DRAFT COPY

Mathematical General Relativity

Evolutionary and Causal Aspects

Willie Wai-Yeung Wong

Revision based on commit 69500cd as of April 10, 2024

DRAFT COPY

Preface

These are lecture notes for a mini-course I am teaching at the National Center for Theoretical Sciences (NCTS) in Taipei, Taiwan. The intended audience are advanced undergraduates, masters students, and PhD candidates in their first two years. The participants are expected to have prior exposure to differential and Riemannian geometry at the advanced undergraduate or beginning graduate level. In particular, they should be comfortable with the concept of tensors, know the distinction between contravariance and covariance, and can carry out basic differential calculus computations using the Levi-Civita connection both in invariant/index notation and in local coordinates. These expectations are reflected in the contents of these notes, which starts from a much more advanced place mathematically compared to standard textbooks on general relativity. Students meeting the basic background outlined above should find the bulk of these notes quite approachable. A number of paragraphs and sections are however marked with the symbol "†"; these sections are slightly farther afield from the main thread of discussion, and may involve more advanced concepts. They can often be skipped without detracting too much from the course.

The selection of topics reflect my research interest and specialty. In particular, very little in these notes concern the quite interesting Riemannian geometric analysis problems that arise from general relativity. Instead, the purpose of these notes are to expose students to techniques and results that are firmly *Lorentzian* geometric in nature, which one must contend with if one were to study the theory in the dynamical setting.

In preparing these notes, I drew heavily from Choquet-Bruhat, *General relativity and the Einstein equations* on general relativity; readers are encouraged to consult it as a fairly modern and exhaustive reference on the subject matter. For more geometric topics I owe much to O'Neill, *Semi-Riemannian geometry: with applications to relativity*; for more advanced developments Beem, Ehrlich, and Easley, *Global Lorentzian geometry* is highly recommended. Other, more precise, references are included inline to specific results in that book or to other works in the literature.

I must, at this point, express my gratitude to the Mathematics Division at the NCTS, for so graciously hosting me during my sabbatical year. In addition to the wonderful staff at the NCTS (especially Murphy Yu, who helped significantly with the organization of the mini-course), I need to thank also Professors Ye-Kai Wang, Mao-Pei Tsui, and Yng-Ing Lee for helping to organize my visit, and for giving me the strong encouragement to offer the mini-course for which these notes are being written.

DRAFT COPY

Willie WY Wong
2023.10.1, Taipei.

Preface for the second semester

Thanks to the encouragement of Prof. Yng-Ing Lee and the immense help of Ms. Murphy Yu, a second semester of the course was offered. The intended audience are still students that are "mathematically younger", though in practice several of the participants in the course are further along in their respective mathematical careers. I hope I succeeded in not letting their excellent questions drag the course too far from my introductory intentions. For the second semester course students should be comfortable performing geometric computations using both invariant notations and abstract index notations, and have a basic command with concepts from differential topology. Additionally, students should have a reasonable familiarity with advanced multivariable calculus; this reflects the change of focus between the first and second semesters.

Whereas the first semester focused almost exclusively on matters of *Lorentzian geometry*, the second semester targets the *analytic* aspects of Einstein's equations. The underlying theme is the initial value problem for Einstein's equations, through which various wave equations techniques are introduced and applied. The opus Choquet-Bruhat, *General relativity and the Einstein equations* is again invaluable as a reference, as is Ringström, *The Cauchy Problem in General Relativity*.

Willie WY Wong
2024.4.8, Taipei.

© Willie Wai-Yeung Wong

DRAFT COPY

Contents

DRAFT COPY

Part I

First Semester

DRAFT COPY

DRAFT COPY

Topic 1 (2023/10/16)

What is General Relativity?

1.1 "General relativity" is a theory of gravity. In many popular accounts and in the minds of many mathematicians, it is equated to Einstein's equations

$$\mathrm{Ric} - \frac{1}{2}\mathrm{R}g = T.$$

While the association is indisputable, it does not capture fully Einstein's contributions. Roughly speaking, the formulation of general relativity arises from, or perhaps is reliant upon, three physical postulates, which we list below in order of specificity.

- The principle of general covariance;
- the weak equivalence principle; and
- Einstein's equations.

The first is a general *philosophy* on how physical laws ought to be represented mathematically, and arguably predates Einstein (see discussion below). The second is a statement on the *nature* of gravity, insofar as how it acts on particles. The third is the law that informs us on how gravity changes through the behavior of particles. Each builds upon the previous, and as we shall see below, one can accept the lower levels of the theory while replacing the higher levels with something else entirely.

1.2 **(†Aside: Maxwell's theory)** A similar hierarchy exists in Maxwell's theory of electromagnetism. At a lower level, Maxwell postulates that the electromagnetic field is given by a two-form F on Minkowski space satisfying $\mathrm{d}F = 0$ (this is the *Faraday–Maxwell law*). Additionally, he proposed that there exists an "electromagnetic displacement two-form" M which satisfies $\mathrm{d}M = J$, where J is the current density (this is the *Ampére–Coulomb–Maxwell law*). These two ideas form a general *framework* for electromagnetism. To connect the electromagnetic field (which influences how charged particles move) and the displacement (which is the influence left by charged particles) requires a specific *law of æther*: Maxwell proposed the linear relation $M = *F$ where $*$ is the Hodge operator. In modern physics alternative perspectives have been given that reject Maxwell's linear law of æther, replacing it by other, potentially nonlinear models, all the while retaining the general Maxwellian framework. For a more detailed account see Kiessling, "Electromagnetic field theory without divergence problems. I. The Born legacy".

DRAFT COPY

1.3 (General covariance) In broad strokes, the principle of general covariance codifies the general belief that there is an objective reality that is independent of the observers' frames of reference, and extends that to state that a *mathematical model of physics should be expressible in frame-independent formulae*. This is an extension of what Truesdell and Noll calls the "principle of material frame indifference"; explicit discussion of this principle can be traced at least to the work of Zaremba, Jaumann, and separately Cosserat in the first decade of the 20th century, but one can argue that implicitly this principle goes back to Hooke and earlier. To have faith that the scientific method is valid and reproducible requires one to believe that different observers performing the same experiment should obtain, up to experimental error, the same result.

Truesdell and Noll, The non-linear field theories of mechanics

The lasting impact of this principle is the geometrization of modern mathematical physics. Accepting Newton's dictum *"Data aequatione quotcunque fluentes quantitates involvente, fluxiones invenire; et vice versa;"* we would ask our model of physics to be one where differential calculus can be applied. The natural setting in which this can be done in an observer-independent (coordinate-independent) way is the differential geometry of vector bundles equipped with connections.

Philosophically, one is also driven to accept the converse: that if two physical processes are indistinguishable by frame-independent quantities, then they should represent the same underlying physical phenomenon. This has implications when it comes to the next postulate.

There is one other important consequence of general covariance, which concerns how physics in the absence of gravity can be lifted to physics in the presence of gravity; this requires talking about local models of space-time and we will omit the rather technical discussion here.

1.4 (Weak equivalence) The idea of the weak equivalence principle is captured in Einstein's elevator *Gedankenexperiment*: "given an observer trapped in an elevator, can he, through conducting experiments and taking measurements only within the elevator, determine if the elevator is floating outside the influence of any external gravitational field, or if the elevator is free-falling within an external gravitational field?"

The conclusion (or rather, supposition) that the observer *cannot* distinguish between the two scenario, together with the principle of general covariance, leads to the postulate that the gravitational field itself is *not* a material quantity. This is in direct contrast with Newton's theory, where the gravitational field is expressed as a vector field that acts on particles as a force.

1.5 (A framework for gravity) The combination of the principles of general covariance and weak equivalence leads to, after a small leap of faith, the following mathematical model for how gravitational forces act on particles. Accepting that gravity should dictate how test particles move, yet be not subject to an invariant description of a tensor, leads naturally to the modeling of gravity using an affine connection on the tangent bundle. More precisely, we shall postulate that the space-time is modeled after

This general framework was first described in Cartan, "Sur les variétés à connexion affine et la théorie de la relativité généralisée (première partie)".

$$M \qquad \text{— a smooth manifold.} \tag{1.5.1}$$

The effect of gravitation is captured by

$$\nabla \qquad \text{— an affine connection on } TM. \tag{1.5.2}$$

And finally, *Newton's First Law* be modified to state that "particles will travel along geodesics of ∇ unless subjected to (non gravitational) external forces."

© Willie Wai-Yeung Wong

1.6 (Lorentzian geometry and Einstein's equation) To complete the description of what gravity is, one needs to specify how massive particles in space-time will generate their "gravitational field". Toward this, Einstein made a further choice in setting the affine connection used for modeling gravity to be the Levi-Civita connection given by a Lorentzian metric. This is based on his earlier work on special relativity, where *in the absence of gravity* the space-time is taken to be Minkowski space. One should note that this choice is not required by the framework described in ¶ 1.5; see also ¶ 1.7.

The mechanical laws of special relativity associates to a distribution of moving particles a corresponding "energy–momentum tensor" T; an analogous quantity can be defined in the gravitational setting. (Recall that a crowning achievement of special relativity is that the "mass" of a particle is *not* frame-independent, and that the frame-independent quantity is in fact the energy–momentum four-vector.) Observationally the gravitational influence generated by a particle is determined by its energy–momentum. To settle on a candidate for the law that describes how particles generate their gravitational field, Einstein was then led to look for a tensor quantity

- that is an invariant (or natural geometric) quantity associated to the Lorentzian metric of the space-time;
- that has the same basic algebraic and differential properties of T (specifically, T is a symmetric two-tensor, and in the special relativistic setting T should be divergence free);
- that reduces to the well-tested Newtonian gravity in a suitable limit.

The candidate that Einstein (together with Marcel Grossman) settled on was the *Einstein tensor*

$$\mathrm{Einst} := \mathrm{Ric} - \frac{1}{2}\mathrm{R}g.$$

Note that as a result of the second Bianchi identities, G is automatically divergence free. Trivially, of course, the metric g itself also satisfies the first two properties, and has no impact on the third. And hence we are led to *Einstein's equations*

$$\mathrm{Einst} + \Lambda g = T.$$

Before diving more into general relativity proper, we will take a detour to see how Newton's theory can be made to fit within the framework of ¶ 1.5, even though Newton's formulation, on the surface, does not appear to be compatible with the principle of weak equivalence.

We shall not launch into a full discussion of Newton–Cartan Theory, for two reasons. First is that to appreciate the full theory requires a higher level of sophistication with differential geometry than I am willing to assume of students for this course, and second is that the theory is surprisingly rigid, and the bulk of the interesting behavior is already captured in the abbreviated discussion here. For a more complete discussion of the geometric features, one can consult Cartan, "Sur les variétés à connexion affine et la théorie de la relativité généralisée (première partie)". For a more physics-oriented presentation, §12.1 of Misner, Thorne, and Wheeler, Gravitation *is a good reference.*

Newtonian Theory Geometrized

1.7 (Newtonian space-time) Our usual conception of the Newtonian universe has the underlying manifold $\mathbb{R}\times\mathbb{R}^3$, with coordinates (t, x^1, x^2, x^3). Newton's theory of gravity essentially states that the gravitational acceleration experienced by a particle is given by the gradient of a gravitational potential, which itself is the solution to Poisson's equation:

$$\ddot{\gamma} + \nabla U = 0, \qquad U = \triangle^{-1}\rho. \tag{1.7.1}$$

In the preceding equation:

- ρ is the mass distribution generating the gravitational field;

DRAFT COPY

- $\gamma = (\gamma^t, \gamma^1, \gamma^2, \gamma^3) : \mathbb{R} \to \mathbb{R} \times \mathbb{R}^3$ is the particle trajectory, and we normalize the parametrization so that $\dot\gamma^t = 1$.

The equation of motion can be written in the form of the geodesic equation for a connection $\nabla = \partial + \Gamma$ (here Γ is the Christoffel symbols), given by

$$\Gamma^t_{\mu\nu} = \Gamma^\nu_{\mu i} = 0, \qquad \Gamma^i_{tt} = \partial_i U. \tag{1.7.2}$$

(Here $\mu, \nu \in \{t, 1, 2, 3\}$ and $i \in \{1, 2, 3\}$.) Of course, the Christoffel symbols are coordinate-dependent; we therefore seek to express these conditions invariantly. One way to do so is to observe that the corresponding Riemann curvature tensor is given by

$$\mathrm{Riem}_{\mu\nu\lambda}{}^\rho = -\partial_\mu \Gamma^\rho_{\nu\lambda} + \partial_\nu \Gamma^\rho_{\mu\lambda} - \Gamma^\rho_{\mu\kappa}\Gamma^\kappa_{\nu\lambda} + \Gamma^\rho_{\nu\kappa}\Gamma^\kappa_{\mu\lambda};$$

observe then the quadratic terms automatically vanish, and the derivative terms is only non-zero when exactly two out of μ, ν, λ is equal to t, for which we find

$$\mathrm{Riem}_{itt}{}^j = -\partial_i \partial_j U. \tag{1.7.3}$$

And so we may characterize Newtonian theory as requiring

$$\mathrm{Ric}_{tt} = \rho \text{ and } \mathrm{Ric}_{i\mu} = 0. \tag{1.7.4}$$

1.8 For (1.7.4) to be truly invariant, we need to remove the dependence on the standard frame. One way to do so is to postulate certain underlying structure to the space-time. We may assume that our space-time M is a four-dimensional manifold equipped with a function $\tau : M \to \mathbb{R}$. This is expected as Newtonian gravity is developed under *Galilean relativity* where a universal notion of time is assumed to exist. We assume further that $\tau^{-1}(\{t\})$ is a three-dimensional manifold that can be identified with $\mathbb{R}^3$, in particular it carries a flat Riemannian metric δ; we will denote by h the pushforward of the inverse metric from $\tau^{-1}(\{t\})$ to M. The requirement that $\tau^{-1}(\{t\})$, the hypersurfaces of instantaneity, carry Riemannian structures is natural in view of Newton's law of gravitation which involve a Poisson equation; the Laplace-Beltrami operator requires a Riemannian structure to define. In the discussion above we often freely lowered and raised indices, see e.g. (1.7.3). That h is flat can actually be derived as a consequence our assumptions below, but we omit that discussion here.

It turns out that the full Newtonian theory of gravity can be recovered from the following: we look for a linear, torsion-free connection ∇ on M such that:

- $\nabla(\mathrm{d}\tau) = 0$ and $\nabla h = 0$ (preserves space-time structures);
- The Ricci curvature of ∇ satisfies $\mathrm{Ric} = \rho\, \mathrm{d}\tau \otimes \mathrm{d}\tau$; here ρ is the mass density of the matter distribution;
- The Riemann curvature of ∇ satisfies the following: for any pair of vectors V, W, the $(2,0)$-tensor given in index notation as $h^{as}\mathrm{Riem}_{svw}{}^b V^v W^w$ is symmetric in a, b.

The second requirement generalizes (1.7.4); the third requirement is a version of (1.7.3).

1.9 (†Sketch of proof) First, the fact that $\nabla(\mathrm{d}\tau) = 0$ implies that $\Sigma_t := \tau^{-1}(\{t\})$ is totally geodesic for every t, and hence ∇, restricted to Σ_t, is equal to the Levi-Civita connection of h. Now choose $\gamma : (a, b) \to M$ a curve satisfying $\tau(\gamma(t)) = t$; choose further that γ to

DRAFT COPY

be geodesic. Fix e_1, e_2, e_3 an orthonormal frame of Σ_t, first by defining it at $\gamma(c)$ for some $c \in (a,b)$, and then parallelly transporting it along γ. Since Σ_t is flat, these extend (by exponential map) to a coordinate system for $\tau^{-1}((a,b))$, identifying it with $(a,b)\times\mathbb{R}^3$. We have $h = \partial_1\otimes\partial_1 + \partial_2\otimes\partial_2 + \partial_3\otimes\partial_3$ in this coordinate system. The Christoffel symbols of ∇ in this coordinate system can then be checked to satisfy $\Gamma^t_{\mu\nu} = 0$ and $\Gamma^i_{jk} = 0$ for $i,j,k \in \{1,2,3\}$. Furthermore, the Christoffel symbols vanish along $(a,b)\times\{0\}$ by construction.

Now, the fact that $\nabla_{\partial_t} h = 0$ implies $\Gamma^j_{ti} + \Gamma^i_{tj} = 0$. Computing the Riemann curvature, we see that $\mathrm{Riem}_{itj}{}^k = -\partial_i\Gamma^k_{tj}$. The symmetry property of the Riemann curvature implies that $\partial_i\Gamma^k_{tj} = \partial_j\Gamma^k_{ti}$, which in turn implies that for each t there exists three functions f^1, f^2, f^3 such that $\Gamma^k_{tj} = \partial_j f^k$. That $\partial_j f^k + \partial_k f^j = 0$ implies $f^k\partial_k$ is Killing, and hence $\partial_j f^k$ is constant (since the Killing fields of $\mathbb{R}^3$ are fully classified). Using that $\Gamma^k_{tj} = 0$ on γ, we conclude that $\Gamma^k_{tj} = 0$ also in this coordinate system, and hence also $\mathrm{Riem}_{itj}{}^k$.

A similar argument shows that as a result, $\mathrm{Riem}_{itt}{}^j$ reduces to be $-\partial_i\Gamma^j_{tt}$; the symmetry property implies then Γ^j_{tt} can be written as a gradient. And the Ricci condition is now exactly Newton's law of gravity.

Lorentzian Geometry, a Summary

1.10 I assume readers are familiar with basic definitions of smooth manifolds.

Assumption
For these notes, unless otherwise stated, *all manifolds will be*

finite dimensional, Hausdorff, paracompact, connected, and orientable.

◊

1.11 Returning to Einstein's theory, let us now officially introduce the object used to model space-time.

Definition (Lorentzian manifold)
A Lorentzian manifold *is a pair (M,g), where M is a smooth manifold and g is a smooth section of $T^{0,2}M$ with the properties that*

- *g is symmetric, i.e. $g(X,Y) = g(Y,X)$;*
- *g is non-degenerate, i.e. X is such that $g(X,Z) = 0$ for all Z if and only if $X = 0$;*
- *g has signature $(-++\ldots+)$.*

The tensor field g is the Lorentzian metric *on M.* ◊

Finite smoothness versions of this definition are also possible, but we will not consider them here.

Some authors use the opposite sign convention $(+--\ldots-)$.

As g is non-degenerate, it induces an *inverse metric* g^{-1} that is a section of $T^{2,0}M$, also with signature $(-++\ldots+)$, acting on covectors. In index notation we will often suppress the notation for the inverse, and write g_{ab} for the metric tensor and g^{ab} for the inverse metric tensor.

DRAFT COPY

DRAFT COPY

1.12 (Index notation) In general, we will adopt the abstract index notation. A type (p, q) tensor will be given p superscript indices and q subscript indices. Recalling the definition of a tensor as a multi-linear function, each index corresponds to one "input slot"; this is particularly useful when taking linear combinations of tensors. A "term" comprising multiple tensors juxtaposed, e.g. $A^{ab}{}_{cd}B^{df}$, is interpreted as a tensor product. In a term, the same symbol appearing once in the superscript and once in the subscript will be used to indicate tensor contraction between those two slots. Different terms within the same formula should, after applying contraction of repeated indices, have the same numbers of indices with the same sets of (uncontracted) symbols used.

1.13 (Musical isomorphism) Since g is non-degenerate, the mapping $X \mapsto g(-, X)$ gives an isomorphism between TM and T^*M. We call the operation sending an element of TM to an element of T^*M "index lowering" after its effect on the index notation. In invariant notation we also use the musical notation $X \mapsto X^\flat$. Similarly, the operation sending an element of T^*M to an element of TM is "index raising" and is denoted invariantly by $\omega \mapsto \omega^\sharp$. For higher-rank tensors we will generally eschew the invariant notation and use index notation.

1.14 An important use of the musical isomorphisms is the construction of the *gradient* of a function.

Definition
On a Lorentzian manifold (M, g), *when given a function* $f : M \to \mathbb{R}$, *its* gradient *is the vector field* $(\mathrm{d}f)^\sharp$. ◊

1.15 The main difference between Riemannian and Lorentzian geometry then is the minus sign in the signature, which for Riemannian manifolds is $(+ + \ldots +)$, so that the metric is positive definite. The possibility of a different sign allows us to classify tangent vectors X based on the sign of $g(X, X)$.

Definition (Causal adjectives)
On a Lorentzian manifold, a tangent vector $X \in T_pM$ *is said to be*
timelike *if* $g(X, X) < 0$;
null / lightlike *if* $g(X, X) = 0$;
spacelike *if* $g(X, X) > 0$;
causal *if* $g(X, X) \leq 0$ *and* $X \neq 0$.
For covectors $\omega \in T_pM$, *we define the causal adjectives similarly using the inverse metric* g^{-1} *instead. Finally, a codimension 1 subspace* $V \subseteq T_pM$ *is said to be*
timelike or co-spacelike *if its conormal covector* ω *is space-like;*
spacelike or co-timelike *if its conormal covector* ω *is time-like;*
null *if its conormal covector* ω *is null.*
(Recall that the conormal covector of a codimension 1 subspace V *is the nonzero covector* ω, *unique up to scaling, satisfying* $\omega(X) = 0$ *for every* $X \in V$.*)* ◊

Some authors admit 0 *as a causal vector.*

DRAFT COPY

1.16 Example
Given a Lorentzian manifold (M, g) and a point $p \in M$, by elementary multilinear algebra there always exists a basis $\{e_0, e_1, \dots, e_n\}$ of T_pM with respect to which g is diagonal:

$$g(e_0, e_0) = -1, \quad g(e_1, e_1) = g(e_2, e_2) = \cdots = g(e_n, e_n) = 1, \quad g(e_i, e_k) = 0 \text{ if } i \neq k.$$

Every vector $X \in T_pM$ can be written as

$$X = \sum_{j=0}^{n} X_j e_j$$

with $X_j \in \mathbb{R}$. We see then

- X is timelike if and only if $|X_0| > \sqrt{\sum_{j=1}^{n} |X_j|^2}$;
- X is null if and only if $|X_0| = \sqrt{\sum_{j=1}^{n} |X_j|^2}$;
- X is spacelike if and only if $|X_0| < \sqrt{\sum_{j=1}^{n} |X_j|^2}$.

In particular, the set of all null vectors defines a bi-cone $\mathcal{C}$ in T_pM, with the "interior" being timelike vectors and the "exterior" being spacelike ones. ◊

1.17 Exercise
Let (M, g) be a Lorentzian manifold with total space-time dimension $(1 + n)$, with n (the spatial dimension) at least 1. Fix $p \in M$. Let $\mathcal{C} \subseteq T_pM$ be the set of null vectors at p.

1. Prove that when $n = 1$, the set $T_pM \setminus \mathcal{C}$ has four connection components, two of them comprising timelike vectors and two of them spacelike vectors.
2. Prove that, when $n \geq 2$, the set of spacelike vectors is a connected subset of T_pM, while the set of timelike vectors has two connected components.
3. Prove that if X and Y are two causal vectors, then $g(X, Y) = 0$ if and only if $X = \lambda Y$ for some $\lambda \neq 0$ and X is a non-zero null vector.
4. Let $\{v_1, \dots, v_k\}$ be a set of causal vectors, such that $g(v_i, v_j) < 0$ for any $i \neq j$. Prove that given positive numbers $a_1, \dots, a_k$, the vector $a_1 v_1 + \cdots + a_k v_k$ is time-like. Give a counterexample showing that the requirement $g(v_i, v_j) < 0$ is needed. ♠

1.18 The previous exercise enables us to make the following definition.

Definition
A Lorentzian manifold (M, g) is said to be time-orientable *if it supports a continuous time-like vector field T. In this case, given a causal vector X, we say that X is* future-directed *if $g(X, T) < 0$ and* past-directed *if $g(X, T) < 0$.* ◊

Note that if T is a continuous time-like vector field, then so is $-T$, and the two define opposite notions of future and past.

DRAFT COPY

1.19 A basic theorem in manifold theory guarantees the existence of a Riemannian metric for paracompact manifolds.

Sketch of proof. Cover the manifold with countably many local coordinate charts U_I. Extract from this a subordinate partition of unity Φ_I. On each coordinate chart, the pull-back of the Euclidean metric, which we denote g_I, is a Riemannian metric. The sum $\sum_I \Phi_I g_I$ is a Riemannian metric. □

A key observation is that *convex combinations of positive definite matrices is positive definite*.

Exercise
Give a counterexample for Lorentzian metrics. That is to say, produce an open set $U \subseteq \mathbb{R}^{1+n}$ and two Lorentzian metrics g and g', such that the bilinear form given by $(X,Y) \mapsto g(X,Y) + g'(X,Y)$ is no longer Lorentzian. ♠

1.20 The previous exercise shows that the argument given previously cannot be used to guarantee the existence of a Lorentzian metric on an arbitrary paracompact manifold. The following theorem provides a topology characterization of those manifolds that admit a Lorentzian metric.

Theorem
Let M be a smooth manifold (assumed to be paracompact and orientable). Then
1. *If M admits a nowhere-vanishing vector field, then M admits a Lorentzian metric.*
2. *If M is compact and admits a Lorentzian metric, then M admits a nowhere-vanishing vector field.* ■

The existence of a nowhere-vanishing vector field is equivalent to the vanishing of the Euler-Poincaré characteristic on compact orientable manifolds; this is a topological condition.

Proof. First we prove the forward direction: since M is assumed to be paracompact, it admits a Riemannian metric g. Let v denote the nonwhere-vanishing vector field that we assume exists. Then the tensor field

$$\tilde{g} = g - 2\frac{v^\flat \otimes v^\flat}{g(v,v)}$$

is Lorentzian. (On the right, the musical operations is with respect to g.)

For the reverse direction: since M is assumed to be compact, it admits a Riemannian metric g. By assumption it also has a Lorentzian metric $\tilde{g}$. The operator given by $V^a \mapsto g^{ab}\tilde{g}_{bc}V^c$ is self-adjoint with respect to g, and hence by the spectral theorem can be diagonalized. Our assumption on $\tilde{g}$ shows it has a unique negative eigenvalue, and since g and $\tilde{g}$ vary smoothly as the base point $p \in M$ changes, the corresponding eigenspace for the negative eigenvalue may be chosen smoothly. This shows that there exists a one-dimensional distribution in TM. A little bit of topological degree theory upgrades this to the existence of a nowhere-vanishing vector field (see Chapter 5, Proposition 37 in O'Neill, *Semi-Riemannian geometry: with applications to relativity* for more details). □

© Willie Wai-Yeung Wong

DRAFT COPY

1.21 (†Remarks)

1. A note concerning the result about topological degree theory. The result roughly goes something like this: the existence of a one-dimensional distribution in TM implies that a double cover of M has a nowhere-vanishing vector field, which implies that the double cover has vanishing Euler characteristic. Since M is orientable, this shows the Euler characteristics of M is zero too. This in turn implies that M admits a nowhere-vanishing vector field. However, this *does not imply* that the original one-dimensional distribution has a non-vanishing section: the nowhere-vanishing vector field could be something else altogether.
2. In particular, Definition 1.18 is *meaningful*: not all Lorentzian manifolds are time-orientable.
3. The fact that Riemannian structures are often more tame than Lorentzian structures means that a recurring theme in the study of Lorentzian geometry is the use of auxiliary Riemannian structures as a way to access otherwise inaccessible theorems, or as a way to provide additional coercivity for technical constructions. This idea will show up a few more times in these notes.
4. If M is non-compact, obstruction theory guarantees that M admits a nowhere-vanishing vector field. And hence non-compact manifolds always admit Lorentzian metrics.

Milnor, Topology from the differentiable viewpoint

1.22 Example

To give an explicit example of a non-time-orientable Lorentzian manifold, we may consider $M = \mathbb{T}^3$, given the explicit coordinates $(x, y, z) \in [0, 2\pi)^3$. We may choose the metric to be given by

$$-\Big(\cos(\tfrac{1}{2}z)\, \mathrm{d}x + \sin(\tfrac{1}{2}z)\, \mathrm{d}y\Big)^2 + \Big(-\sin(\tfrac{1}{2}z)\, \mathrm{d}x + \cos(\tfrac{1}{2}z)\, \mathrm{d}y\Big)^2 + \mathrm{d}z^2.$$

(Note that after expanding everything out, the metric coefficients are 2π periodic, and so the metric tensor is well-defined.) Note that at $x = y = z = 0$ the metric is $-\mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2$. Choose any time-like vector with *positive* x-component. Now travel around the curve $x = y = 0$ with increasing z, and select along the curve continuously a time-like vector field extending the initial choice, we will see that necessarily, by the time we reach to $z = 2\pi \cong 0$, the vector field must now have a *negative* x-component. Of course, one nontheless sees that M admits at least one non-vanishing vector field, as promised by Theorem 1.20. ◊

Einstein's Equation

1.23 Having provided a model for the space-time, we next need to prescribe the how gravity arises from the matter fields. As mentioned above, gravity will be modelled through a linear connection; in the case of a Lorentzian manifold, the natural one to choose is the *Levi-Civita* connection, the unique torsion free, metric-compatible linear connection. Below we will use ∇ to denote the associated covariant derivative. We record here our sign conventions for the Riemann tensor:

$$\mathrm{Riem}(X, Y)Z = \mathrm{Riem}_{XY}Z = \nabla_{[X,Y]}Z - [\nabla_X, \nabla_Y]Z \tag{1.23.1}$$

© Willie Wai-Yeung Wong

DRAFT COPY

in index notation we write

$$(\mathrm{Riem}(X,Y)Z)^a = \mathrm{Riem}_{bcd}{}^a X^b Y^c Z^d. \tag{1.23.2}$$

By definition we have antisymmetry in the first two indices:

$$\mathrm{Riem}_{bcd}{}^a + \mathrm{Riem}_{cbd}{}^a = 0. \tag{1.23.3}$$

In general, for torsion-free linear connections, we have both the first and second Bianchi identities hold

Both are closely related to Jacobi's identity for the Lie bracket.

$$\mathrm{Riem}_{bcd}{}^a + \mathrm{Riem}_{cdb}{}^a + \mathrm{Riem}_{dbc}{}^a = 0, \tag{1.23.4}$$

$$\nabla_e \mathrm{Riem}_{bcd}{}^a + \nabla_b \mathrm{Riem}_{ced}{}^a + \nabla_c \mathrm{Riem}_{ebd}{}^a = 0. \tag{1.23.5}$$

If we lower the final index for the Riemann tensor to obtain a $(0,4)$-tensor, the fact that Riem arises from a *metric* connection implies that

$$\mathrm{Riem}_{bcda} = \mathrm{Riem}_{dabc}. \tag{1.23.6}$$

The Ricci tensor is the trace

$$\mathrm{Ric}(X,Z) = \mathrm{tr}(Y \mapsto \mathrm{Riem}(X,Y)Z), \qquad \mathrm{Ric}_{ab} = \mathrm{Riem}_{acb}{}^c. \tag{1.23.7}$$

Note that as a consequence of (1.23.6) the Ricci tensor is symmetric. Taking the *metric* trace of Ric we obtain the scalar curvature

$$\mathrm{R} = \mathrm{Ric}_a{}^a. \tag{1.23.8}$$

If we take (1.23.5) and contract the a and e indices, we find

$$\nabla_a \mathrm{Riem}_{bcd}{}^a + \nabla_b \mathrm{Ric}_{cd} - \nabla_c \mathrm{Ric}_{bd} = 0. \tag{1.23.9}$$

Taking the trace again between the b and d indices via the metric (here we are also implicitly using the various symmetries of the Riemann tensor described above), we find

$$\nabla^a \mathrm{Ric}_{ac} + \nabla^b \mathrm{Ric}_{bc} - \nabla_c \mathrm{R} = 0. \tag{1.23.10}$$

This motivates us to define the *Einstein tensor*

$$\mathrm{Einst} := \mathrm{Ric} - \frac{1}{2}\mathrm{R}g \implies \nabla^a \mathrm{Einst}_{ab} = 0. \tag{1.23.11}$$

Traditionally the letter G is often used for the Einstein tensor.

We make the passing remark that for two dimensional Lorentzian (and Riemannian) manifolds, it is a fact that $\mathrm{Einst} \equiv 0$ always.

Historically the implication is reversed. It was first observed that the energy momentum tensor should (based on what happens in special relativistic physics) have the two properties, from which the choice of Einst *emerged among other curvature quantities.*

1.24 (Einstein's equations) Modern formulations of general relativity is based on the following law relating certain tensorial quantities, termed Einstein's equation(s):

$$\mathrm{Einst} + \Lambda g = T. \tag{1.24.1}$$

The left side of the equation is made up entirely of geometry: the Einstein tensor Einst is a curvature quantity defined by (1.23.11), g is the space-time metric, and Λ is a chosen real

DRAFT COPY

constant (may be zero) called the "cosmological constant". The right side of the equation describes the matter source for gravity. We assume that all the matter in the universe together generates an *energy momentum tensor* T (sometimes also called the *stress energy tensor*); note that for the equation to make sense, the energy momentum tensor must satisfy two properties:

- T must be divergence free, in that $\nabla^a T_{ab} = 0$;
- T must be symmetric.

Starting with a general matter model, arranging for a symmetric two tensor is fairly easy. But coming up with something that is *automatically divergence free* due to the laws governing matter evolution is not always easy, especially if one approaches this phenomenologically based on physical experiments and measurements. Einstein resolved this difficulty by supplementing his theory with the **strong equivalence principle**. In a nutshell, this principle states that the evolution equations governing matter evolution "look the same" on a general relativistic background as they do on a special relativistic background, with the partial differentiation on Minkowski space replaced by the covariant differentiation on our curved space-time. Additionally, the energy momentum tensor for the matter fields will be exactly their special relativistic counterpart, with and partial differentiation replaced by covariant differentiation as above.

1.25 Exercise

Prove that, when the space-time is $(n+1)$-dimensional with $n \geq 2$, Einstein's equation (1.24.1) is equivalent to

$$\mathrm{Ric} = T - \frac{1}{n-1}\mathrm{tr}_g(T)g + \frac{2}{n-1}\Lambda g. \qquad \spadesuit$$

1.26 (Lagrangian formulation) In modern discussion, the requirement on the form of T is largely swept-aside once one adopts a Lagrangian formulation of Einstein's equation. We summarize the formal computations here. We postulate that the equations of physics is governed by an *action principle*; the action consists of two parts: one for gravity and one for the matter

$$\mathcal{S} = \mathcal{S}_{\mathrm{grav}} + \mathcal{S}_{\mathrm{matt}}.$$

For general relativity, the gravitational term $\mathcal{S}_{\mathrm{grav}}$ is postulated to come from the Einstein–Hilbert action

$$\mathcal{S}_{\mathrm{grav}} = \int_M \mathrm{R} - 2\Lambda \,\mathrm{dvol}_g. \tag{1.26.1}$$

Note that in most cases M is non-compact with infinite volume, so that this integral is only a formal integral used to define its variation. The matter term contributes

$$\mathcal{S}_{\mathrm{matt}} = \int_M L \,\mathrm{dvol}_g; \tag{1.26.2}$$

here we are assuming that the field is modeled after some sections of certain vector bundles over M, and that the only metric information that enters into the Lagrangian density L is through the metric g itself (and not the connection or curvature).

I claim that (1.24.1) arises as the formal Euler–Lagrange equation of this action, if we take the variation with respect to the metric tensor g.

DRAFT COPY

First we consider the variation of the volume form with respect to g. With respect to a fixed coordinate system, the volume form reads $\mathrm{dvol}_g = \sqrt{|\det(g)|}dx^1 \dots dx^n$. Recall now Jacobi's identity

$$\delta(\det A) = \det(A)\mathrm{tr}(A^{-1}\delta A)$$

we find

$$\delta \mathrm{dvol}_g = \frac{1}{2}\mathrm{tr}_g(\delta g)\ \mathrm{dvol}_g. \tag{1.26.3}$$

Next we consider the variation of the scalar curvature R. Again in local coordinates, first we observe that the inverse metric changes by

$$\delta(g^{-1})^{\mu\nu} = -g^{\mu\lambda}(\delta g)_{\lambda\rho}g^{\rho\nu}$$

using that $g^{\mu\nu}g_{\nu\lambda} = \delta^{\mu}_{\lambda}$ is *constant*. A long computation shows that the Christoffel symbols change by

$$g_{\lambda\mu}(\delta\Gamma)^{\lambda}_{\alpha\beta} = \frac{1}{2}\Big[\nabla_\alpha(\delta g)_{\mu\beta} + \nabla_\beta(\delta g)_{\mu\alpha} - \nabla_\mu(\delta g)_{\alpha\beta}\Big];$$

As the difference between two connections is given by a tensor, this quantity (an infinitesimal difference) is naturally tensorial, so the use of local coordinates is not required.

here ∇ is still the Levi-Civita connection for the metric g. Plugging this into the definition of the Riemann curvature tensor, we see that the quadratic terms can be absorbed into another covariant differentiation, and find that

$$(\delta\mathrm{Riem})_{\alpha\beta\gamma}{}^{\delta} = \nabla_\beta(\delta\Gamma)^{\delta}_{\alpha\gamma} - \nabla_\alpha(\delta\Gamma)^{\delta}_{\beta\gamma}.$$

And hence

$$(\delta\mathrm{Ric})_{\alpha\gamma} = \nabla_\beta(\delta\Gamma)^{\beta}_{\alpha\gamma} - \frac{1}{2}\nabla_\alpha\nabla_\gamma\mathrm{tr}_g(\delta g). \tag{1.26.4}$$

Finally, writing $\delta\mathrm{R} = \delta(g^{-1})^{\alpha\gamma}\mathrm{Ric}_{\alpha\gamma} + g^{\alpha\gamma}(\delta\mathrm{Ric})_{\alpha\gamma}$ we find

$$\delta\mathrm{R} = -\mathrm{Ric}^{\alpha\gamma}(\delta g)_{\alpha\gamma} + \nabla_\beta\Big[\underbrace{g^{\alpha\gamma}(\delta\Gamma)^{\beta}_{\alpha\gamma} - \frac{1}{2}\nabla^\beta\mathrm{tr}_g(\delta g)}_{H^\beta}\Big]. \tag{1.26.5}$$

Summarizing we have

$$\delta\mathcal{S}_{\mathrm{grav}} = \int_M \nabla_\beta H^\beta\ \mathrm{dvol}_g - \int_M \mathrm{Ric}^{\alpha\gamma}(\delta g)_{\alpha\gamma}\ \mathrm{dvol}_g + \int_M \frac{1}{2}(\mathrm{R} - 2\Lambda)g^{\alpha\gamma}(\delta g)_{\alpha\gamma}\ \mathrm{dvol}_g;$$

the first term vanishes for compactly supported variations since it is the integral of a total divergence, the remaining terms can be regrouped to read

$$\delta\mathcal{S}_{\mathrm{grav}} = -\int_M (\mathrm{Einst}^{\alpha\gamma} + \Lambda g^{\alpha\gamma})(\delta g)_{\alpha\gamma}\ \mathrm{dvol}_g. \tag{1.26.6}$$

When one considers the situation on a manifold with boundary, the integral of a total divergence gives a boundary term. This causes significant headaches that is partially resolved by the introduction of the Gibbons–Hawking–York term *into the Einstein–Hilbert action. For more details see S W Hawking and Horowitz, "The gravitational Hamiltonian, action, entropy and surface terms".*

For the matter term, notice that the metric g appears implicitly again in the space-time volume form; from which we obtain

$$\delta\mathcal{S}_{\mathrm{matt}} = \int_M \frac{\delta L}{\delta g_{\alpha\gamma}}(\delta g)_{\alpha\gamma} + \frac{1}{2}Lg^{\alpha\gamma}(\delta g)_{\alpha\gamma}\ \mathrm{dvol}_g.$$

© Willie Wai-Yeung Wong

So setting the energy momentum tensor to be precisely the expression

$$T^{\alpha\gamma} = \frac{\delta L}{\delta g_{\alpha\gamma}} + \frac{1}{2} L g^{\alpha\gamma} \tag{1.26.7}$$

(which is automatically symmetric, since the metric g and its variation δg are assumed so) we find that the Euler–Lagrange equations for our action (relative to variations in the metric) to be exactly (1.24.1).

1.27 Note that we can also derive the equations of motion for the matter fields through (1.26.2), this time perturbing the section (of the vector bundle) one takes to represent the matter field.

1.28 (T is divergence free) The Lagrangian formulation actually guarantees that the derived energy momentum tensor will be automatically *divergence free*. To see this, we will consider a very special form of variation, that takes advantage of the *diffeomorphism invariance* of the "geometry half" of the theory. Let g be a critical point of action $\mathcal{S}$. Start with a one-parameter family of diffeomorphisms Φ_s generated by a compactly supported vector field v, and let our one-parameter family of metrics g_s on M be given by the *pullbacks* $\Phi_s^* g$ of the given critical g. The corresponding variation is then

$$(\delta g)_{\alpha\beta} = (\mathcal{L}_v g)_{\alpha\beta} = \nabla_\alpha v_\beta + \nabla_\beta v_\alpha. \tag{1.28.1}$$

(Here $\mathcal{L}_v$ is the Lie derivative in the direction v.)

Note that (M, g_s) and (M, g) are isometric by definition. This means that their Einstein–Hilbert functions agree, since they depend only on geometric quantities like the volume form and curvature. And hence

$$\delta \mathcal{S}_{\text{grav}} = 0.$$

Since g is a critical point, we must have then

$$\delta \mathcal{S}_{\text{matt}} = \int_M T^{\alpha\beta} (\delta g)_{\alpha\beta} \, \mathrm{dvol}_g = 2 \int_M T^{\alpha\beta} \nabla_\alpha v_\beta \, \mathrm{dvol}_g = 0.$$

Using now that v has compact support, we can integrate by parts and obtain

$$\int_M (\nabla_\alpha T^{\alpha\beta}) v_\beta \, \mathrm{dvol}_g = 0.$$

Using finally that v is arbitrary, this implies that T must be divergence free.

Newtonian Limit of General Relativity

1.29 In this section we consider how Newtonian gravity may be seen as a limiting case of general relativity. Based on historical developments, we are led to believe that, when

- the space-time metric is approximately Minwkoski, and
- the motion of the particles are slow compared to the speed of light,

then Newtonian theory is a good approximation. Our goal is to "justify" this claim.

For a more detailed discussion, including some philosophical caveats of these arguments, see Section 4.4a in Wald, General Relativity.

DRAFT COPY

DRAFT COPY

1.30 Let $\varepsilon > 0$ be a small parameter. We will now assume that there exists a parametrization of our space-time as $\mathbb{R} \times \mathbb{R}^3$ with respect to which our metric can be written in the form

$$g = g_{\text{Mink}} + h \tag{1.30.1}$$

where $g_{Mink} = \text{diag}(-1, 1, 1, 1)$ is the Minkowski metric, and h is a perturbation. That the space-time is approximately Minkowski we will capture in the assumption

$$|h|, |\partial h| \leq \varepsilon \tag{1.30.2}$$

uniformly (relative to the coordinates). That the motion of particles are slow compared to the speed of light we will model by requiring that, every time one takes a *time* derivative, the resulting quantity is smaller by the original by a factor of ε. So in particular

$$|\partial_0 h| \leq \varepsilon^2. \tag{1.30.3}$$

Under this ansatz, we see that the Christoffel symbols (see also the linearization computations given in ¶ 1.26) satisfy (for $i, j, k \in \{1, 2, 3\}$)

$$\begin{aligned}
\Gamma^0_{00} &= O(\varepsilon^2) & \Gamma^i_{jk} &= \frac{1}{2}(\partial_j h_{ik} + \partial_k h_{ij} - \partial_j h_{ik}) + O(\varepsilon^2) \\
\Gamma^0_{0i} &= -\frac{1}{2}\partial_i h_{00} + O(\varepsilon^2) & \Gamma^i_{0j} &= \frac{1}{2}(\partial_j h_{0i} - \partial_i h_{0j}) + O(\varepsilon^2) \\
\Gamma^0_{ij} &= -\frac{1}{2}(\partial_i h_{0i} + \partial_j h_{0i}) + O(\varepsilon^2) & \Gamma^i_{00} &= -\frac{1}{2}\partial_i h_{00} + O(\varepsilon^2)
\end{aligned}$$

The leading order terms of the Ricci curvature tensor provides

$$\text{Ric}_{00} = \sum_{i=1}^{3} \partial_i \Gamma^i_{00} + O(\varepsilon^2) = -\frac{1}{2}\triangle h_{00} + O(\varepsilon^2). \tag{1.30.4}$$

Now, in our setting, we expect our matter content to be slowly moving. This means that the spatial components of the energy momentum tensor should be much smaller than the temporal components. In other words, we expect, for some μ (which we will interpret as a mass-energy density) that is size ε,

$$T_{00} = \mu \geq 0, \quad T_{0i} = O(\varepsilon^2), \quad T_{ij} = O(\varepsilon^3). \tag{1.30.5}$$

In particular, we have $\text{tr}_g(T) = -\mu + O(\varepsilon^2)$. So by Exercise 1.25, Einstein's equation implies (setting the cosmological constant to $\Lambda = 0$ and marking the spatial dimension $n = 3$)

$$\text{Ric}_{00} = \frac{1}{2}\mu + O(\varepsilon^2).$$

And so we expect

$$h_{00} = -\triangle^{-1}\mu$$

is essentially the gravitational potential.

