GLOBAL LORENTZIAN GEOMETRY

JAMES D.E. GRANT

Contents

Part 1. Preliminary material and notational conventions 2
- Vector bundles 2
- Connections 2
- Semi-Riemannian manifolds 3
- Maps between manifolds 4

Part 2. Riemannian geometry 5
- 1. Examples 5
- 2. Metric space structure on Riemannian manifold 6
- 3. First variation of arc-length 7
- 4. The exponential map and normal coordinates 9
- 5. Hopf-Rinow theorem 11
- 6. Bishop-Gromov comparison theorem 11
- 7. Second variation of arc-length 18
- 8. Cut locus 24

Part 3. Lorentzian geometry 28
- 9. Inner products of Lorentzian signature 28
- 10. Lorentzian manifolds 28
- 10.1. Examples 29
- 11. Geodesics 33
- 12. Lengths of causal curves 35
- 13. Variation of arc-length 37
- 13.1. First variation of arc-length 37
- 13.2. Second variation of arc-length 37
- 14. Causality 42
- 14.1. Causality relations 42
- Appendix A. Proof of Proposition 14.8 (Non-examinable) 45
- References 49
- Index 50

Date: 31 January, 2008 (Corrected: 12 February, 2009).

Lecture 1

Part 1. Preliminary material and notational conventions

You should be familiar with everything in this section from previous Differential Geometry courses. Just in case, however, I will (eventually) include an Appendix that reviews any previous material that I have used. If you still have problems, standard textbooks on this subject include [GHL, O'N, Cha, W].

Throughout, let M be a finite-dimensional manifold of dimension $n \geq 2$. We will assume that M is Hausdorff, connected and, except where explicitly stated otherwise, that M is smooth (i.e. C^∞).

Vector bundles. Let $\pi : V \to M$ be a vector bundle. (Again, except where stated otherwise, we will assume that vector bundles are smooth.) The space of sections of V will be denoted by $\Gamma(V)$, or $\Gamma(M, V)$ if we need to emphasise the manifold in question, or $C^l(M, V)$ ($l \leq \infty$) if we wish to emphasise the amount of differentiability required. Particular cases, with special notation, are:

- TM, the tangent bundle of M. We denote the space of smooth sections of TM by $\mathfrak{X}(M)$. An element of $\mathfrak{X}(M)$ therefore corresponds to a vector field on M. Recall that we view vector fields as differential operators acting on smooth functions. So, for each $\mathbf{X} \in \mathfrak{X}(M)$ and $f \in C^\infty(M)$, we have $\mathbf{X}f \in C^\infty(M)$. Given $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M)$, we define the *Lie bracket* $[\mathbf{X}, \mathbf{Y}] \in \mathfrak{X}(M)$ by
$$[\mathbf{X}, \mathbf{Y}]\, f = \mathbf{X}\left(\mathbf{Y}\left(f\right)\right) - \mathbf{X}\left(\mathbf{Y}\left(f\right)\right),$$
for all $f \in C^\infty(M)$.
- $\wedge^k T^*M$ is the bundle of k-forms on M. Its space of sections, the space of (differential) k-forms, is denoted $\Omega^k(M)$. Given a vector bundle $V \to M$, we will denote by $\Omega^k(M, V)$ the space of k-forms with values in V (i.e. the space of sections of the bundle $\wedge^k T^*M \otimes V$).
- $T^r_s(M)$ will denote the bundle of (r, s) tensors on M (i.e. $(\otimes^r_1 TM) \otimes (\otimes^s_1 T^*M)$). The set of sections of this bundle (i.e. (r, s) tensor fields on M) will be denoted by $\mathcal{T}^r_s(M)$.

Notation: Given a vector field $\mathbf{X} \in \mathfrak{X}(M)$ and a 1-form $\boldsymbol{\sigma} \in \Omega^1(M)$, we will denote the action of $\boldsymbol{\sigma}$ on $\mathbf{X}$ by $\langle \boldsymbol{\sigma}, \mathbf{X} \rangle$ or $\boldsymbol{\sigma}(\mathbf{X})$.

Connections. A *connection* on a vector bundle, V, is equivalent to a *covariant derivative*. This is a map $\nabla : \mathfrak{X}(M) \times \Gamma(V) \to \Gamma(V); (\mathbf{X}, s) \mapsto \nabla_{\mathbf{X}} s$ with the properties that

$$\begin{aligned}
\nabla_{f\mathbf{X}+g\mathbf{Y}} s &= f \nabla_{\mathbf{X}} s + g \nabla_{\mathbf{Y}} s, &&\forall f, g \in C^\infty(M), \quad \mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M), \quad s \in \Gamma(V),\\
\nabla_{\mathbf{X}}(as + bt) &= a \nabla_{\mathbf{X}} s + b \nabla_{\mathbf{X}} t, &&\forall a, b \in \mathbb{R}, \quad \mathbf{X} \in \mathfrak{X}(M), \quad s, t \in \Gamma(V),\\
\nabla_{\mathbf{X}}(fs) &= \mathbf{X}(f) s + f \nabla_{\mathbf{X}} s, &&\forall f \in C^\infty(M), \quad \mathbf{X} \in \mathfrak{X}(M), \quad s \in \Gamma(V).
\end{aligned}$$

Given a connection on a vector bundle, the *curvature* is the 2-form with values in the bundle of endomorphisms of V, $\operatorname{End} V$, defined by

$$\left(\nabla_{\mathbf{X}} \nabla_{\mathbf{Y}} - \nabla_{\mathbf{Y}} \nabla_{\mathbf{X}} - \nabla_{[\mathbf{X},\mathbf{Y}]}\right) s = \mathbf{F}(\mathbf{X}, \mathbf{Y}) s, \qquad \forall \mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M), \quad s \in \Gamma(V).$$

Hence, given vector fields $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M)$ and a section, s, of V, the object $\mathbf{F}(\mathbf{X}, \mathbf{Y})s$ is also a section of V.

An *affine connection* is a connection on the tangent bundle TM. In this case, we define the *torsion* tensor $\mathbf{T} \in \mathcal{T}^1_2(M)$, by

$$\nabla_{\mathbf{X}} \mathbf{Y} - \nabla_{\mathbf{Y}} \mathbf{X} - [\mathbf{X}, \mathbf{Y}] = \mathbf{T}(\mathbf{X}, \mathbf{Y}) \in \mathfrak{X}(M), \qquad \forall \mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M).$$

In the case of an affine connection, we have a special notation for the curvature tensor $\mathbf{R} \in \mathcal{T}^1_3(M)$:

$$\left(\nabla_{\mathbf{X}} \nabla_{\mathbf{Y}} - \nabla_{\mathbf{Y}} \nabla_{\mathbf{X}} - \nabla_{[\mathbf{X},\mathbf{Y}]}\right) \mathbf{Z} = \mathbf{R}(\mathbf{X}, \mathbf{Y}) \mathbf{Z}, \qquad \forall \mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathfrak{X}(M).$$

As usual, we may extend an affine connection to define a connection on each tensor bundle $T^r_s(M)$.

Semi-Riemannian manifolds. A case that we will be particularly interested in is when M has a Riemannian or semi-Riemannian metric. This is a section, $\mathbf{g}$, of $T_2^0(M)$ that is

- symmetric: $\mathbf{g}(\mathbf{X}, \mathbf{Y}) = \mathbf{g}(\mathbf{Y}, \mathbf{X})$, $\forall \mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M)$;
- non-degenerate: $\mathbf{g}(\mathbf{X}, \mathbf{Y}) = 0$ for all $\mathbf{Y} \in \mathfrak{X}(M)$ if and only if $\mathbf{X} = 0$;
- defines an inner product of signature (p, q) (p plusses, q minuses) on each tangent space T_xM, for each $x \in M$. (For example, the canonical example of an inner product of signature (p, q) would be on $\mathbb{R}^n$, with $n = p + q$, where, given $\mathbf{v} = (v^1, \ldots, v^n)$, $\mathbf{w} = (w^1, \ldots, w^n) \in \mathbb{R}^n$ we define
$$\langle \mathbf{v}, \mathbf{w} \rangle := v^1 w^1 + \cdots + v^p w^p - v^{p+1} w^{p+1} - \cdots - v^{p+1} w^{p+1}.)$$
The cases we will be most interested in are:
a). $p = n$, $q = 0$, in which case $\mathbf{g}$ is a *Riemannian metric*;
b). $p = n - 1$, $q = 1$, in which case $\mathbf{g}$ is a *Lorentzian metric*.
By a semi-Riemannian manifold, we will mean a pair $(M, \mathbf{g})$ consisting of a manifold M and a semi-Riemannian metric, $\mathbf{g}$, on M[1].

Remark 0.1. Given a point $x \in M$ we may pick an orthonormal basis for TM on a neighbourhood, U, of x i.e. a basis $\{\mathbf{e}_i(y)\}_{i=1}^n$ for T_yM, for each $y \in U$, such that
$$\mathbf{g}_y(\mathbf{e}_i(y), \mathbf{e}_j(y)) = \begin{cases} \epsilon_i & i = j, \\ 0 & i \neq j, \end{cases} \qquad \forall y \in U,$$
where $\epsilon_1 = \cdots = \epsilon_p = +1$, $\epsilon_{p+1} = \ldots \epsilon_{p+q} = -1$ for a semi-Riemannian metric of signature (p, q).

Definition 0.2. An affine connection, ∇, on $(M, \mathbf{g})$ is *metric* if $\nabla \mathbf{g} = 0$. It is *torsion-free* if $\mathbf{T} = 0$.

The following result is fundamental:

Theorem. *Let $(M, \mathbf{g})$ be a semi-Riemannian manifold. Then there exists a unique affine connection, ∇, that is torsion-free and metric.*

Sketch of Proof. From the explicit form of the torsion-free and metric conditions, we deduce the *Koszul formula*
$$\begin{aligned} 2\mathbf{g}(\nabla_{\mathbf{X}} \mathbf{Y}, \mathbf{Z}) = {} & \mathbf{X}\left(\mathbf{g}(\mathbf{Y}, \mathbf{Z})\right) + \mathbf{Y}\left(\mathbf{g}(\mathbf{X}, \mathbf{Z})\right) - \mathbf{Z}\left(\mathbf{g}(\mathbf{X}, \mathbf{Y})\right) \\ & + \mathbf{g}(\left[\mathbf{X}, \mathbf{Y}\right], \mathbf{Z}) - \mathbf{g}(\left[\mathbf{X}, \mathbf{Z}\right], \mathbf{Y}) - \mathbf{g}(\left[\mathbf{Y}, \mathbf{Z}\right], \mathbf{X}), \end{aligned}$$
for all $\mathbf{X}, \mathbf{Y}, \mathbf{Z} \in \mathfrak{X}(M)$. We then show that this uniquely determines the connection, and check that the connection defined by this formula is metric and torsion-free. □

The torsion-free metric connection given by this theorem is called the *Levi-Civita connection* for the metric $\mathbf{g}$. For the rest of this course, unless explicitly stated otherwise, the only affine connection that we will use on a semi-Riemannian manifold will be the Levi-Civita connection.

In the case of a semi-Riemannian manifold, it is also useful to define a $(4, 0)$ version of the curvature tensor, also denoted $\mathbf{R} \in \mathcal{T}_4^0(M)$, by
$$\mathbf{R}(\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = \mathbf{g}(\mathbf{R}(\mathbf{W}, \mathbf{X})\mathbf{Z}, \mathbf{Y}).$$
(Note the change in the order of vector fields.)

We also define the *Ricci curvature tensor* (or simply *Ricci tensor*). This is a symmetric $(0, 2)$ tensor field on M defined as the trace $\mathbf{Ric}(\mathbf{X}, \mathbf{Y}) := \mathrm{tr}\ [\mathbf{Z} \to \mathbf{R}(\mathbf{Z}, \mathbf{X})\mathbf{Y}]$. In this case, for $\mathbf{X}, \mathbf{Y} \in \mathfrak{X}(U)$, the Ricci tensor is defined on U by
$$\mathbf{Ric}(\mathbf{X}, \mathbf{Y}) = \sum_{i=1}^n \epsilon_i\, \mathbf{g}(\mathbf{e}_i, \mathbf{R}(\mathbf{e}_i, \mathbf{X})\mathbf{Y}).$$

[1]For brevity, we will often abbreviate the term "semi-Riemannian metric" to just "metric".

Finally, we define the *scalar curvature* (or *Ricci scalar*), s, as the trace of **Ric**:

$$s = \sum_{i=1}^{n} \epsilon_i \, \mathbf{Ric}(\mathbf{e}_i, \mathbf{e}_i).$$

Exercise 0.3. Prove that the Ricci tensor as defined in the last equation is symmetric. (You will need to use the fact that the Levi-Civita connection is torsion-free.)

Maps between manifolds. Let M, N be smooth manifolds, and $\varphi : M \to N$ be a smooth map. The differential of this map is the tangent map $D_p\varphi : T_pM \to T_{\varphi(p)}N$, and we define the corresponding bundle map $D\varphi : TM \to TN$ by $D\varphi(\mathbf{v}_p) := D_p\varphi(\mathbf{v}_p)$, for $\mathbf{v}_p \in T_pM$, $p \in M$.

Given a $(0, s)$ tensor field on N, $\mathbf{S} \in \mathcal{T}^0_s(N)$, the formula

$$\left.(\varphi^*\mathbf{S})\right|_p (\mathbf{v}_1, \ldots, \mathbf{v}_s) = \mathbf{S}_{\varphi(p)} \left(D_p\varphi(\mathbf{v}_1), \ldots, D_p\varphi(\mathbf{v}_s)\right), \qquad \forall \mathbf{v}_1, \ldots, \mathbf{v}_s \in T_pM, \qquad p \in M$$

defines a $(0, s)$ tensor field on M, the *pull-back of* $\mathbf{S}$ *by* φ, which we denote by $\varphi^*\mathbf{S} \in \mathcal{T}^0_s(M)$.

Part 2. Riemannian geometry

We begin by studying some global properties of Riemannian manifolds[2]. Therefore, for the remainder of this part of the course, we will assume that $(M, \mathbf{g})$ is a Riemannian manifold, so $\mathbf{g} \in \mathcal{T}_2^0(M)$ defines an inner product on $T_x M$ for each $x \in M$.

1. Examples

Example 1.1. Let $M = \mathbb{R}^n$, and take a chart $U = M$, $\varphi = Id$, and local Cartesian coordinates $(x^1, \ldots, x^n)$. The flat metric on $\mathbf{R}^n$ is given by

$$\mathbf{g} = \sum_{i=1}^{n} \left(dx^i\right)^2,$$

where we have introduced the notation $\left(dx^i\right)^2 := dx^i \otimes dx^i$. In this coordinate system, $g_{ij} = 1$ for $i = j$ and 0 otherwise. The Christoffel symbols and the components of the Riemann tensor are zero, so the metric is flat. **Lecture 2**

In the case $n = 2$, we introduce polar coordinates $(r, \theta) \in [0, \infty) \times [0, 2\pi)$ by $x^1 = r\cos\theta$, $x^2 = r\sin\theta$. In terms of these coordinates, the metric takes the form

$$\mathbf{g} = dr^2 + r^2 d\theta^2.$$

An example of an orthonormal basis would then be

$$\mathbf{e}_1 = \frac{\partial}{\partial r}, \qquad \mathbf{e}_2 = \frac{1}{r}\frac{\partial}{\partial \theta}.$$

More generally, the n-dimensional flat metric may be written in the form

$$\mathbf{g} = dr^2 + r^2 \mathbf{g}_{S^{n-1}}.$$

where $\mathbf{g}_{S^{n-1}}$ is the induced metric on the unit $(n-1)$ sphere in $\mathbb{R}^n$.

