The Space of Null Geodesics (and a New Causal Boundary)

Robert J. Low

Mathematics Group, School of MIS, Coventry University, Priory Street, Coventry CV1 5FB, U.K.
mtx014@coventry.ac.uk

Abstract. The space of null geodesics, G, of a space-time, $\mathcal{M}$, carries information on various aspects of the causal structure $\mathcal{M}$. In this contribution, we will review the space of null geodesics, G, and some natural structures which it carries, and see how aspects of the causal structure of $\mathcal{M}$ are encoded there. If $\mathcal{M}$ is strongly causal, then G has a natural contact manifold structure, points are represented in G by smooth Legendrian S^2s, and the relationships between these S^2s reflect causal relationships between the points of $\mathcal{M}$. One can also attempt to pass in the opposite direction with the intention of constructing a space-time from a family of S^2s in G; this process suggests a means of attaching end-points to null geodesics of $\mathcal{M}$, and thereby constructing a causal boundary. We close by summarizing some open questions in this general area.

1 Introduction

In Newtonian physics, the structure of space-time is fairly straight-forward. Although there is no notion of absolute rest, there is a notion of absolute time, and it always makes sense to say whether two events are simultaneous, and if not, which of them occurs first. Simultaneity is an equivalence relation, and can be used to slice space-time up into surfaces in a canonical way. Whether the situation at one event, p, can influence that at another, q, is determined by which of them happens at the later time (unless they are simultaneous, in which case they are independent). However, although this is conceptually straight-forward, it suffers from a major drawback: namely, it does not agree with the available data.

In general relativity, on the other hand, we model space-time as a smooth differentiable manifold, $\mathcal{M}$, equipped with a smooth Lorentz metric, g. Quantities associated with the metric, for example the connection, the Riemann tensor, the Ricci tensor and others, have physical interpretations and can be used to develop physics in this setting, and make testable - and tested! - predictions, which to date have proven consistent with the available experimental data [1]. There is also considerable activity in developing and carrying out new experimental tests of effects which have so far been too subtle to detect; two programmes of particular importance are Gravity Probe B [2] and LIGO [3].

R.J. Low: *The Space of Null Geodesics (and a New Causal Boundary)*, Lect. Notes Phys. **692**, 35–50 (2006)
www.springerlink.com

However, the question of when information at one event can affect that at another is rather less straight-forward than in the case of Newton's universe. The Lorentz metric implicitly provides the answer to this question: for it determines whether a curve in space-time can describe the path of a material particle, or influence. But in the general case, the situation can be much more complicated than in the Newtonian one.

For example, one can find space-times which locally look perfectly acceptable, but have the property that a particle can have a closed space-time trajectory, i.e. it can travel in such a way that it meets its own past self. The first, and perhaps the most famous, solution of this type was discovered by Gödel [4]. Such space-times are philosophically problematic, and the subject of much debate [5]; they raise awkward questions about the nature of free will, or nonlocal constraints on initial data. But even if we simply exclude such awkward behaviour by fiat, and restrict our attention to space-times with more acceptable behaviour, the situation remains complicated. Indeed, if a metric is simply expressed in terms of coordinate patches, it may be a difficult task to check whether the space-time is causally acceptable.

But considerations of causal structure are further reaching than just providing a reason for rejecting certain space-times as unacceptable. One would also like to know whether apparently reasonable initial data has a reasonable time evolution, or does some kind of singular behaviour occur eventually? If so, is this singularity decently concealed or can it be naked [6]? What are the properties of the edge, or boundary, of space-time? Even stating such questions precisely requires a good deal of causal machinery.

So, let us recall the basic question: given two points, p and q in $\mathcal{M}$ are p and q causally related in the sense that a physical influence can propagate from one to the other? This causal structure is distinctly more primitive than the metric structure on $\mathcal{M}$, since space-times with different metrics may have the same causal structure. It is, on the other hand, inextricably bound up with the metric structure since two space-times have the same causal structure if and only if they have the same null geodesics. Indeed, so crucial to the causal structure are the null geodesics that one can take the null geodesics themselves as primitive objects, regard points of space-time as derived objects, and profitably study aspects of the causal structure of space-time in this context instead.

In the remaining sections of this contribution I will describe the space of null geodesics, and in particular its topological and geometric structure. We will see how notions of causality in space-time are reflected in this new setting, providing elegant reinterpretations of familiar ideas, and also a powerful way of considering the development of wave fronts. Finally, I will suggest a means of attaching endpoints to endless null geodesics to provide a new type of conformal boundary.

First, however, I will give a very brief review of some causal theory, establishing standard terminology and definitions. This material is developed in depth in Penrose's lecture notes on differential topology [7] and (with slightly different conventions) in the classical text of Hawking and Ellis [8].

We denote space-time by $\mathcal{M}$ (and, by a standard abuse of notation, will use $\mathcal{M}$ when $\langle \mathcal{M}, g\rangle$ would be correct). The tangent bundle of $\mathcal{M}$ is $T\mathcal{M}$, with fibre $T_p\mathcal{M}$ at p, and the cotangent bundle is $T^*\mathcal{M}$ with fibre $T^*_p\mathcal{M}$ at p. The isomorphism beteen $T\mathcal{M}$ and $T^*\mathcal{M}$ provided by the metric g will be used freely. We will use the convention that the metric g has signature $(+,-,\dots,-)$, and say that a vector $v \in T_p\mathcal{M}$ is timelike if $g_p(v,v) > 0$, causal if $g_p(v,v) \geq 0$, null if $g_p(v,v) = 0$ and spacelike if $g_p(v,v) < 0$. Unless otherwise stated, $\mathcal{M}$ will be four-dimensional.

