.

We will show how both of these points can be achieved. In particular, the reconciliation of the ADM approach with ours will use the synchronous condition that was established earlier in section 4.5.3. The upshot of these results is that any degeneracy from our product induces a degeneracy in the corresponding ADM product, from which it can then be determined that the degeneracy is pure gauge. We now state the main theorem. However, its proof is postponed until after we have given a full description of the ADM formalism and our generalisation of the Moncrief decomposition.

Theorem 5.1.6 *If $(M, \boldsymbol{g})$ admits compact Cauchy surfaces then the radical of ω is precisely the subspace of pure gauge solutions $\mathscr{G}(M;\mathbb{C})$. That is, given $\boldsymbol{\gamma}' \in \mathscr{S}(M;\mathbb{C})$ such that $\omega(\boldsymbol{\gamma}', \boldsymbol{\gamma}) = 0$ for all $\boldsymbol{\gamma} \in \mathscr{S}(M;\mathbb{C})$, then $\boldsymbol{\gamma}' \in \hat{\mathscr{G}}(M;\mathbb{C})$.*

5.2 Results from the Arnowitt-Deser-Misner formalism

To prove, for the case that the background spacetime admits a compact Cauchy surface, that the space of degeneracies of the pre-symplectic product are pure gauge, we need to use results from the Arnowitt-Deser-Misner (ADM) formalism, which will now be introduced.

5.2.1 The Arnowitt-Deser-Misner formalism

The goal of the ADM formalism is to cast the general theory of relativity into a Hamiltonian form. This then facilitates the attainment of a quantum theory through canonical quantization; this quantum theory is known as canonical quantum gravity. We will not consider this theory here, though we shall just note that this approach has not proved fully successful as a quantum theory of gravity, yet it has spawned numerous other approaches. Most notably it led Ashtekar [8] to consider new variables for gravity, now known as Ashtekar variables, which subsequently led to the development of the theory of loop quantum gravity that currently has a large research community.

Although the ADM method was not the first approach to placing general relativity into a Hamiltonian form, see for instance Dirac's approach [30] and other references therein, it is the most widely recognised and used. We will now give a brief exposition of the ADM formalism based on the approach of Fischer and Marsden [41, Ch. 2] who use techniques from geometrical analysis to re-write the evolution equations in a more compact form, which uses the adjoint of the linearized constraint operator. These techniques are further utilised, as we will show, in the decomposition theorems of Moncrief. For the original ADM formulation and references, see the review article of Arnowitt, Deser and Misner [7].

Following [41, Sec. 2], let Σ be a three-dimensional smooth manifold without boundary that is assumed to be compact, connected, Hausdorff and orientable. One assumes that Σ can be embedded within a spacetime $(M, \boldsymbol{g})$ such that its image in M is spacelike with respect to $\boldsymbol{g}$. Now assume that on an open subset $U \subset M$ there exists a timelike vector field $\boldsymbol{t} \in C^\infty(T^1_0(U;\mathbb{R}))$ such that the level surfaces of the congruence of integral curves of this vector field are spacelike embeddings of Σ. Therefore we have a flow of time in spacetime and this parameterises the dynamics. Using the future-pointing unit normal vector $\boldsymbol{n}$ to the embedded hypersurface, one may decompose a vector field into its normal and tangential parts. The normal and tangential components of $\boldsymbol{t}$ are called [7], respectively, the lapse function and the shift vector field, and they characterise the slicing.

In this dynamical approach, the state of the spacetime metric will be specified by two quantities associated with the hypersurface. The first is the induced Riemannian metric $\boldsymbol{h} \in C^\infty(S^0_2(\Sigma;\mathbb{R}))$, also known as the first fundamental form; it will act as a 'position' variable. The spacetime metric $\boldsymbol{g}$ can be written purely in terms of the lapse function, shift vector field and the induced Riemannian metric. Note that the lapse function and shift vector field are non-dynamical entities, in that they merely specify the slicing used to describe the evolution, and are freely specifiable.

The other variable is $\boldsymbol{h}$'s canonically conjugate momentum $\boldsymbol{\varpi} \in C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R}))$, where $\tilde{S}^2_0(\Sigma;\mathbb{R})$ denotes the space of symmetric, second rank contravariant tensor densities on Σ. We will now describe what this momentum is and how it arises. The hypersurface will possess a second fundamental form or extrinsic curvature $\boldsymbol{k}$, which is a symmetric covariant two-tensor. It characterises how the surface has been embedded within the ambient space by measuring the difference between the actions of the connection associated with the ambient spacetime metric and the connection associated with the induced Riemannian metric of the surface, see [24, pp. 312-315] for further details. In our case, we choose $k_{ab} = q^c{}_a q^d{}_b \nabla_{(c} n_{d)}$, where $\boldsymbol{n}$ is as above and $q^c{}_a$ projects tensors onto the hypersurface; this convention agrees with [91, eq. (10.2.13)] but is opposite to the convention selected in [41, p. 328]. One may expand the Einstein-Hilbert action (4.2.25) in terms of the quantities of the hypersurface. From this one finds that the canonically conjugate momentum to $\boldsymbol{h}$ is given by the tensor density

$$\varpi^{ab} = \sqrt{h}\left(k^{ab} - \frac{1}{2}h^{ab}k\right).$$

This corresponds to minus the momentum of Fischer and Marsden [41], but recall that their extrinsic curvature convention is minus ours, so in fact the two momentums agree.

The two quantities $(\boldsymbol{h}, \boldsymbol{\varpi})$ evolve according to the ADM equations, which will be stated later, see (5.2.7), in a form first noted by Fischer and Marsden [39, p. 917]. The initial data for this system is a pair $(\boldsymbol{h}, \boldsymbol{\varpi}) \in C^\infty(S^0_2(\Sigma; \mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma; \mathbb{R}))$, but one cannot freely choose arbitrary data. If the data are to determine a solution of the Einstein equation, then the data must satisfy certain initial value constraints. These constraints are given by the map $\Phi : C^\infty(S^0_2(\Sigma; \mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma; \mathbb{R})) \to C^\infty(\Sigma; \mathbb{R}) \times C^\infty(T^1_0(\Sigma; \mathbb{R}))$, where

$$\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) = (\mathscr{H}(\boldsymbol{h}, \boldsymbol{\varpi}), \boldsymbol{\delta}(\boldsymbol{h}, \boldsymbol{\varpi})) \tag{5.2.1}$$

and the Hamiltonian and momentum constraints are

$$\mathscr{H}(\boldsymbol{h}, \boldsymbol{\varpi}) = -R^{(3)}(\boldsymbol{h}) + \frac{\varpi^{ab}\varpi_{ab}}{h} - \frac{\varpi^2}{2h} + 2\Lambda$$

and

$$\delta^a(\boldsymbol{h}, \boldsymbol{\varpi}) = D_b\left(\frac{\varpi^{ab}}{\sqrt{h}}\right)$$

respectively. Here $R^{(3)}(\boldsymbol{h})$ is the Ricci scalar for the metric $\boldsymbol{h}$ and D_a is the covariant derivative associated with $\boldsymbol{h}$. Together, the ADM equations (5.2.7) and the constraint Φ are equivalent [41, Thm 2.1], if Σ is compact, to the Einstein equation $G_{ab} + \Lambda g_{ab} = 0$ holding on the region covered by the slicing.

We now discuss how the linearized theory is treated in the ADM formalism. One linearizes the system using standard techniques (discussed in section 4.2), and obtains quantities that are analogous to the background ADM ones. A linear perturbation $\boldsymbol{\gamma}$ can be split up into a linearized lapse function, linearized shift vector field and a three-perturbation $\boldsymbol{\gamma}^{(3)} \in C^\infty(S^0_2(\Sigma;\mathbb{R}))$. The Cauchy data for $\boldsymbol{\gamma}$ consists of $(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) = \left(\frac{\partial \boldsymbol{h}(\lambda)}{\partial\lambda}, \frac{\partial \boldsymbol{\varpi}(\lambda)}{\partial\lambda}\right)_{\lambda=0}$, where $(\boldsymbol{h}(\lambda), \boldsymbol{\varpi}(\lambda))$ are a one parameter family of data for the non-linear equations. One also has linearized constraints and a system of linearized ADM equations, which together are equivalent [41, Thm 4.5] (for the case that Σ is compact) to the linearized Einstein equation (4.2.4) holding. One should note that to solve the linearized ADM equations one needs to specify a linearized lapse function and linearized shift vector field; as in the background case, they are non-dynamical and freely specifiable.

From now on we assume that the background is a solution to the vacuum Einstein equation with cosmological constant, so $\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) = 0$, where $(\boldsymbol{h}, \boldsymbol{\varpi})$ are the Cauchy data for the background spacetime. The linearized constraints are the derivative, evaluated at $(\boldsymbol{h}, \boldsymbol{\varpi})$, of the constraint map (5.2.1),

$$D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) : C^\infty(S^0_2(\Sigma;\mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R})) \to C^\infty(\Sigma;\mathbb{R}) \times C^\infty(T^1_0(\Sigma;\mathbb{R})),$$

where

$$D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) = \big(D\mathscr{H}(\boldsymbol{h}, \boldsymbol{\varpi})(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}), D\boldsymbol{\delta}(\boldsymbol{h}, \boldsymbol{\varpi})(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p})\big) \tag{5.2.2}$$

and these components are

$$\begin{aligned} D\mathscr{H}(\boldsymbol{h}, \boldsymbol{\varpi})(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) = \frac{1}{h}\Bigg[&-\left(\varpi^{ab}\varpi_{ab} - \frac{1}{2}\varpi^2\right)\gamma^{(3)} + 2\left(\varpi_{ab}p^{ab} - \frac{1}{2}\varpi p\right) \\ &+2\left(\varpi^{ac}\varpi_{cb} - \frac{1}{2}\varpi\varpi^{ab}\right)\gamma^{(3)}_{ab}\Bigg] - \left(D^aD^b\gamma^{(3)}_{ab} - D^aD_a\gamma^{(3)} - R^{(3)ab}\gamma^{(3)}_{ab}\right) \end{aligned} \tag{5.2.3}$$

and

$$D\boldsymbol{\delta}(\boldsymbol{h}, \boldsymbol{\varpi})(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) = \frac{1}{\sqrt{h}}\left[2D_bp^{ab} + \varpi^{bc}\left(D_c\gamma^{(3)a}_{\ \ b} + D_b\gamma^{(3)a}_{\ \ c} - D^a\gamma^{(3)}_{bc}\right)\right],$$

where $\gamma^{(3)} = h^{ab}\gamma^{(3)}_{ab}$, $\varpi = h_{ab}\varpi^{ab}$ and $p = h_{ab}p^{ab}$. To get the components of (5.2.2) into the form of those in [41, 72], we evaluate (5.2.3) on the constraint surface $\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) = 0$, which gives

$$D\mathscr{H}(\boldsymbol{h},\boldsymbol{\varpi})\begin{pmatrix}\boldsymbol{\gamma}^{(3)}\\ \boldsymbol{p}\end{pmatrix} = \frac{1}{h}\left[-\frac{1}{2}\left(\varpi^{ab}\varpi_{ab}-\frac{1}{2}\varpi^2\right)\gamma^{(3)}+2\left(\varpi_{ab}p^{ab}-\frac{1}{2}\varpi p\right)\right.$$
$$\left.+2\left(\varpi^{ac}\varpi_c{}^b-\frac{1}{2}\varpi\varpi^{ab}\right)\gamma^{(3)}_{ab}\right]-\left[D^aD^b\gamma^{(3)}_{ab}-D^aD_a\gamma^{(3)}\right.$$
$$\left.-\left(R^{(3)ab}-\frac{1}{2}h^{ab}R^{(3)}+\Lambda h^{ab}\right)\gamma^{(3)}_{ab}\right].$$

Note that the difference between this and the $\Lambda = 0$ case considered in equation (2.8) of [72] is just the cosmological constant term.

Following [72], see equations (2.4) and (2.6) of that reference, we now introduce two inner products, which will allow one to calculate an adjoint to $D\Phi(\boldsymbol{h},\boldsymbol{\varpi})$. The first product acts on the vector space $C^\infty(S^0_2(\Sigma;\mathbb{R}))\times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R}))$, whereby, given $(\boldsymbol{\gamma}^{(3)},\boldsymbol{p}),(\tilde{\boldsymbol{\gamma}}^{(3)},\tilde{\boldsymbol{p}})\in C^\infty(S^0_2(\Sigma;\mathbb{R}))\times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R}))$, one defines their inner product by

$$\langle(\boldsymbol{\gamma}^{(3)},\boldsymbol{p});(\tilde{\boldsymbol{\gamma}}^{(3)},\tilde{\boldsymbol{p}})\rangle := \int_\Sigma\left(\sqrt{h}\gamma^{(3)}_{ab}\tilde{\gamma}^{(3)}_{cd}h^{ac}h^{bd}+\frac{1}{\sqrt{h}}h_{ac}h_{bd}p^{ab}p^{cd}\right)d^3x. \tag{5.2.4}$$

The second product acts on elements in $C^\infty(\Sigma;\mathbb{R})\times C^\infty(T^1_0(\Sigma;\mathbb{R}))$; given $f,\tilde{f}\in C^\infty(\Sigma;\mathbb{R})$ and $\boldsymbol{V},\tilde{\boldsymbol{V}}\in C^\infty(T^1_0(\Sigma;\mathbb{R}))$, then

$$\langle\!\langle(f,\boldsymbol{V});(\tilde{f},\tilde{\boldsymbol{V}})\rangle\!\rangle := \int_\Sigma(f\cdot\tilde{f}+h_{ab}V^aV^b)dvol_{\boldsymbol{h}}. \tag{5.2.5}$$

One may now calculate the adjoint $D\Phi^*(\boldsymbol{h},\boldsymbol{\varpi})$ of the differential operator $D\Phi(\boldsymbol{h},\boldsymbol{\varpi})$ with respect to these inner products (5.2.4) and (5.2.5). Specifically we have

$$\langle\langle(f,\boldsymbol{V}),D\Phi(\boldsymbol{h},\boldsymbol{\varpi})(\boldsymbol{\gamma}^{(3)},\boldsymbol{p})\rangle\rangle = \langle D\Phi^*(\boldsymbol{h},\boldsymbol{\varpi})(f,\boldsymbol{V}),(\boldsymbol{\gamma}^{(3)},\boldsymbol{p})\rangle. \tag{5.2.6}$$

Using integration by parts, one finds that

$$D\Phi^*(\boldsymbol{h},\boldsymbol{\varpi})(f,\boldsymbol{V}) = (D\mathscr{H}^*(\boldsymbol{h},\boldsymbol{\varpi})(f),D\boldsymbol{\delta}^*(\boldsymbol{h},\boldsymbol{\varpi})(\boldsymbol{V})),$$

where $D\mathscr{H}(\boldsymbol{h},\boldsymbol{\varpi})^*(f) = (\boldsymbol{\alpha},\boldsymbol{\beta})$ with

$$\alpha_{ab} = \frac{1}{h}\left[-\frac{1}{2}\left(\varpi^{cd}\varpi_{cd} - \frac{1}{2}\varpi^2\right)h_{ab}f \ +2\left(\varpi_{ac}\varpi^c{}_b - \frac{1}{2}\varpi_{ab}\varpi\right)f\right] - \left[D_aD_bf - h_{ab}D^cD_cf - \left(R^{(3)}_{ab} - \frac{1}{2}h_{ab}R^{(3)} + \Lambda h_{ab}\right)f\right]$$

and

$$\beta^{ab} = 2f\left(\varpi^{ab} - \frac{1}{2}\varpi h^{ab}\right).$$

The final component of $D\Phi^*(\boldsymbol{h},\boldsymbol{\varpi})$ is calculated to be

$$D\boldsymbol{\delta}^*(\boldsymbol{h},\boldsymbol{\varpi})(\boldsymbol{V}) = \left(\frac{1}{\sqrt{h}}\left(D_c(V^c\varpi_{ab}) - 2\varpi^c{}_{(a}D_{|c|}V_{b)}\right), -\sqrt{h}(D^aV^b + D^bV^a)\right).$$

Again, the difference between this and the $\Lambda = 0$ case, see equation (2.10) of [72], is the cosmological constant term present in $\boldsymbol{\alpha}$.

We introduce a unitary operator $U : C^\infty(S^0_2(\Sigma;\mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R})) \to C^\infty(S^0_2(\Sigma;\mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R}))$, defined by

$$U(\boldsymbol{\gamma}^{(3)},\boldsymbol{p}) := \left(\frac{-1}{\sqrt{h}}\boldsymbol{p}^{\flat\flat}, \sqrt{h}(\gamma^{(3)})^{\sharp\sharp}\right).$$

This operator is unitary with respect to the inner product $\langle\cdot,\cdot\rangle$. The reason for introducing this operator is so that $U \circ D\Phi^*(\boldsymbol{h},\boldsymbol{\varpi})$ corresponds to the operator $\gamma(\boldsymbol{h},\boldsymbol{\varpi}) \equiv \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \circ D\Phi(\boldsymbol{h},\boldsymbol{\varpi})^\dagger$ from [72, p. 1558], where $D\Phi(\boldsymbol{h},\boldsymbol{\varpi})^\dagger$ is the 'new form of the adjoint' defined in equation (4.2) of that reference. The inverse map $U^{-1} : C^\infty(S^0_2(\Sigma;\mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R})) \to C^\infty(S^0_2(\Sigma;\mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R}))$ is given by

$$U^{-1}(\boldsymbol{\gamma}^{(3)},\boldsymbol{p}) = \left(\frac{1}{\sqrt{h}}\boldsymbol{p}^{\flat\flat}, -\sqrt{h}(\boldsymbol{\gamma}^{(3)})^{\sharp\sharp}\right).$$

The ADM evolution equations, which govern how the background data $(\boldsymbol{h},\boldsymbol{\varpi})$ evolve according to the slicing parameter, may be written as

$$\frac{\partial}{\partial\lambda}\left(\boldsymbol{h}(\lambda),\boldsymbol{\varpi}(\lambda)\right) = U^{-1} \circ D\Phi^*(\boldsymbol{h},\boldsymbol{\varpi})(N,-\boldsymbol{N}), \tag{5.2.7}$$

where the lapse function N and the shift vector field $\boldsymbol{N}$ of the chosen slicing are arbitrary.

