14

Causality and the Cauchy problem in general relativity

The study of the Cauchy problem in general relativity was initiated by the seminal work by Fourès-Bruhat (1952). The extensions and refinements of this work and, in particular, the analysis of the existence of maximal Cauchy developments by Choquet-Bruhat and Geroch (1969) bring to the forefront the delicate interplay between geometry and the theory of partial differential equations arising in Einstein's theory of general relativity.

This chapter provides a discussion of two aspects of the Cauchy problem in general relativity: (i) the connection between the notions of causality originating from the theory of symmetric hyperbolic equations and those derived from the existence of a Lorentzian metric on the underlying spacetime manifold – the so-called *Lorentzian causality*, and (ii) the existence and uniqueness of a so-called *maximal Cauchy development* of an initial value problem for the Einstein field equations. This chapter sets the context for the discussion in Part IV of this book where asymptotically simple spacetimes are constructed by means of suitably posed initial value problems.

14.1 Basic elements of Lorentzian causality

In Section 2.5 some basic notions of Lorentzian geometry have already been introduced. These ideas are now further elaborated to present the notions of ***Lorentzian causal theory***. The summary presented here is adapted from the discussion in Ringström (2009).

In what follows, the discussion is restricted to four-dimensional Lorentzian manifolds $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ which are *orientable and time orientable*. In particular, *time orientability is equivalent to the existence of a smooth timelike vector* $\boldsymbol{t}$; see Section 2.1. The Lorentzian manifold $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ is not assumed to satisfy the Einstein field equations.

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

Chronological future, causal future, and so on

A vector $\boldsymbol{v} \in T(\tilde{\mathcal{M}})$ is said to be ***causal*** if $\boldsymbol{v} \neq 0$ and $\boldsymbol{v}$ is either timelike or null. Consistent with the discussion of Section 2.5, $\boldsymbol{v}$ is said to be ***future pointing*** if $\tilde{\boldsymbol{g}}(\boldsymbol{v}, \boldsymbol{t}) > 0$ and ***past pointing*** if $\tilde{\boldsymbol{g}}(\boldsymbol{v}, \boldsymbol{t}) < 0$. A ***future-pointing causal curve*** on $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ is one for which its tangent vector is everywhere future pointing causal. The notion of ***past-pointing causal curve*** is defined in an analogous manner.

Causal curves can be used to define *order relations* between points of the manifold $\tilde{\mathcal{M}}$. Given p, $q \in \tilde{\mathcal{M}}$, one writes $p \prec\prec q$ if there is a *future-pointing timelike curve* in $\tilde{\mathcal{M}}$ from p to q; $p \prec q$ if there is a *future causal curve* from p to q; and $p \preceq q$ if either $p = q$ or $p \prec q$. Given a subset $\mathcal{U} \subset \tilde{\mathcal{M}}$ one defines the ***chronological future*** and ***chronological past*** of $\mathcal{U}$, respectively, as

$$I^+(\mathcal{U}) \equiv \{p \in \tilde{\mathcal{M}} \,|\, q \prec\prec p \text{ for some } q \in \mathcal{U}\},$$
$$I^-(\mathcal{U}) \equiv \{p \in \tilde{\mathcal{M}} \,|\, p \prec\prec q \text{ for some } q \in \mathcal{U}\}.$$

Moreover, the ***causal future*** and ***causal past*** of $\mathcal{U}$ are defined, respectively, as

$$J^+(\mathcal{U}) \equiv \{p \in \tilde{\mathcal{M}} \,|\, q \preceq p \text{ for some } q \in \mathcal{U}\},$$
$$J^-(\mathcal{U}) \equiv \{p \in \tilde{\mathcal{M}} \,|\, p \preceq q \text{ for some } q \in \mathcal{U}\}.$$

A schematic depiction of the sets $I^{\pm}(\mathcal{U})$ and $J^{\pm}(\mathcal{U})$ is given in Figure 14.1. The sets $I^+(\mathcal{U})$ and $I^-(\mathcal{U})$ can be shown to be open. No general statements of this type can be made about $J^+(\mathcal{U})$ and $J^-(\mathcal{U})$. However, one has that $I^+(\mathcal{U}) \subseteq J^+(\mathcal{U})$ and $I^-(\mathcal{U}) \subseteq J^-(\mathcal{U})$.

Global hyperbolicity

The natural class of spacetimes for which an initial value problem can be formulated is that of globally hyperbolic ones.

A spacetime $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ without closed timelike curves is called ***causal***. A causal spacetime $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ is said to be ***globally hyperbolic*** if for any pair of points p, $q \in \tilde{\mathcal{M}}$ with $p \prec q$ the ***causal diamond*** $J^+(p) \cap J^-(q)$ is compact; see Figure 14.2. The classical definition of global hyperbolicity as given, for example, in Wald

Figure 14.1 Schematic representation of the sets $I^{\pm}(\mathcal{U})$ and $J^{\pm}(\mathcal{U})$ for a subset $\mathcal{U} \subset \tilde{\mathcal{M}}$.

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

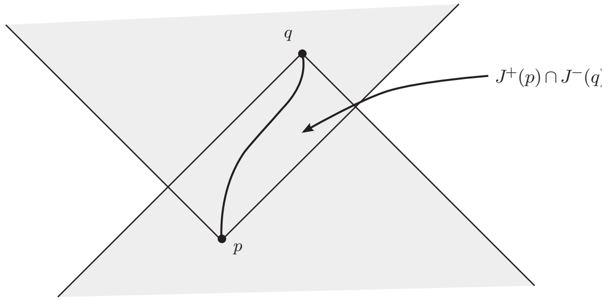

Figure 14.2 The causal diamond $J^+(p) \cap J^-(q)$: the points $p, q \in \tilde{\mathcal{M}}$ satisfy $p \prec q$. In a globally hyperbolic spacetime any such diamond is compact.