Also, $\mathcal{M}$ is said to be time-orientable if there exists a continuous timelike vector field t on $\mathcal{M}$; we will always assume that $\mathcal{M}$ is time-orientable. This does not in fact require any significant loss of generality: any space-time which is not time-orientable has a time orientable double cover [9]. Clearly, if t is such a timelike vector field, so is $-t$. We arbitrarily choose one of these as determining the future direction. As a consequence, we can distinguish between future pointing causal vectors (whose inner product with t is positive) and past pointing ones (whose inner product with t is negative).

A smooth curve is timelike (future pointing) if its tangent vector is everywhere timelike (future pointing), and similarly for causal, null, future or past pointing, or spacelike.

If $p, q \in \mathcal{M}$, then q is in the chronological future of p, written $q \in I^+(p)$, if there is a timelike future pointing curve $\gamma : [0,1] \to \mathcal{M}$ with $\gamma(0) = p$, and $\gamma(1) = q$; similarly, q is in the causal future of p, written $q \in J^+(q)$, if there is a future pointing causal curve from p to q. For any point, p, $I^+(p)$ is open; but $J^+(p)$ need not, in general, be closed. $J^+(p)$ is, however, always a subset of the closure of $I^+(p)$.

One can use properties of these two ordering relations to define a causal space in the absence of any notion of metric, and study causal structure in this more general setting [10]; more recently, similar ideas have been used in Sorkin's causal set program of quantum gravity [11].

In addition, $E^+(p) = J^+(p) \setminus I^+(p)$ is the future horismos of p, and is ruled by segments of null geodesics emanating from p. A null geodesic originating at p can leave $E^+(p)$ and enter $I^+(p)$ if it intersects another, or passes a conjugate point [8] (intuitively, a point where infinitesimally separated null geodesics starting at p cross one another). Then we have the inclusion $E^+(p) \subseteq \partial I^+(p) = \partial J^+(p)$. Denoting by $N^+(p)$ all those points lying on a future pointing null geodesic starting at p, we also have $E^+(p) \subseteq N^+(p)$. In general, neither of $N^+(p)$ or $\partial I^+(p)$ is a subset of the other.

If $K \subset \mathcal{M}$, then $D^+(K)$ is the set of all points p such that any past-endless causal curve through p intersects K. $D^+(K)$ is called the future domain of dependence of K, and is the region where physics is entirely determined by data on K (in the absence of material which allows faster-than-light effects to propagate), since no material influence can reach any element of $D^+(K)$ without passing through K.

One can use the causal relations I^+ and J^+ to impose conditions on space-time. Indeed, Carter has shown [14] that there is an infinite hierarchy of distinct conditions, each implied by all its successors, which can be imposed in terms of causal relations. I will restrain myself to listing a few of immediate relevance. Each of the following conditions is implied by its successor.

1. If there is no point p such that $p \in I^+(p)$, $\mathcal{M}$ is said to satisfy the chronological condition.
2. A space-time $\mathcal{M}$ which has no point p with a non-degenerate causal curve which starts and ends at p is said to satisfy the causal condition.
3. If each point p has arbitrarily small neighbourhoods which any causal curve intersects in a single component, $\mathcal{M}$ satisfies the condition of strong causality.
4. If $\mathcal{M}$ is causal and remains causal under small changes of g, it is stably causal.
5. If $\mathcal{M}$ is strongly causal and $J^\pm(p)$ is the topological closure of $I^\pm(p)$ for every $p \in \mathcal{M}$ (so $E^\pm(p) = \partial I^\pm(p)$, then $\mathcal{M}$ is causally simple.
6. If there is a spacelike surface $\mathcal{S}$ (i.e. a surface of codimension one whose tangent plane at each point contains only spacelike vectors) which every endless causal curve intersects in exactly one point, $\mathcal{M}$ is globally hyperbolic, and is the topological product of $\mathcal{S}$ with $\mathbb{R}$. $\mathcal{S}$ is called a Cauchy surface for $\mathcal{M}$.

We note that all of these concepts depend only on the conformal class of g, i.e. g may be replaced by Ωg, where Ω is a strictly positive function on $\mathcal{M}$, without any effect on causal properties.

2 Space of Null Geodesics

Even from the brief review above, it is clear that null geodesics are fundamental to the causal structure of $\mathcal{M}$. Motivated by this observation, we can consider the space of all null geodesics, and in particular investigate the relationships between its topology and geometry and the causal structure of $\mathcal{M}$.

In the following development, we will use the cotangent bundle of $\mathcal{M}$; for some purposes, it would be more natural to use the tangent bundle and the geodesic flow on the tangent bundle. Indeed, this approach has been used in the study of the space of geodesics of a Riemannian manifold [12] and in the more general case of the space of geodesics of a manifold with affine connection [13]. However, in the section after this one I wish to make use of some structures which naturally arise on the cotangent bundle, and so will work with the cotangent bundle from the beginning. As mentioned above, free use will be made of the isomorphism which g gives between the tangent and cotangent bundles. We will now consider how the cotangent bundle structure provides natural structure on the space of null geodesics.

So let $T^*\mathcal{M}$ be the cotangent bundle of $\mathcal{M}$, and $\pi : T^*\mathcal{M} \to \mathcal{M}$ the canonical projection. There are two vector fields on $T^*\mathcal{M}$ of interest.

Let $\alpha \in T^*\mathcal{M}$, and define $f : \mathbb{R} \to T^*\mathcal{M}$ by $f(t) = t\alpha$. Then Δ, the Euler field, is defined by $\Delta(\alpha) = f_*(\partial/\partial t)(1)$, or, more concretely, if α has coordinates (q^i, n_i) then $\Delta(\alpha)$ is $n_i\partial/\partial n_i$.

