1 Introduction

Global spacetime structure concerns the more foundational aspects of general relativity (e.g. the topological and causal structure of spacetime). Upon investigation, it is often the case that seemingly plausible statements concerning global spacetime structure turn out to be false. Indeed, even after the shift to a relativistic worldview it seems "we are still somewhat over-conditioned to Minkowski spacetime" (Geroch & Horowitz, 1979, p. 215). This Element can be viewed as a kind of manual to help us unlearn what we think we know concerning the global structure of spacetime. A large number of example spacetimes (with diagrams) are central to the presentation and serve to demonstrate just how much is permitted under general relativity. Along the way, open questions are highlighted and periodic exercises can be used to test one's understanding (sample solutions are given in the Appendix).

Section 2 concerns the basic structure of spacetime. A number of preliminary definitions are presented to get things started. The cut-and-paste method is also introduced, which is used throughout to construct a vast array of example spacetimes. Although such spacetimes may seem artificial in some sense, we find that "the mere existence of a space-time having certain global features suggests that there are many models – some perhaps quite reasonable physically – with very similar properties" (Geroch, 1971a, p. 78). Section 3 covers the causal structure of spacetime. It follows a fairly conventional presentation of the hierarchy of causality conditions (Hawking & Ellis, 1973; Wald, 1984). But some nonstandard topics of interest are also explored including the so-called Malament-Hogarth spacetimes allowing for "supertasks" of a certain kind (Earman & Norton, 1993).

Section 4 concerns the singular structure of spacetime. An example singularity theorem is presented showing a sense in which some "physically reasonable" spacetimes have singularities (cf. Hawking & Penrose, 1970). This raises a difficulty in how to sort singular spacetimes into physically reasonable and physically unreasonable varieties. Two families of conditions are investigated that are meant to do the sorting. One family primarily concerns the causal structure of spacetime and forbids "naked" singularities of various types; the other family primarily concerns the modal structure of spacetime and forbids spacetime "holes" of various types. After considering a rich collection of examples, the upshot seems to be that what counts as a physically reasonable spacetime is far from clear (Earman, 1995, p. 86).

As we leave old intuitions behind, a rather basic question arises: What can we know concerning the global structure of spacetime? Building on a trio of papers from Geroch (1977), Glymour (1977), and Malament (1977a), Section 5 explores the epistemic structure of spacetime. It seems that even after we have (i) taken into consideration all possible observational data we could ever (even in principle) gather and (ii) inductively fixed the local features of any unobservable regions of spacetime, a type of "cosmic underdetermination" keeps us from pinning down the global structure of the universe. And if we take seriously the idea that we cannot come to know the global structure of spacetime through observation, queer possibilities present themselves. Does our universe allow for "time travel" of a certain kind? Do spacetime "holes" exist in our universe? This suggests that perhaps we have been too quick to discount as physically unreasonable some of the more peculiar global spacetime properties since, for all we know, such properties obtain in our own universe.

In Section 6, the modal structure of spacetime is explored through the lens of the inextendibility condition. This is the requirement that the universe be as large as possible relative to a standard background collection of spacetimes. But the inextendibility condition would seem to be physically significant only insofar as the background collection coincides with physically reasonable possibilities (Geroch, 1970a). And because what counts as a physically reasonable spacetime is not clear – especially given the underdetermination results just mentioned – it seems natural to consider various nonstandard definitions of inextendibility in a pluralistic way. Upon investigation, we find that foundational claims concerning inextendibility can fail to hold up under some modified definitions. For example, it can happen that a spacetime is "extendible" and yet has no "inextendible extension" – a strange state of affairs with the potential to clash with various Leibniz-inspired metaphysical principles in favor of the "maximality" of spacetime (Earman, 1995, p. 32). In addition, the demand for modified forms of inextendibility can lead to situations in which a spacetime is forced into having global properties of interest. A so-called time machine represents one example along these lines, but other "machine" spacetimes can also be studied (cf. Earman et al., 2016). Stepping back, we find that the prospect of a clear distinction between physically reasonable and physically unreasonable spacetimes is more elusive than ever.