The space $C^\infty(S_2^0(\Sigma;\mathbb{R}))\times C^\infty(\tilde{S}_0^2(\Sigma;\mathbb{R}))$ carries a natural symplectic product, which will be known as the ADM symplectic product. On the background $(\boldsymbol{h},\boldsymbol{\varpi})$ it is, see [41, p. 333], given by

$$\omega^{ADM}_{(\boldsymbol{h},\boldsymbol{\varpi})}((\boldsymbol{\gamma}^{(3)},\boldsymbol{p});(\tilde{\boldsymbol{\gamma}}^{(3)},\tilde{\boldsymbol{p}}))=\int_\Sigma(\gamma^{(3)}_{ab}\tilde{p}^{ab}-\tilde{\gamma}^{(3)}_{ab}p^{ab})d^3x.$$

Observe that

$$\omega^{ADM}_{(\boldsymbol{h},\boldsymbol{\varpi})}((\boldsymbol{\gamma}^{(3)},\boldsymbol{p});(\tilde{\boldsymbol{\gamma}}^{(3)},\tilde{\boldsymbol{p}}))=\langle(\boldsymbol{\gamma}^{(3)},\boldsymbol{p});U^{-1}(\tilde{\boldsymbol{\gamma}}^{(3)},\tilde{\boldsymbol{p}}))\rangle. \tag{5.2.8}$$

Note that the form of the linearized ADM equations is not required here and so we do not include them, but refer the reader to [41, p. 357].

5.2.2 A generalisation of Moncrief's splitting theorems

We now consider a result due to Moncrief, which shows how the space of initial data, $C^\infty(S_2^0(\Sigma;\mathbb{R}))\times C^\infty(\tilde{S}_0^2(\Sigma;\mathbb{R}))$, for the linearized ADM equations may be decomposed into three distinct subspaces. The upshot of this is that it allows one to show that on the space of initial data satisfying the linearized constraints, the subspace of degeneracies of the ADM symplectic product consists entirely of pure gauge initial data. This result will be adapted to show that the same is true for our symplectic product.

In [72] it is shown that, if the background spacetime obeys the vacuum Einstein equation with vanishing cosmological constant and if it admits a compact Cauchy surface Σ, then the space of initial data to the linearized ADM equations can be decomposed, with respect to the inner product (5.2.4), into three orthogonal subspaces. The key point of this decomposition is that it preserves smoothness, that is, each subspace consists entirely of smooth fields. The preservation of smoothness is achieved by the properties of elliptic operators, which is why the compactness criterion arises. We will now describe the splitting and its generalisation to include the case of a vacuum spacetime with a non-zero cosmological constant. This will mean that the decomposition will be valid for the entire class of spacetimes that we consider.

The first splitting is

$$C^\infty(S_2^0(\Sigma;\mathbb{R}))\times C^\infty(\tilde{S}_0^2(\Sigma;\mathbb{R}))=\ker D\Phi(\boldsymbol{h},\boldsymbol{\varpi})\oplus\operatorname{range} D\Phi^*(\boldsymbol{h},\boldsymbol{\varpi}), \tag{5.2.9}$$

where $\ker D\Phi(\boldsymbol{h},\boldsymbol{\varpi})$ is the subspace of data satisfying the linearized constraints and $\operatorname{range} D\Phi^*(\boldsymbol{h},\boldsymbol{\varpi})$ is the unphysical data. Both subspaces consist entirely of elements which are smooth. We now briefly elaborate on how this is achieved. Orthogonality of the subspaces follows directly from (5.2.6). Therefore one is left with showing that an arbitrary $(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) \in C^\infty(S_2^0(\Sigma;\mathbb{R})) \times C^\infty(\tilde{S}_0^2(\Sigma;\mathbb{R}))$ may be uniquely decomposed into

$$(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) = (\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}}) + D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})(C, \boldsymbol{X}), \tag{5.2.10}$$

where $(\widetilde{\boldsymbol{\gamma}^{(3)}}, \tilde{\boldsymbol{p}}) \in \ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$ and $(C, \boldsymbol{X}) \in C^\infty(\Sigma;\mathbb{R}) \times C^\infty(T_0^1(\Sigma;\mathbb{R}))$. By applying $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$ to both sides of this equation, one obtains a partial differential equation for $(C, \boldsymbol{X})$,

$$D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) = D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})(C, \boldsymbol{X}). \tag{5.2.11}$$

If a unique solution $(C, \boldsymbol{X})$ exists, then the splitting will be proved. To obtain a solution, the theory of elliptic operators will be used, in particular, the results of [12] who discuss the properties of differential operators on compact Riemannian manifolds. For the case of $\Lambda = 0$, see [72, Sec. 3], the operator $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi}) : C^\infty(\Sigma;\mathbb{R}) \times C^\infty(T_0^1(\Sigma;\mathbb{R})) \to C^\infty(\Sigma;\mathbb{R}) \times C^\infty(T_0^1(\Sigma;\mathbb{R}))$ is elliptic [12, Lem. 4.4] because $D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$ has injective principal symbol[1]. In the $\Lambda \neq 0$ case, our modifications to the linearized constraint map $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$ and its adjoint $D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$ only results in the addition of a $\Lambda \boldsymbol{h}$ term in both cases. Hence, this does not introduce any further second-order derivative terms and so the principal symbol of $D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$ will be unaffected. Therefore the operator $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$ is still elliptic, and so by [12, Thm 4.3] one has the following decomposition,

$$\begin{aligned} C^\infty(\Sigma;\mathbb{R}) \times C^\infty(T_0^1(\Sigma;\mathbb{R})) &= \operatorname{range}(D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})) \oplus \ker(D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})) \\ &= \operatorname{range}(D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})) \oplus \ker D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi}). \end{aligned} \tag{5.2.12}$$

The remainder of Moncrief's argument [72, Sec. 3] shows that the source term (the left-hand side of (5.2.11)) lies in $\operatorname{range}(D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi}))$, which is because the source is evidently an element of $\operatorname{range} D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$, but this space is orthogonal to $\ker D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$, and so by (5.2.12) it lies in $\operatorname{range}(D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi}))$. Hence a solution $(C, \boldsymbol{X})$ exists and is unique up to an element of $\ker D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$. Since any element of $\ker D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$ does not affect the split (5.2.10), the splitting itself will be unique even though the solution to (5.2.11) might not be. Therefore the first Moncrief decomposition (5.2.9) also holds for cosmological vacuum spacetimes too.

[1] The principal symbol of a second-order differential operator, such as $D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$, is discussed in section 2.4.1 of this thesis.

The second splitting decomposes the constraint subspace, $\ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$, into a pure gauge subspace, meaning data for pure gauge solutions, and a physical subspace. In [73, Sec. IV] it is shown that the initial data for a pure gauge solution $\pounds_{\boldsymbol{w}}\boldsymbol{g}$ to the linearized equations, on a vacuum spacetime with $\Lambda = 0$, is given by

$$(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p})_{\text{gauge}} = U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})(C, \boldsymbol{X}), \tag{5.2.13}$$

where $C = n_a w^a$ and $X^a = q^a{}_b w^b$ are respectively the normal (with respect to the future-pointing unit normal vector $\boldsymbol{n}$) and tangential projections, relative to Σ (using the associated projection tensor $q^a{}_b$), of the gauge vector field. This result was initially proved via a lengthy calculation, and later by more geometrical methods [41, Thm 4.7] using the adjoint form of the ADM equations (5.2.7). By following the methodology given in [41, Thm 4.7] except that one uses the ADM equations (5.2.7), which include the $\Lambda \neq 0$ case, then the result continues to hold for cosmological vacuum spacetimes as well.

Before performing the final split, one also needs to check that the pure gauge subspace actually lies in the constraint subspace. Again, one could check this by the means of a lengthy calculation, as was done in [72, Thm 4.1] for the $\Lambda = 0$ case; instead, we appeal to the geometrical method of [41, Prop. 3.2], which again, just like the expression for the pure gauge initial data, utilises the adjoint form of the ADM evolution equations. By using the ADM equations (5.2.7) that include the $\Lambda \neq 0$ case and the methodology of [41, Prop. 3.2], then it remains true for all cosmological vacuum spacetimes that the subspace of pure gauge data is contained within the subspace of data which satisfy the linearized constraints.

With the two preceding results and, as argued earlier, ellipticity of $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})\circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$ unaffected by the addition of a cosmological constant, the subspace $\ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$ can be decomposed into

$$\ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) = \operatorname{range}(U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})) \oplus \ker((U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi}))^* \cap \ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}),$$

where the first space is pure gauge and the second space is the physical space. To prove this, one first needs to obtain the orthogonal complement, within $\ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$, to the pure gauge space, $\operatorname{range}(U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi}))$. Let $(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) \in \ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$ be orthogonal to all pure gauge data, therefore, for all $(C, \boldsymbol{X}) \in C^\infty(\Sigma;\mathbb{R}) \times C^\infty(T^1_0(\Sigma;\mathbb{R}))$,

$$\begin{aligned} 0 &= \langle (\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}), U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})(C, \boldsymbol{X})\rangle \\ &= \langle U^{-1}(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}), D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})(C, \boldsymbol{X})\rangle \\ &= \langle\langle D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})(U^{-1}(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p})), (C, \boldsymbol{X})\rangle\rangle. \end{aligned}$$

By non-degeneracy of $\langle\langle\cdot,\cdot\rangle\rangle$, this entails that $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})(U^{-1}(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p})) = 0$. Hence, one seeks to split a $(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) \in \ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$ as

$$(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) = (\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}}) + U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})(C, \boldsymbol{X})$$

such that $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})(U^{-1}(\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}})) = 0$. Applying the operator $D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ U$ to this decomposition, one obtains

$$D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})(U^{-1}(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p})) = D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})(C, \boldsymbol{X}). \tag{5.2.14}$$

Using exactly the same argument as for the first split and noting that $\operatorname{range}(D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}) \circ U^{-1})$ is always orthogonal to $\ker D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$, then the unique splitting exists for the $\Lambda \neq 0$ case as well.

Therefore the final decomposition of the initial data is

$$\begin{aligned} C^\infty(S^0_2(\Sigma;\mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R})) = \operatorname{range} D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi}) &\oplus \operatorname{range}(U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})) \\ &\oplus \ker((U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi}))^* \cap \ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi}), \end{aligned}$$

which takes the same form as the $\Lambda = 0$ case from [72, Thm 4.2]. This decomposition allows one to prove that on the space of initial data obeying the constraints, the only degeneracies of the ADM symplectic product are pure gauge. We will now show this by giving the analogue, for our $\Lambda \neq 0$ case, of [40, Prop. 4.38].

Theorem 5.2.1 *The ADM symplectic orthogonal complement to the subspace* $\ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$ *is the pure gauge space* $\operatorname{range}(U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})) \subset \ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$.

Proof. Let $(\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}}) \in C^\infty(S^0_2(\Sigma;\mathbb{R}) \times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R}))$ satisfy $\omega^{ADM}_{(\boldsymbol{h},\boldsymbol{\varpi})}((\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}); (\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}})) = 0$ for all $(\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) \in \ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$. Then by (5.2.8),

$$\langle (\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}); U^{-1}(\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}})\rangle\rangle = 0 \tag{5.2.15}$$

and so $U^{-1}(\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}})$ is orthogonal to $\ker D\Phi(\boldsymbol{h}, \boldsymbol{\varpi})$. By the first Moncrief split (5.2.9) this means that $U^{-1}(\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}}) \in \operatorname{range} D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})$. Hence,

$$(\tilde{\boldsymbol{\gamma}}^{(3)}, \tilde{\boldsymbol{p}}) \in \operatorname{range}(U \circ D\Phi^*(\boldsymbol{h}, \boldsymbol{\varpi})) \tag{5.2.16}$$

and is therefore pure gauge. ■

5.3 Symplectic space

We are now in a position to prove Theorem 5.1.6.

5.3.1 Proof of Theorem 5.1.6

The main issue is to translate Theorem 5.2.1 into the setting studied in the main body of the paper. Begin by taking an arbitrary smooth spacelike Cauchy surface Σ and denote by $\mathscr{N}$ a normal neighbourhood of Σ. (For details about normal neighbourhoods, see section 4.5.3.)

Any element of $\mathscr{S}(M;\mathbb{C})$ can be split into its real and imaginary parts, in particular, the degeneracy can be split. As a degeneracy has vanishing pre-symplectic product with every element of $\mathscr{S}(M;\mathbb{C})$, then if we examine its behaviour against purely real or purely imaginary solutions, the form of both the real and imaginary parts of the degeneracy can be found by discovering the form of a degeneracy on just the space of real solutions. We will show, using the ADM results, that a degeneracy for the case of real solutions has to be pure gauge and so our complex degeneracy is pure gauge too.

Therefore without loss of generality, assume that the solution $\boldsymbol{\gamma}' \in \mathscr{S}(M;\mathbb{R})$ is a degeneracy of the symplectic form ω, that is, $\omega(\boldsymbol{\gamma}', \boldsymbol{\gamma}) = 0$ for all $\boldsymbol{\gamma} \in \mathscr{S}(M;\mathbb{R})$. Also, without loss of generality, $\boldsymbol{\gamma}'$ may be chosen synchronous near Σ; Theorem 4.5.6 entails that we may gauge transform any solution to the synchronous gauge near Σ, and since, by Lemma 5.1.5, pure gauge is a degeneracy, then $\boldsymbol{\gamma}'$ will still be a degeneracy of ω. It will also be sufficient to restrict attention to synchronous $\boldsymbol{\gamma}$ as well.

We now restrict attention to the normal neighbourhood $\mathscr{N}$, where we can introduce Gaussian normal coordinates. In such coordinates the spacetime metric takes the form $\boldsymbol{g} = -dt \otimes dt + \tilde{h}_{ij} dx^i \otimes dx^j$ and the synchronous gauge condition is precisely $\gamma_{0\mu} = 0$. The solutions $\boldsymbol{\gamma}', \boldsymbol{\gamma}$ correspond to solutions to the linearized ADM equations about the background $(\mathscr{N}, \boldsymbol{g}|_{\mathscr{N}})$ in the slicing given by the Gaussian normal coordinates: thus we have unit lapse, vanishing shift (and vanishing linearizations thereof). The corresponding ADM Cauchy data are $(\boldsymbol{\gamma}'^{(3)}, \boldsymbol{p}'), (\boldsymbol{\gamma}^{(3)}, \boldsymbol{p}) \in C^\infty(S^0_2(\Sigma;\mathbb{R})) \times C^\infty(\tilde{S}^2_0(\Sigma;\mathbb{R}))$ respectively, where in these coordinates

$$\gamma^{(3)}_{ij} = \gamma_{ij}|_\Sigma, \tag{5.3.1}$$

$$p^{ij} = \sqrt{h}\left(\frac{\gamma^{(3)}}{4}\left(h^{im}h^{jn} - h^{ij}h^{mn}\right) - \frac{1}{2}\left(\gamma^{im}_{(3)}h^{jn} + h^{im}\gamma^{jn}_{(3)} - \gamma^{ij}_{(3)}h^{mn} - h^{ij}\gamma^{mn}_{(3)}\right)\right)\partial_0 h_{mn} + \frac{\sqrt{h}}{2}\left(h^{im}h^{jn} - h^{ij}h^{mn}\right)\partial_0\gamma^{(3)}_{mn}, \tag{5.3.2}$$

and

$$\pi^{ij}|_\Sigma = \frac{1}{2}\left(h^{im}h^{jn} - h^{ij}h^{mn}\right)\partial_0\gamma^{(3)}_{mn} + \frac{1}{4}h^{ij}h^{mn}h^{kl}(\partial_0 h_{nl})\gamma^{(3)}_{mk}. \tag{5.3.3}$$

We see that the expressions for $\frac{\boldsymbol{p}}{\sqrt{h}}$ and $\boldsymbol{\pi}|_\Sigma$ do not coincide and there is no obvious connection between them. (However, on Minkowski spacetime, in global inertial coordinates, the two expressions do coincide.)