Example 1.2. Let $M = S^2$ with the metric induced by viewing M as a sphere of radius a in flat $\mathbb{R}^3$. In spherical polar coordinates $(\theta, \phi) \in [0, \pi] \times [0, 2\pi)$, the metric takes the form

$$\mathbf{g}_{S^2(a)} = a^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right).$$

An example of an orthonormal basis would be

$$\mathbf{e}_1 = \frac{1}{a}\frac{\partial}{\partial \theta}, \qquad \mathbf{e}_2 = \frac{1}{a\sin\theta}\frac{\partial}{\partial \phi}.$$

Letting $K := 1/a^2$, we may define a new coordinate $r := \theta/\sqrt{K}$ with $0 \leq r \leq \pi/\sqrt{K}$, in terms of which the metric is

$$\mathbf{g}_{S^2(a)} = dr^2 + \left(\frac{1}{\sqrt{K}}\sin\left(\sqrt{K}r\right)\right)^2 d\phi^2.$$

More generally, if $M = S^n$, with the metric induced by viewing M as a sphere of radius a in flat $\mathbb{R}^{n+1}$, then

$$\mathbf{g}_{S^n(a)} = dr^2 + \left(\frac{1}{\sqrt{K}}\sin\left(\sqrt{K}r\right)\right)^2 \mathbf{g}_{S^{n-1}(1)}$$

where, again, $\mathbf{g}_{S^{n-1}}$ denotes the induced metric on the unit $(n-1)$ sphere in $\mathbb{R}^n$.

Example 1.3. Let $K \in \mathbb{R}$. We define the functions

$$sn_K(r) := \begin{cases} \frac{1}{\sqrt{K}}\sin\left(\sqrt{K}r\right) & K > 0, \\ r & K = 0, \\ \frac{1}{\sqrt{|K|}}\sinh\left(\sqrt{|K|}r\right) & K < 0 \end{cases} \tag{1.1}$$

[2]The majority of the material in this part of the course can be found in the books on Riemannian geometry by Gallot *et al* [GHL] and Chavel [Cha]. Our exposition and notation largely follows that in the first chapter of Cheeger and Ebin [CE], and our treatment of the comparison theorems is partly based on some sections of Petersen [P].

where $r \in [0, \pi/\sqrt{K}]$ for $K > 0$, and $r \in [0, \infty)$ for $K \leq 0$. We then define the metric[3]

$$\mathbf{g}_K := dr^2 + sn_K(r)^2 \mathbf{g}_{S^{n-1}}. \tag{1.2}$$

These metrics have the property that they are of *constant curvature*. In particular, the $(0, 4)$ form of the curvature tensor takes the form

$$\mathbf{R}(\mathbf{W}, \mathbf{X}, \mathbf{Y}, \mathbf{Z}) = K \left(\mathbf{g}_K(\mathbf{W}, \mathbf{Y}) \mathbf{g}_K(\mathbf{X}, \mathbf{Z}) - \mathbf{g}_K(\mathbf{W}, \mathbf{Z}) \mathbf{g}_K(\mathbf{X}, \mathbf{Y}) \right). \tag{1.3}$$

For later use, we note that (1.3) implies that the Ricci tensor of the metric $\mathbf{g}_K$ takes the form

$$\mathbf{Ric}_{\mathbf{g}_K} = K \, (n-1) \, \mathbf{g}_K.$$

Although we will not use the concept of sectional curvature extensively, we introduce it here for the sake of completeness. Given a point $p \in M$, and a two-dimensional plane $\sigma \subseteq T_pM$, we define the *sectional curvature* of σ at p to be

$$K_p(\sigma) := \frac{\mathbf{R}_p(\mathbf{X}, \mathbf{Y}, \mathbf{X}, \mathbf{Y})}{\mathbf{g}_p(\mathbf{X}, \mathbf{X}) \mathbf{g}_p(\mathbf{Y}, \mathbf{Y}) - \mathbf{g}_p(\mathbf{X}, \mathbf{Y})^2},$$

where $\mathbf{X}, \mathbf{Y} \in T_pM$ are any two vectors at p that span the plane σ. (It is straightforward to show that $K_p(\sigma)$ depends only on the plane $\sigma \subseteq T_pM$, and is independent of the vectors $\mathbf{X}, \mathbf{Y}$ chosen to span σ in the above expression.) Although, in general, the sectional curvature will depend on the point p and the plane $\sigma \subseteq T_pM$, in the case of our metrics $\mathbf{g}_K$, it follows from (1.3) that $K_p(\sigma) = K$ for all $p \in M$ and all two-planes $\sigma \subseteq T_pM$.

A consequence of the Gauss Lemma (see later) is that, given any Riemannian metric, we can find local coordinates where the metric takes the form

$$\mathbf{g} = dr^2 + \mathbf{g}_{n-1},$$

with $\mathbf{g}_{n-1}$ is a metric on the $(n-1)$ sphere, depending on (r, θ), where θ are coordinates on S^{n-1}. One of the topics that we will investigate is that if we can put bounds on the curvature of such a metric on a manifold, then we can compare geometrical properties of the corresponding manifold (e.g. diameter, volume of metric balls) with those of the model metrics defined in equations (1.1) and (1.2).

2. Metric space structure on Riemannian manifold

Given a vector field $\mathbf{X} \in \mathfrak{X}(M)$, we define the function

$$|\mathbf{X}| := \sqrt{\mathbf{g}(\mathbf{X}, \mathbf{X})} \in C(M),$$

which is smooth if $\mathbf{X} \neq 0$.

A *curve* in M is a smooth map $c : [a, b] \to M$. The *tangent vector* to c is the vector field along c (see Remark 2.2) defined by

$$c'(t) := D_t c \left(\frac{\partial}{\partial t} \right) \in T_{c(t)}M, \qquad t \in [a, b].$$

A curve is *regular* if $|c'(t)| \neq 0$ (equivalently, $c'(t) \neq 0$), for all $t \in [a, b]$. The *length* (or *arc-length*) of the curve c is given by

$$L[c] := \int_a^b |c'(t)| dt = \int_a^b \sqrt{\mathbf{g}(c'(t), c'(t))} dt.$$

Remark 2.1. We have restricted ourselves to the case of smooth curves. It is straightforward to generalise our discussion to the case of, for example, piecewise smooth curves, at the expense of introducing extra terms into the first and second variation formulae, which we will derive later.

Remark 2.2. Note that c' is not a vector field on M, so objects such as, for example, $\nabla_{c'} c'$ are not really well-defined on M. For this reason, we introduce the following concept.

[3]If you find the $\mathbf{g}_{S^{n-1}}$ a little intimidating, just think of the case with $n = 2$, where $\mathbf{g}_{S^1} = d\theta^2$ with $\theta \in [0, 2\pi)$.

Definition 2.3. Let M, N be smooth manifolds and $\varphi : N \to M$ be a smooth map. A *vector field along* φ is a map $\mathbf{X} : N \to TM$, such that $\mathbf{X}(p) \in T_{\varphi(p)}M$, for all $p \in N$. The space of vector fields along φ will be denoted $\Gamma(\varphi, TM)$.

In the case where M has an affine connection, ∇, we may define an induced covariant derivative, $\widetilde{\nabla} : \mathfrak{X}(N) \times \Gamma(\varphi, TM) \to \Gamma(\varphi, TM)$ (see Differential Geometry II), which allows us to define objects such as $\widetilde{\nabla}_{c'} c'$. We will implicitly use this concept when discussing geodesics and variation of arc-length but, for convenience, we will suppress the notation $\widetilde{\nabla}$ and proceed as if vector fields along φ were actually defined on M.

Remark 2.4. It follows from the Chain Rule that $L[c]$ does not depend on the choice of parametrisation of the curve c. In particular, if we have a regular curve $c : [a, b] \to M$, and a C^1 map $\gamma : [a, b] \to [a', b']$ with $\gamma' > 0$ then, defining $\tilde{c} := c \circ \gamma^{-1} : [a', b'] \to M$, we find that

$$L[\tilde{c}] = \int_{a'}^{b'} |\tilde{c}'(s)| ds = \int_a^b |c'(t)| dt = L[c].$$

Given a regular curve c, we may use this invariance to find a reparametrisation with the property that $\tilde{c} := c \circ \gamma^{-1} : [a', b'] \to M$ has the property that $|\tilde{c}'| = \text{constant}$. Such a curve is said to be *parametrised proportional to arc-length.*

Definition 2.5. The *distance function* $d : M \times M \to \mathbb{R}$ is given by

$$(p, q) \mapsto d(p, q) := \inf \left\{ L[c] \,\middle|\, c : [a, b] \to M \text{ smooth with } c(a) = p, c(b) = q \right\}.$$

Proposition 2.6. *Let $(M, \mathbf{g})$ be a connected Riemannian manifold. Then the distance function $d : M \times M \to \mathbb{R}$ defines a metric on M.*

Proof. See Differential Geometry II. □

Since we now know that a Riemannian manifold is automatically a metric space, we may define such concepts as convergence, Cauchy sequences, and completeness as for general metric spaces. One of our interests will be how geometric properties of Riemannian manifolds are related to their metric space structures. For example, the Hopf-Rinow theorem relates completeness as a Riemannian manifold with completeness as a metric space.

3. First variation of arc-length

Definition 3.1. Let $c : [a, b] \to M$ be a smooth curve. A *smooth variation of* c is a smooth map $\alpha : [a, b] \times (-\epsilon, \epsilon) \to M$, for some $\epsilon > 0$, with the property that $\alpha|\,[a, b] \times \{0\} = c : [a, b] \to M$.

Lecture 3

Let (t, s) be coordinates on $Q := [a, b] \times (-\epsilon, \epsilon)$. We wish to calculate how the length of the curves $c_s : [a, b] \to M$ defined by $c_s := \alpha|\,[a, b] \times \{s\}$ varies with $s \in (-\epsilon, \epsilon)$. To this end, we define the vector fields along α

$$\mathbf{T} := D\alpha \left(\frac{\partial}{\partial t} \right), \qquad \mathbf{V} := D\alpha \left(\frac{\partial}{\partial s} \right).$$

In this case, the vector field (along c_s) $\mathbf{T}(t, s)$ is the tangent vector to the curves c_s, while $\mathbf{V}$ is called the *variation vector field* of α. We also define the vector fields along c:

$$\mathbf{t} := Dc \left(\frac{\partial}{\partial t} \right) = \mathbf{T}|_{s=0}, \qquad \mathbf{v} := Dc \left(\frac{\partial}{\partial s} \right) = \mathbf{V}|_{s=0}.$$

We will assume that the curve c is parametrised proportional to arc-length, with $|c'(t)| = |\mathbf{t}(t)| = l$, for all $t \in [a, b]$, where $l > 0$ is a constant.

Lemma 3.2. *Let $c : [a, b] \to M$ be a smooth curve, parametrised such that $|c'(t)| = l > 0$ for $t \in [a, b]$. If $\alpha : [a, b] \times (-\epsilon, \epsilon) \to M$ is a smooth variation of c then*

$$\frac{d}{ds} L[c_s] \bigg|_{s=0} = \frac{1}{l} \left[\mathbf{g}(\mathbf{t}, \mathbf{v}) \Big|_a^b - \int_a^b \mathbf{g}(\nabla_{\mathbf{t}} \mathbf{t}, \mathbf{v})\, dt \right]. \tag{3.1}$$

This expression is called the first variation formula.

Proof. First note that (t, s) are coordinates on $[a, b] \times (-\epsilon, \epsilon)$, so $[\partial_t, \partial_s] = 0$ on $[a, b] \times (-\epsilon, \epsilon)$. Therefore, $D\alpha\left([\partial_t, \partial_s]\right) = [D\alpha(\partial_t), D\alpha(\partial_s)] = [\mathbf{T}, \mathbf{V}] = 0$. Hence $\nabla_{\mathbf{T}}\mathbf{V} = \nabla_{\mathbf{V}}\mathbf{T}$. Therefore,

$$\begin{aligned}\frac{d}{ds}L[c_s] &= \frac{d}{ds}\int_a^b |c_s'|\, dt = \int_a^b \mathbf{V}\left(\mathbf{g}(\mathbf{T}, \mathbf{T})^{1/2}\right) dt = \int_a^b \nabla_{\mathbf{V}}\left(\mathbf{g}(\mathbf{T}, \mathbf{T})^{1/2}\right) dt \\ &= \int_a^b \frac{1}{2|\mathbf{T}|}\nabla_{\mathbf{V}}\left(\mathbf{g}(\mathbf{T}, \mathbf{T})\right) dt = \int_a^b \frac{1}{|\mathbf{T}|}\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{V}}\mathbf{T}) dt = \int_a^b \frac{1}{|\mathbf{T}|}\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{T}}\mathbf{V}) dt \\ &= \int_a^b \mathbf{T}\left(\frac{1}{|\mathbf{T}|}\mathbf{g}(\mathbf{T}, \mathbf{V})\right) dt - \int_a^b \left[\mathbf{T}\left(\frac{1}{|\mathbf{T}|}\right)\mathbf{g}(\mathbf{T}, \mathbf{V}) + \frac{1}{|\mathbf{T}|}\mathbf{g}(\nabla_{\mathbf{T}}\mathbf{T}, \mathbf{V})\right] dt.\end{aligned}$$

Setting $s = 0$, integrating the first term and using the fact that $|\mathbf{t}| = l$ is constant, we deduce that

$$\frac{d}{ds}\int_a^b |c_s'|\, dt\bigg|_{s=0} = \left(\frac{1}{l}\mathbf{g}(\mathbf{t}, \mathbf{v})\right)\bigg|_a^b - \frac{1}{l}\int_a^b \mathbf{g}\left(\nabla_{\mathbf{t}}\mathbf{t}, \mathbf{v}\right) dt$$

as required. □

Remark 3.3. If the variation α has fixed endpoints (i.e. $\alpha(a, s) = p, \alpha(b, s) = q$ for all $s \in (-\epsilon, \epsilon)$) then $\mathbf{v}(a) = \mathbf{v}(b) = 0$, and the first term in equation (3.1) vanishes.

Definition 3.4. Let M be a manifold with affine connection ∇. A curve $c : [a, b] \to M$ is a *geodesic* if its tangent vector $c' := Dc\left(\frac{\partial}{\partial t}\right) \in \Gamma(c, TM)$ satisfies

$$\nabla_{c'}c' = 0.$$

In the case where $(M, \mathbf{g})$ is a semi-Riemannian manifold, and c is a geodesic (with respect to the Levi-Civita connection), then

$$\nabla_{c'}\left(\mathbf{g}(c', c')\right) = 2\mathbf{g}(c', \nabla_{c'}c') = 0,$$

therefore $\mathbf{g}(c', c')$, and hence $|c'|$, is constant along c. An affinely parametrised geodesic with $|c'| = 1$ is said to be *parametrised by arc-length.*

Corollary 3.5. *A smooth curve segment c of constant speed $|c'| = l > 0$ is a critical point of the arc-length functional under fixed end-point variations if and only if it is a geodesic.*

Proof. In the case where the variation has fixed endpoints, the first variation formula becomes

$$\frac{d}{ds}L[c_s]\bigg|_{s=0} = -\frac{1}{l}\int_a^b \mathbf{g}(\nabla_{\mathbf{t}}\mathbf{t}, \mathbf{v})\, dt. \tag{3.2}$$

If c is a geodesic then $\nabla_{\mathbf{t}}\mathbf{t} = 0$, so $\frac{d}{ds}L[c_s]|_{s=0} = 0$.