© Willie Wai-Yeung Wong

We can check this against the geodesic motion. Let γ denote a geodesic representing a slowly moving particle, so that $\dot\gamma^0 = 1 + O(\varepsilon)$ and $\dot\gamma^i = O(\varepsilon)$. Then the geodesic equation reads

$$\ddot\gamma^\mu + \underbrace{\Gamma^\mu_{\nu\rho}\dot\gamma^\nu\dot\gamma^\rho}_{=\Gamma^\mu_{00} + O(\varepsilon^2)} = 0.$$

So we find

$$\ddot\gamma^0 = O(\varepsilon^2)$$
$$\ddot\gamma^i = \frac{1}{2}\partial_i h_{00} + O(\varepsilon^2)$$

which shows that to leading order, the predicted gravitational acceleration is in agreement with Newton's theory.

DRAFT COPY

DRAFT COPY

Topic 2 (2023/10/23)

Geodesics

2.1 A huge amount of information concerning Lorentzian (and similarly Riemannian manifolds) is captured in the behavior of geodesics. We will focus on several aspects: the notion of completeness for a manifold, the variational description of geodesics, and probing geometry through families of geodesics. As we shall see, our focus will be on *causal* geodesics for a good reason.

Basics Concerning Geodesics

2.2 **(Notational convention on curves)** Given M a smooth manifold, by a *curve* we will refer to a *continuous* function $\gamma : I \to M$ whose domain I is a non-degenerate interval (open or closed or half-open). For $k \in \mathbb{N} \cup \{\infty\}$, a curve γ is C^k if the mapping is k-times continuously differentiable on the interior $\mathring{I}$ of its domain *and* the derivatives extends continuously to $I \setminus \mathring{I}$, the derivative $\dot{\gamma}$ is often referred to as its velocity vector. We use "smooth" and "C^∞" as synonyms. A C^k curve is said to be *regular* if $\dot{\gamma} \neq 0$ on its domain.

2.3 **Definition (Geodesics)**
Let (M, g) be a Lorentzian manifold, and $\gamma : (a, b) \to M$ be a smooth regular curve.
- *We say that γ is a* geodesic *if the acceleration vector field $\ddot{\gamma} := \nabla_{\dot{\gamma}}\dot{\gamma}$ (which is defined along γ) is proportional to $\dot{\gamma}$;*
- *We say a geodesic γ is* affinely parametrized *if $\ddot{\gamma} \equiv 0$. We sometimes shorten "affinely parametrized geodesic" to simply "affine geodesic".*

Similarly, given a non-vanishing vector field V on (an open subset of) M, we say that V is a geodesic vector field *if $\nabla_V V \propto V$ everywhere, and furthermore we say that V is* affinely geodesic *if $\nabla_V V \equiv 0$.* ◇

In Riemannian geometry we frequently prefer speaking of geodesics "parametrized by arc-length. In Lorentzian geometry this becomes problematic for null geodesics for which arc-length is not a well-defined notion. Hence we prefer affine parametrization as a more natural concept.

2.4 **Exercise**
1. Letting $\gamma : (a, b) \to M$ be a geodesic. Prove that there exists a smooth mapping $\sigma : (0, 1) \to (a, b)$ such that $\gamma \circ \sigma$ is an affine geodesic.
2. Prove that every integral curve of a geodesic / affinely-geodesic vector field is a geodesic / affine geodesic.

3. Prove that along an affine geodesic, the quantity $g(\dot\gamma,\dot\gamma)$ is constant; conclude therefore that geodesics cannot change causal type (timelike, spacelike, null). ♠

2.5 Example (Eikonal functions)
Given a Lorentian manifold (M,g), a function $f : M \to \mathbb{R}$ is said to be *eikonal* if $g^{-1}(\mathrm{d}f,\mathrm{d}f)$ is constant. Let $X = (\mathrm{d}f)^\sharp$ be the gradient of one such function. I claim that X is affinely geodesic. Let V be any vector field. We compute

$$g(\nabla_X X, V) = g(\nabla_X (\mathrm{d}f)^\sharp, V) = [\nabla_X(\mathrm{d}f)](V) = [\nabla_V(\mathrm{d}f)](X)$$

where in the last step we used that ∇ is torsion free and hence the Hessian of a function is symmetric. We can rewrite this expression as

$$= [\nabla_V(\mathrm{d}f)]((\mathrm{d}f)^\sharp) = g^{-1}(\nabla_V(\mathrm{d}f),\mathrm{d}f) = \frac{1}{2}V(g^{-1}(\mathrm{d}f,\mathrm{d}f)) = 0.$$

Since this holds for any V, we conclude that $\nabla_X X = 0$. ◇

2.6 Some standard facts from Riemannian geometry carry through also in the Lorentzian (and more generally, the pseudo-Riemannian setting). We record a few of them here without proof. First, Picard's Theorem on the existence and uniqueness of solutions to ordinary differential equations holds in general, regardless of the signature of the manifold. So we have

The results listed here have complete proofs in Chapters 3 and 5 of O'Neill, Semi-Riemannian geometry: with applications to relativity.

Proposition
Let (M,g) be a Lorentzian manifold, and $p \in M$. Given $v \in T_pM$, there exists numbers $-\infty \leq a < 0 < b \leq \infty$ and a unique function $\gamma : (a,b) \to M$ such that
- *γ is an affinely-parameterized geodesic;*
- *$\gamma(0) = p$ and $\dot\gamma(0) = v$;*
- *γ is* maximally extended*; that is, whenever $\tilde\gamma : (\tilde a,\tilde b) \to M$ is another function that satisfies the previous two points, then $(\tilde a,\tilde b) \subseteq (a,b)$ and $\tilde\gamma$ is the restriction of γ to the subinterval.* ■

Letting (M,g) and p as in the statement of the previous proposition. We can let $U_p \subseteq T_pM$ be given by

$$U_p := \{v \in T_pM \mid (a,b) \supseteq [0,1]\}; \tag{2.6.1}$$

where (a,b) is the maximal interval of existence as defined above. Note that U_p is necessarily star-shaped. Smooth dependence on initial data for Picard's Theorem implies that U_p is open, and we can define a mapping

$$\exp_p : U_p \to M, \quad \exp_p(v) = \gamma(1) \text{ where } \gamma \text{ is as in the previous proposition.} \tag{2.6.2}$$

By the previous proposition $\exp_p$ is unique, and its domain is maximal. Note that line segments through the origin in U_p gets mapped to geodesic segments in M. An application of the inverse function theorem shows that

Proposition
There exists a subset $V_p \subseteq U_p$ containing the origin in T_pM such that $\exp_p|_{V_p}$ is a diffeomorphism onto its image. Furthermore, V_p can be chosen to be star-shaped, in which case given $v \in V_p$, the geodesic $[0,1] \ni t \mapsto \exp_p(tv)$ is the unique geodesic joining p to $\exp_p(v)$ that lies entirely in $\exp_p(V_p)$. ■

DRAFT COPY

DRAFT COPY

We can choose a basis of T_pM such that the metric g is diagonalized, and using this basis we can identify V_p with a start-shaped subset of $\mathbb{R}^n$, with p corresponding to the origin. This provides a *geodesic normal coordinate chart* on M near p.

Proposition

In a geodesic normal coordinate chart, $\Gamma_{ij}^k(0) = 0$. ■

Another key property of the exponential map states that orthogonality to radial directions is preserved under the exponential map. This result is usually called the *Gauss Lemma* and below is one version of its statement:

Proposition

Since T_pM is a Lorentzian vector space, its tangent space at $x \in T_pM$ may be canonically identified $T_xT_pM \cong T_pM$ and be given the same scalar product g_p. Now let $x \in T_pM$, the differential of the exponential map at x is $D_x \exp_p : T_xT_pM \cong T_pM \mapsto T_{\exp_p(x)}M$ under the identification above. We claim that

$$g_{\exp_p(x)}(D_x \exp_p(v), D_x \exp_p(x)) = g_p(v, x)$$

for every $v \in T_pM$ (note that the second vector is fixed to be the radial vector x). ■

2.7 Definition (Convex normal neighborhood)

Given a Lorentzian manifold (M, g), an open subset $V \subseteq M$ is said to be convex *if, given any $p \in V$, there exists a star-shaped subset $V_p \subseteq T_pM$ such that $\exp_p|_{V_p}$ defines a diffeomorphism from $V_p \to V$.*

Given a $q \in M$, we say that V is a convex normal neighborhood *of q if V is convex and $q \in V$.* ♢

2.8 Lemma

On a Lorentzian manifold (M, g), every point $p \in M$ has a convex normal neighborhood. ■

A detailed proof of this Lemma is given as Proposition 7 in Chapter 5 of O'Neill, *Semi-Riemannian geometry: with applications to relativity*; here we just give a sketch of the basic strategy. The key idea is that in a normal coordinate system, the Christoffel symbols vanish at the origin and so is size $O(r)$, where r is the "coordinate distance" of a point to the origin. By sticking to a sufficiently small ball (in the coordinate sense), we can ensure that the Christoffel symbols are sufficiently small, and hence all geodesics are almost straight lines. Observe that the round ball in a vector space is strictly convex; as the size of the Christoffel symbols places restriction on the maximum curvature (relative to the coordinates) of the geodesics, this shows that if a geodesic curve starts within the small ball, once it exits the ball it will take a while before it can return again.

2.9 A key consequence of the previous lemma is the following characterization of inextensible geodesics.

Corollary

Let (M, g) be a Lorentzian manifold, $p \in M$, $v \in T_pM$, and $\gamma : (a, b) \to M$ an affine geodesic satisfying $\gamma(0) = p$ and $\dot\gamma(0) = v$. Then the domain (a, b) is maximal if and only if there exists a decreasing sequence t_i and an increasing sequence s_i in (a, b) with $\lim t_i = a$ and $\lim s_i = b$, such that the sequence of points $\gamma(t_i)$ (and similarly $\gamma(s_i)$) fail to converge in the manifold topology. ■

© Willie Wai-Yeung Wong

Completeness

2.10 Thus far we've only discussed concepts where there is no distinction between the Riemannian and Lorentzian settings (indeed, the discussion in the previous section holds for any signature). Now we move on to where the Lorentzian setting introduces complications.

2.11 The following technical lemma is sometimes useful.

Lemma
Let (M, g) be a Lorentzian manifold, and U be a convex open set. Suppose $\gamma : (a, b) \to U$ is a smooth regular curve with $\dot\gamma$ always timelike. Let $c \in (a, b)$. Then for every $s \in (a, b) \setminus \{c\}$, the vector $\exp^{-1}_{\gamma(c)} \gamma(s) \in T_{\gamma(c)}M$ is timelike. ■

Proof. For convenience we introduce some additional notations:
- the point $o = \gamma(c)$. We shall also assume without loss of generality that $c = 0$.
- $\tilde U$ will denote the star-shaped set $\exp_o^{-1}(U) \subseteq T_oM$.
- $\tilde\gamma : (a, b) \to \tilde U$ is defined to be $\tilde\gamma = \exp_o^{-1} \circ \gamma$.
- $\tilde P$ on $\tilde U$ refers to the "identity vector field"; more precisely, since T_oM is a vector space, we have a canonical identification of $T_xT_oM \cong T_oM$. Under this identification the vector field $\tilde P_x = x$. P is the pushforward of $\tilde P$ to U.
- $q : (a, b) \to \mathbb{R}$ is the function $q(s) = g_o(\tilde\gamma(s), \tilde\gamma(s))$.

Notice that $q(0) = 0$, and our goal is to prove that $q(s) < 0$ for all $s \neq 0$.

First, we have that the desired result holds for sufficiently small s: observe that $\dot q(0) = 0$ and

$$\ddot q(0) = 2g_o(\dot{\tilde\gamma}(0), \dot{\tilde\gamma}(0)) = 2g_o(\dot\gamma(0), \dot\gamma(0)) < 0$$

as γ is time-like. By smoothness of q we see that the claim holds.

It remains to show that this extends for larger s. The first derivative is

$$\dot q(s) = 2g_o(\tilde\gamma(s), \dot{\tilde\gamma}(s)) = 2g_{\gamma(s)}(P, \dot\gamma(s)); \tag{2.11.1}$$

the latter equality due to the Gauss Lemma. Notice that for sufficiently small s, by continuity we have $\dot{\tilde\gamma}(s) \approx \dot{\tilde\gamma}(0)$ and $\tilde\gamma(s) \approx s\dot{\tilde\gamma}(0)$, which implies therefore that for all sufficiently small non-zero s,

$$s\dot q(s) < 0.$$

Our proof would be concluded if we can show that this final inequality extends to all s, not just those small.

Suppose now that $s\dot q(s) < 0$ holds on the interval $(0, b')$. This implies, using that $q(s)$ is initially negative near 0, that $q(s) < 0$ on $(0, b']$ (we include b' by continuity). Hence $\tilde\gamma$ is time-like on $(0, b']$. Continuity ensures that $\tilde\gamma|_{(0,b']}$ all belong to the same connected component of the set of time-like vectors. Hence the restriction of P to $\gamma|_{(0,b']}$ is a continuous nonvanishing time-like vector field; since initially $\dot q(s) < 0$ we find that it has the same time-orientation as $\dot\gamma(s)$ on $(0, b']$. But this immediately implies that $\dot q(b') < 0$ as it is the product of two time-like vectors with the same orientation.

Our claim follows then by continuous induction. □

DRAFT COPY

© Willie Wai-Yeung Wong

DRAFT COPY

2.12 (Remarks) The proof above used implicitly that a Lorentzian metric has signature with only negative sign, which provides some amount of convexity to the set of time-like directions. The analogous statement is *not* true for space-like curves: an easy counterexample comes from from a helix in Minkowski space. Consider $\gamma : \mathbb{R} \to \mathbb{R}^{1+2}$ given by

$$\gamma(s) = (s, 5\cos s, 5\sin s).$$

Then $g(\dot\gamma, \dot\gamma) = -1 + 25(\sin^2 s + \cos^2 s) = 24 > 0$. But note that $\gamma(0)$ and $\gamma(2\pi)$ can be connected by the timelike geodesic $s \mapsto (s, 5, 0)$.

2.13 The usual definition of geodesic completeness on Riemannian manifolds can still be given, but now we wish to split the definition based on the causal type of curves being considered (recall from Exercise 2.4 that each geodesic curve has a fixed type).

Definition (Geodesic completeness)
Let (M, g) be a Lorentzian manifold.

- *We say that it is* timeike / spacelike / null geodesically complete *if for every $p \in M$ and $v \in T_pM$, with v being timelike / spacelike / null, the maximally extended affinely-parametrized geodesic through (p, v) has domain $= \mathbb{R}$.*
- *We say that it is* geodesically complete *if all it satisfies the condition for all three causal types of geodesics.* ◊

2.14 The reason that we split completeness by causal type is because they are independent properties. Here we give an example that is null and spacelike complete, but not timelike complete.

Example (Geroch)
Let our underlying manifold be $\mathbb{R}^2$ with coordinates (t, x). Choose a smooth positive function f with the following properties

- On $\{|x| \geq 1\}$, we have $f \equiv 1$.
- $f(t, x) = f(t, -x)$.
- $\int_{-\infty}^{\infty} f(t, 0)\, \mathrm{d}t$ is finite.

Consider the metric g given by

$$f^2 \cdot (-\mathrm{d}t^2 + \mathrm{d}x^2).$$

First we prove that this manifold is timelike geodesically incomplete. It is sufficient to find one inextensible timelike geodesic which, when affinely parameterized, has domain strictly smaller than $\mathbb{R}$. Observe that the reflection $x \mapsto -x$ generates an isometry of our manifold (since we chose f to be even); hence its fixed-point set, the t-axis, is totally geodesic in our manifold. Since it is one dimensional, the curve is geodesic. To find an affine parametrization, notice that in affine parametrization the velocity is constant in length. Hence we can choose the velocity field to be $v = \frac{1}{f(t,0)}\partial_t$ along the t-axis. In other words, in affine parametrization we are looking for $\gamma = \gamma(s)$ satisfying

$$\gamma(0) = (0, 0), \quad \dot\gamma(s) = \Big(\frac{1}{f(\gamma(s))}, 0\Big).$$

From which we see $\gamma^{-1}((t, 0)) = \int_0^t f(\tau, 0)\ \mathrm{d}\tau$ and so the maximal domain for γ is (a, b) with $a = \int_0^{-\infty} f(\tau, 0)\ \mathrm{d}\tau$ and $b = \int_0^{\infty} f(\tau, 0)\ \mathrm{d}\tau$ both finite.

Next we show null and space-like completeness. Given a null or space-like geodesic γ that is not constant, necessarily its x-component is monotonic; this is because our metric is conformal to the Minkowski metric and so have the same sets of timelike, spacelike, and null vectors. I claim that this means every maximally extended such γ (when affinely parametrized), will exit the region $\{|x| < 1\}$. More precisely, I claim that if γ is maximally extended with domain (a, b), then there exists a subdomain $(c, d) \subseteq (a, b)$ such that γ is outside $\{|x| < 1\}$ when the parameter is outside (c, d). Assuming this claim, this means that γ is eventually in the region where $f \equiv 1$, and the metric is exactly Minkowski there; the geodesic equation can therefore be solved explicitly and completeness shown.

It remains to prove the claim. Suppose, for contradiction, that for some $c \in (a, b)$, we have that $\gamma|_{[c,b)}$ is always within the set $\{|x| < 1\}$. Since the x-component of γ is monotonic, this means that the x component of γ has a limit as the parameter approaches b. Without loss of generality assume that the x component is increasing, and the limit is $x_f \in (-1, 1)$. Since we know that $\dot\gamma$ is not time-like, by the mean-value theorem we must have that, for any two parameter values s, s', that

$$\frac{|\gamma^t(s) - \gamma^t(s')|}{|\gamma^x(s) - \gamma^x(s')|} \leq 1.$$

But since $\gamma^x(s)$ is Cauchy, so must be $\gamma^t(s)$, and hence $\lim_{s\to b-} \gamma$ exists. But by Corollary 2.9 this contradicts the assumption that γ is maximally extended. ◊

2.15 For those of use familiar with Riemannian geometry, a powerful theorem concerning the completeness of a manifold is:

Theorem (Hopf-Rinow)
The following four statements are equivalent for a Riemannian manifold (M, g):
- *(M, g) is geodesically complete.*
- *For some point $p \in M$, the maximal domain U_p of the exponential map $\exp_p$ is the whole of T_pM.*
- *When equipped with the Riemannian length metric ($d(x,y) = \inf \int_0^1 \sqrt{g(\dot\gamma(s), \dot\gamma(s))}\, ds$ where the infimum is taken among all C^1 curves γ joining $x = \gamma(0)$ to $y = \gamma(1)$), the corresponding metric space is metrically complete.*
- *Sets which are closed and bounded with respect to the Riemannian length metric are compact.* ■

Evidently an analogue theorem for Lorentzian geometry is impossible: for starters, the formula defining the Riemannian length of a curve simply doesn't make sense, as $g(\dot\gamma, \dot\gamma)$ may be negative. (But as we shall see below some aspects of a length can be recovered in a Lorentzian case.) So an understanding of what it means for a Lorentzian manifold to be incomplete is more subtle.

2.16 The previous paragraph should be coupled to two additional points:
1. Many explicit and physically important solutions in general relativity are geodesically incomplete.
2. Mathematical general relativity is often concerned with the evolution problem where the manifold arises dynamically from solving Einstein's equations, rather than being "chosen" by the mathematician.

DRAFT COPY

DRAFT COPY

Together this means it is much less common to see a Lorentzian geometer incorporate completeness as a hypothesis into her theorem, in contrast with what happens in the Riemannian setting.

2.17 A closely related concept to completeness is the question of whether two points in a manifold can be joined by a geodesic. In Riemannian geometry, a consequence of Theorem 2.15 is "On a (connected) geodesically complete Riemannian manifold, any two points are connected by a geodesic." Modelled after this result we make the following definitions for Lorentzian manifolds.

One may ask why there isn't a form for "space-like" in this definition. This is due to the fact that every pair of points in a connected Lorentzian manifold can be connected by at least one space-like curve. To see this, observe that one can approximate a timelike curve with a tightly-coiled helix, which is a space-like curve. (Basically, one can go upstairs using either an elevator or a set of stairs.)

Definition (Geodesic connectedness)
Let (M,g) be a Lorentzian manifold.
- *We say that (M,g) is* timelike / causal geodesically connected *if every pair of points p,q that can be joined by a timelike / causal curve can be joined by a timelike / causal geodesic.*
- *We say that (M,g) is* geodesically connected *if every pair of points p,q can be joined by a geodesic.* ◇

2.18 It turns out that geodesic completeness is much less useful in Lorentzian geometry, in terms of its various implications. A quintessential counterexample is *de Sitter space*, which, as a symmetric space, is a heavily used model in cosmology.

Example (de Sitter space)
The de Sitter space dS^{1+n} is the subset of $(1+n+1)$-dimensional Minkowski space given by

$$\mathsf{dS}^{1+n} := \{x \in \mathbb{R}^{1+(n+1)} \mid \eta(x,x) = 1\} \tag{2.18.1}$$

here η is the Minkowski metric $\mathrm{diag}(-1,1,\dots)$. Graphically it is the exterior hyperboloid to the cone of null vectors based at 0. As its defining function is $x \mapsto \eta(x,x)$, which has differential $x^\flat$, we see that dS^{1+n} is cospacelike. As a hypersurface, the induced metric is also Lorentzian.

TODO: A picture or three would be nice here

Its definition by the quadratic form $\eta(x,x) = 1$ makes dS^{1+n} a *hyperquadric*. A theorem concerning hyperquadrics (which also applies to the hyperbolic space $\mathbb{H}^n$ realized as the hyperboloid $\eta(x,x) = -1$ in $\mathbb{R}^{1+n}$ and to the spheres $\mathbb{S}^n$ as the set of unit euclidean distance from the origin) is:

See Proposition 4.28 and 5.38 in O'Neill, Semi-Riemannian geometry: with applications to relativity.

> Geodesics in hyperquadrics are exactly those curves formed by intersecting the hyperquadric with a two-dimensional plane through the origin.

A direct computation then shows that, as a result, hyperquadrics are always *geodesically complete*.

In the special case of de Sitter space, the characterization above allows us to produce two counterexamples, Both of them are based on the following observation: fix $p = (0,1,0,0\dots) \in \mathsf{dS}^{1+n} \subseteq \mathbb{R}^{1+(n+1)}$, and let $q = (1,\sqrt{2},0,0,\dots) \in \mathsf{dS}^{1+n}$. Then the only two-dimensional plane Π that passes through the origin, p, and q, is the plane spanning the x^0 and x^1 directions. Note, however, that $\Pi \cap \mathsf{dS}^{1+n}$ has two components, with p and q situated on different components. We conclude, therefore,

1. dS^{1+n} is a geodesically complete Lorentzian manifold that is *not* geodesically connected. (Note, however, that it is time-like geodesically connected.)

2. Consider the manifold $M = \mathsf{dS}^{1+n} \setminus \{q\}$. Then the exponential map $\exp_p$ is well-defined on the entirety of T_pM, but M is not geodesically complete. (This is a counterexample to a part of Theorem 2.15.) ◇

Variationally Speaking

2.19 As alluded to earlier, because the metric is no longer signed, the formula typically used to define the length of a curve no longer makes sense for Lorentzian manifolds. However, we can separately consider those curves that are time-like and those curves that are space-like. (The reason that we don't also consider curves that are null is because we are aiming for a variational discussion. Being time-like and space-like are open conditions: if a C^1 curve is time-like / space-like, then any small C^1 perturbations would also be time-like / space-like. The same is not true for null curves. Also, the naive definition of length will always yield a value of 0 for null curves.)

2.20 Definition (Length and proper time)
Let (M,g) be a Lorentzian manifold:

- *The* length *of a* space-like *C^1-curve $\gamma : [a,b] \to M$ is*

$$\ell_+(\gamma) := \int_a^b \sqrt{g(\dot\gamma,\dot\gamma)}.$$

- *The* proper time elapsed *for a time-like C^1 curve $\gamma : [a,b] \to M$ is*

$$\ell_-(\gamma) := \int_a^b \sqrt{-g(\dot\gamma,\dot\gamma)}.$$ ◇

2.21 To incorporate null geodesics, we will use the energy, which is well-defined for all curves.

Definition (Energy)
Let (M,g) be a Lorentzian manfiold, and $\gamma : [a,b] \to M$ a C^1 regular curve. Its energy *is*

$$\mathcal{E}(\gamma) := \int_a^b g(\dot\gamma,\dot\gamma).$$ ◇

2.22 Exercise

1. Check that ℓ_+ and ℓ_- are independent of parametrization. That is, if $\sigma : [c,d] \to [a,b]$ is a C^1-regular bijection, then $\ell_\pm(\gamma) = \ell_\pm(\gamma \circ \sigma)$.
2. Show that $\mathcal{E}$ is not independent of parametrization through an explicit example. ♠

2.23 The following proposition is proved exactly in the same way as in Riemannian geometry, through the computation of the first variation of the three functionals across a one-parameter family of curves.

Proposition
Let (M,g) be a Lorentzian manifold, and $\gamma : [a,b] \to M$ a smooth regular curve. Then:

DRAFT COPY

1. *γ is an affinely parametrized geodesic if and only if it is a critical point of the energy $\mathscr{E}$ among all smooth regular curves $\tilde{\gamma} : [a,b] \to M$ satisfying $\tilde{\gamma}(a) = \gamma(a)$ and $\tilde{\gamma}(b) = \gamma(b)$.*
2. *γ is a time-like geodesic if and only if it is a critical point of ℓ_- among all smooth regular time-like curves connecting $\gamma(a)$ to $\gamma(b)$.*
3. *γ is a space-like geodesic if and only if it is a critical point of ℓ_+ among all smooth regular space-like curves connecting $\gamma(a)$ to $\gamma(b)$.* ■

The first variation formulae are standard, so we just record the result here: throughout let $\gamma_s : [a,b] \to M$ be a one-parameter family of smooth regular curves of the appropriate causal type, and denote by $\alpha = \gamma_0$ and $V = \frac{\partial}{\partial s}\gamma_s|_{s=0}$ the variation field (a vector field along α). The symbols $\dot\alpha$ and $\ddot\alpha$ will denote the velocity and acceleration vector fields of α (latter being $\nabla_{\dot\alpha}\dot\alpha$), and $\dot V = \nabla_{\dot\alpha} V$. We find:

$$\frac{d}{ds}\mathcal{E}(\gamma_s)|_{s=0} = 2\int_a^b g(\dot V, \dot\alpha) = -2\int_a^b g(V, \ddot\alpha)\,\mathrm{d}t + g(V,\dot\alpha)|_a^b.$$

If we further assume that $|g(\dot\alpha,\dot\alpha)| = c^2$ is constant (which can always be achieved by reparametrization, then

$$\frac{d}{ds}\ell_\pm(\gamma_s)|_{s=0} = \mp\frac{1}{c}\int_a^b g(V,\ddot\alpha)\,\mathrm{d}t \pm \frac{1}{c}g(V,\dot\alpha)|_a^b.$$

2.24 Exercise (Critical points in endmanifold case)
Let (M,g) be a Lorentzian manifold, $\Sigma \subseteq M$ a hypersurface, and p a point in $M \setminus \Sigma$. Let $\Gamma_{p,\Sigma}$ be the set of all curves $\gamma : [a,b] \to M$ satisfying:

- $\dot\gamma$ is timelike everywhere;
- $\gamma(b) = p$ and $\gamma(a) \in \Sigma$.

Prove that $\gamma_0 \in \Gamma_{p,\sigma}$ is a critical point of ℓ_- if and only if γ_0 is geodesic and $\dot\gamma_0(a)$ is orthogonal to Σ. ♠

2.25 A well-known result in Riemannian geometry is: "for every $p \in M$, there is a convex neighborhood V of p such that for every $q \in V$, the geodesic joining p to q in V is the shortest curve among all curves joining p to q." The analogue in Lorentzian geometry is false: in fact, given p and q two points in a Lorentzian manifold M such that there exists a spacelike curve joining them, then there exists a spacelike curve joining p to q with arbitrarily small length. This can be achieved by "wiggling the curve" in the time direction. (The opposite construction can be done for time-like curves too.) There also exists a spacelike curve joining p to q with arbitrary large length (as one expects from the Riemannian theory). And so space-like geodesics are *never* "length minimizers" in Lorentzian geometry, even after localizing to small neighborhoods.

TODO: add a picture illustrating this wiggling

DRAFT COPY

© Willie Wai-Yeung Wong

DRAFT COPY

Topic 3 (2023/10/30)

Geodesics and Jacobi Fields

3.1 In the end of the previous lecture, we showed that spacelike geodesics are generally not even locally length minimizers. But our argument hinged upon doing some "wiggling", and the heavily wiggling curves are not C^1 close to the space-like geodesic (since we want the wiggling to be such that the velocity vector is "almost null" at most points). Naturally, one wants to ask whether the process is better if we work only "variationally". For this, we will need the second variation formulae.

3.2 (Second variations) As an aside: in general given a function f defined on a manifold, its "second derivative" is not well-defined as a tensor. At issue is the following observation: letting x and y be local coordinate systems, then we can write

$$\frac{Df}{Dy} = \frac{Df}{Dx} \cdot \frac{Dx}{Dy}$$

but then

$$\frac{D^2 f}{Dy^2} = \frac{Df}{Dx} \cdot \frac{D^2 x}{Dy^2} + \frac{D^2 f}{Dx^2} \cdot \frac{Dx}{Dy} \cdot \frac{Dx}{Dy}$$

does not transform tensorially. Hence to speak of second variations of functions on manifolds, one generally needs to specify a connection. The exception is when the second derivative is evaluated at a critical point; here, the first term that depends on $\frac{D^2x}{Dy^2}$ vanishes.

Bringing it back to our problem at hand, this means that the second variation of $\ell_\pm$ and $\mathcal{E}$ are only meaningful to study at critical points of the functional. Here we will again let $\gamma_s : [a,b] \to M$ be a one-parameter family of curves, with $\alpha = \gamma_0$. Now we will require α to be an affinely-parameterized geodesic with speed $|g(\dot\alpha, \dot\alpha)| = c$. Still we will denote by V the variation $(\frac{\partial}{\partial s}\gamma_s|_{s=0})$. We will now introduce the notion of the *transverse acceleration field*: for each $t \in [a,b]$, we can extend the vector $V|_{\alpha(t)}$ to a vector field along the curve $s \mapsto \gamma_s(t)$. We can take its derivative again (using the Levi-Civita connection as needed) to obtain the vector field A. Finally, given a non-null vector X, we will denote by

$$\mathrm{pr}_X^\perp(V) = V - \frac{g(X,V)}{g(X,X)} X \tag{3.2.1}$$

DRAFT COPY

the projection operator to the orthogonal complement of X.

For the variation of the energy, the formula is almost trivial:

$$\frac{d^2}{ds^2}\mathcal{E}(\gamma_s)\big|_{s=0} = 2\int_a^b g(\dot V, \dot V) - g(\mathrm{Riem}_{V\dot\alpha} V, \dot\alpha) \ \mathrm{d}t + g(\dot\alpha, A)\Big|_a^b. \tag{3.2.2}$$

For variation of the proper time and length, the following formula is due to Synge

$$\frac{d^2}{ds^2}\ell_\pm(\gamma_s)\big|_{s=0} = \pm\frac{1}{c}\int_a^b g(\mathrm{pr}_{\dot\alpha}^\perp \dot V, \mathrm{pr}_{\dot\alpha}^\perp \dot V) - g(\mathrm{Riem}_{V\dot\alpha} V, \dot\alpha)\,\mathrm{d}t \pm \frac{1}{c} g(\dot\alpha, A)\Big|_a^b. \tag{3.2.3}$$

3.3 As a result of the second variation formula, we find the following, in stark contrast to the Riemannian setting.

Proposition

Given (M, g) a Lorentzian manifold of dimension $1+n$, with $n \geq 2$. Let $\alpha : [a,b] \to M$ be a space-like geodesic, then there exists one-parameter families $\gamma_s^\pm$ of curves such that

- *$\gamma_0 = \alpha$;*
- *$\gamma_s^\pm(a) = \alpha(a)$ and $\gamma_s^\pm(b) = \alpha(b)$ for all s;*
- *the corresponding second variation of $\ell_+(\gamma_s^\pm)$ are positive and negative respectively.*

In particular, variationally we see that α is a saddle point *of ℓ_+.* ∎

When $n = 1$ the negative metric gives the same geodesics but with the roles of time-like and space-like reversed. Many analytical properties have special simplifications in this setting.

Proof. The key observation here is that our assumption means that $\dot\alpha$ is space-like, and that its orthogonal complement contains both time-like and space-like vectors. Without loss of generality we can assume $[a,b] = [0,1]$. On $T_{\alpha(1/2)}M$ choose $\tilde V^\pm$ two vectors, with $\tilde V^+$ unit space-like and $\tilde V^-$ unit time-like, and both orthogonal to $\dot\alpha(1/2)$. Extend them to vector fields along α by parallel translation: this guarantees that $\tilde V^\pm$ remain both orthogonal to α and maintain the same causal character and length. For $k \in \mathbb{N}$ to be determined late, for each $t \in [0,1]$, define $V^\pm = \frac{1}{k}\sin(2\pi k t)\tilde V^\pm$; so $V^\pm$ is highly oscillatory in the direction of $\tilde V^\pm$. Choose $\gamma_s^\pm$ so that its variation vector field is $V^\pm$, and such that $\gamma_s^\pm(a) = \alpha(a)$ and $\gamma_s^\pm(b) = \alpha(b)$.

The corresponding second variations can be computed to be

$$\frac{d^2}{ds^2}\ell_+(\gamma_s^\pm)\big|_{s=0} = \frac{1}{c}\int_0^1 (2\pi\cos(2\pi k t))^2 g(\tilde V^\pm, \tilde V^\pm) - \frac{1}{k^2}\sin(2\pi k t)^2 g(\mathrm{Riem}_{\tilde V^\pm \dot\alpha}\tilde V^\pm, \dot\alpha)\,\mathrm{d}t.$$

By taking k sufficiently large, the curvature term is dominated by the first integral, which evaluates to $\pm 2\pi^2$ respectively. □

3.4 Note that the same argument cannot work if α is assumed to be timelike initially, as in this case its orthogonal complement would only have spacelike vectors and we can no longer generate a negative second variation. In fact, this situation makes the time-like case very similar to what happens in Riemannian theory; this turns out to be a general theme in Lorentzian geometry, where nice properties in Riemannian geometry are inherited by time-like (and not space-like) objects in Lorentzian geometry. When this happens invariably it is due to the crucial coercivity / definiteness properties used coming from the orthogonal complement, and not the object itself.

© Willie Wai-Yeung Wong

DRAFT COPY

3.5 Proposition
Every time-like geodesic is locally *a proper-time maximizer.*

More precisely, given (M,g) a Lorentzian manifold and $p \in M$. Let U be a convex normal neighborhood of p and denote by $\tilde{U}$ its corresponding star-shaped subset of T_pM. Then if v is a time-like vector in T_pM, then the curve

$$\gamma : [0,1] \to U, \quad \gamma(s) = \exp_p(sv)$$

is the unique (up to reparametrization) proper-time maximizer *among all smooth regular timelike curves joining p to $\exp_p(v)$ that are entirely contained in U.* ■

Proof. Let $\alpha : [0,1] \to U$ be a smooth regular time-like curve joining p to $\exp_p(sv)$; then α lifts to $\tilde{\alpha} : [0,1] \to \tilde{U}$ a smooth regular curve joining 0 to v. By Lemma 2.11 we have $\tilde{\alpha}|_{(0,1]}$ all are time-like vectors. Consider the function $\rho : [0,1] \to \mathbb{R}$ given by

$$\rho(s) = \sqrt{-g_p(\tilde{\alpha}(s), \tilde{\alpha}(s))}.$$

Its derivative we compute to be

$$\dot{\rho}(s) = -\frac{1}{\rho(s)} g_p(\tilde{\alpha}(s), \dot{\tilde{\alpha}}(s)).$$

(Note that the same argument as the proof of Lemma 2.11 shows that $\dot{\rho}(s) > 0$ for $s > 0$.) Observe that $\tilde{\alpha}(s)/\rho(s)$ is, when $s > 0$, a unit time-like vector; we may decompose

$$\dot{\tilde{\alpha}}(s) = \tilde{v}^{\|}(s) + \tilde{v}^{\perp}(s)$$

the former being exactly $-\frac{1}{\rho(s)^2} g_p(\tilde{\alpha}(s), \dot{\tilde{\alpha}}(s)) \cdot \tilde{\alpha}(s)$ the "radial component". Let $v^{\|}$ and $v^{\perp}$ be their pushforwards to U along γ. By the Gauss Lemma, we find $v^{\perp}$ and $v^{\|}$ to be orthogonal, with $v^{\|}$ time-like. And hence

$$-g_{\alpha(s)}(\dot{\alpha}(s), \dot{\alpha}(s)) \leq -g_{\alpha(s)}(v^{\|}(s), v^{\|}(s)) = -g_p(\tilde{v}^{\|}(s), \tilde{v}^{\|}(s)), = \dot{\rho}(s)^2$$

with equality only when $v^{\perp}(s) = 0$. And so we find

$$\ell_-(\alpha) = \int_0^1 \sqrt{-g(\dot{\alpha}(s), \dot{\alpha}(s))} \leq \int_0^1 \dot{\rho}(s) = \rho(1) = \sqrt{-g_p(v,v)} = \ell_-(\gamma)$$

with equality only when $\dot{\tilde{\alpha}} \parallel \tilde{\alpha}$ which is the same as saying that α is a radial geodesic. □

3.6 Exercise
We say that a curve $\gamma : [a,b] \to M$ is "piecewise regularly smooth and timelike" if there exists a finite set of points $S \subsetneq (a,b)$ such that

- on each connected component of $[a,b] \setminus S$ the curve is smooth with $\dot{\gamma}$ being timelike;
- at each $s \in S$, the limits $\lim_{s^-} \dot{\gamma}$ and $\lim_{s^+} \dot{\gamma}$ both exist, are time-like, and have the same time-orientation.

Prove the following extension of the previous Proposition: "Let U be a convex normal neighborhood, and $p, q \in U$. Let $v = \exp_p^{-1}(q)$. Then up to reparametrization, the geodesic $[0,1] \ni s \mapsto \exp_p(sv)$ is the unique maximizer of the proper time functional, where the maximum is taken over all piecewise regularly smooth and timelike curves connecting p to q that remain within U." ♠

© Willie Wai-Yeung Wong

DRAFT COPY

3.7 (Index form) The previous proposition is only local: a natural question is whether this can be extended to larger intervals, and if not, what are the characterizations? The question turns out to be slightly easier to address variationally: given a time-like geodesic $\gamma : [a,b] \to M$, is it variationally a local maximum of the proper time functional, among all time-like curves that connects $\gamma(a)$ to $\gamma(b)$? The fact that we are fixing the end points means that we can rid ourselves of the transverse acceleration term from (3.2.3). The second variation is then given by a *quadratic form* on vector fields along γ vanishing at the endpoints. We call the corresponding *bilinear form* the *index form*: given γ a time-like geodesic, affinely parameterized, and V, W vector fields along γ that vanish at the endpoints, we write

$$I_\gamma(V,W) := -\int_\gamma g(\mathrm{pr}^\perp_{\dot\gamma}\dot V, \mathrm{pr}^\perp_{\dot\gamma}\dot W) - g(\mathrm{Riem}_{V\dot\gamma}W, \dot\gamma). \tag{3.7.1}$$

Observe that if we replace $V \mapsto V + f\dot\gamma$ for any function f vanishing at the endpoints, we find that

$$\nabla_{\dot\gamma}(V + f\dot\gamma) = \dot V + \dot f\dot\gamma$$

using that γ is affinely parameterized. Its projection is therefore equal to $\mathrm{pr}^\perp_{\dot\gamma}\dot V$. Similarly, we also have

$$\mathrm{Riem}_{(V+f\dot\gamma)\dot\gamma}W = \mathrm{Riem}_{V\dot\gamma}W.$$

And so we see that we can restrict the domain of I_γ to those vector fields that are orthogonal to $\dot\gamma$.

The same argument as Proposition 3.3 shows that there exists V for which $I_\gamma(V,V) < 0$. Our variational problem therefore only concerns whether there exists V for which $I_\gamma(V,V) \geq 0$.

3.8 Exercise

In Lorentzian geometry, the sectional curvature is defined to factor in the signature of the plane. Given a Lorentzian manifold (M,g) and two linearly independent vectors V, W, the *sectional curvature of the plane spanned by V and W* is the quantity

$$K(V,W) := \frac{g(\mathrm{Riem}_{VW}V, W)}{g(V,V)g(W,W) - g(V,W)^2}. \tag{3.8.1}$$

Prove that, if (M,g) has *non-negative sectional curvature*, then every time-like geodesic is a local maximum for the proper time functional, variationally speaking. (In other words, prove that the index form I_γ is negative definite.) (This can be relaxed to requiring only that, for every time-like V and space-like W, the sectional curvature $K(V,W) \geq 0$.) ♠

Examples of Lorentzian manifolds with non-negative sectional curvature include Minkowski space and de Sitter space (see Example 2.18).

Note that the analogous statement in Riemannian geometry requires replacing appropriately with "length functional", "non-positive sectional curvature", and "local minimum".

3.9 (Jacobi field) The form of I_γ in (3.7.1) suggests that looking at the following differential equation may be fruitful:

$$\ddot V - \mathrm{Riem}_{V\dot\gamma}\dot\gamma = 0. \tag{3.9.1}$$

Solutions to this ODE are called *Jacobi fields*, and they also arise as the variation vector fields for the one-parameter family $\gamma_s : [a,b] \to M$, where each γ_s is an affine geodesic. Note that with t being the affine parameter, any vector field of the form $(c_1 t + c_2)\dot\gamma$ is automatically a solution to the equation. Similarly, we can compute

$$\frac{d}{dt} g(\dot V, \dot\gamma) = 0, \qquad \frac{d}{dt} g(V, \dot\gamma) = g(\dot V, \dot\gamma);$$

and so if a Jacobi field and its first derivative are both orthogonal to γ at some point, it will remain so throughout. Finally, note also that this is a *linear* differential equation, and so can be solved as long as the underlying geodesic exists.

3.10 The linearity also allows us to exploit well-known results of linear dynamical systems. For example, the Wronskian between two Jacobi fields is constant.