Using these results we have

$$\omega(\boldsymbol{\gamma}', \boldsymbol{\gamma}) = \int_\Sigma (\gamma'_{ij}\pi^{ij} - \gamma_{ij}\pi'^{ij})\sqrt{h}d^3x.$$

If one expands the out the integrand using (5.3.3) and compares this, using (5.3.2), to the expansion of the integrand, with the $\sqrt{h}$ removed, of

$$\omega^{ADM}_{(\boldsymbol{h},\boldsymbol{\varpi})}((\boldsymbol{\gamma}'^{(3)}, \boldsymbol{p}'); (\boldsymbol{\gamma}^{(3)}\boldsymbol{p})) = \int_\Sigma (\gamma'_{ij}p^{ij} - \gamma_{ij}p'^{ij})d^3x,$$

then one finds that they are equal. Hence, we have

$$\omega(\boldsymbol{\gamma}', \boldsymbol{\gamma}) = \omega^{ADM}_{(\boldsymbol{h},\boldsymbol{\varpi})}((\boldsymbol{\gamma}'^{(3)}, \boldsymbol{p}'); (\boldsymbol{\gamma}^{(3)}\boldsymbol{p})). \tag{5.3.4}$$

The result of this is that if a solution $\boldsymbol{\gamma}'$ is a degeneracy for ω, then its ADM Cauchy data $(\boldsymbol{\gamma}'^{(3)}, \boldsymbol{p}')$ will be a degeneracy for ω^{ADM} on the subspace of initial data obeying the constraints. Hence, by Theorem 5.2.1, $(\boldsymbol{\gamma}'^{(3)}, \boldsymbol{p}')$ is data for a pure gauge solution, and so on the region $\mathscr{N}$, $\boldsymbol{\gamma}' = \pounds_{\boldsymbol{w}}\boldsymbol{g}$ for some $\boldsymbol{w} \in C^\infty(T^1_0(\mathscr{N};\mathbb{R}))$. Now perform a global gauge transformation on $\boldsymbol{\gamma}'$ using a vector field $\boldsymbol{v} \in C^\infty(T^1_0(M;\mathbb{R}))$, which satisfies $\boldsymbol{v} = -\boldsymbol{w}$ on an open neighbourhood of Σ within $\mathscr{N}$. The result will still be both a solution and a degeneracy in $\mathscr{S}(M;\mathbb{R})$, but it satisfies $\mathrm{Data}_\Sigma(\boldsymbol{\gamma}' - \pounds_{\boldsymbol{v}}\boldsymbol{g}) = (0,0)$ and therefore by

Theorem 4.6.3, $\boldsymbol{\gamma}' = \pounds_{\boldsymbol{u}}\boldsymbol{g}$ for some $\boldsymbol{u} \in C^\infty(T^1_0(M;\mathbb{R}))$. Note that due to the compactness of Σ, all three vector fields $\boldsymbol{w}$, $\boldsymbol{v}$ and $\boldsymbol{u}$ will be spacelike-compact and hence so will their associated pure gauge perturbation. Therefore $\boldsymbol{\gamma}' \in \mathscr{G}(M;\mathbb{R}) = \hat{\mathscr{G}}(M;\mathbb{R})$. ■

5.3.2 Phase space for linearized gravity

It has just been proven that for the case that the background spacetime admits a compact Cauchy surface, the space of degeneracies consists only of the pure gauge perturbations. Unfortunately, we do not, as yet, possess a proof for the non-compact case, and so we conjecture that this result also holds for the non-compact case as well.

If the space of pure gauge perturbations $\hat{\mathscr{G}}(M;\mathbb{C})$ is the space of degeneracies, then the phase space of the theory is obtained by quotienting the space of solutions by this subspace. This entails that the phase space is the quotient vector space

$$\mathscr{P}_{\mathbb{C}}(M) := \mathscr{S}(M;\mathbb{C})/\hat{\mathscr{G}}(M;\mathbb{C}), \tag{5.3.5}$$

consisting of gauge equivalence classes $[\boldsymbol{\gamma}]$ of solutions to the linearized Einstein equation. On $\mathscr{P}_{\mathbb{C}}(M)$, we have a weakly non-degenerate[2] symplectic product,

$$\omega([\boldsymbol{\gamma}^1],[\boldsymbol{\gamma}^2]) = \int_\Sigma (\gamma^1_{ab}\pi_2^{ab} - \gamma^2_{ab}\pi_1^{ab})dvol_{\boldsymbol{h}}. \tag{5.3.6}$$

Note that under complex conjugation, this satisfies

$$\omega([\boldsymbol{\gamma}^1],[\boldsymbol{\gamma}^2])^* = \omega([\boldsymbol{\gamma}^{1*}],[\boldsymbol{\gamma}^{2*}]),$$

and the real phase space $\mathscr{P}_{\mathbb{R}}(M)$ is obtained by restricting attention to real-valued solutions and real-valued gauge transformations.

The right-hand side of (5.3.6) is independent of the choice of representative from the equivalence class, therefore we may freely select a de Donder representative in each case. In fact, the ability to select de Donder representatives will be exploited to establish Theorem 5.3.3. We begin by introducing the differential operator $\mathscr{D} : C^\infty(T^0_2(M;\mathbb{C})) \to C^\infty(T^3_0(M;\mathbb{C}))$ whose action on an arbitrary $\boldsymbol{\beta} \in C^\infty(T^0_2(M;\mathbb{C}))$ is given by

$$\mathscr{D}^{cab}(\boldsymbol{\beta}) := \frac{1}{2}\nabla^c\beta^{ab} - \frac{1}{2}\nabla^b\beta^{ca} - \frac{1}{2}\nabla^a\beta^{cb}. \tag{5.3.7}$$

[2]Weak non-degeneracy entails that if $\omega([\gamma],[\gamma']) = 0$ for all $[\gamma'] \in \mathscr{P}_{\mathbb{C}}(M)$, then $[\gamma] = [0]$.

For the case of de Donder perturbations, the divergence of this differential operator reduces to a rather elegant result.

Lemma 5.3.1 *If $\boldsymbol{\gamma} \in C^\infty(S_2^0(M;\mathbb{C}))$ satisfies the de Donder condition, $\nabla^a \overline{\gamma}_{ab} = 0$, and the background spacetime is a cosmological vacuum solution, then we have*

$$\nabla_c \mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}}) = \frac{1}{2} P^{abcd} \overline{\gamma}_{cd} - \Lambda \overline{\gamma}^{ab}. \tag{5.3.8}$$

Proof. Expanding out the left-hand side gives

$$\nabla_c \mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}}) = \frac{1}{2} \nabla_c \nabla^c \overline{\gamma}^{ab} - \frac{1}{2} \nabla_c \nabla^b \overline{\gamma}^{ca} - \frac{1}{2} \nabla_c \nabla^a \overline{\gamma}^{cb}. \tag{5.3.9}$$

The order of the derivatives in the final two terms may be exchanged using the Riemann tensor as follows:

$$\begin{aligned}
\nabla_c \nabla^b \overline{\gamma}^{ca} = g^{be} \delta^d{}_c \nabla_d \nabla_e \overline{\gamma}^{ca} &= g^{be} \delta^d{}_c (\nabla_e \nabla_d \overline{\gamma}^{ca} - R_{def}{}^c \overline{\gamma}^{fa} - R_{def}{}^a \overline{\gamma}^{cf}) \\
&= \nabla^b \nabla_c \overline{\gamma}^{ca} + R^b{}_f \overline{\gamma}^{fa} - R_c{}^b{}_f{}^a \gamma^{cf} \\
&= \nabla^b \nabla_c \overline{\gamma}^{ca} + \Lambda \overline{\gamma}^{ba} + R_c{}^{ba}{}_f \gamma^{cf}.
\end{aligned}$$

As $\boldsymbol{\gamma}$ obeys the de Donder condition, then the above simplifies to

$$\nabla_c \nabla^b \overline{\gamma}^{ca} = \Lambda \overline{\gamma}^{ba} + R_c{}^{ba}{}_f \overline{\gamma}^{cf}.$$

Substituting this result back into (5.3.9) and using the symmetry of $\boldsymbol{\gamma}$ yields

$$\begin{aligned}
\nabla_c \mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}}) &= \frac{1}{2} \Box \overline{\gamma}^{ab} - \Lambda \overline{\gamma}^{ab} - R_c{}^{ba}{}_f \overline{\gamma}^{cf} \\
&= \frac{1}{2} P^{abcd} \overline{\gamma}_{cd} - \Lambda \overline{\gamma}^{ab},
\end{aligned}$$

where in the final line we have used the definition of P from (4.2.9). ■

Now we establish an expression for the momentum, $\boldsymbol{\pi}$, of a de Donder perturbation.

Lemma 5.3.2 *If $\boldsymbol{\gamma} \in C^\infty(S_2^0(M;\mathbb{C}))$ is de Donder, then its associated momentum is given by*

$$\pi^{ab}(\boldsymbol{\gamma}) = n_c \mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}}). \tag{5.3.10}$$

Proof. Using the definition of $\boldsymbol{\pi}$ from (5.1.2) and the expression for $\boldsymbol{\Pi}$ from (4.2.27), we have

$$\pi^{ab} = \frac{1}{2}n_c\nabla^c\gamma^{ab} - \frac{1}{2}g^{ab}n_c\nabla^c\gamma + \frac{1}{2}g^{ab}n_c\nabla_d\gamma^{cd} + \frac{1}{4}n^b\nabla^a\gamma + \frac{1}{4}n^a\nabla^b\gamma - \frac{1}{2}n_c\nabla^a\gamma^{cb} - \frac{1}{2}n_c\nabla^b\gamma^{ca}. \tag{5.3.11}$$

As $\boldsymbol{\gamma}$ obeys the de Donder condition, meaning that $\nabla^a\gamma_{ab} = \frac{1}{2}\nabla_b\gamma$, the third term on the right-hand side can be re-expressed to give

$$\pi^{ab} = \frac{1}{2}n_c\nabla^c\gamma^{ab} - \frac{1}{4}g^{ab}n_c\nabla^c\gamma + \frac{1}{4}n^b\nabla^a\gamma + \frac{1}{4}n^a\nabla^b\gamma - \frac{1}{2}n_c\nabla^a\gamma^{cb} - \frac{1}{2}n_c\nabla^b\gamma^{ca}, \tag{5.3.12}$$

which is equal to $n_c\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}})$. ∎

These results will now be used to prove the result alluded to earlier. This result will be important when we consider the observables of the theory.

Theorem 5.3.3 *Given a $\boldsymbol{\gamma} \in \mathscr{S}(M;\mathbb{C})$ and an $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{C}))$ satisfying $\nabla \cdot \boldsymbol{f} = 0$, then*

$$\omega([\boldsymbol{E}\overline{\boldsymbol{f}}],[\boldsymbol{\gamma}]) = -\frac{1}{2}\int_M \gamma^{dD}_{ab} f^{ab} dvol_{\boldsymbol{g}}, \tag{5.3.13}$$

where $\boldsymbol{\gamma}^{dD}$ denotes a de Donder representative of $[\boldsymbol{\gamma}]$.

Proof. Given the above assumption regarding $\boldsymbol{f}$, then by Theorem 4.7.4, $\boldsymbol{E}\overline{\boldsymbol{f}}$ is a de Donder solution and the proof of that theorem also shows that $\boldsymbol{E}^{\pm}\overline{\boldsymbol{f}}$ will obey the de Donder condition. By selecting a de Donder representative $\boldsymbol{\gamma}^{dD}$ of $[\boldsymbol{\gamma}]$ we can use the result of Lemma 5.3.2 to expand out the left-hand side of (5.3.13) as

$$\omega([\boldsymbol{E}\overline{\boldsymbol{f}}],[\boldsymbol{\gamma}]) = \int_\Sigma \Big((\boldsymbol{E}\overline{\boldsymbol{f}})_{ab}\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}^{dD}}) - \gamma^{dD}_{ab}\mathscr{D}^{cab}(\boldsymbol{E}\boldsymbol{f}) \Big) n_c dvol_{\boldsymbol{h}}, \tag{5.3.14}$$

where we have used that $\overline{\boldsymbol{E}\overline{\boldsymbol{f}}} = \boldsymbol{E}\overline{\overline{\boldsymbol{f}}} = \boldsymbol{E}\boldsymbol{f}$ from Lemma 4.7.3.

Given that $\operatorname{supp}\boldsymbol{f}$ is compact, we may choose Cauchy surfaces Σ and Σ' such that $\Sigma \subset I^+(\Sigma')$ and $\operatorname{supp}\boldsymbol{f} \subset I^+(\Sigma') \cap I^-(\Sigma)$. The region bounded between these Cauchy surfaces will henceforth be denoted by V, and the future-pointing unit normals to these Cauchy surfaces will be denoted by $\boldsymbol{n}$ and $\boldsymbol{n}'$ respectively. To prove the result, we will apply Gauss' Theorem to the region V and the vector field

$$v^c = \gamma^{dD}_{ab}\mathscr{D}^{cab}(\boldsymbol{E}^+\boldsymbol{f}) - (\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab}\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}^{dD}}). \tag{5.3.15}$$

The divergence of this vector field is given by

$$\nabla_c v^c = (\nabla_c \gamma^{dD}_{ab})\mathscr{D}^{cab}(\boldsymbol{E}^+\boldsymbol{f}) + \gamma^{dD}_{ab}\nabla_c\mathscr{D}^{cab}(\boldsymbol{E}^+\boldsymbol{f}) - (\nabla_c(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab})\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}^{dD}}) - (\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab}\nabla_c\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}^{dD}}).$$

As both $\boldsymbol{E}^+\boldsymbol{f}$ and $\boldsymbol{\gamma}^{dD}$ obey the de Donder condition, then we may apply Lemma 5.3.1 and obtain

$$\nabla_c v^c = (\nabla_c \gamma^{dD}_{ab})\mathscr{D}^{cab}(\boldsymbol{E}^+\boldsymbol{f}) + \frac{1}{2}\gamma^{dD}_{ab}P^{abcd}(\boldsymbol{E}^+\boldsymbol{f})_{cd} - \Lambda\gamma^{dD}_{ab}(\boldsymbol{E}^+\boldsymbol{f})^{ab} - (\nabla_c(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab})\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}^{dD}}) - \frac{1}{2}(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab}P^{abcd}\overline{\gamma^{dD}}_{cd} + \Lambda(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab}\overline{\gamma^{dD}}^{ab}.$$

The third and sixth terms on the right-hand side cancel each other. Now, as $\boldsymbol{\gamma}^{dD}$ is a linearized gravity solution, then $P^{abcd}\overline{\gamma^{dD}}_{cd} = 0$, and we know that $P^{abcd}(\boldsymbol{E}^+\boldsymbol{f})_{cd} = f^{ab}$, so therefore

$$\nabla_c v^c = (\nabla_c \gamma^{dD}_{ab})\mathscr{D}^{cab}(\boldsymbol{E}^+\boldsymbol{f}) + \frac{1}{2}\gamma^{dD}_{ab}f^{ab} - (\nabla_c(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab})\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}^{dD}}). \tag{5.3.16}$$

We now examine the first and third terms on the right-hand side. The first term is simply

$$(\nabla_c \gamma^{dD}_{ab})\mathscr{D}^{cab}(\boldsymbol{E}^+\boldsymbol{f}) = \frac{1}{2}\nabla_c\gamma^{dD}_{ab}\nabla^c(\boldsymbol{E}^+\boldsymbol{f})^{ab} - \frac{1}{2}\nabla_c\gamma^{dD}_{ab}\nabla^b(\boldsymbol{E}^+\boldsymbol{f})^{ca} - \frac{1}{2}\nabla_c\gamma^{dD}_{ab}\nabla^a(\boldsymbol{E}^+\boldsymbol{f})^{cb},$$

and the third term is

$$(\nabla_c(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab})\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}^{dD}}) = \frac{1}{2}\nabla_c(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab}\nabla^c\overline{\gamma_{dD}}^{ab} - \frac{1}{2}\nabla_c(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab}\nabla^b\overline{\gamma_{dD}}^{ca} - \frac{1}{2}\nabla_c(\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab}\nabla^a\overline{\gamma_{dD}}^{cb}. \tag{5.3.17}$$

The terms from the right-hand side of (5.3.16) can be re-expressed using the following two lemmas.

Lemma 5.3.4 *Given* $\boldsymbol{u}, \boldsymbol{v} \in C^\infty(S^0_2(M;\mathbb{C}))$ *then*

$$\nabla_c u_{ab}\nabla^c\overline{v}^{ab} = \nabla_c\overline{u}_{ab}\nabla^c v^{ab}.$$

Proof. Using the definition of trace-reversal, we have

$$\nabla_c u_{ab} \nabla^c \overline{v}^{ab} = \nabla_c u_{ab} \nabla^c v^{ab} - \frac{1}{2} \nabla_c u_{ab} g^{ab} \nabla^c v = \nabla_c u_{ab} \nabla^c v^{ab} - \frac{1}{2} \nabla_c u \nabla^c v$$
$$= \nabla_c u_{ab} \nabla^c v^{ab} - \frac{1}{2} g_{ab} \nabla_c u \nabla^c v^{ab} = \nabla_c \overline{u}_{ab} \nabla^c v^{ab}.$$

■

Lemma 5.3.5 *Given $\boldsymbol{u}, \boldsymbol{v} \in C^\infty(S^0_2(M;\mathbb{C}))$, both obeying the de Donder condition, then*

$$\nabla_c u_{ab} \nabla^a \overline{v}^{cb} = \nabla_c \overline{u}_{ab} \nabla^a v^{cb}. \tag{5.3.18}$$

Proof. Expanding out the left-hand side of (5.3.18) using the definition of trace-reversal and then the de Donder condition yields

$$\nabla_c u_{ab} \nabla^a \overline{v}^{cb} = \nabla_c u_{ab} \nabla^a v^{cb} - \frac{1}{2} \nabla_c u_{ab} g^{cb} \nabla^c v = \nabla_c u_{ab} \nabla^a v^{cb} - \frac{1}{4} \nabla_c u \nabla^c v.$$

Again, using first the de Donder condition and then the definition of trace-reversal gives

$$\nabla_c u_{ab} \nabla^a \overline{v}^{cb} = \nabla_c u_{ab} \nabla^a v^{cb} - \frac{1}{2} \nabla_c u \nabla_b v^{cb} = \nabla_c u_{ab} \nabla^a v^{cb} - \frac{1}{2} g_{ab} \nabla_c u \nabla^a v^{cb} = \nabla_c \overline{u}_{ab} \nabla^a v^{cb}.$$

■

Applying Lemmas 5.3.4 and 5.3.5 to (5.3.17) and using $\overline{\boldsymbol{E}\overline{\boldsymbol{f}}} = \boldsymbol{E}\overline{\overline{\boldsymbol{f}}} = \boldsymbol{E}\boldsymbol{f}$ gives

$$(\nabla_c (\boldsymbol{E}^+ \overline{\boldsymbol{f}})_{ab}) \mathscr{D}^{cab} (\overline{\boldsymbol{\gamma}^{dD}}) = \frac{1}{2} \nabla_c (\boldsymbol{E}^+ \boldsymbol{f})_{ab} \nabla^c \gamma^{ab}_{dD} - \frac{1}{2} \nabla_c (\boldsymbol{E}^+ \boldsymbol{f})_{ab} \nabla^b \gamma^{ca}_{dD} - \frac{1}{2} \nabla_c (\boldsymbol{E}^+ \boldsymbol{f})_{ab} \nabla^a \gamma^{cb}_{dD}$$
$$= (\nabla_c \gamma^{dD}_{ab}) \mathscr{D}^{cab} (\boldsymbol{E}^+ \boldsymbol{f}).$$