The non-trivial direction is to assume $\frac{d}{ds}L[c_s]\big|_{s=0} = 0$ for any variation with fixed endpoints. In particular, the expression on the right-hand-side of (3.2) must vanish for any vector field along c, $\mathbf{v}$, that is generated by a smooth variation and has $\mathbf{v}(a) = \mathbf{v}(b) = 0$. If we assume, for the moment, that any vector field $\mathbf{v} \in \Gamma(c, TM)$ can arise from a smooth variation, then let $\mathbf{v}(t) = f(t)\left(\nabla_{\mathbf{t}}\mathbf{t}\right)(t)$, for f a smooth function with $f(a) = f(b) = 0$. We then have

$$\frac{d}{ds}L[c_s]\bigg|_{s=0} = -\frac{1}{l}\int_a^b f(t)|\nabla_{\mathbf{t}}\mathbf{t}|^2\, dt.$$

In order that the right-hand-side vanish for any function f, we require $\nabla_{\mathbf{t}}\mathbf{t} = 0$ on (a, b) i.e. c is a geodesic.

It remains to be shown that any vector field along c, $\mathbf{v}$, arises from a smooth variation of c. However, given $\mathbf{v}$, we simply define $\alpha(t, s) := \exp_{c(t)}\left(s\mathbf{v}(t)\right)$ (the exponential map will be discussed in the Section 4), and check that $D_{(t,0)}\alpha\left(\frac{\partial}{\partial s}\right) = \mathbf{v}(t)$. □

As mentioned earlier, we could have considered piecewise smooth curves instead of smooth curves. We then have the more general result:

Proposition 3.6. *A piecewise smooth curve segment c of constant speed $|c'| = l > 0$ is a critical point of the arc-length functional under fixed end-point variations if and only if it is a geodesic.*

Proposition 3.7. *Let N be a submanifold of M without boundary and $p \in M$ ($p \notin N$). Let $c : [a,b] \to M$ be a geodesic such that $c(a) \in N$, $c(b) = p$, and such that c is a shortest curve from N to p. Then $c'(a)$ is perpendicular to $T_{c(a)}N$ (i.e. $\mathbf{g}_{c(a)}(c'(a), \mathbf{v}) = 0$, for all $\mathbf{v} \in T_{c(a)}N$.)*

Proof. If $c'(a)$ is not perpendicular to $T_{c(a)}N$, then let $x \in T_{c(a)}N$ be such that $\mathbf{g}_{c(a)}(c'(a), \mathbf{x}) > 0$. Let γ be a curve in N with $\gamma(0) = c(a)$, $\gamma'(0) = x$. Then construct any variation $\alpha : [a,b] \times (-\epsilon, \epsilon) \to M$ such that

$$\begin{aligned} \alpha| \, [a,b] \times \{0\} &= c : [a,b] \to M; \\ \alpha(a,s) &= \gamma(s), \qquad s \in (-\epsilon,\epsilon); \\ \alpha(b,s) &= p, \qquad s \in (-\epsilon,\epsilon). \end{aligned}$$

Let $c_s := \alpha| \, [a,b] \times \{s\} : [a,b] \to M$. Then the first variation formula (3.1) implies that

$$\frac{d}{ds} L[c_s]\bigg|_{s=0} = -\frac{1}{l}\mathbf{g}(c'(a), \mathbf{x}) < 0.$$

Hence, for sufficiently small $s > 0$, we have $L[c_s] < L[c]$. This contradicts the fact that c is the shortest curve from N to p. □

Remark 3.8. Note that the curves c_s that we construct in the proof of the previous proposition need not be geodesics. As long as they are smooth, our argument shows that $L[c_s] < L[c]$.

4. The exponential map and normal coordinates

Proposition 4.1. *Let $p \in M$ and $\mathbf{v} \in T_pM$. Then there exists a unique geodesic $\gamma_{\mathbf{v}}$ in M such that*

1. *the initial velocity of $\gamma_{\mathbf{v}}$ is $\mathbf{v}$; i.e. $\frac{d}{dt}\gamma_{\mathbf{v}}(t)\big|_{t=0} = \mathbf{v}$;*
2. *the domain $I_{\mathbf{v}}$ of $\gamma_{\mathbf{v}}$ is the largest possible. Hence if $\alpha : J \to M$ is a geodesic with initial velocity $\mathbf{v}$ then $J \subseteq I_{\mathbf{v}}$ and $\alpha = \gamma_{\mathbf{v}}\big|_J$.*

The geodesic $\gamma_{\mathbf{v}}$ is said to be maximal *or* geodesically inextendible.

Proof. See Differential Geometry II for details. The main idea is to write the geodesic equations in a local coordinate chart (U, φ) on a neighbourhood of p. The geodesic equations then become the collection of ordinary differential equations

$$\frac{d^2}{dt^2}\left(x^i \circ c\right) + \sum_{j,k}\left({\Gamma^i}_{jk} \circ c\right)\left(\frac{d\left(x^j \circ c\right)}{dt}\right)\left(\frac{d\left(x^k \circ c\right)}{dt}\right) = 0,$$

for maps $x^i \circ c : \mathbb{R} \to \mathbb{R}$, $i = 1, \ldots, n$, with $\left(x^i \circ c\right)(0) = 0$, $\frac{d}{dt}\left(x^i \circ c\right)\big|_{t=0} = v^i$, where v^i are the components of $\mathbf{v}$ in this coordinate system. The result then follows from the standard existence and uniqueness results for ODEs. □

Definition 4.2. Let $p \in M$. We define $\mathcal{D}_p \subseteq T_pM$ to be the set of vectors $\mathbf{v} \in T_pM$ such that the inextendible geodesic $\gamma_{\mathbf{v}}$ is defined at least on the interval $[0,1]$. The *exponential map* of M at p is the map **Lecture 4**

$$\exp_p : \mathcal{D}_p \to M, \quad \mathbf{v} \mapsto \gamma_{\mathbf{v}}(1).$$

Proposition 4.3. *Let $(M, \mathbf{g})$ be a Riemannian manifold. For each $p \in M$ there exists a neighbourhood, U, of $0 \in T_pM$ and a neighbourhood, V, of p in M such that the map $\exp_p\big|_U : U \to V$ is a diffeomorphism (i.e. a smooth homeomorphism).*

Let $\{\mathbf{e}_i\}$ be a basis for T_pM, orthonormal with respect to $\mathbf{g}_p$ (i.e. $\mathbf{g}_p(\mathbf{e}_i, \mathbf{e}_j) = \delta_{ij}$, where $\delta_{ij} = 1$ if $i = j$ and 0 otherwise.) We may then define a coordinate system on the neighbourhood V by assigning the point $\exp_p\left(\sum_{i=1}^n x^i \mathbf{e}_i\right)$ the coordinates $(x^1, \ldots, x^n)$. In particular, let $q \in V$. Then $\exp_p^{-1} q \in U$, and therefore may be written in the form $\sum_{i=1}^n x^i \mathbf{e}_i$, for some real numbers $x^1, \ldots, x^n$, which we define to be the coordinates of q. Such coordinates are called *normal coordinates* at p. Note that $x^i : U \to \mathbb{R}$ are functions on $U \subseteq T_pM$. We may also define the maps $y^i := x^i \circ \exp_p^{-1} : V \to \mathbb{R}$ so that, given $q \in V$, the $y^i(q)$ are the coordinates of the point q.

Gauss Lemma. *Let $p \in M$, $\mathbf{x} \in \mathcal{D}_p \setminus \{0\}$. Let $\mathbf{v_x}, \mathbf{w_x} \in T_{\mathbf{x}}(T_pM)$ with $\mathbf{v_x}$ radial. Then*

$$\mathbf{g}_p(\mathbf{v_x}, \mathbf{w_x}) = \mathbf{g}_{\exp_p \mathbf{x}}\left(\left(D_{\mathbf{x}} \exp_p\right)(\mathbf{v_x}), \left(D_{\mathbf{x}} \exp_p\right)(\mathbf{w_x})\right). \tag{4.1}$$

Proof. See Differential Geometry II. □

Remark 4.4. A remark may be in order at this point concerning the left-hand-side of equation (4.1). Since T_pM is a vector space, given any $x \in T_pM$, we may identify the tangent space $T_x(T_pM)$ with T_pM itself. More precisely, let $\mathbf{v} \in T_pM$, then $\frac{d}{dt}(x + t\mathbf{v})\big|_{t=0}$ defines a tangent vector by $\mathbf{v_x} \in T_x(T_pM)$. This map defines an isomorphism $\psi_{\mathbf{x}} : T_pM \to T_{\mathbf{x}}(T_pM)$. We may, in effect, now turn T_pM into a Riemannian manifold, by defining a metric $\mathbf{G} \in \mathcal{T}^0_2(T_pM)$ by

$$\mathbf{G_x}(\mathbf{v_x}, \mathbf{w_x}) = \mathbf{g}_p(\mathbf{v}, \mathbf{w}), \qquad \forall \mathbf{v}, \mathbf{w} \in T_pM.$$

By the left-hand side of (4.1), we really mean $\mathbf{G_x}(\mathbf{v_x}, \mathbf{w_x})$ or, equivalently, $\mathbf{g}_p(\mathbf{v}, \mathbf{w})$.

We can then identify the vector fields $\mathbf{x} \to \psi_{\mathbf{x}}(\mathbf{e}_i)$ on T_pM with the coordinate derivatives $\partial/\partial x^i$. (Proof: Look at their actions as directional derivatives at any point $\mathbf{x} \in T_pM$.) Therefore, we find that

$$\mathbf{G}\left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j}\right) = \delta_{ij}. \tag{4.2}$$

If we now switch to polar coordinates (ρ, ψ^i) on $U \setminus \{0\} \subseteq T_pM$, where $\rho := \sqrt{\sum (x^i)^2}$ and ψ^i are coordinates on the $(n-1)$-spheres $\rho = \textit{constant}$, then we find from (4.2) that

$$\mathbf{G}\left(\frac{\partial}{\partial \rho}, \frac{\partial}{\partial \rho}\right) = 1, \qquad \mathbf{G}\left(\frac{\partial}{\partial \rho}, \frac{\partial}{\partial \psi^i}\right) = 0.$$

If we perform the same coordinate transformation on $V \setminus \{p\}$, with $r := \sqrt{\sum (x^i)^2}$ and θ^i the corresponding coordinates on the spheres $r = \textit{constant}$, then we have

$$\rho = r \circ \exp_p, \qquad \psi^i = \theta^i \circ \exp_p.$$

In particular, $d\rho = \left(\exp_p\right)^* dr$, $d\psi^i = \left(\exp_p\right)^* d\theta^i$. Since $\exp_p : U \to V$ is a diffeomorphism, we therefore have

$$\left(D \exp_p\right)\left(\frac{\partial}{\partial \rho}\right) = \frac{\partial}{\partial r}, \qquad \left(D \exp_p\right)\left(\frac{\partial}{\partial \psi^i}\right) = \frac{\partial}{\partial \theta^i}.$$

Since $\partial/\partial\rho$ is a radial vector field, we may apply the Gauss Lemma to deduce that

$$\mathbf{g}(\partial_r, \partial_r) = \mathbf{G}(\partial_\rho, \partial_\rho) = 1, \qquad \mathbf{g}(\partial_r, \partial_{\theta_i}) = \mathbf{G}(\partial_\rho, \partial_{\psi_i}) = 0.$$

In particular, in terms of these coordinates, the metric takes the form

$$\mathbf{g}|_{V \setminus \{p\}} = dr^2 + \mathbf{g}_{S^{n-1}}(r),$$

on $V \setminus \{p\}$, where $\mathbf{g}_{S^{n-1}}(r)$ is the metric induced, by $\mathbf{g}$, on the hypersurface $S(r) := \{q \in V \setminus \{p\} : r(q) = r\}$.

Definition 4.5. Let $(M, \mathbf{g})$ be a Riemannian manifold, and $f \in C^\infty(M)$. We define the vector field[4] $\operatorname{grad} f \in \mathfrak{X}(M)$, *gradient* of f by

$$\mathbf{g}(\mathbf{X}, \operatorname{grad} f) = \langle \mathbf{X}, df \rangle = \mathbf{X}(f), \qquad \forall \mathbf{X} \in \mathfrak{X}(M).$$

Corollary 4.6 (to Remark 4.4)**.** *On $V \setminus \{p\}$ we have*

$$\operatorname{grad} r = \frac{\partial}{\partial r}.$$

[4]Some people denote this vector field by ∇f. However, given our use of ∇ for connections, this would, perhaps, be slightly confusing.

Proof. Given the form of the metric, $\mathbf{g}|_{V\setminus\{p\}} = dr^2 + \mathbf{g}_{S^{n-1}}(r)$, we deduce that

$$\mathbf{g}\left(\frac{\partial}{\partial r}, \cdot\right) = dr$$

on $V \setminus \{p\}$. Hence, for any vector field, $\mathbf{X}$, on $V \setminus \{p\}$ we have

$$\mathbf{g}(\mathbf{X}, \operatorname{grad} r) = \langle dr, \mathbf{X}\rangle = \mathbf{g}\left(\frac{\partial}{\partial r}, \mathbf{X}\right).$$

Hence, $\operatorname{grad} r = \partial/\partial r$, as required. □

Remark 4.7. Let $q \in V \setminus \{p\}$, then there exists a unique $\mathbf{v} \in T_pM$ such that $q = \exp_p \mathbf{v}$. Moreover, $r(q) = |\mathbf{v}| := \sqrt{\mathbf{g}_p(\mathbf{v}, \mathbf{v})}$. If we consider the geodesic $\gamma_{\mathbf{v}} : [0, 1] \to M$ with $\gamma(0) = p$, $\gamma'(0) = \mathbf{v}$, then $\gamma(1) = q$, and the length of this curve is

$$L[\gamma_{\mathbf{v}}] = \int_0^1 |\gamma'_{\mathbf{v}}(t)|dt = \int_0^1 |\mathbf{v}|dt = |\mathbf{v}|,$$

where the second inequality follows from the fact that $|\gamma'_{\mathbf{v}}|$ is constant, since $\gamma_{\mathbf{v}}$ a geodesic. Hence

$$r(q) = |\mathbf{v}| = L[\gamma_{\mathbf{v}}].$$

5. Hopf-Rinow theorem

Hopf-Rinow theorem. *Let $(M, \mathbf{g})$ be a Riemannian manifold. The following are equivalent:*

(a) *The metric space (M, d) is complete;*
(b) *For some $p \in M$, $\exp_p$ is defined on all of T_pM (i.e. M is geodesically complete at p);*
(c) *For all $p \in M$, $\exp_p$ is defined on all of T_pM (i.e. M is geodesically complete);*
(d) *Every closed, bounded subset of M is compact.*