The other vector field we require is the geodesic vector field, X_G. To define this on $T^*\mathcal{M}$ we first define the Hamiltonian function $H : T^*\mathcal{M} \to \mathbb{R}$ which sends each covector to its squared length given by g, and then X_G is the corresponding Hamiltonian vector field determined by $i_{X_G}(\omega) = -dH$. In terms of the usual coordinates,

$$X_G = n^i \frac{\partial}{\partial q^i} - \Gamma^i_{jk} n^j n_i \frac{\partial}{\partial n_k}$$

where indices are raised and lowered using g, and Γ^i_{jk} is the usual Christoffel symbol.

If $c : \mathbb{R} \to \mathcal{M}$ is a smooth curve, given in coordinates by $c(t) = q^i(t)$, then it has a natural lift to $T^*\mathcal{M}$ given by $(q^i(t), n_i(t))$, where $n_i = g_{ij}\dot{q}^i$. This is an integral curve of X_G iff c is an affinely parameterised geodesic in $\mathcal{M}$.

Now, we can restrict our attention to $N^*\mathcal{M}$, the subset of $T^*\mathcal{M}$ given by the future pointing null vectors (excluding the zero vector at each point). Since each of Δ and X_G are tangent to $N^*\mathcal{M}$, we can regard them as vector fields on this manifold. Furthermore, since the two are never linearly independent, the vector space spanned by X_G and Δ at each point gives a two-dimensional distribution on $N^*\mathcal{M}$, i.e. a two-dimensional subspace of $T(N^*\mathcal{M})$ at each point. In addition, since $[\Delta, X_G] = X_G$, this distribution is integrable, i.e. $N^*\mathcal{M}$ is foliated by two-surfaces whose tangent surfaces are the subspaces spanned by X_G and Δ at each point [15]. We can therefore consider the quotient space of integral surfaces.

It is perhaps more geometrically intuitive to construct this quotient space in stages. First, we can take the quotient space of integral curves of Δ in $N^*\mathcal{M}$, resulting in the bundle of future pointing null directions, $\mathbb{P}N^*\mathcal{M}$. Now, although X_G itself does not descend to this quotient space, the one-dimensional distribution of subspaces spanned by X_G does, and we again obtain a distribution. This time, the integral curves are the lifts of null geodesics in $\mathcal{M}$ to the bundle of null directions of $\mathcal{M}$, and the quotient space is obtained by identifying points on the same (lifted) null geodesics. We therefore call this quotient space the space of null geodesics, and denote it by $\mathcal{N}$.

Alternatively, we can take the quotient of $N^*\mathcal{M}$ under the action of X_G, to obtain the space of scaled null geodesics, and then take the further quotient which corresponds to forgetting the scaling.

Now that $\mathcal{N}$ has been provided with a topology, we can consider convergence of a sequence of null geodesics as curves in space-time. So let γ_n be a sequence of points in $\mathcal{N}$, and denote by Γ_n the corresponding curves in $\mathcal{M}$. Suppose $\gamma \in \mathcal{N}$. Then, since a neighbourhood of γ in $\mathcal{N}$ is the image under the projection from $\mathbb{P}N^*\mathcal{N}$ of a neighbourhood of a point on the lift of Γ to $\mathbb{P}N^*\mathcal{N}$, we see that $\gamma_n \to \gamma$ if there is a sequence of points $p_n \in \Gamma_n$ and a point $p \in \Gamma$ such that $p_n \to p$, and the tangent direction to Γ_n at p_n tends to the tangent direction to Γ at p. More naively, two null geodesics are close if they pass close to each other and the tangent directions are also close.

Note that this is not the topology we obtain by insisting that, for any neighbourhood U of Γ, each Γ_n eventually lies inside U - it may well be that for each Γ_n there are points which are very far from Γ.

Example 1. Let $\mathcal{M}$ be the Minkowski space, with the usual coordinates (t, x, y, z), and let γ_n be the null geodesic through the origin with tangent $(1, \cos(1/n), \sin(1/n), 0)$. Then the limit null geodesic has tangent $(1, 1, 0, 0)$, but the distance between the points where Γ_n and Γ intersect the surface $t = T$ can be made arbitrarily large by taking T large enough.

Furthermore, although Frobenius' theorem guarantees the existence of integral surfaces, and hence of a quotient space which inherits a topological structure, this need not in general be a manifold.

Example 2. Let $\mathbb{M}^2$ be two-dimensional Minkowski space, with the usual coordinates (t, x), and let $\mathcal{M}$ be obtained from $\mathbb{M}^2$ by identifying t with $t + 1$ and x with $x + \sqrt{2}$ for all x, t. Then the space of null geodesics, $\mathcal{N}$ has two components: $\mathcal{L}$, the space of left directed null geodesics, and $\mathcal{R}$, the space of right directed ones. All right directed null geodesics are parallel, and each is dense in $\mathcal{M}$. As a consequence, each point of $\mathcal{R}$ is dense in $\mathcal{R}$, and similarly for $\mathcal{L}$.

It is a useful exercise to investigate the structure of $\mathcal{N}$ for two-dimensional toroidal space-times, where we identify x with $x + \alpha$ for various values of α.

3 Structures on the Space of Null Geodesics

Although $\mathcal{N}$ as a topological space need not in general be compatible with any manifold structure, we can guarantee that it is in fact a quotient manifold of $N^*\mathcal{M}$ by imposing a standard causal condition.

