THE CAUCHY PROBLEM FOR EINSTEIN'S VACUUM EQUATIONS

ARTHUR LEI QIU

1. Introduction

Non-quantum physicists are wont to seek out theories which are *deterministic*: when given a physical system's state at some fixed time, it should be possible to fully predict how the system will evolve in time, at least in theory. In particular, solutions to the equations of motion describing a system's evolution should be unique in some sense upon fixing suitable initial data, such as the initial positions and velocities of the system's constituent particles at some fixed time.

When trying to adapt this idea to the framework of general relativity, however, one is immediately faced with several (surmountable) challenges. To start with, the ambient space modelling the universe is not given in advance—one must simultaneously find the spacetime manifold in which dynamics takes place, as well as the Lorentzian metric on this manifold which determines the dynamics and satisfies Einstein's equations. As such, the notion of "initial data" is not well-defined a priori. Moreover, it is not immediately clear how to quantify the notion of "uniqueness of solutions".

In this essay, we will describe how to make sense of the Cauchy problem for the Einstein vacuum equations. In Section 2, we explain what it means to specify initial data for Einstein's equations, what that initial data should consist of, and what it means to solve Einstein's equations with prescribed initial data. We will then state a result (Theorem 2.5) which guarantees local existence and uniqueness of solutions to Einstein's equations for any admissible initial data set, thereby reassuring the physicist that general relativity is at least locally deterministic. This result, however, does not exclude the possibility that a relativistic system could evolve in two different ways beyond some point in spacetime. In Section 3, we upgrade this result by showing that, in fact, every initial data set admits a *unique maximal* solution. The proof we sketch, however, will have a potentially undesirable feature, in that it is not a consequence of Zermelo–Fraenkel set theory alone; in fact, it will rely on Zorn's lemma twice. Section 4 is dedicated to a much more recent proof of the same fact, but which does not require Zorn's lemma.

2. Preliminary definitions

For simplicity, we assume that all manifolds are smooth (C^∞), Hausdorff, and without boundary, and that all maps between manifolds are smooth unless specified otherwise. Henceforth, by a *spacetime* we mean a 4-dimensional Lorentzian manifold (M, g) with signature $(-,+,+,+)$; if g satisfies the Einstein vacuum equations (i.e., has vanishing Ricci curvature), then we call (M, g) a *vacuum spacetime*.

Definition 2.1. *A* ***Cauchy hypersurface*** *in a spacetime (M, g) is a 3-dimensional submanifold $\Sigma \subset M$ such that every inextensible timelike curve in M intersects Σ exactly once. We say that (M, g) is* ***globally hyperbolic*** *if it admits a spacelike Cauchy hypersurface.*

Very loosely speaking, a Cauchy hypersurface can be thought of as an "instantaneous snapshot of the entire universe at a moment of time"—indeed, the prototypical examples of Cauchy hypersurfaces are the constant-t hyperplanes in Minkowski space $(\mathbb{R}^{1+3}, -dt^2 + dx^2 + dy^2 + dz^2)$.

Suppose (M, g) is a globally hyperbolic vacuum spacetime with spacelike Cauchy hypersurface $\Sigma \subset M$. Let K denote the second fundamental form of Σ taken with respect to the future-oriented unit normal. Then by taking suitable traces of the Gauss–Codazzi equations, one finds that

$$\begin{cases} R_{\mathrm{scal}} - |K|^2 + (\operatorname{tr} K)^2 = 0, \\ \qquad \operatorname{div}(K) - d(\operatorname{tr} K) = 0, \end{cases} \tag{1}$$

where the scalar curvature R_{scal}, divergence, etc. are defined using the induced Riemannian metric on Σ.

On the other hand, if we *start* with any Riemannian 3-manifold (Σ, g_0) (that is, not necessarily embedded as a submanifold of a larger manifold) and a symmetric $(0, 2)$-tensor K on Σ, then we can use the metric g_0 to make sense of Equations (1). In fact, these equations provide a sufficient condition for Σ to be realizable as a spacelike Cauchy hypersurface with induced metric g_0 and second fundamental form K in some vacuum spacetime. For this reason, Equations (1) are known as the *constraint equations*, and we refer to a triple (Σ, g_0, K) satisfying the constraint equations as an *initial data set*.