Proof. See Differential Geometry II. □

Corollary 5.1. *Let $(M, \mathbf{g})$ be a complete, connected Riemannian manifold. Given any two points $p, q \in M$, there exists a geodesic from p to q the length of which is equal to $d(p, q)$.*

Proof. See Differential Geometry II. □

6. Bishop-Gromov comparison theorem

Lecture 5

Let $(M, \mathbf{g})$ be a Riemannian manifold and $f \in C^\infty(M)$. We have already defined the vector field $\operatorname{grad} f \in \mathfrak{X}(M)$ by

$$\mathbf{g}(\mathbf{X}, \operatorname{grad} f) = \langle \mathbf{X}, df\rangle = \mathbf{X}(f), \qquad \forall \mathbf{X} \in \mathfrak{X}(M).$$

The *Hessian* of f is the $(0, 2)$ tensor field on M defined by

$$\operatorname{Hess} f\,(\mathbf{X}, \mathbf{Y}) = \langle \nabla_{\mathbf{X}} df, \mathbf{Y}\rangle, \qquad \forall \mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M).$$

Proposition 6.1.

$$\operatorname{Hess} f\,(\mathbf{X}, \mathbf{Y}) = \operatorname{Hess} f\,(\mathbf{Y}, \mathbf{X}), \qquad \forall \mathbf{X}, \mathbf{Y} \in \mathfrak{X}(M).$$

Proof.

$$\begin{aligned}
\operatorname{Hess} f\,(\mathbf{X}, \mathbf{Y}) - \operatorname{Hess} f\,(\mathbf{Y}, \mathbf{X}) &= \langle \nabla_{\mathbf{X}} df, \mathbf{Y}\rangle - \langle \nabla_{\mathbf{Y}} df, \mathbf{X}\rangle \\
&= \nabla_{\mathbf{X}}\langle df, \mathbf{Y}\rangle - \langle df, \nabla_{\mathbf{X}}\mathbf{Y}\rangle - \nabla_{\mathbf{Y}}\langle df, \mathbf{X}\rangle + \langle df, \nabla_{\mathbf{Y}}\mathbf{X}\rangle \\
&= \mathbf{X}\left(\mathbf{Y}\left(f\right)\right) - \mathbf{Y}\left(\mathbf{X}\left(f\right)\right) + \langle df, [\mathbf{Y}, \mathbf{X}]\rangle \\
&= 0.
\end{aligned}$$

□

We define the $(1,1)$ tensor field $\mathbf{S}$ by

$$\mathbf{S}(\mathbf{X}) := \nabla_{\mathbf{X}}\left(\operatorname{grad} f\right), \qquad \forall \mathbf{X} \in \mathfrak{X}(M),$$

in which case we have

$$\operatorname{Hess} f\left(\mathbf{X}, \mathbf{Y}\right) = \mathbf{g}(\mathbf{S}(\mathbf{X}), \mathbf{Y}) = \mathbf{g}(\mathbf{X}, \mathbf{S}(\mathbf{Y})),$$

so $\mathbf{S}$ is symmetric with respect to the metric $\mathbf{g}$. The trace of the operator $\mathbf{S} : \mathfrak{X}(M) \to \mathfrak{X}(M)$ is the *Laplacian* of f:

$$\Delta f := \operatorname{tr} \mathbf{S} = \sum_{i=1}^{n} \mathbf{g}(\mathbf{e}_i, \mathbf{S}(\mathbf{e}_i)) = \sum_{i=1}^{n} \operatorname{Hess} f(\mathbf{e}_i, \mathbf{e}_i),$$

where $\{\mathbf{e}_i\}_{i=1}^n$ is any orthonormal basis with respect to $\mathbf{g}$.

Definition 6.2. Let $(M, \mathbf{g})$ be a Riemannian manifold. A *distance function* on an open set $U \subseteq M$ is a smooth map $r : U \to \mathbb{R}$ with the property that $|\operatorname{grad} r| = 1$. For a distance function, r, we will denote the vector field $\operatorname{grad} r \in \mathfrak{X}(U)$ by ∂_r.

Example 6.3. Fix a point $p \in M$. Then the function $r(q) := d(p, q)$ is a distance function, for q in a sufficiently small neighbourhood of p. This is closely related to the discussion of geodesic coordinates in Section 4 and, in fact, the function r defined here coincides with that used in geodesic polar coordinates. This is due to the fact (which we have not proved) that for sufficiently small neighbourhoods, U, of $0 \in T_pM$ the geodesic $\gamma_{\mathbf{v}}$ will be a minimising curve between p and $\exp_p \mathbf{v}$, and hence $d(p, \exp_p \mathbf{v}) = |\mathbf{v}| = r(\exp_p \mathbf{v})$.

Proposition 6.4. *Let $U \subseteq M$ be an open set and $r : U \to \mathbb{R}$ be a distance function. Then*

$$\nabla_{\partial_r} \partial_r = 0$$

on U.

Proof. Let $\mathbf{X} \in \mathfrak{X}(U)$, then

$$\mathbf{g}(\nabla_{\partial_r}\partial_r, \mathbf{X}) = \mathbf{g}(\mathbf{S}(\partial_r), \mathbf{X}) = \mathbf{g}(\partial_r, \mathbf{S}(\mathbf{X})) = \mathbf{g}(\partial_r, \nabla_{\mathbf{X}}\partial_r) = \frac{1}{2}\nabla_{\mathbf{X}}\mathbf{g}(\partial_r, \partial_r) = \frac{1}{2}\nabla_{\mathbf{X}}(1) = 0.$$

Non-degeneracy of $\mathbf{g}$ then implies that $\nabla_{\partial_r}\partial_r = 0$, as required. □

Proposition 6.5. *Let r be a distance function on an open set $U \subseteq M$. Then, on U,*

$$\partial_r\left(\Delta r\right) + \frac{1}{n-1}\left(\Delta r\right)^2 \leq \partial_r\left(\Delta r\right) + \operatorname{tr}(\mathbf{S}^2) = -\mathbf{Ric}(\partial_r, \partial_r). \tag{6.1}$$

Proof. Let $p \in U$, and $\{\mathbf{e}_i\}_{i=1}^n$ be an orthonormal basis on a neighbourhood of p. Then

$$\begin{aligned}
\partial_r(\Delta r) &= \nabla_{\partial_r}\left(\sum_{i=1}^n \mathbf{g}(\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r)\right), \\
&= \sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r) + \sum_{i=1}^n \mathbf{g}(\mathbf{e}_i, \nabla_{\partial_r}\nabla_{\mathbf{e}_i}\partial_r) \\
&= \sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r) + \sum_{i=1}^n \mathbf{g}\big(\mathbf{e}_i, \nabla_{\mathbf{e}_i}\nabla_{\partial_r}\partial_r + \nabla_{[\partial_r,\mathbf{e}_i]}\partial_r + \mathbf{R}(\partial_r,\mathbf{e}_i)\partial_r\big) \\
&= \sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r) + 0 + \sum_{i=1}^n \mathbf{g}(\mathbf{e}_i, \mathbf{S}([\partial_r,\mathbf{e}_i])) - \mathbf{Ric}(\partial_r,\partial_r) \\
&= -\mathbf{Ric}(\partial_r,\partial_r) + \sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r) + \sum_{i=1}^n \mathbf{g}(\mathbf{S}(\mathbf{e}_i), [\partial_r,\mathbf{e}_i]) \\
&= -\mathbf{Ric}(\partial_r,\partial_r) + \sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r) + \sum_{i=1}^n \mathbf{g}(\mathbf{S}(\mathbf{e}_i), \nabla_{\partial_r}\mathbf{e}_i - \nabla_{\mathbf{e}_i}\partial_r) \\
&= -\mathbf{Ric}(\partial_r,\partial_r) - \sum_{i=1}^n \mathbf{g}(\mathbf{S}(\mathbf{e}_i), \nabla_{\mathbf{e}_i}\partial_r) + 2\sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r) \\
&= -\mathbf{Ric}(\partial_r,\partial_r) - \sum_{i=1}^n \mathbf{g}(\mathbf{S}(\mathbf{e}_i), \mathbf{S}(\mathbf{e}_i)) + 2\sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r) \\
&= -\mathbf{Ric}(\partial_r,\partial_r) - \sum_{i=1}^n \mathbf{g}\big(\mathbf{e}_i, \mathbf{S}^2(\mathbf{e}_i)\big) + 2\sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r) \\
&= -\mathbf{Ric}(\partial_r,\partial_r) - \operatorname{tr}\mathbf{S}^2 + 2\sum_{i=1}^n \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \nabla_{\mathbf{e}_i}\partial_r). \qquad (6.2)
\end{aligned}$$

To calculate the final term on the right-hand-side, we note that, since $\nabla_{\partial_r}\mathbf{e}_i \in \mathfrak{X}(U)$, there exist functions ϕ_{ij} on U such that

$$\nabla_{\partial_r}\mathbf{e}_i = \sum_{j=1}^n \phi_{ij}\mathbf{e}_j.$$

Since $\{\mathbf{e}_i\}_{i=1}^n$ is an orthonormal basis, it follows that $\mathbf{g}(\mathbf{e}_i,\mathbf{e}_j)$ are constant, and therefore

$$\begin{aligned}
0 = \nabla_{\partial_r}(\mathbf{g}(\mathbf{e}_i,\mathbf{e}_j)) &= \mathbf{g}(\nabla_{\partial_r}\mathbf{e}_i, \mathbf{e}_j) + \mathbf{g}(\mathbf{e}_i, \nabla_{\partial_r}\mathbf{e}_j) \\
&= \mathbf{g}\left(\sum_k \phi_{ik}\mathbf{e}_k, \mathbf{e}_j\right) + \mathbf{g}\left(\mathbf{e}_i, \sum_k \phi_{jk}\mathbf{e}_k\right) = \phi_{ij} + \phi_{ji}.
\end{aligned}$$

Hence $\phi_{ij} = -\phi_{ji}$. Therefore, the final term in (6.2) takes the form

$$\begin{aligned}
2\sum_{i,j=1}^n \phi_{ij}\mathbf{g}(\mathbf{e}_j, \mathbf{S}(\mathbf{e}_i)) &= -2\sum_{i,j=1}^n \phi_{ji}\mathbf{g}(\mathbf{e}_j, \mathbf{S}(\mathbf{e}_i)) && \text{(skew-symmetry of } \phi) \\
&= -2\sum_{i,j=1}^n \phi_{ji}\mathbf{g}(\mathbf{S}(\mathbf{e}_j), \mathbf{e}_i) && (\mathbf{S} \text{ symmetric w.r.t. } \mathbf{g}) \\
&= -2\sum_{i,j=1}^n \phi_{ij}\mathbf{g}(\mathbf{S}(\mathbf{e}_i), \mathbf{e}_j) && \text{(relabelling summation indices)} \\
&= -2\sum_{i,j=1}^n \phi_{ij}\mathbf{g}(\mathbf{e}_j, \mathbf{S}(\mathbf{e}_i)) && \text{(symmetry of } \mathbf{g}).
\end{aligned}$$

Since the first line equals (-1) times the last line, we deduce that $\sum_{i,j=1}^{n} \phi_{ij}\mathbf{g}\left(\mathbf{e}_j, \mathbf{S}(\mathbf{e}_i)\right) = 0$, so (6.2) reduces to

$$\partial_r\left(\Delta r\right) = -\mathbf{Ric}(\partial_r, \partial_r) - \operatorname{tr}\mathbf{S}^2,$$

as required.

To deduce the required inequality in (6.1), we note that $\mathbf{S}$ is symmetric with respect to $\mathbf{g}$ and therefore, at any fixed point $p \in M$, we may choose an orthonormal basis for T_pM in which $\mathbf{S}|_p$ is diagonal, with the diagonal entries being the eigenvalues of $\mathbf{S}|_p$. Noting that $\mathbf{S}(\partial_r) = \nabla_{\partial_r}\partial_r = 0$, we deduce that $\mathbf{S}|_p$ has a zero eigenvalue, and therefore can be put in the form $\operatorname{diag}[0, \lambda_1, \dots, \lambda_{n-1}]$. We define an inner product on $(n-1)\times(n-1)$ matrices by the formula

$$\langle a, b\rangle := \operatorname{tr}\,(ab)\,, \qquad a, b \in \mathbb{R}(n-1).$$

Viewing $\mathbf{S}|_p$ as a $n\times n$ matrix of the form $\begin{pmatrix} 0 & 0 \\ 0 & s\end{pmatrix}$, where $s = \operatorname{diag}[\lambda_1, \dots, \lambda_{n-1}]$, we deduce, from the Cauchy-Schwartz inequality that

$$|\operatorname{tr}\ \mathbf{S}|_p\,| = |\operatorname{tr} s| = |\langle 1_{n-1}, s\rangle| \le \|1\|\cdot\|s\| = \sqrt{n-1}\sqrt{\operatorname{tr}\,(s^2)} = \sqrt{n-1}\sqrt{\operatorname{tr}\,(\,S|_p^2)}.$$

Since p is arbitrary, we deduce that

$$\operatorname{tr}\,(\mathbf{S}^2) \ge \frac{1}{n-1}\left(\operatorname{tr}\mathbf{S}\right)^2 = \frac{1}{n-1}\left(\Delta r\right)^2.$$

The inequality in (6.1) then follows. □

Remark 6.6. An equation, like (6.1), of the general form $\partial_r\Theta + \Theta^2 + f(r) = 0$, with f a fixed smooth function, is referred to as a *Riccati equation*. Equations of this type will be useful in both Riemannian and Lorentzian geometry.

Remark 6.7. More generally, on a Riemannian manifold $(M, \mathbf{g})$ a formula of Bochner asserts that for any smooth function u on M we have the identity

$$\frac{1}{2}\Delta|\operatorname{grad} u|^2 = \operatorname{tr}\mathbf{S}^2 + \mathbf{g}\left(\operatorname{grad} u, \operatorname{grad}\left(\Delta u\right)\right) + \mathbf{Ric}(\operatorname{grad} u, \operatorname{grad} u).$$

The derivation of this formula is similar to that of equation (6.1), but a little more involved.

Lecture 6

We now apply the formula (6.1) to the metric in normal coordinates that we derived earlier. In particular, let $(M, \mathbf{g})$ be a Riemannian manifold and $p \in M$. We know that there exists a neighbourhood, $U \subseteq T_pM$, of $0 \in T_pM$ and a neighbourhood, $V \subseteq M$, of p in M such that $\exp_p : U \to V$ a diffeomorphism, and such that the metric, $\mathbf{g}$, restricted to $V \setminus \{p\}$ takes the form

$$\mathbf{g}|\, V \setminus \{p\} = dr^2 + \mathbf{h}(r, \theta), \tag{6.3}$$

where $r \in (0, R)$, for some $R > 0$, θ denotes dependence on the $(n-1)$-sphere $S(r) := \exp_p\{\mathbf{v} \in T_pM : |\mathbf{v}| = r\}$ $(r < R)$ and $\mathbf{h}(r, \theta)$ is the induced metric on $S(r)$. (In particular, $\mathbf{h}(\cdot, \partial_r) = 0$.) Since $\operatorname{grad} r = \partial_r$ and $\mathbf{g}(\partial_r, \partial_r) = 1$, it follows that r is a distance function on $V \setminus \{p\}$. (Note that, for notational simplicity, we drop the restriction to $V \setminus \{p\}$ on $\mathbf{g}$.)