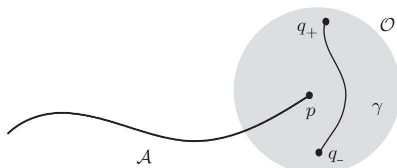

Figure 14.3 The edge of a closed achronal set $\mathcal{A}$: for the point $p \in \mathcal{A}$ there exists an open neighbourhood $\mathcal{O}$ containing p and $q_+ \in I^+(p)$, $q_- \in I^-(p)$ such that q_- and q_+ can be joined by a timelike curve γ not intersecting $\mathcal{A}$.

(1984), makes use of the *stronger* notion of ***strongly causal spacetimes***, that is, the non-existence of "almost closed" causal curves. The classical definition and the one given here have been shown to be equivalent in Bernal and Sánchez (2007).

In physical terms, global hyperbolicity is closely connected to the idea of ***classical determinism***, that is, the prediction or retrodiction of future or past states, respectively, from a set of initial conditions. Pathologies like the existence of closed timelike curves are not present in globally hyperbolic spacetimes.

Cauchy surfaces

A subset $\mathcal{A}$ of a Lorentzian manifold $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ is said to be ***achronal*** if there is no pair of points $p, q \in \mathcal{A}$ that can be connected by a timelike curve. Spacelike hypersurfaces are examples of achronal sets. For $\mathcal{A}$ closed and achronal, one defines its ***edge*** as the set of points $p \in \mathcal{A}$ such that every open neighbourhood $\mathcal{O}$ of p contains points $q_+ \in I^+(p)$, $q_- \in I^-(p)$ and a timelike curve γ from q_- to q_+ which does not interset $\mathcal{A}$; see Figure 14.3.

Given $\mathcal{A} \subset \tilde{\mathcal{M}}$ achronal, the ***future domain of dependence*** of $\mathcal{A}$ is the set $D^+(\mathcal{A})$ of all points $p \in \tilde{\mathcal{M}}$ such that every past inextendible causal curve through p intersects $\mathcal{A}$. The ***past domain of dependence*** of $\mathcal{A}$ is defined in an analogous manner by considering future inextendible causal curves. The ***(full)***

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

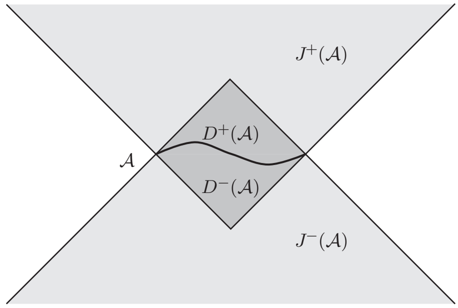

Figure 14.4 Domain of dependence of an achronal set $\mathcal{A}$ and its relation to the causal past and future $J^{\pm}(\mathcal{A})$.

domain of dependence of $\mathcal{A}$ is then defined as

$$D(\mathcal{A}) \equiv D^+(\mathcal{A}) \cup D^-(\mathcal{A}).$$

In some accounts, the sets $D^+(\mathcal{A})$ and $D^-(\mathcal{A})$ are called, respectively, the ***future*** and ***past Cauchy development of*** $\mathcal{A}$. The reason for these alternative names is clarified by the discussion in Section 14.2. It can be verified that $\mathcal{A} \subset D^+(\mathcal{A}) \subset J^+(\mathcal{A})$; see Figure 14.4. From the achronality of $\mathcal{A}$ it follows that $D^+(\mathcal{A}) \cap I^-(\mathcal{A}) = \emptyset$.

Since information travels along causal curves, a point $p \in D^+(\mathcal{S})$ receives information only from $\mathcal{S}$. Accordingly, if *physical laws are causal* – as in the case of general relativity – initial data should determine the physics in $D^+(\mathcal{S})$ – and, in fact, in all of $D(\mathcal{S})$.

A ***Cauchy hypersurface*** in $\tilde{\mathcal{M}}$ is a hypersurface $\tilde{\mathcal{S}}$ such that

$$D(\tilde{\mathcal{S}}) = \tilde{\mathcal{M}}.$$

Cauchy hypersurfaces are characterised by the fact that they are intersected exactly once by every inextendible timelike curve in $\tilde{\mathcal{M}}$; see, for example, Ringström (2009). Cauchy hypersurfaces are continuous three-dimensional submanifolds of the spacetime manifold $\tilde{\mathcal{M}}$; see, for example, Wald (1984). Cauchy hypersurfaces provide an alternative description of globally hyperbolic spacetimes: *any globally hyperbolic spacetime possesses a Cauchy hypersurface.* Global hyperbolicity restricts the topology of a spacetime. More precisely, one has that:

Proposition 14.1 (***topology of globally hyperbolic spacetimes***) *Let* $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ *denote a connected, time-oriented globally hyperbolic Lorentzian manifold and let* $\tilde{\mathcal{S}}$ *be a Cauchy hypersurface thereof. Then*

$$\tilde{\mathcal{M}} \approx \mathbb{R} \times \tilde{\mathcal{S}}.$$

In other words, $\tilde{\mathcal{M}}$ *can be foliated by Cauchy hypersurfaces. Moreover, if* $\tilde{\mathcal{S}}'$ *is another Cauchy hypersurface, then* $\tilde{\mathcal{S}} \approx \tilde{\mathcal{S}}'$.