Definition 2.2. *Let (Σ, g_0, K) be an initial data set. A **globally hyperbolic development** (or **GHD**) of (Σ, g_0, K) is a triple (M, g, ι), where (M, g) is a time-oriented, globally hyperbolic vacuum spacetime and $\iota\colon (\Sigma, g_0) \to (M, g)$ is an isometric embedding*[1] *such that*

- *$\iota(\Sigma)$ is a spacelike Cauchy hypersurface, and*
- *the pullback along ι of the second fundamental form of $\iota(\Sigma)$ with respect to the future-oriented unit normal is given by K.*

We will sometimes suppress the embedding map ι and refer to a GHD by the shorthand notation (M, g), or even just M.

Remark 2.3. Is it possible for two universes to join together? That is, can an initial data set (Σ, g_0, K) where Σ is *disconnected* admit a GHD (M, g) where M is *connected*?

In fact, the answer is no. A 1970 theorem of Geroch [4] shows that, if (M, g) is a globally hyperbolic spacetime and $\Sigma \subset M$ is a spacelike Cauchy hypersurface, then M is homeomorphic to $\mathbb{R} \times \Sigma$. Thus it is not possible for a connected spacetime to admit a disconnected Cauchy hypersurface. Geroch's result was improved in 2003 by Bernal and Sánchez [1], who proved that M is in fact diffeomorphic to $\mathbb{R} \times \Sigma$.

The Cauchy hypersurface $\{t = 0\}$ in Minkowski space $(\mathbb{R}^{1+3}, -dt^2 + dx^2 + dy^2 + dz^2)$ has $\{|t| < T\}$ as a GHD, for every $T > 0$. Of course, there is a GHD of $\{t = 0\}$ which contains all of these GHDs: namely, all of Minkowski space. We now give a name to more general situations in which one GHD can be thought of as "sitting inside" another GHD, without a priori assuming that the "smaller" one is actually a subset of the "bigger" one.

Definition 2.4. *Fix an initial data set (Σ, g_0, K), and suppose that (M_1, g_1, ι_1) and (M_2, g_2, ι_2) are GHDs of this initial data. We say that (M_2, g_2, ι_2) is an **extension** of (M_1, g_1, ι_1) if there exists an isometric embedding $\varphi\colon (M_1, g_1) \to (M_2, g_2)$ respecting the time orientations on M_1 and M_2 and such that the following diagram commutes:*

$$\begin{tikzcd} & \Sigma \arrow[dl, "\iota_1"'] \arrow[dr, "\iota_2"] & \\ M_1 \arrow[rr, "\varphi"] & & M_2 \end{tikzcd}$$

[1] "Isometric embedding" is meant in the differential-geometric sense, i.e., ι is an embedding of smooth manifolds and the pullback along ι of g is g_0.

The embedding $M_1 \to M_2$ realizes M_1 as an open subset of M_2 as both manifolds have the same dimension, and thus M_1 can be thought of as an open neighbourhood of (the image of) Σ in M_2.

With preliminary definitions out of the way, we can now precisely state what it means for the Cauchy problem for the Einstein vacuum equations to be locally well-posed.

Theorem 2.5 (Local well-posedness [3]). *Let (Σ, g_0, K) be an initial data set.*

(i) (Local existence) (Σ, g_0, K) admits a GHD.

(ii) (Local uniqueness) Any two GHDs M_1 and M_2 of (Σ, g_0, K) have a common GHD; that is, there exists a GHD M of (Σ, g_0, K) such that M_1 and M_2 are both extensions of M.

A common GHD of M_1 and M_2 can be thought of as an "open neighbourhood of Σ on which M_1 and M_2 agree".

3. From local to global

We now state a theorem by Choquet-Bruhat and Geroch which, in particular, upgrades Theorem 2.5 from local to global existence and uniqueness.

Theorem 3.1 ([2]). *Every initial data set (Σ, g_0, K) admits a* ***maximal*** *GHD, i.e., a GHD that is an extension of every other GHD of (Σ, g_0, K). The maximal GHD is unique up to isometry.*

Choquet-Bruhat and Geroch's proof goes roughly as follows. By Theorem 2.5(i), the collection $\mathcal{M}$ of GHDs of (Σ, g_0, K) is non-empty. Given $M_1, M_2 \in \mathcal{M}$, Theorem 2.5(ii) yields a common GHD of these; thus there are open subsets $U \subseteq M_1$ and $V \subseteq M_2$ which are GHDs of the initial data, and a time orientation-preserving isometry $\varphi\colon U \to V$ making the relevant diagram commute.