2 Preliminaries

A (general relativistic) ***spacetime*** is a pair (M, g_{ab}) where M is a smooth, connected, Hausdorff, paracompact, n-dimensional ($n \geq 2$) manifold and g_{ab} is a smooth metric on M of Lorentz signature $(+, -, \ldots, -)$. Under the assumption of Einstein's equation (see p. 6), a spacetime is a model of general relativity and represents a possible universe compatible with the theory. Details concerning the relevant background mathematics (including the “abstract index” notation used throughout) can be found in Hawking and Ellis (1973), Wald (1984), or Malament (2012). Here, we follow Geroch & Horowitz (1979) in avoiding technical machinery whenever possible.

We begin with the notion of a ***manifold***, which, unless otherwise stated, is taken to be smooth, connected, Hausdorff, and paracompact (see the Appendix for basic topological definitions). All of the topological structure of a spacetime (M, g_{ab}) is given by the manifold M; it fully captures the shape of the model. Locally, a manifold looks like plain old $\mathbb{R}^n$ although globally it may have a very different structure. A number of manifolds are easy to visualize. For example, consider the sphere S^2. Despite its round shape, if one zooms in on the vicinity of any point, one finds it has the same topological structure as the plane $\mathbb{R}^2$ (see Figure 1). Other two-dimensional manifolds include the cylinder $S^1 \times \mathbb{R}$ and the torus $S^1 \times S^1$. In addition, the result of taking any manifold and removing from it a closed proper subset also counts as a manifold. For example, a new manifold $\mathbb{R}^2 - \{(0,0)\}$ can be constructed by excising the origin from the plane.

We say the n-dimensional manifolds M and N are ***diffeomorphic*** if there is a bijection $\varphi : M \to N$ such that both it and its inverse are smooth. Diffeomorphic manifolds have identical topological and smoothness properties. It turns out that every non-compact manifold of two dimensions or more admits some Lorentzian metric. One can also show that the compact manifold S^n for $n \geq 2$ admits a Lorentzian metric if and only if n is odd (Geroch & Horowitz, 1979). We also have the useful result that any manifold $M \times N$ admits a Lorentzian metric if either M or N does. And of course, if M admits a Lorentzian metric, then so does $M - C$ where C is any closed proper subset of M.

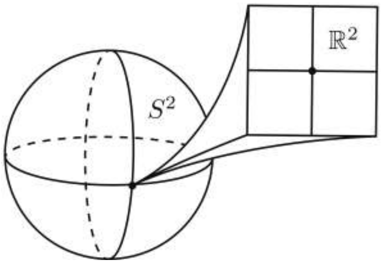

Figure 1 The sphere S^2 has the same topological structure as the plane $\mathbb{R}^2$ in the vicinity of each point.

Exercise 1 Find a manifold M and a point $p \in M$ such that M and $M - \{p\}$ are diffeomorphic.

Each point on a manifold represents an idealized possible event in spacetime (e.g. one's birth). The Lorentzian ***metric*** tells us how such events in spacetime are related to one another. Consider a spacetime (M, g_{ab}). At each point $p \in M$, the metric g_{ab} assigns to each vector ξ^a in the tangent space of p a length given by $\xi^a \xi^b g_{ab} = \xi^a \xi_a \in \mathbb{R}$. This creates a type of double cone structure in the tangent space of each point. Positive-length vectors are ***timelike*** and fall inside the cone, negative-length vectors are ***spacelike*** and fall outside the cone, and zero-length vectors are ***null*** and make up the boundary of the cone (see Figure 2).

One can think of the cone structure at each point as representing the speed of light in all directions there; timelike and spacelike vectors represent, respectively, velocities that are slower and faster than light. For this reason, we often refer to these structures as light cones in what follows. Now consider a smooth curve $\gamma : I \to M$ where I is some connected interval of $\mathbb{R}$. (In what follows, curves are understood to be smooth unless otherwise stated.) If each of its tangent vectors ξ^a is timelike according to g_{ab}, then we say the curve γ is ***timelike***. Timelike curves represent the possible trajectories of massive objects. Analogous definitions can be given for ***spacelike*** and ***null*** curves; a ***causal*** curve has no spacelike tangent vectors (see Figure 3).