Hence, (5.3.16) reduces to

$$\nabla_a v^a = \frac{1}{2} \gamma^{dD}_{ab} f^{ab}.$$

Applying Gauss' Theorem to the region V and the vector field $\boldsymbol{v}$ gives

$$\frac{1}{2} \int_V \gamma^{dD}_{ab} f^{ab} dvol_{\boldsymbol{g}} = \int_V \nabla_a v^a dvol_{\boldsymbol{g}} = - \int_\Sigma n_a v^a dvol_{\boldsymbol{h}} + \int_{\Sigma'} n'_a v^a dvol_{\boldsymbol{h}}. \tag{5.3.19}$$

By the support properties of $\boldsymbol{E}^+ \boldsymbol{f}$ and the way that the Cauchy surfaces were chosen, both $\boldsymbol{E}^+ \boldsymbol{f}$ and its derivative vanish on Σ' and so $\boldsymbol{v}$ vanishes on Σ'. Therefore, upon substituting the expression for $\boldsymbol{v}$, we are left with

$$\frac{1}{2}\int_V \gamma_{ab}^{dD} f^{ab} dvol_{\boldsymbol{g}} = -\int_{\Sigma} n_c \left(\gamma_{ab}^{dD}\mathscr{D}^{cab}(\boldsymbol{E}^+\boldsymbol{f}) - (\boldsymbol{E}^+\overline{\boldsymbol{f}})_{ab}\mathscr{D}^{cab}(\overline{\boldsymbol{\gamma}^{dD}})\right) dvol_{\boldsymbol{h}}. \qquad (5.3.20)$$

Now, by the support properties of $\boldsymbol{E}^-\boldsymbol{f}$, both it and its derivative vanish at Σ and so we may replace $\boldsymbol{E}^+$ by $-\boldsymbol{E}$. As $\frac{1}{2}\int_M \gamma_{ab}^{dD} f^{ab} dvol_{\boldsymbol{g}} = \frac{1}{2}\int_V \gamma_{ab}^{dD} f^{ab} dvol_{\boldsymbol{g}}$, then upon using the expression (5.3.14) the result is obtained. ■

5.4 Observables

As discussed in section 3.2.1, the basic observables of a linear quantum field theory are the smeared quantum fields. Classically, we are led to consider scalar-valued functions on the phase space. Initially, as our prototype observable, we consider integrals of the form $\int_M \gamma_{ab} f^{ab} dvol_{\boldsymbol{g}}$, where $\boldsymbol{\gamma} \in \mathscr{S}(M;\mathbb{C})$ and $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$. One immediately notices a problem with this. If to each equivalence class $[\boldsymbol{\gamma}] \in \mathscr{P}_{\mathbb{C}}(M)$, we were to assign the quantity $\int_M \gamma_{ab} f^{ab} dvol_{\boldsymbol{g}}$ as the value of a function on the phase space, then that function would not be well-defined. Specifically, just making a gauge transformation to move around within the equivalence class would generally change the value of the integral. Therefore to ensure that such objects are well-defined and to ensure that they are physical, in the sense of being gauge invariant, we will have to place a restriction upon the choice of smearing tensors used. The next lemma shows what this restriction is.

Lemma 5.4.1 *For $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$, we have $\int_M \gamma_{ab} f^{ab} dvol_{\boldsymbol{g}} = 0$ for all $\boldsymbol{\gamma} \in \hat{\mathscr{G}}(M;\mathbb{C})$ if and only if $\nabla^a f_{(ab)} = 0$.*

Proof. If $\boldsymbol{\gamma} = \pounds_{\boldsymbol{w}}\boldsymbol{g} \in \hat{\mathscr{G}}(M;\mathbb{C})$, then $\int_M (\pounds_{\boldsymbol{w}}\boldsymbol{g})_{ab} f^{ab} dvol_{\boldsymbol{g}} = 2\int_M (\nabla_{(a}w_{b)}) f^{ab} dvol_{\boldsymbol{g}}$, whereupon if one moves the symmetrization onto $\boldsymbol{f}$ and uses the Leibniz rule, then

$$2\int_M (\nabla_{(a}w_{b)}) f^{ab} dvol_{\boldsymbol{g}} = 2\int_M \nabla_a(w_b f^{(ab)}) dvol_{\boldsymbol{g}} - 2\int_M w_b(\nabla_a f^{(ab)}) dvol_{\boldsymbol{g}}.$$

As $\boldsymbol{f}$ has compact support, the first integral on the right-hand side vanishes by Gauss' Theorem. Hence, we are left with

$$\int_M (\pounds_{\boldsymbol{w}}\boldsymbol{g})_{ab} f^{ab} dvol_{\boldsymbol{g}} = -2\int_M w_b(\nabla_a f^{(ab)}) dvol_{\boldsymbol{g}}. \qquad (5.4.1)$$

For (5.4.1) to vanish it is clearly sufficient that $\nabla^a f_{(ab)} = 0$; as $\boldsymbol{w}$ may, in particular, be any element of $C_0^\infty(T_0^1(M;\mathbb{C}))$, necessity holds as well. This last statement follows from a similar argument to that used at the end of the proof of Theorem 4.5.4. Select an arbitrary point p and a local coordinate neighbourhood of this point, so one has a bump function χ at p. Now choose an orthonormal frame (with associated dual frame) at p. Construct four covector fields, one in each of the frame directions, which at p are equal to the component of $\nabla^a f_{(ab)}$ in the dual direction, multiplied by χ. Now follow the same argument described at the end of Theorem 4.5.4 for each of the four directions. The result is that at p the components of $\nabla^a f_{(ab)}$ in the frame vanish and so $\nabla^a f_{(ab)}|_p = 0$. Since p was arbitrary, we conclude that $\nabla^a f_{(ab)} = 0$ globally. ■

Definition 5.4.2 *For each $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$ satisfying $\nabla^a f_{(ab)} = 0$ there is an associated gauge invariant observable $F_{\boldsymbol{f}} : \mathscr{P}_{\mathbb{C}}(M) \to \mathbb{C}$ whose action is given by*

$$F_{\boldsymbol{f}}([\boldsymbol{\gamma}]) = \int \gamma_{ab} f^{ab} dvol_{\boldsymbol{g}}. \tag{5.4.2}$$

Having defined our observables, we now consider a generalization of Theorem 5.3.3 that displays a link between the observables and the sympletic product.

Theorem 5.4.3 *Given a $[\boldsymbol{\gamma}] \in \mathscr{P}_{\mathbb{C}}(M)$ and an $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{C}))$ satisfying $\nabla^a f_{ab} = 0$, then*

$$F_{\boldsymbol{f}}([\boldsymbol{\gamma}]) = \int_M \gamma_{ab} f^{ab} dvol_{\boldsymbol{g}} = -2\omega([\boldsymbol{E}\overline{\boldsymbol{f}}], [\boldsymbol{\gamma}]). \tag{5.4.3}$$

Proof. As $\nabla^a f_{ab} = 0$ we can use Theorem 5.3.3 to give

$$\omega([\boldsymbol{E}\overline{\boldsymbol{f}}], [\boldsymbol{\gamma}]) = -\frac{1}{2}\int_M \gamma_{ab}^{dD} f^{ab} dvol_{\boldsymbol{g}}. \tag{5.4.4}$$

As $\boldsymbol{f}$ satisfies the requirements of Lemma 5.4.1 we may replace $\boldsymbol{\gamma}^{dD}$ in the integral by any representative of $[\boldsymbol{\gamma}]$, in particular, $\boldsymbol{\gamma}$. ■

We now move on to consider the various relations satisfied by the observables, beginning with the simplest ones first.

Theorem 5.4.4 *Given any $[\boldsymbol{\gamma}] \in \mathscr{P}_{\mathbb{C}}(M)$, the $F_{\boldsymbol{f}}$'s satisfy:*

(i) Complex linearity: $F_{\alpha\boldsymbol{f}+\beta\boldsymbol{f}'}([\boldsymbol{\gamma}]) = \alpha F_{\boldsymbol{f}}([\boldsymbol{\gamma}]) + \beta F_{\boldsymbol{f}'}([\boldsymbol{\gamma}])$ *for all $\alpha, \beta \in \mathbb{C}$ and all $\boldsymbol{f}, \boldsymbol{f}' \in C_0^\infty(T_2^0(M;\mathbb{C}))$ satisfying $\nabla^a f_{(ab)} = 0 = \nabla^a f'_{(ab)}$;*

(ii) Hermiticity: $F_{\boldsymbol{f}}([\boldsymbol{\gamma}])^* = F_{\boldsymbol{f}^*}([\boldsymbol{\gamma}^*])$ *for all $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$ satisfying $\nabla^a f_{(ab)} = 0$;*

(iii) Symmetry: $F_{\boldsymbol{f}}([\boldsymbol{\gamma}]) = 0$ *for all antisymmetric* $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$.

Proof. (i) holds because the integrand is linear in $\boldsymbol{f}$ and because integration is a linear operation, (ii) holds by the properties of complex-conjugation, and (iii) is true because all elements of $[\boldsymbol{\gamma}]$ are symmetric. ■

Another important relation obeyed by the observables will now be discussed. In the simplest possible terms it shows that the equation of motion holds, which in our case is the linearized Einstein equation. However, the result is subtly different from what one might have expected, which, as we will establish, is a direct consequence of having to consider the expanded pure gauge space $\hat{\mathscr{G}}(M;\mathbb{C})$.

One would normally expect, see section 3.2.1 for the scalar field case, that as L is formally self-adjoint, then we should have $F_{L(\boldsymbol{f})} \equiv 0$ for all $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$; recall that L is defined for non-symmetric $\boldsymbol{f}$ by (4.2.8). We will prove that the more general relation $F_{L(\boldsymbol{f})} \equiv 0$ holds for all $\boldsymbol{f} \in C_{TC}^\infty(T_2^0(M;\mathbb{C}))$ satisfying $L(\boldsymbol{f}) \in C_0^\infty(S_2^0(M;\mathbb{C}))$. To establish this, the following sets will be required. Let

$$\begin{aligned}
\mathscr{L}(M;\mathbb{C}) &:= \{L(\boldsymbol{k}) : \boldsymbol{k} \in C_0^\infty(S_2^0(M;\mathbb{C}))\};\\
\hat{\mathscr{L}}(M;\mathbb{C}) &:= \{L(\boldsymbol{k}) : \boldsymbol{k} \in C_{TC}^\infty(S_2^0(M;\mathbb{C}))\} \cap C_0^\infty(S_2^0(M;\mathbb{C}));\\
\mathscr{F}(M;\mathbb{C}) &:= \{\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{C})) \mid \nabla\cdot\boldsymbol{f} = 0\}.
\end{aligned}$$

A consequence of the upcoming result is that $\hat{\mathscr{G}}(M;\mathbb{C}) = \mathscr{G}(M;\mathbb{C})$ if and only if $\hat{\mathscr{L}}(M;\mathbb{C}) = \mathscr{L}(M;\mathbb{C})$.

We consider the kernel of the linear map F, which assigns an observable $F_{\boldsymbol{f}}$ to each $\boldsymbol{f} \in \mathscr{F}(M;\mathbb{C})$. Without loss of generality, we can restrict attention to the space $\mathscr{F}(M;\mathbb{C})$ due to relation (iii) of Theorem 5.4.4. Given any such $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{C}))$ satisfying $\nabla\cdot\boldsymbol{f} = 0$ then, by Theorem 5.4.3, $F_{\boldsymbol{f}}([\boldsymbol{\gamma}]) = 0$ for all $[\boldsymbol{\gamma}] \in \mathscr{P}_{\mathbb{C}}(M)$ if and only if $[\boldsymbol{E}\overline{\boldsymbol{f}}] = [0]$, which means that $\boldsymbol{E}\overline{\boldsymbol{f}}$ is pure gauge. We will now find what form $\boldsymbol{f}$ must take for this to hold.

Lemma 5.4.5 *Suppose* $\boldsymbol{f} \in \mathscr{F}(M;\mathbb{C})$. *Then* $\boldsymbol{E}\overline{\boldsymbol{f}} \in \hat{\mathscr{G}}(M;\mathbb{C})$ *if and only if* $\boldsymbol{f} \in \hat{\mathscr{L}}(M;\mathbb{C})$; *similarly,* $\boldsymbol{E}\overline{\boldsymbol{f}} \in \mathscr{G}(M;\mathbb{C})$ *if and only if* $\boldsymbol{f} \in \mathscr{L}(M;\mathbb{C})$.

Proof. ($\Rightarrow$) Suppose $\boldsymbol{E}\overline{\boldsymbol{f}} = \pounds_{\boldsymbol{w}}\boldsymbol{g} \in \hat{\mathscr{G}}(M;\mathbb{C})$. As $\nabla\cdot\boldsymbol{f} = 0$, $\pounds_{\boldsymbol{w}}\boldsymbol{g}$ is a de Donder solution. In consequence, $\boldsymbol{w} \in C^\infty(T_0^1(M;\mathbb{C}))$ satisfies $(\Box + \Lambda)\boldsymbol{w} = 0$, and so, by the exact sequence of Theorem 2.4.6, it may be written as $\boldsymbol{w} = \tilde{\boldsymbol{E}}\boldsymbol{v}$ for some $\boldsymbol{v} \in C_{TC}^\infty(T_0^1(M;\mathbb{C}))$. Therefore we have $\pounds_{\boldsymbol{w}}\boldsymbol{g} = \pounds_{\tilde{\boldsymbol{E}}\boldsymbol{v}}\boldsymbol{g} = \boldsymbol{E}\pounds_{\boldsymbol{v}}\boldsymbol{g}$, where we have used the analogue of Lemma 4.7.5 for the extended Green's operators. Thus $\boldsymbol{E}\overline{\boldsymbol{f}} = \boldsymbol{E}\pounds_{\boldsymbol{v}}\boldsymbol{g}$ and so, again using Theorem 2.4.6, we have

$$\pounds_{\boldsymbol{v}}\boldsymbol{g} = \overline{\boldsymbol{f}} + P(\boldsymbol{k}) \tag{5.4.5}$$

for some $\boldsymbol{k} \in C^\infty_{TC}(S^0_2(M;\mathbb{C}))$. Taking the trace-reversal and then the divergence of this equation gives

$$(\Box + \Lambda)\boldsymbol{v}^\flat = (\Box + \Lambda)\nabla \cdot \overline{\boldsymbol{k}},$$

where we have used the properties of $\boldsymbol{f}$ and Lemmas 4.2.1, 4.2.3 & 4.5.1. As both $\boldsymbol{v}^\flat$ and $\nabla \cdot \overline{\boldsymbol{k}}$ are time-compact, and their difference solves the homogeneous equation $(\Box + \Lambda)(\boldsymbol{v}^\flat - \nabla \cdot \overline{\boldsymbol{k}}) = 0$, then by [11, Thm 3.3.1] their difference will vanish, meaning $\boldsymbol{v}^\flat = \nabla \cdot \overline{\boldsymbol{k}}$. Substituting this result back into (5.4.5) and using Theorem 4.2.2 gives $\overline{\boldsymbol{f}} = \pounds_{(\nabla\cdot\overline{\boldsymbol{k}})^\sharp}\boldsymbol{g} - P(\boldsymbol{k}) = 2\overline{L(\boldsymbol{k})}$. Therefore $\boldsymbol{f} = L(2\boldsymbol{k}) \in \hat{\mathscr{L}}(M;\mathbb{C})$ as required.

($\Leftarrow$) Conversely, given an $\boldsymbol{f} \in \hat{\mathscr{L}}(M;\mathbb{C})$, then $\boldsymbol{f} = L(2\boldsymbol{k})$ for some $\boldsymbol{k} \in C^\infty_{TC}(S^0_2(M;\mathbb{C}))$. Now, let $\boldsymbol{w} = \tilde{\boldsymbol{E}}(\nabla \cdot \overline{\boldsymbol{k}})^\sharp \in C^\infty(T^1_0(M;\mathbb{C}))$, which gives $\pounds_{\boldsymbol{w}}\boldsymbol{g} = \pounds_{\tilde{\boldsymbol{E}}(\nabla\cdot\overline{\boldsymbol{k}})^\sharp}\boldsymbol{g} = \boldsymbol{E}\pounds_{(\nabla\cdot\overline{\boldsymbol{k}})^\sharp}\boldsymbol{g}$, where we have used the analogue of Lemma 4.7.5 for the extended Green's operators in the final equality. By Theorem 4.2.2 this becomes $\pounds_{\boldsymbol{w}}\boldsymbol{g} = \boldsymbol{E}(2\overline{L(\boldsymbol{k})} + P(\boldsymbol{k})) = \boldsymbol{E}\overline{\boldsymbol{f}}$, because $\boldsymbol{E}P(\boldsymbol{k}) = 0$. As $\boldsymbol{f}$ is compactly supported by assumption, we deduce, from the exact sequence of Theorem 2.4.3, that $\pounds_{\boldsymbol{w}}\boldsymbol{g} \in C^\infty_{SC}(S^0_2(M;\mathbb{C}))$ and hence $\boldsymbol{E}\overline{\boldsymbol{f}} \in \hat{\mathscr{G}}(M;\mathbb{C})$.

The second statement has an exactly analogous proof, replacing C^∞_{TC} by C^∞_0 and C^∞ by C^∞_{SC}, and hatted spaces by their unhatted counterparts throughout. ■

With this result in mind, we have established that the following fourth relation holds.

Theorem 5.4.6 *Given any $[\boldsymbol{\gamma}] \in \mathscr{P}_{\mathbb{C}}(M)$, then we have*

$$F_{L(\boldsymbol{f})}([\boldsymbol{\gamma}]) = 0 \tag{5.4.6}$$

for all $\boldsymbol{f} \in C^\infty_{TC}(T^0_2(M;\mathbb{C}))$ such that $L(\boldsymbol{f}) \in C^\infty_0(T^0_2(M;\mathbb{C}))$.