Lemma
If V, W are two Jacobi fields along an affine geodesic, then $g(\dot V, W) - g(\dot W, V)$ is constant along γ. ■

We leave its proof to the reader. Another interesting statement is

Lemma
Let $\gamma : [a,b] \to M$ be an affine geodesic. Let $V_1, \ldots, V_k$ be Jacobi fields such that $g(V_i, \dot\gamma) = 0$. Suppose $W = \sum \alpha_i V_i$ is a vector field along γ that vanishes both at a and b, then

$$I_\gamma(W,W) = -\sum_{i,j} \int_\gamma g(\dot\alpha_i V_i, \dot\alpha_j V_j) + \dot\alpha_i \alpha_j [g(\dot V_j, V_i) - g(\dot V_i, V_j)]. \tag{3.10.1}$$

■

Proof. We compute

$$\begin{aligned} I_\gamma(W,W) &= -\int_\gamma g(\dot W, \dot W) - g(\mathrm{Riem}_{W\dot\gamma} W, \dot\gamma) = \int_\gamma g(\ddot W, W) - g(\mathrm{Riem}_{W\dot\gamma}\dot\gamma, W) \\ &= \sum \int_\gamma g(\ddot\alpha_i V_i, W) + 2g(\dot\alpha_i \dot V_i, W) + \underbrace{g(\alpha_i \ddot V_i, W) - g(\alpha_i \mathrm{Riem}_{V_i\dot\gamma}\dot\gamma, W)}_{=0} \\ &= \sum \int_\gamma -g(\dot\alpha_i V_i, \dot W) + g(\dot\alpha_i \dot V_i, W) \\ &= \sum_{i,j} \int_\gamma -g(\dot\alpha_i V_i, \dot\alpha_j V_j) + \dot\alpha_i \alpha_j [g(\dot V_i, V_j) - g(\dot V_j, V_i)]. \end{aligned}$$

□

3.11 Definition (Conjugate points)
Given a geodesic γ, we say that two points p and q along γ are conjugate *if there exists a non-trivial Jacobi field along γ that vanishes at both p and q.* ◇

Now given $\gamma : [a,b] \to M$ an affine geodesic, such that $\gamma(a)$ and $\gamma(b)$ are conjugate. Then letting V be the Jacobi field that vanishes at both a and b, we find

$$0 = \int_a^b g(\ddot V, V) - g(\mathrm{Riem}_{V\dot\gamma}\dot\gamma, V) = -\int_a^b g(\dot V, \dot V) - g(\mathrm{Riem}_{V\dot\gamma} V, \dot\gamma) = I_\gamma(V,V)$$

DRAFT COPY

© Willie Wai-Yeung Wong

DRAFT COPY

where we integrated the first term by parts and used the algebraic antisymmetry of the Riemann curvature in the middle equality. We see therefore that the presence of conjugate points stops the index form from being negative definite. In fact, this is a full characterization,

3.12 Theorem
Let $\gamma : [a,b] \to M$ be a time-like geodesic.

- *If there exists no conjugate points of $\gamma(a)$ along γ, then I_γ is negative definite and γ is a variational maximizer of proper time.*
- *If $\gamma(b)$ is the unique conjugate point to $\gamma(a)$, then I_γ is negative semi-definite, but not definite.*
- *If there exists $r \in (a,b)$ such that $\gamma(r)$ is a conjugate point to $\gamma(a)$, then I_γ is indefinite.* ■

We sketch the proof of the first and third claims below; see Theorem 17 in Chapter 10 of O'Neill, *Semi-Riemannian geometry: with applications to relativity* for a complete proof of the entire theorem (which is almost entirely the same as the version in Riemannian geometry).

Sketch of proof of first claim. We can solve the Jacobi equation for $V_1, \dots, V_n$ such that $V_i(a) = 0$, and $\dot V_i(a)$ form a basis of $\{\dot\gamma(a)\}^\perp$ in $T_{\gamma(a)}M$. Note that their pairwise Wronskian vanish. By assumption, no conjugate point exists, and hence on $(a,b]$ the vector fields V_i do not vanish, and must form a basis of $\{\dot\gamma\}^\perp$. Now given any W vanishing at both a and b, we can decompose it in this basis so that $W = \sum \alpha_i V_i$. Applying (3.10.1), and using that γ is time-like, we see that $I_\gamma[W,W] < 0$ unless $\dot\alpha_i \equiv 0$; but the vanishing of W at b requires then $\alpha_i \equiv 0$ and $W \equiv 0$. This proves negative definiteness. □

Sketch of proof of third claim. The main idea is very similar to a trick often used for studying subsolutions in elliptic PDE theory.

We start with a Jacobi field W that vanishes at $\gamma(a)$ and $\gamma(r)$, which necessarily is orthogonal to γ. Since it is non-trivial, we have that $\dot W$ is non-zero at $\gamma(a)$ and $\gamma(r)$, and so we can find a smooth unit space-like vector field U that is orthogonal to γ such that $W \propto U$. We want to show that there exists V such that $I_\gamma(V,V) > 0$, with V vanishing at the boundary. By looking at the definition of index form, it suffices to find V satisfying

$$g(\ddot V, V) - g(\mathrm{Riem}_{V\dot\gamma}\dot\gamma, V) > 0$$

along (a,b). We make, as an ansatz, that $V = fU$, then the problem reduces to the scalar ordinary differential inequality

$$Lf := \ddot f + f\Big(g(\ddot U, U) - g(\mathrm{Riem}_{U\dot\gamma}\dot\gamma, U)\Big) > 0$$

with $f(a) = f(b) = 0$. The existence of W shows that there is a function h satisfying $Lh = 0$ that vanishes at a and r. We assume for simplicity that $h(b) < 0$ (this would be the case if, there are exactly odd number of conjugate points between a and b); the case $h(b) \geq 0$ can be treated with a technical device that is besides the main analytical point. Then setting

$$f = h + \lambda \sinh(\omega t)$$

for ω and λ to be determined, we see that if we choose ω such that ω^2 is a strict upper bound of $|g(\ddot U, U) - g(\mathrm{Riem}_{U\dot\gamma}\dot\gamma, U)|$, then $Lf > 0$. After choosing ω, just set λ so that $f(b) = 0$, and we are done. □

© Willie Wai-Yeung Wong

3.13 The following is an extension of the previous theorem, that applies if we allow one of the base points to run along a hypersurface. The proof is largely similar to what was given above. See also Exercise 2.24.

Theorem

Let (M, g) be a time-orientable Lorentzian manifold, and Σ a spacelike hypersurface, $q \in M \setminus \Sigma$ a point. Denote by $\mathrm{Sh} : T_p\Sigma \to T_p\Sigma$ the future oriented shape operator of Σ (i.e. letting n be the future-oriented unit normal vector field on Σ, we have $\mathrm{Sh}(v) = \nabla_v n$). Suppose $\gamma : [a, b] \to M$ is such that

- *γ is a future-directed unit-speed (hence affine) timelike geodesic;*
- *$\gamma(a) \in \Sigma$ and $\gamma(b) = q$;*
- *$\dot\gamma(a) \perp \Sigma$.*

Suppose further that there is a Jacobi field W satisfying

- *$W(a) \in T_{\gamma(a)}\Sigma$;*
- *$\dot W(a) = \mathrm{Sh}(W(a))$.*
- *$W(r) = 0$ for some $r \in (a, b)$.*

Then there exists some timelike $\tilde\gamma : [a, b] \to M$ with $\tilde\gamma(a) \in \Sigma$, $\tilde\gamma(b) = q$, satisfying $\ell_-(\tilde\gamma) > \ell_-(\gamma)$. ∎

The point $\gamma(r)$ as described by the previous theorem is called a *focal point* of the hypersurface Σ.

3.14 The discussion above highlights the importance of Jacobi fields in understand the behavior of geodesics. Naturally, for studying individual Jacobi fields the sectional curvature plays an important role: in Exercise 3.8 non-negative sectional curvature is associated with time-like geodesics realizing the proper-time between events. However, sectional curvature is too strong a quantity. As Einstein's equation (1.24.1) only involves the *Ricci* and *Scalar* curvatures, one would ideally like to have some sort of control based only on curvature assumptions at the Ricci curvature level. This is particularly significant as it would directly connect behavior of time-like geodesics to statements that we can make concerning the energy momentum tensor T.

One approach to resolving this is by trying to study not a single Jacobi field, but a whole family at once, in a way that would guarantee at least one of the Jacobi fields involved will have a zero.

3.15 The basic observation is this: supposing our space-time is $(1+n)$-dimensional and γ is an affine time-like geodesic, if we let $V_1, \ldots, V_n$ be Jacobi fields that are orthogonal to $\dot\gamma$, then "there exists a non-trivial $\mathbb{R}$-linear combination of $V_1, \ldots, V_n$ that vanish at s" is equivalent to "$V_1 \wedge V_2 \wedge \ldots \wedge V_n$ vanishing at s" (basically: invertible matrices have non-zero determinant). So to probe the presence of zeros, it is useful to look at the wedge product instead.

3.16 (Basic assumptions) The following basic assumptions will be in force in the remainder of this lecture.

- (M, g) is a Lorentzian manifold, of dimension $(1 + n)$.
- $\gamma : (a, b) \to M$ is an affinely parametrized timelike geodesic, with $a < 0 < b$ for convenience.

DRAFT COPY

© Willie Wai-Yeung Wong

DRAFT COPY

- $V_1, V_2, \ldots, V_n$ are Jacobi fields along γ; each is assumed to be orthogonal to $\dot\gamma$, and we further assume that at $s = 0$ they are linearly independent, so that $\dot\gamma(0) \wedge V_1(0) \wedge \cdots \wedge V_n(0) \neq 0$.
- For convenience we will denote by Ω the top-degree form along γ given by $\dot\gamma^\flat \wedge V_1^\flat \wedge \cdots \wedge V_n^\flat$.

Since the space of top-degree forms is one dimensional, there exists a scalar function ϖ along γ such that $\Omega = \varpi \,\mathrm{dvol}_g$, we can assume the orientation has been chosen such that $\varpi(0) > 0$.

3.17 (Relative shape operator) Whenever $\Omega \neq 0$, the vectors $V_1, \ldots, V_n$ form a basis of the orthogonal complement to $\dot\gamma$ in $T_\gamma M$. As γ is affinely geodesic, our assumptions imply that $\dot V_i = \nabla_{\dot\gamma} V_i$ are also orthogonal to $\dot\gamma$, and hence there exists a matrix-valued function B_i^j along γ such that

$$\dot V_i = \sum_j B_i^j V_j. \tag{3.17.1}$$

This mapping defines also a linear transformation B on $\{\dot\gamma\}^\perp$, whenever $\Omega \neq 0$; we call this linear operation the "relative shape operator to the Jacobi fields $\{V_1, \ldots, V_n\}$".

3.18 (Common choices of bases) The description above relies on a particular choice of basis $V_1, \ldots, V_n$ of Jacobi fields. Now, for a timelike curve γ in an $(1+n)$-dimensional Lorentzian manifold, the space of orthogonal solutions to the Jacobi equation is $2n$ (since the equations are second order). How do we choose a basis? Three natural choices exist:

1. The first case is when we aim to study conjugate points. In this case, we are starting at a point $\gamma(a)$ and looking for $\gamma(b)$ that are conjugate to it. In this case, we would want our Jacobi fields all to be vanishing at $\gamma(a)$: so we want $V_1(a) = \cdots = V_n(a) = 0$. We supplement this with a choice of basis $\{W_1, \ldots, W_n\}$ of the orthogonal complement $\{\dot\gamma(a)\}^\perp$ in $T_{\gamma(a)}M$, and set $\dot V_i(a) = W_i$.
2. The second case is when we aim to study focal points; this is the case of Theorem 3.13. In this case we have that the unit-speed γ passes through a spacelike hypersurface Σ at parameter a, and after choosing a basis $\{W_1, \ldots, W_n\}$ of $T_{\gamma(a)}\Sigma$, we will set $V_i(a) = W_i$ and $\dot V_i(a) = \mathrm{Sh}(W_i)$ where Sh is the shape operator of Σ with the same time-orientation as $\dot\gamma(a)$. Notice that in this case the relative shape operator B, at a, is exactly equal to Sh.
3. The third case is a bit of a generalization of the second. Let T denote an affinely geodesic timelike vector field (see Definition 2.3). Let γ be an integral curve, and set $\{W_1, \ldots, W_n\}$ a basis of $\{\dot\gamma(a)\}^\perp$. For each W_i, generate a one-parameter family of geodesics using the integral curves of T, such that its variation at $\gamma(a)$ is exactly W_i; this variation field produces a Jacobi-field along γ.

Such a vector field defines what is called a "timelike geodesic congruence". Its integral curves represent the world-lines of a family of free-falling observers. Thus this notion is useful both for probing the geometry of space-time and for providing a convenient coordinate system for the space-time.

We remark in all three cases the relative shape operator B is *independent of the choice of basis* $\{W_1, \ldots, W_n\}$; this is largely due to the fact that the Jacobi equation is linear. The first two cases is mostly obvious, for the third case see the next exercise.

3.19 Exercise (Relative shape operator for congruences)
In the third case above, let B be the relative shape operator of these Jacobi fields. Prove, along γ, we have that for vector X in $T_{\gamma(s)}M$, that $B(X) = \nabla_X T$. (Note: in particular, this

shows that given T, the definition of B is independent of the choice of $\{W_1, \ldots, W_n\}$.) ♠

3.20 (Evoluton of the relative shape operator) For convenience denote by S the mapping $X \mapsto \mathrm{Riem}_{X\dot\gamma}\dot\gamma$. By the symmetry properties of the Riemann curvature tensor we see that $S(\dot\gamma) = 0$ and $S(X)$ must be orthogonal to $\dot\gamma$, so we can treat S as another linear operator from $\{\dot\gamma\}^\perp$ to itself.

The matrix function B_i^j has an evolution equation: taking the time derivative of its defining equation (3.17.1), and using the Jacobi equation (3.9.1), we find

$$\mathrm{Riem}_{V_i\dot\gamma}\dot\gamma = \sum_j \dot B_i^j V_j + \sum_{j,k} B_i^j B_j^k V_k.$$

Letting S_i^j denote the matrix components of S relative to the basis $\{V_1, \ldots, V_n\}$, we arrive finally at

$$S_i^j = \dot B_i^j + \sum_k B_i^k B_k^j. \tag{3.20.1}$$

3.21 Let's talk about the significance of this relative shape operator to our question of whether Ω (and hence ϖ) becomes zero. Taking the time derivative of the defining equation

$$\dot\gamma^\flat \wedge V_1^\flat \wedge \cdots \wedge V_n^\flat = \varpi \,\mathrm{dvol}_g$$

We find (after using the Leibniz rule and the fact that $\ddot\gamma = 0$, and that the metric volume form is by definition parallel with respect to the Levi-Civita connection):

$$\begin{aligned}\dot\varpi \,\mathrm{dvol}_g &= \dot\gamma^\flat \wedge \sum_j B_1^j V_j^\flat \wedge V_2^\flat \wedge \cdots \wedge V_n^\flat \\ &\quad + \dot\gamma^\flat \wedge V_1 \wedge \sum_j B_2^j V_j^\flat \wedge V_3^\flat \wedge \cdots \wedge V_n^\flat + \cdots + \dot\gamma^\flat \wedge V_1^\flat \wedge \cdots \wedge V_{n-1}^\flat \wedge \sum_j B_n^j V_j^\flat.\end{aligned} \tag{3.21.1}$$

By the antisymmetry of the wedge product we find therefore $\dot\varpi \,\mathrm{dvol}_g = \sum_j B_j^j \varpi \,\mathrm{dvol}_g$ or

$$\frac{\dot\varpi}{\varpi} = \operatorname{tr} B. \tag{3.21.2}$$

3.22 Proposition
Under the assumptions of ¶ 3.16, suppose additionally
- *for an interval $[s_0, s_1] \subseteq (a, b)$ we have $\varpi > 0$ on (s_0, s_1),*
- *the Jacobi fields $\{V_1, \ldots, V_n\}$ are such that their Wronskians $g(\dot V_i, V_j) - g(\dot V_j, V_i) = 0$.*

Then for $s \in \{s_0, s_1\}$ we have that $\varpi(s) = 0$ if and only if the one-sided limit

$$\lim_{\substack{\sigma\to s\\ \sigma\in(s_0,s_1)}} |\operatorname{tr} B(\sigma)| = \infty.$$ ■

If the $\{V_i\}$ are generated according to the first two options in ¶ 3.18, then automatically the Wronskians vanish; in the "conjugate point" case this is because there is a point where all the V_i vanish; in the "focal point" case this is due to the fact that the shape operator is automatically self-adjoint (second fundamental form is symmetric).

DRAFT COPY

DRAFT COPY

Topic 4 (2023/11/6)

More on timelike curves

Raychaudhuri's Equation and Applications

4.1 (Expansion, shear, twist) Before giving the proof of Proposition 3.22, it is convenient to introduce some more terminology by decomposing the operator B. Observe that the space-time metric g induces a positive definite inner product on $\{\dot\gamma\}^\perp$, so we can algebraically decompose

$$B = \frac{1}{n}(\operatorname{tr} B)\cdot \mathrm{Id} + \tilde{B} + \breve{B}, \tag{4.1.1}$$

where

- Id is the identity operator; the scalar $\operatorname{tr} B$ is called the *expansion* of $\{V_1, \dots, V_n\}$;
- observe that $B - \frac{1}{n}(\operatorname{tr} B)\mathrm{Id}$ has zero trace;
- $\tilde{B}$ is the part of $B - \frac{1}{n}(\operatorname{tr} B)\mathrm{Id}$ that is *self-adjoint* relative to the metric g; this is called the *shear* of $\{V_1, \dots, V_n\}$;
- $\breve{B}$ is the part of $B - \frac{1}{n}(\operatorname{tr} B)\mathrm{Id}$ that is *anti-self-adjoint* relative to g; this is called the *twist* of $\{V_1, \dots, V_n\}$.

Notice now

$$B \circ B = \frac{1}{n^2}(\operatorname{tr} B)^2 \mathrm{Id} + \frac{2}{n}(\operatorname{tr} B)(\tilde{B} + \breve{B}) + \tilde{B}\circ\tilde{B} + \breve{B}\circ\breve{B} + \tilde{B}\circ\breve{B} + \breve{B}\circ\tilde{B}.$$

Combining this with the observation that the symmetry properties of the Riemann curvature tensor implies that the operator S is self-adjoint relative to g, we find the following system of equations for the three algebraic parts:

$$\frac{d}{ds}\breve{B} = -\frac{2}{n}(\operatorname{tr} B)\breve{B} - \tilde{B}\circ\breve{B} - \breve{B}\circ\tilde{B}, \tag{4.1.2}$$

$$\frac{d}{ds}(\operatorname{tr} B) = -\mathrm{Ric}_{\dot\gamma\dot\gamma} - \frac{1}{n}(\operatorname{tr} B)^2 - \underbrace{\operatorname{tr}(\tilde{B}\circ\tilde{B})}_{\geq 0} - \underbrace{\operatorname{tr}(\breve{B}\circ\breve{B})}_{\leq 0}, \tag{4.1.3}$$

$$\frac{d}{ds}\tilde{B} = \tilde{S} - \frac{2}{n}(\operatorname{tr} B)\tilde{B} - \tilde{B}\circ\tilde{B} - \breve{B}\circ\breve{B} + \frac{1}{n}\operatorname{tr}(\tilde{B}\circ\tilde{B} + \breve{B}\circ\breve{B})\,\mathrm{Id}. \tag{4.1.4}$$

DRAFT COPY

In the final equation, $\tilde{S}$ is the trace-free part of S.

The equation (4.1.3) is called the *timelike Raychaudhuri equation* and takes a significant role in our understanding of singularity formation in general relativity, since it can be used to provide a monotonicity property. An example of it is in the proof below of Proposition 3.22.

4.2 We remark that $\breve{B}$, being the anti-self-adjoint part, satisfies

$$2g(\breve{B}V_i, V_j) = g(BV_i, V_j) - g(BV_j, V_i) = g(\dot{V}_i, V_j) - g(\dot{V}_j, V_i).$$

And so when the Wronskians of the family $\{V_1, \dots, V_n\}$ vanish, so does the twist.

4.3 Exercise (More on the twist)
Let T be an affinely geodesic timelike vector field, as in Exercise 3.19. Suppose for every integral curve γ we have the corresponding B begin *twist-free*. Prove that T is locally hypersurface orthogonal (meaning that at every point p there exists a hypersurface Σ through p such that T is normal to Σ along the whole hypersurface).

Hint: Frobenius' Theorem. ♠

Proof of Proposition 3.22. Notice that our assumptions imply that $\varpi\, \mathrm{dvol}_g$ is a smooth top-degree form along γ (since the Jacobi equation is linear and can be always globally solved along any geodesic). With $\varpi > 0$ on (s_0, s_1), we have necessarily $\operatorname{tr} B$ is a smooth function on (s_0, s_1).

($\Leftarrow$) Near s our assumption that $\operatorname{tr} B$ diverges implies it is never zero. Hence we can rewrite (3.21.2) to read $|\varpi| = \frac{|\dot{\varpi}|}{|\operatorname{tr} B|}$. Taking the limit and using the smoothness of ϖ implies the result. (Note that this step does not use the assumption on the Wronskians.)

($\Rightarrow$) Without loss of generality assume $s = s_1$. Integrating (3.21.2) gives, for any $\sigma < \sigma'$ in (s_0, s_1),

$$\ln \circ \varpi(\sigma') - \ln \circ \varpi(\sigma) = \int_\sigma^{\sigma'} \operatorname{tr} B.$$

Taking $\sigma' \nearrow s = s_1$ we see that $\int_\sigma^{s_1} \operatorname{tr} B$ must diverge to $-\infty$, which requires that $\liminf \operatorname{tr} B = -\infty$.

Our assumptions on the Wronskians imply that $\breve{B} \equiv 0$, and hence (4.1.3) gives

$$\frac{d}{d\sigma} \operatorname{tr} B \leq -\mathrm{Ric}_{\dot\gamma\dot\gamma} - \frac{1}{n} (\operatorname{tr} B)^2.$$

As $[s_0, s_1]$ is a compact set, we have that the Ricci curvature is bounded on it. So there exists some number $m \geq 0$ such that

$$\frac{d}{d\sigma} \operatorname{tr} B \leq m - \frac{1}{n} (\operatorname{tr} B)^2.$$

But this implies that once $|\operatorname{tr} B|$ exceeds $\sqrt{nm}$ the function $\operatorname{tr} B$ becomes monotonically decreasing. This means that we can upgrade $\liminf \operatorname{tr} B = -\infty$ to $\lim \operatorname{tr} B = -\infty$. ☐

© Willie Wai-Yeung Wong

DRAFT COPY

4.4 (Interpreting Raychaudhuri's equation) Let's return to the expression (4.1.3), which we copy again here:

$$\frac{d}{ds}(\operatorname{tr} B) = -\mathrm{Ric}_{\dot\gamma\dot\gamma} - \frac{1}{n}(\operatorname{tr} B)^2 - \underbrace{\operatorname{tr}(\tilde B \circ \tilde B)}_{\geq 0} - \underbrace{\operatorname{tr}(\breve B \circ \breve B)}_{\leq 0},$$

It turns out that this equation has an interpretation based on largely physically obvious notions. Recall that Jacobi fields represent variation fields of infinitesimally nearby geodesics. So the picture to keep in your head is a central geodesic γ with a "bundle" of geodesics surrounding it. We can interpret this "bundle" as the trajectories of a family of observers (say we normalize the geodesics to be unit speed).

- The form $\Omega = \varpi \operatorname{dvol}_g$ is the amount of volume occupied by our observers; in other words, we should think of $\frac{1}{\varpi}$ as the *density* of local observers in our "bundle". From (3.21.2), we see that the expansion $\operatorname{tr} B$ is the (negative) logarithmic rate of change of this density, justifying the terminology.
- Overall, the matrix B describes how the distribution of observers deform over time, relative to the current distribution. Thus the shear $\tilde B$ captures the shear (volume-preserving) deformation of the "bundle" over time, and the twist $\breve B$ captures how the "bundle" twists (or how nearby observers rotate / orbit around γ).

With these interpretations, then several of the terms in (4.1.3) become clear:

- the presence of the $(\operatorname{tr} B)^2$ term represents the expectation that if a family of observers in free fall starts out drifting towards each other, than this focussing effect will continue (and in fact be reinforced). We see this for observers in flat spacetime too: if a family of observers move uniformly towards each other in free fall, then the volume occupied will shrink at the rate $(T-t)^n$, whose log derivative $\frac{n}{T-t}$ solves exactly the expression $\frac{d}{dt}\frac{n}{T-t} = -\frac{1}{n}\left(\frac{n}{T-t}\right)^2$.
- that the twist or rotation $\breve B$ slows down the volume decrease can be regarded as a sort of *centrifugal force* in the rotating frame.
- finally, through Einstein's equation, Ric is related to the energy-momentum tensor. And so the inclusion of the $\mathrm{Ric}_{\dot\gamma\dot\gamma}$ term *is the only term that directly captures the acceleration coming from matter-generated gravitational effects*.

4.5 Our everyday understanding of gravity is that "gravity only pulls, not pushes". Examining Raychaudhuri's equation, we see that we would want conditions on the energy momentum T that guarantees that matter contribute by *accelerating* the convergence of the "bundle" of observers. Motivated by these types of considerations, physicists proposed several different notions that would capture this idea.

Definition (Energy conditions)
Given (M,g) an $(n+1)$-dimensional Lorentzian manifold equipped with an energy momentum tensor T, we say that T satisfies the

Strong energy condition *if for any timelike vector X, we have $T(X,X) - \frac{1}{n-1}\operatorname{tr}_g(T) g(X,X) \geq 0$; in the absence of cosmological constant this is equivalent to requiring* $\mathrm{Ric}(X,X) \geq 0$.

Dominant energy condition *if for any pair of causal vectors X, Y with the same time orientation, $T(X,Y) \geq 0$.*

Weak energy condition *if for any timelike vector X, we have $T(X,X) \geq 0$.*

Null energy condition *if for any null vector X, we have $T(X,X) \geq 0$.* ◊

4.6 Exercise (Relation between the energy conditions)

1. Prove that if T satisfies the strong energy condition then it satisfies the null energy condition.
2. Prove that if T satisfies the dominant energy condition then it satisfies the weak energy condition.
3. Prove that if T satisfies the weak energy condition then it satisfies the null energy condition.
4. Prove that the above three are the only *general* relationships between the energy conditions by exhibiting explicit counterexamples. (Hint: it is enough to search for your counterexamples among diagonal T.) ♠

4.7 Under the strong energy condition, we see that the contribution of the spacetime Ricci curvature (which is tied *only* to the matter content of the universe) is signed. This allows us to provide, in some cases, an upper bound of the proper time elapsed before timelike geodesics encounter a focal point.

Theorem (Singularity Theorem, version 0)
Let (M,g) be a $(1+n)$-dimensional time-orientable Lorentzian manifold satisfying $\mathrm{Ric}(X,X) \geq 0$ *for all time-like X, and Σ a space-like hypersurface. Suppose there exists a constant $C > 0$ such that the future oriented shape operator (see Theorem 3.13)* $\mathrm{Sh} : T_p\Sigma \to T_p\Sigma$ *of Σ is such that* $\mathrm{tr}(\mathrm{Sh}) \leq -C$ *at every point on Σ. Let $\gamma : [a,b) \to M$ be a timelike geodesic satisfying*

- *$\gamma(a) \in \Sigma$;*
- *$\dot\gamma(a)$ is the future-directed unit timelike normal to Σ at $\gamma(a)$;*
- *$b - a > n/C$.*

Then there is a focal point of Σ along γ. ■

Proof. Returning to ¶ 3.18, option 2, we build a basis of $\{V_i\}$ of Jacobi fields along γ satisfying $\dot V_i(a) = \mathrm{Sh}(V_i(a))$; we find the operator $B = \mathrm{Sh}$, and hence the twist vanishes (as the shape operator is always self-adjoint). Therefore Raychaudhuri's equation (4.1.3), together with the energy condition (i.e. assumption on Ricci curvature) implies

$$\frac{d}{d\sigma}\,\mathrm{tr}\,B \leq -\frac{1}{n}(\mathrm{tr}\,B)^2.$$

Integrating this differential inequality we find

$$\left(-\frac{1}{\mathrm{tr}\,B}\right)\Big|_a^r \leq \frac{a-r}{n} \implies -\frac{1}{\mathrm{tr}\,B(r)} \leq \underbrace{-\frac{1}{\mathrm{tr}\,B(a)}}_{\leq \frac{1}{C}} + \frac{a-r}{n},$$

and hence $\mathrm{tr}\,B \searrow -\infty$ at some $r > a$ with $r - a \leq n/C$. By Proposition 3.22 a non-trivial linear combination of V_i vanishes at r, and $\gamma(r)$ is a focal point of Σ (see Theorem 3.13). □

DRAFT COPY

More about Proper Time, a Singularity Theorem

4.8 As mentioned earlier, between two points space-like curves can be drawn with arbitrarily small and arbitrarily large length, so the notion of a "distance" between two points cannot be defined in analogy to what happens in the Riemannian setting. Proposition 3.5, on the other hand, shows that at least locally the same construction can potentially be meaningful when proper time is used instead. Here we formalize that with some definitions.

Definition (Proper time between two events)
Let (M,g) be a Lorentzian manifold, and $p,q \in M$.

- *We denote by $\mathsf{CsPth}_{\text{reg}}(p,q)$ the set of all smooth regular causal curves $\gamma : [a,b] \to M$ such that $\gamma(a) = p$, $\gamma(b) = q$;*
- *We say that the* proper time *between p,q is*

$$\mathscr{L}_-(p,q) := \sup_{\gamma \in \mathsf{CsPth}_{\text{reg}}(p,q)} \ell_-(\gamma) \in [0,\infty]. \tag{4.8.1}$$

(Recall that if $\mathsf{CsPth}_{\text{Reg}}(p,q) = \emptyset$ then as usual we set the supremum to be $\min[0,\infty] = 0$.*)* ◊

4.9 To streamline some of the discussions below, it is convenient to introduce some terminology capturing "causal relations" between pairs of points in a Lorentzian manifold (M,g). We will return to a more detailed discussion of these relations in the next section.

4.10 Definition
Let (M,g) be a time-orientable Lorentzian manifold. We shall denote:

- *$\mathbb{I}^+ \subseteq TM$ for all future timelike vectors, and $\mathbb{I}^-$ those past timelike vectors; we will use the notation $\mathbb{I}^\pm_p$ for those corresponding vectors in T_pM.*
- *Similarly, $\mathbb{J}^\pm$ will denote those future/past causal vectors.*

Given two points p and q, we write:

- *$p \lneqq q$ if there exists a smooth regular curve $\gamma : [a,b] \to M$, such that $\gamma(a) = p$, $\gamma(b) = q$, and $\dot\gamma \in \mathbb{I}^+$;*
- *$p \leqq q$ if there exists a smooth regular curve $\gamma : [a,b] \to M$, such that $\gamma(a) = p$, $\gamma(b) = q$, and $\dot\gamma \in \mathbb{J}^+$.*

We further denote

- *$\mathcal{I}^+(p) := \{x \in M \mid p \lneqq x\}$, and $\mathcal{I}^-(p) := \{x \in M \mid x \lneqq p\}$; and similarly*
- *$\mathcal{J}^+(p) := \{x \in M \mid p \leqq x\}$ and $\mathcal{J}^-(p) := \{x \in M \mid x \leqq p\}$.* ◊

4.11 The proper time function (see Definition 4.8) does *not* satisfy the axioms of the metric space. While it is by definition symmetric in the points p,q, it is not positive definite, and it rather strongly fails the triangle inequality. In fact, we have the following:

Lemma (Reverse triangle inequality for proper time)
Suppose $p,q,r \in M$ are such that $p \leqq q \leqq r$, then

$$\mathscr{L}_-(p,r) \geq \mathscr{L}_-(p,q) + \mathscr{L}_-(q,r). \tag{4.11.1}$$

■

© Willie Wai-Yeung Wong

DRAFT COPY

Sketch of proof. The basic idea is the following analogue of "corner cutting" in Euclidean geometry: in Euclidean geometry, if you have two (smooth) curves joined at some point p so that the two curves form a non-flat angle there, then by "cutting the corner" one can shorten the total distance. It turns out that the same process, when applied to time-like curves in Lorentzian geometry, can reliably increase (not necessarily strictly) the proper time elapsed. From this idea the lemma follows immediately. We defer a more detailed discussion of corner cutting to Proposition 5.8 and Lemma 5.9 below. □

4.12 Theorem (Singularity Theorem, version 1)
*Let (M, g) be a time-orientable $(1+n)$-dimensional Lorentzian manifold, satisfying $\mathrm{Ric}(X, X) \geq 0$ for every timelike X. Let Σ be a space-like hypersurface in M, and suppose the following **technical assumptions hold**:*

1. *M can be written as the disjoint union $\Sigma \sqcup M_+ \sqcup M_-$ where*

$$M_\pm = \cup_{q\in\Sigma} \mathcal{I}^\pm(q).$$

2. *For every $p \in M_+$, there exists a timelike geodesic $\gamma : [a, b] \to M$ with*
   - $\gamma(a) \in \Sigma$, $\dot\gamma(a) \perp T_{\gamma(a)}\Sigma$;
   - $\gamma(b) = p$;
   - $\ell_-(\gamma) = \sup_{q\in\Sigma} \mathscr{L}_-(p, q)$.

If furthermore there exists $C > 0$ such that the future oriented shape operator Sh of Σ satisfies $\mathrm{tr(Sh)} < -C$ everywhere on Σ, then (M, g) is timelike geodesically incomplete. ■

Proof. In fact, we will prove that there exists *no* complete timelike geodesics through Σ. Suppose $\eta : [a, b') \to M$ is a maximally extended unit-speed timelike geodesic with $\eta(a) \in \Sigma$ and $\dot\eta$ future directed. We first observe then $\eta((a, b')) \subseteq M_+$. By assumption, for every $s \in [a, b')$ there exists a timelike geodesic $\gamma : [a, b] \to M$ with $\gamma(a) \in \Sigma$, $\gamma(b) = \eta(s)$, and such that γ realizes the maximum of the proper time between $q \in \Sigma$ and $\eta(s)$. By reparametrizing γ we may assume it is unit-speed. And hence we have

$$b - a = \ell_-(\gamma) \geq \ell_-(\eta|_{[a,s]}) = s - a.$$

On the other hand, for $\ell_-(\gamma)$ to maximize the proper time, by Theorem 3.13 there cannot be a focal point of Σ along γ. By Theorem 4.7, we see that this requires $b \leq n/C$. Returning to the curve η, we see that this requires $s - a \leq n/C$. Since s is arbitrary, this means that $b' \leq n/C + a$. And hence η is incomplete. □

4.13 Theorem 4.12 is the prototype for the *Hawking singularity theorems*. Ignoring the technical assumptions for a moment, these types of theorem roughly states that "if the universe has an instant (in the sense of a space-like cross section) that is everywhere expanding (contracting), then there must be a singularity toward the past (future)." Singularity here is in a very weak sense: it just states that the universe cannot be geodesically complete. In particular, it does not tell us about why it is that the geodesic cannot be continued beyond proper time n/C from the specified instant. Nevertheless, this theorem is often cited as the justification of the "big bang".

DRAFT COPY

4.14 Example
To illustrate the issue with the interpretation of geodesic incompleteness as singularities: let M be the subset of Minkowski space

$$M := \{(t,x) \in \mathbb{R}^{1+3} \mid t < -\sqrt{(x^1)^2 + (x^2)^2 + (x^3)^2}\};$$

in other words, M is the interior of the past light cone emanating form the origin. Equip M with the induced Minkowski metric. Let $\Sigma := \{t = -\sqrt{1 + (x^1)^2 + (x^2)^2 + (x^3)^2}\}$ be the hypersurface in question. One can check pretty straightforwardly that M and Σ meets the requirements of Theorem 4.12; and indeed, to the future of Σ, no time-like geodesic has proper time more than 1, and hence M is geodesically incomplete. But the geodesic incompleteness is fundamentally due to the fact that M is a proper subset of $\mathbb{R}^{1+3}$ which *is* geodesically complete. One would be hard-pressed to say that M has any sort of intrinsic "singularity". ◊

4.15 The various version of Hawking singularity theorems one can find in the literature are attempts to replace the "technical assumptions" by more reasonable assumptions one could make of the spacetime. A particular very natural assumption that implies the technical assumption listed above is called "global hyperbolicity"; time permitting we will describe this notion in more detail in a later lecture.

Finally, note that our proof shows that a large family of timelike geodesics must all be incomplete. One may ask the question: "can the hypothesis be weakened if one only wishes to show the existence of one incomplete timelike geodesic?" It turns out the answer is yes: in fact, we can even replace the assumption that Σ is "expanding" (or "contracting", depending on the sign of $\mathrm{tr}(\mathrm{Sh})$) with the assumption that Σ is compact and a genericity condition that says, roughly speaking, $\mathrm{Ric}(X,X) > 0$ strictly for "most" points in $\mathbb{I}^{\pm}$. I refer the reader to §8.2 in Stephen William Hawking and Ellis, *The large scale structure of space-time*.

4.16 Returning to the fact that the hypotheses of Theorem 4.12 guarantee that all timelike geodesics through Σ cannot be complete, one is faced with the expectation that the existence of a complete timelike geodesic, under the assumption that it has some proper time maximization properties, should be a very special case indeed. This expectation is realized in the following theorem.

Theorem (Lorentzian splitting theorem)
Let (M,g) be a Lorentzian manifold. Suppose the following conditions holds:
1. *At least one of the following hold:*
   - *(M,g) is globally hyperbolic;*
   - *(M,g) is timelike geodesically complete.*
2. $\mathrm{Ric}(X,X) \geq 0$ *for all timelike X.*
3. *There exists a unit speed timelike geodesic $\gamma : \mathbb{R} \to M$ such that for every $a,b \in \mathbb{R}$ we have $|b-a| = \mathscr{L}_-(\gamma(b),\gamma(a))$.*

Them M splits: in particular, there exists a complete Riemannian manifold (Σ,h) such that $M = \mathbb{R} \times \Sigma$ with $g = -\,\mathrm{d}t^2 + h$. ■

DRAFT COPY

DRAFT COPY

Basic ideas of the proof. The full proof is quite technical, and can be found in Beem, Ehrlich, and Easley, *Global Lorentzian geometry*, Galloway, "The Lorentzian splitting theorem without the completeness assumption", and Flores, "The Riemannian and Lorentzian Splitting Theorems". In one proof, the basic approach is through the use of Busemann functions, similar to the proof of the Riemannian splitting theorem.

Given the exceptional geodesic γ, we can define the Busemann function $b_\pm : M \to \overline{\mathbb{R}}$ by $b_\pm(p) = \lim_{s\to\pm\infty} |s| - \mathscr{L}_-(\gamma(s), p)$; this converges thanks to the reverse triangle inequality Lemma 4.11. If one were to do this operation with the t-axis in Minkowski space, then one would find that the level sets of the Busemann functions are exactly the "constant t hyperplanes".

To replicate this on general manifolds with the Ricci curvature sign condition, in the Riemannian setting one first proves that both $b_\pm$ are weakly subharmonic, and hence pinching the two against each other both are harmonic. Then one uses the Bochner formula to show that the gradients $\nabla b_\pm$ are parallel vector fields, from which the theorem would follow using de Rham's decomposition theorem.

In the Lorentzian setting, the final step is not a problem, as it was shown by Wu that de Rham's decomposition theorem carries over with basically no change. The main difficulty is that comparison results such as subharmonicity and Bochner formula cannot be used when the metric is indefinite. The main trick is to resolve this by doing the comparison analysis on certain well-chosen space-like hypersurfaces. □

Intro to Causal Theory

4.17 In this section we investigate more the properties of the causal relations defined in Definition 4.10. Throughout we shall assume that (M, g) is also *time-orientable* (see Definition 1.18); in particular, the set of all causal vectors can be decomposed into those that are future-directed and those that are past-directed.

4.18 (Basic comments) In spite of the notation, $\lneqq$ and $\leqq$ do not, in general, define any sort of ordering. In general (for example: time-periodic universes) it is possible for both $p \lneqq q$ and $q \lneqq p$ to be true. Our definition where we require the causal curves to be non-trivial also prevents "$p \leqq p$" to hold, and hence $p \notin \mathcal{J}^\pm(p)$.

It is easy to see, from our definitions, that

$$\mathbb{I}^\pm \subsetneq \mathbb{J}^\pm; \quad \overline{\mathbb{I}^\pm} = \mathbb{J}^\pm \cup \{0\}; \tag{4.18.1}$$

here $\{0\}$ refers to the zero-section of TM. It is also easy to see that $\mathcal{I}^\pm(p) \subseteq \mathcal{J}^\pm(p)$. But in general it is possible for the closure of $\mathcal{I}^\pm(p)$ to contain elements not in $\mathcal{J}^\pm(p) \cup \{p\}$.

4.19 Example
Let $M = \mathbb{R}^{1,1} \setminus \{0\}$, the two dimensional Minkowski space with the origin removed; we can still use the standard coordinates (t, x) to refer to points in M. Let $p = (-1, -1)$, then $\mathcal{I}^+(p) = \{(t, x) \mid t + 1 > |x + 1|\}$, and so its closure would include the entire line $\{(s, s) \mid s > 0\}$. But none of those elements belong in $\mathcal{J}^+(p)$ (in $\mathbb{R}^{1,1}$ there is exactly one causal curve connection p to, say, $(1, 1)$, and that causal curve passes through the origin. So by removing the origin from M, there no longer exist any causal curve from p to $(1, 1)$). ◊

© Willie Wai-Yeung Wong

4.20 To better understand the relation between $\mathcal{I}^{\pm}$ and $\mathcal{J}^{\pm}$, we need to do some manipulations of curves. We will capture the needed results in a series of lemmas.

4.21 Lemma
Let $\gamma_u : [a,b] \to M$ be a one parameter family of smooth regular curves. Let $V = \frac{\partial}{\partial u}\gamma_u|_{u=0}$ be the variational vector field along γ_0. Suppose:

- *γ_0 is causal;*
- *$g(\dot V, \dot\gamma_0) < 0$*

then there exists some $\epsilon > 0$ such that for all $u \in (0,\epsilon)$, the curve γ_u is time-like. ■

Proof. Consider the quantity $K(u,s) = g(\dot\gamma_u(s), \dot\gamma_u(s))$. Our first assumption shows $K(0,s) \leq 0$. Then

$$\frac{\partial}{\partial u}K(u,s) = 2g(\dot\gamma_u(s), \frac{\partial}{\partial u}\frac{\partial}{\partial s}\gamma_u(s)) = 2g(\dot\gamma_u(s), \frac{\partial}{\partial s}\frac{\partial}{\partial u}\gamma_u(s))$$

and in particular

$$\frac{\partial}{\partial u}K(0,s) = 2g(\dot\gamma_0(s), \dot V(s)) < 0.$$

The result follows. □

DRAFT COPY

DRAFT COPY

Topic 5 (2023/11/13)

Causal Theory

5.1 In this lecture we continue our earlier discussion on causal theory.

5.2 Corollary
If $\gamma : [a,b] \to M$ is a smooth regular causal curve such that $\dot\gamma$ is future time-like at at least one point, then $\gamma(a) \lneq \gamma(b)$. ■

Proof. By continuity we can assume that there exists $c \in (a,b)$ such that $\dot\gamma(c)$ is future time-like. Let V be the vector field along γ formed by parallel transporting $\dot\gamma(c)$. By continuity again, there exists a subinterval $[a',b']$ along which $\dot\gamma$ is future-time-like. Consider $W = \phi V$ for a function ϕ. Observe that $\dot W = \dot\phi V$ since V is parallel along γ. Choose ϕ such that
- $\phi(a) = \phi(b) = 0$,
- $\dot\phi > 0$ on $[a,a']$ and $[b',b]$; so on those intervals $g(\dot W, \dot\gamma) < 0$ (using that $\dot W$ is time-like future directed, and $\dot\gamma$ is causal future directed).