Associated with g_{ab} is a unique ***derivative operator*** ∇_a on M that is compatible with the metric in the sense that $\nabla_a g_{bc} = \mathbf{0}$. We say that a given curve $\gamma : I \to M$ is a ***geodesic*** if, for each each point along the curve, the tangent vector ξ^a is such that $\xi^a \nabla_a \xi^b = \mathbf{0}$. One can think of a geodesic as a curve that is as straight as possible according to a given metric. Timelike geodesics

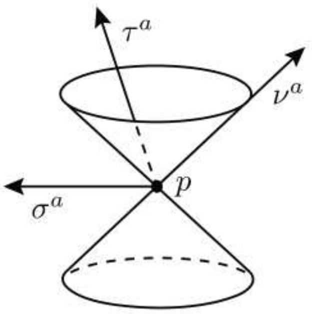

Figure 2 A three-dimensional double cone structure at the point p. A timelike vector τ^a, a null vector ν^a, and a spacelike vector σ^a are depicted.

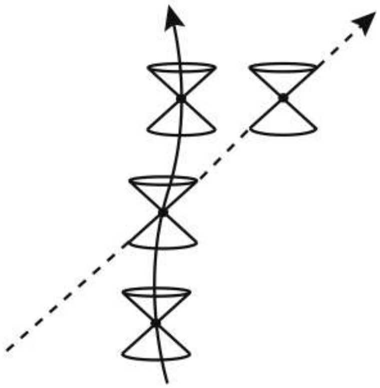

Figure 3 A pair of causal curves in a three-dimensional spacetime. One is timelike (solid line) and one is null (dotted line).

represent the possible trajectories of non-accelerating (freely falling) massive objects; null geodesics represent the possible trajectories of light. In any spacetime (M, g_{ab}), one can always find some open neighborhood $O \subseteq M$ around any point $p \in M$ such that any two points $q, r \in O$ can be connected by a unique geodesic whose image is contained in O.

Exercise 2 Find a spacetime (M, g_{ab}) and a pair of points $p, q \in M$ that can be connected by spacelike and null geodesics but not by a timelike geodesic.

A curve $\gamma : I \to M$ in a spacetime (M, g_{ab}) is ***maximal*** if there is no curve $\gamma' : I' \to M$ such that I is properly contained in I' and $\gamma(s) = \gamma'(s)$ for all $s \in I$. If a maximal geodesic $\gamma : I \to M$ is such that $I \neq \mathbb{R}$, then we say it is ***incomplete***. A spacetime that harbors an incomplete geodesic is ***geodesically incomplete***; otherwise it is ***geodesically complete***. An incomplete timelike geodesic can be considered a type of singularity since it represents a possible trajectory of a freely falling massive object whose existence is cut short in either the past or future direction (cf. Geroch, 1968a; Curiel, 1999). By excising points from the manifold, one can easily create examples of geodesically incomplete spacetimes (see Figure 4).

Given a spacetime (M, g_{ab}), one can use its associated derivative operator ∇_a to define the ***Riemann tensor*** $R^a{}_{bcd}$ where $R^a{}_{bcd}\xi^b = -2\nabla_{[c}\nabla_{d]}\xi^a$ for all smooth vector fields ξ^a. Here, the square brackets indicate the antisymmetrization operation. In this case, we find that $-2\nabla_{[c}\nabla_{d]}\xi^a = -(\nabla_c\nabla_d - \nabla_d\nabla_c)\xi^a$ (see Malament, 2012, p. 33). The Riemann tensor encodes all of the curvature of spacetime at each point in M. A spacetime is ***flat*** if its Riemann tensor vanishes