The fifth and final relation is the Poisson bracket of two of these observables. As was discussed at length for the scalar field case in section 3.2.1, in order to be able to define the Poisson bracket we will need to ensure the presence of a smooth structure on our infinite-dimensional symplectic manifold $\mathscr{P}_{\mathbb{C}}(M)$. Just as for the scalar field case, one can make use of the concept of a Frölicher space [66, Ch. 23]. Following the methodology we utilised in section 3.2.1, we take a curve $c : \mathbb{R} \to \mathscr{P}_{\mathbb{C}}(M)$ to be smooth if the map $t \mapsto \omega(v, c(t))$ is smooth for all $v \in \mathscr{P}_{\mathbb{C}}(M)$. A function $F : \mathscr{P}_{\mathbb{C}}(M) \to \mathbb{C}$ is then deemed to be smooth if $F \circ c : \mathbb{R} \to \mathbb{C}$ is a smooth function for any smooth curve. The generating set of functions, that is, the functions $\omega(v,\cdot)$ (or equally $\omega(\cdot,v)$) for some $v \in \mathscr{P}_{\mathbb{C}}(M)$ are contained within the set of smooth functions by definition and hence the symplectic product is smooth in both of its arguments. We know from Theorem 5.4.3 that the functions $F_{\boldsymbol{f}}$ can be expressed in terms of the symplectic product and so all of the $F_{\boldsymbol{f}}$'s that we consider here are smooth functions. One can define the differential of a smooth function using the standard methods from finite dimensions, see [84, eq. (1.3.1)] for the finite-dimensional case and see section 3.2.1 of this thesis for the Frölicher space definition.

To define the Poisson bracket, we follow the methodology of [1, p. 568], but keep new concepts and notation to a minimum. As such, we take the Poisson bracket of two smooth functions $F, G \in C^{\infty}(\mathscr{P}_{\mathbb{C}}(M))$ to be given in terms of their exterior derivatives by

$$\{F, G\}([\boldsymbol{\gamma}]) = dF(dG^{\sharp_\omega})|_{[\boldsymbol{\gamma}]}, \tag{5.4.7}$$

where the Hamiltionian vector field $dG^{\sharp_\omega}$ induced by G satisfies

$$\omega_{[\boldsymbol{\gamma}]}(dG^{\sharp_\omega}|_{[\boldsymbol{\gamma}]}, v) = dG|_{[\boldsymbol{\gamma}]}(v) \tag{5.4.8}$$

for all $v \in T_{[\boldsymbol{\gamma}]}\mathscr{P}_{\mathbb{C}}(M)$. We will show in the proof of Theorem 5.4.7 that, for our case, $dG^{\sharp_\omega}|_{[\boldsymbol{\gamma}]}$ is uniquely defined by the condition (5.4.8). Here $\omega_{[\boldsymbol{\gamma}]} : T_{[\boldsymbol{\gamma}]}\mathscr{P}_{\mathbb{C}}(M) \times T_{[\boldsymbol{\gamma}]}\mathscr{P}_{\mathbb{C}}(M) \to \mathbb{C}$ is the symplectic form at $[\boldsymbol{\gamma}] \in \mathscr{P}_{\mathbb{C}}(M)$. Under the identification $T_{[\boldsymbol{\gamma}]}\mathscr{P}_{\mathbb{C}}(M) \cong \mathscr{P}_{\mathbb{C}}(M)$, we can replace $\omega_{[\boldsymbol{\gamma}]}$ by $\omega : \mathscr{P}_{\mathbb{C}}(M) \times \mathscr{P}_{\mathbb{C}}(M) \to \mathbb{C}$.

Theorem 5.4.7 *Assuming weak non-degeneracy holds, in particular, if $(M, \boldsymbol{g})$ has compact Cauchy surfaces, then the Poisson bracket of two observables satisfying Definition 5.4.2 is given by*

$$\{F_{\boldsymbol{f}}, F_{\boldsymbol{f}'}\} = -2\boldsymbol{E}(\boldsymbol{f}^s, \overline{\boldsymbol{f}'^s}) = 4\omega([\boldsymbol{E}\overline{\boldsymbol{f}^s}], [\boldsymbol{E}\overline{\boldsymbol{f}'^s}]), \tag{5.4.9}$$

where $\boldsymbol{f}^s$ denotes the symmetric part of $\boldsymbol{f}$, that is, $f^s_{ab} = f_{(ab)}$, and the bi-distribution $\boldsymbol{E}$ is defined by

$$\boldsymbol{E}(\boldsymbol{f}^s, \overline{\boldsymbol{f}'^s}) := \int_M f^{(ab)}(E_{ab}{}^{cd}\overline{f}'_{(cd)})dvol_{\boldsymbol{g}}. \tag{5.4.10}$$

Proof. We note that $dF_{\boldsymbol{f}}|_{[\boldsymbol{\gamma}]}([\boldsymbol{\gamma}']) = F_{\boldsymbol{f}}([\boldsymbol{\gamma}'])$ by linearity of $F_{\boldsymbol{f}}$. Thus, upon using (5.4.8), then relation (iii) of Theorem 5.4.4, and finally Theorem 5.4.3, we have

$$\omega((dF_{\boldsymbol{f}})^{\sharp_\omega}|_{[\boldsymbol{\gamma}]}, [\boldsymbol{\gamma}']) = dF_{\boldsymbol{f}}|_{[\boldsymbol{\gamma}]}([\boldsymbol{\gamma}']) = F_{\boldsymbol{f}}([\boldsymbol{\gamma}']) = F_{\boldsymbol{f}^s}([\boldsymbol{\gamma}']) = -2\omega([\boldsymbol{E}\overline{\boldsymbol{f}^s}], [\boldsymbol{\gamma}']), \tag{5.4.11}$$

for all $[\boldsymbol{\gamma}'] \in \mathscr{P}_{\mathbb{C}}(M)$. By the weak non-degeneracy of ω, this gives $(dF_{\boldsymbol{f}})^{\sharp_\omega}|_{[\gamma]} = -2[\boldsymbol{E}\overline{\boldsymbol{f}^s}]$. Inserting this into the definition of the Poisson bracket (5.4.7), results in

$$\{F_{\boldsymbol{f}}, F_{\boldsymbol{f}'}\}([\gamma]) = -dF_{\boldsymbol{f}}|_{[\gamma]}(2[\boldsymbol{E}\overline{\boldsymbol{f}'^s}]) = -2F_{\boldsymbol{f}}([\boldsymbol{E}\overline{\boldsymbol{f}'^s}]) = -2F_{\boldsymbol{f}^s}([\boldsymbol{E}\overline{\boldsymbol{f}'^s}]), \tag{5.4.12}$$

where we have used that $dF_{\boldsymbol{f}}|_{[\gamma]}([\gamma']) = F_{\boldsymbol{f}}([\gamma'])$ and relation (iii) of Theorem 5.4.4. By using the definition of $F_{\boldsymbol{f}}$, we have

$$\{F_{\boldsymbol{f}}, F_{\boldsymbol{f}'}\}([\gamma]) = -2\boldsymbol{E}(\boldsymbol{f}^s, \overline{\boldsymbol{f}'^s}), \tag{5.4.13}$$

and using Theorem 5.4.3 gives the final equality of (5.4.9). ■

The bi-distribution $\boldsymbol{E}$ will be referred to as the propagator (in an integral representation, its integral kernel would be known as the Pauli-Jordan or Schwinger function [17, p. 20]). This bi-distribution appears in the Poisson bracket, however, one of its arguments contains a trace-reversal. By expanding this out explicitly, we can discover what the Poisson bracket is equal to. We have

$$\begin{aligned}
\boldsymbol{E}(\boldsymbol{f}^s, \overline{\boldsymbol{f}'^s}) &= \int_M f^{(ab)} E_{ab}{}^{c'd'} \overline{f'}_{(c'd')} dvol_{\boldsymbol{g}} = \int_M f^{(ab)} \overline{E_{ab}{}^{c'd'} f'_{(c'd')}} dvol_{\boldsymbol{g}} \\
&= \int_M f^{(ab)} E_{ab}{}^{c'd'} f'_{(c'd')} dvol_{\boldsymbol{g}} - \frac{1}{2}\int_M f^{(ab)} g_{ab} E f' dvol_{\boldsymbol{g}} \\
&= \int_M f^{(ab)} E_{ab}{}^{c'd'} f'_{(c'd')} dvol_{\boldsymbol{g}} - \frac{1}{2}\int_M f E f' dvol_{\boldsymbol{g}},
\end{aligned}$$

where in the second equality we have used Lemma 4.7.3 and the third equality uses the definition of trace-reversal and that the trace commutes with the Green's operators, see Lemma 4.7.2. Note that E is the scalar propagator for $(\Box + 2\Lambda)$. Therefore the Poisson bracket is

$$\{F_{\boldsymbol{f}}, F_{\boldsymbol{f}'}\} = -\left(2\int_M f^{(ab)} E_{ab}{}^{c'd'} f'_{(c'd')} dvol_{\boldsymbol{g}} - \int_M f E f' dvol_{\boldsymbol{g}}\right). \tag{5.4.14}$$

An expression for the commutator of two free graviton fields was previously conjectured by Lichnerowicz [70]. To motivate his definition he used what he had earlier found from treating the case of Minkowski spacetime, analogy with electromagnetism and the results of Fierz and Pauli [38].

To obtain our commutator we will simply apply Dirac quantization in the next chapter, which essentially just means the commutator is i times the Poisson bracket for suitable observables. Comparing what we would get, i times the right-hand side of (5.4.14), with Lichnerowicz's formula [70, eq. (21.3)], it is clear that they are identical. Note that his definition of the symmetrised propagator [70, eqs (13.2) & (13.3)] explains why we have a factor of two and he does not. Therefore his conjectured commutator has actually be found to be what one obtains by applying Dirac's quantization prescription to the classical observables $F_{\boldsymbol{f}}$.

One final aspect to consider is whether the collection of observables that we have defined is in fact 'large enough'. What we mean by this is whether they are sufficient to the task of distinguishing points of the phase space $\mathscr{P}_{\mathbb{C}}(M)$. The following theorem shows that they are.

Theorem 5.4.8 *Assuming that ω is weakly non-degenerate, in particular, for any $(M, \boldsymbol{g})$ with compact Cauchy surfaces, then, for any distinct $[\boldsymbol{\gamma}_1], [\boldsymbol{\gamma}_2] \in \mathscr{P}_{\mathbb{C}}(M)$, there exists an $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{C}))$ with $\nabla^a f_{ab} = 0$ such that $F_{\boldsymbol{f}}([\boldsymbol{\gamma}_1]) \neq F_{\boldsymbol{f}}([\boldsymbol{\gamma}_2])$.*

Proof. As $[\boldsymbol{\gamma}_1] \neq [\boldsymbol{\gamma}_2]$, then, by the weak non-degeneracy of ω, there exists a $[\boldsymbol{\gamma}] \in \mathscr{P}_{\mathbb{C}}(M)$ such that

$$\omega([\boldsymbol{\gamma}], [\boldsymbol{\gamma}_1]) \neq \omega([\boldsymbol{\gamma}], [\boldsymbol{\gamma}_2]). \tag{5.4.15}$$

By Theorem 4.7.6, $[\boldsymbol{\gamma}] = [\boldsymbol{E}\overline{\boldsymbol{f}}]$ for some $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{C}))$ satisfying $\nabla \cdot \boldsymbol{f} = 0$. Using Theorem 5.4.3 together with (5.4.15) gives

$$F_{\boldsymbol{f}}([\boldsymbol{\gamma}_1]) = -2\omega([\boldsymbol{E}\overline{\boldsymbol{f}}], [\boldsymbol{\gamma}_1]) \neq -2\omega([\boldsymbol{E}\overline{\boldsymbol{f}}], [\boldsymbol{\gamma}_2]) = F_{\boldsymbol{f}}([\boldsymbol{\gamma}_2]). \tag{5.4.16}$$

■

Chapter 6

Quantization

The first part of this chapter, which considers the algebra of observables, is based upon section 4.4 of the paper [34] cowritten by the author with Dr C. J. Fewster. The remainder of the chapter is presently unpublished work.

6.1 Algebra of observables

As was shown in section 3.2.2 for the case of the scalar field, a quantum theory can be obtained in the algebraic framework by applying Dirac's prescription for quantization to the minimal collection of classical observables. However, before the algebra is actually constructed, the various relations between the quantum observables need to be addressed. We will let $[\boldsymbol{\gamma}](\boldsymbol{f})$ denote the smeared quantum field describing the graviton, where $[\boldsymbol{\gamma}]$ is not an equivalence class of classical solutions.

Applying Dirac's prescription [29] to the classical observables $F_{\boldsymbol{f}}$, the quantum observables $[\boldsymbol{\gamma}](\boldsymbol{f})$, labelled by test tensors $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$ with divergence-free symmetric part, will have a commutator given by

$$[[\boldsymbol{\gamma}](\boldsymbol{f}_1), [\boldsymbol{\gamma}](\boldsymbol{f}_2)] = i\{F_{\boldsymbol{f}_1}, F_{\boldsymbol{f}_2}\} = -2i\boldsymbol{E}(\boldsymbol{f}_1^s, \overline{\boldsymbol{f}_2^s}).$$

Recall that $\boldsymbol{f}^s$ denotes the symmetric part of the tensor field $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$. The classical relations from Theorems 5.4.4 & 5.4.6 carry straight over to their quantum analogues and as such, the following relations should hold:

(i) *Complex-linearity*: $[\boldsymbol{\gamma}](\alpha\boldsymbol{f}_1 + \beta\boldsymbol{f}_2) = \alpha[\boldsymbol{\gamma}](\boldsymbol{f}_1) + \beta[\boldsymbol{\gamma}](\boldsymbol{f}_2)$ for all $\alpha, \beta \in \mathbb{C}$ and all $\boldsymbol{f}_i \in C_0^\infty(T_2^0(M;\mathbb{C}))$ such that $\nabla^a(f_i)_{(ab)} = 0$;

(ii) *Hermiticity*: $[\boldsymbol{\gamma}](\boldsymbol{f})^* = [\boldsymbol{\gamma}](\boldsymbol{f}^*)$ for all $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$ such that $\nabla^a f_{(ab)} = 0$;

(iii) *Symmetry*: $[\boldsymbol{\gamma}](\boldsymbol{f}) = 0$ for all antisymmetric $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$;

(iv) *Field equation*: $[\boldsymbol{\gamma}](L(\boldsymbol{f})) = 0$ for all $\boldsymbol{f} \in C_{TC}^\infty(T_2^0(M;\mathbb{C}))$ such that $L(\boldsymbol{f}) \in C_0^\infty(S_2^0(M;\mathbb{C}))$;

(v) *Commutation relation*: $[[\boldsymbol{\gamma}](\boldsymbol{f}_1),\ [\gamma](\boldsymbol{f}_2)] = -2i\boldsymbol{E}(\boldsymbol{f}_1^s, \overline{\boldsymbol{f}_2^s})\mathbb{1}$ for all $\boldsymbol{f}_i \in C_0^\infty(T_2^0(M;\mathbb{C}))$ such that $\nabla^a (f_i)_{(ab)} = 0$.

One should note that the hermiticity relation is different from its classical counterpart. Here it is expressing the property that the observable should be self-adjoint. This corresponds to the classical field being real-valued. Also, if the $\boldsymbol{f}$'s are spacelike separated then the commutator vanishes, which reflects the Bose statistics of this field.

The algebra of observables is constructed exactly along the same lines as for the scalar field in section 3.2.2. One generates the free unital $*$-algebra using the $[\boldsymbol{\gamma}](\boldsymbol{f})$'s as the generators. The relations (i)-(v) are imposed on this algebra by constructing a $*$-ideal and then quotienting the algebra by this ideal to give the algebra of observables for the spacetime $(M, \boldsymbol{g})$. Also, as was noted in constructing the algebra for the scalar field, one retrieves the local algebras by restricting to those elements of $\mathcal{A}(M, \boldsymbol{g})$ for which the test tensors labelling them have support contained within a chosen open set $\mathscr{O} \subset M$ that has compact closure.

We now verify that the time-slice condition holds for this algebra. This is achieved by proving that a symmetric, divergence-free, smooth, compactly supported test tensor field on the spacetime can be decomposed into a smooth, symmetric, divergence-free test tensor field, which is compactly supported within a connected causally convex neighbourhood of an arbitrary spacelike Cauchy surface, and a term that is the action of the linearized Einstein tensor with cosmological constant on an arbitrary smooth, compactly supported test tensor field. If one combines this decomposition with the fourth relation, then one sees that the algebra of a connected causally convex neighbourhood of a Cauchy surface coincides with the full spacetime algebra. We can restrict to considering purely symmetric smearing tensors due to the third relation of the observables. The decomposition is as follows.

Theorem 6.1.1 *Given a connected causally convex neighbourhood $\mathscr{N}$ of any spacelike Cauchy surface Σ and a $\boldsymbol{f} \in C_0^\infty(S_2^0(M;\mathbb{C}))$ with $\nabla \cdot \boldsymbol{f} = 0$ then there exists a $\tilde{\boldsymbol{f}} \in C_0^\infty(S_2^0(\mathscr{N};\mathbb{C}))$ with $\nabla \cdot \tilde{\boldsymbol{f}} = 0$ and a $\boldsymbol{h} \in C_0^\infty(S_2^0(M;\mathbb{C}))$ such that*

$$\boldsymbol{f} = \tilde{\boldsymbol{f}} + 2L(\boldsymbol{h}). \tag{6.1.1}$$

This entails that $[\boldsymbol{\gamma}](\tilde{\boldsymbol{f}}) = [\boldsymbol{\gamma}](\boldsymbol{f})$ *in* $\mathcal{A}(M, \boldsymbol{g})$.