Aside on integration: In order to define volumes on an oriented manifold, we need to define a volume form i.e. $d\mathbf{vol} \in \Omega^n(M)$. Given a Riemannian metric, $\mathbf{g}$, on M there is a natural volume form associated to it, which we denote by $d\mathbf{vol}_{\mathbf{g}}$, defined as follows. In local coordinates $\{x^1, \dots, x^n\}$, we have the matrix of components of the metric $g_{ij} := \mathbf{g}(\partial/\partial x^i, \partial/\partial x^j)$. We then define

$$d\mathbf{vol}_{\mathbf{g}} := \sqrt{\det\left(g_{ij}\right)}dx^1 \wedge \dots \wedge dx^n.$$

(Note that I will sometimes drop the $\mathbf{g}$ on $d\mathbf{vol}_{\mathbf{g}}$ if it is clear what metric we are working with.) If we wish to integrate a function $f \in C^\infty(M)$ over a subset $S \subseteq M$, which we assume to lie in the subset of M covered by coordinates $\{x^i\}$, then we define

$$\int_S f(x)d\mathbf{vol}_{\mathbf{g}} = \int_S f(x)\sqrt{\det\left(g_{ij}\right)}dx^1 \wedge \dots \wedge dx^n,$$

which we define to be equal to the multiple integral

$$\int_{x(S)} f(x^1, \ldots, x^n)\sqrt{\det(g_{ij})}dx^1 \ldots dx^n$$

over the relevant subset of $\mathbb{R}^n$.

Example 6.8.

(1) Let $(M, \mathbf{g})$ be $\mathbb{R}^n$ with the flat metric. In Cartesian coordinates, we have $\mathbf{g} = \sum_{i=1}^n \left(dx^i\right)^2$, so $\sqrt{\det(g_{ij})} = 1$ and $d\mathbf{vol}_\mathbf{g} = dx^1 \wedge \cdots \wedge dx^n$. In this case, $\int_S f(x)d\mathbf{vol}_\mathbf{g}$ is simply the multiple integral of the function f over the set $S \subseteq \mathbb{R}^n$.

(2) Let $(M, \mathbf{g})$ two-sphere $\mathbb{S}^2$ with the metric induced from embedding S^2 as the unit sphere in flat $\mathbb{R}^3$. In polar coordinates $x^1 = \theta, x^2 = \phi$, we have

$$\mathbf{g} = d\theta^2 + \sin^2\theta d\phi^2.$$

Therefore the components of the metric are

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2\theta \end{pmatrix},$$

so $\sqrt{\det(g_{ij})} = \sin\theta$ and $d\mathbf{vol}_\mathbf{g} = \sin\theta d\theta \wedge d\phi$. In this case, $\int_S f(x)d\mathbf{vol}_\mathbf{g}$ is the double integral

$$\int_S f(\theta, \phi)\sin\theta d\theta d\phi,$$

over the relevant subset $S \subseteq S^2$. A particular case would be the calculation of $\mathrm{Vol}(S^2)$, where we take $f(\theta, \phi) = 1$ and integrate over $S = S^2$, in which case we have

$$\mathrm{Vol}(S^2) = \int_0^{2\pi}\int_0^{\pi} 1 \sin\theta d\theta d\phi = 4\pi.$$

(3) Let $(M, \mathbf{g})$ be $\mathbb{R}^n$ with the flat metric, written in polar coordinates, so

$$\mathbf{g} = dr^2 + r^2\mathbf{g}_{S^{n-1}},$$

where $\mathbf{g}_{S^{n-1}}$ is the standard round metric on the $(n-1)$-sphere of radius 1. If we adopt coordinates $x^1 = r$, $x^2 = \theta^1, \ldots, x^n = \theta^{n-1}$, where $\theta^1, \ldots, \theta^{n-1}$ are coordinates on S^{n-1}, then the metric takes the local form

$$\mathbf{g} = dr^2 + r^2\sum_{i,j=1}^{n-1} a_{ij}(\theta^1, \ldots, \theta^{n-1})d\theta^i \otimes d\theta^j,$$

for some functions a_{ij}. In this case, the matrix of components of the metric $\mathbf{g}$ decomposes in block form

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 a \end{pmatrix},$$

and we have

$$\sqrt{\det(g_{ij})} = r^{n-1}\sqrt{\det(a_{ij})}.$$

Since $\sqrt{\det(a_{ij})}d\theta^1 \wedge \cdots \wedge d\theta^{n-1} = d\mathbf{vol}_{\mathbf{g}_{S^{n-1}}}$, we deduce that

$$d\mathbf{vol}_\mathbf{g} = r^{n-1}dr \wedge d\mathbf{vol}_{\mathbf{g}_{S^{n-1}}}.$$

(4) More generally, if we consider the constant curvature metrics $\mathbf{g}_K$ introduced in Section 1, then an almost identical calculation to that in the previous example shows that the volume element takes the form

$$d\mathbf{vol}_\mathbf{g} = sn_K(r)^{n-1}dr \wedge d\mathbf{vol}_{\mathbf{g}_{S^{n-1}}}.$$

Consider our general metric in normal coordinates centred at $p \in M$ of the form given in (6.3), where $\mathbf{h}(r, \theta)$ is a general metric on the $(n-1)$-sphere. In this case, a similar calculation to that in Example 3 implies that

$$\begin{aligned} d\mathbf{vol}_{\mathbf{g}} &= dr \wedge d\mathbf{vol}_{\mathbf{h}(r,\theta)} = \sqrt{\det(h_{ij})} dr \wedge d\theta^1 \wedge \cdots \wedge d\theta^n \\ &= \frac{\sqrt{\det(\mathbf{h})}}{\sqrt{\det(\mathbf{g}_{S^{n-1}})}} dr \wedge d\mathbf{vol}_{\mathbf{g}_{S^{n-1}}}. \end{aligned}$$

Therefore, if we define a function $a(r, \theta)$ by

$$a(r, \theta)^{n-1} = \frac{\sqrt{\det(\mathbf{h})}}{\sqrt{\det(\mathbf{g}_{S^{n-1}})}},$$

then the volume element of the metric (6.3) takes the form

$$d\mathbf{vol} = a(r, \theta)^{n-1} dr \wedge d\mathbf{vol}_{S^{n-1}}, \tag{6.4}$$

where $d\mathbf{vol}_{S^{n-1}}$ denotes the volume element of the standard sphere of radius 1.

As $r \to 0$, we require that the metric (6.3) approaches the flat metric on $\mathbb{R}^n$ (so that the metric $\mathbf{g}$ is not singular at $r = 0$), and therefore

$$\mathbf{g} \to dr^2 + r^2 \mathbf{g}_{S^{n-1}} \text{ as } r \to 0.$$

In terms of the volume element, this implies that

$$d\mathbf{vol} \to r^{n-1} dr \wedge d\mathbf{vol}_{S^{n-1}} \text{ as } r \to 0.$$

In particular,

$$a(r, \theta) \to r \text{ as } r \to 0,$$

which implies that the function a obeys the boundary conditions

$$a(0, \theta) = 0, \qquad \frac{\partial a}{\partial r}(0, \theta) = 1.$$

Finally, in order to apply equation (6.1), we need to calculate Δr in the above metric. In local coordinates (r, θ) we have

$$\Delta r = \sum_{i,j} g^{ij} \left(\frac{\partial^2 r}{\partial x^i \partial x^j} - \sum_k \Gamma^k{}_{ij} \frac{\partial r}{\partial x^k} \right)$$

where $\Gamma^i{}_{jk}$ denote the Christoffel coefficients of the metric $\mathbf{g}$. Since $x^1 = r$, the first term vanishes, and only the $k = 1$ term survives in the second term, so we have

$$\begin{aligned} \Delta r &= -\sum_{i,j} g^{ij} \Gamma^r{}_{ij} = -\frac{1}{2} \sum_{i,j,k} g^{ij} g^{rk} \left[\partial_i g_{kj} + \partial_j g_{ki} - \partial_k g_{ij} \right] \\ &= -\frac{1}{2} \sum_{i,j,k} g^{ij} g^{rr} \left[\partial_i g_{rj} + \partial_j g_{ri} - \partial_r g_{ij} \right] \\ &= \frac{1}{2} \sum_{i,j} g^{ij} \partial_r g_{ij} \equiv \frac{1}{2} \mathrm{tr} \left(g^{-1} \partial_r g \right) = \frac{1}{2} \frac{\partial_r (\det g)}{\det g} = \frac{\partial_r (\det g)^{1/2}}{(\det g)^{1/2}}. \end{aligned}$$

Here, we have used the fact that $g^{rr} = 1$, $\partial_i g_{rj} = 0$ for all i, j and the fact that, given any invertible, symmetric matrix $\Lambda(x)$, we have $\mathrm{tr}\,(\Lambda^{-1} \frac{d\Lambda}{dx}) = \frac{d \det \Lambda}{dx} / \det \Lambda$. (If you didn't know this last fact, it can easily be deduced by fixing a basis in which Λ is diagonal.) Since $d\mathbf{vol} = (\det g)^{1/2} d^n x$, we deduce from (6.4) that $(\det g)^{1/2} = a(r, \theta)^{n-1} \cdot (\det h)^{1/2}$, where $\det h$ is the determinant of the metric on S^{n-1}. Since $\det h$ is independent of r, we therefore have

$$\Delta r = \frac{\partial_r (a(r, \theta))^{n-1}}{(a(r, \theta))^{n-1}} = \frac{(n-1)}{a} \frac{\partial a}{\partial r}. \tag{6.5}$$

Bishop-Gromov Comparison Theorem (Version 1). *Let $(M, \mathbf{g})$ be a complete Riemannian manifold with the property that there exists a constant $K \in \mathbb{R}$ such that*

$$\mathbf{Ric}(\mathbf{X}, \mathbf{X}) \geq K\,(n-1)\,\mathbf{g}(\mathbf{X}, \mathbf{X}), \qquad \forall \mathbf{X} \in \mathfrak{X}(M). \tag{6.6}$$

Let $p \in M$ and $B(p,r) := \exp_p(\{\mathbf{v} \in T_pM : |\mathbf{v}| < r\})$, where we restrict to $r \in [0, R)$ such that $B(p,R) \subseteq V$. Then the map

$$r \mapsto \frac{\mathrm{Vol}\,(B(p,r), \mathbf{g})}{\mathrm{Vol}\,(B(p,r), \mathbf{g}_K)}$$

is non-increasing, where $\mathbf{g}_K$ denotes the metric of constant curvature K (see Section 1). In particular,

$$\mathrm{Vol}\,(B(p,r), \mathbf{g}) \leq \mathrm{Vol}\,(B(p,r), \mathbf{g}_K)$$

for $r \in [0, R)$.

Remark 6.9. The statement that $\mathrm{Vol}\,(B(p,r), \mathbf{g}) \leq \mathrm{Vol}\,(B(p,r), \mathbf{g}_K)$ is the original result proved by Bishop, who proved this result for r less than the injectivity radius[5] of the manifold $(M, \mathbf{g})$. Gromov [Gr] later generalised this result to show that the map $r \to \mathrm{Vol}\,(B(p,r), \mathbf{g})/\mathrm{Vol}\,(B(p,r), \mathbf{g}_K)$ is non-increasing and also noted that, with suitable modifications, the result holds for all values of r. (We will extend the following proof to cover Gromov's more general result after we have studied the second variation formula and conjugate points etc.)

Lecture 7

Proof of Theorem 6. Let the metric $\mathbf{g}$ satisfy the curvature condition (6.6). The function r is a distance function on the set V, and therefore satisfies

$$\partial_r\,(\Delta r) + \frac{1}{n-1}\,(\Delta r)^2 \leq -K(n-1).$$

Given that $\Delta r = (n-1)a'/a$ (from now on, $'$ will denote partial differentiation with respect to r) we deduce that a satisfies

$$\frac{1}{a}\frac{\partial^2 a}{\partial r^2} \leq -K.$$

Since $a(r,\theta) \sim r$, for sufficiently small r, it follows that $a(r,\theta) > 0$, for sufficiently small $r > 0$. Therefore,

$$\frac{\partial^2 a(r,\theta)}{\partial r^2} + Ka(r,\theta) \leq 0.$$

We now wish to use this equation to compare the properties of the metric $\mathbf{g}$ with those of the constant curvature metric $\mathbf{g}_K$. To do this, we compare the function a with the function, $sn_K(r)$, defined in equation (1.1) of Section (1). This function satisfies

$$\frac{\partial^2 sn_K(r)}{\partial r^2} + K sn_K(r) = 0, \qquad sn_K(0) = 0, \qquad sn'_K(0) = 1.$$

We then have

$$\begin{aligned}\partial_r\,(sn_K(r)a'(r) - sn'_K(r)a(r)) &= sn_K(r)a''(r) - sn''_K(r)a(r)\\ &\leq -K sn_K(r)a(r) + K sn_K(r)a(r)\\ &= 0.\end{aligned}$$

Hence $sn_K(r)a'(r) - sn'_K(r)a(r)$ is non-increasing. Since, evaluated at $r = 0$, it equals zero, we deduce that

$$sn_K(r)a'(r) - sn'_K(r)a(r) \leq 0.$$

Since sn_K, a are positive for sufficiently small $r > 0$, we deduce that

$$\left(\frac{a(r)}{sn_K(r)}\right)' \leq 0. \tag{6.7}$$

We therefore deduce that

$$\text{the map } r \mapsto \frac{a(r)}{sn_K(r)} \text{ is non-increasing.} \tag{6.8}$$

[5]This term will be defined later.