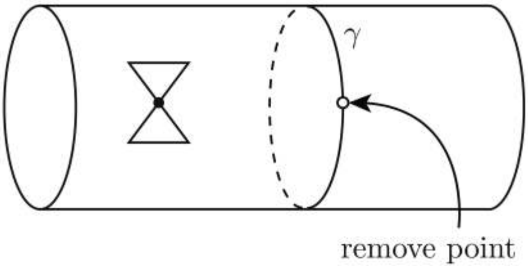

Figure 4 The timelike geodesic γ is maximal but incomplete since it cannot be extended through the missing point.

everywhere. The contraction of the Riemann tensor leads to the ***Ricci tensor*** $R_{ab} = R^c{}_{abc}$ and the ***Ricci scalar*** $R = R^a{}_a$ (see Malament, 2012, 84). The distribution of matter in spacetime can be represented by the ***energy-momentum tensor*** T_{ab} defined via Einstein's equation: $R_{ab} - (1/2)Rg_{ab} = 8\pi T_{ab}$. Here, we have ignored the possibility of a nonzero "cosmological constant" term in Einstein's equation (see Earman, 2001). Indeed, within the field of global structure there is a general lack of concern with the details of Einstein's equation; we find that "things which can happen in the absence of this equation can usually also happen in its presence" (Geroch & Horowitz, 1979, p. 215). If a spacetime is such that its corresponding energy-momentum tensor vanishes everywhere, then it is ***vacuum***. It turns out that any two-dimensional spacetime is vacuum (see Fletcher et al., 2018). In dimension three or greater, a spacetime is vacuum if and only if its associated Ricci tensor vanishes everywhere. Of course, any flat spacetime is necessarily vacuum.

We are now in a position to define ***Minkowski spacetime*** – it is any flat, geodesically complete spacetime with manifold $\mathbb{R}^n$. In standard (t, x) coordinates, two-dimensional Minkowski spacetime comes out as $(\mathbb{R}^2, g_{ab})$ where $g_{ab} = \nabla_a t \nabla_b t - \nabla_a x \nabla_b x$. This is the spacetime of special relativity and the vanilla model of general relativity. In what follows, we use Minkowski spacetime as our basic tool to construct various examples; we cut it, glue it, bend it, and warp it in order to get what we need. In a representation of Minkowski spacetime in standard coordinates, the light cones are uniformly oriented throughout and all geodesics appear as straight lines (see Figure 5).

Exercise 3 Find a flat spacetime such that every maximal timelike geodesic is incomplete but some maximal null and spacelike geodesics are complete.

Some spacetimes (M, g_{ab}) admit a continuous timelike vector field ξ^a on M and some do not. Those that do (e.g. Minkowski spacetime) allow for a

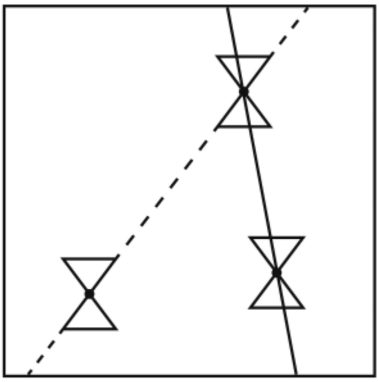

Figure 5 A timelike geodesic (solid line) and a null geodesic (dotted line) in two-dimensional Minkowski spacetime.

consistent global distinction between the "past" and "future" temporal directions since the continuous timelike vector field picks out one of two "lobes" of the light cone at each point. Such spacetimes are said to be ***time-orientable***. One can show that any spacetime (M, g_{ab}) is time-orientable if M is simply connected. A classic example of a spacetime that fails to be time-orientable can be constructed by starting with a Möbius strip manifold and orienting the light cones in such a way that any would-be continuous timelike vector field is flipped when transported around the strip (see Figure 6). In the following, we assume that spacetimes are time-orientable and that a temporal direction has been chosen. A causal curve $\gamma : I \to M$ in a spacetime (M, g_{ab}) is ***future-directed*** if its tangent vector at each point falls in or on the future lobe of the light cone or vanishes; an analogous definition can be given for ***past-directed*** causal curves. Unless otherwise stated, causal curves are understood to be future-directed.