Theorem 1. *Let $\mathcal{M}$ be strongly causal. Then $\mathcal{N}$, the space of null geodesics of $\mathcal{M}$, inherits a manifold structure from $N^*\mathcal{M}$.*

Proof. If $\mathcal{M}$ is strongly causal, then every point in $\mathcal{M}$ has arbitrarily small neighbourhoods which null geodesics intersect in a single connected component. As a consequence, when the geodesics are lifted to the bundle of null directions over $\mathcal{M}$, this is also true of the lifts. Then the distribution is regular, and so the quotient space is a quotient manifold [15]. □

In general, if $\mathcal{M}$ is n-dimensional, $T^*\mathcal{M}$ is $2n$-dimensional, so that $N^*\mathcal{M}$ has $2n-1$ dimensions, $\mathbb{P}N^*\mathcal{M}$ has $2n-2$, and so $\mathcal{N}$ has $2n-3$ dimensions. In the standard case, $n = 4$ and $\mathcal{N}$ is five-dimensional.

Once we can guarantee that $\mathcal{N}$ is a manifold, we can look for other geometric structures on it. In fact, much of the geometry of $T^*\mathcal{M}$ descends to $\mathcal{N}$. The canonical one-form θ on $T^*\mathcal{M}$ is defined at $\alpha \in T^*\mathcal{M}$ by: for $v \in T_\alpha(T^*\mathcal{M})$, $\theta_\alpha(v) = \alpha(\pi_*(v))$. Then the symplectic form ω on $T^*\mathcal{M}$ is defined by $\omega = \mathrm{d}\theta$.

If $\mathcal{M}$ has local coordinates $\{q^i\}$, and $T^*\mathcal{M}$ has associated coordinates $\{q^i, n_i\}$, then $\theta = n_i \mathrm{d}q^i$ and $\omega = \mathrm{d}n_i \wedge \mathrm{d}q^i$.

The canonical form is the annihilator of a field of hyperplanes on $T^*\mathcal{M}$ which is called a contact structure, see Appendix 4 of Arnold's classical mechanics text [16] for an exposition of contact geometry. Although the form itself is not preserved by dilatations, the field of hyperplanes is, and it is also preserved by the geodesic flow. Consequently, one obtains a field of hyperplanes on $\mathcal{N}$, and in fact also a contact structure on $\mathcal{N}$. Indeed, there is a one-form on $\mathcal{N}$ whose pull-back to $N^*\mathcal{M}$ is proportional to θ, and hence determines the same distribution of hyperplanes, and such a form is a contact form for $\mathcal{N}$.

One can also consider the space $\mathcal{N}'$ of scaled null geodesics and in this case use ω to obtain a symplectic structure on $\mathcal{N}'$ [17]; we will not make use of that structure here.

So let γ be a point on a smooth curve in $\mathcal{N}$, and let j be the tangent to that curve at γ. What does it mean for j to lie in the contact hyperplane at γ?

The vector j at γ determines a Jacobi field, J, along Γ in $\mathcal{M}$ (up to a multiple of the tangent to Γ). Denoting the tangent covector to Γ by n, we then see that j lies in the contact hyperplane at γ iff $n(J) = 0$ at some (and hence any) point of Γ. In other words, the vector connecting two infinitesimally separated null geodesics in $\mathcal{N}$ lies in the contact hyperplane if and only if the vector connecting points on the null geodesics as curves in $\mathcal{M}$ is orthogonal to the tangent to those null geodesics. We say that such null geodesics are abreast.

Penrose has given a detailed exposition of the meanings of θ and ω in the context of null congruences in space-time [18]; the interested reader is referred to this work for more detail.

Clearly, a two-dimensional submanifold of $\mathcal{N}$ corresponds to a smooth two-dimensional family of null geodesics in $\mathcal{M}$, i.e. a three-dimensional surface (perhaps with singularities) ruled by null geodesics.

Note, in passing, that such a surface need not be a null hypersurface:

Example 3. The surface in four-dimensional Minkowski space (with the usual coordinates) consisting of all points on the null geodesics with tangent vector $(1, 1, 0, 0)$ through the surface given by $t = z = 0$ is the timelike surface $z = 0$.

Also, recall that a surface whose tangents all lie in the hyperplanes of the contact structure is called a Legendre surface [16]. Then we finally have

Theorem 2. *Let Σ be a two-dimensional submanifold of $\mathcal{N}$. Then iff Σ is a Legendre surface in $\mathcal{N}$, the surface $\tilde{\Sigma}$ in $\mathcal{M}$ ruled by the null geodesics of Σ is hypersurface-orthogonal; i.e. $\tilde{\Sigma}$ is an orthogonal null congruence to its intersection with any spacelike three-surface in $\mathcal{M}$.*

Proof. From the above discussion we see that a vector connecting neighbouring points of Σ lies in the contact hyperplane if and only if the vector connecting points of nearby null generators of $\tilde{\Sigma}$ is orthogonal to the tangents to the null generators. Hence the tangent vector to the intersection of a spacelike three-surface with $\tilde{\Sigma}$ is orthogonal to the tangent to any null generator of $\tilde{\Sigma}$, i.e. $\tilde{\Sigma}$ is an orthogonal null congruence to this intersection. □

In particular, if $p \in \mathcal{M}$, then we can find the subset of $\mathcal{N}$ consisting of all null geodesics through p. This subset is the image of the S^2 fibre over p in the bundle of null directions over $\mathcal{M}$, and is itself a smooth S^2 in $\mathcal{N}$, which we denote P; the S^2 in $\mathcal{N}$ corresponding to a point of $\mathcal{M}$ is called the sky of that point. Every sky is a Legendre surface in $\mathcal{N}$; but a Legendre surface need not be a sky.