Next build a one-parameter family of variations γ_u with variational field W, by the previously lemma for all sufficiently small positive u the curve γ_u is timelike on $[a,a']$ and $[b,b']$. On the other hand, since γ_0 is already time-like on $[a',b']$, for sufficiently small u we also have γ_u is timelike on $[a',b']$. □

5.3 Corollary
If $\gamma : [a,b] \to M$ is a smooth regular null *curve with $\dot\gamma \in \mathbb{J}^+$, that is* not *everywhere geodesic, then $\gamma(a) \lneq \gamma(b)$.* ■

Proof. Since γ is a null curve, we have $g(\dot\gamma,\dot\gamma) = 0$; taking derivative again we find

$$g(\ddot\gamma,\dot\gamma) = 0$$

Since $\dot\gamma$ is causal, this means that either $\ddot\gamma$ is causal and parallel to $\dot\gamma$, or that it is space-like. In particular, since we assumed that γ is *not* a geodesic, we must have $g(\ddot\gamma,\ddot\gamma) > 0$ somewhere along γ.

Taking the derivative once more, we find

$$g(\ddot\gamma,\ddot\gamma) + g(\dddot\gamma,\dot\gamma) = 0 \implies g(\dddot\gamma,\dot\gamma) \le 0.$$

DRAFT COPY

Now, choose a parallel future time-like vector field V along γ. The above discussion prepared us to choose, for some pair of functions ϕ and ψ vanishing at a and b

$$W = \phi V + \psi \ddot{\gamma}. \tag{5.3.1}$$

To apply our Lemma, we need to arrange

$$g(\dot{W}, \dot{\gamma}) = g(\dot{\phi} V + \dot{\psi} \ddot{\gamma} + \psi \dddot{\gamma}, \dot{\gamma}) < 0.$$

Our previous computations show that it suffices

$$\dot{\phi} g(V, \dot{\gamma}) - \psi g(\ddot{\gamma}, \ddot{\gamma}) < 0. \tag{5.3.2}$$

We now proceed similarly to the proof of Corollary 5.2. Let $[a', b']$ be a subinterval on which $g(\ddot{\gamma}, \ddot{\gamma}) > 0$. Choose ϕ so that $\dot{\phi} > 0$ on $[a, a']$ and $[b', b]$. Next choose ψ to be non-negative, such that on $[a', b']$

$$\psi > \frac{\dot{\phi} g(V, \dot{\gamma})}{g(\ddot{\gamma}, \ddot{\gamma})}.$$

The conclusion follows. □

5.4 Putting together the two previous Corollaries, we arrive at:

Proposition
$q \in \mathcal{J}^+(p) \setminus \mathcal{I}^+(p)$, if and only if the set of smooth causal curves connecting p to q is non-empty and consists only of null geodesics. ■

5.5 Exercise
Let (M, g) be Lorentzian, and $\gamma : [0, b] \to M$ a future-directed null geodesic. Let $\mu : (-1, 1) \to M$ be a smooth curve, such that $\mu(0) = \gamma(0)$ and $g(\dot{\mu}0, \dot{\gamma}(0)) \neq 0$. Prove that for every $\epsilon > 0$, there exists some $\delta \in (-\epsilon, \epsilon)$ such that $\mu(\delta) \lneqq \gamma(b)$.

(Hint: use Lemma 4.21 by choosing a V wisely.) ♠

5.6 Our next theorem concerns the transitivity of the relations $\lneqq$ and $\leqq$. While the results are intuitively obvious, the proof requires some technical tricks, and we will present them in a series of Lemmas.

Theorem
Given (M, g) a Lorentzian manifold, suppose $p, q, r \in M$ satisfy $p \leqq q$ and $q \leqq r$, then $p \leqq r$. Furthermore, $r \in \mathcal{J}^+(p) \setminus \mathcal{I}^+(p)$ if and only if all of the following are true:
- *there exists only one causal curve γ_1 joining p to q, and that curve is a null geodesic;*
- *there exists only one causal curve γ_2 joining q to r, and that curve is a null geodesic;*
- *γ_2 and γ_1 point in the same direction at q.* ■

5.7 Lemma
If (M, g) is a two-dimensional Lorentzian manifold, near every point there is a local coordinate system (u, v) with respect of which the metric takes the form $\Omega(\mathrm{d}u \otimes \mathrm{d}v + \mathrm{d}v \otimes \mathrm{d}u)$ for some non-vanishing function Ω. ■

Proof. There are exactly two null directions in TM for a two-dimensional Lorentzian manifold; locally one can therefore choose (uniquely, up to swapping labels and scalar multiplication) two linearly independent non-vanishing null vector fields L and N. Their integral curves locally form one dimensional foliations that intersect transversely. Select one integral curve of each of the two fields, call them λ and ν respectively. Define u as follows: along λ set u to be the parameter for the curve λ (as an integral curve), and require $N(u) = 0$. Define v similarly: along ν set v to be the parameter for ν, and require $L(v) = 0$. One checks then the level sets of u and v are null, and hence the metric takes the desired form. □

5.8 Proposition
Theorem 5.6 holds for 2-dimensional Lorentzian manifolds. ■

Proof. First we prove that $p \leqq r$. Let γ and η be smooth causal curves that witness $p \leqq q$ and $q \leqq r$ respectively. Take a local coordinate system of the form given by Lemma 5.7 centered at q, such that $q = (0, 0)$. Assume without loss of generality that $\Omega < 0$. Assume without loss of generality that our parametrization is such that $\gamma(0) = \eta(0) = q$, and that $\gamma|_{[-1,0]}$ and $\eta|_{[0,1]}$ maps into our local coordinate system.

Now let γ^u, γ^v, η^u, and η^v denote the components of γ and η in our coordinate system. That γ, η are future-directed causal translates to the four functions γ^u, γ^v, η^u, and η^v all being monotonically increasing. Choose now a smooth function $\chi : [-1, 1] \to [-1, 0]$ satisfying

- $\chi' \geq 0$, with $\chi'(s) = 0 \iff s \geq \frac{1}{2}$
- $\chi|_{[-1,-1/2]}(s) = s$
- $\chi|_{[1/2,0]} = 0$

Set $\zeta : [-1, 1] \to M$ as

$$\zeta(s) = \gamma(\chi(s)) + \eta(-\chi(-s)).$$

(With addition performed using the coordinate system.) Then

- ζ is smooth
- $\zeta|_{[-1,-1/2]} = \gamma$ and $\zeta|_{[1/2,1]} = \eta$
- $\dot\zeta(s) = \dot\gamma(\chi(s))\chi'(s) + \dot\eta(-\chi(s))\chi'(-s)$ is a linear combination of $\dot\gamma$ and $\dot\eta$ with non-negative coefficients at least one of which is positive.

This last condition implies that ζ is a regular curve, and that the components ζ^u and ζ^v are all monotonically increasing. Hence ζ is a smooth causal curve that is future-directed.

Next we characterize the case that $r \in \mathcal{J}^+(p) \setminus \mathcal{I}^+(p)$. By Proposition 5.4, the negation of the three required conditions imply at least one of the following holds:

- $p \gneqq q$
- $q \gneqq r$
- the curves γ and η are both null geodesics, but point in different directions at q.

DRAFT COPY

© Willie Wai-Yeung Wong

DRAFT COPY

In the first case, we can assume that γ is time-like, and hence both $\dot\gamma^u$ and $\dot\gamma^v$ are everywhere positive. This implies that ζ is time-like when $s < -\frac{1}{2}$, and hence by Corollary 5.2 we can perturb ζ to be everywhere time-like implying that $p \lneqq r$. The second case is treated in exactly the same way.

In the third case, we can assume without loss of generality that $\dot\gamma^u = \dot\eta^v = 0$ everywhere, but both $\dot\gamma^v$ and $\dot\eta^u$ are everywhere positive. Then we see from our construction that $\dot\zeta^u$ and $\dot\zeta^v$ are both positive on $(-1/2, 1/2)$, which implies in particular that $\dot\zeta(0)$ is time-like. Using Corollary 5.2 we again see that $p \lneqq r$. □

5.9 Lemma
If $p \lneqq q$ and $q \lneqq r$, then $p \lneqq r$. ■

Proof. Let γ and η be the witnesses again. Construct ζ as in the previous proof, except this time,

- instead of having χ transition on $[-1/2, 1/2]$, it transitions on $[-\epsilon, \epsilon]$ for some small ϵ;
- instead of using a double-null coordinate system, we just work in a convex normal neighborhood of q.

By choosing ϵ sufficiently small, we have that for $s \in (-\epsilon, \epsilon)$

- $\zeta(s)$ can be required to live in a small coordinate neighborhood of q;
- $\dot\gamma(\chi(s)) = \dot\gamma(0) + o(\epsilon)$
- $\dot\eta(-\chi(-s)) = \dot\eta(0) + o(\epsilon)$
- and hence $\dot\zeta(s)$ is, up to an $o(\epsilon)$ error, always a convex combination of $\dot\gamma(0)$ and $\dot\eta(0)$.

Continuity and the fact that being time-like is an open condition implies then for $s \in (-\epsilon, \epsilon)$, the ζ thus constructed is still time-like. This proves the claim. □

Proof of Theorem 5.6. In view of Lemma 5.9, it suffices to consider the case where at least one of $q \in \mathcal{J}^+(p) \setminus \mathcal{I}^+(p)$ and $r \in \mathcal{J}^+(q) \setminus \mathcal{I}^+(q)$ is true.

$q \in \mathcal{J}^+(p) \setminus \mathcal{I}^+(p)$ and $q \lneqq r$. By Proposition 5.4 we can let γ be a null geodesic joining p to q, and η a time-like curve joining q to r. Note that it suffices to show that for some r' on η between q and r we have $p \lneqq r'$, as then we can apply Lemma 5.9 to the points p, r', r.

By Lemma 2.11 we can find some r' on η and a time-like geodesic joining q to r'. So by an abuse of notation we shall now assume that η is the geodesic joining q to r'. Consider the geodesic normal coordinate system centered at q. We have γ and η are now radial lines. Let Π be the plane spanned by those two directions; Π represents a two-dimensional submanifold of M. As it contains η, a time-like curve, we have that Π is Lorentzian (at least near q). We can therefore apply Proposition 5.8 to γ and η on the two-dimensional manifold Π to conclude that there exists a time-like curve (in Π) connecting p to r'; as the metric on Π is the induced metric from M this means that the same curve is still time-like in M.

$r \in \mathcal{J}^+(q) \setminus \mathcal{I}^+(q)$ and $p \lneqq q$. This case is analogous to the previous, and omitted.

Both $q \in \mathcal{J}^+(p) \setminus \mathcal{I}^+(p)$ and $r \in \mathcal{J}^+(q) \setminus \mathcal{I}^+(q)$. Let γ and η be the two null geodesics. If $\dot\gamma$ and $\dot\eta$ are parallel at q, then they join up to a null geodesic, showing $p \leqq r$. It suffices to show that if the two are not parallel at q, then $p \lneqq r$. But now we can proceed as before: when the two are not parallel, using geodesic normal coordinates we can find a two dimensional Lorentzian submanifold near q that contains both γ and η, but then Proposition 5.8 kicks in and we are again done. □

© Willie Wai-Yeung Wong

DRAFT COPY

5.10 Theorem
Let (M,g) be a Lorentzian manifold, and $p \in M$. Then
1. *$\mathcal{I}^+(p)$ is open;*
2. *The interior of $\mathcal{J}^+(p)$ is $\mathcal{I}^+(p)$;*
3. *$\mathcal{J}^+(p) \cup \{p\} \subseteq \overline{\mathcal{I}^+(p)}$;*
4. *Equality of the previous point holding if and only if $\mathcal{J}^+(p) \cup \{p\}$ is closed.* ■

Proof. We prove each claim in turn.
1. Let $p \lneqq q$, then there exists a timelike curve γ joining p to q. Take a convex normal neighborhood U of q. Let q' be a point along γ, such that $q' \lneqq q$ and $q' \in U$. By definition there exists a star-shaped open set $\tilde{U} \subseteq T_{q'}M$ such that $\exp_{q'} : \tilde{U} \to U$ is a diffeomorphism. The subset $\tilde{V} = \tilde{U} \cap \mathbb{I}^+_{q'}$ is also open, and contains $\exp^{-1}_{q'}(q)$. Let $V = \exp_{q'}(\tilde{V})$; it is open. Observe that for any $r \in V$, the geodesic joining q' to r is future directed and time-like, hence $q' \leqq r$. By Theorem 5.6 we find $p \lneqq r$, or that the open set $V \subseteq \mathcal{I}^+(p)$.
2. Let q be an interior point of $\mathcal{J}^+(p)$, then there exists a convex normal neighborhood U of q in $\mathcal{J}^+(p)$. In particular, there exists an element $q' \in U$ such that the geodesic joining q to q' is past directed and time-like. But as $p \leqq q'$ and $q' \lneqq q$, Theorem 5.6 implies $q \in \mathcal{I}^+(p)$. This shows one inclusion.
   The other inclusion follows immediately after noting that $\mathcal{I}^+(p)$ is open and $\mathcal{I}^+(p) \subseteq \mathcal{J}^+(p)$ by definition, as timelike curves are causal.
3. That $p \in \overline{\mathcal{I}^+(p)}$ is obvious.
   Given $q \in \mathcal{J}^+(p)$, let U be a convex normal neighborhood of q, and $\tilde{U}$ the corresponding star shaped set in T_qM. We see that any neighborhood of 0 in $\tilde{U}$ contains elements of $\mathbb{I}^+_q$, which maps to elements r satisfying $q \lneqq r$ and hence $r \in \mathcal{I}^+(p)$. This shows that any open neighborhood of q intersects $\mathcal{I}^+(p)$ as needed.
4. Trivial. (Recall the characterization of the closure of a set being the smallest closed set containing said set.) □

5.11 The results above gives a pretty satisfactory discussion on the causal implications when one perturbs a causal curve that is *not* a null geodesic. Our next topic studies the perturbation of null geodesics.

5.12 (Lack of projections) Before we start, however, I will mention a conceptual stumbling block that often trips up new students to Lorentzian geometry, namely that in contrast with the analysis of timelike or spacelike vectors and subspaces, in the null case there is a lack of a canonical projection operator. When $v \in T_pM$ is a timelike or a spacelike vector, we have a well-defined notion of orthogonal projection to the direction of v, given by

$$X \mapsto \frac{g(v,X)}{g(v,v)} v.$$

This formula obviously fails when v is null.

More spectacularly, even the concept of an "orthogonal projection" is problematic when v is a null vector. One way to understand what orthogonal projection is doing (when

© Willie Wai-Yeung Wong

DRAFT COPY

v is timelike or spacelike) is the decomposition

$$X = X^{\parallel} + X^{\perp}$$

where $X^{\parallel}$ is in the span of $\{v\}$ and $X^{\perp}$ satisfies $g(X^{\perp}, v) = 0$. However, as null vectors are orthogonal to themselves, such a decomposition when v is null, even when it exists, cannot possibly be unique.

However, this points to a possible way out: note that given $v \in T_pM$, the non-degeneracy of the metric implies that the set $N_v = \{X \in T_pM \mid g(X, v) = 0\}$ is a co-dimension 1 subspace. What we are missing, compared to the timelike and spacelike settings, is a canonical choice of a vector that is transverse to this subspace: in the timelike and spacelike settings, since $g(v, v) \neq 0$ we find v is transverse to N_v, and in a sense it is the fact that $v \in N_v$ that is driving the difficulty in the null case.

5.13 (To gauge or not to gauge) In view of the above diagnosis, there are two natural ways of handling things, depending on the end goal.

1. The first is to *restrict* our consideration only to those vectors in N_v. This approach especially makes sense when the theory is consistent, in the sense that the naturally defined operations only send vectors from N_v to other vectors in N_v.
2. The second is to impose a gauge condition: given a null vector v, we can choose a vector k that is transverse to N_v. *With respect to k* we can then uniquely decompose every $X \in T_pM$ as $\alpha k + X^{\perp}$, where $\alpha \in \mathbb{R}$ and $X^{\perp} \in N_v$. We even have the convenient formula
$$\alpha = \frac{g(X, v)}{g(k, v)}.$$
This option becomes necessary if the natural objects that we need to study may not be guaranteed to lie in N_v.

Keep these in mind as you consider the discussions below.

5.14 Recall that Jacobi fields V along an affinely parametrized geodesic γ satisfy

$$\ddot{V} = \mathrm{Riem}_{V\dot{\gamma}}\dot{\gamma},$$

and that Jacobi fields can be generated as infinitesimal variations: given a one parameter family of affinely parametrized geodesics γ_s, the vector field $\partial_s \gamma_s|_{s=0}$ is a Jacobi field along γ_0. Taking the scalar product again $\dot{\gamma}$ we have

$$\frac{d}{dt} g(\dot{V}, \dot{\gamma}) = 0, \qquad \frac{d}{dt} g(V, \dot{\gamma}) = g(\dot{V}, \dot{\gamma}). \tag{5.14.1}$$

5.15 In the timelike case, we can decompose each Jacobi field using orthogonal projection to a portion parallel to $\dot{\gamma}$ and a portion orthogonal to it, with the parallel portion automatically taking the form $(c_1 + c_2 t)\dot{\gamma}$. As discussed above, this is no longer possible when $\dot{\gamma}$ is null. Nevertheless, much information can still be gleaned if we just *restrict* our attention only to those Jacobi fields that are always orthogonal to γ, as we will see below.

© Willie Wai-Yeung Wong

DRAFT COPY

5.16 We begin with an analogue of Theorem 3.12, this time for null geodesics. In that Theorem, we found that a time-like geodesic $\gamma : [a,b] \to M$ is variationally the local maximizer of proper-time if and only if there does *not* exist a conjugate point $c \in (a,b]$ to a along γ. Assuming the underlying ideas carry through, then starting with a null geodesic (which has zero proper time) $\gamma : [a,b] \to M$ and a conjugate point c to a along γ, we should be able to generate a curve with *positive* proper time connecting $\gamma(a)$ to $\gamma(b)$, thereby showing that $a \lneqq b$, strengthening the hypothesis that $a \leqq b$.

5.17 First we see that it suffices to only consider those variations that are orthogonal to γ for this problem.

Lemma
Suppose $\gamma_s : [a,b] \to M$ is a one-parameter family of smooth regular curves, with γ_0 an affine null geodesic and $\gamma_s(a) = \gamma_0(a)$ and $\gamma_s(b) = \gamma_0(b)$. If there exists a sequence $s_i \to 0$ such that each γ_{s_i} is a timelike curve, then the variational vector field $V = \partial_s \gamma_s|_{s=0}$ is orthogonal to γ_0. ■

Proof. Without loss of generality and by taking a subsequence, we can assume that s_i decreases to zero; our assumption that γ_0 is null also means s_i is never 0. We have along the sequence $g(\dot\gamma_{s_i}, \dot\gamma_{s_i}) < 0$. Observe that as a result

$$0 \geq \frac{d}{ds} g(\dot\gamma_s, \dot\gamma_s)|_{s=0} = 2g(\dot\gamma_0, \dot V).$$

Integrating from a to b we find, on the other hand

$$\int_a^b g(\dot\gamma_0, \dot V) = g(\dot\gamma_0, V)\Big|_a^b = 0$$

by our assumption that V vanishes at the end points. For a signed function to integral to zero it must be everywhere zero, and hence we find $g(\dot\gamma_0, \dot V) = 0$. But since γ_0 is an affine null geodesic, we have $\frac{d}{dt} g(\dot\gamma_0, V) = g(\dot\gamma_0, \dot V) = 0$, and the desired result follows. □

5.18 (Interpretation) The previous lemma, if you think about it, is slightly counter-intuitive: why is it that to change from timelike curves to a null curve the limiting variation must be non-timelike (recall that no timelike vectors can be orthogonal to a causal vector)? To this we can give two explanations:

1. The rigidity property is strongly dependent on the fact that our family of variations *fixes the end points*. Indeed, even in Minkowski space we can see situation where null geodesics get perturbed to timelike curves via timelike variations: start with the curve in $\mathbb{R}^{1+1}$ given by $x = t$. The family of curves $x = mt$ with $m < 1$ approach $x = t$ as $m \nearrow 1$. Only that in these settings we cannot fix both endpoints.
2. A better mental picture is that of *gravitational lensing*. Imagine a strong enough gravitational well (say, a black hole even) where light emerging from a point refocuses back at a later time due to the gravitational bending of the spacetime. In this case one can easily imagine spacetime trajectories taken by timelike observers (not necessarily inertial) between when-where the light is emitted and where the light refocusses.

Indeed, the later case is prototypical of how we understand conjugate points and focal points of null geodesics.

© Willie Wai-Yeung Wong

DRAFT COPY

5.19 Lemma 5.17 should be regarded as a counterpoint to Lemma 4.21. In the start of this lecture we used the latter Lemma extensively to perturb causal curves to timelike curves. Here we see a limitation to this method, that it can never work if the starting curve is a null geodesic. On the other hand, this means that for null geodesics we need to take advantage of the second variation. To do so, we need to ensure that for a given pair (V, A) of vector fields along γ, we can realize them as the first and second variations of a one-parameter family. This we can accomplish in general using the exponential map.

More precisely, let $\gamma : [a,b] \to M$ be a smooth regular curve. Let V and A be vector fields along γ that vanish on the end points. Then we can define

$$\gamma_s : [a,b] \to M, \qquad \gamma_s(t) = \exp_{\gamma(t)}\Big(sV(t) + \frac{1}{2}s^2 A(t)\Big). \tag{5.19.1}$$

Since the exponential map is smooth and depends smoothly on the base point, we have $\gamma_s(t)$ is a smooth one-parameter family of curves. As $\dot\gamma \neq 0$, by continuity the same holds for all sufficiently small s. And taking the s derivative (for fixed t) is simple using the definition of the exponential map.

5.20 For the one parameter family of curves γ_s with first variation $V = \partial_s \gamma_s|_{s=0}$ and acceleration $A = \partial^2_{ss}\gamma_s|_{s=0}$, now let us compute the second derivative (see also (3.2.2))

$$\frac{d^2}{ds^2} g(\dot\gamma_s, \dot\gamma_s)|_{s=0} = 2g(\dot V, \dot V) - 2g(\mathrm{Riem}_{V\dot\gamma_0} V, \dot\gamma_0) + 2g(\dot\gamma_0, \dot A).$$

Our goal then is to find two vector fields V and A along our null affine geodesic γ such that

- V and A vanish at the endpoints of γ;
- V is orthogonal to γ;
- $g(\dot V, \dot V) - g(\mathrm{Riem}_{V\dot\gamma} V, \dot\gamma) + g(\dot\gamma, \dot A) < 0$.

Were we successful in finding such vector fields, then the one-parameter family given by (5.19.1) will be such that all sufficiently small s yields γ_s that is timelike on (a,b).

5.21 Theorem
Let $\gamma : [a,b] \to M$ be a future-directed affine null geodesic. Suppose there exists a conjugate point $c \in (a,b)$ to a along γ, then $\gamma(a) \lneqq \gamma(b)$. ∎

Proof. It suffices to show that there exists some $d \in (c,b]$ such that $\gamma(a) \lneqq \gamma(d)$.

Let J be the Jacobi field along γ that vanishes at a and c. We may also assume that c is the first conjugate point, so that J is non-zero between (a,c).

(1) J must be orthogonal to γ. Furthermore, J is space-like on (a,c).

By (5.14.1), since J vanishes at both a and c and $g(\dot J, \dot\gamma)$ is constant, we must have that $g(\dot J, \dot\gamma) \equiv 0$ which implies then $g(J, \dot\gamma) = 0$. Since γ is null, the only non-space-like vector in the orthogonal complement of $\dot\gamma$ is $\dot\gamma$ itself. Suppose $J(c') = \kappa\dot\gamma(c')$ at some $c' \in (a,c)$, then noting that $\tilde J(s) = \frac{s-a}{c'-a}\kappa\dot\gamma(s)$ is another Jacobi field, we find that either $J = \tilde J$, or that $J - \tilde J$ is another non-trivial Jacobi field that vanishes at $c' < c$. The first case is ruled out as $\tilde J(c) \neq 0$ while $J(c) = 0$ by assumption. The second case is ruled out as we assumed c is the first conjugate point.

(2) There exists a unit-length space-like vector field U along γ such that J is in the direction of U everywhere.

The same argument as part (1) shows that the claim holds on any subinterval on which J is non-vanishing, simply by taking the unit vector in the J direction. It suffices to glue across points where J is zero. But this is doable as necessarily $\dot J \neq 0$ whenever $J = 0$, and so we can use Malgrange's Preparation Theorem to conclude. Note that we can choose an orientation for U such that $g(J,U) > 0$ on (a,c).

(3) Strategy for finding V and A.

Recall our goal is to make the quantity $g(\dot V, \dot V) - g(\mathrm{Riem}_{V\dot\gamma}V, \dot\gamma) + g(\dot A, \dot\gamma) < 0$. Recall further that the Jacobi field J satisfies $-g(J, \ddot J) - g(\mathrm{Riem}_{J\dot\gamma}J, \dot\gamma) = 0$. We can rewrite our goal as

$$\frac{d}{dt}\Big[g(\dot V, V) + g(A, \dot\gamma)\Big] - g(V, \ddot V) - g(\mathrm{Riem}_{V\dot\gamma}V, \dot\gamma) < 0.$$

So provided we can find a $d \in (c,b)$ and a V such that

- $g(V, \ddot V) + g(\mathrm{Riem}_{V\dot\gamma}V, \dot\gamma) > 0$ on (a,d);
- $V(a) = V(d) = 0$;

then if we choose A to be any vector field satisfying $g(A, \dot\gamma) = -g(\dot V, V)$, we would be done. This final step can be completed by first choosing a vector K at $\gamma(a)$ satisfying $g(K, \dot\gamma(a)) = -1$, extend it to a vector field along γ by parallel transport, and setting $A = g(\dot V, V)K$.

(4) Finding a V.

We shall proceed very similarly to the proof of part 3 of Theorem 3.12. By way of an *ansatz* we shall seek a vector field V that is in the direction of the unit vector U. More precisely, we can assume

$$V = hU$$

for some real valued function h. The expression

$$g(V, \ddot V) + g(\mathrm{Riem}_{V\dot\gamma}V, \dot\gamma) = h\Big[g(U, \ddot h U + 2\dot h \dot U + h\ddot U) + hg(\mathrm{Riem}_{U\dot\gamma}U, \dot\gamma)\Big].$$

Using that U is unit and hence $g(U, \dot U) = 0$, we see that for our construction it suffices we find a scalar function h on $[a,b]$ and a $d \in (c,b]$, satisfying the following properties:

- $Lh := \ddot h + h\Big[g(U, \ddot U) + g(\mathrm{Riem}_{U\dot\gamma}U, \dot\gamma)\Big] > 0$ on (a,d);
- $h(a) = h(d) = 0$.

We note that the function $h_0 = g(J,U)$ satisfies $Lh_0 = 0$ and vanishes at both a and c. Furthermore, as J is a non-trivial Jacobi field we have $\dot J$ is non-zero at a and c; having chosen U to be oriented such that $h_0 > 0$ on (a,c), this means that $\dot h_0 > 0$ at a and $\dot h_0 < 0$ at c. By continuity, this means that $h_0 < 0$ on $(c, d']$ for some $d' \in (c,b]$.

We will use a standard trick of elliptic PDEs to build h from h_0. Consider the ansatz

$$h(s) = \lambda \sinh(k(s-a)) + h_0$$

for some λ and k to be specified. We can compute

$$Lh = \lambda \sinh(k(s-a))\Big[k^2 + g(U, \ddot U) + g(\mathrm{Riem}_{U\dot\gamma}U, \dot\gamma)\Big].$$

DRAFT COPY

The terms involving U is a C^∞ function on the compact set $[a,b]$, and hence is uniformly bounded. Therefore for all k sufficiently large, we have $Lh > 0$ whenever $s > a$. After fixing k, by choosing λ sufficiently small, we can ensure that

$$h(d') = \lambda \sinh(k(d' - a)) + h_0(d') < 0.$$

On the other hand, as sinh is positive on the positive reals, and $h_0 > 0$ on (a,c), we see that $h > 0$ on $(a,c]$. We can therefore choose $d = \inf\{s \in (a,d'] \mid h(s) \leq 0\}$ and our construction is done. $\square$

DRAFT COPY

DRAFT COPY

Topic 6 (2023/11/27)

Null Geometry

Null Hypersurfaces

6.1 Previously we were able to extend Theorem 3.12 concerning time-like geodesics joining points p and q and arrive at Theorem 3.13 concerning the proper time between a space-like hypersurface Σ and a point q. It turns out a similar sort of extension is also possible in the null case for Theorem 5.21. However, with null objects in play, several of the statements need technical modifications before they can even make sense (for example, the definition of the shape operator for space-like hypersurfaces is based on the derivative of its *unit* normal vector field; but in the case of null vector fields we cannot normalize, which causes an issue). In preparation, in this section we discuss some of the basic relevant properties of null hypersurfaces.

6.2 To recall: given an $(1+n)$-dimensional Lorentzian manifold (M, g), we say that the n-dimensional hypersurface Σ is a *null* hypersurface if one (and hence both) of the following equivalent conditions hold:

- Σ is locally given as the 0-level set of a defining function f, with the property that df is a null covector along Σ.
- The restriction of the metric g to $T\Sigma$ is degenerate.

One can check that the following is true: taking the defining function f, its gradient $(df)^\sharp$ is tangent to Σ, and is the unique degenerate direction of the restriction of the metric g to $T\Sigma$.

Sketch of proof. First, on any subspace of T_pM the metric can have at most one degenerate direction. To see this, suppose v and w are degenerate directions. This would imply that $g(v,v) = g(w,w) = g(v,w) = 0$. By Exercise 1.17 this implies v, w are collinear.

Now, by definition: $(df)((df)^\sharp) = g^{-1}(df, df) = 0$, and hence $(df)^\sharp \in T\Sigma$ is a tangent vector. Next, given any tangent vector to Σ, we have $g(V, (df)^\sharp) = V(f) = 0$ since f is a defining function. Hence $(df)^\sharp$ is a degenerate direction. □

An important consequence of this definition is:

6.3 Proposition (Null hypersurfaces are ruled)
If Σ is a null hypersurface and (M,g) is time-orientable, then Σ is the union of null geodesic segments. ■

Proof. What we will prove is that there exists a vector field L along Σ, that is tangent to Σ, and such that when considered as a vector field in M, L is geodesic. Then we can decompose Σ into the integral curves of L.

Locally select a defining function f; by replacing f by $-f$ if necessary, we can assume that $(df)^\sharp$ is future-oriented. First we claim that $(df)^\sharp$ is geodesic. The same computations as in Example 2.5 shows that $g(V, \nabla_{(df)^\sharp}(df)^\sharp) = 0$ *for any V that is tangent to* Σ (but not necessarily transverse vector fields, since $g^{-1}(df, df)$ can vanish to order 1 only at Σ). This implies that $\nabla_{(df)^\sharp}(df)^\sharp$ is orthogonal to $T\Sigma$, and hence must live in the span of $(df)^\sharp$, rendering $(df)^\sharp$ geodesic.

Observe now that if f and f' are two defining functions on overlapping domains, then the uniqueness of the degenerate direction of $g|_{T\Sigma}$ implies that $(df)^\sharp$ and $(df')^\sharp$ must be collinear on the overlap. And since we chose them both to be future-oriented, they differ by a positive factor. From here, a standard partition of unity argument allows us to build L. □

6.4 As a consequence of the construction above, we see that the object that we really want is a non-vanishing section of $T\Sigma$ that is degenerate relative to the restricted metric g. We give such vector fields a name.

Definition
Given a null hypersurface Σ, we say that L is a null geodesic generator of Σ *if L is a nonvanishing section of $T\Sigma$ that satisfies $g(L,L) = 0$.* ◊

In addition to being automatically a geodesic vector field, we note that the null geodesic generator is also *orthogonal* to the null hypersurface Σ, with respect to the space-time metric g.

6.5 Example
One way to *build* a null hypersurface is then by "shooting out" a family of null geodesics. Here we give two specific examples of this process. Throughout let (M,g) be a fixed Lorentzian manifold.

1. Given $p \in M$, let $\tilde{U} \subseteq T_pM$ be an open set on which $\exp_p$ is a diffeomorphism. If we denote by
$$\tilde{C} := \{v \in \tilde{U} \mid v \neq 0, g_p(v,v) = 0\}$$
then $C := \exp_p(\tilde{C})$ is a null hypersurface. It is ruled by geodesic segments emanating from p.
2. Given S a co-dimension 2 submanifold of M, with TS containing only space-like vectors, we see that at $p \in S$ there exists two null directions in T_pM that are orthogonal to S (since T_pS is spacelike, its orthogonal complement in T_pM is time-like and two dimensional, so contains exactly two null directions). Assume that there exists a vector field L along S such that L is orthogonal to S and $g(L,L) = 0$. Then we

© Willie Wai-Yeung Wong

DRAFT COPY

can generate a null hypersurface by starting on S and launching a geodesic in the direction of L.

Note that in the second construction, co-dimension 2 is necessary, as space-like hypersurfaces do not admit null normal vectors. $\diamond$

6.6 Definition
The second case of Example 6.5 is used frequently enough that we would like to formally define the construction. Let (M, g) be a fixed Lorentzian manifold with dimension $1+n$. Suppose Σ is a null hypersurface, and L a geodesic generator. Let $\mathring{\Sigma} \subsetneq \Sigma$ is a submanifold with dimension $n-1$. We say that the pair $(\mathring{\Sigma}, L|_{\mathring{\Sigma}})$ generates *Σ, or that $\mathring{\Sigma}$ is* a spatial section *of Σ, if every maximally extended integral curve of L on Σ intersects $\mathring{\Sigma}$ exactly once.* $\diamond$

6.7 Note that in spite of the name used, we did *not* assume that $T\mathring{\Sigma}$ is space-like. This turns out to be a requirement from the definition:

Lemma
Let Σ be a null hypersurface, and $\mathring{\Sigma}$ a submanifold that is transverse to the null geodesic generators of Σ, then $\mathring{\Sigma}$ is space-like. $\blacksquare$

Proof. If $v \in T_p\mathring{\Sigma}$, then $v \in T_p\Sigma$ and so v is orthogonal to the null geodesic generator of Σ. Suppose v is not zero, so v is either space-like or v is causal. By Exercise 1.17, in the latter case v must be parallel to the null geodesic generator of Σ, which is impossible as we assumed $\mathring{\Sigma}$ is transverse to said generators. $\square$

6.8 The following related lemma is obvious and we omit the proof. (Roughly: choose a space-like hypersurface through p, and set $\mathring{\Sigma}$ to be its intersection with Σ.)

Lemma (Existence of local spatial sections)
Let Σ be a null hypersurface and $p \in \Sigma$. Then there exists an open set U and a space-like submanifold $\mathring{\Sigma}$ with $\dim(\mathring{\Sigma}) = \dim(\Sigma) - 1$ *such that $\mathring{\Sigma}$ is a spatial section of $U \cap \Sigma$.* $\blacksquare$

6.9 Given a hypersurface, in Riemannian geometry a key concept is its extrinsic geometry, defined by the second fundamental form, which is the normal projection of ambient connection. However, with null objects, as discussed in ¶ 5.12, we don't have a normal projection. So the notion of a second fundamental form does not really make sense. On the other hand, the notion of a shape operator, which is defined by the connection acting on the unit normal vector field of a hypersurface, does survive in a form.

The terminology "null second fundamental form" that one sees in the literature really refer to something more akin to the null shape operator to be defined below.

Before giving the definition, we first perform some computations. Let V be a tangent vector field to Σ, and L a null geodesic generator, we find

$$g(\nabla_V L, L) = \frac{1}{2} V(g(L,L)) = 0.$$

And hence we see that $\nabla_V L$ is also a tangent vector field. Similarly,

$$g(\nabla_L V, L) = L(g(V,L)) - g(V, \nabla_L L) = 0$$

and so $\nabla_L V$ is also tangent to Σ.

© Willie Wai-Yeung Wong

DRAFT COPY

DRAFT COPY

Now, given a function h, let's compute

$$\nabla_{V+hL} L = \nabla_V L + h \nabla_L L.$$

This tells us that if V and W are equal up to a multiple of L, then $\nabla_V L$ and $\nabla_W L$ are equal up to a multiple of L. Similarly, we also have

$$\nabla_V (hL) = h \nabla_V L + V(h) L.$$

These computations ensure that the formula for $\mathring{\mathrm{Sh}}$ given in the next definition is well-defined.

6.10 Definition (Null shape operator)
Let (M,g) be a Lorentzian manifold, and Σ be a null hypersurface.

1. *We denote by $N_p\Sigma$ the unique one dimensional subspace of $T_p\Sigma$ corresponding to the degenerate direction of the restricted metric g; we write $N\Sigma$ for the corresponding vector bundle. We call it the* null subspace.
2. *We denote by $\mathring{T}_p\Sigma := T_p\Sigma / N_p\Sigma$ the quotient, and $\mathring{T}\Sigma$ the corresponding vector bundle. Given a vector $v \in T_p\Sigma$, we write $[v]$ for its equivalence class in $\mathring{T}_p\Sigma$.*
3. *By the* null shape operator *we refer to* $\mathring{\mathrm{Sh}}$, *which is a section of $N^*\Sigma \otimes \mathring{T}^*\Sigma \otimes \mathring{T}\Sigma$, satisfying*

$$\mathring{\mathrm{Sh}}(L, [v]) = [\nabla_v L].$$

Here L is a section of $N\Sigma$ and v a section of $T\Sigma$. ◇

In the literature, this is also sometimes called the null Wirtinger operator.

6.11 Exercise
Let Σ be a null hypersurface.

1. Prove that if $v, w, v', w' \in T_p\Sigma$ are such that $v' \in [v]$ and $w' \in [w]$, then $g(v,w) = g(v',w')$. Conclude from this that g restricts to a well-defined scalar product $\mathring{g}$ on $\mathring{T}_p\Sigma$.
2. Prove that $\mathring{g}$ as defined above is positive definite.
3. Fix a choice of null geodesic generator L. Prove that $\mathring{\mathrm{Sh}}(L, \text{-}\,)$ is self-adjoint relative to $\mathring{g}$. ♠

6.12 The null shape operator is closely related to the extrinsic geometry of the spatial cross sections of a null hypersurface. We have the following simple formula:

Proposition
Let (M,g) be a Lorentzian manifold, Σ a null hypersurface, and L a null geodesic generator of Σ. Take $p \in \Sigma$, and $\mathring{\Sigma}$ a local spatial cross section (as in Lemma 6.8) through p. Denote by II *the (vector-valued) second fundamental form of $\mathring{\Sigma}$ as a submanifold of M (so given V, W sections of $T\mathring{\Sigma}$ we have $\mathrm{II}(V,W)$ is the component of $\nabla_V W$ orthogonal to $\mathring{\Sigma}$). Then for v, w tangent to $\mathring{\Sigma}$ we have*

$$\mathring{g}(\mathring{\mathrm{Sh}}(L,[v]),[w]) = -g(\mathrm{II}(v,w), L).$$ ■

Proof. Observe that

$$g(\mathrm{II}(v,w), L) = g(\nabla_v w, L) = \nabla_v (g(w,L)) - g(w, \nabla_v L).$$

Since $L \perp w$ by assumption, the result follows. □

DRAFT COPY

Null Jacobi Fields

6.13 As it turns out, many of the discussions above do not require having a full null hypersurface; we just need such a thing "infinitesimally". This is similar to how in the proof of Theorem 5.21, we only consider those Jacobi fields that remain always orthogonal to a fixed null geodesic γ.

6.14 (Local null structure) Now let (M, g) be a Lorentzian manifold and γ an affine null geodesic. Similar to Definition 6.10, we can introduce the notations for $p \in \gamma$

- $N_p\gamma = \operatorname{span}\{\dot\gamma\} \subseteq T_pM$;
- $T_p^\perp\gamma = \{v \in T_pM \mid g(v, \dot\gamma) = 0\}$;
- $\mathring{T}_p\gamma := T_p^\perp\gamma / N_p\gamma$; we will also use $[v]$ to denote the equivalence class of $v \in T_p^\perp\gamma$.

It turns out this splitting is compatible with a lot of geometry.

6.15 Proposition

1. *The space-time metric g induces a positive definite inner product $\mathring{g}$ on $\mathring{T}_p\gamma$.*
2. *The space-time Levi-Civita connection ∇ induces a linear connection $\mathring{\nabla}$ on the vector bundle $\mathring{T}\gamma$, with respect to which $\mathring{g}$ is parallel.*
3. *The space-time Riemman curvature tensor induces a linear mapping $\mathring{S} : \mathring{T}_p\gamma \to \mathring{T}_p\gamma$ that satisfies*

$$[\mathrm{Riem}_{v\dot\gamma}\dot\gamma] = \mathring{S}([v]). \qquad \blacksquare$$

Proof. The first statement is trivial. For the third statement it suffices to note that by virtue of the symmetries of the Riemann curvature tensor, for any vector v, we have $\mathrm{Riem}_{v\dot\gamma}\dot\gamma$ is orthogonal to $\dot\gamma$. Note additionally that $\mathrm{Riem}_{\dot\gamma\dot\gamma}\dot\gamma = 0$, then the claim follows from linearity.

We focus our attention on the second claim. First observe that the base γ is one dimensional, so we are only looking at $\nabla_{\dot\gamma}$. Using that parallel transport preserves orthogonality, we have that the spacetime ∇ restricts to a linear connection on $T^\perp\gamma$. Now, for any vector field of the form $f(t)\dot\gamma(t)$ we have $\nabla_{\dot\gamma}(f\dot\gamma) = \dot f\dot\gamma \propto \dot\gamma$. This implies that for any two sections v, w of $T^\perp\gamma$, we have that $[v] = [w] \implies [\nabla_{\dot\gamma}v] = [\nabla_{\dot\gamma}w]$; hence this factors through a linear connection $\mathring{\nabla}$ on the quotient. Metric compatibility follows immediately from the fact that g is compatible with ∇. $\square$

6.16 Earlier, we saw that null hypersurfaces are ruled by null geodesics. As the Jacobi fields represent variations of one parameter families of geodesics, they represent infinitesimal versions of null hypersurfaces near the geodesic γ.

6.17 Definition (Null Jacobi Fields, Relative Null Shape Operator)
Given (M, g) an $(1+n)$-dimensional Lorentzian manifold and γ an affine null geodesic.

- *By a* null Jacobi field *we mean a Jacobi field that is a section of $T^\perp\gamma$.*
- *Given $V_1, \ldots, V_{n-1}$ null Jacobi fields along γ. If t is such that $[V_1], \ldots, [V_{n-1}]$ are linearly independent at $\gamma(t)$, then there exists a linear operator $\mathring{B} : \mathring{T}_{\gamma(t)}\gamma \to \mathring{T}_{\gamma(t)}\gamma$ satisfying*

$$\mathring{B}([V_i]) = [\dot V_i].$$

We call $\mathring{B}$ the relative null shape operator of $\{V_1, \ldots, V_n\}$ at t. $\diamond$

Observe that if a Jacobi field V is such that, at some t_0, both $V(t_0)$ and $\dot V(t_0) \in T^\perp_{\gamma(t_0)}\gamma$, then V is a section of $T^\perp\gamma$ (exactly the same as in the timelike case). Hence the set of all null Jacobi fields form a $(2n)$-dimensional $\mathbb{R}$-vector space.