Let $C(M_1, M_2)$ denote the set of all such pairs (U, φ). A short argument shows that, if (U, φ) and (U', φ') are elements of $C(M_1, M_2)$, then φ and φ' agree on $U \cap U'$. Thus, φ and φ' can be glued together to define an isometric embedding of $U \cup U'$ into M_2 as a GHD. It follows that $C(M_1, M_2)$ is partially ordered by declaring that $(U, \varphi) \leq (U', \varphi')$ whenever $U \subseteq U'$, and every chain in $C(M_1, M_2)$ has an upper bound. Thus, by Zorn's lemma, $C(M_1, M_2)$ has a maximal element (U, φ), which is necessarily unique by the discussion above. If $U = M_1$ (i.e., M_2 is an extension of M_1), write $M_1 \preceq M_2$.

In fact, $\preceq$ is a partial order on $\mathcal{M}$, and we claim that every chain $\{M_\alpha\}_{\alpha \in A} \subseteq \mathcal{M}$ has an upper bound. To see this, we first introduce the notation $\varphi_{\beta\alpha}\colon M_\alpha \to M_\beta$ for $\alpha, \beta \in A$ such that $M_\alpha \preceq M_\beta$. Then the uniqueness argument above shows that the cocycle condition is satisfied:

$$\varphi_{\gamma\alpha} = \varphi_{\gamma\beta} \circ \varphi_{\beta\alpha} \text{ whenever } M_\alpha \preceq M_\beta \preceq M_\gamma.$$

Thus, we may form the direct limit topological space

$$\varinjlim_{\alpha \in A} M_\alpha = \left(\bigsqcup_{\alpha \in A} M_\alpha \right) \Big/ \sim,$$

where the equivalence relation $\sim$ is defined by $(\alpha, p) \sim (\beta, q)$ if $M_\alpha \preceq M_\beta$ and $q = \varphi_{\beta\alpha}(p)$. In fact, $\varinjlim_\alpha M_\alpha$ admits the structure of a smooth Lorentzian 4-manifold, and there is an isometric embedding making it a GHD of (Σ, g_0, K). This yields an upper bound for the chain $\{M_\alpha\}_{\alpha \in I}$.

Applying Zorn's lemma again, one obtains a maximal element $\widetilde{M}$ of $\mathcal{M}$. We claim that $\widetilde{M}$ is an extension of every other GHD; thus, $\widetilde{M}$ this is the desired maximal GHD of (Σ, g_0, K). Indeed, given any other GHD M, let (U, φ) be the (unique) maximal element of $C(M, \widetilde{M})$ and consider the adjunction space

$$M \cup_\varphi \widetilde{M} = \left(M \sqcup \widetilde{M} \right) \Big/ \sim,$$

where $x \sim \varphi(x)$ for $x \in U$. It is straightforward to see that $M \cup_\varphi \widetilde{M}$ satisfies the requirements to be a GHD except for possibly the Hausdorff condition, which must be verified separately. Assuming by way of contradiction that $M \cup_\varphi \widetilde{M}$ is not Hausdorff, one can show that points which fail the Hausdorff condition come in pairs $p \neq p'$ where $p \in \partial U$ and $p' \in \partial\varphi(U)$; in fact, p uniquely determines p' and vice versa. Given such a pair, one can construct a spacelike 3-dimensional hypersurface $T \subset M$ passing through p such that $T \setminus \{p\} \subseteq U$. Then $T' := \varphi(T \setminus \{p\}) \cup \{p'\}$ is a spacelike 3-dimensional hypersurface in $\widetilde{M}$. But these hypersurfaces come with isometric initial data, and so by Theorem 2.5(ii), the maximal common GHD of M and $\widetilde{M}$ can actually be extended to a neighbourhood of p. This contradicts maximality of the maximal common GHD. (An extensively detailed proof of Hausdorffness can be found in [5, Chapter 23].)

Having shown that $M \cup_\varphi \widetilde{M}$ is Hausdorff, it follows that this space is a GHD which is an extension of both M and $\widetilde{M}$. On the other hand, $\widetilde{M}$ being a maximal element with respect to $\preceq$ implies that $M \cup_\varphi \widetilde{M} = \widetilde{M}$. Thus $\widetilde{M}$ is an extension of M, as desired. Finally, uniqueness of the maximal GHD follows from uniqueness of the maximal elements in each set $C(M_1, M_2)$.