Even though the space of null geodesics of a strongly causal space-time is naturally a manifold, it may still have pathologies: in particular, it may fail to be Hausdorff.

Example 4. Consider Minkowski space with the usual coordinates (t, x, y, z), and let its cotangent bundle have coordinates $(t, x, y, z, p_t, p_x, p_y, p_z)$. Let $\mathcal{M}$ be Minkowski space minus the origin. Then the sequence of null geodesics given by the covectors $(0, 1/n, 0, 0, 1, 1, 0, 0)$ has as limit points both the null geodesic determined by the covector $(1, 1, 0, 0, 1, 1, 0, 0)$ and that determined by $(-1, -1, 0, 0, 1, 1, 0, 0)$. (In Minkowski space these covectors determine the same null geodesic, which passes through the origin.)

Such a pathology cannot arise in the case where $\mathcal{M}$ is globally hyperbolic. In fact, if $\mathcal{S}$ is a Cauchy surface for $\mathcal{M}$, then $\mathcal{S}$ inherits a Riemannian structure from g, and $\mathcal{N}$ is diffeomorphic to the tangent unit sphere bundle to $\mathcal{S}$ with this Riemannian metric; furthermore, the contact structure on $\mathcal{N}$ agrees with the natural one on the tangent unit sphere bundle of $\mathcal{S}$. In this case, being the tangent unit sphere bundle of a Hausdorff (Riemannian) manifold, $\mathcal{N}$ is automatically Hausdorff.

This immediately tells us that any space-time whose space of null geodesics is not Hausdorff cannot be globally hyperbolic: so in particular, neither

Minkowski space with a point removed, nor the impulsive gravitational plane wave space-time [19] can be globally hyperbolic.

Global hyperbolicity is a sufficient, but not a necessary condition for $\mathcal{N}$ to be Hausdorff. For example, the region of Minkowski space given by $x^2 + y^2 + z^2 < 1$ has a Hausdorff space of null geodesics, but is not globally hyperbolic.

4 Insight into Space-Time

A natural question to ask is how the causal structure of $\mathcal{M}$ is reflected in $\mathcal{N}$.

We will use the convention of taking lower case Latin letters to refer to points of $\mathcal{M}$, and the corresponding upper case letter to represent the sky of a point; also, a lower case Greek letter will represent a point of $\mathcal{N}$, and the corresponding upper case letter the corresponding null geodesic curve in $\mathcal{M}$ (or, more precisely, its image).

The situation is simplest in Minkowski space, where $\mathcal{N}$ is projective null twistor space, excluding the line at infinity, $\mathbb{PN}^I$, which has been extensively studied from the point of view of projective geometry [18]. In this case $\mathcal{N}$ has a great deal of extra structure; in particular, it is naturally a real submanifold of $\mathbb{CP}^3$, and skies are characterised as the holomorphic surfaces with topology S^2.

Then if $x_1, x_2 \in \mathcal{M}$, x_1 and x_2 lie on a common null geodesic (i.e. are null separated) iff $X_1 \cap X_2$ is non-empty; and dually, if $\gamma_1, \gamma_2 \in \mathcal{N}$, then γ_1 and γ_2 lie on a common sky iff $\Gamma_1 \cap \Gamma_2$ is non-empty.

Null separation, then, is neatly described in this picture, and there is a natural duality between points and null geodesics in space-time and points and skies in the space of null geodesics.

But it is not only null separation which can be given an elegant characterization. $\mathbb{PN}^I$ is topologically $\mathbb{R}^3 \times S^2$, and it is possible to define a linking number L for two skies in $\mathbb{PN}^I$. This linking number may be computed in Minkowski space. Given p and q, let $\mathcal{S}$ be a surface of constant time containing p; then the light cone of q, $N(q)$, intersects $\mathcal{S}$ in an S^2 in general. The linking number of Q round P is simply the winding number of $N(q) \cap \mathcal{S}$ round p in $\mathcal{S}$. By the appropriate choice of orientation and sign convention [20], we have $p \in I^{\pm}(q)$ iff $L(P, Q) = \pm 1$.

In more general, curved, space-times, we lose much of the structure of twistor theory: in compensation, new phenomena arise.

As before, we observe that a Jacobi field, J, along the null geodesic Γ arising from a one-parameter family of null geodesics in $\mathcal{M}$ determines a tangent vector, j, at $\gamma \in \mathcal{N}$, and vice versa. Furthermore, J is tangent to Γ at p if and only if $j \in T_\gamma P$. Recall that two points, p and q are conjugate [8] along Γ if and only if there is a non-trivial Jacobi field along Γ which is tangent to Γ at p and q. Because of this, we have:

Theorem 3. *Let $x, y \in \mathcal{M}$ lie on the null geodesic Γ. Then $\gamma \in X \cap Y$. Furthermore, X and Y interesect transversally in γ unless x is conjugate to y along Γ, and in this case the dimension of $T_\gamma X \cap T_\gamma Y$ is the number of linearly independent Jacobi fields along Γ vanishing at both x and y.*

The property of Minkowski space, that points p and q are chronologically related iff P and Q are linked in $\mathbb{PN}^I$, fails in general, even in globally hyperbolic space-times. It still holds if p and q are close together, in the following sense:

Theorem 4. *Let $\mathcal{M}$ be strongly causal, and for $p \in \mathcal{M}$ let U be a causally convex neighbourhood of p (so U is geodesically convex, and causal curves intersect U in a single connected component). Then denoting the space of null geodesics of U by $\mathcal{N}(U)$, chronological relations of points in U are encoded in the linking of their skies in $\mathcal{N}(U)$ in just the same way as in Minkowski space.*

In other words, linking of skies still encodes chronological relations locally; but globally it need not. If $N^+(p)$ has self-intersections, there can be points $q \in I^+(p)$ such that $\mathrm{Link}(P, Q) = 0$. In fact, if we use the equivalence of $\mathcal{N}$ with the unit tangent sphere bundle to $\mathcal{S}$ in a globally hyperbolic space-time with Cauchy surface $\mathcal{S}$, one can have points p and q with $q \in I^+(p)$ but such that P and Q can be simultanously deformed to tangent spheres [21]. If $\mathcal{S}$ is a Cauchy surface containing p, this will occur when $N(q) \cap \mathcal{S}$ has winding number zero round p.