© Willie Wai-Yeung Wong

6.18 (Comment on the terminology) Given Σ a null hypersurface, we can take L to be an affine null geodesic generator. Fix γ one of the integral curves. The other integral curves of L near γ induces null Jacobi fields $V_1, \dots, V_{n-1}$ along γ as their variations. Note that the choice $V_1, \dots, V_{n-1}$ is not unique: if we fix p along γ, for each set $\{W_1, \dots, W_{n-1}\}$ of $T_p^\perp \gamma$ vectors such that $\{W_1, \dots, W_{n-1}, \dot\gamma\}$ forms a basis of $T_p^\perp \gamma$, there is a set of null Jacobi fields $V_1, \dots, V_{n-1}$ such that $V_i|_p = W_i$. One can compute in this case that the relative null shape operator $\mathring{B}$ is identical to the null shape operator $\mathring{\mathrm{Sh}}(\dot\gamma, -)$.

6.19 To connect arguments using sections of $\mathring{T}\gamma$ to those using section of $T^\perp\gamma$, we have the following useful result concerning Jacobi fields.

Lemma

Let (M, g) be Lorentzian and γ and null affine geodesic. Suppose a, b belongs the domain of γ.

1. *If there exists a null Jacobi field V along γ satisfying $[V(a)] = [V(b)] = 0$, then there exists a null Jacobi field W along γ satisfying $W(a) = W(b) = 0$ (i.e., $\gamma(a)$ and $\gamma(b)$ are conjugate).*
2. *Suppose there exists a null hypersurface Σ such that $\gamma(a) \in \Sigma$ and $\dot\gamma(a) \in N_{\gamma(a)}\Sigma$. If there exists a Jacobi field V along γ satisfying $[V(b)] = 0$ and $[\dot V(a)] = \mathring{\mathrm{Sh}}(\dot\gamma(a), [V(a)])$, then there exists a Jacobi field W along γ with $W(b) = 0$, $W(a) = V(a)$, and $[\dot W(a)] = \mathring{\mathrm{Sh}}(\dot\gamma(a), [W(a)])$.* ■

Proof.

1. Our assumption implies there exists real number λ_a and λ_b such that $V(a) = \lambda_a \dot\gamma(a)$ and $V(b) = \lambda_b \dot\gamma(b)$. So choosing
$$W(t) = V(t) - \dot\gamma(t) \cdot \left[\frac{t-a}{b-a}\lambda_b + \frac{t-b}{a-b}\lambda_a \right]$$
we obtain a Jacobi field with the desired conditions.
2. Similarly, we have that $V(b) = \lambda\dot\gamma(b)$, so we can set
$$W(t) = V(t) - \lambda\dot\gamma(t) \cdot \frac{t-a}{b-a}.$$ □

6.20 Definition (Null focal point)

Let (M, g) be a Lorentzian manifold, Σ a null hypersurface, and $\gamma : (a, b) \to M$ an affine null geodesic with $a < 0 < b$, such that $\gamma(0) \in \Sigma$ and $\dot\gamma(0) \in N_{\gamma(0)}\Sigma$. Given $0 \neq c \in (a, b)$, we say that $\gamma(c)$ is a null focal point *of Σ if there exists a null Jacobi field V along γ such that $[V(c)] = 0$ and $[\dot V(0)] = \mathring{\mathrm{Sh}}(\dot\gamma(0), [V(0)])$.* ◊

6.21 Having introduced these language, now we can properly record the generalization of Theorem 5.21 to the case where we allow one endpoint to vary among a null hypersurface. The proof is largely similar, and so we omit it here; interested readers can consult Proposition 48 in Chapter 10 of O'Neill, *Semi-Riemannian geometry: with applications to relativity*.

Theorem

Let (M, g) be Lorentzian, Σ a null hypersurface, and $\gamma : (a, b) \to M$ a future-directed affine null geodesic with $a < 0 < b$, such that $\gamma(0) \in \Sigma$ and $\dot\gamma(0) \in N_{\gamma(0)}\Sigma$. Suppose for some $c \in (0, b)$ we have that $\gamma(c)$ is a null focal point of Σ, then for every $d \in (c, b)$, and for every open neighborhood U of $\gamma(0)$, there exists $p \in U \cap \Sigma$ satisfying $p \lneqq \gamma(d)$. ■

DRAFT COPY

© Willie Wai-Yeung Wong

DRAFT COPY

6.22 Our next task is to provide sufficient conditions to make use of Theorem 5.21 and Theorem 6.21, by exhibiting the existence of the relevant Jacobi fields. We will follow the same philosophy as outlined in ¶ 3.15, and detect the existence of a vanishing null Jacobi field by studying the wedge product of a number of them.

Now let $V_1, \dots, V_{n-1}$ be null Jacobi fields along the null geodesic γ in an $(1+n)$-dimensional Lorentzian manifold (M, g). We can define

$$\Omega = [V_1] \wedge [V_2] \wedge \cdots \wedge [V_{n-1}]$$

which vanishes if and only if $[V_1], \dots, [V_{n-1}]$ are linearly dependent. We will again study its time-evolution. As $\mathring{T}\gamma$ is $(n-1)$ dimensional, the space of $(n-1)$-alternating-vectors over $\mathring{T}\gamma$ is one-dimensional, and hence where Ω is non-vanishing we have

$$\dot{\Omega} = \kappa \Omega$$

for some real valued κ. Directly computing we find

$$\dot{\Omega} = [\dot{V}_1] \wedge [V_2] \wedge \cdots \wedge [V_{n-1}] + \cdots + [V_1] \wedge [V_2] \wedge \cdots \wedge [V_{n-2}] \wedge [\dot{V}_{n-1}]$$

and hence

$$\dot{\Omega} = \operatorname{tr}(\mathring{B}) \Omega.$$

(Here $\mathring{B}$ is the relative null shape operator of $V_1, \dots, V_{n-1}$.)

6.23 Just as in the time-like case, we need to examine the evolution equation for $\mathring{B}$. This we find to read

$$\frac{d}{dt}\mathring{B} + \mathring{B} \circ \mathring{B} = \mathring{S}.$$

We can also decompose

$$\mathring{B} = \frac{1}{n-1} \operatorname{tr}(\mathring{B}) \cdot \mathrm{Id} + \tilde{\mathring{B}} + \check{\mathring{B}}. \tag{6.23.1}$$

Here $\tilde{\mathring{B}}$ and $\check{\mathring{B}}$ are the traceless self-adjoint and anti-self-adjoint parts of $\mathring{B}$ respectively, with respect to the inner product $\mathring{g}$. Note that in comparison to (4.1.1), we have a factor of $\frac{1}{n-1}$ instead of $\frac{1}{n}$; this arises due to $\mathring{T}_p\gamma$ being only $(n-1)$-dimensional.

6.24 (The trace of $\mathring{S}$) It is worth discussing what is the trace of the curvature term $\mathring{S}$. Let $[e_1], \dots, [e_{n-1}]$ be an orthonormal basis of $\mathring{T}\gamma$ (using that $\mathring{g}$ is positive definite), the trace can be written as

$$\operatorname{tr} \mathring{S} = \sum_{i=1}^{n-1} \mathring{g}([e_i], \mathring{S}([e_i])) = \sum_{i=1}^{n-1} g(e_i, \mathrm{Riem}_{e_i \dot{\gamma}} \dot{\gamma}).$$

Now set $e_n = \dot{\gamma}$, and choose e_0 to be a null vector that is orthogonal to $e_1, \dots, e_{n-1}$, satisfying $g(e_0, e_n) = 1$, we find

$$\mathrm{Ric}_{\dot{\gamma}\dot{\gamma}} = -\sum_{i=1}^{n-1} g(e_i, \mathrm{Riem}_{e_i \dot{\gamma}} \dot{\gamma}) - g(e_0, \mathrm{Riem}_{e_n \dot{\gamma}} \dot{\gamma}) - g(e_n, \mathrm{Riem}_{e_0 \dot{\gamma}} \dot{\gamma}).$$

Both final terms vanish from the symmetries of the Riemann tensor! As a result we've shown that

$$\operatorname{tr} \mathring{S} = -\mathrm{Ric}_{\dot{\gamma}\dot{\gamma}}. \tag{6.24.1}$$

DRAFT COPY

6.25 (Vanishing twist) For simplicity, we will assume that the null twist $\check{\mathring{B}}$ vanishes; this is always the case when we are studying null conjugate points and null focal points, where we apply the results in this discussion. In this case, we have the following *null* version of the Raychaudhuri equation:

$$\frac{d}{dt}(\operatorname{tr}\mathring{B}) = -\mathrm{Ric}_{\dot\gamma\dot\gamma} - \frac{1}{n-1}(\operatorname{tr}\mathring{B})^2 - \operatorname{tr}(\tilde{\mathring{B}} \circ \tilde{\mathring{B}}). \tag{6.25.1}$$

The final term is non-negative since $\tilde{\mathring{B}}$ is self-adjoint. And hence the following two counterparts to Proposition 3.22 and Theorem 4.7 hold, with almost identical proofs.

6.26 Proposition
Let (M,g) be Lorentzian with dimension $(1+n)$, γ a null affine geodesic, and $V_1,\dots,V_{n-1}$ null Jacobi fields. Assume that for an interval (s_0,s_1) in the domain of γ we have $[V_1],\dots,[V_{n-1}]$ are linearly independent, and that the null twist $\check{\mathring{B}} = 0$. Then for $s \in \{s_0,s_1\}$, we have $\Omega(s) = 0$ if and only if the one-sided limit

$$\lim_{\substack{\sigma\to s\\ \sigma\in(s_0,s_1)}} |\operatorname{tr}\mathring{B}| = \infty.$$ ■

6.27 Theorem (Null singularity theorem, version 0)
Let (M,g) be Lorentzian with dimension $(1+n)$, such that for every null *vector X we have* $\mathrm{Ric}(X,X) \geq 0$. *Take Σ a null hypersurface, and set L a null geodesic generator of Σ. Suppose the null shape operator of Σ satisfies*

$$\operatorname{tr}(\mathring{\mathrm{Sh}}(L, \text{-})) \leq -C$$

for some constant $C > 0$. Then letting $\gamma : [0,b) \to M$ be an affine geodesic satisfying

- $\gamma(0) \in \Sigma$;
- $\dot\gamma(0) = L|_{\gamma(0)}$;
- $b > \frac{n-1}{C}$.

Then there is a null focal point of Σ along γ. ■

6.28 Exercise
Write out the proofs of Proposition 6.26 and Theorem 6.27. ♠

6.29 (More on the null Raychaudhuri equation) Focusing now on the case of null hypersurfaces, we can give the null Raychaudhuri equation (6.25.1) a geometric interpretation. By way of ¶ 6.18, given a null hypersurface Σ with affine null geodesic generator L, we have that the trace $\operatorname{tr}\mathring{B}$ can be identified with $\operatorname{tr}\mathring{\mathrm{Sh}}(L, \text{-})$. This motivates the following definition.

6.30 Definition (Null expansion)
Let Σ be a null hypersurface. Its null expansion *is the section θ of $N^*\Sigma$ given by $\theta(L) =$* $\operatorname{tr}\mathring{\mathrm{Sh}}(L, \text{-})$. ◊

Just as how the null shape operator should be regarded as the correct replacement of notion of the second fundamental form for null hypersurfaces, the null expansion is the correct replacement for the *mean curvature* for null hypersurfaces.

© Willie Wai-Yeung Wong

6.31 With the above definition, we see that (6.25.1) implies, for any null geodesic generator L (whether affine or not),

$$\nabla_L(\theta(L)) - \theta(\nabla_L L) \leq -\mathrm{Ric}(L,L) - \frac{1}{n-1}(\theta(L))^2. \tag{6.31.1}$$

In the presence of the null energy condition that $\mathrm{Ric}(L,L) \geq 0$ for any null L, this further implies the monotonicity property

$$\nabla_L(\theta(L)) - \theta(\nabla_L L) \leq -\frac{1}{n-1}(\theta(L))^2. \tag{6.31.2}$$

By Proposition 6.12, for *any* spatial cross section $\mathring{\Sigma}$, if we define its mean curvature vector $H = \mathrm{tr}_{\mathring{g}} \mathrm{II}$, we see that $\theta(L) = -g(H,L)$. The inequality can be used to prove:

6.32 Proposition (No-return for null expansion)
Let (M,g) be a time-orientable Lorentzian manifold satisfying $\mathrm{Ric}(X,X) \geq 0$ for every null vector X. Let Σ be a null hypersurface, with L any future directed null geodesic generator of Σ. Let p,q lie on the same null geodesic in Σ, with $p \leqq q$. Suppose a local cross section $\mathring{\Sigma}_p$ through p has mean curvature such that $g(H,L)|_p \geq 0$ (resp. > 0), then any *local cross section $\mathring{\Sigma}_q$ through q will have mean curvature with $g(H,L)|_q \geq 0$ (resp. > 0).* ■

6.33 Exercise
Write out the proof of Proposition 6.32.

Hint: there are two ways to take advantage of the fact that L is geodesic. First is to use $\nabla_L L \propto L$ and argue using integrating factors. Second is to reparametrize and consider L' an affine geodesic vector field in the same direction of L. ♠

DRAFT COPY

Topic 7 (2023/12/4)

Penrose Singularity Theorem and Black Holes

7.1 In popular discourse, perhaps the two most striking predictions of general relativity are the cosmological big bang and black holes. Previously we discussed how Theorem 4.7 (and its extension, Theorem 4.12) provides the foundations for the Hawking singularity theorems. The Hawking theorems are tightly associated with the *big bang* singularity in cosmology. In this lecture we will discuss the *Penrose singularity theorem*, as well as the notion of black holes that it significantly clarified. The foundations of the Penrose theorem are Theorem 6.27 and Theorem 6.21.

A bit more Causal Theory

7.2 In Theorem 5.6 we proved the transitivity of $\leqq$ and $\lneqq$. We can then use some of the language and ideas from *order theory* to study them. A lot more can be said on this topic, see Chapter 14 in O'Neill, *Semi-Riemannian geometry: with applications to relativity*. We only present a portion of the discussion necessary for stating (our version of) the Penrose Singularity Theorem.

7.3 **Definition**
Let (M, g) be a time-orientable Lorentzian manifold.

1. *Given a subset $\Gamma \subseteq M$, for convenience we denote by $\mathcal{I}^\pm(\Gamma)$ (and similarly $\mathcal{J}^\pm(\Gamma)$) the set $\cup_{p\in\Gamma}\mathcal{I}^\pm(p)$ (and similarly $\cup_{p\in\Gamma}\mathcal{J}^\pm(p)$).*
2. *We say that a set $\Gamma \subseteq M$ is* future-closed *if $\mathcal{I}^+(\Gamma) \subseteq \Gamma$; similarly we may define* past-closed *sets.*
3. *Conversely, we say that a set $\Gamma \subseteq M$ is* achronal *if $\mathcal{I}^+(\Gamma) \cap \Gamma = \emptyset$.* ◊

7.4 A few simple properties of this definition include:

Proposition
Let (M, g) be a time-orientable Lorentzian manifold.

1. *Given any subset Γ, both $\mathcal{I}^+(\Gamma)$ and $\mathcal{J}^+(\Gamma)$ are future-closed.*

DRAFT COPY

2. *If Γ is future-closed, then $M \setminus \Gamma$ is past-closed.*
3. *If Γ is an achronal set, and γ is a time-like curve, then γ intersects Γ at most once; if Γ is furthermore a (smooth) submanifold, then the intersection must be transverse.*
4. *If $q \in \mathcal{J}^+(\Gamma) \setminus \mathcal{I}^+(\Gamma)$, then any causal curve γ starting from Γ and ending at q must be a null geodesic with no conjugate points before q, such that $\gamma \subseteq \mathcal{J}^+(\Gamma) \setminus \mathcal{I}^+(\Gamma)$.* ■

These statement are easy consequences of transitivity Theorem 5.6 and the characterization Proposition 5.4; we omit detailed proofs.

7.5 Our previous result on the topology of $\mathcal{I}^+(p)$ and $\mathcal{J}^+(p)$ (Theorem 5.10) also extends to this setting.

Theorem (Causal Topology)
Let (M,g) be a time-orientable Lorentzian manifold, and Γ a subset of M.
1. *The set $\mathcal{I}^+(\Gamma)$ is open.*
2. *The interior of $\mathcal{J}^+$ is equal to $\mathcal{I}^+$.*
3. *The set $\mathcal{J}^+(\Gamma) \cup \Gamma \subseteq \overline{\mathcal{I}^+(\Gamma)}$.* ■

7.6 Exercise
Prove that if (M,g) is a time-orientable compact Lorentzian manifold, then there exists some $p \in M$ satisfying $p \gneq p$.

Hint: using compactness find a finite cover of (M,g) by $\{\mathcal{I}^+(p_1),\ldots,\mathcal{I}^+(p_k)\}$. Suppose this cover is minimal: that removing any of the sets involved results in a non-cover of M. Prove that this requires $p_1 \in \mathcal{I}^+(p_1)$. ♠

7.7 Theorem (More Causal Topology)
Let (M,g) be a time-orientable Lorentzian manifold, and Γ a future-closed set. Then

$$\overline{\mathcal{I}^+(\Gamma)} \underset{(a)}{=} \overline{\Gamma} \underset{(b)}{\supseteq} \Gamma \underset{(c)}{\supseteq} \mathcal{I}^+(\Gamma) \underset{(d)}{=} \mathrm{int}(\Gamma) \underset{(e)}{=} \mathcal{I}^+(\mathrm{int}(\Gamma)) \underset{(f)}{=} \mathcal{I}^+(\overline{\Gamma}).$$

(Here $\mathrm{int}(\Gamma)$ denotes the topological interior of the set Γ.) As a consequence:
1. *Both $\mathrm{int}(\Gamma)$ and $\overline{\Gamma}$ are future-closed.*
2. *The three sets $\mathrm{int}(\Gamma), \Gamma, \overline{\Gamma}$ all have the same interior and the same closure.* ■

Proof. Relation (b) is trivial, and (c) is by definition of Γ being future-closed. It remains to prove the four equalities.

For (a), by Theorem 7.5 we have $\overline{\mathcal{I}^+(\Gamma)} \supseteq \Gamma$, and taking the closure on both sides gives $\supseteq$. Taking the closure of (c) gives the other containment.

We prove (d) and (e) together. As $\mathrm{int}(\Gamma)$ is the largest open set contained in Γ, it must contain $\mathcal{I}^+(\Gamma)$, since (c) and Theorem 7.5 shows $\mathcal{I}^+(\Gamma)$ is an open set contained in Γ. This shows $\mathcal{I}^+(\Gamma) \subseteq \mathrm{int}(\Gamma)$. Next, observe that if U is an open neighborhood of p, then there exists (using exponential map) points $p_- \lneq p \lneq p_+$ in U. In particular, $p \in \mathcal{I}^-(U) \cap \mathcal{I}^+(U)$. This shows $\mathrm{int}(\Gamma) \subseteq \mathcal{I}^+(\mathrm{int}(\Gamma))$. We last note that the inclusion $\mathcal{I}^+(\mathrm{int}(\Gamma)) \subseteq \mathcal{I}^+(\Gamma)$ is trivial.

Finally, we prove (f). Trivially we have $\mathcal{I}^+(\Gamma) \subseteq \mathcal{I}^+(\overline{\Gamma})$. Now suppose $p \in \overline{\Gamma} \setminus \Gamma$. Let $q \gneq p$, so that $p \in \mathcal{I}^-(q)$. By Theorem 7.5 we find $\mathcal{I}^-(q)$ is open, and hence there exists an open set $U \ni p$ such that $U \subseteq \mathcal{I}^-(q)$. Since $p \in \overline{\Gamma}$, the set $U \cap \Gamma$ is non-empty. And hence there exists $p' \in \Gamma$ such that $q \gneq p'$, or that $q \in \mathcal{I}^+(\Gamma)$.

The two final consequences are easy and left to the reader. □

DRAFT COPY

DRAFT COPY

7.8 Theorem
Let (M, g) be a time-orientable Lorentzian manifold, and $\Gamma \subseteq M$ be future-closed. Then its boundary $\partial\Gamma$ is an achronal Lipschitz hypersurface. ■

A set N is a Lipschitz hypersurface of M^{k+1} if, near every point $p \in N$, there exists an open set $U \ni p$ and a local coordinate system $(z, y_1, \ldots, y_k)$ of U and a Lipschitz function $f : \mathbb{R}^k \to \mathbb{R}$ such that $N \cap U = \{z = f(y_1, \ldots, y_k)\}$. In other words, N can be expressed locally as the graph of a Lipschitz function.

Proof. By Theorem 7.7 we have $\partial\Gamma = \overline{\Gamma} \setminus \mathcal{I}^+(\overline{\Gamma})$. So

$$\mathcal{I}^+(\partial\Gamma) \cap \partial\Gamma \subseteq \mathcal{I}^+(\overline{\Gamma}) \cap (\overline{\Gamma} \setminus \mathcal{I}^+(\overline{\Gamma})) = \emptyset,$$

showing achronality. Now let $p \in \partial\Gamma$. Choose geodesic normal coordinates $x_0, \ldots, x_n$ near p with p at the origin and the metric

$$g_p = -\,\mathrm{d}x_0^2 + \,\mathrm{d}x_1^2 + \ldots + \,\mathrm{d}x_n^2,$$

normalized so that ∂_0 is future-directed. Restrict to a cylindrical subset $U_\epsilon := \{|x_0| < \epsilon, \sum_{i=1}^n |x_i|^2 < \epsilon^2\}$, with ϵ chosen small enough such that every vector $v = v_0\partial_0 + \cdots v_n\partial_n$ satisfying $|v_0|^2 > 2\sum_{i=1}^n |v_i|^2$ is time-like. Denote by

$$U_\epsilon^\pm := \left\{ \pm x_0 > \sqrt{\textstyle\sum_{i=1}^n |x_i|^2} \right\} \cap U_\epsilon.$$

Then we have

1. $U_\epsilon^\pm \subseteq \mathcal{I}^\pm(p)$ respectively.
2. For fixed $(x_1, \ldots, x_n)$ the curve $\gamma : (-\epsilon, \epsilon) \ni s \mapsto (s, x_1, \ldots, x_n) \in U_\epsilon$ is time-like, starts in $U_\epsilon^- \subseteq M \setminus \Gamma$, and ends in $U_\epsilon^+ \subseteq \Gamma$. Hence it intersects $\partial\Gamma$. Furthermore, the intersection is unique (as $\partial\Gamma$ has already been proven to be achronal).
3. Therefore, there exists a function f such that $(f(x_1, \ldots, x_n), x_1, \ldots, x_n)$ is said intersection; in other words, $\partial\Gamma$ is locally a graph of f.
4. We note that necessarily f has a Lipschitz constant no more than $\sqrt{2}$; for the straight line joining $(x_0, x_1, \ldots, x_n)$ and $(y_0, y_1, \ldots, y_n)$ with $|x_0 - y_0| > 2\sum_{i=1}^n |x_i - y_i|^2$ is a time-like curve, and achronality forbids it from intersecting $\partial\Gamma$ more than once. □

Penrose Singularity Theorem

7.9 The Penrose Singularity Theorem provides a *sufficient condition* for a space-time to be *null geodesic incomplete*. This should be compared to the Hawking theorems (see Theorem 4.12) which leads to *timelike geodesic incompleteness*. As with our statement of Theorem 4.12, the statement here also contains certain technical assumptions on the regularity of the space-time.

A better name would be the Penrose Incompleteness *Theorem, but we will follow historical precedent here for the name.*

7.10 Assumption (for Penrose Theorem)
We assume (M, g) is time-orientable (with τ the non-vanishing timelike vector field), and that the following two conditions hold.

"Continuity of causal relation" *if the sequences of points $p_i \to p$ and $q_i \to q$, with $p \neq q$, satisfy $p_i \leqq q_i$ for all i, then $p \leqq q$.*

"Space-time splitting" *Let $p \sim q$ denote the condition "p and q lies on the same integral curve of τ". Then $M/\sim$ can be given the structure of a smooth manifold, with the projection map a submersion.* ◊

Recall our Assumption 1.10 on properties of manifolds.

DRAFT COPY

Observe that a consequence of "continuity of causal relation" is that for any *closed* subset $\Gamma \subseteq M$, we have

$$\overline{\mathcal{I}^+(\Gamma)} = \overline{\mathcal{J}^+(\Gamma)} = \mathcal{J}^+(\Gamma) \cup \Gamma. \tag{7.10.1}$$

7.11 The purpose of the two assumptions are for ruling out "trivial" ways in which a manifold can be geodesically incomplete. An example of spacetimes ruled out by the assumptions above is Example 4.19; it satisfies neither of the two assumptions given above. Both of the assumptions serve to rule out certain types of "holes" in the space-time. The second one, in particular, states that the topology of "space" cannot change over time.

Besides being intuitively reasonable, both of the assumptions are consequences of "global hyperbolicity". This latter condition is intimately tied to the solvability of the initial value problem (and hence the question of whether physics is "deterministic"), and hence we will take it as physically appropriate that the assumptions listed above are satisfied.

7.12 The two assumptions listed above are merely technical, and are generally satisfied for physically interesting space-times. The key driving force of the Penrose theorem is the following definition.

Definition (Trapped and Marginally Trapped Surfaces)
Let (M, g) be a time-orientable Lorentzian manifold. Given a co-dimension 2 space-like submanifold $\mathring{\Sigma}$, we let H to be its mean-curvature vector (trace of the second fundamental form relative to the induced metric $\mathring{g}$ on $\mathring{\Sigma}$). We say that $\mathring{\Sigma}$ is:
future/past trapped *if $-H$ is everywhere future/past time-like;*
future/past marginally trapped *if $-H$ is everywhere future/past null or zero.* ◊

7.13 The significance of this definition lies in the following lemma, which is an easy consequence of Proposition 6.12, which shows that the null expansion $\theta(L) = g(-H, L)$.

Lemma
Let Σ be a null hypersurface, with L its future-directed null geodesic generator. Take $\mathring{\Sigma}$ a spatial section. Then:

- *If $\mathring{\Sigma}$ is future trapped, then the null expansion $\theta(L) < 0$ along $\mathring{\Sigma}$.*
- *If $\mathring{\Sigma}$ is future marginally trapped, then the null expansion $\theta(L) \leq 0$ along $\mathring{\Sigma}$.* ■

What's important to note about this statement is that, locally, since $\mathring{\Sigma}$ is co-dimension 2 and space-like, its normal bundle contains two null directions. This means locally, it is possible to generate *two* null hypersurfaces from $\mathring{\Sigma}$. This lemma states that both hypersurfaces will have negative expansion to the future.

This is a somewhat counter-intuitive concept: in our everyday experience, if we are given a curved mirror that is concave on one side, then it must be convex on the other. Light rays coming off the concave side will tend to focus, while light rays coming off the convex side will disperse. That we can have "focusing on both sides" depends strongly on the fact that we are working in space-*time*, where the geometry is changing dynamically over time. This is illustrated in the following exercise.

DRAFT COPY

7.14 Exercise

Consider $(M,g) = (\mathbb{R}\times N, -\,\mathrm{d}t^2 + h)$ a product Lorentzian manifold. Let $\mathring{\Sigma}$ a co-dimension 2 submanifold such that $\mathring{\Sigma} \subseteq \{0\}\times N$.

1. Prove that the second fundamental form $\mathrm{II}_{\mathring{\Sigma}}$ takes values in the N direction only.
2. Prove that if $\mathring{\Sigma}$ is closed and compact, then $\mathring{\Sigma}$ is neither trapped nor marginally trapped. ♠

7.15 Exercise

Let (M,g) be the Minkowski space $\mathbb{R}^{1+3}$. Choose

$$\mathring{\Sigma} := \{(t,x,y,z) \mid -t = \sqrt{x^2+y^2+z^2} \wedge t = x-1\}.$$

1. Prove that $T\mathring{\Sigma}$ is space-like;
2. Prove that $\mathring{\Sigma}$ is future marginally trapped.

(In fact, a similar analysis can be extended to arbitrary conic sections of the light cone in Minkowski space. Those that are compact [ellipsoids] turn out to always be neither trapped nor marginally trapped. Those that are paraboloids [like above] are marginally trapped either to the future or to the past. And those that are hyperboloids are trapped.) ♠

7.16 Example

The previous exercise is the best we can do: compact trapped surfaces are very much a *gravitational* phenomenon, where the driver for the focusing effect for light rays is the strong gravitational pull. We illustrate this fact by showing that

> In Minkowski space, there does not exist any *compact* trapped surface.

Let $M = \mathbb{R}^{1,n}$ where $n \geq 2$, and take $\mathring{\Sigma}$ a compact (closed) $n-1$ dimensional space-like submanifold. We will use $(t, x_1, \dots, x_n)$, the standard coordinates of M, relative to which the Minkowski metric is $-\,\mathrm{d}t^2 + \,\mathrm{d}x_1^2 + \cdots + \,\mathrm{d}x_n^2$. It suffices to show that for some future-directed null vector L orthogonal to $\mathring{\Sigma}$, the inner product $g(H,L) < 0$. (By time-symmetry, this means there also exist a past-directed null vector L' orthogonal to $\mathring{\Sigma}$ with inner product $g(H,L') < 0$, but as $-L'$ is now future-directed, we see also that $g(H,-L') > 0$, so that $\mathring{\Sigma}$ is not future/past trapped, nor future/past marginally trapped.)

A device we will use is the following formula for the mean curvature of a immersed pseudo-Riemannian submanifold of Minkowski space (similar formula holds for any pseudo-Riemannian vector space $\mathbb{R}^{p,q}$ too). Since $\mathring{\Sigma}$ is a space-like submanifold, the Minkowski metric induces on it a Riemannian metric $\mathring{g}$. Then the *mean curvature vector* H along $\mathring{\Sigma}$ has the following expression: We can decompose H using the standard frame of $\mathbb{R}^{1,n}$:

$$H = H_t\partial_t + H_1\partial_1 + \ldots + H_n\partial_n$$

then the components satisfy

$$H_t = \triangle_{\mathring{g}} t, \quad H_1 = \triangle_{\mathring{g}} x_1, \quad \ldots, \quad H_n = \triangle_{\mathring{g}} x_n. \tag{7.16.1}$$

This expression is also why the minimal surface equation is a quasilinear elliptic PDE.

Here, $\triangle_{\mathring{g}}$ is the Laplace-Beltrami operator for the metric $\mathring{g}$ on $\mathring{\Sigma}$, acting on scalar functions. We consider $t, x_1, \dots, x_n$ as the restriction of the coordinate functions to $\mathring{\Sigma}$.

© Willie Wai-Yeung Wong

DRAFT COPY

Since $\mathring{\Sigma}$ has co-dimension > 1 and is compact, we can assume that it avoids the t-axis. Then the function $|x| - t$ restricts to a smooth function on $\mathring{\Sigma}$. Let us compute its Laplacian:

$$\triangle_{\mathring{g}}(|x| - t) = -\triangle_{\mathring{g}} t + \sum_i \frac{x_i}{|x|} \triangle_{\mathring{g}} x_i + \sum_i \frac{\mathring{g}^{\mu\nu} \nabla_\mu x_i \nabla_\nu x_j}{|x|} - \sum_{i,j} \frac{\mathring{g}^{\mu\nu} x_i \nabla_\mu x_i \; x_j \nabla_\nu x_j}{|x|^3}. \tag{7.16.2}$$

Since $|x| - t$ is a continuous function on a compact manifold, it attains a global maximum. At such a point p, its Laplacian is non-positive, and its gradient vanishes. The vanishing of the gradient implies that at p,

$$\sum_i \frac{x_i}{|x|} \nabla_\mu x_i = \nabla_\mu t.$$

So we find that at p,

$$0 \geq -H_t + \sum_i \frac{x_i}{|x|} H_i + \sum_i \frac{1}{|x|} \mathring{g}(\nabla x_i, \nabla x_i) - \frac{1}{|x|} \mathring{g}(\nabla t, \nabla t). \tag{7.16.3}$$

Since $\mathring{g}$ is induced from the metric g, the final two factors evaluate to $\frac{1}{|x|} \mathrm{tr}_{\mathring{g}} \mathring{g} = (n-1)$. Now set L to be the null vector with $L_t = 1$ and $L_i = \frac{x_i}{|x|}$, we find finally that at p,

$$-\frac{n-1}{|x|} \geq g(H, L). \tag{7.16.4}$$

It remains to check that L is orthogonal to $\mathring{\Sigma}$, but this holds as $L(|x| - t) = 0$ so is a geodesic null generator of the level sets of $|x| - t$ (which are null hypersurfaces). At p, since $|x| - t$ reached a global maximum, we have that $\mathring{\Sigma}$ is tangent to one such level set. And hence $T_p \mathring{\Sigma}$ is orthogonal to L. ◊

7.17 The abbreviated version of Penrose's theorem states that "if a space-time admits a compact trapped surface, then it is geodesically incomplete." The detailed version, which we are now in a position to prove, states:

Theorem (Penrose Incompleteness)
Let (M, g) be a Lorentzian manifold satisfying Assumption 7.10. Suppose additionally that

1. *for every null vector X, we have* $\mathrm{Ric}(X, X) \geq 0$*;*
2. *the smooth manifold $M/\sim$ of the "space-time splitting" assumption is* connected *and* non-compact;
3. *there exists a co-dimension 2 submanifold $\Gamma \subseteq M$ that is: compact, space-like, achronal, and future-trapped.*

Then (M, g) is null geodesically incomplete. ■

Proof. We will argue by contradiction. Assume now that (M, g) is null geodesically complete. This implies that exp is well-defined on some open set $O \subseteq TM$ that contains all null vectors.

(1) Compactness of $\partial\mathcal{I}^+(\Gamma)$

DRAFT COPY

Under our hypothesis, Γ is achronal so $\Gamma \cap \mathcal{I}^+(\Gamma) = 0$. Hence by (7.10.1) we find

$$\partial\mathcal{I}^+(\Gamma) = \Gamma \cup \Big(\mathcal{J}^+(\Gamma) \setminus \mathcal{I}^+(\Gamma)\Big).$$

By Proposition 7.4, every element of $\partial\mathcal{I}^+(\Gamma)$ is of the form $\exp_p(v)$, where $p \in \Gamma$ and $v \in T_pM$ is either zero or a future directed null vector. That there exists no time-like curves from Γ to $\exp_p(v)$ poses some restrictions:

- By Exercise 5.5 we must have $v \perp T\Gamma$.
- We can now locally select continuously null vectors orthogonal to Γ, extending v; this generate locally a null hypersurface. By Lemma 7.13 we conclude that these null hypersurfaces have negative expansion. Hence arguing using the null Raychaudhuri equation (6.25.1) in an analogous fashion to Theorem 6.27 (which requires the energy condition that $\mathrm{Ric}(X,X) \geq 0$ for all null X), we see that we must have $g(H,v) \in [0, n-1]$, where H is the mean curvature vector of Γ. For otherwise we will see a focal point, and beyond which Theorem 6.21 shows that $\exp_p(v) \in \mathcal{I}^+(\Gamma)$.

Now considering the exponential map as a smooth mapping $\exp : TM \supseteq O \to M$, we see that $\partial\mathcal{I}^+(\Gamma)$ is a subset of

$$\exp\{(p,v) \in TM \mid p \in \Gamma, g(v,v) = 0, g(H,v) \in [0,n-1]\}.$$

The relevant subset of O is compact, hence also its image under exp.

Finally, as $\partial\mathcal{I}^+(\Gamma)$ is a topological boundary, hence a closed set, we see that it is a closed set that is a subset of a compact set, hence it is compact.

(2) Local homeomorphism to $M/\sim$

By Proposition 7.4 $\mathcal{I}^+(\Gamma)$ is a future set, so by Theorem 7.8 its boundary is an achronal Lipschitz hypersurface. The projection $\Pi : M \to M/\sim$ is assumed to be a submersion, hence smooth. Thus Π restricts to a continuous mapping from $\mathcal{I}^+(\Gamma)$ to $M/\sim$; note that the two are both (Hausdorff) topological manifolds of the same dimension. Now, our definition of Π states that for $\Pi(p) = \Pi(q)$ if and only if they are connected by an integral curve of τ, a time-like vector field. Since $\partial\mathcal{I}^+(\Gamma)$ is achronal, we have that Π acts injectively on it. We can therefore apply Brouwer's Invariance of Domain Theorem to conclude that $\Pi : \partial\mathcal{I}^+(\Gamma) \to M/\sim$ is a homeomorphism onto its image.

(3) Deriving a contradiction

As an open mapping, we have $\Pi(\partial\mathcal{I}^+(\Gamma))$ is open. But since $\partial\mathcal{I}^+(\Gamma)$ is compact, so is this image; in particular the image is closed. As $M/\sim$ is connected, this implies the image must be all of $M/\sim$, and hence $\partial\mathcal{I}^+(\Gamma)$ and $M/\sim$ must be homeomorphic. But this is a contradiction, since one is compact and the other is non-compact. □

7.18 We previously mentioned the Lorentzian Splitting Theorem 4.16 as a counterpart to the Hawking-type singularity theorems. A similar heuristic argument indicates that for "generic" space-times with the null energy condition $\mathrm{Ric}(X,X) \geq 0$ for every null X, we expect generic null geodesics to be either incomplete, or contain conjugate points. This is realized in the Null Splitting Theorem of G. Galloway, which we state without proof below.

Galloway, "Null geometry and the Einstein equations"

DRAFT COPY

7.19 Theorem
Let (M, g) be a time-orientable Lorentzian manifold. Assume
- *(M, g) is null geodesically complete;*
- *$\mathrm{Ric}(X, X) \geq 0$ for every null vector X.*

Suppose γ is a complete affine null geodesic in M, such that γ is achronal, *then there exists a null hypersurface Σ such that*
- *$\gamma \subseteq \Sigma$;*
- *The null shape operator of Σ vanishes (i.e. Σ is totally geodesic).* ■

What is a black hole?

7.20 The popular science explanation of a black hole is a region in space whenceforth neither observers nor signals can escape. This may work well intuitively, but as it presupposes a preferred frame of reference (for isolating space from time), faces two issues.
1. General space-times do not admit obvious splittings into a spatial component and a temporal component.
2. For space-times that do admit a product splitting, Exercise 7.14 shows that there can be no compact trapped surfaces residing in a fixed time-slice, and so Penrose's theorem is not applicable.

Instead, it is better to define using only causal properties of the space-time.

It should be noted that there is not yet universal consensus on how to define a black hole space-time; on the other hand there is consensus that several of the well-known solutions in general relativity should be considered to be black holes. Our attempts at defining what a black hole is are informed by this latter consensus.

7.21 First, observe that if $\Omega \subseteq M$ is future-closed, then by Theorem 7.7 its closure $\overline{\Omega}$ has the property that "every future-directed causal curve, emanating from within $\overline{\Omega}$, cannot every reach its complement". This means that the space-time region $\overline{\Omega}$ satisfies the basic inescapability requirements of what we imagine to be a black hole. On the other hand, by Proposition 7.4 future-closed sets are dime a dozen, so we need to rely on something else to characterize a black hole.

7.22 The following definition is intuitively satisfying. We start by defining what should be considered to be outside of the black hole, and then define the black hole to be its complement. The basic idea stems from the expectations that when outside of the black hole, one should be able to send signals to observers who are arbitrarily far away. We codify this with:

This definition is inspired by Christodoulou, "On the global initial value problem and the issue of singularities".

Definition
Let (M, g) be a time-orientable Lorentzian manifold. We say a point $p \in M$ can outer communicate *if there exists a null geodesic ray γ emanating from p that is (a) future complete and (b) achronal.* ◇

The curve γ represents a signal sent from p. The completeness assumption captures the idea that the signal doesn't get cut-off pre-maturely by running into a singularity of the space-time or whatnot. The acrhronality is to capture the idea that the signal, as it travels, keeps "covering new ground". Note that if p can outer communicate via the geodesic ray γ, so can all points along γ. We should therefore treat the locations along such γ as "distant observers". In particular, we can imagine some sort of ideal "limiting point at infinity" for each such curve. The portion of the space-time that is not in a black hole are those that can "escape to infinity".

DRAFT COPY

Definition
The domain of outer communication *of a time-orientable Lorentzian manifold* (M, g) *is defined to be* $\mathcal{I}^-(\{p \in M \mid p \text{ can outer communicate}\})$. *Correspondingly, we define the* black hole region *to be the complement of the closure of the domain of outer communication.* ◊

Note that, by definition, the domain of outer communication (and its closure) is a past-closed set. Therefore the black hole region is future-closed.

Definition
The event horizon *is the boundary of the black hole region.* ◊

By Theorem 7.8, we see that we can expect the event horizon to be an achronal Lipschitz hypersurface. In fact, we expect in general the event horizon to be a null hypersurface. This can be proved under some technical assumptions on the global structure of space-time, but the basic idea is this: supposing the event horizon were to be *space-like*, then for a point p just outside the event horizon (meaning, in the domain of outer communication, and hence slightly to the past of the event horizon), by studying the geometry in a convex neighborhood we can see that every causal curve emanating from p must cross the event horizon. But this would cause a contradiction: since p is in the domain of outer communication, there must be some point $q \gneqq p$ such that q can outer communicate.

7.23 While I do not claim that the definitions above is "correct", I note that standard examples of black hole space-times (including those with positive cosmological constant) can be described appropriately by the definitions given.

7.24 Notice that the definition above depends on geodesic completeness. This means that it is automatically a *global* definition, that depends on knowing the entirety of the space-time (M, g). In fact, truncating the space-time can destroy geodesic completeness, so this is a definition that can only be used when one already knows that the space-time is "maximally extended" (in some suitable sense), so we don't get "false black hole regions".

7.25 This global definition also is not 100% compatible with Penrose's theorem. That a point in the regular region need only have one (null, say) geodesic ray that is future complete, means that there is still the possibility for a future-trapped surface to pass through said point. Similarly, as the definition is global, there is no guarantee that the black hole region contains any trapped surfaces. This motivates an alternative description using trapped and marginally trapped surfaces.

7.26 Definition
We say that a hypersurface Σ *of* (M, g) *is an* apparent horizon *if it admits a foliation by compact space-like sections* $\mathring{\Sigma}$, *each of which is marginally future-trapped.* ◊

7.27 How are the two definitions related? This is a question that is far from answered in the general setting. What we do know, and what is leading to active research, are:

1. For stationary black hole solutions (e.g. Schwarzschild and Kerr), the event horizon is also an apparent horizon.

© Willie Wai-Yeung Wong

2. In the spherically symmetric case, one can prove (for many matter models), that there exists a spherically symmetric apparent horizon, on one side of which every point p lies on a future trapped surface, and on the other side no spherically symmetric trapped surface exists. For some matter models one can further prove that this apparent horizon is achronal (though for other matter models this is known to be false), and hence the apparent horizon must lie within the closure of the black hole region.

Our lack of knowledge of the connection between the two cases is significant because frequently practitioners associate the "existence of a trapped surface" with "existence of a black hole", even though the mutual implication between these two statements is largely unknown.

DRAFT COPY

DRAFT COPY

Topic 8 (2023/12/11)

Black Holes in Spherical Symmetry

8.1 At the end of last lecture we mentioned that the notions of trapped surfaces, apparent horizons, and event horizons are hard to relate in general Lorentzian manifolds. A lot of our current understanding and expectations are based on a case where such relationships are much more clear cut; this is the setting of *spherical symmetric solutions*. We will explore these solutions in this lecture, specifically in relation to black hole solutions. Much of the material discussed in this lecture have been generalized significantly in An and W. Wong, "Warped Product Space-times".