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

The above result is complemented by the following:

Proposition 14.2 (***existence of a global time function***) *Let* $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ *be an oriented, time-oriented, connected and globally hyperbolic spacetime and let* $\tilde{\mathcal{S}}$ *be a Cauchy hypersurface thereof. Then there is a smooth function* t *on* $\tilde{\mathcal{M}}$ *such that* $\mathbf{d}t$ *is timelike and future directed everywhere and satisfies the property that* $t^{-1}(T_\bullet)$ *is a Cauchy hypersurface for every* $T_\bullet \in \mathbb{R}$*. Furthermore,* $t^{-1}(0) = \tilde{\mathcal{S}}$ *and for every inextendible causal curve* $\gamma : (\mathrm{s}_-, \mathrm{s}_+) \to \tilde{\mathcal{M}}$ *one has* $t(\gamma(\mathrm{s})) \to \pm\infty$ *as* $\mathrm{s} \to \mathrm{s}_\pm$*.*

For a proof of these results, see, for example, Ringström (2009), proposition 11.3 and theorem 11.27. Finally, one has the following:

Proposition 14.3 (***asymptotic simplicity and global hyperbolicity***) *An asymptotically simple and empty spacetime* $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ *is globally hyperbolic.*

The reader interested in a proof is referred to Hawking and Ellis (1973), proposition 6.9.2.

Cauchy horizons

In what follows let $\overline{D^+(\mathcal{A})}$ denote the *closure of the future domain of dependence* of an achronal set $\mathcal{A} \subset \tilde{\mathcal{M}}$. This set is characterised by the fact that for $p \in \overline{D^+(\mathcal{A})}$ every past inextendible timelike curve from p intersects $\mathcal{A}$; see proposition 8.3.2 in Wald (1984). The achronal set $\mathcal{A}$ is not necessarily a Cauchy hypersurface. To characterise how much $\mathcal{A}$ deviates from being a Cauchy hypersurface, it is convenient to introduce the set

$$H^+(\mathcal{A}) \equiv \overline{D^+(\mathcal{A})} \setminus I^-\big(D^+(\mathcal{A})\big),$$

the so-called ***future Cauchy horizon*** of $\mathcal{A}$. The ***past Cauchy horizon*** is defined in an analogous manner as $H^-(\mathcal{A}) \equiv \overline{D^-(\mathcal{A})} \setminus I^+\big(D^-(\mathcal{A})\big)$. It can be shown that $H^+(\mathcal{A})$ is achronal. Moreover, one has that $\mathcal{A} \subset D^+(\mathcal{A})$ and $\partial D^+(\mathcal{A}) = H^+(\mathcal{A}) \cup \mathcal{A}$; see Figure 14.5. Similar properties hold for $D^-(\mathcal{A})$.

The ***(full) Cauchy horizon*** is then defined as $H(\mathcal{A}) \equiv H^+(\mathcal{A}) \cup H^-(\mathcal{A})$. It can be proved that $H(\mathcal{A}) = \partial\big(D(\mathcal{A})\big)$ and that the achronal set $\mathcal{A}$ is a Cauchy surface for $(\tilde{\mathcal{M}}, \boldsymbol{g})$ if and only if $H(\mathcal{A}) = \emptyset$; see proposition 8.3.6 and its corollary in Wald (1984).

The following property of Cauchy horizons will be used at various points in this book (cf. theorem 8.3.5 in Wald (1984)):

Proposition 14.4 (***structure of Cauchy horizons***) *Every point* $p \in H^+(\mathcal{A})$ *lies on a null geodesic contained entirely in* $H^+(\mathcal{A})$ *which is either inextendible or has a past endpoint on the edge of* $\mathcal{A}$*.*

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

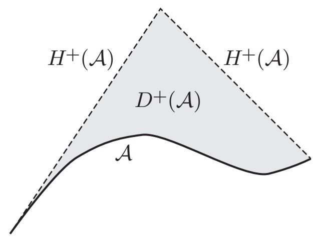

Figure 14.5 The Cauchy horizon of an achronal set $\mathcal{A}$. Observe that $\partial D^+(\mathcal{A}) = H^+(\mathcal{A}) \cup \mathcal{A}$.

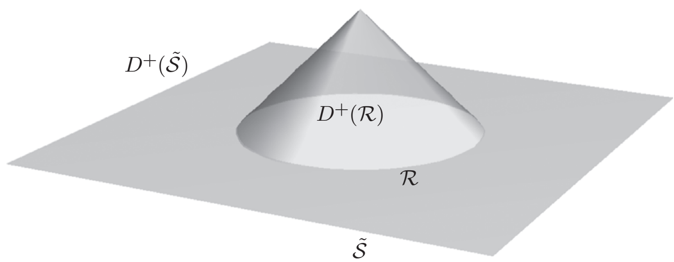

Figure 14.6 Schematic representation of the causal domains of Theorem 14.1. The hypersurface $\tilde{\mathcal{S}}$ is a Cauchy hypersurface and $\mathcal{R} \subseteq \tilde{\mathcal{S}}$ is a region within such that $\mathbf{u} = 0$.

14.2 PDE causality versus Lorentzian causality

Two different notions of causality have been discussed so far in this book: *partial differential equation (PDE) causality* based on the uniqueness of solutions to symmetric hyperbolic systems – Theorem 12.1 – and *Lorentzian causality*, discussed in the first sections of this chapter. These notions of causality are conceptually different from each other. However, they are linked by the following result (see also Figure 14.6):

Theorem 14.1 (***the relation between PDE and Lorentzian causalities***) *Let $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ be a connected, oriented, time-oriented, globally hyperbolic spacetime and let $\tilde{\mathcal{S}}$ be a smooth spacelike Cauchy hypersurface. Let $\mathcal{R} \subseteq \tilde{\mathcal{S}}$ and let $\mathcal{U}$ be an open set containing $\overline{D^+(\mathcal{R})}$. Assume that $\mathbf{u} : \mathcal{U} \to \mathbb{C}^N$ solves the symmetric hyperbolic system*

$$\mathbf{A}^\mu(x, \mathbf{u})\partial_\mu \mathbf{u} + \mathbf{B}(x, \mathbf{u}) = 0.$$

Moreover, assume that the above equation has a characteristic polynomial which contains the factor $(\tilde{g}^{\mu\nu}\xi_\mu\xi_\nu)$ where $\tilde{g}^{\mu\nu}$ denotes the contravariant components of the metric $\tilde{\boldsymbol{g}}$. If $\mathbf{u}$ vanishes on $\mathcal{R}$, then $\mathbf{u}$ vanishes on $D^+(\mathcal{R})$. There is a similar statement for $D^-(\mathcal{R})$.