So there is no sense in which topological linking encodes causal relations in general.

At least, not in four (or more) space-time dimensions. If we consider the case of three space-time dimensions, the situation is rather different. As is well known, topology in three dimensions has special properties; in particular, it is possible to have two S^1s embedded in S^3 in such a way that they have linking number 0, but are nevertheless non-trivially linked [22]. One particular example of such a link is the Whitehead link; and the example alluded to above reduces precisely to the Whitehead link if we reduce the number of spatial dimensions by one.

Conjecture 1. Let $\mathcal{M}$ be a globally hyperbolic space-time with two spatial dimensions, and Cauchy surface $\mathcal{S}$ diffeomorphic to $\mathbb{R}^2$, so that $\mathcal{N}$ is diffeomorphic to $\mathbb{R}^2 \times S^1$. Then points p and q are causally related if and only if their skies cannot be simultaneously deformed to unit tangent spheres of $\mathcal{S}$.

A proof of this conjecture in a non-trivial class of space-times has been found [23], although the general case remains elusive.

To return to the more physically interesting case of four-dimensional space-time, we see that the problem is caused by light cones developing self-intersections. In fact, the apparatus we now have available provides a powerful tool for investigating just how the light cone can develop in a general space-time. This provides information on the types of caustic that can arise due to gravitational lensing.

First, we recall that, by Theorem 2, a Legendre surface Σ in $\mathcal{N}$ corresponds to a hypersurface-orthogonal null hypersurface $\tilde{\Sigma}$ in $\mathcal{M}$; but saying that $\tilde{\Sigma}$ is hypersurface-orthogonal is just saying that it is the wave front obtained by instantaneously lighting up an initial space-like two-surface and tracing out the resulting light rays. So Legendre surfaces of $\mathcal{M}$ correspond to wave fronts in $\mathcal{M}$. Furthermore, if $\mathcal{S}$ is a Cauchy surface of $\mathcal{M}$, and we consider $\mathcal{N}$ as the unit tangent sphere bundle to $\mathcal{S}$, then the natural projection of Σ to $\mathcal{S}$ gives the intersection of $\tilde{\Sigma}$ with $\mathcal{S}$; this projection is a Legendre map. We can then deduce from the properties of such mappings presented by Arnold [16] that the only singularities which are present in the intersection of a wave front with a Cauchy surface and stable under small perturbations are those of type A_2 or D_4. In particular, these are the only stable singularities which appear on a light cone at a given instant.

One can alternatively carry out an analysis in terms of the full structure of the cotangent bundle of space-time [24]. Such an analysis allows for the investigation of other properties of wave front evolution, but at the expense of requiring far more technical machinery. An extensive review of gravitational lensing and wave front evolution has now been provided by Perlick [25].

5 Recovering Space-Time

We have considered how points of $\mathcal{M}$ are represented in $\mathcal{N}$ as skies, and seen how this gives a new approach to considerations of the causal structure of $\mathcal{M}$. A natural question to ask is whether we can use such a setting to define a space-time as a family of skies in a suitable five-manifold; and in such a setting, to look for characterizations of a space-time being conformal to an Einstein vacuum, or to a space-time satisfying a standard energy condition.

Unfortunately, such a characterization is not (yet) available. Indeed, there may be problems even with recovering the original space-time if we know all the skies in $\mathcal{N}$.

Example 5. Let $\mathcal{M}$ be the Einstein static universe. Then the space of null geodesics is precisely that of compactified and identified Minkowski space [18], and all the skies of points in $\mathcal{M}$ are skies of points of compactified, identified Minkowski space.

The phenomenon at the heart of this problem is that the null cone of a point in the Einstein static universe converges back to a point again; so distinct points can have the same sky. Clearly, this is a bad thing from the point of view of regarding a space-time as arising from a set of skies.

In fact, even if the null cone does not converge back exactly to a point, there may still be a problem. For suppose p is a point of $\mathcal{M}$, and p_n is a sequence of points which remain strictly outside some neighbourhood K of p, but have the property that if U is a neighbourhood of p then for n sufficiently large, all the null geodesics through p_n pass through U. We call a space-time exhibiting this behaviour a *refocussing* space-time.

Then in $\mathcal{N}$, no matter how small a neighbourhood V of P we choose, there will be infinitely many P_n lying inside V. In this case, even if all the points of $\mathcal{M}$ have distinct skies, $\mathcal{N}$ cannot provide $\mathcal{M}$ with the correct topology.

Fortunately, this phenomenon cannot occur in a large class of space-times of interest.

Theorem 5. *Let $\mathcal{M}$ be globally hyperbolic, with non-compact Cauchy surface $\mathcal{S}$. Then $\mathcal{M}$ cannot be refocussing.*

Proof. In brief outline: We can suppose without loss of generality that each $p_n \in I^+(p)$. Now, if the entire light cone of p_n focusses back into a small neighbourhood of p, then all null geodesics through p_n must meet a conjugate point within some finite time. In this case, it follows that all of $\mathcal{S}$ lies in $J^-(p_n)$, which is impossible because $J^-(p_n) \cap \mathcal{S}$ must always be compact. □

If we are given the set of all skies in the space of null geodesics, $\mathcal{N}$, of a globally hyperbolic space-time, $\mathcal{M}$, with non-compact Cauchysurface, $\mathcal{S}$, then we can reconstruct the original space-time up to a conformal factor.