Spherically Symmetric Lorentzian Manifolds

8.2 Letting (M, g) be a Lorentzian manifold of dimension $(n+1)$. One way to define spherical symmetry is to ask that the group $SO(n)$ acts by isometry on (M, g). We will take instead the following much more pedestrian (yet equivalent) description:

Definition
We say that the $(n+1)$ *dimensional Lorentzian manifold* (M, g_M) *is* spherical symmetric *if an open, dense subset of* M *can be decomposed (topologically) as* $Q \times \mathbb{S}^{n-1}$*, and the metric* g_M *can be written in the form*

$$g_M = g_Q + r^2 g_{\mathbb{S}^{n-1}}$$

where (Q, g_Q) *is a 2-dimensional Lorentzian manifold, and* $r : Q \to (0, \infty)$ *is called the* area radius *function.* ◊

In more geometric language, an equivalent definition is that the open dense subset of M *takes the form of a "warped product" with Lorentzian base* (Q, g_Q) *and spherical fiber.*

8.3 A standard computation shows that

DRAFT COPY

Lemma

For a spherically symmetric Lorentzian manifold $(M, g_M) = (Q \times \mathbb{S}^{n-1}, g_Q + r^2 g_{\mathbb{S}^{n-1}})$, if X, Y are Q-tangent vectors and V, W are $\mathbb{S}^{n-1}$-tangent vectors, we have

$$\begin{aligned}
\mathrm{Ric}_M(X,Y) &= \mathrm{Ric}_Q(X,Y) - \frac{n-1}{r}\nabla^2_{X,Y} r \\
\mathrm{Ric}_M(X,V) &= 0 \\
\mathrm{Ric}_M(V,W) &= \mathrm{Ric}_{\mathbb{S}^{n-1}}(V,W) - (r\triangle_Q r + (n-2) g_Q(\nabla r, \nabla r)) g_{\mathbb{S}^{n-1}}(V,W).
\end{aligned}$$

Here $\nabla^2 r$ is the Hessian of r on the Lorentzian manifold (Q, g_Q) and $\triangle_Q$ is the Laplace-Beltrami operator on (Q, g_Q); note also that the Ricci curvature of the $(n-1)$ sphere is $(n-2)g_{\mathbb{S}^{n-1}}$.

Additionally, we have that for each $q \in \mathbb{S}^{n-1}$, the submanifold $Q \times \{q\}$ is totally geodesic in M. For each $p \in Q$, the submanifold $\{p\} \times \mathbb{S}^{n-1}$ has (vector-valued) second fundamental form

$$\mathrm{II}(V,W) = -r g_{\mathbb{S}^{n-1}}(V,W)\nabla r,$$

where ∇r is the gradient (on Q) of r. ■

An immediate consequence is that the spheres have mean curvature vectors

$$H = -\frac{n-1}{r}\nabla r. \tag{8.3.1}$$

8.4 (Einstein tensor) The Einstein tensor $\mathrm{Einst} = \mathrm{Ric} - \frac{1}{2}Rg$ can be expressed using the Lemma above. First we compute

$$R = R_Q - \frac{2(n-1)}{r}\triangle_Q r + \frac{(n-2)(n-1)}{r^2}(1 - g_Q(\nabla r, \nabla r)). \tag{8.4.1}$$

To simplify notation, we will denote by

$$\varpi := 1 - g_Q(\nabla r, \nabla r). \tag{8.4.2}$$

From this we find, given X, Y vectors that are Q-tangent, and V, W vectors that are $\mathbb{S}^{n-1}$-tangent,

$$\mathrm{Einst}(X,Y) = \frac{n-1}{r}\Big((\triangle_Q r) g_Q(X,Y) - \nabla^2_{X,Y} r\Big) - \frac{(n-2)(n-1)}{2r^2}\varpi g_Q(X,Y) \tag{8.4.3}$$

$$\mathrm{Einst}(X,V) = 0 \tag{8.4.4}$$

$$\mathrm{Einst}(V,W) = \left[(n-2) r \triangle_Q r - \frac{r^2}{2}R_Q - \frac{(n-2)(n-3)}{2}\varpi\right] g_{\mathbb{S}^{n-1}}(V,W) \tag{8.4.5}$$

Birkhoff's Theorem and the Schwarzschild Solution

8.5 In 1916, soon after the publication of Einstein's manuscript on general relativity, Karl Schwarzschild wrote down an explicit solution to the *vacuum equations*. While Schwarzshild originally intended his solution to model the gravitational force *exterior* to a compact body (say, a star), we nowadays recognize this as the first demonstration

© Willie Wai-Yeung Wong

DRAFT COPY

of the fact that within Einstein's nonlinear theory of gravity, feedback of gravitational interaction among itself is sufficient to sustain static gravitational well. In 1923, George Birkhoff showed that Schwarzschild's discovery of this static solution is almost inevitable, as Schwarzschild's solutions form the *only family of* spherically symmetric solutions to the vacuum Einstein equations. In this section we discuss these two mathematical discoveries using modern geometric techniques.

We shall, additionally, take a slightly more general view: instead of treating the problem in pure vacuum, we will allow the presence of a cosmological constant. That is, throughout this section we will assume there is some $\Lambda \in \mathbb{R}$ such that

For lecturing, perhaps just do $\Lambda = 0$ to save time.

$$\mathrm{Einst} + \Lambda g = 0. \tag{8.5.1}$$

8.6 From (8.4.3) and (8.5.1), we find that
- $\nabla^2 r$ must be *pure trace*
- after some algebraic manipulation: $\triangle_Q(r^{n-1}) - (n-1)(n-2)r^{n-3} + 2r^{n-1}\Lambda = 0$.

We now examine the consequences of the two statements.

8.7 (Hodge operator) Assume that Q is orientable. Let ε denote the volume 2-form corresponding to the metric g_Q. The *Hodge operator* $\star : TQ \to TQ$ is given by

$$\star X := (\iota_X \varepsilon)^\sharp, \qquad (\star X)^b = X^a \varepsilon_a{}^b. \tag{8.7.1}$$

In the Riemannian case, the Hodge operator corresponds to "rotation by 90°"; in the Lorentzian case this is replaced by a hyperbolic rotation. And hence we have that
- $g(X,X) = -g(\star X, \star X)$ and $g(X, \star X) = 0$;
- $\star\star X = X$.

Alternatively, identifying a two-dimensional Riemannian manifold with a Riemann surface, the Hodge operator acts as complex multiplication by the imaginary unit.

8.8 (Kodama vector field) On a spherically symmetric Lorentzian manifold, the *Kodama vector field* K is the Q-tangent vector field that is given by

$$K = \star(\nabla r). \tag{8.8.1}$$

By definition, it is orthogonal to ∇r and hence $K(r) = 0$. This implies that its action by Lie derivation on the space-time metric satisfies

$$\mathcal{L}_K g_M = \mathcal{L}_K g_Q.$$

In index notation, we have

$$(\mathcal{L}_K g_Q)_{ab} = \nabla_a K_b + \nabla_b K_a = \nabla_a \varepsilon_{cb} \nabla^c r + \nabla_b \varepsilon_{ca} \nabla^c r.$$

And so

$$(\mathcal{L}_K g_Q)(X,X) = -2(\nabla^2 r)(X, \star X). \tag{8.8.2}$$

From ¶ 8.6 we know that $\nabla^2 r$ is pure trace, so that it is proportional to g_Q, and therefore the orthogonality of X and $\star X$ tells us that $(\mathcal{L}_K g_Q)(X,X) = 0$ for any vector X. As the tensor $\mathcal{L}_K g_Q$ is symmetric, we conclude then

DRAFT COPY

8.9 Theorem (Birkhoff, part 1)
In a spherical symmetric Lorentzian manifold (M, g_M) that solves (8.5.1), the Kodama vector field K must be a Killing vector field for both the base metric g_Q and the total metric g_M. ■

8.10 Applying now the general geometric fact that for a non-trivial Killing vector field, its vanishing set has codimension at least 2, we find:

Corollary
Either r is constant (and hence (M, g_M) is a Cartesian product), or ∇r vanishes only on isolated points of Q. ■

8.11 (The case of constant r) We first remark that this case is only possible when the cosmological constant Λ is strictly positive, thanks to the second consequence in ¶ 8.6; and only the specific value $r = \sqrt{(n-1)(n-2)/(2\Lambda)}$ is possible. Inserting this into (8.4.5) and applying (8.5.1) we see additionally that R_Q must equal $\frac{4\Lambda}{n-1}$. As the base manifold would have constant curvature, it is a space-form for which explicit parametrizations are available.

8.12 (The case of non-constant r) Let $Q_0 = Q \setminus \{\nabla r = 0\}$; we have that Q_0 is open and dense. Since ∇r is non-vanishing, we can use r to provide a coordinate function. The following provides an "inspired" choice of the second coordinate. Consider the one form

$$\frac{K^\flat}{g_Q(K,K)} = \frac{(\star \nabla r)^\flat}{-g_Q(\nabla r, \nabla r)}.$$

We shall show that it is a closed form, this will show that it can be integrated to give a scalar coordinate function. This form being closed is equivalent to its Hodge dual being divergence free, so we are led to looking at whether $\nabla r / g_Q(\nabla r, \nabla r)$ is divergence free. But this we can verify by a computation: first

$$\mathrm{div}_Q\Big(\frac{\nabla r}{g_Q(\nabla r, \nabla r)}\Big) = \frac{\triangle_Q r}{g_Q(\nabla r, \nabla r)} - \frac{g_Q(\nabla r, \nabla g_Q(\nabla r, \nabla r))}{g_Q(\nabla r, \nabla r)^2} = \frac{\triangle_Q r}{g_Q(\nabla r, \nabla r)} - \frac{2 g_Q(\nabla r, g_Q(\nabla r, \nabla^2 r))}{g_Q(\nabla r, \nabla r)^2}.$$

Now, using that $\nabla^2 r$ is pure trace, we have it can be replaced by $\frac{1}{2}\triangle_Q r g_Q$, we find finally

$$\mathrm{div}_Q\Big(\frac{\nabla r}{g_Q(\nabla r, \nabla r)}\Big) = \frac{\triangle_Q r}{g_Q(\nabla r, \nabla r)} - \frac{2\nabla^2 r(\nabla r, \nabla r)}{g_Q(\nabla r, \nabla r)^2} = 0.$$

Hence there exists a coordinate function s such that $ds = K^\flat / g_Q(K,K)$. By construction, $K(s) = 1$ and $K(r) = 0$, so we have that $\partial_s = K$ is the Killing vector field and that the metric is diagonalized in the (r,s) coordinates. So we can write

$$g_Q = \frac{1}{g_Q(\nabla r, \nabla r)} dr^2 + g_Q(K,K) ds^2 = \frac{1}{g_Q(\nabla r, \nabla r)} dr^2 - g_Q(\nabla r, \nabla r) ds^2. \tag{8.12.1}$$

Importantly, the coefficient $\mu = g_Q(\nabla r, \nabla r)$ should be *independent of* s. This means that the equation $\triangle_Q(r^{n-1}) - (n-1)(n-2)r^{n-3} + 2r^{n-1}\Lambda = 0$ from ¶ 8.6 can be rewritten as an ODE:

$$(n-1)\partial_r(\mu r^{n-2}) = (n-1)(n-2)r^{n-3} - 2\Lambda r^{n-1} \tag{8.12.2}$$

© Willie Wai-Yeung Wong

DRAFT COPY

which we can integrate to obtain (using customarily the letter m for the arbitrary constant of integration

$$\mu = \mu(r) = g_Q(\nabla r, \nabla r) = 1 - \frac{2m}{r^{n-2}} - \frac{2\Lambda}{n(n-1)} r^2. \tag{8.12.3}$$

8.13 Definition
The spherical symmetric Lorentzian solutions to (8.5.1) *given by*

$$g_M = -\mu(r) ds^2 + \frac{1}{\mu(r)} dr^2 + r^2 g_{\mathbb{S}^{n-1}}$$

with $\mu(r)$ as in (8.12.3) *are the Schwarzschild(-(Anti-)de Sitter) solutions. (The "de Sitter" case refers to $\Lambda > 0$; the "anti de Sitter" case refers to $\Lambda < 0$.)* ◊

8.14 Theorem (Birkhoff, part 2)
The only *spherically symmetric Lorentzian solutions to* (8.5.1) *are those given by the Schwarzschild(-(Anti-)de Sitter) family.* ■

8.15 (Comments) Birkhoff's theorem has two parts. The first part shows that for solutions to (8.5.1), the Kodama vector field K is a Killing vector field; additionally, if K is trivial, then the solution must admit another non-trivial Q-tangent Killing vector field. This is often presented as "spherical symmetry implies one extra symmetry". The second part shows that with this extra symmetry, it is possible to solve Einstein's equations explicitly by integrating an ODE, and we have the precise form of the solution given in the definition above.

8.16 (Coordinate singularity) Observe that $K = \partial_s$ is a *timelike symmetry* when $\mu(r) > 0$, and *spacelike* when $\mu(r) < 0$. But what about when $\mu(r) = 0$? We see that the form of the metric given becomes singular, since the coefficients of dr^2 is given in terms of $1/\mu$. Where does this singularity come from? This singularity is a side-effect of how we chose the *coordinate s*. Our construction of s relied on integrating the one-form $K^\flat/\mu(r)$, which is singular when $\mu(r) = 0$. Hence we should not expect that the coordinate charts we thus generated to cover any portion of the space-time where $\mu(r) = g(\nabla r, \nabla r) = 0$. This is *not to say* that $\{\mu(r) = 0\}$ cannot occur: one can in fact cover the set $\mu(r) = 0$ in the Schwarzschild(-(Anti-)de Sitter) solutions with a different choice of coordinate systems; this enables the exhibition of the *Kruskal extension* of these solutions. We omit this construction as it can be found in most textbooks on general relativity.

8.17 (Genuine singularity) In contrast, we see that $\mu(r)$ approaches ∞ as $r \searrow 0$. This turns out to be an actual singularity. This is easiest to see from (8.4.5); when combined with (8.5.1), we find

$$R_Q - 2\Lambda = \frac{2(n-2)}{r} \triangle_Q r - \frac{(n-2)(n-3)}{r^2}(1 - \mu(r)).$$

An explicit computation shows that the leading order term is given by

$$R_Q - 2\Lambda = \frac{2(n-1)(n-2)m}{r^n} + \dots$$

indicating that (Q, g_Q) has a curvature singularity as one approaches $r = 0$.

© Willie Wai-Yeung Wong

DRAFT COPY

8.18 (Interpretation as black hole solution) Let us now connect the Schwarzschild(-(Anti-)de Sitter) solution the notion of black holes introduced in the last lecture. To start with, we need to define the *domain of outer communication*. Our description in the previous lecture asks us to think about whether a point $p \in M$ can shoot out a null geodesic ray that is future-complete and achronal. Intuitively, for this to work we want to choose a ray that tries to "leave any compact region of the space time as fast as possible"; and this should be accomplished by looking for rays that have no spherical components. Mathematically this is possible due to the fact that Q embeds in M as a totally geodesic submanifold.

For our analysis, we rely on a clever trick: let γ be an affine null geodesic. To study how the $r \circ \gamma$ changes, we can first ask about the value of $g_Q(\dot\gamma, \nabla r)$, which is the rate of change of $r \circ \gamma$. Now, observe that K and ∇r are orthogonal, and have the same length, as $\dot\gamma$ is null and tangent to Q, we must have then

$$|g_Q(\dot\gamma, \nabla r)| = |g_Q(\dot\gamma, K)|.$$

Now, K is a Killing vector field, and hence for affine geodesics the value $g_Q(\dot\gamma, K)$ is a *constant of motion*. Therefore we have shown that $r \circ \gamma$ is a *linear function of the affine parameter*. This shows, in particular, if $r \circ \gamma$ is initially increasing, then the geodesic ray emanating from it must be future-complete. (Achronality also follows, but require some additional technical arguments.)

8.19 Proposition
In a Schwarzschild(-(Anti-)de Sitter) solution, a sufficient condition for a point to belong to the domain of outer communication is for there to be a future-directed null vector L at that point with $L(r) \geq 0$. In particular, the set $\{\mu(r) > 0\}$ belongs to the domain of outer communication. ■

Proof. The first claim is proved above the proposition statement. For the second: note that $\mu(r) > 0$ implies that ∇r is space-like. Assuming that we have chosen our orientation such that K is future-directed at this point, then one of $K + \nabla r$ and $K - \nabla r$ is a future null vector along which r is increasing. □

8.20 Next we consider those points where $\mu(r) < 0$. These we can split into those points for which ∇r is future time-like and those for which ∇r is past time-like. Where ∇r is past timelike, we see that $g_Q(L, \nabla r) > 0$ for any future null vector L. This indicates that such points must also belong to the domain of outer communication.

It remains to consider those points for which ∇r is future time-like. As ∇r is proportional to $-H$, the mean curvature vector of the corresponding sphere, we see that ∇r being future time-like indicates the sphere as being *future trapped*. This gives strong indications that we should consider those points with $\mu(r) < 0$ and ∇r future time-like correspond to the *black hole region*, by virtue of Penrose's theorem.

We can make good on this intuition with a direct computation: letting γ be an arbitrary future-directed affine null geodesic, which may be no longer tangent to Q. In the region where ∇r is future-time-like, the Killing field K is space-like. That $\dot\gamma$ is null then requires

$$|g_Q(\dot\gamma, \nabla r)| \geq |g_Q(\dot\gamma, K)|.$$

The conservation of $g_Q(\dot\gamma, K)$ implies therefore that $r \circ \gamma$ must decay faster than linearly. This means one of two things must happen:

© Willie Wai-Yeung Wong

DRAFT COPY

1. $r \circ \gamma$ may decay in such a way that it exits the region where $\mu(r) < 0$ in finite affine parameter. In this case, the initial point would belong to the causal past of the domain of outer communication, and hence is also in the domain of outer communication. This possibility is only possible with $\Lambda > 0$.
2. $r \circ \gamma$ decay to zero with finite affine parameter, in which case the geodesic γ is incomplete and terminate in a curvature singularity.

INSERT FIGURE HERE OF CAUSAL DIAGRAM

8.21 Finally, let us talk about apparent and event horizons. Our discussions above shows that $\mu(r) < 0$ has at least two connected regions. The black hole region corresponds to the connected components where

- $\mu(r) < 0$,
- ∇r is future time-like,
- $\inf r$ on the region is 0.

Therefore we conclude that the event horizon is an appropriate subset of $\{\mu(r) = 0\}$. On the other hand, the set $\{\mu(r) = 0\}$ corresponds precisely to those points in Q whose corresponding spheres have null mean curvature vectors. And it is not too hard to see here that this means the event horizon *can be foliated by marginally future trapped surfaces*. So this for the Schwarzschild(-(Anti)-de Sitter) solutions, the event horizon and apparent horizons coincide.

Spherically Symmetric, Asymptotically Flat Black Holes

8.22 While the Schwarzschild solution is very simple (at least in terms of its explicit formula), one may question its usefulness for helping us understand gravitational collapse. After all, in the domain of outer communication ("outside the black hole"), the Schwarzschild solution has a *time like* Killing vector field K. This means it can only model entirely static scenarios and not anything that is evolving in time.

This shortcoming is fundamental to the theory, as Birkhoff's Theorem ensures the symmetry. On the other hand, our discussion above shows that Birkhoff's Theorem is only in play if the stress energy tensor is such that its Q-tangent component is "pure trace". While this is certainly the case for vacuum ($T = 0$ is pure trace), it can be violated for most matter models. That is to say, spherical symmetry does *not* preclude dynamics once we add matter into consideration.

8.23 One should note that while the spherically symmetric models have a venerable history in driving our early understanding of gravitational collapse, modern studies are more often based on numerical simulations outside of spherical symmetry, to better align with the more generic situation with non-zero total angular momentum. Nevertheless, these models provide a setting where we can actually rigorously discuss mathematically expected behaviors for gravitational collapse, and still serves as inspiration for more modern investigations in mathematical general relativity.

8.24 For simplicity, we will consider below the case where $\Lambda = 0$; some of the discussion also carry to cases where $\Lambda \neq 0$, but not always. Readers interested in non-zero

© Willie Wai-Yeung Wong

DRAFT COPY

cosmological constants should tread carefully. Note that results of current astronphysical observations still do not rule out $\Lambda = 0$.

8.25 As our goal is to provide some intuition for gravitational collapse of an *isolated* system, we shall make the following assumptions on our system.

- (M, g_M) is spherically symmetric;
- There is an "instant in time" where a spherically symmetric spatial cross section of M has topology $\mathbb{R}^n$, with $\{r = 0\}$ a single point corresponding to the center of symmetry; (note that this is *not* satisfied by the Schwarzschild solution)
- At points along this "instant in time", $\mu(r) > 0$; and
- On the spatial cross section, there is a radius R such that the geometry at $r > R$ is that of Schwarzschild, and the energy momentum tensor vanishes there.

A precise description of the second and third conditions can be found in Christodoulou, "A mathematical theory of gravitational collapse". The final condition is not obviously satisfiable, in view of the fact that initial data for the Einstein equations cannot be freely prescribed, thanks to the so-called *constraint equations* that need to be satisfied. This is only understood more recently due to the gluing construction originally considered by Justin Corvino and Rick Schoen, a modern exposition of which can be found in Delay, "Smooth compactly supported solutions of some underdetermined elliptic PDE, with gluing applications".

The reason that we make the second and third assumptions is to ensure that no black hole is present *initially* in our system. Note that those two conditions are satisfied by Minkowski space; the assumptions ensure that the initial situation of our space-time is still somewhat similar to Minkowski space, so that it can be interpreted as in a situation where the gravitational effects are not yet too pronounced. The third condition specifically ensures that none of the symmetry spheres at the initial instant are trapped.

The final condition is one way to ensure that the solution is "asymptotically flat". It ensures that as $r \to \infty$ the space-time metric converges to that of Minkowski space. It can be replaced by weaker assumptions on asymptotic behavior, but we shall not explore that in this lecture.

8.26 In the quotient Q, the initial instant in time we shall denote by the curve $\mathring{Q}$. Denote by Q_+ the subset where $\mu(r) > 0$. Our assumptions imply that $\mathring{Q} \subseteq Q_+$. Assume our choice of orientation is such that on Q_+ the Kodama vector field K is future-directed. Since Q is two-dimensional and Lorentzian, it admits two future directed null vector fields $L_\pm$. We choose L_+ so that on Q_+ it is parallel to $K + \nabla r$, and L_- such that it is parallel to $K - \nabla r$.

By definition then $L_-(r) < 0$ on Q_+ and $L_+(r) > 0$. Note that $L_\pm(r) = g(L_\pm, \nabla r)$ is exactly the *null expansion* of the $\{p\} \times \mathbb{S}^{n-1}$ spheres in the direction $L_\pm$.

8.27 Now we impose the assumption that our matter model satisfies the null energy condition. This in particular implies (since $\Lambda = 0$) that $\mathrm{Ric}(L_+, L_+) = \mathrm{Ric}(L_-, L_-) \geq 0$. From our set-up it is clear that every $q \in Q$ to the future of $\mathring{Q}$ can be joined to $\mathring{Q}$ via an integral curve of L_-. So using Proposition 6.32 (or simply the null Raychaudhuri equation) we conclude that to the future of $\mathring{Q}$, the null expansion in the direction of L_- is always negative.

Draw a picture!

© Willie Wai-Yeung Wong

This means the sign of $\mu(r)$ is entirely determined in this region by the sign of the null expansion in the direction of L_+, or the value of $L_+(r)$. If $L_+(r)$ is positive, then $\mu(r) > 0$, and if $L_+(r) < 0$, then $\mu(r) < 0$.

INSERT FIGURE FOR THE NEXT PARAGRAPH

8.28 If we follow through the proof of Theorem 7.17, we would see that under the hypotheses provided above, if q is a point to the future of $\mathring{Q}$ at which $L_+(r) < 0$, then the null geodesic in the direction of L_+ must be future incomplete. Based on the idea that the L_+ null geodesics representing "radial" motion and hence has the best chance "escaping a gravitational well", we can relax the notion of the domain of outer communication to only requiring the L_+ geodesics to be future complete. Then we see the following conclusions:

- A necessary condition for a point q to be in the domain of outer communications is that the L_+ geodesic emanating from it remain always in Q_+.
- This implies, in particular, that the black hole region must contain any point at which $L_-(r) < 0$ and $L_+(r) < 0$; call this the "trapped set" $\mathcal{T}$.
- By Proposition 6.32, and the assumption that $L_-(r) < 0$ always to the future of $\mathring{Q}$, we see that if $p \in \mathcal{T}$, then the L_+ geodesic ray emanating from p must also belong to $\mathcal{T}$.
- It is however possible (in general) to have points of Q_+ in the future points of $\mathcal{T}$.
- The boundary $\partial\mathcal{T}$ is foliated by marginally future-trapped spheres, and hence form an apparent horizon.
- Therefore in general it is possible to the event horizon an apparent horizon to not coincide; but the apparent horizon must not enter the domain of outer communication.

We emphasize that these conclusions are only true / provable in the context of spherical symmetry.

DRAFT COPY

Part II

Second Semester

DRAFT COPY

DRAFT COPY

Topic 9 (2024/04/01)

Einstein's Equations as Partial Differential Equations

9.1 In the first semester course, while we introduced the notion of *Einstein's equation*, our mathematical discussion never really depended on it. Indeed, the results proven are largely about the geometry of *geodesics* in arbitrary Lorentzian manifolds, perhaps with certain control on its *Ricci curvature*. The (minimal) extent to which Einstein's equation is involved comes from the fact that it relates the Ricci curvature of a solution to the physical *stress–energy tensor*; thus positivity conditions placed on the stress–energy tensor, which in turn are often justified either through the explicit models or through physical arguments, can be transformed into various control on the Ricci curvature. Indeed, we referred to such assumptions on Ricci curvature as *energy conditions*.

9.2 Our goal in this semester is to now focus more on Einstein's equation proper, and we shall restrict ourselves to thinking about solutions (and in some cases approximate solutions) to these equations. To get the appropriate quantitative understanding, our tools will be those from the theory of partial differential equations.

9.3 Our first problem, however, is *how* to formulate Einstein's equation as a system of partial differential equations. Recall that Einstein's equation is written as

$$\mathrm{Ric} - \frac{1}{2}Rg + \Lambda g = T; \tag{9.3.1}$$

for mathematical simplicity we will focus on the *vacuum* case where the stress–energy tensor $T \equiv 0$. In this case taking the trace we find $R - \frac{n}{2}R + n\Lambda = 0$, where n is the space-time dimension, so $R = \frac{2n}{n-2}\Lambda$. We have therefore the *Einstein–vacuum equations*

$$\mathrm{Ric} = \frac{2}{n-2}\Lambda g, \tag{9.3.2}$$

the Riemannian analogue of which having been studied extensively by many authors, including the pseudonymous Arthur Besse.

DRAFT COPY

9.4 Recall that our unknowns are: a space-time manifold M and a Lorentzian metric g, so we would like to write (9.3.2) as a partial differential equation for the quantity g. This we expect since morally speaking the curvature should be a "second derivative quantity in g". At this point, of course we can just throw in a local coordinate system and compute everything in Christoffel symbols. We however take the following philosophical points:

- Many manifolds do not admit single global coordinate patches. On the other hand, solutions to partial differential equations are known to be dependent on the choice of boundary conditions. Precisely what boundary conditions one should supply on the boundaries of the coordinate patches is unclear, and introduce an additional complication for the analysis.
- In general one should avoid prematurely converting things to local coordinates, as the formulae often become more complicated and obscure important structures that one may exploit from the invariant formulation.

Taking for granted we wish to work invariantly, we now re-encounter the fundamental motivation for differential geometry: given a vector bundle $V \to M$, while the fibers V_p and V_q are isomorphic for any $p, q \in M$, they are not *canonically* so. So to do differential calculus for sections of vector bundles require choosing a (linear) connection. Since we are working on a Lorentzian manifold, we do happen to have a connection lying around, namely the Levi-Civita connection ∇ of the metric g. But it is of course hopeless to try to use ∇ to study g: by definition $\nabla g = 0$ and no information could be gleamed from examining the "derivatives"! This should be interpreted as ∇ being itself part of the unknown (being a derived quantity from g), and so studying g relative to ∇ runs the risk of being "circular".

9.5 One way to resolve this issue and still maintain our invariant formulation would be to equip our space-time M with a "reference" linear connection, and perform our computations relative to this connection. As every manifold M can be equipped with a Riemannian metric, it is convenient to just fix a Riemannian metric $\mathring{h}$ and denote by $\mathring{\nabla}$ its Levi-Civita connection.

9.6 (The manifold M) One may worry at this point that we have been thoroughly ignoring the manifold M, which also is part of the "unknown" in the theory. This turns out *not* to be a worry. As we will see, the "correct" problem to ask for solving Einstein's equations is the *initial vale problem*. And the topology type of the manifold M for this problem can be completely determined by that of the initial value. And so for the analytical aspects it is valid to assume that we have a given manifold M on which we are solving for an unknown metric g (that may only be regularly defined on some open subset of M).

9.7 Example

In fact, coordinate charts can also be interpreted using this language.

A coordinate chart is given by a mapping $\Phi : U \to V$ where $U \subseteq M$ and $V \subseteq \mathbb{R}^n$ are open sets. Restricting now to the domain U, the diffeomorphism Φ can be used to pull back the standard Euclidean metric e from $\mathbb{R}^n$ to $\mathring{h} = \Phi^* e$. If we let ∂_i to be the coordinate vector fields of the coordinate chart, and dx^i the coordinate one-forms, the Euclidean

© Willie Wai-Yeung Wong

DRAFT COPY

Levi-Civita connection is such that

$$\mathring{\nabla}_{\partial_i}\partial_j = 0, \qquad \mathring{\nabla}_{\partial_i} dx^j = 0.$$

So given a (p,q)-tensor field T we have

$$(\nabla_{\partial_i} T)(\partial_{i_1},\dots,\partial_{i_q}; dx^{j_1},\dots,dx^{j_p}) = \partial_i\Big[T(\partial_{i_1},\dots,\partial_{i_q}; dx^{j_1},\dots,dx^{j_p})\Big]$$

showing that we can interpret computations in a local coordinate system as the same as computations with respect to the pullback Euclidean connection. ◊

9.8 Proposition

Let M be a manifold. Suppose we are given a Riemannian metric $\mathring{h}$ and a pseudo-Riemannian metric g on M. Denote by $\mathring{\nabla}$ and ∇ their corresponding Levi-Civita connections, then $\nabla - \mathring{\nabla}$ is tensorial. ■

Proof. It is clear that given X, Y vector fields, the mapping $(X,Y) \mapsto \nabla_X Y - \mathring{\nabla}_X Y$ is $\mathbb{R}$-bilinear. It suffices to show that for any scalar function ψ, we have $\nabla_{\psi X} Y - \mathring{\nabla}_{\psi X} Y = \psi\Big(\nabla_X Y - \mathring{\nabla}_X Y\Big) = \nabla_X(\psi Y) - \mathring{\nabla}_X(\psi y)$. The first equality is obvious. For the second, observe that by the Leibniz rule we have

$$\nabla_X(\psi Y) = X(\psi)Y + \psi\nabla_X Y$$

and as all connections act identically on scalar functions, those terms cancel when we subtract. □

9.9 Definition

Let M be a manifold, and suppose we are given a Riemannian metric $\mathring{h}$ and a pseudo-Riemannian metric g on M. We will denote by Γ the $(1,2)$-tensor field on M satisfying $\Gamma(X,Y) = \nabla_X Y - \mathring{\nabla}_X Y$. We call it the Christoffel symbol of g relative to $\mathring{h}$. ◊

9.10 Computations using the relative Christoffel symbol are largely the same as those using the Christoffel symbols in coordinates. For example, we can extend its usage to covariant tensors by noting (for ω a one form)

$$X(\omega(Y)) = (\nabla_X\omega)(Y) + \omega(\nabla_X Y) = (\mathring{\nabla}_X\omega)(Y) + \omega(\mathring{\nabla}_X Y)$$

and hence

$$(\nabla_X\omega - \mathring{\nabla}_X\omega)(Y) = -\omega(\nabla_X Y - \mathring{\nabla}_X Y) = -\omega(\Gamma(X,Y)).$$

(In abstract index notation, we have therefore $\nabla_a Y^b = \mathring{\nabla}_a Y^b = \Gamma^b_{ac} Y^c$ and $\nabla_a\omega_b = \mathring{\nabla}_a\omega_b - \Gamma^c_{ab}\omega_c$.)

9.11 Exercise

Prove (using that $\nabla g = 0$ and expanding this in terms of $\mathring{\nabla}$ and relative Christoffel symbols)

$$\Gamma^a_{bc} = \frac{1}{2}(g^{-1})^{ad}\Big(\mathring{\nabla}_b g_{dc} + \mathring{\nabla}_c g_{db} - \mathring{\nabla}_d g_{bc}\Big).$$ ♠

© Willie Wai-Yeung Wong

DRAFT COPY

9.12 Recall now that we defined

$$\mathrm{Riem}(X,Y)Z = \nabla_{[X,Y]}Z - [\nabla_X, \nabla_Y]Z$$

Expanding this formula using the relative Christoffel symbol, we find

$$\begin{aligned}\mathrm{Riem}(X,Y)Z &= \mathring{\nabla}_{[X,Y]}Z + \Gamma([X,Y],Z) - \Big[\mathring{\nabla}_X + \Gamma(X,\text{-}), \mathring{\nabla}_Y + \Gamma(Y,\text{-})\Big]Z \\ &= \mathring{\mathrm{Riem}}(X,Y)Z - (\mathring{\nabla}_X\Gamma)(Y,Z) + (\mathring{\nabla}_Y\Gamma)(X,Z) - \Gamma(X,\Gamma(Y,Z)) + \Gamma(Y,\Gamma(X,Z)). \end{aligned} \tag{9.12.1}$$

As from Exercise 9.11 we see that Γ is bilinear in g^{-1} and $\mathring{\nabla}g$, our (9.12.1) reflects our expectation that the curvature involves second derivatives of the metric. Taking the trace we can obtain a similar formula for the Ricci curvature; it is more convenient to record this using abstract index notation.

$$\mathrm{Ric}_{ac} = \mathring{\mathrm{Ric}}_{ac} + \mathring{\nabla}_b\Gamma^b_{ac} - \mathring{\nabla}_a\Gamma^b_{bc} + \Gamma^b_{bd}\Gamma^d_{ac} - \Gamma^b_{ad}\Gamma^d_{bc}. \tag{9.12.2}$$

9.13 As now we see that the Einstein vacuum equation (9.3.2) is a system of second order partial differential equations, it is natural to ask about its type. Specifically, we want to know whether this is elliptic, parabolic, or hyperbolic. For this we will focus on only the second derivative terms that show up in (9.12.2), which we find after expanding using Exercise 9.11.

$$\mathrm{Ric}_{ac} = \mathring{\mathrm{Ric}}_{ac} + \mathcal{P}(g^{-1}, \mathring{\nabla}g) + \frac{1}{2}(g^{-1})^{bd}\Big[\mathring{\nabla}_b\mathring{\nabla}_a g_{dc} + \mathring{\nabla}_b\mathring{\nabla}_c g_{da} - \mathring{\nabla}_b\mathring{\nabla}_d g_{ac} - \mathring{\nabla}_a\mathring{\nabla}_c g_{db}\Big]. \tag{9.13.1}$$

The second order derivative terms in (9.13.1) *do not have an obvious type*. Before we explain how this is resolved, let us do some seemingly pointless manipulations first. Interchanging the derivatives we find

$$\begin{aligned}\mathrm{Ric}_{ac} = \mathring{\mathrm{Ric}}_{ac} &+ \mathcal{P}(g^{-1}, \mathring{\nabla}g) - \frac{1}{2}(g^{-1})^{bd}\mathring{\nabla}_b\mathring{\nabla}_d g_{ac} \\ &+ \frac{1}{4}(g^{-1})^{bd}\Big[2\mathring{\nabla}_a\mathring{\nabla}_b g_{dc} + 2\mathring{\nabla}_c\mathring{\nabla}_b g_{da} - \mathring{\nabla}_c\mathring{\nabla}_a g_{db} - \mathring{\nabla}_a\mathring{\nabla}_c g_{db}\Big] \\ &+ \frac{1}{2}(g^{-1})^{bd}\Big[\mathring{\mathrm{Riem}}_{bad}{}^f g_{fc} + \mathring{\mathrm{Riem}}_{bac}{}^f g_{fd} + \mathring{\mathrm{Riem}}_{bcd}{}^f g_{fa} + \mathring{\mathrm{Riem}}_{bca}{}^f g_{fd}\Big]\end{aligned}$$

This allows us to write

$$\begin{aligned}\mathrm{Ric}_{ac} = \mathring{\mathrm{Ric}}_{ac} + (g^{-1}\cdot\mathring{\mathrm{Riem}}\cdot g)_{ac} + \mathcal{P}(g^{-1}, \mathring{\nabla}g) - \boxed{\frac{1}{2}(g^{-1})^{bd}\mathring{\nabla}_b\mathring{\nabla}_d g_{ac}} \\ + \frac{1}{2}g_{fc}\mathring{\nabla}_a\Big[(g^{-1})^{bd}\Gamma^f_{bd}\Big] + \frac{1}{2}g_{fa}\mathring{\nabla}_c\Big[(g^{-1})^{bd}\Gamma^f_{bd}\Big]\end{aligned} \tag{9.13.2}$$

Now, if it were to happen *for some still mysterious reason* that the terms on the final line can be completely ignored, then the only remaining second order derivative in the expression, which is boxed up, would be of *hyperbolic type*, as the coefficients are given by a matrix with Lorentzian signature. (In fact, one may even say that in this setting the Einstein-vacuum equation is can be expressed as a system of quasilinear wave equations.)

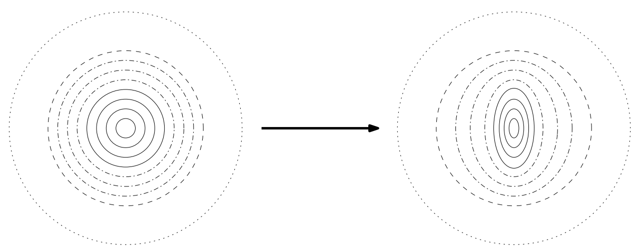
© Willie Wai-Yeung Wong

DRAFT COPY

9.14 The justification for this simplification falls upon the double-edged sword known as *gauge ambiguity*. Fundamentally, the issue is that the equation (9.3.2) is *diffeomorphism invariant*: if $\phi : M \to M$ is a diffeomorphism, and g a solution, then $\phi^* g$ is also a solution. Now, as in the smooth category morphisms abound, we obtain the following negative result:

Proposition
Let (M, g) *be a solution to the Einstein–vacuum equations* (9.3.2) *(with dimension* $n \geq 3$*). Then given any point* $p \in M$*, there exists an open set* $U \ni p$ *with* $U \subsetneq M$ *such that there are* continuum–many *solutions* g' *that satisfy (a)* $g = g'$ *on* $M \setminus U$ *and (b)* $g \neq g'$ *at* p. ■

Proof. The construction is entirely local, so we will work in a coordinate chart, with p identified as the origin. Assume without loss of generality that our coordinate chart contains the ball of (coordinate) radius 2 centered at the origin. We can construct diffeomorphisms ϕ of the ball of radius 2 to itself, that is identity outside the ball of radius 1, but corresponds to a shear with in the ball of radius $\frac{1}{2}$; see illustration below.



It is clear that there are continnum–many different "aspect ratios", each of which giving different differential at the origin. Letting $g' = \phi^* g$ we see that the proposition holds. □

9.15 An immediate consequence is that the initial, initial-boundary, and boundary value problems for the Einstein vacuum equation *cannot* have unique solutions. In terms of the PDE (9.13.1) this means that the equation is *vastly underdetermined*. The practical upshot are two fold:

1. If we wish to study the *uniqueness* of solutions of Einstein's equations, it only makes sense to consider uniqueness *up to diffeomorphism*.
2. If we wish to study the *existence* of solutions to Einstein's equations, we can try to impose *extra gauge conditions* that kill the diffeomorphism degree of freedom.

Note that in the latter, what's important is *consistency*. Imposing an additional gauge condition may allow us to obtain a simplified equation from (9.13.1), which can then be solved. But in order to guarantee that this still generates a *bona fide* solution to the original Einstein vacuum equations, we need to show that the gauge conditions together with the reduced equations are equivalent to the original PDE *and* that the gauge conditions can be satisfied.

9.16 For convenience, let us define the *reduced Ricci operator*

$$\widetilde{\mathrm{Ric}}[g, \mathring{h}]_{ac} := \mathring{\mathrm{Ric}}_{ac} + (g^{-1} \cdot \mathring{\mathrm{Riem}} \cdot g)_{ac} + \mathcal{P}(g^{-1}, \mathring{\nabla} g) - \frac{1}{2}(g^{-1})^{bd} \mathring{\nabla}_b \mathring{\nabla}_d g_{ac} \qquad (9.16.1)$$

© Willie Wai-Yeung Wong

DRAFT COPY

and *gauge operator*

$$\mathfrak{X}^f := (g^{-1})^{bd}\Gamma^f_{bd}. \tag{9.16.2}$$

So that

$$\mathrm{Ric}_{ac} = \widetilde{\mathrm{Ric}}[g,\mathring{h}]_{ac} + \frac{1}{2}\Big(g_{fc}\mathring{\nabla}_a + g_{fa}\mathring{\nabla}_c\Big)\mathfrak{X}^f.$$

Then

Lemma

Given a fixed Riemannian metric $\mathring{h}$ on M, a sufficient condition for a Lorentzian metric g to be a solution to the Einstein–vacuum equation is that both $\widetilde{\mathrm{Ric}}[g,\mathring{h}] = \frac{2}{n-2}\Lambda g$ and $\mathfrak{X} = 0$. ■

9.17 (Equation for $\mathfrak{X}$) Of the two conditions in the previous Lemma, we have already established that the condition concerning the reduced Ricci operator is given by quasilinear wave equation. But what about the condition concerning $\mathfrak{X}$, which depends on the relative Christoffel symbols? For this, we shall appeal to the Bianchi identities for the metric g, one of whose consequence is that the tensor $\mathrm{Ric} - \frac{1}{2}Rg$ is divergence free. We can write the Einstein tensor as

$$\underbrace{\widetilde{\mathrm{Ric}}[g,\mathring{h}]_{ac} - \frac{1}{2}(g^{-1})^{bd}\widetilde{\mathrm{Ric}}[g,\mathring{h}]_{bd}g_{ac}}_{\tilde{G}[g,\mathring{h}]_{ac}} + \frac{1}{2}\Big(g_{fc}\mathring{\nabla}_a + g_{fa}\mathring{\nabla}_c\Big)\mathfrak{X}^f - \frac{1}{2}g_{ac}\mathring{\nabla}_f\mathfrak{X}^f. \tag{9.17.1}$$

Hence the second Bianchi identity implies

$$(g^{-1})^{ce}\nabla_e\tilde{G}[g,\mathring{h}]_{ac} + \frac{1}{2}\Big(\nabla_f\mathring{\nabla}_a + g_{fa}(g^{-1})^{ce}\nabla_e\mathring{\nabla}_c - \nabla_a\mathring{\nabla}_f\Big)\mathfrak{X}^f = 0.$$

This identity holds for *all* pairs of Lorentzian metric g and Riemannian metric $\mathring{h}$. We can simplify it further. First

$$\nabla_f\mathring{\nabla}_a\mathfrak{X}^f - \nabla_a\mathring{\nabla}_f\mathfrak{X}^f = \mathring{\mathrm{Ric}}_{ae}\mathfrak{X}^e + \Gamma^f_{fe}\mathring{\nabla}_a\mathfrak{X}^e - \Gamma^f_{ae}\mathring{\nabla}_f\mathfrak{X}^e$$

So we simplify to

$$(g^{-1})^{ce}\nabla_e\tilde{G}[g,\mathring{h}]_{ac} + \frac{1}{2}\Big(\mathring{\mathrm{Ric}}_{ae}\mathfrak{X}^e + \Gamma^f_{fe}\mathring{\nabla}_a\mathfrak{X}^e - \Gamma^f_{ae}\mathring{\nabla}_f\mathfrak{X}^e - g_{fa}\mathfrak{X}^m\mathring{\nabla}_m\mathfrak{X}^f + g_{fa}(g^{-1})^{ec}\Gamma^f_{em}\mathring{\nabla}_c\mathfrak{X}^m\Big) + \frac{1}{2}g_{fa}(g^{-1})^{ce}\mathring{\nabla}_e\mathring{\nabla}_c\mathfrak{X}^f = 0.$$

So schematically, $\mathfrak{X}$ solves a wave equation of the form

$$(g^{-1})^{ce}\mathring{\nabla}_e\mathring{\nabla}_c\mathfrak{X} = \mathcal{Q}(\mathring{\mathrm{Ric}},\Gamma,g^{-1},\mathfrak{X})\cdot(\mathfrak{X},\mathring{\nabla}\mathfrak{X}) + \mathrm{div}\tilde{G}[g,\mathring{h}]. \tag{9.17.2}$$

From this we see that the existence of solutions to the Einstein–vacuum equations hinges on the existence and uniqueness theory of *nonlinear wave equations*.