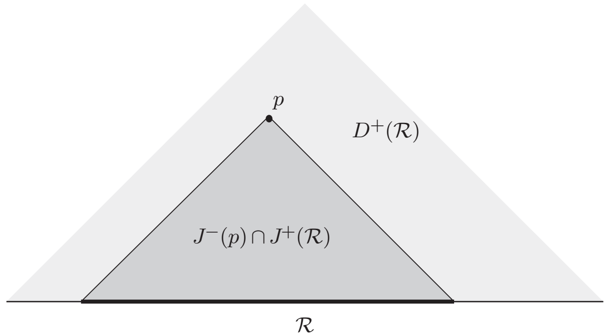

Figure 14.7 Schematic representation of the causal domains in the proof of Theorem 14.1. On $D^+(\mathcal{R})$ one considers for arbitrary $p \in D^+(\mathcal{R})$ the associated domain $\mathcal{D}(p) \equiv J^-(p) \cap J^+(\mathcal{R})$.

The interested reader is referred to chapter 12 of Ringström (2009) for a detailed account of the proof of an analogous result for quasilinear wave equations. It is, nevertheless, useful to discuss some of the ideas behind the proof.

Theorem 12.1 and its Corollary 12.1 – ensuring the uniqueness of solutions to symmetric hyperbolic systems – can be applied only on *lens-shaped domains*. The main idea behind Theorem 14.1 is then to construct a cover of $D^+(\mathcal{R})$ consisting of lens-shaped domains. The metric $\tilde{\boldsymbol{g}}$ provides a natural way of constructing the required cover. Accordingly, the Lorentzian metric allows the introduction of the notions of Lorentzian causality discussed in the first sections of this chapter.

One begins by considering points $p \in D^+(\mathcal{R})$ which are *suitably close* to $\mathcal{R}$ and aims to conclude that $\mathbf{u} = 0$ on $\mathcal{D}(p) \equiv J^-(p) \cap J^+(\mathcal{R})$;—see Figure 14.7. By means of the exponential map $\exp_p : T|_p(\tilde{\mathcal{M}}) \supset \bar{\mathcal{V}} \to \mathcal{V} \subset \tilde{\mathcal{M}}$ – see Section 11.6.2 – the metric $\tilde{\boldsymbol{g}}$ allows the introduction of normal coordinates in some neighbourhood of p – these coordinates can be seen as providing a diffeomorphism between a neighbourhood $\bar{\mathcal{V}}$ of the origin in $T|_p(\tilde{\mathcal{M}})$ to a neighbourhood $\mathcal{V}$ of p. By considering p sufficiently close to $\mathcal{R}$ one can ensure that $\mathcal{D}(p)$ is compact and completely contained in $\mathcal{V}$. On $T|_p(\tilde{\mathcal{M}})$ one can define a function $\bar{f} : T|_p(\tilde{\mathcal{M}}) \to \mathbb{R}$ via $\bar{f}(\boldsymbol{v}) = \tilde{\boldsymbol{g}}(\boldsymbol{v}, \boldsymbol{v})$. Hence, for the present purposes, the neighbourhood $\mathcal{V}$ can be regarded as a subset of the Minkowski spacetime coordinatised by standard Cartesian coordinates. One also defines $f : \mathcal{V} \to \mathbb{R}$ such that $f \equiv \bar{f} \circ \exp_p^{-1}$. Now, given a constant $c > 0$, the condition $\tilde{\boldsymbol{g}}(\boldsymbol{v}, \boldsymbol{v}) = c$ defines (spacelike) hyperboloids on $\mathcal{V}$. More precisely, for given c, the locus of points in $\mathcal{V}$ corresponding to the hyperboloid is given by $f^{-1}(c) \equiv \{q \in \mathcal{V} \,|\, \tilde{\boldsymbol{g}}(\exp_p^{-1}(q), \exp_p^{-1}(q)) = c\}$ – observe that both $\bar{f}$ and f are not injective so that f^{-1} is a set consisting of more than a single point. Now, $f^{-1}(c)$ has two components: one associated to future-directed vectors and the other associated to past-directed vectors. For $c > 0$, let $\mathcal{Q}_c(p)$ denote the component of $f^{-1}(c)$ associated to past-directed vectors and let $\mathcal{Q}_0(p)$ denote the past null cone through p. One can use the hyperboloids

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

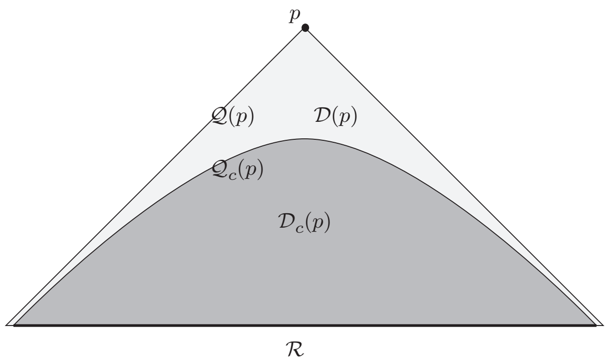

Figure 14.8 Schematic representation of the causal domains in the proof of Theorem 14.1. Given $p \in D^+(\mathcal{R})$, the sets $\mathcal{Q}(p)$ and $\mathcal{D}(p)$ describe, respectively, the past light cone through p and the region between the light cone and $\mathcal{R}$. For $c > 0$, the set $\mathcal{Q}_c(p)$ describes a hyperboloid inside the past light cone of p, while $\mathcal{D}_c(p)$ is the region between the hyperboloid and $\mathcal{R}$. The set $\mathcal{D}_c(p)$ is a lens-shaped domain. See the main text for further details.