Proof. Just as for the scalar case, described in Theorem 3.2.7, since $\mathscr{N}$ is a causally convex subset, then it will be a globally hyperbolic subset of M. Therefore [37, Lem 3.1] one can choose two Cauchy surfaces $\Sigma^{\pm}$, which lie to the future/past of Σ and that are both contained within $\mathscr{N}$. Now, take two scalar functions $\chi^{\pm} \in C^{\infty}(M; \mathbb{C})$ satisfying $\chi^{+} = 1$ in $J^{+}(\Sigma^{+})$, $\chi^{+} = 0$ in $J^{-}(\Sigma^{-})$, and $\chi^{+} + \chi^{-} = 1$. (Since $J^{+}(\Sigma^{+})$ and $J^{-}(\Sigma^{-})$ are closed and disjoint sets, then the existence of χ^{+} is guaranteed by [1, Prop. 5.5.8]. One then defines χ^{-} by the condition $\chi^{+} + \chi^{-} = 1$.)

We now follow the method used in the electromagnetic case from [35, Prop. A.3(b)]. Define

$$\tilde{\boldsymbol{f}} := 2L(\chi^{+}\boldsymbol{E}\overline{\boldsymbol{f}}), \tag{6.1.2}$$

which satisfies $\nabla \cdot \tilde{\boldsymbol{f}} = 0$ by the linearized Bianchi identity and is compactly supported within $\mathscr{N}$ (it evidently vanishes to the past of $\mathscr{N}$ and coincides with a de Donder solution to the linearized Einstein equation to the future of $\mathscr{N}$ by the hypothesis placed on $\boldsymbol{f}$). Note that (6.1.2) implies that $2L(\chi^{-}\boldsymbol{E}\overline{\boldsymbol{f}}) = -\tilde{\boldsymbol{f}}$. By Lemma 4.7.7 we have $-\boldsymbol{E}^{+}\overline{\tilde{\boldsymbol{f}}} \sim \chi^{+}\boldsymbol{E}\overline{\boldsymbol{f}}$, $\boldsymbol{E}^{-}\overline{\tilde{\boldsymbol{f}}} \sim \chi^{-}\boldsymbol{E}\overline{\boldsymbol{f}}$ and hence

$$\boldsymbol{E}\overline{\tilde{\boldsymbol{f}}} = \boldsymbol{E}\overline{\boldsymbol{f}} + \pounds_{\boldsymbol{w}}\boldsymbol{g} \tag{6.1.3}$$

for some $\pounds_{\boldsymbol{w}}\boldsymbol{g} \in \mathscr{G}(M; \mathbb{C})$. As $\nabla \cdot \tilde{\boldsymbol{f}} = \nabla \cdot \boldsymbol{f} = 0$, then by Theorem 4.7.4, both $\boldsymbol{E}\overline{\tilde{\boldsymbol{f}}}$ and $\boldsymbol{E}\overline{\boldsymbol{f}}$ are de Donder solutions, so $\boldsymbol{w}$ solves $(\Box + \Lambda)\boldsymbol{w} = 0$ (see the remarks following Theorem 4.5.2); hence by Theorem 2.4.3, $\boldsymbol{w} = \tilde{\boldsymbol{E}}\boldsymbol{v}$ for some $\boldsymbol{v} \in C_0^{\infty}(T_0^1(M; \mathbb{C}))$. Substituting this result into (6.1.3) and using Lemma 4.7.5 gives $\boldsymbol{E}(\overline{\tilde{\boldsymbol{f}}} - \overline{\boldsymbol{f}} - \pounds_{\boldsymbol{v}}\boldsymbol{g}) = 0$. Using Lemma 4.7.8 gives the desired result. ■

We will now consider how the quantum field theory that we have constructed circumvents the problems arising when one attempts to construct a Wightman quantum field theory for the free graviton in Minkowski spacetime. It was pointed out by Strocchi [86] that the quantized linear spacetime perturbation cannot exist as a Lorentz covariant operator-valued distribution. Soon after, another paper [18] showed how the perturbation does not generally satisfy commutativity for spacelike-separations as an operator-valued distribution. The implications of these two results are that one must either abandon two standard requirements, which is in fact what happens when one works in a totally fixed gauge that isolates the true degrees of freedom, or if one insists on their retention, then, as shown in [19], one is forced to use the methodology of the Gupta-Bleuler formalism. This formalism entails the use of a Hilbert space with an indefinite inner product, which thus leads to the introduction of unphysical particles. One then isolates the subspace of physical states by using a global condition, namely that the physical states be annihilated by the positive frequency part of a gauge condition. Note that the requirement of such a condition had already been noted by other authors, see, for instance, the end of [70, Sec. 21]. However, the imposition of a global condition like this raises serious issues when we move away from the convenient setting of Minkowski spacetime and to a general curved spacetime, where it is unclear how to define positive frequency due to the lack of a global Fourier transform.

However, in our case, the restriction placed on our class of smearing tensors, namely that they have divergence-free symmetric part, ensures that our smeared quantum fields will only be defined on the subspace of physical states in the Gupta-Bleuler Hilbert space and that their action leaves the subspace invariant, see [19, Sec. 10]. Therefore our approach of restricting the smearing tensors is much more generally applicable and also immediately picks out the physical subspace.

There exist exactly analogous issues for electromagnetism [85, 87], which again can be circumvented, as above, by Dimock's treatment [28] of restricting the class of smearing tensors. In both Dimock's approach to electromagnetism and the approach to linearized gravity taken here, the serious issues that arise when one allows arbitrary smearings is removed by restricting the smearing tensors to ensure that one only deals with gauge invariant objects.

One should note that, although we have only discussed the Gupta-Bleuler formalism, there do exist other approaches to quantizing gauge theories such as: BRST and Batalin-Vilkovisky, but these methods also involve the introduction of negative-norm states and/or ghost fields.

We now briefly comment on how our approach fits into the framework of locally covariant quantum field theory [22]. As we illustrated in section 3.1, in this framework one deals with two distinct categories: the set of globally hyperbolic spacetimes with a collection of suitable embeddings between them, and the category of unital $*$-algebras with suitable embeddings between them. For the case of linearized gravity, we are forced to place further restrictions upon the category of spacetimes. One has to restrict attention to the subcategory, for which the spacetimes obey the vacuum Einstein equation with cosmological constant, and for which embeddings satisfy the restriction that: if $\psi : M \to N$ is an embedding, then $\psi_* \hat{\mathscr{L}}(M;\mathbb{C}) \subset \hat{\mathscr{L}}(N;\mathbb{C})$ (recall that $\hat{\mathscr{L}}(M;\mathbb{C})$ was defined in section 5.4). This extra condition on the embeddings is due to the consideration of the enlarged pure gauge subspace $\hat{\mathscr{G}}(M;\mathbb{C})$ and with it the consideration of time-compact tensor fields, see [34, Sec. 4.5] for further details. However, even with these restrictions, one still obtains a (covariant) functor from the restricted category of spacetimes to the category of unital $*$-algebras, and so the theory is locally covariant in this sense. Note that the restriction that our spacetimes be cosmological vacuum solutions entails that one cannot formulate the relative Cauchy evolution, where [22, Sec. 4] one would seek to perturb a spacetime using an arbitrary compactly supported perturbation to obtain an expression for the smeared stress-energy tensor of the quantum field. Being unable to formulate the relative Cauchy evolution is therefore linked to the lack of a local stress-energy tensor for gravity. For further details on the local covariance of the graviton field, see [34, Sec. 4.5].

6.2 States

As discussed in section 3.1, a state is a linear functional $\omega : \mathcal{A}(M, \boldsymbol{g}) \to \mathbb{C}$ on the algebra satisfying: (i) the positivity condition, $\omega(A^* A) \geq 0$ for all $A \in \mathcal{A}(M, \boldsymbol{g})$, and (ii) the normalization condition, $\omega(\mathbb{1}) = 1$.

6.2.1 Quantum linearization instabilities

We now briefly discuss how the classical issues related to linearization instability pass over to the quantum case. As discussed in section 4.3, issues arise concerning the physical admissibility of solutions to the linearized Einstein equation if the background spacetime admits a compact Cauchy surface and global Killing vector fields. In which case, a perturbation is deemed physically admissible if certain conserved quantities associated with it vanish.

It was Moncrief [75] who first considered how linearization instabilities would manifest themselves in the quantum case. He utilised Dirac's methodology and imposed the constraints as operator equations on the class of physical states. Due to the quadratic nature of these constraints, a suitable renormalisation prescription will in general be required to define them, although for the case of de Sitter spacetime it has been shown [58, Sec. 4] that such a prescription is unnecessary. With the potential issue of renormalisation in mind we shall postpone considering the linearization instability issue until after we have fully addressed the issue of Hadamard states. However, the punch line of Moncrief's result is that because the classical conserved quantities form a Poisson algebra that is isomorphic to the Lie algebra of the Killing vector fields [75], the annihilation of the physical states by the operator versions of the conserved quantities is equivalent to the physical states being invariant under a unitary representation of the background isometry group of Killing vectors, which generate the conserved quantities. This result is much stronger than the usual requirement of invariance of only the vacuum in Minkowski spacetime, and it is related to the issue of having to consider an observer as part of the system, see [75, 76] for further discussions.

6.2.2 Hadamard condition

In section 3.3, for the case of the scalar field, we gave a discussion and motivation of the Hadamard condition, which places a restriction on the physical states by demanding that the singular behaviour of their two-point function be of a specified form.

The scalar field case had considerable advantages, namely that the equation of motion has hyperbolic form. As we have found out throughout our consideration of linearized gravity, issues are complicated by the linearized Einstein equation being non-hyperbolic. However, we have consistently exploited the close relationship, established in Theorem 4.2.2, between the linearized Einstein tensor with cosmological constant and the hyperbolic differential operator P, defined in (4.2.9). To define the Hadamard condition we shall exploit this relationship again.

To begin with we discuss Hadamard form bi-distributions, $\boldsymbol{W} : C_0^\infty(T_2^0(M;\mathbb{C})) \times C_0^\infty(T_2^0(M;\mathbb{C})) \to \mathbb{C}$, which will also be assumed to be bi-solutions to P. As P is a wave-operator, the results of Sahlmann and Verch [82], who studied the Hadamard condition for vector bundle distributions obeying wave equations, are immediately available to us. Just as in the scalar case, the equivalence relation in the wavefront set is defined by: $(x,\boldsymbol{k}) \sim (x',\boldsymbol{k}')$ if and only if there exists a null geodesic connecting x to x' and $\boldsymbol{k}'$ is the parallel propagation of $\boldsymbol{k}$ along this null geodesic; if $x = x'$, then this reduces to the requirement that $\boldsymbol{k} = \boldsymbol{k}'$. On a spacetime $(M,\boldsymbol{g})$, we will say that a P bi-solution $\boldsymbol{W}$ has Hadamard form if [82, Thm 5.8] its wavefront set takes the prescribed form

$$\mathrm{WF}(\boldsymbol{W}) = \left\{((x,\boldsymbol{k});(x',-\boldsymbol{k}')) \in \dot{T}^*(M\times M) \mid (x,\boldsymbol{k}) \sim (x',\boldsymbol{k}') \text{ with } \boldsymbol{k} \in \overline{V}_x^+\right\},$$

where $\overline{V}_x^+$ denotes the set of future-pointing covectors at x, and the antisymmetric part of $\boldsymbol{W}$ is, modulo[1] smooth terms that vanish on symmetric divergence-free smearing tensor

[1] This condition is necessary to deal with the problems that arise from zero-modes, see chapter 7 for further details, and it has no influence on the wavefront set.

fields, given by

$$\boldsymbol{W}(\boldsymbol{f},\tilde{\boldsymbol{f}}) - \boldsymbol{W}(\tilde{\boldsymbol{f}},\boldsymbol{f}) \doteq -2i\boldsymbol{E}(\boldsymbol{f},\tilde{\boldsymbol{f}}),$$

where $\boldsymbol{E}$ is the advanced-minus-retarded solution operator associated with P, and '$\doteq$' denotes equality up to smooth terms that vanish when smeared against symmetric divergence-free tensor fields.

In order to give our definition of Hadamard states, we need to define two new concepts, namely the trace and trace-reversal of a bi-distribution that acts on rank $(0,2)$ test-tensor fields. On a spacetime $(M,\boldsymbol{g})$, given a bi-distribution $\boldsymbol{W} : C_0^\infty(T_2^0(M;\mathbb{C})) \times C_0^\infty(T_2^0(M;\mathbb{C})) \to \mathbb{C}$, one defines its trace to be the scalar bi-distribution $Tr\boldsymbol{W} : C_0^\infty(M;\mathbb{C}) \times C_0^\infty(M;\mathbb{C}) \to \mathbb{C}$ given by

$$Tr\boldsymbol{W}(f_1,f_2) := \boldsymbol{W}(f_1\boldsymbol{g}, f_2\boldsymbol{g}), \tag{6.2.1}$$

where $f_1, f_2 \in C_0^\infty(M;\mathbb{C})$. This notion of trace allows us to define the trace-reversal of $\boldsymbol{W}$ to be the bi-distribution $\overline{\boldsymbol{W}} : C_0^\infty(T_2^0(M;\mathbb{C})) \times C_0^\infty(T_2^0(M;\mathbb{C})) \to \mathbb{C}$ given by

$$\overline{\boldsymbol{W}}(\boldsymbol{f}_1,\boldsymbol{f}_2) := \boldsymbol{W}(\boldsymbol{f}_1,\boldsymbol{f}_2) - \frac{1}{8}Tr\boldsymbol{W}(Tr\boldsymbol{f}_1, Tr\boldsymbol{f}_2), \tag{6.2.2}$$

where $Tr\boldsymbol{f} := g^{ab}f_{ab}$. The choice of coefficient of the $Tr\boldsymbol{W}$ term in $\overline{\boldsymbol{W}}$ comes from the requirement that $\overline{\overline{\boldsymbol{W}}} = \boldsymbol{W}$. To see that $\overline{\overline{\boldsymbol{W}}} = \boldsymbol{W}$, we show that $Tr\overline{\boldsymbol{W}} = -Tr\boldsymbol{W}$, from which the result $\overline{\overline{\boldsymbol{W}}} = \boldsymbol{W}$ is then immediate. Using the definition of Tr from (6.2.1), we see that, given any $f_1, f_2 \in C_0^\infty(M;\mathbb{C})$,

$$\begin{aligned} Tr\overline{\boldsymbol{W}}(f_1,f_2) = \overline{\boldsymbol{W}}(f_1\boldsymbol{g}, f_2\boldsymbol{g}) &= \boldsymbol{W}(f_1\boldsymbol{g}, f_2\boldsymbol{g}) - 2Tr\boldsymbol{W}(f_1,f_2) \\ &= Tr\boldsymbol{W}(f_1,f_2) - 2Tr\boldsymbol{W}(f_1,f_2) = -Tr\boldsymbol{W}(f_1,f_2), \end{aligned}$$

where in the second equality we have used the definition of $\overline{\boldsymbol{W}}$ from (6.2.2) and that $g_{ab}g^{ab} = 4$. The third equality just uses the definition of Tr from (6.2.1).

One also needs to consider: if $\boldsymbol{W}$ is a P bi-solution, then will $Tr\boldsymbol{W}$ be a bisolution to $(\Box + 2\Lambda)$ [the trace of P], and will $\overline{\boldsymbol{W}}$ be a P bi-solution too? The answer in both cases is yes and we now prove this. Let $\boldsymbol{W}$ be a P bi-solution, therefore $\boldsymbol{W}(P(\boldsymbol{f}),\tilde{\boldsymbol{f}}) = 0$ for all $\boldsymbol{f},\tilde{\boldsymbol{f}} \in C_0^\infty(T_2^0(M;\mathbb{C}))$. Using the properties of the trace of P, from (4.2.10), one can show that

$$Tr\boldsymbol{W}((\Box + 2\Lambda)f, \tilde{f}) = \boldsymbol{W}((\Box + 2\Lambda)f\boldsymbol{g}, \tilde{f}\boldsymbol{g}) = \boldsymbol{W}(P(f\boldsymbol{g}), \tilde{f}\boldsymbol{g}) = 0,$$

for all $f, \tilde{f} \in C_0^\infty(M;\mathbb{C})$. Hence, $Tr\boldsymbol{W}$ is a $(\Box + 2\Lambda)$ bi-solution. For the trace-reversal,

$$\overline{\boldsymbol{W}}(P(\boldsymbol{f}), \tilde{\boldsymbol{f}}) = \boldsymbol{W}(P(\boldsymbol{f}), \tilde{\boldsymbol{f}}) - \frac{1}{8} Tr\boldsymbol{W}(Tr(P(\boldsymbol{f})), Tr(\tilde{\boldsymbol{f}}) = -\frac{1}{8} Tr\boldsymbol{W}((\Box+2\Lambda)Tr(\boldsymbol{f}), Tr(\tilde{\boldsymbol{f}}) = 0$$

for all $\boldsymbol{f}, \tilde{\boldsymbol{f}} \in C_0^\infty(T_2^0(M;\mathbb{C}))$. Therefore $\overline{\boldsymbol{W}}$ is a P bi-solution too.