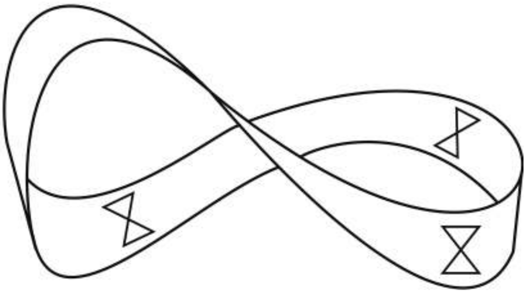

Figure 6 A spacetime that fails to be time-orientable since the flip in the Möbius strip precludes any continuous timelike vector field.

Exercise 4 Find a spacetime (M, g_{ab}) for some $M \subset \mathbb{R}^2$ that fails to be time-orientable.

Consider a spacetime (M, g_{ab}) and a pair of points p and q in M that, respectively, represent the past event of one's birth and the future event of one's reading of this sentence. One's trajectory through spacetime from the first event to the second can be represented by a future-directed timelike curve $\gamma : I \to M$ connecting p to q. The metric g_{ab} assigns a ***length*** $\|\gamma\| = \int (g_{ab}\xi^a\xi^b)^{1/2} ds$ to this curve by adding up the lengths of all the tangent vectors ξ^a along the curve. This length represents the elapsed time between p and q along γ. It follows that the elapsed time between any two events will depend on how one moves through spacetime from one to the other. Some trajectories with velocity vectors "close to the speed of light" will have a short elapsed time relative to others. Indeed, continuity considerations require that if two points can be connected by a timelike curve, then for any $\epsilon > 0$, there is a timelike curve connecting the points with length less than ϵ. It turns out that some spacetimes (e.g. Minkowski spacetime) are such that if two points can be connected by a timelike curve, then there is a longest curve connecting the points that must be a geodesic (see Figure 7).

A point $p \in M$ in a spacetime (M, g_{ab}) is a ***future endpoint*** of a future-directed causal curve $\gamma : I \to M$ if, for every open neighborhood O of p, there exists a point $s' \in I$ such that $\gamma(s) \in O$ for all $s > s'$. A ***past endpoint*** is defined analogously. We say that a causal curve is ***future-inextendible*** if it has no future endpoint and analogously for ***past-inextendible***. A causal curve is ***inextendible*** if it is both future-inextendible and past-inextendible. A causal curve that is inextendible must be maximal, but the converse is false. In Minkowski

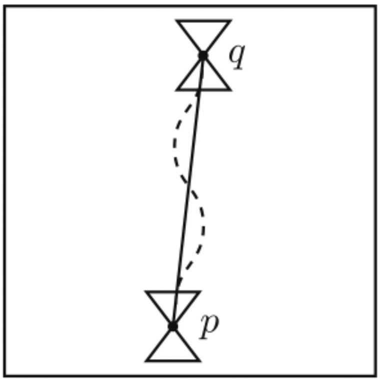

Figure 7 The points p and q can be connected by a short timelike curve (dotted line), but the longest such curve will be a geodesic (solid line).

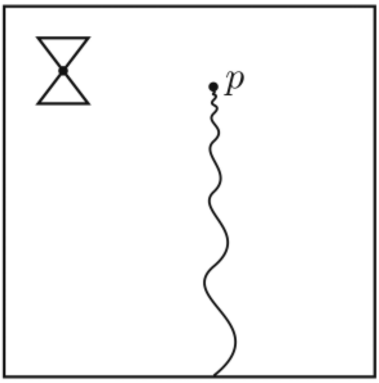

Figure 8 A maximal timelike curve with future endpoint p.

spacetime, a timelike curve can "wiggle" faster and faster as a future endpoint (which is not part of the curve) is approached (cf. Penrose, 1972, p. 3). The curve counts as maximal since any extension through the endpoint must fail to be smooth (see Figure 8).