$\mathcal{Q}_c(p)$ to foliate the interior of the past light cone passing through p. One defines $\mathcal{D}_c(p) \equiv J^-(\mathcal{Q}_c(p)) \cap J^+(\mathcal{R})$. The domain $\mathcal{D}_c(p)$ describes the region between the hyperboloid $Q_c(p)$ and $\mathcal{R}$, while $\mathcal{D}(p)$ describes the region between the past light cone and $\mathcal{R}$; see Figure 14.8. Now, $\mathcal{D}(p)$ can be shown to be compact. Moreover, it can be seen that $\mathcal{Q}_c(p) \subset I^-(p)$ for $c > 0$ so that $J^-(\mathcal{Q}_c(p)) \subset J^-(p)$ and, in addition, that $\mathcal{D}_c(p) \subset \mathcal{D}(p)$. A further argument allows one to verify that $\mathcal{Q}_c(p)$ for $c > 0$ is a lens-shaped domain on which, modulo some technical details, Corollary 12.1 can be applied. Thus, if $\mathbf{u} = 0$ on $\mathcal{R}$, one concludes that $\mathbf{u} = 0$ on $\mathcal{D}_c(p)$.

To show that $\mathbf{u} = 0$ on $\mathcal{D}(p)$ one now considers a sequence $\{c_l\}$, $l \in \mathbb{N}$, of positive numbers converging to zero. It can then be shown that

$$\operatorname{int} \mathcal{D}(p) \subset \cup_l \mathcal{D}_{c_l}(p) \subset \mathcal{D}(p)$$

– intuitively, by choosing smaller and smaller c_l's one obtains hyperboloids which are, successively, "closer" to the light cone $\mathcal{Q}_p$ thus "filling" $\mathcal{D}(p)$. From this observation and given that $\mathbf{u} = 0$ on each of the $\mathcal{D}_{c_l}(p)$ one can conclude that, indeed, $\mathbf{u} = 0$ on $\mathcal{D}(p)$.

Now, an adaptation of Proposition 14.2 ensures the existence of a time function t on $D^+(\mathcal{R})$. Given $c > 0$, $t^{-1}([0, c])$ denotes a ***slab*** in $D^+(\tilde{\mathcal{S}}) \supset D^+(\mathcal{R})$. Considering points suitably close to $\tilde{\mathcal{S}}$, it is possible to construct a slab $K_\varepsilon \equiv t^{-1}\big([0, \varepsilon]\big) \cap D^+(\mathcal{R})$ in $D^+(\mathcal{R})$ for some $\varepsilon > 0$ – this slab can be thought of as the union of domains of the type $\mathcal{D}(p)$ on each of which one already knows that $\mathbf{u} = 0$. The ***top*** of the slab, $t^{-1}(\varepsilon) \cap D^+(\mathcal{R})$ – on which $\mathbf{u} = 0$ – can now be used as a new initial surface from which one constructs a further slab. The rest of the proof consists of showing that $D^+(\mathcal{R})$ can be fully covered by slabs of the type described above so that $\mathbf{u} = 0$ everywhere on $D^+(\mathcal{R})$.

Remark. An important observation is that

$$D^+(\mathcal{R}) \cap I^+(\tilde{\mathcal{S}} \setminus \mathcal{R}) = \emptyset.$$

Accordingly, the value of $\mathbf{u}$ on $D^+(\mathcal{R})$ is determined only by the data on $\mathcal{R}$ – that is, whatever data is prescribed on $\tilde{\mathcal{S}} \setminus \mathcal{R}$, it has no influence on the behaviour of $\mathbf{u}$ on $D^+(\mathcal{R})$. The proof of this statement follows by contradiction: let $q \in D^+(\mathcal{R}) \cap I^+(\tilde{\mathcal{S}} \setminus \mathcal{R})$; on the one hand we have that $q \in I^+(\tilde{\mathcal{S}} \setminus \mathcal{R})$ so that there exists a future timelike curve γ from $p \in \tilde{\mathcal{S}} \setminus \mathcal{R}$ to q. On the other hand $q \in D^+(\mathcal{R})$ so that every past inextendible causal curve through q intersects $\mathcal{R}$. As a consequence one has that $p \in \mathcal{R}$. This is a contradiction since $p \in \tilde{\mathcal{S}} \setminus \mathcal{R}$.

14.3 Cauchy developments and maximal Cauchy developments

For ease of presentation, the subsequent discussion is restricted to the case of *standard Cauchy initial value problems* where initial data is prescribed on a Cauchy hypersurface $\tilde{\mathcal{S}}$. For a detailed account of the Cauchy problem in general relativity the reader is referred to the monograph by Ringström (2009).