Before we dive into this theory in a subsequent lecture, let me explain the geometric meaning behind the condition $\mathfrak{X} = 0$ and make some historical remarks.

© Willie Wai-Yeung Wong

DRAFT COPY

9.18 (Nomenclature) In the literature the case where one uses a single coordinate chart and $\mathring{h}$ as the Euclidean metric on the coordinate chart (see Example 9.7) is known as either the "wave coordinates condition", the "harmonic coordinates condition", or the "harmonic gauge condition". The case with more general $\mathring{h}$ on the entire manifold M is often called the "wavemap gauge". These names combine two observations:

- The requirement that $\mathfrak{X} \equiv 0$ is equivalent to requiring that the *identity map* $M \to M$, when considered as a mapping from a Lorentzian manifold (M, g) to a Riemannian manifold $(M, \mathring{h})$, is a "wave map".
- When the target manifold is flat, so locally diffeomorphic to $\mathbb{R}^n$, the coordinate functions of $\mathbb{R}^n$, when considered now as a scalar function on Lorentzian manifold (M, g), must solve the linear wave equation.

9.19 (Wave maps) Given a Lorentzian manifold (M, g) and a Riemannian manifold (N, h), we can consider the formal action on mappings $\phi : M \to N$ given by

$$\mathcal{S}[\phi] = \int \operatorname{tr}_g \phi^* h \, \mathrm{dvol}_g.$$

A formal critical point of this function is called a "wave map" (when (M, g) is a Riemannian manifold, this is the action for "harmonic maps"). Observe when (N, h) is the standard real line, we can treat ϕ as a real-valued function and

$$\operatorname{tr}_g \phi^* h = g^{-1}(d\phi, d\phi)$$

so the action is the standard Dirichlet integral, and the corresponding Euler-Lagrange equation is just the linear wave equation on the Lorentzian manifold (M, g). Essentially the same thing happens when (N, h) is flat, whereby the components of ϕ solve independently a linear wave equation each.

When (N, h) is not flat, the resulting Euler–Lagrange equations are more complicated. If we let y^A denote local coordinates on (N, h) and x^μ the local coordinates for (M, g), the action can be written in terms of this coordinate representation as

$$\mathcal{S}[\phi] = \int (g^{-1})^{\mu\nu} \partial_\mu \phi^A \partial_\nu \phi^B h_{AB} \circ \phi \, \mathrm{dvol}_g.$$

Taking a formal variation we find that we need

$$-h_{AB} \Box_g \phi^B - (g^{-1})^{\mu\nu} \partial_\nu \phi^B \partial_\mu \phi^C \partial_C h_{AB} \circ \phi + \frac{1}{2} (g^{-1})^{\mu\nu} \partial_\mu \phi^B \partial_\nu \phi^C \partial_A h_{BC} \circ \phi = 0.$$

Here $\Box_g$ is the Laplace-Beltrami operator of the Lorentzian metric g. We can write this in terms of the Christoffel symbol of h as

$$\Box_g \phi^A + (g^{-1})^{\mu\nu} \partial_\nu \phi^B \partial_\mu \phi^C \cdot {}^{(h)}\Gamma^A_{BC} \circ \phi = 0. \tag{9.19.1}$$

9.20 Now consider the case where $N = M$ and ϕ is the identity map. In this case we can consider x and y to be the same coordinate system. The wave maps equation becomes in this case the equation

$$\Box_g \phi^\kappa + (g^{-1})^{\mu\nu} \cdot {}^{(\mathring{h})}\Gamma^\kappa_{\mu\nu} = 0.$$

© Willie Wai-Yeung Wong

DRAFT COPY

The first term can be written as

$$(g^{-1})^{ab}\nabla_a(d\phi^\kappa)_b$$

where $(d\phi^\kappa)_b$ is the family of *one-forms* indexed by κ. Using the relative Christoffel symbol we have therefore the wave maps equation is equivalent to

$$(g^{-1})^{ab}\mathring{\nabla}_a(d\phi^\kappa)_b - \mathfrak{X}^c(d\phi^\kappa)_c + (g^{-1})^{\mu\nu}\cdot {}^{(\mathring{h})}\Gamma^\kappa_{\mu\nu} = 0.$$

The first and third terms combine to form the *covariant derivative of the* $(1,1)$*-tensor* $d\phi$, which as ϕ is the identity map is also the identity map on the tangent space. In particular, it has zero covariant derivative. Therefore we have found that the identity map from M to itself being a wave map is exactly equivalent to $\mathfrak{X} \equiv 0$.

9.21 (Historical remarks) This method of casting the Einstein–vacuum equations as a coupled system of nonlinear wave equations was first treated seriously (in the case where M has a single coordinate chart and $\mathring{h}$ is the Euclidean metric) by Choquet-Bruhat in her seminal work establishing the local well-posedness of the initial value problem in general relativity, and is also used in the work of Lindblad and Rodnianski re-proving the stability of Minkowski space.

Fourès-Bruhat, "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires"

Lindblad and Rodnianski, "Global Existence for the Einstein Vacuum Equations in Wave Coordinates"; Lindblad and Rodnianski, "The global stability of Minkowski space-time in harmonic gauge"

The Riemannian analogue also has wide applications. In direct analogy of our discussion above, for a Riemannian Einstein manifold, by using harmonic coordinates one sees that the Einstein condition becomes a quasilinear elliptic partial differential equation on the metric components. Using this DeTurck and Kazdan showed that all Riemannian Einstein metrics are real analytic in harmonic coordinates.

DeTurck and Kazdan, "Some regularity theorems in Riemannian geometry"

This idea of exploiting the gauge degree of freedom in geometric equations has also been used in great effect in the study of Ricci flow. Naive formulations of the Ricci flow equation gives a partial differential equation with no obvious type. The so-called DeTurck trick is essentially the same as the use of the wave map gauge described above, and this allows one to kill the gauge ambiguity and at the same time obtain a parabolic system of partial differential equations as a result.

© Willie Wai-Yeung Wong

DRAFT COPY

Topic 10 (2024/04/08)

The Initial Value Problem; Constraint Equations

In the previous lecture we showed that, under a convenient choice of gauge, Einstein's equations can be viewed as a coupled system of nonlinear wave equations. This suggest that the equations of general relativity are fundamentally *evolutionary*. Hence the natural question to ask (also motivated by its stature as a *physical* theory) is the *initial value problem*, where we predict what happens in the future given knowledge of the current state of the universe. The main purpose of this lecture is to explore this idea further. Specifically, we will formulate what constitutes admissible initial data for the Einstein–vacuum equations, and sketch the proof of (local) existence and uniqueness of solutions.

Additionally, in the previous lecture the use of the wave maps gauge was motivated from the expression of the Ricci curvature not taking a well-defined type. Many may consider this very flimsy as a justification. In this lecture we will provide an alternative point of view that may be more satisfactory, that is tied to certain geometric facts about submanifold embeddings in pseudo-Riemannian manifolds.

The "$1 + n$" splitting

10.1 Having accepted that Einstein's equations as fundamentally evolutionary, let's think back to how one usually models evolutionary behavior in mathematics. One typically has a time parameter t, and a space of configurations $\mathcal{C}$ represented all possible states of the system at any given time. This space $\mathcal{C}$ may be finite dimensional or infinite dimensional (as in the case of partial differential equations); it may be a vector space or a manifold. The dynamics is then represented by a vector field X on the configuration space, and solutions are curves γ in $\mathcal{C}$ with $\dot{\gamma} = X \circ \gamma$.

10.2 Now this approach seems ill-suited as first sight for general relativity, since the theory of relativity is all about the lack of a canonical choice of time. Nevertheless, this approach is also how modern mathematical physics interprets *quantum mechanics*, and hence is especially favored by those researchers interested in quantum gravity. Additionally, one can argue that even though physically the notion of a global time may not be canonical, its use mathematically to study the solutions can still be valid.

To implement such an approach, one needs to first *define* a notion of time. We do so by specifying a time function $t : M \to \mathbb{R}$ with dt being time-like (so in particular non-zero). From causality considerations we may assume (as is physically reasonable) that the level sets $\{t = const\}$ are all diffeomorphic to the same smooth manifold Σ. To distinguish between the different leaves, we will write Σ_t for the different level sets.

We shall also denote by

$$n = -\frac{1}{\sqrt{|g(dt,dt)|}} dt^\sharp \tag{10.2.1}$$

the unique, unit timelike vector field on M that is orthogonal to each of the Σ_t and satisfies $n(t) > 0$.

10.3 Notice that the Σ_t are diffeomorphic but not canonically so. We can make a *choice* of diffeomorphism by choosing a diffeomorphism $M \cong \mathbb{R} \times \Sigma$ (assuming without loss of generality that t is onto). For convenience we will make one such choice, but for maximum flexibility we will not consider this choice as fixed or canonical. With this diffeomorphism the tangent space TM splits into $T\mathbb{R} \oplus T\Sigma$, and we will also denote by ∂_t the vector field that is in the $T\mathbb{R}$ direction and satisfies $\partial_t(t) = 1$.

10.4 Definition
For each choice of a diffeomorphism $M \cong \mathbb{R} \times \Sigma$ we define two quantities, capturing its main geometric features. Since we know that TM also decomposes into $\operatorname{span}\{n\} \oplus T\Sigma$, *we can write*

$$\partial_t = Nn + \zeta \tag{10.4.1}$$

- *The scalar N is called the* lapse function *and measures how "fast" the time function t flows with respect to the proper-time as defined by the space-time metric g.*
- *The Σ-tangent vector field ζ is called the* shift *and measures how much curves which represent "constant spatial location" (as defined by the diffeomorphism $M \cong \mathbb{R} \times \Sigma$) fail to be orthogonal to the surfaces Σ_t.* ◊

10.5 The space-time metric g can also be decomposed relative to this diffeomoprhism. First note that g induces on the level sets $\Sigma_t \cong \Sigma$ a time-dependent Riemannian metric, which we label as γ; we therefore have $g = -n^\flat \otimes n^\flat + \gamma$. Given vector fields $A\partial_t + X$ and $B\partial_t + Y$, we can compute

$$\begin{aligned} g(A\partial_t + X, B\partial_t + Y) &= g(ANn + A\zeta + X, BNn + B\zeta + Y) \\ &= -ABN^2 + AB\gamma(\zeta,\zeta) + A\gamma(\zeta, Y) + B\gamma(\zeta, X) + \gamma(X, Y) \end{aligned}$$

So

$$g = \left(-N^2 + \gamma(\zeta,\zeta)\right) dt \otimes dt + dt \otimes \zeta^\flat + \zeta^\flat \otimes dt + \gamma. \tag{10.5.1}$$

DRAFT COPY

DRAFT COPY

10.6 For our discussions below, it is helpful to keep in mind the following idea: the induced metric γ on Σ_t are well-defined regardless of the choice of the diffeomorphism $M \cong \mathbb{R} \times \Sigma$, and so is in some sense a more fundamental object. The only criterion for its definition is the choice of the time function t. On the other hand, the lapse and shift do depend on the specific diffeomorphism, and hence are less fundamental.

As a consequence, we will focus first on discussing the induced metric γ and its evolution. The lapse and shift we expect will be tied to the choice of diffeomoprhism and will be set by such "gauge choices" later.

10.7 As described last lecture, Einstein's equation defines a second order partial differential equation for the metric g. So from the first order point of view, the elements of the configuration space $\mathcal{C}$ should include both the spatial metric γ and its "first time derivative". From the lessons from last time we see that we shouldn't take derivatives using the Levi-Civita connection of the metric g, since it contains the unknown to be solved for. Last time we introduced a background connection to solve this problem. This time we can use a different idea: associated to the slicing of M by the time function t is a natural vector field n. (We prefer n to ∂_t as the former is canonically associated to the time function t, while the latter requires the explicit choice of diffeomorphism.) Given that n is time-like, a notion of time differentiation can be *Lie differentiation by n*.

The Lie derivative of γ can be related to that of g through the expression $\gamma = g + n^\flat \otimes n^\flat$, so

$$\mathcal{L}_n \gamma = \mathcal{L}_n g + \mathcal{L}_n g(\,\text{-}\,, n) \otimes n^\flat + n^\flat \otimes \mathcal{L}_n g(\,\text{-}\,, n);$$

here we used that $\mathcal{L}_n n = [n, n] = 0$. Furthermore, using that $g(n,n) = -1$, we have

$$\mathcal{L}_n g(n,n) = n(g(n,n)) - 2g(\mathcal{L}_n n, n) = 0$$

and hence

$$\mathcal{L}_n \gamma(n, \,\text{-}\,) = \mathcal{L}_n \gamma(\,\text{-}\,, n) = 0.$$

So we find that *$\mathcal{L}_n \gamma$ is exactly the orthogonal projection of $\mathcal{L}_n g$ to $T\Sigma$*. This quantity has a nice geometric interpretation. As is well-known, the Lie derivative of a metric by a vector field satisfies

$$(\mathcal{L}_X g)_{ab} = \nabla_a X^\flat_b + \nabla_b X^\flat_a.$$

With n the unit normal to a hypersuface, we find that for Y, Z tangent to Σ that

$$(\mathcal{L}_n g)(Y,Z) = g(\nabla_Z n, Y) + g(\nabla_Y n, Z) = -g(\nabla_Z Y + \nabla_Z Y, n) = -II(Y,Z).$$

Here II is the *second fundamental form of the embedding Σ into M*.

10.8 (Initial data) The analysis above suggests that we may expect that to specify the instantaneous state of the system at a fixed time t_0 for Einstein's vacuum equations, we would need to specify a triple (Σ, γ, k) where

- Σ is an n-dimensional smooth manifold giving the spatial topology;
- γ is a Riemannian metric on Σ;
- k is a symmetric two tensor on Σ; this represents the second fundamental form of the embedding of Σ into the full space-time, and can be thought of as the initial "velocity" of the metric.

© Willie Wai-Yeung Wong

DRAFT COPY

This expectation will turn out to be correct, in the sense that correspond to each such initial data set, there exists (and in some sense uniquely) a space-time (M,g) solving the Einstein–vacuum equations into which Σ embeds as a spatial slice with γ and k its first and second fundamental forms. We will return to this later in this lecture.

Constraint equations and under-determinism

10.9 We will first revisit the idea that Einstein's equations are in some sense under-determined, and hence *requires* choosing a gauge or coordinate system before it is solvable. This time, however, we will discuss a manifestation of this under-determinism that is purely geometric in nature. Our starting point will be the Gauss and Codazzi equations for non-degenerate hypersurfaces in a pseudo-Riemannian manifold.

10.10 Proposition (Gauss and Codazzi Equations)
Let (M,g) be a pseudo-Riemannian manifold, and let $\Sigma \hookrightarrow M$ a hypersurface, such that the induced metric γ on Σ is non-degenerate. *Denote by* ${}^{(g)}\mathrm{Riem}$ *and* ${}^{(\gamma)}\mathrm{Riem}$ *their Riemann curvature tensors respectively. Let n denote a unit normal vector field to Σ, and k the second fundamental form of the embedding, so that for X,Y tangent to Σ we have $k(X,Y) = g(n,\nabla_X Y)$.*
Gauss equation *Given X,Y,Z,W vector fields tangent to Σ, we have*

$$g({}^{(g)}\mathrm{Riem}(V,W)X,Y) = \gamma({}^{(\gamma)}\mathrm{Riem}(V,W)X,Y) \\ - g(n,n)\Big[k(V,X)k(W,Y) - k(V,Y)k(W,X)\Big].$$

Codazzi Equation *Given X,Y,Z vector fields tangent to Σ, we have*

$$g(n,n)g({}^{(g)}\mathrm{Riem}(X,Y)Z,n) = -({}^{(\gamma)}\nabla_X k)(Y,Z) + ({}^{(\gamma)}\nabla_Y k)(X,Z). \qquad \blacksquare$$

The proof is identical to that which can be found in any standard textbook in Riemannian geometry.

10.11 (Contracted Gauss equation) Now let $\{e_i\}$ be an orthonormal basis for $T\Sigma$, so we can write $\gamma^{-1} = \sum \gamma(e_i,e_i)e_i \otimes e_i$. Using this to take double contractions on both sides of the Gauss equation we find

$$\sum_{i,j} \gamma(e_i,e_i)\gamma(e_j,e_j)g({}^{(g)}\mathrm{Riem}(e_i,e_j)e_i,e_j) = {}^{(\gamma)}R - g(n,n)\Big[(\operatorname{tr} k)^2 - \operatorname{tr}(k^2)\Big]$$

The left side is almost the full double contraction of the space-time Riemann curvature; we are just missing terms of the form $g(n,n)n \otimes n$. Using that $g({}^{(g)}\mathrm{Riem}(n,n)n,n) = 0$ we can therefore obtain

$${}^{(g)}R - 2g(n,n){}^{(g)}\mathrm{Ric}(n,n) = {}^{(\gamma)}R - g(n,n)[(\operatorname{tr} k)^2 - \operatorname{tr}(k^2)].$$

Slightly reorganizing we finally find

$${}^{(g)}\mathrm{Ric}(n,n) - \frac{1}{2}{}^{(g)}R g(n,n) = -\frac{1}{2}g(n,n){}^{(\gamma)}R + \frac{1}{2}[(\operatorname{tr} k)^2 - \operatorname{tr}(k^2)]. \qquad (10.11.1)$$

The left hand side is the Einstein tensor evaluated at the $n \otimes n$ component.

© Willie Wai-Yeung Wong

10.12 (Contracted Codazzi equation) Arguing similarly, after taking one contraction of the Codazzi equation we find for Y tangent to Σ that

$$^{(g)}\mathrm{Ric}(Y,n) = g(n,n)\Big[-(\mathrm{div}_\gamma k)(Y) + \nabla_Y(\mathrm{tr}\,k)\Big]. \tag{10.12.1}$$

Observe now that the left hand side is the Einstein tensor evaluated at the $Y\otimes n$ component.

10.13 The key observation is this: our original goal for solving Einstein's equation is to solve for the unknown metric g from the Einstein–vacuum equation

$$\mathrm{Ric} - \frac{1}{2}g + \Lambda g = 0.$$

However, if we wish to study the corresponding initial value problem, we see that the components of the equation corresponding to the $n\otimes n$ and $n\otimes Y$ slots do *not* carry any evolutionary information at all, and in fact can be entirely computed from information prescribed as initial data: the submanifold Σ, and its first and second fundamental forms γ and k.

This has two consequences:

1. This shows that the Einstein–vacuum equation must be under-determined, at least as an evolution equation. We expect it to provide a second order partial differential equation on the metric components, and so evolutionary terms should involve the second time derivatives of the metric. The contracted Gauss and Codazzi equations show that a number of terms simply do not involve second time derivatives at all.
2. This also shows that initial data for the Einstein equations cannot be freely prescribed. In order for us to generate a solution, the data must satisfy what are called the "constraint equations". In the vacuum case, this means that we need to satisfy
   Hamiltonian constraint $^{(\gamma)}R + (\mathrm{tr}\,k)^2 - \mathrm{tr}(k^2) + 2\Lambda = 0$;
   Momentum constraint $\mathrm{div}_\gamma k = \nabla(\mathrm{tr}\,k)$.
   In the non-vacuum cases additional terms relating to the stress-energy tensor will also need to be included.

Note that once we have obtained a solution (M,g) to the Einstein equations, if you take any space-like hypersurface Σ in M, by virtue of the Gauss and Codazzi equations the constraint equations must be satisfied on Σ. So the constraint equations only occur as an issue when prescribing initial data; once suitable data is prescribed and a solution has been constructed, then the constraint becomes automatically satisfied at all other constant-time slices. This idea is known as the *propagation of the constraint*.

10.14 (Remarks on the constraint equations) The constraint equations are equations concerning a Riemannian manifold (Σ,γ) augmented with a symmetric covariant 2-tensor k. The study of solutions to the constraint equations therefore rely more on techniques of *elliptic* partial differential equations and *Riemannian geometric analysis*. We will not be discussing these problems in our lectures, but I feel it to be necessary to mention some of the related research directions. When faced with an equation such as the constraint equations, there are two obvious types of questions one can ask. The first is: "what are some properties common to all solutions?" The second is: "how can we parametrize the set of solutions?"

A recent book exploring these questions is Lee, Geometric Relativity.

DRAFT COPY

© Willie Wai-Yeung Wong

10.15 For simplicity, we will set $\Lambda = 0$, and assume that the second fundamental form $k \equiv 0$ (in reality, the former is a bit hard to relax, the latter a bit easier). To include a bit more generality, we allow a non-trivial matter model so that there is non-trivial contribution from the stress-energy tensor. The assumption that $k \equiv 0$ is often called the "time symmetric" setting, as k represents the time derivative of the spatial metric γ, so with $k \equiv 0$ we have that the universe is "instantaneously at rest", and the future and past should therefore look identical. In this case the momentum constraint trivializes.

If we assume that the stress-energy satisfies the dominant energy condition, then the Hamiltonian constraint becomes the statement ${}^{(\gamma)}R = J \geq 0$, where J is the $n \otimes n$ component of the stress-energy tensor. And thus from the constraint equations, we are naturally led to the study of **Riemannian manfiolds with lower bounds on scalar curvature**. Especially in the non-compact setting, one is interested in understanding what the scalar curvature lower bound says about the asymptotics of the metric γ. This led to the development of the *Positive Mass Theorem* and the study of the *Riemannian Penrose inequality*.

10.16 Due to the constraint equations, one cannot prescribe arbitrarily a Riemannian metric g and a symmetric 2-tensor field k and ask that they be the initial data for a solution to the Einstein–vacuum equations. A natural question to ask is "what is the set of all pairs (γ, k) that satisfy the two constraint equations?" Even better would be **finding ways to parametrize the set of all solutions to the constraint equations**; the ideal goal would be to be able to reduce the degrees of freedom to a set of "free variables", preferably geometric in origin, that one can freely prescribe. This should be combined with an algorithm to generate a corresponding (γ, k) that solves the constraints.

This problem was first attacked by André Lichnerowicz using what is now called the "conformal method". Focusing again on the case $\Lambda = 0$, but now looking only at the vacuum equations and admitting non-trivial second fundamental form, our goal is to solve both the Hamiltonian and momentum constraints. In the discussion below we will use $d = \dim(\Sigma)$.

The first step in Lichnerowicz's method is to observe that we can split k into its trace part and trace-free part algebraically

$$k = \frac{\kappa}{d}\gamma + \hat{k}$$

where $\operatorname{tr}_\gamma \hat{k} = 0$. The constraint equations becomes

$$\begin{gathered} {}^{(\gamma)}R + \frac{d-1}{d}\kappa^2 - \operatorname{tr}(\hat{k}^2) = 0 \\ \frac{d-1}{d}\nabla\kappa = \operatorname{div}\hat{k}. \end{gathered}$$

We shall now describe the simplest case of the conformal method, where κ is a constant. This is known as the "constant mean curvature case". (The conformal method can also be extended to settings where the mean curvature is not constant, see the works of David Maxwell from the 2010s and 2020s for modern developments and generalizations.)

Under the assumption that κ is a constant, the momentum constraint reduces to the equation $\operatorname{div}\hat{k} = 0$. The key observation is the following fact, which can be proven with a computation. (Note that the assumption that $\hat{k}$ is traceless is crucial.)

DRAFT COPY

DRAFT COPY

Let γ be a Riemannian metric and $\hat{k}$ a symmetric 2-tensor that is traceless and divergence free (relative to the metric γ). For any strictly positive smooth function ϕ, if we set $\tilde{\gamma} = \phi^{\frac{4}{d-2}}\gamma$ and $\tilde{k} = \phi^{-2}\hat{k}$, then $\tilde{k}$ is traceless and divergence free (relative to the metric $\tilde{\gamma}$).

The basic idea of the conformal method then proceeds to solve the constraint equations by reducing it to a *single scalar PDE*. Start with a manifold Σ and a choice of the constant κ. Choose an arbitrary pair $(\mathring{\gamma}, \mathring{k})$ where $\mathring{\gamma}$ is a Riemannian metric and $\mathring{k}$ is traceless and divergence free relative to $\mathring{\gamma}$. We then seek a factor ϕ such that $\gamma = \phi^{\frac{4}{d-2}}\mathring{\gamma}$ and $\hat{k} = \phi^{-2}\mathring{k}$ solves the Hamiltonian constraint (the momentum constraint will be trivially satisfied by the previous fact.

Using the well-known formula for the scalar curvature of the conformal metric, we find the equation to solve becomes

$$\phi^{-\frac{4}{d-2}}\left[\mathring{R} - \frac{4(d-1)}{d-2}\phi^{-1}\mathring{\Delta}\phi\right] + \frac{d-1}{d}\kappa^2 - \phi^{-4-\frac{8}{d-2}}\mathring{\mathrm{tr}}(\mathring{k}^2) = 0.$$

Which is a scalar elliptic PDE with mixed power nonlinearities. This suggests that those initial data which have constant mean curvature can be "parametrized" by the free data

- A choice of a constant mean curvature κ;
- A choice of a conformal class $[\mathring{\gamma}]$;
- A choice of a "conformal class" $[\mathring{k}]$ where for some pair of representatives $\mathring{\gamma}$ and $\mathring{k}$ we have that the latter is traceless and divergence free relative to the former;

together with a derived quantity ϕ, the conformal factor(s) that will allow the Hamiltonian constraint to hold.

Local existence

10.17 We finally return to the existence of solutions to the Einstein–vacuum equations. We shall take as a "black box" the following result from nonlinear wave equations.

Precisely this theorem is hard to find in the literature. Among "experts" such a result is widely regarded as folklore. Its proof strongly hinges on the so-called "finite speed of propagation" property for wave type equations.

10.18 Theorem (Local existence and uniqueness for nonlinear wave equations) *Let M be a smooth manifold equipped with a linear connection $\mathring{\nabla}$. Let $\mathcal{V}$ be a vector bundle over M, and abuse notation to denote also by $\mathring{\nabla}$ a linear connection for $\mathcal{V}$. Consider the partial differential equation*

$$\mathfrak{g}(\phi)^{ab}\mathring{\nabla}_a\mathring{\nabla}_b\phi = \mathfrak{S}(\phi, \mathring{\nabla}\phi) \tag{10.18.1}$$

where

- $\mathfrak{g}$ *is a smooth (possibly nonlinear) bundle map from $\mathcal{V}$ to $T^{2,0}M$, whose values are symmetric two-tensors;*
- $\mathfrak{S}$ *is a smooth (possibly nonlinear) bundle map from $\mathcal{V} \oplus (T^*M \otimes \mathcal{V})$ to $\mathcal{V}$.*

For each triple (Σ, ϕ_0, ϕ_1) where Σ is an embedded submanifold of M (not necessarily closed), ϕ_0 a smooth section of $\mathcal{V}$ over Σ, and ϕ_1 a smooth section of $T^*M \otimes V$ over Σ, if it holds that

- (spatial compatibility) *for vector field X tangent to Σ, we have $\mathring{\nabla}_X\phi_0 = \phi_1(X)$;*
- (hyperbolicity) *along Σ the $T^{2,0}M$ tensor field $\mathfrak{g}(\phi_0)$ is Lorentzian;*
- (spacelike data) *if τ is a local defining function of Σ, then $\mathfrak{g}(\phi_0)(d\tau, d\tau) < 0$;*

© Willie Wai-Yeung Wong

then there exists an open set $U \subseteq M$ with Σ an embedded submanifold of U, and an unique smooth section ϕ of $\mathcal{V}$ over U, such that ϕ solves (10.18.1)*, and $\phi|_\Sigma = \phi_0$, and $\mathring{\nabla}\phi|_\Sigma = \phi_1$.* ■

10.19 The statement of "uniqueness" in the theorem above is not its most general form; the correct statement is a bit more subtle and requires developing some additional ideas from Lorentzian geometry first. We defer its discussion to the next lecture.

10.20 For the purposes of these lecture notes, I am not assuming familiarity with Sobolev spaces, and so stated the theorem only for smooth objects. The local existence and uniqueness theorem is usually stated and proved in the literature for initial data in L^2-based Sobolev spaces with sufficiently many derivatives (usually one assumes $\lfloor \frac{\dim \Sigma}{2} \rfloor + 3$ to ensure classical solutions); see e.g. Hughes, Kato, and Marsden, "Well-posed quasi-linear second-order hyperbolic systems with applications to nonlinear elastodynamics and general relativity". These equations enjoy "persistence of higher regularity", which roughly says that any additional higher regularity on the initial data assumed beyond that which is needed to prove the basic local existence theorem will be reflected as additional higher regularity available for the solution.

DRAFT COPY

DRAFT COPY

Topic 11 (2024/04/15)

Local Existence and Uniqueness

Local existence continued

11.1 (Quasilinearity) The equation (10.18.1) is said to be *quasilinear*.

The word *quasilinear* refers to the fact that the top order derivatives terms appear linearly in the equation. In our equation, the highest order derivative to appear is second order, and we see that the second derivatives $\mathring{\nabla}^2\phi$ appears linearly in the expression, as it is multiplied by a coefficient that depends (possibly nonlinearly) only on ϕ. In general, quasilinear equations can have the coefficients $\mathfrak{g}$ depend on the first derivative $\mathring{\nabla}\phi$ too; we can however assume the system is of the form we wrote down without loss of generality by *prolongation*. This is easiest to see through an explicit computation.

11.2 Example

Suppose we have the following differential equation on $\mathbb{R}^n$

$$g(\phi, \partial\phi)^{ij}\partial^2_{ij}\phi = 0,$$

where ϕ takes values in $\mathbb{R}$. By taking a derivative of the equation, we see that the $\mathbb{R}^n$-valued function $\nabla\phi$ satisfies the differential equation

$$g(\phi, \partial\phi)^{ij}\partial^2_{ij}\nabla\phi + (\partial_\phi g)(\phi, \partial\phi)\cdot\nabla\phi\cdot\partial^2\phi + (\partial_{\partial\phi}g)(\phi, \partial\phi)\cdot\partial\nabla\phi\cdot\partial^2\phi = 0.$$

Note that in the expression above, every second derivative may be regarded as a first derivative of $\nabla\phi$. And so if we enlarge our set of unknowns to be the pair $(\phi, \nabla\phi) = \Phi$, we see that the equations above can be expressed in the form

$$g(\Phi)^{ij}\partial^2_{ij}\Phi = S(\Phi, \partial\Phi)$$

where the coefficients in front of the second derivative no longer depend on the first derivative of the unknown. ◊

11.3 (Quasi-diagonality) Another feature of (10.18.1) is its *quasi-diagonality*. To understand this concept, it helps to return first to a general second order (linear) differential operator acting on sections of $\mathcal{V}$. In general, such a differential operator will take the form

$$P\phi = A^{ab}\mathring{\nabla}_a\mathring{\nabla}_b\phi + B^a\mathring{\nabla}_a\phi + C\phi.$$

Where the coefficient tensors A, B, C *take value in the set of continuous linear maps from* $\mathcal{V}$ *to itself*. (If this feels confusing, you should go ahead to the next example first before reading the rest of this paragraph.) Relative to a basis of $\mathcal{V}$, the i-th component of $P\phi$ may depend on the values of the derivatives of the other components.

The word quasi-diagonal refers to the fact that in our equation (10.18.1), the operator on the left side of the equation acts diagonally on ϕ, so that the second derivative contributions to the i-th component of the equation only involve the i-th component of the field. This is realized by the fact that we assumed $\mathfrak{g}$ is scalar-valued (a section of $T^{2,0}M$ instead of a section of $T^{2,0}M \otimes \mathcal{V} \otimes \mathcal{V}^*$ as would be the case in general).

Quasi-diagonal equations are "almost scalar", and most techniques developed for scalar wave equations can be carried out unchanged for their analysis. Non-quasi-diagonal equations are more complicated: even defining what it means for such a system to be hyperbolic is tricky, and proving local existence statements doubly so.

11.4 Example

Consider the operator $\nabla \times (\nabla \times v)$ acting on vector fields v in $\mathbb{R}^3$. It is well-known that this has the explicit coordinate form

$$[\nabla \times (\nabla \times v)]_i = \partial_i(\partial_1 v_1 + \partial_2 v_2 + \partial_3 v_3) - (\partial^2_{11} + \partial^2_{22} + \partial^2_{33})v_i.$$

We can write it in the following alternative way:

$$\nabla \times (\nabla \times v) = A^{ij}\partial^2_{ij} v$$

where for each (i, j) the object A^{ij} is a 3×3 matrix. More precisely we have

$$A^{11} = \begin{pmatrix} 0 & & \\ & -1 & \\ & & -1 \end{pmatrix}, \quad A^{12} = \begin{pmatrix} 0 & 1 & \\ & 0 & \\ & & 0 \end{pmatrix}, \quad A^{13} = \begin{pmatrix} 0 & 0 & 1 \\ & 0 & \\ & & 0 \end{pmatrix}$$

and so on. ◊

11.5 Corollary

If in (10.18.1), *the inhomogeneous term is such that* $\mathfrak{S}(0, 0) = 0$, *then when* $\phi_0 = 0$ *and* $\phi_1 = 0$ *the only solution is the* 0 *solution.* ■

Proof. Under the hypothesis of the corollary, one sees immediately that the 0 section is a solution. Uniqueness follows from Theorem 10.18. □

DRAFT COPY

© Willie Wai-Yeung Wong

DRAFT COPY

11.6 Theorem
Let Σ be a smooth manifold of dimension $d \geq 2$, equipped with a Riemannian metric γ and a symmetric two tensor k such that the constraint equations hold. Then there exists a $(d+1)$-dimensional manifold M, a Lorentzian metric g on M, and an embedding $\Sigma \hookrightarrow M$ such that g solves the Einstein–vacuum equations and γ and k are the first and second fundamental forms of the embedding. ■

Proof. First let $\tilde{M}$ be the smooth manifold $\mathbb{R} \times \Sigma$, we can equip it with a Riemannian metric $\mathring{h}$. This can be chosen arbitrarily, but for concreteness consider $\mathring{h} = ds^2 + \gamma$ the product metric, where γ is the given data. We shall fix the embedding so that Σ embeds into $\tilde{M}$ as the slice $\{0\} \times \Sigma$. It therefore suffices to find an open set $M \subseteq \tilde{M}$ containing $\{0\} \times \Sigma$ and a Lorentzian metric g on M such that the Einstein–vacuum equation is satisfied on M.

By Lemma 9.16, it suffices that we solve the system $\widetilde{\mathrm{Ric}}[g, \mathring{h}] = \frac{2}{n-2}\Lambda g$ and $\mathfrak{X} = 0$ on M. We will do so by solving the system of quasilinear quasi-diagonal wave equations (see ¶ 9.17) for the pair $(g, \mathfrak{X})$ (so in the language of Theorem 10.18, we have $\mathcal{V} = \tilde{T}^{0,2}M \oplus TM$, where $\tilde{T}^{0,2}M$ is the bundle of symmetric covariant 2 tensors):

$$(g^{-1})^{bd} \mathring{\nabla}_b \mathring{\nabla}_d g = 2\mathring{\mathrm{Ric}} + \frac{2}{n-2}\Lambda g + (g^{-1} \cdot \mathring{\mathrm{Riem}} \cdot g) + \mathcal{P}(g^{-1}, \mathring{\nabla} g)$$
$$(g^{-1})^{bd} \mathring{\nabla}_b \mathring{\nabla}_d \mathfrak{X} = \mathcal{Q}(\mathring{\mathrm{Ric}}, \Gamma, g^{-1}, \mathfrak{X}) \cdot (\mathfrak{X}, \mathring{\nabla}\mathfrak{X}) + \mathrm{div}\tilde{G}[g, \mathring{h}].$$

Note that once we are able to provide appropriate initial data for g and $\mathfrak{X}$, then Theorem 10.18 guarantees the existence of a solution. Furthermore, the first equation is equivalent to $\tilde{G}[g, \mathring{h}] = \Lambda g$, so if g were to solve the first equation, the second equation becomes equivalent to

$$(g^{-1})^{bd} \mathring{\nabla}_b \mathring{\nabla}_d \mathfrak{X} = \mathcal{Q}(\mathring{\mathrm{Ric}}, \Gamma, g^{-1}, \mathfrak{X}) \cdot (\mathfrak{X}, \mathring{\nabla}\mathfrak{X}) + \Lambda \underbrace{\mathrm{div}(g)}_{=0}$$

where the final term vanishes since crucially the differentiation there is made with respect to the Levi-Civita connection of the g metric. We therefore see that the equation for $\mathfrak{X}$ satisfies the hypothesis of Corollary 11.5. If we can ensure that the initial data for $\mathfrak{X}$ is trivial, then we may appeal to the Corollary to conclude that $\mathfrak{X} \equiv 0$ throughout M.

So it remains to show that we can start from γ and k and generate appropriate initial data for g and $\mathfrak{X}$ for the wave equations.

Our goal is to specify $g_0, g_1, \mathfrak{X}_0, \mathfrak{X}_1$ so that the latter two vanishes and the first two are compatible with each other, and all are compatible with the given data in terms of γ and k. Let's start by recording what we already know, in terms of constraints that must be obeyed by the initial data. Below we will again use n to denote the unit (with respect to g) time-like normal to Σ.

- The restriction of g_0 to $T\Sigma$ is equal to γ; this means that there exists a scalar ν and a one-form ξ satisfying $\xi(\partial_s) = 0$ such that

$$g_0 = \nu ds \otimes ds + ds \otimes \xi + \xi \otimes ds + \gamma.$$

- That k is the second fundamental form of the embedding of Σ, so that for Σ-tangent vector fields X, Y we have

$$k(X,Y) = g(\nabla_X Y, n) = g(\mathring{\nabla}_X Y, n) + g(\Gamma(X,Y), n).$$

© Willie Wai-Yeung Wong

Since Σ is totally geodesic in the metric $\mathring{h}$, we have $\mathring{\nabla}_X Y$ must be tangent to Σ, and hence the first term drops out. We conclude

$$k(X,Y) = g(\Gamma(X,Y), n).$$

- That Σ-tangent derivatives of g_0 agrees with g_1.
- $\mathfrak{X}_0 = (g^{-1})^{bd}\Gamma_{bd}$ so there is a non-trivial relationship between $\mathfrak{X}_0$, g_0, and g_1.
- The Σ-tangent derivative of $\mathfrak{X}_0$ agrees with $\mathfrak{X}_1$.
- As $\mathfrak{X}$ is at the level of the relative Christoffel symbol, the quantities $\mathfrak{X}_1$, being at its derivative, is a "curvature level" object. And so we pick up the additional constraint that $\mathfrak{X}_1$ must be such that the Ricci curvature of g satisfies Einstein's vacuum equations.

It turns out that the final condition is not significant. Our process of solving the reduced Einstein's equation using $\widetilde{\mathrm{Ric}}[g,\mathring{h}]$ will automatically guarantee that the reduced Einstein tensor $\tilde{G}[g,\mathring{h}] = \Lambda g$. Hence (9.17.1) means that compatibility requires only that $\frac{1}{2}(g_{fc}\mathring{\nabla}_a\mathfrak{X}^f + g_{fa}\mathring{\nabla}_c\mathfrak{X}^f) - \frac{1}{2}g_{ac}\mathring{\nabla}_f\mathfrak{X}^f$ which is certainly compatible with $\mathfrak{X}_1 \equiv 0$. (Note that here the fact that γ and k satisfies the constraint equations play a role, since those components of $\tilde{G}$ are not dynamical, and depends only on the choice of g_0 and g_1.) The second to last condition is also satisfied if we can show that $\mathfrak{X}_0 \equiv 0$ is viable. So we are down to the first four.

It turns out we can simplify our discussion somewhat by posing the ansatz $\nu \equiv -1$ and $\xi \equiv 0$ on Σ. This implies that along Σ we have $n = \partial_s$. That this is allowed is fundamentally part of our choice of the initial parametrization of $\tilde{M}$ by $\mathbb{R}\times\Sigma$; while we have already prescribed how Σ should embed, we have not specified how the transversal directions over Σ are related between $\tilde{M}$ and $\mathbb{R}\times\Sigma$. We make use of this freedom now. With this choice we see that for any X tangent to Σ, we have $\mathring{\nabla}g_0 = 0$. Hence

$$k(X,Y) = g(\Gamma(X,Y), n) = -\frac{1}{2}(\mathring{\nabla}_{\partial_s} g)(X,Y).$$

And hence we should set g_1 so that $g_1(X; -, -) = 0$ for all X tangent to Σ, and $g_1(n; X, Y) = -2k(X,Y)$. It remains to specify $g_1(n; n, X)$ and $g_1(n; n, n)$.

For these last conditions, we use $\mathfrak{X}_0$. For it to vanish, we need $(g^{-1})^{bd}(\mathring{\nabla}_b g_{cd} - \frac{1}{2}\mathring{\nabla}_c g_{bd}) = 0$. Taking an orthonormal frame with $e_0 = n = \partial_s$ and $e_1, \dots, e_d$ spanning $T\Sigma$, we see that this implies

$$-\frac{1}{2}(\mathring{\nabla}_n g)(n,n) - \frac{1}{2}\sum_{i=1}^{d}(\mathring{\nabla}_n g)(e_i, e_i) = 0$$

when evaluated against n^c; when evaluated against a Σ-tangent direction we find

$$-(\mathring{\nabla}_n g)(e_j, n) = 0.$$

And so we see that choosing $g_1(n; n, n) = \mathrm{tr}_\gamma k$ and $g_1(n; n, X) = 0$ for X tangent to Σ provides us now with a self-consistent choice of initial values. (Note: this choice is not unique; especially in view of our initial choice that $g_0 = -ds\otimes ds + \gamma$.) □

DRAFT COPY

11.7 The theorem that we just proved is fundamentally an *existence* result. It does not prove uniqueness, in view of the diffeomorphism invariance of Einstein's equations. To obtain something that resembles a uniqueness statement, we need two ingredients.

1. We can fix the diffeomorphism invariance issue by establishing a (partial) diffeomorphism between any two solutions.
2. The uniqueness statement in Theorem 10.18 turns out to be too weak; a stronger, more geometric statement is needed.

The first ingredient turns out to just require some cleverness: using the fact that the wave map gauge is related to solving the wave map equation, which is itself a nonlinear wave equation, when given two solutions we can solve the wave map equation using the *same* target Riemannian manifold and use it to define a diffeomorphism with which to compare the two solutions. We will describe in more detail this implementation later on when we prove the local uniqueness theorem.

The second ingredient is fundamental to the geometric understanding of nonlinear wave equations, and has significant physical implications. We will therefore focus on it first.

Local Uniqueness for Wave Equations

11.8 It is helpful to first get into the right frame of mind about the question of uniqueness. Note that we assume in Theorem 10.18 only that the initial data is prescribed on a hypersurface Σ, without requiring any "completeness" of this hypersurface. In PDE language, the analogue would be akin to specifying a partial differential equation on $\mathbb{R}\times\mathbb{R}^d$, prescribing the initial data on $\Sigma = \{0\}\times\Omega$ where Ω is an open subset of $\mathbb{R}^d$, and asking for a solution to exist on an open set U of the space-time that contains Σ. Notably, Ω may have compact closure, and we have made no attempt to prescribe boundary data on U.

In general, there are two obstacles to any sort of uniqueness. The first comes from the "type" of the PDE being studied. The second comes from the danger of trying to make U "too large".