Given an n-dimensional spacetime (M, g_{ab}), a set $S \subset M$ is a ***spacelike surface*** if S is an $(n-1)$-dimensional sub-manifold of M such that every curve whose image is contained in S is spacelike. A set $S \subset M$ in a spacetime (M, g_{ab}) is ***achronal*** if no two points in S can be connected by a timelike curve. The ***edge*** of a closed, achronal set $S \subset M$ is the collection of points $p \in S$ for which every open neighborhood O of p contains points q and r such that future-directed timelike curves exist from q to p, from p to r, and from q to r where the last curve fails to intersect S (see Figure 9). A ***slice*** is a closed, achronal set with an empty edge. In Minkowski spacetime in standard (t, x) coordinates, each $t =$ constant surface counts as a slice. But not all spacetimes admit slices. For example, consider the spacetime $(S^1 \times \mathbb{R}, g_{ab})$ where $g_{ab} = \nabla_a t \nabla_b t - \nabla_a x \nabla_b x$ and $0 \leq t \leq 2\pi$; this is just Minkowski spacetime that has been "rolled up" along the time direction. Let this spacetime be called ***time-rolled Minkowski spacetime***. In an analogous way, one can also construct other two-dimensional models: ***space-rolled***, ***null-rolled***, and ***(time and space)–rolled Minkowski spacetimes***.

Exercise 5 Find a spacelike surface in Minkowski spacetime that fails to be achronal.

A diffeomorphism $\varphi : M \to M'$ between the spacetimes (M, g_{ab}) and (M', g'_{ab}) is an ***isometry*** if $\varphi^*(g'_{ab}) = g_{ab}$ where φ^* is the map associated with φ, which,

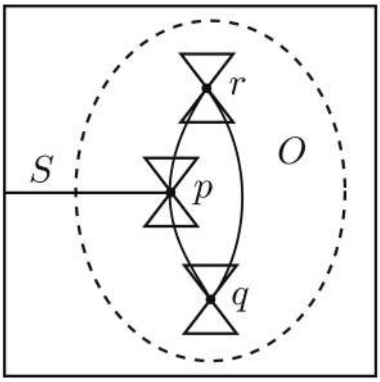

Figure 9 The point p is in the edge of the closed, achronal set S since every open neighborhood O of p contains points q and r such that future-directed timelike curves exist from q to p, from p to r, and from q to r where the last curve fails to intersect S.

for any point $p \in M$, pulls back the tensor g'_{ab} at $\varphi(p) \in M'$ to the tensor $\varphi^*(g'_{ab})$ at $p \in M$ (Malament, 2012, p. 36). Spacetimes (M, g_{ab}) and (M', g'_{ab}) are ***isometric*** if there is an isometry between them. Isometric spacetimes have fully equivalent structure and share all of the same physical properties; indeed, when no confusion arises, we often take isometric spacetimes to be the same spacetime in what follows. Consider the spacetimes (M, g_{ab}) and (M', g'_{ab}). If there is a proper subset O of M' such that (M, g_{ab}) and (O, g'_{ab}) are isometric, then we say that (M, g_{ab}) is ***extendible*** and (M', g'_{ab}) is an ***extension*** of (M, g_{ab}). A spacetime that is not extendible is ***inextendible***.

Exercise 6 Find a pair of non-isometric spacetimes such that each counts as an extension of the other.

It turns out that every geodesically complete spacetime (e.g. Minkowski spacetime) is inextendible. But the other direction does not hold. To see this, consider Minkowski spacetime in standard (t, x) coordinates and remove two slits $S_n = \{(0, n) : n \leq x \leq n + 1/2\}$ for $n = 1, 2$. Excluding the four slit boundary points, identify the top edge of each slit with the bottom edge of the other (Hawking & Ellis, 1973, p. 58; Geroch, 1977, p. 89). The resulting spacetime is such that an observer entering one slit from below must emerge from the other slit from above. Because the four slit boundary points are "missing" from the spacetime, there are incomplete geodesics (see Figure 10). But one can show that this spacetime cannot be extended.