As discussed in Section 11.3, the (say, vacuum) Einstein field equations imply on $\tilde{\mathcal{S}}$ a set of constraint equations: the so-called Hamiltonian and momentum constraints for a Riemannian metric $\tilde{\boldsymbol{h}}$ and a symmetric trace-free tensor $\tilde{\boldsymbol{K}}$. Assume one is given a solution $(\tilde{\boldsymbol{h}}, \tilde{\boldsymbol{K}})$ to the Hamiltonian and momentum constraint Equations (11.13a) and (11.13b) on $\tilde{\mathcal{S}}$. To discuss the relation between a solution to the Einstein constraint equations and a solution to the Einstein field equations one needs to introduce the notion of a ***Cauchy development***:

Definition 14.1 (***Cauchy development***) *A* ***Cauchy development*** *of the initial data set* $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{K}})$ *consists of a solution* $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ *of the vacuum Einstein field equations, an embedding* φ *of* $\tilde{\mathcal{S}}$ *into* $\tilde{\mathcal{M}}$ *and a choice of a unit normal vector such that* $\varphi(\tilde{\mathcal{S}})$ *is a Cauchy hypersurface and the pull-backs by* φ *of the induced metric and the second fundamental form for the prescribed unit normal coincide with* $\tilde{\boldsymbol{h}}$ *and* $\tilde{\boldsymbol{K}}$.

The seminal work in Fourès-Bruhat (1952) has shown that, given a solution to the constraint equations on $\tilde{\mathcal{S}}$, it is always possible to obtain a Cauchy development. More precisely, one has the following (see also Figure 14.9):

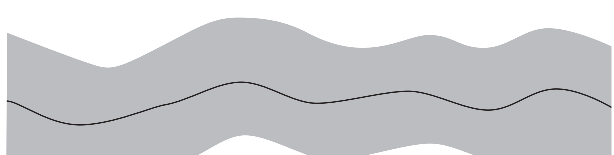

Figure 14.9 Schematic representation of a Cauchy development (in gray) of some initial data set $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{K}})$ for the Einstein field equations.

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

Theorem 14.2 (***existence of a development of an initial data set***) *Given an initial data set $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{K}})$ for the Einstein field equations it is always possible to find a corresponding development.*

The above result is a cornerstone of the mathematical study of the Einstein field equations as it shows that it is meaningful to formulate a Cauchy problem for the Einstein field equations. The original proof of the theorem used the hyperbolic reduction of the Einstein field equations based on *wave coordinates*; see the Appendix to Chapter 13. The hyperbolic reductions for the conformal Einstein field equations discussed in Chapter 13 readily lead to an alternative proof which is briefly sketched for completeness.

Proof The proof of the theorem amounts to a local existence result for the Cauchy problem for the Einstein field equations. For convenience, consider the spinorial version of the standard conformal Einstein field equations; see Section 8.3.2. *Setting* $\Xi = 1$ *and* $\Phi_{\boldsymbol{AA'BB'}} = 0$ *one obtains a spinorial representation of the vacuum Einstein field equations.* For these equations, the hyperbolic reduction procedure summarised in Proposition 13.1 shows that given a choice of coordinate and frame gauge source functions $F^{\boldsymbol{a}}(x)$ and $F_{\boldsymbol{AB}}(x)$, the Einstein field equations imply a symmetric hyperbolic system for the frame coefficients, connection coefficients and the Weyl spinor. Smooth initial data $\mathbf{u}_\star$ for these evolution equations can be obtained from the pair $(\tilde{\boldsymbol{h}}, \tilde{\boldsymbol{K}})$ using the procedure leading to Lemma 11.1. The basic existence and uniqueness result for symmetric hyperbolic systems given in Theorem 12.2 ensures the existence of a solution $\mathbf{u}$ to the *evolution equations* in a slab of the form $\tilde{\mathcal{M}}_T \equiv (-T, T) \times \tilde{\mathcal{S}}$ for some $T > 0$. In what follows, for conceptual clarity, the Riemannian 3-manifold $\tilde{\mathcal{S}}$ regarded as a submanifold of $\tilde{\mathcal{M}}_T$ will be denoted as $\tilde{\mathcal{S}}_\star$; that is, one has $\tilde{\mathcal{S}}_\star = \varphi(\tilde{\mathcal{S}})$. On $\tilde{\mathcal{S}}_\star$ the solution $\mathbf{u}$ coincides with the initial data $\mathbf{u}_\star$. In view of the homogeneous structure of the subsidiary evolution equations as described in Proposition 13.2, the solution $\mathbf{u}$ implies a solution to the *conformal Einstein field equations* with $\Xi = 1$ and $\Phi_{\boldsymbol{AA'BB'}} = 0$. From the components of $\mathbf{u}$ one can construct a Lorentzian metric $\tilde{\boldsymbol{g}}$ which will be a solution to the Einstein field equations on $\tilde{\mathcal{M}}_T$; compare Proposition 8.1. To conclude, it is observed that the hyperbolic procedure leading to the evolution equations is based on an adapted frame tetrad $\{\boldsymbol{e_a}\}$ such that $\boldsymbol{e_0}$ on $\tilde{\mathcal{S}}_\star$ gives the unit normal of the initial hypersurface; see Section 11.4. From this observation it follows that the pull-back of $\tilde{\boldsymbol{g}}$ to $\tilde{\mathcal{S}}_\star$ coincides with the Riemannian metric $\tilde{\boldsymbol{h}}$. Moreover, by construction, the extrinsic curvature of $\tilde{\mathcal{S}}_\star$ coincides with the tensor $\tilde{\boldsymbol{K}}$. Accordingly $(\tilde{\mathcal{M}}_T, \tilde{\boldsymbol{g}})$ provides the required Cauchy development. □

An important aspect of the notion of a Cauchy development is its non-uniqueness. A different choice of gauge source functions will, in general, lead to *a different Cauchy development for the same initial data.* Observe, however, that as one is constructing solutions to tensorial equations in the regions where two different Cauchy developments $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ and $(\tilde{\mathcal{M}}', \tilde{\boldsymbol{g}}')$ overlap $\tilde{\mathcal{M}} \cap \tilde{\mathcal{M}}'$, these