These definitions allow one to understand more clearly what is happening in the commutator, where one of the arguments in the bi-distribution $\boldsymbol{E}$ has a trace-reversal. Relation (v) in section 6.1 states that the commutator is equal to the bi-distribution $-2i\boldsymbol{E}(\boldsymbol{f}_1^s, \overline{\boldsymbol{f}_2^s}) = -2i\boldsymbol{E}(\boldsymbol{f}_1^s, \boldsymbol{f}_2^s) + iE(Tr\boldsymbol{f}_1^s, Tr\boldsymbol{f}_2^s)$, for the details of this expansion see equation (5.4.14). We will now show how this is in fact just the trace-reversal of the bi-distribution $\boldsymbol{E}$. First we calculate $Tr\boldsymbol{E}$. Using the definition of $\boldsymbol{E}$, given any $\boldsymbol{f}_1, \boldsymbol{f}_2 \in C_0^\infty(T_2^0(M;\mathbb{C}))$, one sees that

$$\begin{aligned} Tr\boldsymbol{E}(Tr\boldsymbol{f}_1, Tr\boldsymbol{f}_2) = \boldsymbol{E}((Tr\boldsymbol{f}_1)\boldsymbol{g}, (Tr\boldsymbol{f}_2)\boldsymbol{g}) &= \int_M (Tr\boldsymbol{f}_1) g^{ab} E_{ab}{}^{c'd'} (Tr\boldsymbol{f}_1) g_{c'd'} dvol_{\boldsymbol{g}} \\ &= 4\int_M (Tr\boldsymbol{f}_1) E(Tr\boldsymbol{f}_2) dvol_{\boldsymbol{g}} \\ &= 4E(Tr\boldsymbol{f}_1, Tr\boldsymbol{f}_2), \end{aligned}$$

where we have used Lemma 4.7.2 in the third equality, and the fourth equality defines the scalar bi-distribution E. Combining this result with the definition (6.2.2) of the trace-reversal of a bi-distribution, we see that

$$\overline{\boldsymbol{E}}(\boldsymbol{f}_1, \boldsymbol{f}_2) = \boldsymbol{E}(\boldsymbol{f}_1, \boldsymbol{f}_2) - \frac{1}{2} E(Tr\boldsymbol{f}_1, Tr\boldsymbol{f}_2) = \boldsymbol{E}(\boldsymbol{f}_1, \overline{\boldsymbol{f}_2}).$$

Hence, the commutator, from relation (v) of our list of algebraic relations, is in fact given by

$$[[\gamma](\boldsymbol{f}_1), [\gamma](\boldsymbol{f}_2)] = -2i\overline{\boldsymbol{E}}(\boldsymbol{f}_1^s, \boldsymbol{f}_2^s).$$

We are now in a position to define the Hadamard states for the free graviton quantum field. We will restrict attention to quasi-free states, that is, states whose n-point functions vanish if n is odd, and are completely specified by the two-point function if n is even.

Definition 6.2.1 *A quasi-free state $\omega : \mathcal{A}(M, \boldsymbol{g}) \to \mathbb{C}$ will be said to be a Hadamard state if there exists a Hadamard form P bi-solution $\boldsymbol{W} : C_0^\infty(T_2^0(M;\mathbb{C})) \times C_0^\infty(T_2^0(M;\mathbb{C})) \to \mathbb{C}$ such that*

$$\omega([\gamma](\boldsymbol{f}_1)[\gamma](\boldsymbol{f}_2)) = \overline{\boldsymbol{W}}(\boldsymbol{f}_1^s, \boldsymbol{f}_2^s)$$

for all $\boldsymbol{f}_1, \boldsymbol{f}_2 \in C_0^\infty(T_2^0(M;\mathbb{C}))$ satisfying $\nabla \cdot \boldsymbol{f}_1^s = \nabla \cdot \boldsymbol{f}_2^s = 0$.

Having given a definition of Hadamard states, the focus now shifts to finding whether there exist any states that satisfy the definition. Unfortunately, unlike for the cases of the scalar field [47] and the electromagnetic field [35, Sec. IV.E], the deformation arguments of Fulling, Narcowich, Wald [47, App. C] are not available here due to the restriction to cosmological vacuum spacetimes. This entails that we cannot just prove existence of Hadamard states on ultrastatic globally hyperbolic spacetimes to obtain existence of Hadamard states on general globally hyperbolic spacetimes. With the lack of availability of an alternative method, we are forced to consider existence on a case by case basis. We will show, using methods from Fourier analysis, that the standard Fock vacuum state (this is a quasi-free state) from Minkowski spacetime is a Hadamard state.

In Minkowski spacetime, the Riemann tensor vanishes and so P reduces to $\Box$. We will work in global inertial coordinates (t, x, y, z). To construct a tensor Hadamard $\Box$ bi-solution, we will use the standard scalar $\Box$-Hadamard bi-solution, which is just the massless-scalar field two-point function:

$$W^{scalar}(x, x') = \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3 2\omega} e^{-ik\cdot(x-x')}, \tag{6.2.3}$$

where $\omega := |\vec{k}|$. We define our tensor Hadamard $\Box$ bi-solution to be

$$\boldsymbol{W}(\boldsymbol{f}, \tilde{\boldsymbol{f}}) := 2\eta^{\mu\mu'}\eta^{\nu\nu'} W^{scalar}(f_{\mu\nu}, \tilde{f}_{\mu'\nu'}),$$

where $\boldsymbol{f}, \tilde{\boldsymbol{f}} \in C_0^\infty(T_2^0(M; \mathbb{C}))$, and W^{scalar} acts on the individual components $f_{\mu\nu}$, which are scalar functions. The wavefront set of $\boldsymbol{W}$ is the same as the wavefront set for W^{scalar} and therefore has the prescribed form for a Hadamard bi-solution, but we must check that it has the correct antisymmetric part.

The antisymmetric part of $\boldsymbol{W}$ is

$$\begin{aligned}\boldsymbol{W}(\boldsymbol{f}, \tilde{\boldsymbol{f}}) - \boldsymbol{W}(\tilde{\boldsymbol{f}}, \boldsymbol{f}) &= 2\eta^{\mu\mu'}\eta^{\nu\nu'} \left[W^{scalar}(f_{\mu\nu}, \tilde{f}_{\mu'\nu'}) - W^{scalar}(\tilde{f}_{\mu\nu}, f_{\mu'\nu'})\right] \\ &= 2\eta^{\mu\mu'}\eta^{\nu\nu'} \int_{\mathbb{R}^4\times\mathbb{R}^4} d^4x d^4x' \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3 2\omega} e^{-ik\cdot(x-x')} \left[f_{\mu\nu}(x)\tilde{f}_{\mu'\nu'}(x') - \tilde{f}_{\mu\nu}(x) f_{\mu'\nu'}(x')\right].\end{aligned}$$

On the second collection of terms, one can interchange the integration variables x and x', relabel the indices $\mu \leftrightarrow \mu'$ and $\nu \leftrightarrow \nu'$, and use symmetry of the metric $\boldsymbol{\eta}$ to obtain

$$\boldsymbol{W}(\boldsymbol{f}, \tilde{\boldsymbol{f}})-\boldsymbol{W}(\tilde{\boldsymbol{f}}, \boldsymbol{f}) = 2\eta^{\mu\mu'}\eta^{\nu\nu'} \int_{\mathbb{R}^4\times\mathbb{R}^4} d^4x d^4x' \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3 2\omega} \left[e^{-ik\cdot(x-x')} - e^{ik\cdot(x-x')}\right] f_{\mu\nu}(x)\tilde{f}_{\mu'\nu'}(x'). \tag{6.2.4}$$

Using the form of the scalar two-point function from (6.2.3) and the scalar commutation relation (relation (iv) from section 3.2.2), it is clear that:

$$-iE^{scalar}(f, \tilde{f}) = \int_{\mathbb{R}^4\times\mathbb{R}^4} d^4x d^4x' \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3 2\omega} \left[e^{-ik\cdot(x-x')} - e^{ik\cdot(x-x')}\right] f(x)\tilde{f}(x'),$$

where $f, \tilde{f} \in C_0^\infty(M; \mathbb{C})$. Substituting this expression into (6.2.4) yields

$$\boldsymbol{W}(\boldsymbol{f}, \tilde{\boldsymbol{f}}) - \boldsymbol{W}(\tilde{\boldsymbol{f}}, \boldsymbol{f}) = -2i\eta^{\mu\mu'}\eta^{\nu\nu'}E^{scalar}(f_{\mu\nu}, \tilde{f}_{\mu'\nu'}).$$

In this instance, it holds that[2] $\boldsymbol{E}(\boldsymbol{f}, \tilde{\boldsymbol{f}}) = \eta^{\mu\mu'}\eta^{\nu\nu'}E^{scalar}(f_{\mu\nu}, \tilde{f}_{\mu'\nu'})$ and so $\boldsymbol{W}$ has the correct antisymmetric part, namely:

$$\boldsymbol{W}(\boldsymbol{f}, \tilde{\boldsymbol{f}}) - \boldsymbol{W}(\tilde{\boldsymbol{f}}, \boldsymbol{f}) = -2i\boldsymbol{E}(\boldsymbol{f}, \tilde{\boldsymbol{f}})$$

for all $\boldsymbol{f}, \tilde{\boldsymbol{f}} \in C_0^\infty(T_2^0(M; \mathbb{C}))$, as there is no zero-mode problem in this instance.

Having established that our $\boldsymbol{W}$ has the correct antisymmetric part, we now need to compute its trace-reversal to see if it agrees with the two-point function. Given any $f, \tilde{f} \in C_0^\infty(M; \mathbb{C})$, the trace of $\boldsymbol{W}$ is

$$Tr\boldsymbol{W}(f, \tilde{f}) = \boldsymbol{W}(f\boldsymbol{\eta}, \tilde{f}\boldsymbol{\eta}) = 8W^{scalar}(f, \tilde{f}),$$

and therefore, using the definition (6.2.2) of the trace-reversal, we have

$$\overline{\boldsymbol{W}}(\boldsymbol{f}, \tilde{\boldsymbol{f}}) = (2\eta^{\mu\mu'}\eta^{\nu\nu'} - \eta^{\mu\nu}\eta^{\mu'\nu'})W^{scalar}(f_{\mu\nu}, \tilde{f}_{\mu'\nu'}).$$

Note that when the $\boldsymbol{f}$'s are symmetric, one can write the expression of $\boldsymbol{\eta}$'s as three terms, by moving the symmetrisation from the $\boldsymbol{f}$'s to the $\boldsymbol{\eta}$'s. This ensures that it takes a form just like the two-point function, which is stated below in (6.2.5).

The graviton two-point function in the Fock vacuum state in Minkowski spacetime

[2] This equality holds by uniqueness of Green's operators. The right-hand side of this expression is an advanced-minus-retarded solution operator on elements of $C_0^\infty(T_2^0(M; \mathbb{C}))$.

is [62, eq. (A1)] given by

$$\omega_{\mu\nu\mu'\nu'}(x,x') = \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3 2\omega} e^{-ik\cdot(x-x')} (h_{\mu\mu'}h_{\nu\nu'} + h_{\mu\nu'}h_{\nu\mu'} - h_{\mu\nu}h_{\mu'\nu'}), \tag{6.2.5}$$

where

$$h_{\mu\nu} = \begin{cases} \eta_{\mu\nu} - \frac{k_\mu k_\nu}{\omega^2} & \text{if } \mu \neq 0 \neq \nu, \\ 0 & \text{otherwise.} \end{cases}$$

For $\boldsymbol{f}$ such that $\nabla \cdot \boldsymbol{f}^s = 0$, or equally, $\boldsymbol{k} \cdot \hat{\boldsymbol{f}}^s = 0$, one can easily show that

$$\int_{\mathbb{R}^4 \times \mathbb{R}^4} \omega_{\mu\nu\mu'\nu'}(x,x') f^{\mu\nu}(x) \tilde{f}^{\mu'\nu'} d^4x d^4x' = \overline{\boldsymbol{W}}(\boldsymbol{f}^s, \tilde{\boldsymbol{f}}^s)$$

for all $\boldsymbol{f}, \tilde{\boldsymbol{f}} \in C_0^\infty(T_2^0(M;\mathbb{C}))$ such that $\nabla \cdot \boldsymbol{f}^s = 0 = \nabla \cdot \tilde{\boldsymbol{f}}^s$. Therefore we have shown that the two-point function in the Fock vacuum state agrees with our Hadamard bi-solution as per Definition 6.2.1. What remains to be verified is that the Fock vacuum is a state on our algebra of observables.

We now verify that this two-point function does define a state on our algebra $\mathcal{A}(M, \boldsymbol{g})$. We choose an element $[\boldsymbol{\gamma}](\boldsymbol{f}) \in \mathcal{A}(M, \boldsymbol{g})$, where $\boldsymbol{f} \in C_0^\infty(T_2^0(M;\mathbb{C}))$ satisfies $\nabla \cdot \boldsymbol{f}^s = 0$, to be represented by the smearing of the standard Fock space construction of the field operator in the transverse-traceless and synchronous gauge,

$$\hat{\gamma}(\boldsymbol{f}) = \int_{\mathbb{R}^4} d^4x \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^3 2\omega} \sum_j \left(\epsilon^j_{\mu\nu}(k) a_j(k) e^{-ik.x} + \epsilon^{j*}_{\mu\nu}(k) a^\dagger_j(k) e^{ik.x} \right) f^{\mu\nu}(x), \tag{6.2.6}$$

where the polarisation tensors $\epsilon^j_{\mu\nu}$ are symmetric on their μ, ν indices, and they satisfy $\epsilon^i_{\mu\nu}(k)k^\mu = 0$, $\eta^{\mu\nu}\epsilon^i_{\mu\nu}(k) = 0$ and $\epsilon^i_{\mu\nu}(k)n^\mu = 0$, which respectively represent the gauge conditions: transverse, traceless and synchronous. Note that $\boldsymbol{n}$ is the future-pointing unit normal to the $t = \textit{constant}$ Cauchy surfaces. To verify that the field operators (6.2.6) provide a representation of the algebra $\mathcal{A}(M, \boldsymbol{g})$, one needs [35, Sec. IV.C] to confirm that they obey all five relations from section 6.1 on a dense domain of the Fock space. This is easily done, and so upon using the standard GNS construction, the Fock vacuum $|0\rangle$ will define a state on $\mathcal{A}(M, \boldsymbol{g})$, and its two-point function satisfies

$$\langle 0 | [\boldsymbol{\gamma}](\boldsymbol{f}_1)[\boldsymbol{\gamma}](\boldsymbol{f}_2) | 0 \rangle = \overline{\boldsymbol{W}}(\boldsymbol{f}_1^s, \boldsymbol{f}_2^s)$$

for all $\boldsymbol{f}_1, \boldsymbol{f}_2 \in C_0^\infty(T_2^0(M;\mathbb{C}))$ obeying $\nabla \cdot \boldsymbol{f}_1^s = 0 = \nabla \cdot \boldsymbol{f}_2^s$. Hence by Definition 6.2.1, $|0\rangle$

is a Hadamard state.

Chapter 7

Conclusion

In this thesis we have achieved all but one of the goals set out in the introduction. We constructed a quantum field theory of the free graviton within the rigorous mathematical framework of algebraic quantum field theory. This required us to fully consider the classical theory of linearized gravity in chapter 4. In particular, we considered when various choices of gauge can be made. We found that there exists a topological obstruction to achieving the well-known transverse-traceless gauge whenever one works with spacelike-compact perturbations on vacuum spacetimes with vanishing cosmological constant. We also found that within a normal neighbourhood of any Cauchy surface, one can gauge transform a perturbation to the synchronous gauge with the synchronous condition set by the normal field on that neighbourhood.

In chapter 5, we initially formulated the classical phase space for linearized gravity as the complexified space of solutions to the linearized Einstein equation with a pre-symplectic product. When it came to examining the form of the degeneracies of this product, we found, by generalising Moncrief's splitting results [72], that if a spacetime admits a compact Cauchy surface, then the degeneracies are just the pure gauge solutions. Therefore for the case of a spacetime that admits a compact Cauchy surface, the phase space is just the collection of gauge equivalence classes of solutions. Unfortunately, for the case of a spacetime only admitting non-compact Cauchy surfaces, a proof was not forthcoming. This is due to the Moncrief splitting using results from elliptic theory, namely Sobolev spaces, that rely on compactness of the underlying manifold. However, it may be possible to establish such a splitting for certain types of non-compact Cauchy surface, such as asymptotically flat ones, where it is possible to introduce weighted Sobolev spaces and certain decomposition theorems exist [23]. However, for a general Cauchy surface this is probably not possible, but the author hopes to consider these issues further in the future.

We were thus led to conjecture that the degeneracies were pure gauge in the non-compact case too, and consequently that our symplectic product was weakly non-degenerate on the space of gauge equivalence classes of solutions to the linearized Einstein equation.

Having established a phase space, our attention moved to the observables of the theory. We defined the standard smeared fields and found that for them to be gauge invariant, the class of smearing tensors needed to be restricted to those whose symmetric part has vanishing divergence. The Poisson bracket of the observables was explicitly calculated and found to agree with a result previously posited by Lichnerowicz [70].

Dirac's quantization prescription was then used to obtain the algebraic relations obeyed by our quantum observables and we constructed an algebra of observables for arbitrary globally hyperbolic cosmological vacuum spacetimes. We described how this circumvents the problems described by Strocchi [18, 86], and how it fits into the framework of locally covariant quantum field theory established by Brunetti, Fredenhagen and Verch [22].

This setup allowed for a precise definition of Hadamard states to be given using techniques from microlocal analysis. We then showed how the vacuum state in the standard Fock space construction in Minkowski spacetime is, by our definition, a Hadamard state on our algebra. Unfortunately, we were not able to show here whether or not the de Sitter invariant vacuum state is a Hadamard state on our algebra. There have been numerous papers written about the graviton two-point function in de Sitter spacetime [2, 4, 59, 60], and our goal is construct a suitable Hadamard P bi-solution $\boldsymbol{W}$, which agrees with the two-point function on our restricted class of smearing tensor fields. The idea would be to proceed along the lines set out in [35, App. B] for the case of electromagnetism and construct the bi-solution as a mode expansion. One might be concerned about potential issues with the mode expansion if there are zero modes present. However, even if they are present, one can still obtain a well-defined Hadamard form P bi-solution by cutting out these troublesome zero-modes. These modes are smooth, so their removal will not affect the wavefront set and hence not affect whether the bi-solution is Hadamard. In a mode expansion of the propagator the zero-modes will still be present, and so the antisymmetric part of the Hadamard bi-solution without zero-modes will not agree with the propagator expansion, as they will differ by the zero-mode terms. This is rectified by noticing that when we restrict to using tensors with divergence-free symmetric part in the propagator, the terms with a zero-mode smeared against such tensors vanish and so the two expressions will in fact agree on the class of smearing tensors that we use. This explains our definition of Hadamard bi-solutions given earlier. Note that there is an analogous situation concerning zero-modes in electromagnetism, see [35]. Therefore if one can construct the Hadamard bi-solution $\boldsymbol{W}$ and show that it agrees with the graviton two-point function for the de Sitter invariant vacuum state for our class of smearing tensors, and show that the associated state does define a state on our algebra, then the de Sitter invariant vacuum state will be a Hadamard state on our algebra. The author hopes to finally resolve this in the very near future.