Moreover, $a(r), sn_K(r) \sim r$ as $r \to 0$, so $\lim_{r\to 0}(a(r)/sn_K(r)) = 1$. Therefore

$$\frac{a(r)}{sn_K(r)} \leq 1 \text{ for } r > 0.$$

Finally, we consider the volumes

$$\begin{aligned}
\mathrm{Vol}\,(B(p,r),\mathbf{g}) &:= \int_0^r \int_{S^{n-1}} 1 d\mathbf{vol}_{\mathbf{g}} = \int_0^r \int_{S^{n-1}} a(t,\theta)^{n-1} dt \wedge d\mathbf{vol}_{\mathbf{g}_{S^{n-1}}} \\
\mathrm{Vol}\,(B(p,r),\mathbf{g}_K) &:= \int_0^r \int_{S^{n-1}} 1 d\mathbf{vol}_{\mathbf{g}_K} = \int_0^r \int_{S^{n-1}} sn_K(t)^{n-1} dt \wedge d\mathbf{vol}_{\mathbf{g}_{S^{n-1}}} \\
&= \mathrm{Vol}\,(S^{n-1}) \int_0^r sn_K(t)^{n-1} dt
\end{aligned}$$

and the ratio (abbreviating $d\mathbf{vol}_{\mathbf{g}_{S^{n-1}}}$ with just $d\theta$)

$$\frac{\mathrm{Vol}\,(B(p,r),\mathbf{g})}{\mathrm{Vol}\,(B(p,r),\mathbf{g}_K)} = \frac{\int_0^r \int_{S^{n-1}} a(t,\theta)^{n-1} d\theta dt}{\int_0^r \int_{S^{n-1}} sn_K(t)^{n-1} d\theta dt} = \frac{1}{Area(S^{n-1})} \frac{\int_0^r \int_{S^{n-1}} a(t,\theta)^{n-1} d\theta dt}{\int_0^r sn_K(t)^{n-1} dt}.$$

Taking the limit as $r \to 0$, we see that $\mathrm{Vol}\,(B(p,r),\mathbf{g})/\mathrm{Vol}\,(B(p,r),\mathbf{g}_K) \to 1$, since $a(r,\theta), sn_K(r) \sim r$. To prove that the above ratio is non-increasing, we take is derivative:

$$\begin{aligned}
\frac{d}{dr}\left[\frac{\mathrm{Vol}\,(B(p,r),\mathbf{g})}{\mathrm{Vol}\,(B(p,r),\mathbf{g}_K)}\right] &= \frac{1}{\mathrm{Vol}\,(S^{n-1})} \frac{1}{\left(\int_0^r sn_K(t)^{n-1} dt\right)^2} \\
&\quad \times \left[\int_{S^{n-1}} a(r,\theta)^{n-1} d\theta \times \int_0^r sn_K(t)^{n-1} dt - sn_K(r)^{n-1} \int_0^r \int_{S^{n-1}} a(t,\theta)^{n-1} d\theta dt\right] \\
&= \frac{1}{\mathrm{Vol}\,(S^{n-1})} \frac{1}{\left(\int_0^r sn_K(t)^{n-1} dt\right)^2} \int_0^r sn_K(t)^{n-1} sn_K(r)^{n-1} \\
&\quad \times \int_{S^{n-1}} \left[\left(\frac{a(r,\theta)}{sn_K(r)}\right)^{n-1} - \left(\frac{a(t,\theta)}{sn_K(t)}\right)^{n-1}\right] d\theta dt. \qquad (6.9)
\end{aligned}$$

From (6.8), we know that $\frac{a(t,\theta)}{sn_K(t)}$, and hence $\left(\frac{a(t,\theta)}{sn_K(t)}\right)^{n-1}$, is a non-increasing function of t. Hence, in the integral in (6.9), since $0 \leq t \leq r$, we have

$$\left(\frac{a(r,\theta)}{sn_K(r)}\right)^{n-1} - \left(\frac{a(t,\theta)}{sn_K(t)}\right)^{n-1} \leq 0.$$

Since $sn_K(r) \geq 0$ for sufficiently small r, it follows from (6.9) that

$$\frac{d}{dr}\left[\frac{\mathrm{Vol}\,(B(p,r),\mathbf{g})}{\mathrm{Vol}\,(B(p,r),\mathbf{g}_K)}\right] \leq 0.$$

Therefore the ratio of volumes is non-increasing. Since we know that $\mathrm{Vol}\,(B(p,r),\mathbf{g})/\mathrm{Vol}\,(B(p,r),\mathbf{g}_K) \to 1$ as $r \to 0$, we also deduce that $\mathrm{Vol}\,(B(p,r),\mathbf{g}) \leq \mathrm{Vol}\,(B(p,r),\mathbf{g}_K)$.

□

7. Second variation of arc-length

Definition 7.1. Let $c : [a,b] \to M$ be a smooth curve. A *(two parameter) smooth variation of* c is a smooth map $\alpha : [a,b] \times (-\delta,\delta) \times (-\epsilon,\epsilon) \to M$, for some $\epsilon, \delta > 0$, with the property that $\alpha|\,[a,b] \times \{0\} \times \{0\} = c : [a,b] \to M$. For $(r,s) \in (-\delta,\delta) \times (-\epsilon,\epsilon)$, we define the curve $c_{r,s} := \alpha|\,[a,b] \times \{r\} \times \{s\} : [a,b] \to M$.

In this case, taking coordinates (t,r,s) on $[a,b] \times (-\delta,\delta) \times (-\epsilon,\epsilon)$, we define the vector fields along α:

$$\mathbf{T} := D\alpha\left(\frac{\partial}{\partial t}\right), \qquad \mathbf{V} := D\alpha\left(\frac{\partial}{\partial r}\right), \qquad \mathbf{W} := D\alpha\left(\frac{\partial}{\partial s}\right),$$

and the vector fields along c:

$$\mathbf{t} := \mathbf{T}|_{r=s=0}, \qquad \mathbf{v} := \mathbf{V}|_{r=s=0}, \qquad \mathbf{w} := \mathbf{W}|_{r=s=0}.$$

We will assume that the curve c is parametrised proportional to arc-length, with $|c'(t)| = |\mathbf{t}(t)| = l$, for all $t \in [a,b]$, where $l > 0$ is a constant.

Lemma 7.2. *Let $c : [a,b] \to M$ be a geodesic with $|c'| = l > 0$. If α is a variation of c as above then*

$$\left.\frac{\partial^2 L}{\partial r \partial s}\right|_{r=s=0} = \frac{1}{l}\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{w}}\mathbf{v})\Big|_a^b + \frac{1}{l}\int_a^b \left[\mathbf{g}((\nabla_{\mathbf{t}}\mathbf{v})^{\perp}, (\nabla_{\mathbf{t}}\mathbf{w})^{\perp}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{w},\mathbf{t})\mathbf{v})\right] dt, \tag{7.1}$$

where, given any vector field $\mathbf{X}$ along c, we define

$$\mathbf{X}^{\perp} := \mathbf{X} - \frac{\mathbf{g}(\mathbf{X},\mathbf{t})}{|\mathbf{t}|^2}\mathbf{t},$$

the component of $\mathbf{X}$ perpendicular to $\mathbf{t} \equiv c'$. The expression in (7.1) is called the second variation formula.

Proof. As in the proof of the first variation formula, note that $\nabla_{\mathbf{T}}\mathbf{V} = \nabla_{\mathbf{V}}\mathbf{T}$, $\nabla_{\mathbf{T}}\mathbf{W} = \nabla_{\mathbf{W}}\mathbf{T}$ and $\nabla_{\mathbf{V}}\mathbf{W} = \nabla_{\mathbf{W}}\mathbf{V}$. We therefore have

$$\begin{aligned}
\frac{\partial^2}{\partial r \partial s} L[c_{r,s}] &= \frac{\partial^2}{\partial r \partial s}\int_a^b \left|c'_{r,s}\right| dt = \int_a^b \nabla_{\mathbf{V}}\nabla_{\mathbf{W}}\left(\mathbf{g}(\mathbf{T},\mathbf{T})\right)^{1/2} dt = \int_a^b \nabla_{\mathbf{V}}\left(\frac{1}{|\mathbf{T}|}\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{W}}\mathbf{T})\right) dt. \\
&= \int_a^b \nabla_{\mathbf{V}}\left(\frac{1}{|\mathbf{T}|}\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{T}}\mathbf{W})\right) dt. \\
&= \int_a^b \left[-\frac{1}{|\mathbf{T}|^3}\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{V}}\mathbf{T})\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{T}}\mathbf{W}) + \frac{1}{|\mathbf{T}|}\mathbf{g}(\nabla_{\mathbf{V}}\mathbf{T}, \nabla_{\mathbf{T}}\mathbf{W}) + \frac{1}{|\mathbf{T}|}\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{V}}\nabla_{\mathbf{T}}\mathbf{W})\right] dt \\
&= \int_a^b \left[-\frac{1}{|\mathbf{T}|^3}\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{T}}\mathbf{V})\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{T}}\mathbf{W}) + \frac{1}{|\mathbf{T}|}\mathbf{g}(\nabla_{\mathbf{T}}\mathbf{V}, \nabla_{\mathbf{T}}\mathbf{W}) + \frac{1}{|\mathbf{T}|}\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{V}}\nabla_{\mathbf{T}}\mathbf{W})\right] dt.
\end{aligned}$$

Lecture 8

We may rewrite the final term in the integrand using the following

$$\begin{aligned}
\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{V}}\nabla_{\mathbf{T}}\mathbf{W}) &= \mathbf{g}\left(\mathbf{T}, \nabla_{\mathbf{T}}\nabla_{\mathbf{V}}\mathbf{W} + \nabla_{[\mathbf{T},\mathbf{V}]}\mathbf{W} + \mathbf{R}(\mathbf{V},\mathbf{T})\mathbf{W}\right) \\
&= \nabla_{\mathbf{T}}\left(\mathbf{g}(\mathbf{T}, \nabla_{\mathbf{V}}\mathbf{W})\right) - \mathbf{g}(\nabla_{\mathbf{T}}\mathbf{T}, \nabla_{\mathbf{V}}\mathbf{W}) + \mathbf{g}(\mathbf{T}, \mathbf{R}(\mathbf{V},\mathbf{T})\mathbf{W}).
\end{aligned}$$

Setting $r = s = 0$, and using the fact that $\nabla_{\mathbf{t}}\mathbf{t} = 0$ and $|\mathbf{t}| = l$ is constant, we deduce that

$$\begin{aligned}
\left.\frac{\partial^2}{\partial r \partial s} L[c_{r,s}]\right|_{r=s=0} &= \int_a^b \Big[-\frac{1}{l^3}\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{t}}\mathbf{v})\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{t}}\mathbf{w}) + \frac{1}{l}\mathbf{g}(\nabla_{\mathbf{t}}\mathbf{v}, \nabla_{\mathbf{t}}\mathbf{w}) \\
&\qquad\qquad + \frac{1}{l}\left(\nabla_{\mathbf{t}}\left(\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{v}}\mathbf{w})\right) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{v},\mathbf{t})\mathbf{w})\right)\Big] dt \\
&= \frac{1}{l}\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{v}}\mathbf{w})\Big|_a^b + \frac{1}{l}\int_a^b \left[\mathbf{g}((\nabla_{\mathbf{t}}\mathbf{v})^{\perp}, (\nabla_{\mathbf{t}}\mathbf{w})^{\perp}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{v},\mathbf{t})\mathbf{w})\right] dt.
\end{aligned}$$

We now note that

- $\nabla_{\mathbf{v}}\mathbf{w} = \nabla_{\mathbf{w}}\mathbf{v}$;
- $\mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{v},\mathbf{t})\mathbf{w}) = \mathbf{R}(\mathbf{v},\mathbf{t},\mathbf{t},\mathbf{w}) = \mathbf{R}(\mathbf{t},\mathbf{w},\mathbf{v},\mathbf{t}) = \mathbf{R}(\mathbf{w},\mathbf{t},\mathbf{t},\mathbf{v}) = \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{w},\mathbf{t})\mathbf{v})$

Therefore,

$$\left.\frac{\partial^2}{\partial r \partial s} L[c_{r,s}]\right|_{r=s=0} = \frac{1}{l}\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{w}}\mathbf{v})\Big|_a^b + \frac{1}{l}\int_a^b \left[\mathbf{g}((\nabla_{\mathbf{t}}\mathbf{v})^{\perp}, (\nabla_{\mathbf{t}}\mathbf{w})^{\perp}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{w},\mathbf{t})\mathbf{v})\right] dt.$$

as required. □

Remarks 7.3.

(1) Note that, by the argument at the end of the preceding proof, it follows that the right-hand-side of (7.1) is symmetric under interchange of $\mathbf{v}$ and $\mathbf{w}$.
(2) The endpoint contribution to the above expression vanishes if $\nabla_{\mathbf{w}}\mathbf{v}$ vanishes at a and b. In particular, if $\mathbf{v}$ and/or $\mathbf{w}$ vanishes at $c(a)$ and $c(b)$, then this term is zero.

Let $c : [a, b] \to M$ be a geodesic, with $|c'(t)| = l$ for $t \in [a, b]$. Let $\Gamma_0(c, TM)$ be the space of vector fields, $\mathbf{X}$, along c such that $\mathbf{X}(a) = \mathbf{X}(b) = 0$ that are orthogonal to c i.e.

$$\Gamma_0(c, TM) := \left\{ \mathbf{X} \in \Gamma(c, TM) \,\middle|\, \begin{array}{l} \mathbf{X}(a) = 0 \\ \mathbf{X}(b) = 0 \\ \mathbf{g}_{c(t)}(\mathbf{X}(t), c'(t)) = 0, \forall t \in [a, b] \end{array} \right\}.$$

Definition 7.4. We define the symmetric bilinear form $I : \Gamma_0(c, TM) \times \Gamma_0(c, TM) \to \mathbb{R}$, called the *index form*, by

$$I(\mathbf{X}, \mathbf{Y}) := \frac{1}{l} \int_a^b \left[\mathbf{g}((\nabla_{\mathbf{t}} \mathbf{X})^{\perp}, (\nabla_{\mathbf{t}} \mathbf{Y})^{\perp}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{X}, \mathbf{t})\mathbf{Y}) \right] dt,$$

for $\mathbf{X}, \mathbf{Y} \in \Gamma_0(c, TM)$.

Remark 7.5. Note that if $\mathbf{g}(\mathbf{X}, \mathbf{t}) = 0$ along c then

$$0 = \nabla_{\mathbf{t}} \left(\mathbf{g}(\mathbf{X}, \mathbf{t}) \right) = \mathbf{g}(\nabla_{\mathbf{t}} \mathbf{X}, \mathbf{t}).$$

Therefore $\nabla_{\mathbf{t}} \mathbf{X}$ is orthogonal to $\mathbf{t}$, so

$$\nabla_{\mathbf{t}} \mathbf{X} = (\nabla_{\mathbf{t}} \mathbf{X})^{\perp}.$$

Therefore, the index form may be written in the simpler form

$$I(\mathbf{X}, \mathbf{Y}) := \frac{1}{l} \int_a^b \left[\mathbf{g}(\nabla_{\mathbf{t}} \mathbf{X}, \nabla_{\mathbf{t}} \mathbf{Y}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{X}, \mathbf{t})\mathbf{Y}) \right] dt, \qquad \mathbf{X}, \mathbf{Y} \in \Gamma_0(c, TM).$$

Proposition 7.6. *Given* $\mathbf{X}, \mathbf{Y} \in \Gamma_0(c, TM)$,

$$I(\mathbf{X}, \mathbf{Y}) := -\frac{1}{l} \int_a^b \mathbf{g}(\mathbf{X}, \nabla_{\mathbf{t}} \nabla_{\mathbf{t}} \mathbf{Y} + \mathbf{R}(\mathbf{Y}, \mathbf{t})\mathbf{t}) \, dt. \tag{7.2}$$

Proof. We have

$$\mathbf{g}(\nabla_{\mathbf{t}} \mathbf{X}, \nabla_{\mathbf{t}} \mathbf{Y}) = \nabla_{\mathbf{t}} \mathbf{g}(\mathbf{X}, \nabla_{\mathbf{t}} \mathbf{Y}) - \mathbf{g}(\mathbf{X}, \nabla_{\mathbf{t}} \nabla_{\mathbf{t}} \mathbf{Y}).$$

Upon integration, this yields

$$\int_a^b \mathbf{g}(\nabla_{\mathbf{t}} \mathbf{X}, \nabla_{\mathbf{t}} \mathbf{Y}) \, dt = \mathbf{g}(\mathbf{X}, \nabla_{\mathbf{t}} \mathbf{Y})\big|_a^b - \int_a^b \mathbf{g}(\mathbf{X}, \nabla_{\mathbf{t}} \nabla_{\mathbf{t}} \mathbf{Y}) \, dt = -\int_a^b \mathbf{g}(\mathbf{X}, \nabla_{\mathbf{t}} \nabla_{\mathbf{t}} \mathbf{Y}) \, dt,$$

since $\mathbf{X}(a) = \mathbf{X}(b) = 0$. Using the symmetries of the $(0, 4)$ version of the Riemann tensor, we have

$$\mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{X}, \mathbf{t})\mathbf{Y}) \equiv \mathbf{R}(\mathbf{X}, \mathbf{t}, \mathbf{t}, \mathbf{Y}) = \mathbf{R}(\mathbf{t}, \mathbf{Y}, \mathbf{X}, \mathbf{t}) = -\mathbf{R}(\mathbf{Y}, \mathbf{t}, \mathbf{X}, \mathbf{t}) \equiv -\mathbf{g}(\mathbf{X}, \mathbf{R}(\mathbf{Y}, \mathbf{t}), \mathbf{t})$$

Hence

$$I(\mathbf{X}, \mathbf{Y}) := -\frac{1}{l} \int_a^b \left[\mathbf{g}(\mathbf{X}, \nabla_{\mathbf{t}} \nabla_{\mathbf{t}} \mathbf{Y}) + \mathbf{g}(\mathbf{X}, \mathbf{R}(\mathbf{Y}, \mathbf{t})\mathbf{t}) \right] dt,$$

as required. ☐

Corollary 7.7. *The bilinear form* $I : \Gamma_0(c, TM) \times \Gamma_0(c, TM) \to \mathbb{R}$ *is degenerate if and only if there exists* $\mathbf{Y} \in \Gamma_0(c, TM)$, *not identically zero, such that*

$$\nabla_{\mathbf{t}} \nabla_{\mathbf{t}} \mathbf{Y} + \mathbf{R}(\mathbf{Y}, \mathbf{t})\mathbf{t} = 0.$$

Proof. For I to be degenerate implies that there exists $\mathbf{Y} \in \Gamma_0(c, TM)$, not identically zero, such that $I(\mathbf{X}, \mathbf{Y}) = 0$ for all $X \in \Gamma_0(c, TM)$. The result then follows from equation (7.2) by a standard calculus of variations argument. ☐

Definition 7.8. Let $c : [a, b] \to M$ be a geodesic parametrised proportional to arc-length. A vector field along a geodesic c, $\mathbf{J} \in \Gamma(c, TM)$, that satisfies the *Jacobi equation*

$$\nabla_{\mathbf{t}} \nabla_{\mathbf{t}} \mathbf{J} + \mathbf{R}(\mathbf{J}, \mathbf{t})\mathbf{t} = 0$$

is called a *Jacobi field* along c.