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

must be related to each other by a diffeomorphism, that is, a coordinate transformation. This non-uniqueness of Cauchy developments of a given initial data creates a tension with the notion of ***geometric uniqueness***, that is, the expectation that a given initial data set should give rise to a unique solution to the Einstein field equations. To deal with this issue one introduces the notion of a ***maximal Cauchy development***:

Definition 14.2 (***maximal Cauchy development***) *Let $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{K}})$ be an initial data set for the vacuum Einstein equations. A Cauchy development $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ with embedding $\varphi : \tilde{\mathcal{S}} \to \tilde{\mathcal{M}}$ of this data is said to be* ***maximal*** *if for any other Cauchy development $(\tilde{\mathcal{M}}', \tilde{\boldsymbol{g}}')$ with embedding $\varphi' : \tilde{\mathcal{S}} \to \tilde{\mathcal{M}}'$, there is a smooth map $\psi : \tilde{\mathcal{M}}' \to \tilde{\mathcal{M}}$ which is a diffeomorphism onto its image such that $\varphi = \psi \circ \varphi'$ and $\psi^* \tilde{\boldsymbol{g}} = \tilde{\boldsymbol{g}}'$.*

The maximal Cauchy development describes the biggest spacetime that can be recovered from a given initial data set for the Einstein field equations. Any other Cauchy development must be contained in it. For this notion to be of utility it should satisfy some existence and uniqueness properties. Indeed, one has the following result, first proven in Choquet-Bruhat and Geroch (1969):

Theorem 14.3 (***existence of a maximal Cauchy development***) *Given some initial data $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{K}})$ for the Einstein field equations, there exists a maximal Cauchy development which is unique up to isometries.*

The original proof of this theorem famously relies on *Zorn's lemma*. Alternative proofs not depending on this axiom of set theory have been given more recently in Sbierski (2013) and Wong (2013).

Remark. The maximal Cauchy development of an initial data set is, in general, different from the so-called ***maximal analytic extension*** of the solution to the initial value problem, that is, the biggest spacetime that can be associated to a given metric allowing for analytic changes of coordinates. As an example, compare the Penrose diagram of the maximal analytical extension of the Reissner-Nordström spacetime given in Figure 6.14 and the Penrose diagram of its maximal Cauchy development in Figure 14.10.

The characterisation and construction of the maximal Cauchy development of an arbitrary initial data set $(\tilde{\mathcal{S}}, \tilde{\boldsymbol{h}}, \tilde{\boldsymbol{K}})$ is a challenging endeavour. It requires controlling the evolution dictated by the Einstein field equations under very general conditions. Generically, one expects the following to be true:

Conjecture 14.1 (***strong cosmic censorship***) *The maximal Cauchy development of generic initial data for the vacuum Einstein field equations is inextendible.*

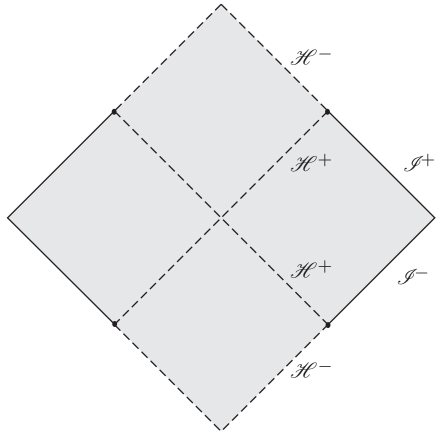

Figure 14.10 Maximal Cauchy development of the Reissner-Nordström spacetime. In this case, the spacetime extends only up to the Cauchy horizon $\mathscr{H}^-$. Notice that the timelike singularities of the spacetime do not appear in the diagram.

A concise discussion of the above conjecture and its various caveats can be found in Chruściel (1991), Rendall (2008) and Ringström (2009).

14.4 Stability of solutions

A problem simpler than cosmic censorship is the construction of the development of initial data sets which are, in some sense, close to initial data for some exact solution (the ***background solution***) whose global structure is well known. Such initial data are called a ***perturbation of the initial data for the exact solution***. Under suitable circumstances one expects the maximal Cauchy development of the perturbed initial data to have a global structure similar to that of the maximal Cauchy development of the exact solution. The resulting spacetime is called a ***perturbation of the exact solution***. This notion of perturbations is a non-linear one: the perturbed solutions are required to satisfy the Einstein field equations without any approximation – as opposed, say, to linearised perturbations where one considers solutions to evolution equations which are linearised with respect to some background exact solution. The underlying strategy behind the analysis of non-linear perturbations is to use the knowledge of the global properties of a solution to the equations of general relativity to infer the existence of other solutions with analogous properties. This point of view leads to the notion of ***stability***.

When discussing the stability of solutions to the Einstein field equations one typically distinguishes between the notions of orbital and asymptotical stability. A solution is said to be ***orbitally stable*** if the global geometry of the perturbed evolution exhibits the same features as the original (background) solution – for example, the existence of a complete null infinity. The stronger notion

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

of ***asymptotic stability*** requires, in addition, that the perturbed solution converges to the background solution for late times. The stability results to be discussed in the remainder of this book will be of the orbital type.

Remark. Although the notion of stability has a strong physical motivation – see, for example, the discussion in Section 12.3.2 – the precise formulation of *closeness* to a certain exact solution is dictated by the details of the PDE theory used to analyse the evolution equations – for example, Sobolev norms – and it may be difficult to provide it with a direct physical interpretation. In particular, statements about closeness may not be *gauge independent*.