At this junction, we must note two small fibs in the above proof sketch:

- The collection $\mathcal{M}$ of GHDs is *not* a set (which is why we avoided calling it such), but a proper class—thus, one would have to invoke a "Zorn's lemma for proper classes" to make their proof go through fully. This can be rectified by redefining $\mathcal{M}$ to be the *set* of GHDs (M, g, ι) of (Σ, g_0, K) such that M is an open subset of $\Sigma \times \mathbb{R}$ which contains $\Sigma \times \{0\}$, and such that $\iota(x) = (x, 0)$ for all $x \in \Sigma$.
- The "partial order" $\preceq$ on $\mathcal{M}$ is not actually a partial order, as it fails to be antisymmetric: if $M_1 \preceq M_2$ and $M_2 \preceq M_1$, then M_1 and M_2 are isometric,[2] but they need not be literally equal. This is fixed by identifying isometric GHDs and working with isometry classes.

Remark 3.2. As noted by Choquet-Bruhat and Geroch in [2], this local-to-global result does not rely on the particular form of the Einstein vacuum equations; it only relies on the local well-posedness property (Theorem 2.5). Thus their argument also works for more general Cauchy problems, such as those for the Einstein field equations with sources given by perfect fluids or electromagnetic fields.

4. De-Zornification

It is interesting to note that Choquet-Bruhat and Geroch's proof of Theorem 3.1 invokes Zorn's lemma twice: first to guarantee the existence of a maximal common GHD of two GHDs, and second to guarantee the existence of a maximal element of $\mathcal{M}$. From a philosophical point of view, this can seem somewhat objectionable: if a maximal GHD is supposed to model a physical universe, one would hope to have an explicit description of it, and yet the proof presented does not offer any insight into what the maximal GHD looks like. Moreover, our models of physics appear to be mostly agnostic towards accepting or rejecting Zorn's lemma.[3] One might therefore reasonably ask if Theorem 3.1 truly requires Zorn's lemma.

In fact, Sbierski [6] has given a proof of Theorem 3.1 which, after blackboxing some results from PDE theory, explicitly constructs the maximal GHD in the setting of $\mathsf{ZF} + \mathsf{CC}$ (Zermelo–Fraenkel set theory with the axiom of countable choice); in particular, the proof does not rely on Zorn's lemma. The black boxes use the axiom of *dependent* choice, which is stronger than the axiom of

[2]This fact is not immediate; it can be deduced from [6, Corollary 3.2].

[3]Although quantum mechanics uses functional analysis to great extent and many results in functional analysis rely on Zorn's lemma/the axiom of choice in some way, it seems that the spaces encountered in quantum mechanics are at worst separable (to the best of the author's knowledge). For such spaces, the full power of Zorn's lemma is not needed to prove most functional-analytic results; one can get away with weaker axioms, such as the axiom of dependent choice.

countable choice but weaker than the full axiom of choice, thus making the existence of the maximal GHD a theorem in $\mathsf{ZF}+\mathsf{DC}$. For the remainder of this section, all mentions of GHDs will be meant in reference to a fixed initial data set (Σ, g_0, K).

To begin, we introduce a slightly different definition of "common GHD" from the one referred to in Theorem 2.5(ii). The main difference is that, whereas before, a common GHD was a GHD with preferred embeddings into two given GHDs, now a common GHD will actually be a *subset* of one of the two GHDs and come with an embedding into the other.

Definition 4.1. *A **common GHD** of two GHDs (M_1, g_1, ι_1) and (M_2, g_2, ι_2) is a GHD $(U, g_1|_U, \iota_1)$ where $U \subseteq M_1$ is an open neighbourhood of $\iota_1(\Sigma)$, and such that (M_2, g_2, ι_2) is an extension of $(U, g_1|_U, \iota_1)$ in the sense of Definition 2.4.*

Of course, a common GHD in the sense of Theorem 2.5(ii) yields a common GHD in the sense of Definition 4.1. One might wonder why we introduce this definition—after all, it seems to break the symmetry between M_1 and M_2 by requiring that a common GHD be an actual subset of M_1. The reason is that this approach avoids the use of Zorn's lemma to ensure the existence of a maximal common GHD as is done in Choquet-Bruhat and Geroch's proof. With this definition, one can instead take the literal union of all common GHDs of M_1 and M_2 to construct the maximal common GHD as a submanifold of M_1 [6, Theorem 3.4]—all the desired properties then follow immediately.