First, we have $\mathcal{M}$ simply as the point set of all skies in $\mathcal{N}$.

Next, if P is a sky in $\mathcal{N}$, we take a neighbourhood U of P, sufficiently small, that any two skies in P intersect transversally. Then the set of all skies lying in U give a neighbourhood, V, of p. Neighbourhoods constructed in such a way give a basis for the topology on $\mathcal{M}$.

Recovering the differentiable structure of this neighbourhood is a little more involved. First, we construct the Grassmannian bundle of two-planes over U; then we lift each sky in U to this bundle by lifting $\gamma \in Q$ to $T_\gamma Q$. This gives a six-dimensional submanifold $\tilde{U}$ of our Grassmannian manifold which is (because of the smoothness of the geodesic flow and the absence of non-transversal intersections) diffeomorphic to the bundle of null directions over V. Each sky lifts to the fibre over a point of V, and so taking the quotient manifold of lifted skies (fibres here are compact, so the distribution is automatically regular) gives us back V as a differentiable manifold.

It is now a simple matter to find the metric, g, up to a conformal factor. Given a point $\gamma \in U$, we obtain a curve in $\tilde{U}$ given by the tangent planes to all skies in U containing γ. This projects to a curve in V under the quotient above, namely Γ. But once we know all the null geodesics in V, we have the null cone at each point, and as is well known [8] this determines the metric up to a conformal factor.

We can then recover $\mathcal{M}$ by constructing a neighbourhood of each point by this means, and using the overlap maps induced by the intersections of neighbourhoods of skies in $\mathcal{N}$.

Alternatively, we can make use of an alternative topology on $\mathcal{M}$ which was observed by Hawking et al. [26] to capture all of the relevant structure of $\mathcal{M}$.

Again, we begin with a $\mathcal{M}$ as the point set of skies in $\mathcal{N}$, but no other structure, and consider a small neighbourhood U of the sky P in $\mathcal{M}$. This time, we take as a neighbourhood of $p \in \mathcal{M}$ all those skies in U which are nontrivially linked to P. This gives a basis for the topology on $\mathcal{M}$ generated by neighbourhoods of the form $(V \cap I(p)) \cap \{p\}$, where V is a neighbourhood of p in the original manifold topology, and I is determined by the chronology relation of the original Lorentz metric on $\mathcal{M}$. This topology, called the path topology, uniquely determines the causal, differential, and conformal structure of $\mathcal{M}$.

Thus, at least in the case of non-refocussing space-times, it is possible to regard space-time as given by a structure (namely the set of skies) in the space of null geodesics. In the final section we will consider an approach to providing null geodesics with endpoints, in an attempt to provide space-time with a causal boundary.

6 A (New?) Causal Boundary

Let $\mathcal{M}$ be a strongly causal space-time, and let γ be a null geodesic of $\mathcal{M}$. How can we attach a future endpoint to Γ? The idea behind what will be done here is to find all null geodesics which focus at the same point at infinity, and treat this set of null geodesics as the light cone of the (common) future endpoint of these null geodesics.

To this end, we let s be an affine parameter for Γ, so that $p(s)$ traces out Γ as s ranges from $-\infty$ to ∞. (By an appropriate choice of conformal factor, we can assume that all null geodesics have affine parameters with this range of values [27].) Then as s increases, $T_\gamma P(s)$ traces out a curve in the Grassmannian manifold of two-dimensional subspaces of $T_\gamma\mathcal{N}$. Since Grassmannian manifolds are compact, this curve has a limit point as $s \to \infty$.

This limiting two-plane is supposed to be the tangent plane to the sky of the future endpoint of Γ. However, it need not be unique. One would expect that $T_\gamma P(s)$ would settle down if curvature decayed along Γ; ascertaining appropriate conditions for uniqueness is the subject of current work. Thus uniqueness should be related to the asymptotic flatness of the (perhaps conformally rescaled) space-time.

Then at worst, we have some subset - denote it by B_γ - of $T_\gamma\mathcal{N}$ for each $\gamma \in \mathcal{N}$. In general, this need not be a distribution; its dimension may vary from point to point, and it need not be continuous. Now regard γ_1 and γ_2 as equivalent if they can be connected by a curve whose tangent everywhere lies in some B_γ. By definition, null geodesics which are equivalent under this relationship focus to a common future endpoint.

Then we obtain a topological space of future endpoints to the null geodesics of $\mathcal{M}$ by taking the quotient space under this relation: call this topological space B^+ (for future boundary). An equivalence class (point at infinity) will be denoted by B when we are thinking of it as a subset of $\mathcal{N}$, and by b when we think of it as a point of the boundary of $\mathcal{M}$. What we still lack is a topology for $\mathcal{M} \cup B^+$; so how do we decide if a sequence of points p_n in $\mathcal{M}$ converges to a point, b, of B^+?

We require two conditions on the sequence of points p_n to say that $p_n \to b$. First, p_n should eventually leave any compact set in $\mathcal{M}$; and second, the light cone of p_n should approach $N^-(b)$ (which is, by definition, the set of null geodesics defining b). To make this latter condition precise, we require that there exists $\gamma \in B$ and $\gamma_n \in P_n$ such that $\gamma_n \to \gamma$, and the limit set of T_{γ_n} lies in B_γ as $n \to \infty$.