We have seen throughout our discussion of linearized gravity that there is a close relationship between it and the case of electromagnetism. The point where the two systems seem to diverge is when it comes to proving non-degeneracy of their symplectic products. As we briefly discussed, recently a new method [67] has been proposed that claims to prove non-degeneracy, for the case of compactly supported initial data on a non-compact Cauchy surface, by using methods from cohomology theory. It would be interesting to find out whether a similar theory exists for the case of the symmetric tensor fields used in linearized gravity. One can see the beginnings of such a theory if, for instance, one defines for any $\boldsymbol{\gamma} \in C_0^\infty(S_2^0(M;\mathbb{C}))$, a 'codifferential' δ to be

$$\delta\gamma := -\nabla \cdot \overline{\boldsymbol{\gamma}}.$$

Similarly, one can define an 'exterior derivative' d on vector fields to be

$$d\boldsymbol{w} := \pounds_{\boldsymbol{w}}\boldsymbol{g}.$$

Combining these two, one finds that

$$d\delta\boldsymbol{\gamma} = -\pounds_{\nabla\cdot\overline{\boldsymbol{\gamma}}}\boldsymbol{g}.$$

Now using our decomposition of the linearized Einstein tensor with cosmological constant from Theorem 4.2.2, one sees that

$$2\overline{L(\boldsymbol{\gamma})} = -P(\boldsymbol{\gamma}) - d\delta\boldsymbol{\gamma}.$$

Using analogy with electromagentism, see (4.2.11), one is led to define

$$\delta d\boldsymbol{\gamma} := 2\overline{L(\boldsymbol{\gamma})},$$

and attempt to obtain δ and d for higher rank tensors. However, as yet we do not have any explicit formulas for these objects. A concern here is that symmetry will prevent the standard result from differential forms that $d^2 = 0$, and so any possible link with cohomology theory would be broken. Therefore it remains to be seen whether there is something in this, and it is something which the author would like to return to in the future.

Bibliography

[1] Abraham, R., Marsden, J. E., and Ratiu, T., *Manifolds, tensor analysis, and applications*, *Applied Mathematical Sciences*, Vol. 75, 2nd edn. (Springer-Verlag, New York, 1988).

[2] Allen, B., Graviton propagator in de Sitter space, *Phys. Rev. D (3)* **34** (1986) 3670–3675.

[3] Allen, B., Folacci, A., and Ottewill, A. C., Renormalized graviton stress-energy tensor in curved vacuum space-times, *Phys. Rev. D* **38** (1988) 1069–1082.

[4] Allen, B. and Turyn, M., An evaluation of the graviton propagator in de sitter space, *Nuclear Physics B* **292** (1987) 813 – 852.

[5] Araki, H., *Mathematical theory of quantum fields*, *International Series of Monographs on Physics*, Vol. 101 (Oxford University Press, Oxford, 2009), Reprint of the 1999 edition.

[6] Arms, J. M., Marsden, J. E., and Moncrief, V., The structure of the space of solutions of Einstein's equations. II. Several Killing fields and the Einstein-Yang-Mills equations, *Ann. Physics* **144** (1982) 81–106.

[7] Arnowitt, R., Deser, S., and Misner, C. W., The dynamics of general relativity, in *Gravitation: An introduction to current research* (Wiley, New York, 1962), pp. 227–265.

[8] Ashtekar, A., New variables for classical and quantum gravity, *Phys. Rev. Lett.* **57** (1986) 2244–2247.

[9] Ashtekar, A. and Magnon-Ashtekar, A., On the symplectic structure of general relativity, *Commun. Math. Phys.* **86** (1982) 55–68.

[10] Bär, C. and Fredenhagen, K. (eds.), *Quantum field theory on curved spacetimes: Concepts and mathematical foundations*, *Lecture Notes in Physics*, Vol. 786 (Springer-Verlag, Berlin, 2009).

[11] Bär, C., Ginoux, N., and Pfäffle, F., *Wave equations on Lorentzian manifolds and quantization*, ESI Lectures in Mathematics and Physics (European Mathematical Society (EMS), Zürich, 2007).

[12] Berger, M. and Ebin, D., Some decompositions of the space of symmetric tensors on a Riemannian manifold, *J. Differential Geometry* **3** (1969) 379–392.

[13] Bernal, A. N. and Sánchez, M., On smooth Cauchy hypersurfaces and Geroch's splitting theorem, *Commun. Math. Phys.* **243** (2003) 461–470.

[14] Bernal, A. N. and Sánchez, M., Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes, *Commun. Math. Phys.* **257** (2005) 43–50.

[15] Bernal, A. N. and Sánchez, M., Further results on the smoothability of Cauchy hypersurfaces and Cauchy time functions, *Lett. Math. Phys.* **77** (2006) 183–197.

[16] Bernal, A. N. and Sánchez, M., Globally hyperbolic spacetimes can be defined as 'causal' instead of 'strongly causal', *Class. Quantum Grav.* **24** (2007) 745–749.

[17] Birrell, N. D. and Davies, P. C. W., *Quantum fields in curved space*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 1984), corrected reprint of the 1982 original.

[18] Bracci, L. and Strocchi, F., Einstein's equations and locality, *Commun. Math. Phys.* **24** (1972) 289–302.

[19] Bracci, L. and Strocchi, F., Local and covariant quantization of linearized Einstein's equations, *J. Math. Phys.* **13** (1972) 1151–1163.

[20] Brill, D. R. and Deser, S., Instability of closed spaces in general relativity, *Commun. Math. Phys.* **32** (1973) 291–304.

[21] Brunetti, R., Fredenhagen, K., and Köhler, M., The microlocal spectrum condition and Wick polynomials of free fields on curved spacetimes, *Commun. Math. Phys.* **180** (1996) 633–652.

[22] Brunetti, R., Fredenhagen, K., and Verch, R., The generally covariant locality principle: A new paradigm for local quantum physics, *Commun. Math. Phys.* **237** (2003) 31–68.

[23] Cantor, M., Elliptic operators and the decomposition of tensor fields, *Bull. Amer. Math. Soc. (N.S.)* **5** (1981) 235–262.

[24] Choquet-Bruhat, Y., DeWitt-Morette, C., and Dillard-Bleick, M., *Analysis, manifolds and physics*, 2nd edn. (North-Holland Publishing Co., Amsterdam, 1982).

[25] DeWitt, B. S., Quantum field theory in curved spacetime, *Phys. Rep.* **19** (1975) 295 – 357.

[26] Dimock, J., Algebras of local observables on a manifold, *Commun. Math. Phys.* **77** (1980) 219–228.

[27] Dimock, J., Dirac quantum fields on a manifold, *Trans. Amer. Math. Soc.* **269** (1982) 133–147.

[28] Dimock, J., Quantized electromagnetic field on a manifold, *Rev. Math. Phys.* **4** (1992) 223–233.

[29] Dirac, P. A. M., *The Principles of Quantum Mechanics* (Oxford, at the Clarendon Press, 1958), 4th ed.

[30] Dirac, P. A. M., The theory of gravitation in Hamiltonian form, *Proc. Roy. Soc. (London) Ser. A* **246** (1958) 333–343.

[31] Duistermaat, J. J. and Hörmander, L., Fourier integral operators. II, *Acta Math.* **128** (1972) 183–269.

[32] Fewster, C. J., Lectures on quantum energy inequalities, arXiv:1208.5399.

[33] Fewster, C. J., Lectures on quantum field theory in curved spacetime (2008).

[34] Fewster, C. J. and Hunt, D. S., Quantization of linearized gravity in cosmological vacuum spacetimes, arXiv:1203.0261, submitted to Reviews in Mathematical Physics.

[35] Fewster, C. J. and Pfenning, M. J., A quantum weak energy inequality for spin-one fields in curved space-time, *J. Math. Phys.* **44** (2003) 4480–4513.

[36] Fewster, C. J. and Verch, R., Dynamical locality and covariance: What makes a physical theory the same in all spacetimes?, arXiv:1106.4785, to appear in Annales H. Poincaré.

[37] Fewster, C. J. and Verch, R., Dynamical locality of the free scalar field, arXiv:1109.6732, to appear in Annales H. Poincaré.

[38] Fierz, M. and Pauli, W., On relativistic wave equations for particles of arbitrary spin in an electromagnetic field, *Proc. Roy. Soc. (London) Ser. A.* **173** (1939) 211–232.

[39] Fischer, A. and Marsden, J., A new hamiltonian structure for the dynamics of general relativity, *General Relativity and Gravitation* **7** (1976) 915–920.

[40] Fischer, A. E. and Marsden, J. E., The initial value problem and the dynamical formulation of general relativity, in *General Relativity: An Einstein centenary survey*, ed. S. W. Hawking & W. Israel (1979), pp. 138–211.

[41] Fischer, A. E. and Marsden, J. E., Topics in the dynamics of general relativity, in *Isolated Gravitating Systems in General Relativity*, ed. J. Ehlers (1979), pp. 322–395.

[42] Ford, L. H. and Parker, L., Infrared divergences in a class of Robertson-Walker universes, *Phys. Rev. D* **16** (1977) 245–250.

[43] Ford, L. H. and Parker, L., Quantized gravitational wave perturbations in Robertson-Walker universes, *Phys. Rev. D* **16** (1977) 1601–1608.

[44] Friedlander, F. G., *The wave equation on a curved space-time* (Cambridge University Press, Cambridge, 1975), Cambridge Monographs on Mathematical Physics, No. 2.

[45] Friedman, J. L., Generic instability of rotating relativistic stars, *Commun. Math. Phys.* **62** (1978) 247–278.

[46] Fulling, S. A., Nonuniqueness of canonical field quantization in Riemannian space-time, *Phys. Rev. D* **7** (1973) 2850–2862.

[47] Fulling, S. A., Narcowich, F. J., and Wald, R. M., Singularity structure of the two-point function in quantum field theory in curved spacetime. II, *Ann. Physics* **136** (1981) 243–272.

[48] Furlani, E. P., Quantization of massive vector fields in curved space–time, *J. Math. Phys.* **40** (1999) 2611–2626.

[49] Geroch, R., Spinor structure of space-times in general relativity. I, *J. Math. Phys.* **9** (1968) 1739–1744.

[50] Geroch, R., Limits of spacetimes, *Commun. Math. Phys.* **13** (1969) 180–193.

[51] Geroch, R., Domain of dependence, *J. Math. Phys.* **11** (1970) 437–449.

[52] Haag, R., *Local quantum physics*, Texts and Monographs in Physics, 2nd edn. (Springer-Verlag, Berlin, 1996), fields, particles, algebras.

[53] Haag, R. and Kastler, D., An algebraic approach to quantum field theory, *J. Math. Phys.* **5** (1964) 848–861.

[54] Haag, R. and Schroer, B., Postulates of quantum field theory, *J. Math. Phys.* **3** (1962) 248–256.

[55] Hack, T.-P. and Schenkel, A., Linear bosonic and fermionic quantum gauge theories on curved spacetimes, arXiv:1205.3484.

[56] Hawking, S. W., Particle creation by black holes, *Commun. Math. Phys.* **43** (1975) 199–220.

[57] Hawking, S. W. and Ellis, G. F. R., *The large scale structure of space-time* (Cambridge University Press, London, 1973), Cambridge Monographs on Mathematical Physics, No. 1.

[58] Higuchi, A., Quantum linearization instabilities of de sitter spacetime. I, *Class. Quantum Grav.* **8** (1991) 1961.

[59] Higuchi, A. and Kouris, S. S., Large-distance behaviour of the graviton two-point function in de Sitter spacetime, *Class. Quantum Grav.* **17** (2000) 3077–3090.

[60] Higuchi, A. and Kouris, S. S., The covariant graviton propagator in de Sitter spacetime, *Class. Quantum Grav.* **18** (2001) 4317–4327.

[61] Higuchi, A., Marolf, D., and Morrison, I. A., de Sitter invariance of the dS graviton vacuum, *Class. Quantum Grav.* **28** (2011) 245012.

[62] Higuchi, A. and Weeks, R. H., The physical graviton two-point function in de Sitter spacetime with S^3 spatial sections, *Class. Quantum Grav.* **20** (2003) 3005–3021.

[63] Hörmander, L., *The analysis of linear partial differential operators. I*, Classics in Mathematics (Springer-Verlag, Berlin, 2003), Distribution theory and Fourier analysis, Reprint of the second (1990) edition.

[64] Jackson, J. D., *Classical electrodynamics*, 2nd edn. (John Wiley & Sons Inc., New York, 1975).

[65] Kay, B. S. and Wald, R. M., Theorems on the uniqueness and thermal properties of stationary, nonsingular, quasifree states on spacetimes with a bifurcate Killing horizon, *Phys. Rep.* **207** (1991) 49–136.

[66] Kriegl, A. and Michor, P. W., *The convenient setting of global analysis*, *Mathematical Surveys and Monographs*, Vol. 53 (American Mathematical Society, Providence, RI, 1997).

[67] Lang, B., Private communication.

[68] Lee, J. and Wald, R. M., Local symmetries and constraints, *J. Math. Phys.* **31** (1990) 725–743.

[69] Lee, J. M., *Introduction to smooth manifolds*, *Graduate Texts in Mathematics*, Vol. 218 (Springer-Verlag, New York, 2003).

[70] Lichnerowicz, A., Propagateurs et commutateurs en relativité générale, *Inst. Hautes Études Sci. Publ. Math.* (1961) 56.

[71] Miao, S. P., Tsamis, N. C., and Woodard, R. P., Gauging away Physics, *Class. Quantum Grav.* **28** (2011) 245013.

[72] Moncrief, V., Decompositions of gravitational perturbations, *J. Math. Phys.* **16** (1975) 1556–1560.

[73] Moncrief, V., Spacetime symmetries and linearization stability of the Einstein equations. I, *J. Math. Phys.* **16** (1975) 493–498.

[74] Moncrief, V., Space-time symmetries and linearization stability of the Einstein equations. II, *J. Math. Phys.* **17** (1976) 1893–1902.

[75] Moncrief, V., Invariant states and quantized gravitational perturbations, *Phys. Rev. D* **18** (1978) 983–989.

[76] Moncrief, V., Quantum linearization instabilities, *General Relativity and Gravitation* **10** (1979) 93–97.

[77] O'Neill, B., *Semi-Riemannian geometry*, *Pure and Applied Mathematics*, Vol. 103 (Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1983).

[78] Pfenning, M. J., Quantization of the Maxwell field in curved spacetimes of arbitrary dimension, *Class. Quantum Grav.* **26** (2009) 135017.

[79] Radzikowski, M. J., Micro-local approach to the Hadamard condition in quantum field theory on curved space-time, *Commun. Math. Phys.* **179** (1996) 529–553.

[80] Reed, M. and Simon, B., *Methods of modern mathematical physics. I*, 2nd edn. (Academic Press Inc. [Harcourt Brace Jovanovich Publishers], New York, 1980), Functional analysis.

[81] Sachs, R. K., Gravitational radiation, in *Relativité, Groupes et Topologie (Lectures, Les Houches, 1963 Summer School of Theoret. Phys., Univ. Grenoble)* (Gordon and Breach, New York, 1964), pp. 521–562.

[82] Sahlmann, H. and Verch, R., Microlocal spectrum condition and Hadamard form for vector-valued quantum fields in curved spacetime, *Rev. Math. Phys.* **13** (2001) 1203–1246.

[83] Stewart, J., *Advanced general relativity*, Cambridge Monographs on Mathematical Physics (Cambridge University Press, Cambridge, 1990).

[84] Stewart, J. M. and Walker, M., Perturbations of space-times in general relativity, *Proc. Roy. Soc. (London) Ser. A* **341** (1974) 49–74.

[85] Strocchi, F., Gauge problem in quantum field theory, *Phys. Rev. (2)* **162** (1967) 1429–1438.

[86] Strocchi, F., Gauge problem in quantum field theory. II. Difficulties of combining Einstein equations and Wightman theory, *Phys. Rev.* **166** (1968) 1302–1307.

[87] Strocchi, F., Gauge problem in quantum field theory. III. Quantization of Maxwell equations and weak local commutativity, *Phys. Rev. D (3)* **2** (1970) 2334–2340.

[88] Taylor, M. E., *Partial differential equations I. Basic theory*, *Applied Mathematical Sciences*, Vol. 115, 2nd edn. (Springer, New York, 2011).

[89] Unruh, W. G., Notes on black-hole evaporation, *Phys. Rev. D* **14** (1976) 870–892.

[90] Wald, R. M., The back reaction effect in particle creation in curved spacetime, *Commun. Math. Phys.* **54** (1977) 1–19.

[91] Wald, R. M., *General relativity* (University of Chicago Press, Chicago, IL, 1984).

[92] Wald, R. M., *Quantum field theory in curved spacetime and black hole thermodynamics*, Chicago Lectures in Physics (University of Chicago Press, Chicago, IL, 1994).

[93] Warner, F. W., *Foundations of differentiable manifolds and Lie groups*, *Graduate Texts in Mathematics*, Vol. 94 (Springer-Verlag, New York, 1983).

[94] Weinberg, S., *Cosmology* (Oxford University Press, Oxford, 2008).

[95] Woodhouse, N. M. J., *Geometric quantization*, Oxford Mathematical Monographs, 2nd edn. (The Clarendon Press Oxford University Press, New York, 1992), Oxford Science Publications.