Definition 4.2. *Let $U \subseteq M_1$ be a common GHD of M_1 and M_2 with embedding $\varphi\colon U \to M_2$. Two points $p \in \partial U$ and $p' \in \partial\varphi(U)$ are called **corresponding boundary points** if for all neighbourhoods V of p and V' of p', respectively, we have $\varphi^{-1}(V' \cap \varphi(U)) \cap V \neq \varnothing$.*

This condition should be thought of as meaning that the images of p and p' in the adjunction space $M_1 \cup_\varphi M_2$ form a pair of points which fail the Hausdorff condition. Analogously to Choquet-Bruhat and Geroch's proof that the adjunction space is Hausdorff, Sbierski shows [6, Theorem 3.6] that the maximal common GHD does not have any corresponding boundary points by proving that any common GHD that *does* have corresponding boundary points can be extended to a strictly larger common GHD. (The axiom of countable choice is used in the proof of this fact.)

As a stepping stone to constructing the maximal GHD, Sbierski first proves the following global uniqueness result.

Theorem 4.3. *For any two GHDs M_1 and M_2, there exists a GHD which is a common extension of M_1 and M_2.*

Indeed, let $U \subseteq M_1$ be the maximal common GHD of M_1 and M_2, and let $\varphi\colon U \to M_2$ be its embedding into M_2. Then $M_1 \cup_\varphi M_2$ is Hausdorff by the discussion in the preceding paragraph, and this furnishes the desired common extension once equipped with a suitable smooth structure, Ricci-flat Lorentzian metric, and time orientation.

To construct the maximal GHD of (Σ, g_0, K), the goal is to "glue *all* GHDs together along their maximal common GHDs". This cannot work on the nose, however: as we mentioned at the end of Section 3, the collection $\mathcal{M}$ of all GHDs is not a set, but a proper class. To sidestep this issue, we again replace $\mathcal{M}$ with the set(!) $\{M_\alpha \mid \alpha \in A\}$ of GHDs which are an open neighbourhood of $\Sigma \times \{0\}$ in $\Sigma \times \mathbb{R}$, and for which the embedding $\Sigma \to \Sigma \times \{0\}$ maps $x \mapsto (x, 0)$. Given two such GHDs M_α and M_β, we let $U_{\alpha\beta} \subseteq M_\alpha$ denote the their maximal common GHD, and $\varphi_{\alpha\beta}\colon U_{\alpha\beta} \to M_\beta$ the corresponding embedding. Then the maximal GHD is given by

$$\widetilde{M} := \left(\bigsqcup_{\alpha \in A} M_\alpha \right) \Big/ \sim,$$

where we identify $(\alpha, p) \sim (\beta, q)$ if $p \in U_{\alpha\beta}$ and $q = \varphi_{\alpha\beta}(p)$. Verifying that this quotient space can be endowed with the structure of a GHD is done similarly to before; to construct an embedding from any other GHD M into $\widetilde{M}$, one uses the flow of a globally timelike vector field on M (which is provided by the time orientation on M). Uniqueness of the maximal GHD follows, unsurprisingly, from maximality.

We conclude by remarking that Wong [7] has given a construction of the maximal GHD in an even weaker framework than $\mathsf{ZF} + \mathsf{CC}$ (again, after blackboxing some background results); namely, Wong's proof is valid in ZF + "every countable union of countable sets is countable".

References

1. Antonio N. Bernal and Miguel Sánchez, *On smooth Cauchy hypersurfaces and Geroch's splitting theorem*, Communications in Mathematical Physics **243** (2003), no. 3, 461–470.
2. Yvonne Choquet-Bruhat and Robert Geroch, *Global aspects of the Cauchy problem in general relativity*, Communications in Mathematical Physics **14** (1969), no. 4, 329–335.
3. Yvonne Fourès-Bruhat, *Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires*, Acta Mathematica **88** (1952), 141–225.
4. Robert Geroch, *Domain of dependence*, Journal of Mathematical Physics **11** (1970), no. 2, 437–449.
5. Hans Ringström, *On the topology and future stability of the universe*, 1. ed., Oxford mathematical monographs, Oxford Univ. Press, Oxford, 2013.
6. Jan Sbierski, *On the existence of a maximal Cauchy development for the Einstein equations: a Dezornification*, Annales Henri Poincaré **17** (2016), 301–329.
7. Willie Wai-Yeung Wong, *A comment on the construction of the maximal globally hyperbolic Cauchy development*, Journal of Mathematical Physics **54** (2013), no. 11, 113511.