This certainly works well in certain simple cases.

If $\mathcal{M}$ is the Einstein static cylinder, then all null geodesics share a single future endpoint, which lies to the future of every point of $\mathcal{M}$.

More generally, if $\mathcal{M}$ can be conformally embedded into a strongly causal space-time as a subspace with compact closure, then each null geodesic will acquire a future endpoint. Furthermore, each of the equivalence classes defined above will be a subset of the sky of such a future endpoint. If we further require that $\mathcal{M}$ be globally hyperbolic, then the equivalence classes will coincide with the skies of endpoints, and the boundary points are precisely those of the usual causal boundary [10].

So this attempt at constructing a boundary is well behaved in certain simple cases where we have a good idea of what the boundary “ought” to be. Furthermore, properties of space-time which complicate the intuitive notion of boundary also complicate this construction.

Note that one can similarly attempt to add a past endpoint to each null geodesic, and thereby construct a past boundary, B^-. However, the usual problem of identification of appropriate points in the past and future boundary remains.

In conclusion, I will list some of the questions which arise in the context of this proposal:

1. How is this boundary related to the Geroch, Kronheimer and Penrose boundary?
2. For example, if null geodesics lie in the same equivalence class, need they have the same chronological past in general?
3. How is this boundary related to the Geroch boundary [28]? Is his equivalence relationship strictly weaker than this one, vice versa, or are they incomparable?
4. Does good/bad behaviour of the limiting “distribution” match to good/bad asymptotic properties of the original space-time?
5. In particular, can we gain any insight into space-time in the vicinity of a strong curvature singularity from this point of view?

References

1. C. Will: *Theory and Experiment in Gravitational Physics* (Cambridge University Press, Cambridge 1993)
2. `http://www.gravityprobeb.com/`
3. `http://www.ligo.caltech.edu/`
4. K. Gödel: An example of a new type of cosmological solution of Einstein's field equations of gravitation. Rev. Mod. Phys. **21**, 447–450 (1943)
5. J. Earman: *Bangs, Crunches, Whimpers and Shrieks: Singularities and Acausalities in Relativistic Spacetimes* (Oxford University Press, Oxford 1995)
6. R. Penrose: Singularities and time asymmetry. In: *General Relativity: An Einstein Centenary Survey*, ed by S.W. Hawking, W. Israel (Cambridge University Press, Cambridge 1979)
7. R. Penrose: *Techniques of Differential Topology in Relativity*, Regional Conference Series in Applied Math. **7** (Society for Industrial and Applied Mathematics, Philadelphia 1972)
8. S.W. Hawking, G.F.R. Ellis: *The Large Scale Structure of Space-Time* (Cambridge University Press, Cambridge 1973)
9. J.K. Beem, P.E. Ehrlich: *Global Lorentzian Geometry* (Marcel Dekker, New York 1981)
10. R.P. Geroch, E.H. Kronheimer, R. Penrose: Ideal points of space-time. Proc. Roy. Soc. London A **327**, 545–567 (1972)
11. R.D. Sorkin: Forks in the road, on the way to quantum gravity. Int. J. Theor. Phys. **36**, 2759–2781 (1997)
12. J.F. Cariñena, C. López: Symplectic structure on the set of geodesics of a Riemannian manifold. Int. J. Modern Phys. A **6**, 431–444 (1991)
13. J.K. Beem, R.J. Low, P.E. Parker: Spaces of geodesics: Products, coverings, connectedness. Geometriae Dedicata **59**, 51–64 (1996)
14. B. Carter: Causal structure in space-time. Gen. Rel. Grav. **1**, 349–391 (1971)
15. F. Brickell, R.S. Clark: *Differentiable Manifolds: An Introduction* (Van Nostrand Reinhold, London 1970)
16. V.I. Arnold: *Mathematical Methods of Classical Mechanics*, 2nd edn (Springer, New York 1991)
17. R. Penrose: On the nature of quantum geometry. In: *Magic Without Magic: J. A. Wheeler Festschrift*, ed by J. R. Klauder (Freeman, New York 1972)
18. R. Penrose, W. Rindler: Spinors and space-time. Vol 2: Spinor and Twistor Methods in space-time Geometry (Cambridge University Press, Cambridge 1986)
19. R. Penrose: A remarkable property of plane waves in general relativity. Rev. Mod. Phys. **37**, 215–220 (1965)
20. R.J. Low: Twistor linking and causal relations. Class. Quantum Grav. **7**, 177–187 (1990)
21. R.J. Low: Causal relations and spaces of null geodesics. D. Phil. Thesis, Mathematical Institute, Oxford University (1988)
22. D. Rolfsen: *Knots and Links* (AMS, Chelsea 2003)
23. J. Natário, K. P. Tod: Linking, Legendrian linking and causality. Proc. London Math. Soc. **88**, 251–272 (2004)
24. W. Hasse, M. Kriele, V. Perlick: Caustics of wavefronts in general relativity. Class. Quantum Grav. **13**, 1161–1182 (1996)

25. V. Perlick: *Gravitational Lensing from a Spacetime Perspective*, Living Rev. Relativity **7** (2004), 9. `http://www.livingreviews.org/lrr-2004-9`
26. S.W. Hawking, A.R. King, P.T. McCarthy: A new topology for space-time which incorporates the causal, differential and conformal structures. J. Math. Phys. **17**, 174–181 (1976)
27. J.K. Beem: Conformal changes and geodesic completeness. Comm. Math. Phys. **49**, 179–186 (1976)
28. R. Geroch: Local characterization of singularities in general relativity. J. Math. Phys. **9**, 450–465 (1967)