11.9 Example

Consider the 1D heat equation $\partial_t u = \partial^2_{xx} u$ for $u : [0,\infty)\times\mathbb{R}\to\mathbb{R}$. This equation has an explicit solution by convolution. Suppose u_0 is a continuous function with compact support, then the function

$$u(t,x) = \frac{1}{\sqrt{4\pi t}}\int_{\mathbb{R}} \exp\Big(-\frac{(x-y)^2}{4t}\Big) u_0(y)\, dy$$

defined on $(0,\infty)\times\mathbb{R}$ extends continuously to $\{0\}\times\mathbb{R}$ and solves the heat equation, with $u(0,x) = u_0(x)$. Note that a consequence of the strict positivity of the Gaussian kernel means that if $u_0\geq 0$ and is strictly positive somewhere, then the solution $u(t,x) > 0$ for all $t > 0$.

On the other hand, the zero solution is another solution to the heat equation.

Both of these solutions have the same initial data on along the set $\Sigma = \{0\}\times(-1,0)$, yet for no open neighborhood of Σ are the two solutions identical. This provides a strong lack of uniqueness. $\diamond$

DRAFT COPY

For those aware of PDE terminology: one of the fundamental differences between the heat equation and the wave equation is that the latter enjoy "finite speed of propagation" (a concept to which we will return) while the former exhibit "infinite speed of propagation". That we can have some semblance of a uniqueness statement without boundary conditions in Theorem 10.18 is a hallmark of the "finite speed of propagation" property.

11.10 The finite speed of propagation property is a bit of a double-edged sword. On the one hand, it says that perturbations that are initially far away will take some minimum of time to reach us, and therefore the local (in space-time) behavior of the solution is entirely determined by the local (in space) behavior of the initial data. This gives a form of uniqueness. On the other hand, finite speed of propagation also indicates that local (in space) behavior of the initial data cannot influence the behavior of the solution that is suitably "far away". This gives a form of non-uniqueness.

11.11 Example
Consider the wave equation $\partial^2_{tt} u = \partial^2_{xx} u$ on $\mathbb{R} \times \mathbb{R}$. For initial data u_0, u_1 prescribed along $\{0\} \times \mathbb{R}$, we have the d'Alembert's representation formula

Recall that u_1 in our notation should be an element of $T^*\mathbb{R}^2$ over the $\Sigma = \{0\} \times \mathbb{R}$.

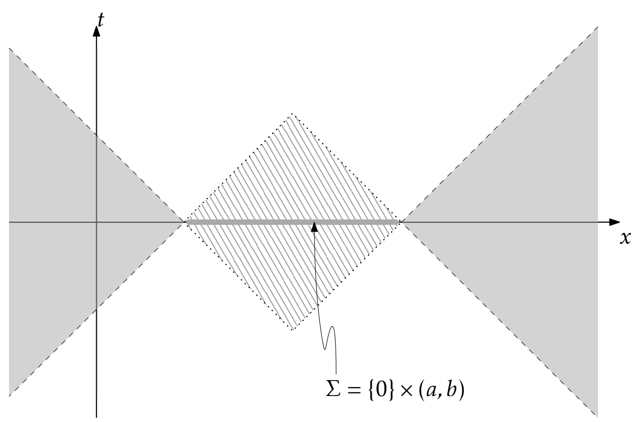
$$u(t,x) = \frac{1}{2}\Big(u_0(x-t) + u_0(x+t)\Big) + \frac{1}{2}\int_{x-t}^{x+t} u_1(\partial_t).$$

As a result, if u_0, u_1 both vanish on an interval $\{0\} \times (a,b)$ (the segment labeled Σ in the illustration below, shown as the thickened gray line), the corresponding solution vanishes on the diamond-shaped region (hatched in the illustration below)

$$\Big\{(t,x) \in \mathbb{R}^2 \,\Big|\, (a \le x - t \le b) \wedge (a \le x + t \le b)\Big\}.$$

On the other hand, knowing that the initial data vanish on $\{0\} \times [a,b]$ provides us with absolutely no knowledge on what the solution should be like on the region (shaded in the illustration below)

$$\Big\{(t,x) \in \mathbb{R}^2 \,\Big|\, (x + |t| < a) \vee (x - |t| > b)\Big\}.$$



◇

DRAFT COPY

11.12 Our next question is: without solving the wave equation explicitly, how can we determined what would be the region on which uniqueness hold? The fundamental ideas are two fold. First, we expect that when solving a wave equation, perturbations travel in trajectories that are either light-like or time-like with respect to the Lorentzian metric that defines the wave equation. Second, we expect that we can quantitatively measure the first idea through looking at the "energy" carried by the solution.

11.13 Definition (Lens-shaped domains)
Let M be a $(d+1)$-dimensional manifold. By a lens-shaped domain *we refer to*

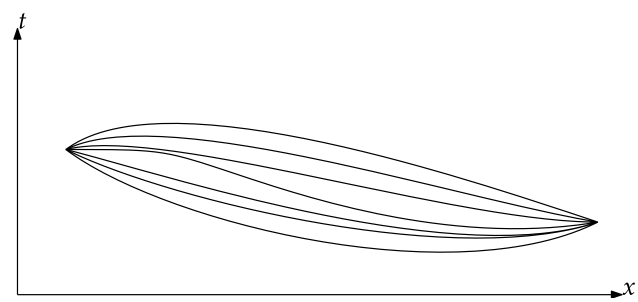
- *a d-dimensional compact manifold with boundary S, with boundary ∂S; and*
- *a smooth function $\Xi : (-1,1) \times S \to M$*

such that

1. *$\Xi|_{(-1,1)\times(S\setminus\partial S)}$ is a diffeomorphism to its image,*
2. *for every $t \in (-1,1)$ and $p \in \partial S$, the mapping $\Xi(t,p) = \Xi(0,p)$.*

If M is equipped with a Lorentzian metric g, we say that the lens-shaped domain is space-like *if $\Xi(t, S\setminus\partial S)$ is a space-like hypersurface for every $t \in (-1,1)$.* ◊

11.14 For a caricature of a lens-shaped domain when $d = 1$, see the illustration below.



Lens-shaped domains are fundamental to understanding the uniqueness properties of solutions of nonlinear wave equations.

11.15 Theorem (Local uniqueness for nonlinear wave equations)
Let $\mathcal{V} \to M$ be a vector bundle and $\mathring{\nabla}$ denote a linear connection. Suppose now ϕ is a section of $\mathcal{V}$ that satisfies the wave equation

$$\mathfrak{g}(\phi)^{ab}\mathring{\nabla}_a\mathring{\nabla}_b\phi = \mathfrak{S}(\phi, \mathring{\nabla}\phi)$$

where $\mathfrak{g}$ and $\mathfrak{S}$ are as in Theorem 10.18. Assume further that $\mathfrak{g}(\phi)$ has Lorentzian signature at every point in M, and that $\mathfrak{S}(0,0) = 0$. Let S and Ξ define a Lens-shaped domain in M that is space-like with respect to $\mathfrak{g}(\phi)$. If ϕ and $\mathring{\nabla}\phi$ vanish identically along $\Xi(0,S)$, then ϕ vanishes identically on $\Xi(t,S)$ for all $t \in (-1,1)$. ■

© Willie Wai-Yeung Wong

DRAFT COPY

11.16 First we remark that since we are studying the property of a fixed solution, we can *freeze coefficients* and consider the Lorentzian metric g whose inverse metric $g^{-1} = \mathfrak{g}(\phi)$ for the given solution ϕ. (Note that this is only possible since our system of equations is quasi-diagonal; for non quasi-diagonal systems the analogous theorem is much harder to prove.) Denote now by ∇ the Levi-Civita connection of the metric g, extended to agree with $\mathring{\nabla}$ on sections of $\mathcal{V}$. If Γ is the relative Christoffel symbol, we can write the wave equation as

$$(g^{-1})^{ab}\nabla_a\nabla_b\phi - (g^{-1})^{ab}\Gamma^c_{ab}\nabla_c\phi = \mathfrak{S}(\phi, \nabla\phi).$$

The second term depends on $\nabla\phi$ and not second derivatives, so we have equivalently that ϕ solves the wave equation

$$(g^{-1})^{ab}\nabla_a\nabla_b\phi = \tilde{\mathfrak{S}}(\phi, \nabla\phi). \tag{11.16.1}$$

Here $\tilde{\mathfrak{S}}$ still satisfies the condition that $\tilde{\mathfrak{S}}(0,0) = 0$. Note that this transformation is not necessary. We do so merely for convenience since it simplifies some of the computations that we need to carry out below.

11.17 The proof will be based on the energy method. For convenience of exposition we will only prove the case where $\mathcal{V}$ is the trivial $\mathbb{R}$ bundle over M, so that sections are merely scalar functions. Throughout we will briefly describe some ideas needed to extend the proof to the general case. The energy method is based on studying the *energy-momentum tensor* $\mathfrak{Q}$ of the wave equation, which is defined by

$$\mathfrak{Q}(\phi, \nabla\phi)^a_b = (g^{-1})^{ac}\nabla_c\phi\nabla_b\phi - \frac{1}{2}\delta^a_b(g^{-1})^{cd}\nabla_c\phi\nabla_d\phi - \delta^a_b\phi^2 \tag{11.17.1}$$

Observe that given any one form ω on M and vector field X on M, we have

$$\mathfrak{Q}(\phi, \nabla\phi)(\omega, X) = \Big((g^{-1})^{ac}\omega_c X^b - \frac{1}{2}\omega(X)(g^{-1})^{ab}\Big)\nabla_a\phi\nabla_b\phi - \omega(X)\phi^2. \tag{11.17.2}$$

A bit of linear algebra shows that

11.18 Lemma
If g is a Lorentzian product on a vector space, ω a time-like covector, X a time-like vector, with $\omega(X) < 0$, then $g^{ac}\omega_c X^b - \frac{1}{2}\omega(X)g^{ab}$ defines a positive definite *quadratic form.* ■

11.19 In the general case, the expression (11.17.1) defines an element of $T^{1,1}M \otimes \mathcal{V} \otimes \mathcal{V}$. For energy estimates we wish to *scalarize* the expression so we can do comparisons (of real numbers). To do so we should choose a positive definite inner product on $\mathcal{V}$ represented as a section of $\mathcal{V}^* \otimes \mathcal{V}^*$, and incorporate its action in the definition of the energy-momentum tensor.

© Willie Wai-Yeung Wong

DRAFT COPY

Topic 12 (2024/04/29)

More on uniqueness

Local Uniqueness for Wave Equations, Continued

12.1 Next, we compute the divergence of $\mathfrak{Q}(\phi, \nabla\phi)^a_b$, which gives

$$\nabla_a \mathfrak{Q}(\phi, \nabla\phi)^a_b = (g^{-1})^{ac}\nabla_a\nabla_c\phi\nabla_b\phi + (g^{-1})^{ac}\nabla_c\phi\nabla_a\nabla_b\phi - (g^{-1})^{ac}\nabla_c\phi\nabla_b\nabla_a\phi - 2\phi\nabla_b\phi.$$

The first term we can replace using the wave equation satisfied by ϕ. The second and third terms cancel as metric Hessians of scalar fields are symmetric. (In the general case there are two differences. First there will also be terms when the derivative hits the inner product of $\mathcal{V}$, since we have not assumed that the connection on $\mathcal{V}$ is compatible with the inner product. In the specific case where we are applying these arguments to the reduced Einstein–vacuum equations in the wave-map gauge, we have $\mathcal{V} = \tilde{T}^{0,2}M \oplus TM$ (see the proof of Theorem 11.6); since we have chosen $\mathring{\nabla}$ to be the Levi-Civita connection for a background Riemannian metric, this Riemannian metric extends to an inner product on $\mathcal{V}$ and the action of $\mathring{\nabla}$ is compatible with this inner product. The second difference is that when ϕ is vector-valued, the Hessian of ϕ is no longer symmetric, and so we replace $\nabla_a\nabla_b\phi - \nabla_b\nabla_a\phi$ with a term that captures the curvature of the $\mathcal{V}$ connection. Note that such a term will depend linearly on ϕ.)

We therefore find

$$\nabla_a \mathfrak{Q}(\phi, \nabla\phi)^a_b = \tilde{\mathfrak{S}}(\phi, \nabla\phi)\nabla_b\phi - 2\phi\nabla_b\phi. \tag{12.1.1}$$

(In the general case there will be a few additional terms on the right hand side, but the expressions will depend only on ϕ and $\nabla\phi$, and not higher derivatives. Furthermore, the additional terms will all be quadratic in $(\phi, \nabla\phi)$.)

12.2 Now choose an arbitrary time-like (relative to the metric g) vector field T on M. Denote by J

$$\mathrm{J}^a = \mathfrak{Q}(\phi, \nabla\phi)^a_b T^b$$

the *energy current as observed by T*. Then from (12.1.1) we find

$$\operatorname{div}\mathrm{J} = \tilde{\mathfrak{S}}(\phi, \nabla\phi)T(\phi) - 2\phi T(\phi) + \mathfrak{Q}(\phi, \nabla\phi)^a_b \nabla_a T^b. \tag{12.2.1}$$

DRAFT COPY

12.3 Next fix $[t_0, t_1] \subsetneq (-1, 1)$. Let us consider the region $\Xi([t_0, t_1] \times S)$, the interior of which we can parametrize using $(t_0, t_1) \times (S \setminus \partial S)$. Fix a volume form ω_S on $S \setminus \partial S$, then the Lorentzian volume form can be decomposed as

$$\mathrm{dvol}_g = \sqrt{|g|}\, dt \wedge \omega_S.$$

The divergence theorem applied to $\mathfrak{J}$ yields

$$\int_{\{t_1\}\times S} dt(\mathfrak{J})\, \sqrt{|g|}\omega_S - \int_{\{t_0\}\times S} dt(\mathfrak{J})\, \sqrt{|g|}\omega_S \\ = \int_{t_0}^{t_1} \int_S \Big(\tilde{\mathfrak{S}}(\phi, \nabla\phi) T(\phi) - 2\phi T(\phi) + \mathfrak{Q}(\phi, \nabla\phi)^a_b \nabla_a T^b\Big) \sqrt{|g|} dt \wedge \omega_S. \tag{12.3.1}$$

The key point now is that, after possibly replacing T by $-T$ we can assume $dt(T) < 0$ and hence Lemma 11.18 can be used to evaluate $dt(\mathfrak{J}) = \mathfrak{Q}(\phi, \nabla\phi)(dt, T)$. The positive definiteness guaranteed by the Lemma implies that the terms appearing on the right side of (12.3.1) can, after applying Cauchy-Schwartz, be controlled uniformly. More precisely, using now that $\Xi([t_0, t_1] \times S)$ is compact there exists some constant $C > 0$ (this C may depend on ϕ and $\nabla\phi$; but they are uniformly bounded on $\Xi([t_0, t_1] \times S)$) such that

$$\left| \tilde{\mathfrak{S}}(\phi, \nabla\phi) T(\phi) - 2\phi T(\phi) + \mathfrak{Q}(\phi, \nabla\phi)^a_b \nabla_a T^b \right| \leq C dt(\mathfrak{J})$$

pointwise; here we used the fact that $\tilde{\mathfrak{S}}$ is a smooth mapping and that $\tilde{\mathfrak{S}}(0,0) = 0$. And hence we have the inequality

$$\left| \int_{\{t_1\}\times S} dt(\mathfrak{J})\, \sqrt{|g|}\omega_S - \int_{\{t_0\}\times S} dt(\mathfrak{J})\, \sqrt{|g|}\omega_S \right| \leq \int_{t_0}^{t_1} \int_S C dt(\mathfrak{J})\, \sqrt{|g|}\omega_S\, dt.$$

To make the next steps more explicit, define

$$\mathcal{E}(t) = \int_{\{t\}\times S} dt(\mathfrak{J})\, \sqrt{|g|}\omega_S \geq 0,$$

we have the *energy inequality*

$$|\mathcal{E}(t_1) - \mathcal{E}(t_0)| \leq C \int_{t_0}^{t_1} \mathcal{E}(t)\, dt. \tag{12.3.2}$$

To this we may apply Grönwall's inequality to obtain

$$\mathcal{E}(t_1) \leq e^{C(t_1 - t_0)} \mathcal{E}(t_0), \quad \mathcal{E}(t_0) \leq e^{C(t_1 - t_0)} \mathcal{E}(t_1)$$

for any $-1 < t_0 < t_1 < 1$.

Finally, setting one of t_0 and t_1 to be 0, we may use the hypothesis that ϕ and $\nabla\phi$ both vanish along $\Xi(0, S)$ to see that $\mathfrak{Q}(\phi, \nabla\phi)|_{\{0\}\times S} = 0$, and hence $\mathcal{E}(0) = 0$. The inequalities above then imply that $\mathcal{E}(t) = 0$ for all $t \in (-1, 1)$. By the positive definiteness guaranteed by Lemma 11.18 we see therefore $dt(\mathfrak{J}) = 0$ everywhere and hence ϕ and $\nabla\phi$ both vanish throughout the lens-shaped domain. This proves our claim.

© Willie Wai-Yeung Wong

DRAFT COPY

Detour: the universal speed limit

12.4 The energy method used to prove Theorem 11.15 is a powerful one that applies to many other hyperbolic PDEs. In the context of general relativity, it can also be used to relate the dominant energy condition to the idea that no matter can travel faster than the speed of light (or rather, the speed of gravity).

12.5 Consider now Einstein's equation with unspecified matter

$$\mathrm{Ric} - \frac{1}{2} R g + \Lambda g = T;$$

necessarily the energy–momentum tensor T is divergence-free. Suppose now that T satisfies the dominant energy condition. That is, given two time-like vectors X, Y with $g(X,Y) < 0$, we have $T(X,Y) \geq 0$.

Suppose now we are given a lens-shaped domain given by S and Ξ that is space-like relative to the space-time metric g. Suppose further that T vanishes along $\Xi(0,S)$. We can choose an arbitrary time-like vector field X and set $J^a = (g^{-1})^{ab} T_{bc} X^c$, which satisfies $\mathrm{div} J = T_{bc} (g^{-1})^{ab} \nabla_a X^c$. Then exactly the same argument as the proof of Theorem 11.15 (by considering the divergence/Stokes' theorem applied to J) shows that $dt(J)$ must vanish everywhere in our lens-shaped domain.

If we further impose the coercivity condition that $T(X,Y) = 0$ for time-like vectors X, Y if and only if $T = 0$ (which holds for many physical models and can be regarded as a strict form of the dominant energy condition), then this result would indicate that if T vanishes along a single slice of a lens-shaped domain, it must vanish within the entire lens-shaped domain. The physical interpretation of this is that "the edge of the vacuum region cannot receded too fast".

Dependence Domains

12.6 We start by taking some generalizations of Theorem 11.15. First, while the theorem was stated for the uniqueness of the zero solution, we can in fact use it to prove the uniqueness of general solutions.

12.7 Corollary
Let $\mathcal{V} \to M$ be a vector bundle with connection $\mathring{\nabla}$, suppose ϕ and ψ are two sections of $\mathcal{V}$ that solve

$$\mathfrak{g}(\phi)^{ab} \mathring{\nabla}_a \mathring{\nabla}_b \phi = \mathfrak{S}(\phi, \nabla\phi)$$

where $\mathfrak{g}$ and $\mathfrak{S}$ are as in Theorem 10.18. Assume that the solution ϕ is such that $\mathfrak{g}(\phi)$ has Lorentzian signature everywhere. Suppose S, Ξ defines a lens-shaped domain in M that is space-like with respect to $\mathfrak{g}(\phi)$. If $(\phi, \mathring{\nabla}\phi)$ and $(\psi, \mathring{\nabla}\psi)$ agree on $\Xi(0,S)$, then the two solutions agree on $\Xi(t,S)$ for all $t \in (-1,1)$. ∎

Proof. Let $\Delta = \psi - \phi$, another section of $\mathcal{V}$. We shall consider ϕ as *fixed*. We have

$$\mathfrak{g}(\phi)^{ab} \mathring{\nabla}_a \mathring{\nabla}_b \Delta = \Big[\mathfrak{g}(\phi) - \mathfrak{g}(\phi+\Delta)\Big]^{ab} \mathring{\nabla}_a \mathring{\nabla}_b \psi + \mathfrak{S}(\phi+\Delta, \mathring{\nabla}(\phi+\Delta)) - \mathfrak{S}(\phi, \mathring{\nabla}\phi).$$

The smoothness of $\mathfrak{g}$ and $\mathfrak{S}$ ensures that there exists smooth bundle maps $\mathfrak{h}$ and Υ, both of which depend implicitly on both ϕ, such that

$$\mathfrak{g}(\phi+\Delta)-\mathfrak{g}(\phi)=\mathfrak{h}(\Delta), \quad \mathfrak{S}(\phi+\Delta,\mathring{\nabla}(\phi+\Delta))-\mathfrak{S}(\phi,\mathring{\nabla}\phi)=\Upsilon(\Delta,\mathring{\nabla}\Delta).$$

Note in particular that $\mathfrak{h}(0)=0$ and $\Upsilon(0,0)=0$. So regarding $\mathring{\nabla}_a\mathring{\nabla}_b\psi$ as now a *fixed* smooth section, we have that Δ solves the wave equation

$$\mathfrak{g}(\phi)^{ab}\mathring{\nabla}_a\mathring{\nabla}_b\Delta=-(\mathring{\nabla}_a\mathring{\nabla}_b\psi)\mathfrak{h}(\Delta)^{ab}+\Upsilon(\Delta,\mathring{\nabla}\Delta),$$

here the equation is semilinear, as the coefficients $\mathfrak{g}(\phi)^{ab}$ are *independent* of Δ. Importantly, the right hand side vanishes when Δ and $\mathring{\nabla}\Delta$ vanish. In view of the additional hypothesis that along $\Xi(0,S)$ that ensures Δ and $\mathring{\nabla}\Delta$ vanish there, we may apply Theorem 11.15 to Δ to conclude. □

12.8 The uniqueness theorem depends on a choice of a lens-shaped domain. Let's globalize the statement and hide this dependence using a definition.

Definition (Domain of dependence)
Let (M,g) be a Lorentzian manifold, and Σ a hypersurface. We say that an open subset $U\subseteq M$ is a dependence domain of Σ *if for every $p\in U$, there exists a lens-shaped domain (S_p,Ξ_p), space-like relative to g, such that*

- *$\Xi(0,S_p)$ is a subset of Σ,*
- *$\Xi((-1,1)\times S)\subseteq U$, and*
- *there exists some $t\in(-1,1)$ and $q\in S_p\setminus\partial S_p$ such that $p=\Xi(t,q)$.* ◊

Using this concept we can formulate a global version of Corollary 12.7.

12.9 Corollary
Let $\mathcal{V}\to M$ be a vector bundle with connection $\mathring{\nabla}$, suppose ϕ and ψ are two sections of $\mathcal{V}$ that solve

$$\mathfrak{g}(\phi)^{ab}\mathring{\nabla}_a\mathring{\nabla}_b\phi=\mathfrak{S}(\phi,\nabla\phi)$$

where $\mathfrak{g}$ and $\mathfrak{S}$ are as in Theorem 10.18. Assume that the solution ϕ is such that $\mathfrak{g}(\phi)$ has Lorentzian signature everywhere. Suppose there exists a hypersurface $\Sigma\hookrightarrow M$ along which $(\phi,\mathring{\nabla}\phi)$ and $(\psi,\mathring{\nabla}\psi)$ agree. Then for every $U\subseteq M$ that is a dependence domain of Σ with respect to the metric $\mathfrak{g}(\phi)$, we have $\phi|_U=\psi|_U$. ■

12.10 We can now apply this uniqueness result to say something about the uniqueness of the solution to Einstein–vacuum equations, for a given initial data. But before stating and proving the theorem, we state one useful lemma.

Lemma
Let (M,g) be a Lorentzian manifold, Σ a hypersurface, and $U\subseteq M$ an open dependence domain of Σ. Given any $V\subseteq M$ that is an open subset that contains $U\cap\Sigma$, there exists $W\subseteq V$ open, with $W\supseteq U\cap\Sigma$, and such that W is a dependence domain of Σ. ■

Proof. Let $\mathscr{K}_U$ denote the set of all lens-shaped domains (S,Ξ) satisfying:

- (S,Ξ) is space-like relative to g,

DRAFT COPY

DRAFT COPY

- $\Xi(0, S)$ is a subset of Σ,
- $\Xi((-1,1)\times S) \subseteq U$.

Now let V be as given. For any $(S,\Xi) \in \mathscr{K}_U$, we have that $\Xi(0,S) \subseteq V$. Since S is compact and V is open, we see that there exists some $\epsilon > 0$ such that $\Xi((-\epsilon,\epsilon)\times S) \subseteq V$. So if we set $\Xi'(t,q) = \Xi(\epsilon t, q)$, we have that (S,Ξ') belongs to $\mathscr{K}_V$. In particular, we see that $\mathscr{K}_V$ is not empty. Furthermore, the fact that U is a dependence domain of Σ means that every point $p \in U \cap \Sigma$ arises as $\Xi(0,q)$ for some $q \in S \setminus \partial S$ with $(S,\Xi) \in \mathscr{K}_U$. This implies that p is also representable as $\Xi'(0,q)$ for some $q \in S \setminus \partial S$ with $(S,\Xi') \in \mathscr{K}_V$.

Now, define W by

$$W := \bigcup \{\Xi'((-1,1)\times S \setminus \partial S) \mid (S,\Xi') \in \mathscr{K}_V\}.$$

By definition $W \subseteq V$, and that W is a dependence domain of Σ. The previous paragraph shows that $U \cap \Sigma$ is contained in W. □

12.11 Theorem (Local uniqueness for Einstein–vacuum equations)
Let (Σ,γ,k) be an initial data set. Suppose (M,g) and $(\tilde{M},\tilde{g})$ are two solutions to the Einstein–vacuum equation, and suppose further that there exists embeddings $\Phi : \Sigma \to M$ and $\tilde{\Phi} : \Sigma \to \tilde{M}$ such that γ and k are the first and second fundamental forms for both *embeddings. Suppose furthermore that M is a dependence domain of $\Phi(\Sigma)$ with respect to the metric g. Then there exists a neighborhood U of $\Phi(\Sigma)$ and a neighborhood $\tilde{U}$ of $\tilde{\Phi}(\Sigma)$, together with a diffeormorphism $\Psi : U \to \tilde{U}$ such that $g = \Psi^* \tilde{g}$.* ■

Proof. Consider the wave map problem mapping from $(M,g) \to (\mathbb{R}\times\Sigma, ds^2 + \gamma)$; this is described by a nonlinear wave equation. Denote by θ the putative wave map. We will prescribe initial data at $\Phi(\Sigma)$, so that for $x \in \Sigma$, we have $\theta_0(\Phi(x)) = (0,x)$ in $\mathbb{R}\times\Sigma$, and so that $\theta_1(n) = \partial_s$ where n is the unit normal to $\Phi(\Sigma)$. By Theorem 10.18 a solution θ exists on some neighborhood V of $\Phi(\Sigma)$; since θ_1 is unto (as a mapping from TM to $T(\mathbb{R}\times\Sigma)$) we can, after shrinking the neighborhood a little bit if necessary, assume that θ is a diffeomorphism from V to its image. Furthermore, by Lemma 12.10, shrinking V if necessary we can assume that V is a dependence domain of $\Phi(\Sigma)$ relative to g.

Do the same thing with $(\tilde{M},\tilde{g})$ and obtain an open set $\tilde{V}$ and a diffeomorphism $\tilde{\theta}$. (Note for $\tilde{V}$ we do not make any claims about it being a dependence domain.)

By construction, if we push forward the metric g by θ, we obtain on $\theta(V)$ a solution to the Einstein–vacuum equations, and with respect to $\mathring{h} = ds^2 + \gamma$ the wave-map gauge conditions are fulfilled. So the reduced Einstein equations and the equations for $\mathfrak{X}$ are satisfied. Similarly for the pushforward of $\tilde{g}$ by $\tilde{\theta}$. Now our choice that $\theta_1(n) = \partial_s$ means that the pushforward metric $\theta_* g$ has ∂_s as the unit normal to $\{0\}\times\Sigma$, and similarly for $\tilde{\theta}_*\tilde{g}$. And so using the previous proof we see that the two solutions *generate the same initial data* for the reduced Einstein system on $\{0\}\times\Sigma$.

Next, we have $\theta(V)$ and $\tilde{\theta}(\tilde{V})$ are both open neighborhoods of $\{0\}\times\Sigma$ in $\mathbb{R}\times\Sigma$; so is their intersection. Within this intersection by Lemma 12.10 again we can find an open neighborhood W of $\{0\}\times\Sigma$ such that W is a dependence domain of Σ with respect to $\theta_* g$. On W we can now apply Corollary 12.9 to conclude that the two solutions are equal. Finally we define $U = \theta^{-1}(W)$ and $\tilde{U} = \tilde{\theta}^{-1}(W)$, and the diffeomorphism $\Psi = \tilde{\theta}^{-1}\circ\theta$. □

© Willie Wai-Yeung Wong

DRAFT COPY

12.12 One may worry about the requirement in Theorem 12.11 that for at least one of the solutions, M is a dependence domain of $\Phi(\Sigma)$ relative to g. It turns out that this is *automatic* of solutions constructed using the existence theorem Theorem 11.6 (or rather, the general result for quasilinear waves given as Theorem 10.18), as the usual proofs of the exists theorem is also based on energy estimates similar to that which was used in Theorem 11.15.

12.13 In fact, we have also the following result which asserts that locally the construction of lens-shaped domain is possible, which would obviate the assumption imposed in Theorem 12.11.

Proposition
Let (M,g) be a Lorentzian manifold, and Σ a space-like hypersurface. Given any $p \in \Sigma$, there exists a lens-shaped domain (S,Ξ) that is space-like relative to g, such that $\Xi(0,S) \subseteq \Sigma$, and $p \in \Xi(0, S \setminus \partial S)$. ■

Proof. Let x^0 be a local defining function of Σ. Complete this to a coordinate system $\{x^0, x^1, \dots, x^d\}$ for a small neighborhood of p in M, ensure the coordinates are chosen so that p is at the origin of the coordinate system, and at p, the coordinate components of the metric g is precisely that of the Minkowski metric $\mathrm{diag}(-1,1,1,\dots,1)$. Near p we have by continuity that g is approximately the Minkowski metric.

Now consider Minkowski space $\mathbb{R}^{1,d}$, and let B_r to be the closed ball of radius r in $\mathbb{R}^d$ centered at the origin. Consider the mapping $\Xi : (-1,1) \times B_r \to \mathbb{R}^{1,d}$ given by

$$\Xi(t,x) = \Big(\frac{1}{3r}(|x|^2 - r^2)t, x\Big).$$

A direct computation checks that this defines a lens-shaped domain that is space-like with respect to the Minkowski metric.

Now consider Ξ as taking values in M, through the local coordinate system mentioned above. For sufficiently small r we will have $\Xi(0, B_r) \subseteq \Sigma$. Furthermore, when r is sufficiently small, on $\Xi((-1,1) \times B_r)$ the metric g is sufficiently close to the Minkowski metric, which guarantees that (B_r, Ξ) is also space-like with respect to g. □

12.14 In Theorem 12.11 the two solutions are only shown to agree on a subset that is determined by the size of the set W that appears in Lemma 12.10. Since W here was built based on the continuity of the Ξ mapping in the lens-shaped domain definition, one may be concerned that this W may be very tiny.

It turns out that W often be enlarged.

Observe that W is a subset of both $\theta(V)$ and $\tilde{\theta}(\tilde{V})$, and on W the two solutions agree, so by continuity the solutions must also agree on the closure of W. Let $\hat{W} = \theta(V) \cap \tilde{\theta}(\tilde{V})$. Suppose now $q \in \tilde{W}$ lies on the boundary of W. If it were the case that that there exists a space-like hypersurface Σ' that lies in $\overline{W} \cap \tilde{W}$ and passes through q, then we can apply the uniqueness theorem again to Σ' to show that on an open neighborhood of Σ' (and hence an open neighborhood of q) the two solutions should agree (here we used Proposition 12.13 to extract a lens-shaped domain near q).

© Willie Wai-Yeung Wong

12.15 Connected to this notion of enlarging W is the notion of a Cauchy development.

Definition

Let (M, g) be a Lorentzian manifold. Suppose Σ is a hypersurface, and $U \subseteq M$ is an open subset.

1. *We say that Σ is a* Cauchy hypersurface *of U if every time-like inextensible curve $\gamma : (a, b) \to U$ intersects Σ exactly once.*
2. *We say that U is a* Cauchy development *of Σ if $\Sigma \subseteq U$ and Σ is a Cauchy hypersurface of U.* ◊

12.16 Lemma

Let (S, Ξ) define a space-like lens-shaped domain in (M, g). Then $\Xi((-1, 1) \times (S \setminus \partial S))$ is a Cauchy development of $\Xi(t, S \setminus \partial S)$ for any $t \in (-1, 1)$. ■

Sketch of proof. Let γ be a time-like inextensible curve in $\Xi((-1, 1) \times (S \setminus \partial S))$. Our assumption that (S, Ξ) is space-like implies that along γ the "time coordinate" is strictly monotone. And hence we see that γ passes through each $\Xi(t, S \setminus \partial S)$ at most once. Suppose it never hits $\Xi(t_0, S\partial S)$. Then γ must exit through the "side" of the domain, namely that γ will converge toward $\{t_m\} \times \partial S$ for some $t_m \in (-1, 1)$. But this causes a contradiction since, for this to happen, using the fact that Ξ collapses $(-1, 1) \times \partial S$, we must have that $\dot{\gamma}$ is space-like in the limit, this contradicts the assumption that γ is time-like. □

12.17 The following theorem is the culmination of the ideas outlined above. Its proof involves some rather subtle technical arguments, and so we will not detail it here. Interested readers can find accounts given in Ringström, *The Cauchy Problem in General Relativity*; Sbierski, "On the Existence of a Maximal Cauchy Development for the Einstein Equations: a Dezornification".

Theorem

Let Σ be a smooth manifold of dimension $d \geq 2$, equipped with a Riemannian metric γ and a symmetric two tensor k such that the constraint equations hold. Let Sol denote the set of $(d + 1)$-dimensional Lorentzian manifolds (M, g) with the property

- *(M, g) solves the Einstein–vacuum equations;*
- *Σ embeds into M as a hypersurface with γ and k the first and second fundamental forms respectively;*
- *M is a Cauchy development of Σ.*

Given (M, g) and (M', g') in Sol, we say that (M, g) extends *(M', g') if there exists an embedding $\phi : M' \to M$ as an open submanifold, such that $g' = \phi^* g$ and ϕ fixes the embedding of Σ. Then:*

1. *(Existence) The set Sol is non-empty.*
2. *(Uniqueness) Given any pair (M, g) and (M', g') in Sol, there exists $(N, h) \in \mathsf{Sol}$ such that both (M, g) and (M', g') are extensions of (N, h).*
3. *(Directedness) Given any pair (M, g) and (M', g') in Sol, there exists $(N, h) \in \mathsf{Sol}$ that simulatneously extends both (M, g) and (M', g').*
4. *(Maximality) Sol has a maximal element; that is, there exists $(M, g) \in \mathsf{Sol}$ such that it is an extension of any other (M', g') in Sol.* ■

The first two claims are largely the same as what we have already proven in these notes. The most difficult to prove is the third claim on directedness; fundamentally this requires first finding the largest "W" (see ¶ 12.14) and showing that this W is large enough that

DRAFT COPY

© Willie Wai-Yeung Wong

there is no ambiguity when we try to glue the two solutions together to form a larger solution. The argument showing that this gluing will result in a *Hausdorff* manifold is particularly complicated. It turns out that once directedness is proven, maximality follows immediately by abstract nonsense. (See W. W. Wong, "A comment on the construction of the maximal globally hyperbolic Cauchy development".)

DRAFT COPY

DRAFT COPY

Further Reading

An et al.: Warped Product Space-times

X An and WWY Wong. "Warped Product Space-times". In: *Class. Quantum Grav.* 35 (Jan. 2018), p. 025011.

Beem et al.: Global Lorentzian geometry

John K. Beem, Paul E. Ehrlich, and Kevin L. Easley. *Global Lorentzian geometry*. Second. Vol. 202. Monographs and Textbooks in Pure and Applied Mathematics. New York: Marcel Dekker Inc., 1996, pp. xiv+635. ISBN: 0-8247-9324-2.

Cartan: Sur les variétés à connexion affine et la théorie de la relativité généralisée (première partie)

Elie Cartan. "Sur les variétés à connexion affine et la théorie de la relativité généralisée (première partie)". In: *Ann. Sci. École Norm. Sup. (3)* 40 (1923), pp. 325–412. ISSN: 0012-9593. URL: http://www.numdam.org/item?id=ASENS_1923_3_40__325_0.

Choquet-Bruhat: General relativity and the Einstein equations

Yvonne Choquet-Bruhat. *General relativity and the Einstein equations*. Oxford Mathematical Monographs. Oxford: Oxford University Press, 2009, pp. xxvi+785. ISBN: 978-0-19-923072-3.

Christodoulou: A mathematical theory of gravitational collapse

Demetrios Christodoulou. "A mathematical theory of gravitational collapse". In: *Comm. Math. Phys.* 109.4 (Dec. 1987), pp. 613–647. DOI: 10.1007/BF01208960. URL: http://www.springerlink.com/content/k84h510vw04h3831.

Christodoulou: On the global initial value problem and the issue of singularities

Demetrios Christodoulou. "On the global initial value problem and the issue of singularities". In: *Classical Quant. Grav.* 16.12A (1999), A23. URL: http://stacks.iop.org/0264-9381/16/i=12A/a=302.

DRAFT COPY

Delay: Smooth compactly supported solutions of some underdetermined elliptic PDE, with gluing applications

Erwann Delay. "Smooth compactly supported solutions of some underdetermined elliptic PDE, with gluing applications". In: (Mar. 2010). eprint: 1003.0535.

DeTurck et al.: Some regularity theorems in Riemannian geometry

Dennis M. DeTurck and Jerry L. Kazdan. "Some regularity theorems in Riemannian geometry". In: *Ann. Sci. École Norm. Sup. (4)* 14.3 (1981), pp. 249–260. ISSN: 0012-9593. URL: http://www.numdam.org/item?id=ASENS%5C_1981%5C_4%5C_14%5C_3%5C_249%5C_0.

Flores: The Riemannian and Lorentzian Splitting Theorems

José Luis Flores. "The Riemannian and Lorentzian Splitting Theorems". In: *Topics in Modern Differential Geometry*. Ed. by Stefan Haesen and Leopold Verstraelen. Paris: Atlantis Press, 2017, pp. 1–20. ISBN: 978-94-6239-240-3. DOI: 10.2991/978-94-6239-240-3_1. URL: https://doi.org/10.2991/978-94-6239-240-3_1.

Fourès-Bruhat: Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires

Yvonne Fourès-Bruhat. "Théorème d'existence pour certains systèmes d'équations aux dérivées partielles non linéaires". In: *Acta Math.* 88.1 (Dec. 1952), pp. 141–225. DOI: 10.1007/BF02392131. URL: http://www.springerlink.com/content/43234177133б1248.

Galloway: Null geometry and the Einstein equations

Gregory J. Galloway. "Null geometry and the Einstein equations". In: *The Einstein equations and the large scale behavior of gravitational fields*. Birkhäuser, Basel, 2004, pp. 379–400. ISBN: 3-7643-7130-7.

Galloway: The Lorentzian splitting theorem without the completeness assumption

Gregory J. Galloway. "The Lorentzian splitting theorem without the completeness assumption". In: *J. Differential Geom.* 29.2 (1989), pp. 373–387. ISSN: 0022-040X,1945-743X. URL: http://projecteuclid.org/euclid.jdg/1214442881.

S W Hawking et al.: The gravitational Hamiltonian, action, entropy and surface terms

S W Hawking and Gary T Horowitz. "The gravitational Hamiltonian, action, entropy and surface terms". In: *Classical and Quantum Gravity* 13.6 (June 1996), p. 1487. DOI: 10.1088/0264-9381/13/6/017. URL: https://dx.doi.org/10.1088/0264-9381/13/6/017.

Stephen William Hawking et al.: The large scale structure of space-time

Stephen William Hawking and George F. R. Ellis. *The large scale structure of space-time*. Cambridge University Press, 1973.

© Willie Wai-Yeung Wong

DRAFT COPY

Hughes et al.: Well-posed quasi-linear second-order hyperbolic systems with applications to nonlinear elastodynamics and general relativity

Thomas J. R. Hughes, Tosio Kato, and Jerrold E. Marsden. "Well-posed quasi-linear second-order hyperbolic systems with applications to nonlinear elastodynamics and general relativity". In: *Arch. Rational Mech. Anal.* 63.3 (1976), 273–294 (1977). ISSN: 0003-9527.

Kiessling: Electromagnetic field theory without divergence problems. I. The Born legacy

Michael K.-H. Kiessling. "Electromagnetic field theory without divergence problems. I. The Born legacy". In: *J. Stat. Phys.* 116.1-4 (2004), pp. 1057–1122. ISSN: 0022-4715. DOI: 10.1023/B:JOSS.0000037250.72634.2a. URL: http://dx.doi.org/10.1023/B:JOSS.0000037250.72634.2a.

Lee: Geometric Relativity

Dan Lee. *Geometric Relativity*. Graduate Studies in Mathematics 201. American Mathematical Society, 2019.

Lindblad et al.: Global Existence for the Einstein Vacuum Equations in Wave Coordinates

Hans Lindblad and Igor Rodnianski. "Global Existence for the Einstein Vacuum Equations in Wave Coordinates". In: *Comm. Math. Phys.* 256.1 (2005), pp. 43–110. DOI: 10.1007/s00220-004-1281-6. URL: http://www.springerlink.com/content/r6147k78v2946247.

Lindblad et al.: The global stability of Minkowski space-time in harmonic gauge

Hans Lindblad and Igor Rodnianski. "The global stability of Minkowski space-time in harmonic gauge". In: *Ann. of Math. (2)* 171.3 (2010), pp. 1401–1477. ISSN: 0003-486X. DOI: 10.4007/annals.2010.171.1401. URL: http://dx.doi.org/10.4007/annals.2010.171.1401.

Milnor: Topology from the differentiable viewpoint

John W. Milnor. *Topology from the differentiable viewpoint*. Princeton Landmarks in Mathematics. Based on notes by David W. Weaver, Revised reprint of the 1965 original. Princeton University Press, Princeton, NJ, 1997, pp. xii+64. ISBN: 0-691-04833-9.

Misner et al.: Gravitation

Charles W. Misner, Kip S. Thorne, and John Archibald Wheeler. *Gravitation*. San Francisco, Calif.: W. H. Freeman and Co., 1973, ii+xxvi+1279+iipp.

O'Neill: Semi-Riemannian geometry: with applications to relativity

Barrett O'Neill. *Semi-Riemannian geometry: with applications to relativity*. Academic Press, 1983.

© Willie Wai-Yeung Wong

DRAFT COPY

Ringström: The Cauchy Problem in General Relativity

Hans Ringström. *The Cauchy Problem in General Relativity*. European Mathematical Society, 2009.

Sbierski: On the Existence of a Maximal Cauchy Development for the Einstein Equations: a Dezornification

Jan Sbierski. "On the Existence of a Maximal Cauchy Development for the Einstein Equations: a Dezornification". In: *Annales Henri Poincaré* 17 (2016), pp. 301–329.

Truesdell et al.: The non-linear field theories of mechanics

C. Truesdell and W. Noll. *The non-linear field theories of mechanics*. Third. Edited and with a preface by Stuart S. Antman. Springer-Verlag, Berlin, 2004, pp. xxx+602. ISBN: 3-540-02779-3. DOI: 10.1007/978-3-662-10388-3. URL: https://doi.org/10.1007/978-3-662-10388-3.

Wald: General Relativity

Robert M. Wald. *General Relativity*. University of Chicago Press, 1984.

W. W. Wong: A comment on the construction of the maximal globally hyperbolic Cauchy development

Willie WY Wong. "A comment on the construction of the maximal globally hyperbolic Cauchy development". In: *J. Math. Physics* 54 (Nov. 2013), p. 113511.

© Willie Wai-Yeung Wong