Remark 7.9. In local coordinates, $\{x^i\}$, a Jacobi field along a geodesic γ has components $J^i(t)$ that obey the second order ordinary differential equations

$$\frac{d^2}{dt^2}J^i(t)+\sum_{j,k,l}R_{kl}{}^i{}_j(\gamma(t))J^k(t)\frac{d\left(x^l\circ\gamma\right)}{dt}\frac{d\left(x^j\circ\gamma\right)}{dt}=0.$$

where the geodesic γ. It follows that if we fix initial data $\mathbf{J}(0),\mathbf{J}'(0)\in T_{\gamma(0)}M$, then there will exist a unique Jacobi field along γ with this initial data.

Proposition 7.10. *If $c_s:[a,b]\to M$ are geodesics, then $\mathbf{V}$ is a Jacobi field along c_s.*

Proof. If c_s are geodesics, then we have $\nabla_{\mathbf{T}}\mathbf{T}=0$. Hence

$$\nabla_{\mathbf{T}}\nabla_{\mathbf{T}}\mathbf{V}=\nabla_{\mathbf{T}}\nabla_{\mathbf{V}}\mathbf{T}=\nabla_{\mathbf{V}}\nabla_{\mathbf{T}}\mathbf{T}+\mathbf{R}(\mathbf{T},\mathbf{V})\mathbf{T}=\mathbf{R}(\mathbf{T},\mathbf{V})\mathbf{T}.$$

Hence $\nabla_{\mathbf{T}}\nabla_{\mathbf{T}}\mathbf{V}+\mathbf{R}(\mathbf{V},\mathbf{T})\mathbf{T}=0$, so $\mathbf{V}$ is a Jacobi field along c_s. □

Definition 7.11. A point $q\in M$ is *conjugate* to $p\in M$ along a geodesic $c:[0,1]\to M$ with $c(0)=p,c(1)=q$ if there exists a non-zero Jacobi field, $\mathbf{J}$, along c such that $\mathbf{J}(0)=\mathbf{J}(1)=0$.

Proposition 7.12. *A point $q\in M$ is conjugate to $p\in M$ along a geodesic c if and only if q is a singular value of $\exp_p:T_pM\to M$ (i.e. $q=\exp_p\mathbf{t}$ for $\mathbf{t}\in T_pM$ and $D_{\mathbf{t}}\exp_p:T_pM\to T_qM$ has non-trivial kernel). The* order *of a conjugate point is the dimension of the kernel of the $D_{\mathbf{t}}\exp_p:T_{\mathbf{t}}(T_pM)\to T_{\exp_p\mathbf{t}}M$.*

Proof. Let $\mathbf{J}$ be a non-zero Jacobi field along c with $\mathbf{J}(0)=0$ and $\mathbf{J}(1)=0$ and $\mathbf{u}:=\mathbf{J}'(0)\in T_pM$. The smooth variation

$$\alpha(t,s):=\exp_p\left[t\left(\mathbf{t}+s\mathbf{u}\right)\right].$$

of c has variation vector field $\mathbf{v}(t):=D_{(t,s)}\alpha\left(\frac{\partial}{\partial s}\right)\big|_{s=0}$. Since the curves $c_s:[0,1]\to M$ are geodesics, Proposition 7.10 implies that $\mathbf{v}$ is a Jacobi vector field along c. Since $\mathbf{v}(0)=0$ and $\mathbf{v}'(0)=\mathbf{u}=\mathbf{J}'(0)$, it follows that $\mathbf{v}=\mathbf{J}$. We then have

$$\mathbf{J}(1)=D_{(t,s)}\alpha\left(\frac{\partial}{\partial s}\right)\bigg|_{s=0,t=1}=\left(D_{\mathbf{t}}\exp_p\right)\left(\mathbf{J}'(\mathbf{0})\right)=0.$$

Therefore, $D_{\mathbf{t}}\exp_p$ has non-trivial kernel $\mathbf{J}'(\mathbf{0})$.

Reversing the above construction then implies that, given $\mathbf{u}\in T_{\mathbf{t}}(T_pM)$ such that $D_{\mathbf{t}}\exp_p(\mathbf{u})=0$, the Jacobi field $\mathbf{J}$ with $\mathbf{J}(0)=0$, $\mathbf{J}'(0)=\mathbf{u}$ will obey $\mathbf{J}(1)=0$, so p and q are conjugate. □

Proposition 7.13. *Let $c:[a,b]\to M$ be a geodesic, $c'=\mathbf{t}$, and let $\mathbf{J}$ be a Jacobi field that vanishes at $c(a)$ and $c(b)$. Then $\mathbf{g}(\mathbf{t},\mathbf{J})=\mathbf{g}(\mathbf{t},\mathbf{J}')=0$.*

Proof. Note that

$$\mathbf{t}\left(\mathbf{g}(\mathbf{t},\mathbf{J}')\right)=\mathbf{g}(\mathbf{t},\mathbf{J}'')=\mathbf{g}(\mathbf{t},-\mathbf{R}(\mathbf{J},\mathbf{t})\mathbf{t})=0.$$

Therefore $\mathbf{g}(\mathbf{t},\mathbf{J}')$ is constant on c. However, $\mathbf{g}(\mathbf{t},\mathbf{J}')=\mathbf{t}\left(\mathbf{g}(\mathbf{t},\mathbf{J})\right)$, so $\mathbf{g}(\mathbf{t},\mathbf{J})$ (as a function of t) has constant derivative. Since $\mathbf{g}(\mathbf{t},\mathbf{J})|_{c(a)}=\mathbf{g}(\mathbf{t},\mathbf{J})|_{c(b)}=0$, it follows that $\mathbf{g}(\mathbf{t},\mathbf{J})=0$ and, therefore, $\mathbf{g}(\mathbf{t},\mathbf{J}')=0$. □

Lecture 9

Proposition 7.14. *Let $c:[a,b]\to M$ be a geodesic, with conjugate point at $t=t_0\in(a,b)$. Then there exists a vector field along c, $\mathbf{X}$, such that $I(\mathbf{X},\mathbf{X})<0$. (In general, $\mathbf{X}$ is continuous but not differentiable.)*

Proof. Let $\mathbf{Y}$ be a Jacobi field along c with $\mathbf{Y}(a)=\mathbf{Y}(t_0)=0$, but $\mathbf{Y}$ not identically zero. By the previous proposition, $\mathbf{Y}$ and $\nabla_{\mathbf{t}}\mathbf{Y}$ are orthogonal to $\mathbf{t}$. We then define the vector field along c:

$$\widetilde{\mathbf{Y}}(t):=\begin{cases}\mathbf{Y}(t) & t\in[a,t_0],\\ 0 & t\in[t_0,b].\end{cases}$$

Then $\widetilde{\mathbf{Y}}\in\Gamma_0(c,TM)$ and $I(\widetilde{\mathbf{Y}},\widetilde{\mathbf{Y}})=0$. Now let $\mathbf{Z}\in\Gamma_0(c,TM)$ be the vector field along c defined by the conditions

$$\mathbf{Z}(t_0)=-\left(\nabla_{\mathbf{t}}\mathbf{Y}\right)(t_0),\qquad\nabla_{\mathbf{t}}\mathbf{Z}=0.$$

Now let $\varphi : [a,b] \to \mathbb{R}$ be a smooth function with $\varphi(a) = \varphi(b) = 0$ and $\varphi(t_0) = 1$. Letting $\lambda \in \mathbb{R}$ be a real parameter, we then consider

$$\mathbf{X}_\lambda := \widetilde{\mathbf{Y}} + \lambda\varphi\mathbf{Z} \in \Gamma_0(c, TM).$$

We then have

$$\begin{aligned}
I(\mathbf{X}_\lambda, \mathbf{X}_\lambda) &= I(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{Y}}) + 2\lambda I(\widetilde{\mathbf{Y}}, \varphi\mathbf{Z}) + O(\lambda^2) \\
&= 2\lambda \int_a^{t_0} \left[\mathbf{g}(\nabla_\mathbf{t}\mathbf{Y}, \nabla_\mathbf{t}(\varphi\mathbf{Z})) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{Y},\mathbf{t})\mathbf{Z})\right] dt + O(\lambda^2) \\
&= 2\lambda \int_a^{t_0} \nabla_\mathbf{t}\left[\mathbf{g}(\nabla_\mathbf{t}\mathbf{Y}, \varphi\mathbf{Z})\right] dt - 2\lambda \int_a^{t_0} \varphi\left[\mathbf{g}(\mathbf{Z}, \nabla_\mathbf{t}\nabla_\mathbf{t}\mathbf{Y} + \mathbf{R}(\mathbf{Y},\mathbf{t})\mathbf{t})\right] dt + O(\lambda^2) \\
&= 2\lambda\, \mathbf{g}(\nabla_\mathbf{t}\mathbf{Y}, \varphi\mathbf{Z})\Big|_a^{t_0} + O(\lambda^2) \\
&= 2\lambda\, \mathbf{g}(\nabla_\mathbf{t}\mathbf{Y}, \varphi\mathbf{Z})\Big|_{t_0} + O(\lambda^2) \\
&= -2\lambda \left|\left(\nabla_\mathbf{t}\mathbf{Y}\right)(t_0)\right|^2 + O(\lambda^2).
\end{aligned}$$

Hence, for sufficiently small $\lambda > 0$, we have $I(\mathbf{X}_\lambda, \mathbf{X}_\lambda) < 0$. □

Corollary 7.15 (to Proposition 7.14)**.** *A geodesic cannot be minimal after its first conjugate point.*

Proof. Assuming, in the proof of the Proposition 7.14, that $c(t_0)$ is the first point conjugate to $c(0)$ along c, consider the variation generated by $\mathbf{X}_\lambda$:

$$\alpha(t,s) := \exp_{c(t)}\left(s\mathbf{X}_\lambda(t)\right).$$

It then follows that

$$\frac{d}{ds}L[c_s]\Big|_{s=0} = 0, \qquad \frac{d^2}{ds^2}L[c_s]\Big|_{s=0} = I(\mathbf{X}_\lambda, \mathbf{X}_\lambda) < 0$$

for sufficiently small $\lambda > 0$. Therefore, for sufficiently small $s > 0$, we have $L[c_s] < L[c]$, so c is not minimising[6]. □

Recall that a geodesic, $c : [a,b] \to M$, is *minimising* between points $c(a)$ and $c(b)$ if

$$L[c] = d(c(a), c(b)).$$

A basic result is that a geodesic is initially minimising. For example, let M be complete, $p \in M$, $\mathbf{v} \in T_pM$, then there exists a $T > 0$ (T can be ∞) such that the geodesic $c : [0,\infty) \to M$ with $c(0) = p$, $c'(0) = \mathbf{v}$ satisfies

$$d(c(0), c(t)) = L[c|_{[0,t]}], \qquad \forall t \in [0,T].$$

(A similar result is true if M is not complete, but we have to then take care concerning the domain of c.)

Remark 7.16. Note that, on a complete manifold, the interval on which a geodesic is minimising is closed. In particular, let $c : [0,\infty) \to M$ be a geodesic with $L[c_{[0,t]}] = d(c(0), c(t))$ for all $t \in [0,b)$, for some $b > 0$. Since c is continuous, it follows that both $L[c_{[0,t]}]$ and $d(c(0), c(t))$ are continuous functions of t. Hence, taking the limit as $t \to b$, we deduce that $d(c(0), c(b)) = L[c_{[0,b]}]$, so c is minimising on the interval $[0,b]$.

Let $(M, \mathbf{g})$ be a complete Riemannian manifold. For $\mathbf{v} \in T_pM$, we defined $\gamma_\mathbf{v}$ to be the geodesic satisfying $\gamma_\mathbf{v}(0) = p$, $\gamma'_\mathbf{v}(0) = \mathbf{v}$. Let

$$I_\mathbf{v} := \{t \in \mathbb{R} : \gamma_\mathbf{v} \text{ minimal on } [0,t]\}.$$

[6]Stricly speaking, we are assuming that the path c_s is contained in M for sufficiently small values of s. Since, generally, we will be applying such results on complete manifolds, this condition is automatically satisfied.

As mentioned in Remark 7.16, $I_{\mathbf{v}}$ is closed, so we let $I_{\mathbf{v}} = [0, \rho(\mathbf{v})]$, where ρ is a map $T_pM \to (0, +\infty]$. If $\rho(\mathbf{v}) < \infty$, then $\gamma_{\mathbf{v}}(\rho(\mathbf{v}))$ is the *cut point of p along $\gamma_{\mathbf{v}}$*(or *cut point of $\gamma_{\mathbf{v}}$ with respect to p*.

Remarks 7.17.

- If M is compact, then for any $\mathbf{v} \in T_pM$ with $|\mathbf{v}| = 1$, we have $\rho(\mathbf{v}) \leq \operatorname{diam} M$, where
$$\operatorname{diam} M := \sup_{p,q \in M} d(p,q).$$
By the Hopf-Rinow theorem, $\operatorname{diam} M < \infty$.
- If we restrict to the unit sphere $S_p := \{\mathbf{v} \in T_pM : |\mathbf{v}| = 1\}$, then one can show that the map $S_p \to (0, +\infty]; \mathbf{v} \to \rho(\mathbf{v})$ is upper semi-continuous. If $(M, \mathbf{g})$ is complete, then this map is continuous[7].