14.5 Causality and conformal geometry

Let $(\mathcal{M}, \boldsymbol{g})$ denote a conformal extension of a *physical spacetime* $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ with $\boldsymbol{g} = \Xi^2 \tilde{\boldsymbol{g}}$. As a Lorentzian manifold in its own right, the unphysical spacetime $(\mathcal{M}, \boldsymbol{g})$ gives rise to its own causal notions. The causal notions in $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ and $(\mathcal{M}, \boldsymbol{g})$ are, however, related to each other – it is not hard to see that *the causal notions introduced in Section 14.1 are conformally invariant*. More precisely, if $p, q \in \tilde{\mathcal{M}}$ are connected to each other via some particular causal relation with respect to the metric $\tilde{\boldsymbol{g}}$ (e.g. $p \prec q$, $p \preceq q$ or $p \prec\!\prec q$), then they are also causally related in the same way with respect to the metric $\boldsymbol{g}$. Special care is needed, however, when discussing points which lie on the conformal boundary of the conformal extension $(\mathcal{M}, \boldsymbol{g})$ as these points do not exist in the physical spacetime manifold $\tilde{\mathcal{M}}$. Moreover, any compact set in the unphysical manifold $(\mathcal{M}, \boldsymbol{g})$ which intersects the conformal boundary will be, from the perspective of the physical manifold $\tilde{\mathcal{M}}$, non-compact. This observation is of importance for the notion of global hyperbolicity as it is formulated in terms of compactness of domains in the physical spacetime $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$.

A further cautionary note concerns Cauchy horizons in the unphysical spacetime $(\mathcal{M}, \boldsymbol{g})$ which may not correspond to domains in $\tilde{\mathcal{M}}$. The prototypical case of this situation arises in the discussion of Minkowski-like spacetimes. From the point of view of $(\mathcal{M}, \boldsymbol{g})$, the conformal boundary of these spacetimes corresponds to the Cauchy horizon of hyperboloidal hypersurfaces – which from the conformal point of view are compact domains. From the physical perspective of $(\tilde{\mathcal{M}}, \tilde{\boldsymbol{g}})$ the hyperboloids are non-compact and there are no Cauchy horizons. The correspondence between the conformal boundary and Cauchy horizons for Minkowski-like spacetimes is analysed in some detail in Section 16.3.

Penrose diagrams provide a convenient way of visualising the causal properties of spacetimes. For example, an inspection of the Penrose diagram of the anti-de Sitter spacetime, Figure 14.11, readily shows that the spacetime cannot be globally hyperbolic: causal diamonds intersecting the conformal boundary of the conformal representation correspond to non-compact causal diamonds in the physical spacetime. Alternatively, by looking at the Penrose diagram it is easy to draw a timelike curve which does not intersect a putative Cauchy

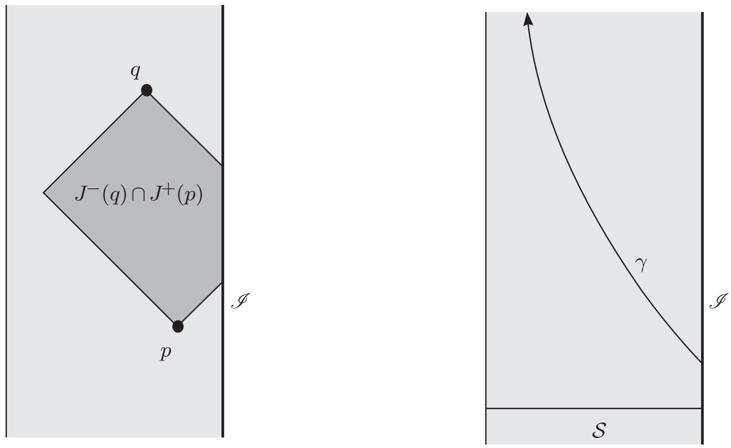

Figure 14.11 Non-global hyperbolicity of the anti-de Sitter spacetime. To the left: causal diamonds intersecting the conformal boundary are non-compact in the physical picture. To the right: given a putative Cauchy hypersurface $\mathcal{S}$, it is always possible to find a timelike curve γ not intersecting $\mathcal{S}$.

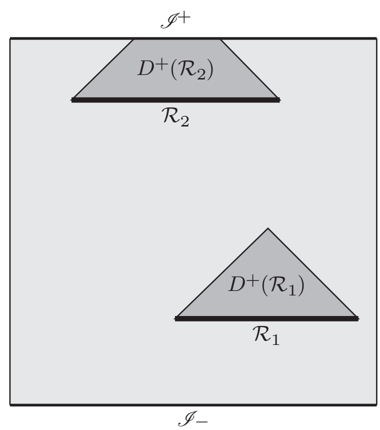

Figure 14.12 Examples of some domains of dependence in the de Sitter spacetime.

hypersurface $\mathcal{S}$ – it is only necessary that in the conformal picture the curve starts at some point of the conformal boundary which lies in the future of $\mathcal{S}$. A second example of the insights provided by the inspection of the Penrose diagrams involves the de Sitter spacetime; see Figure 14.12. A peculiarity of this spacetime is that there exist regions in the spacetime whose domain of dependence is non-compact – to see this, it is only necessary to consider domains which are, from the conformal point of view, sufficiently close to the conformal boundary $\mathscr{I}$.

https://doi.org/10.1017/9781009291347.018 Published online by Cambridge University Press

14.6 Further reading

Detailed accounts of the theory of Lorentzian causality can be found in Hawking and Ellis (1973), chapter 6; O'Neill (1983), chapter 5; or Wald (1984), chapter 8.

An extensive discussion of the Cauchy problem in general relativity can be found in Ringström (2009). A concise presentation is given in Rendall (2008). A related discussion is contained in Friedrich and Rendall (2000). A discussion of various aspects of strong cosmic censorship as well as a number of ancillary results concerning the Cauchy problem can be found in the monograph by Chruściel (1991).