The question of whether this is the only way that a geodesic can cease to be minimising is answered, negatively, by the following.

Proposition 7.18. *Let $(M, \mathbf{g})$ be a complete Riemannian manifold and $c : [a, b] \to M$ a minimising geodesic from $c(a)$ to $c(b)$. If there exists a distinct geodesic from $c(a)$ to $c(b)$ of the same length, then c is not minimal on any larger interval $[a, b + \epsilon]$.*

Proof. If c and γ are geodesics from $c(a)$ to $c(b)$ of the same length, then, given $\epsilon > 0$, we define the curve
$$\varphi(t) := \begin{cases} \gamma(t) & t \in [a, b] \\ c(t) & t \in (b, b + \epsilon] \end{cases}.$$
This curve joins the points $c(a)$ and $c(b+\epsilon)$. By the Hopf-Rinow theorem, there exists a minimising geodesic, s, from $\varphi(a)$ to $\varphi(b + \epsilon)$. Since φ is not a smooth geodesic (it is not smooth at $\varphi(b)$), it follows from the results of first variation of arc-length that φ cannot be minimising. Therefore
$$L[s] < L[\varphi] = L[\gamma|_{[a,b]}] + L[c|_{[b,b+\epsilon]}] = L[c|_{[a,b]}] + L[c|_{[b,b+\epsilon]}] = L[c|_{[a,b+\epsilon]}].$$
Hence c is not minimising from $c(a)$ to $c(b + \epsilon)$. □

Remark 7.19. Note that if $c : [a, b] \to M$ is minimising up to $c(b)$, and there exists another geodesic $\gamma : [a, b] \to M$ from $c(a)$ to $c(b)$, then we know from Remark 7.16 that $d(c(a), c(b)) = L[c|_{[a,b]}]$, so we must have $L[\gamma] \geq L[c]$.

We can deduce this result in an alternative fashion. If $L[\gamma] < L[c]$, then we can define a curve $\varphi : [a, b + \epsilon] \to M$ given by
$$\varphi(t) = \begin{cases} \gamma(t) & t \in [a, b] \\ c(b - (t - b)) & t \in [b, b + \epsilon] \end{cases}.$$
For sufficiently small $\epsilon > 0$, this defines a curve from $c(a)$ to $c(b - \epsilon)$ with length less than $L[c|_{[a,b-\epsilon]}]$, contradicting the fact that c is minimal up to $c(b)$.

Lecture 10

From now on, assume that $(M, \mathbf{g})$ is a *complete* Riemannian manifold

Proposition 7.20. *Let $(M, \mathbf{g})$ be a complete Riemannian manifold, $p \in M$, and $c : [0, \infty) \to M$ a geodesic with $c(0) = p$ and $|c'(t)| = 1$. Given $T \in (0, \infty)$, the point $c(T)$ is the cut point of p along c if and only if one (or both) of the following holds for $t = T$, and neither holds for any smaller value of t:*

1. *there exists a geodesic $\gamma \neq c$ from p to $c(T)$ with $L[\gamma] = L[c]$;*
2. *$c(T)$ is conjugate to p along c.*

[7]See [Cha], Theorem III.2.1 for a proof.

Proof. Firstly note that by Corollary 7.15 and Proposition 7.18, if either of the above conditions holds at $t = T$, then c will not be minimising for $t > T$.

To prove the converse statement, let $\mathbf{v} := c'(0)$. Then $c = \gamma_{\mathbf{v}} : [0, \infty) \to M$ and is minimising for $t \in [0, T]$, but not minimising for $t > T$, where $T \equiv \rho(\mathbf{v}) < \infty$. Given any decreasing sequence ϵ_i with $\epsilon_i \to 0$ as $i \to \infty$, the Hopf-Rinow theorem implies the existence of geodesics $\sigma_i : [0, L_i] \to M$ with $L_i := d(p, \gamma(T + \epsilon_i))$ with the property that $\sigma_i(0) = p$, $\sigma_i(L_i) = \gamma_{\mathbf{v}}(T + \epsilon_i)$ and $|\sigma_i'(0)| = 1$ (i.e. $\sigma_i'(0) \in S_p$). Since S_p is compact, there exists a convergent subsequence of $\{\sigma_i'(0)\}$, which we again denote by $\{\sigma_i'(0)\}$, with $\sigma_i'(0) \to \mathbf{w} \in S_p$ as $i \to \infty$. We then have a limiting geodesic, $\sigma := \gamma_{\mathbf{w}} : [0, \infty) \to M$. By continuity of the exponential map,

$$\lim_{i\to\infty} \exp_p(L_i \sigma_i'(0)) = \exp_p(\lim_{i\to\infty} L_i \sigma_i'(0)) = \exp_p(T\mathbf{w})$$

However, we also have

$$\lim_{i\to\infty} \exp_p(L_i \sigma_i'(0)) = \lim_{i\to\infty} \sigma_i(L_i) = \lim_{i\to\infty} c(T + \epsilon_i) = c(T) = \gamma_{\mathbf{v}}(T) = \exp_p(T\mathbf{v}).$$

Hence $\mathbf{v}, \mathbf{w} \in S_p$ and $\exp_p(T\mathbf{w}) = \exp_p(T\mathbf{v})$. There are now two possibilities:

- $\mathbf{w} \neq \mathbf{v}$,
- $\mathbf{w} = \mathbf{v}$.

In the first case, $\gamma_{\mathbf{w}}$ and $\gamma_{\mathbf{v}}$ are distinct geodesics from p to $\gamma_{\mathbf{v}}(T)$, both of length T.

In the second case, we wish to show that $D_{T\mathbf{v}} \exp_p : T_{\mathbf{v}}(T_pM) \to T_{\gamma_{\mathbf{v}}(T)}M$ is singular. Assume, to the contrary, that $D_{T\mathbf{v}} \exp_p$ is non-singular. Then the inverse function theorem implies that $\exp_p$ defines an injective mapping from a neighbourhood of $T\mathbf{v}$ in T_pM, U, to a neighbourhood, V, of $\gamma_{\mathbf{v}}(T)$ in M. For sufficiently large i, we have $\sigma_i(L_i) \in V$ and $\sigma_i'(0) \in U$, and therefore the relation

$$\exp_p\left((T + \epsilon_i)\,\mathbf{v}\right) = \gamma_{\mathbf{v}}(T + \epsilon_i) = \sigma_i(L_i) = \exp_p L_i\left(\sigma_i'(0)\right)$$

implies that

$$(T + \epsilon_i)\,\mathbf{v} = L_i \sigma_i'(0).$$

Therefore, $\sigma_i'(0) = \mathbf{v}$ and $L_i = T + \epsilon_i$. Hence

$$d(p, \gamma_{\mathbf{v}}(T + \epsilon_i)) = T + \epsilon_i = L[\gamma_{\mathbf{v}}|_{[0, T+\epsilon_i]}].$$

This contradicts the fact that $\gamma_{\mathbf{v}}$ is not minimising beyond the point $\gamma_{\mathbf{v}}(T)$. Hence $D\exp_p$ is singular at $T\mathbf{v} \in T_pM$, so $\gamma_{\mathbf{v}}(T)$ is conjugate to p along $\gamma_{\mathbf{v}}$. □

8. Cut locus

Let

$$U_p := \{t\mathbf{v} \in T_pM : \mathbf{v} \in S_p, t \in [0, \rho(\mathbf{v}))\}.$$

Remarks 8.1.

(1) U_p is an open neighbourhood of $0 \in T_pM$.
(2) $\exp_p : U_p \to \exp_p(U_p)$ is injective[8]. In particular, if $q \in \exp_p U_p$ then there exists a unique $\mathbf{v} \in U_p$ with $\exp_p \mathbf{v} = q$ and $d(p, q) = |\mathbf{v}|$ which, in our earlier notation, is $r(q)$. Hence $r(q) = d(p, q)$ for all $q \in \exp_p U_p$.
(3) $\exp_p : U_p \to \exp_p(U_p)$ is non-singular everywhere i.e. $D_{\mathbf{v}} \exp_p$ is non-singular for all $\mathbf{v} \in U_p$. (Otherwise $\exp_p \mathbf{v}$ is conjugate to p along $\gamma_{\mathbf{v}}$ and thus cannot lie in $\exp_p U_p$.)

[8]*Proof*: If $q \in \exp_p(U_p)$ is the image of two distinct vectors, $\mathbf{v}, \mathbf{w}$, in U_p, then there are two distinct geodesics from p to q. If $|\mathbf{v}| \neq |\mathbf{w}|$, in which case one of these geodesics is non-minimising, which contradicts the fact that both $\mathbf{v}$ and $\mathbf{w}$ lie in U_p. Otherwise, $|\mathbf{v}| = |\mathbf{w}|$, in which case there exist two distinct geodesics between p and q of the same length. However, in this case, the geodesics cannot be minimising after the point q, again contradicting the fact that $\mathbf{v}$, $\mathbf{w}$ lie in U_p.

Definition 8.2. Let $p \in M$. The *cut-locus of p in T_pM* is the set

$$C(p) := \{\rho(\mathbf{v})\mathbf{v} : \mathbf{v} \in S_p \text{ such that } \rho(\mathbf{v}) < \infty\} \subset T_pM.$$

The *cut-locus* of p is the set

$$\mathrm{Cut}(p) := \{\gamma_{\mathbf{v}}(\rho(\mathbf{v})) : \mathbf{v} \in S_p \text{ such that } \rho(\mathbf{v}) < \infty\} \equiv \exp_p C(p)$$

i.e. the union of the cut-points of p. The *injectivity radius of* $(M, \mathbf{g})$ *at* p, $\mathrm{inj}(p)$, is

$$\mathrm{inj}(p) = \inf_{\mathbf{v} \in S_p} \rho(\mathbf{v}).$$

Remark 8.3.

(1) Equivalently,

$$\begin{aligned} \mathrm{inj}(p) &= d(p, \mathrm{Cut}(p)) \\ &= \sup\{R \in (0, \infty] : \{\mathbf{v} \in T_pM : |\mathbf{v}| < R\} \subseteq U_p\} \end{aligned}$$

where we define $d(p, \mathrm{Cut}(p)) = \infty$ if $\mathrm{Cut}(p)$ is empty.

(2) It can be shown that, for any $p \in M$, the set $\mathrm{Cut}(p)$ is a subset of M of measure zero (see, e.g., [GHL, §3.96]).

Given points $p, q \in M$, the Hopf-Rinow theorem implies the existence of a minimal geodesic, $\gamma_{\mathbf{v}} : [0, \infty) \to M$ with $|\mathbf{v}| = 1$ such that $\gamma_{\mathbf{v}}(0) = p$, $\gamma_{\mathbf{v}}(T) = q$. Since $\gamma_{\mathbf{v}}$ is minimising, it follows that $T \leq \rho(\mathbf{v})$. Hence either $T = \rho(\mathbf{v})$, in which case $q \in \mathrm{Cut}(p)$, or $T < \rho(\mathbf{v})$, in which case $q \in \exp_p U_p$.

We therefore have proved the first part of the following result:

Proposition 8.4. *Let $(M, \mathbf{g})$ be a complete Riemannian manifold. Let $p \in M$. Then*

$$M = \left(\exp_p (U_p)\right) \cup \mathrm{Cut}(p). \tag{8.1}$$

The above union is disjoint.

Proof. To prove that the union is disjoint, let $x \in \left(\exp_p (U_p)\right) \cap \mathrm{Cut}(p)$. Since $x \in \exp_p (U_p)$, there exists a geodesic c with $c(0) = p$, $c(1) = x$ that is minimal on the larger interval $[0, 1 + \epsilon]$. However, $x \in \mathrm{Cut}(p)$ implies that there exists a geodesic, γ, with $\gamma(0) = p$, $\gamma(1) = x$ that is not minimal for $t > 1$. Since both of these geodesics are minimal, they must have the same length. Since c remains minimal after x whereas γ does not, then c and γ must be distinct geodesics from p to x of the same length. However, Proposition 7.18 then tells us that c cannot be minimal after the point x, and we have a contradiction. Hence $\left(\exp_p (U_p)\right) \cap \mathrm{Cut}(p) = \emptyset$, and the union in (8.1) is disjoint. $\square$

Remark 8.5. The above result implies that if we remove the point p from M, then the flow along radial geodesics from p defines a map $M \setminus \{p\} \to \mathrm{Cut}(p)$. This map turns out to be a retraction, and $\mathrm{Cut}(p)$ is a deformation retract of $M \setminus \{p\}$[9]. In the case where M is compact, the above result implies that M is the disjoint union of an open ball in $\mathbb{R}^n$ and the continuous image of an $(n-1)$-sphere. Such information allows one to deduce various topological properties of M (e.g. certain homotopy groups) in terms of those of the cut locus $\mathrm{Cut}(p)$. (See, e.g., [K] for more information on this topic.)

Lecture 11

Remark 8.6. Since the exponential map is smooth, it follows that the map $\exp_p : U_p \to \exp_p(U_p)$ is smooth[10] Therefore, the exponential coordinates $y^i : \exp_p(U_p) \to \mathbb{R}$ introduced in Section 4 are smooth. As such, the distance function $r : \exp_p(U_p) \to \mathbb{R}; q \to |\exp_p^{-1} q| \equiv d(p, q)$ will be smooth on $\exp_p(U_p) \setminus \{p\}$.

[9]See, e.g., [HY] for a definition of these terms, if they are unfamiliar.

[10]In particular, the Hopf-Rinow theorem and definition of U_p imply that we are in the domain where the geodesic equations with initial conditions $c(0) = p$, $c'(0) = \mathbf{v} \in U_p$ have a unique solution for $t \in [0, \rho(\mathbf{v})$. As such, within this domain, we can then deduce smooth dependence on initial conditions.

Bishop-Gromov Comparison Theorem (Final Version). *Let $(M, \mathbf{g})$ be a complete Riemannian manifold of dimension n with*

$$\mathbf{Ric}(\mathbf{X}, \mathbf{X}) \geq K\,(n-1)\,\mathbf{g}(\mathbf{X}, \mathbf{X}), \qquad \forall \mathbf{X} \in \mathfrak{X}(M)$$

for some $K \in \mathbb{R}$. Let $V_K(r)$ denote the volume of the ball of radius r in the complete, simply-connected space of constant curvature K and, given $p \in M$, let $B_{\mathbf{g}}(p, r) := \{q \in M : d_{\mathbf{g}}(p, q) < r\}$. Then the map

$$r \mapsto \frac{\mathrm{Vol}\,(B_{\mathbf{g}}(p, r))}{V_K(r)}$$

is non-increasing. In particular,

$$\mathrm{Vol}\,(B_{\mathbf{g}}(p, r)) \leq V_K(r).$$

Proof. The proof follows that of Version 1, until we get to the point where we deduce that

$$\frac{a(r, \theta)}{sn_K(r)}$$

is non-increasing as a function of r. For fixed $\theta \in S_p := \{\mathbf{v} \in T_pM : |\mathbf{v}| = 1\}$, this holds up to $r = \rho(\mathbf{v})$ i.e. up to the point where the geodesic γ_θ intersects the cut-locus of p. To extend the result beyond the cut-locus, we define

$$a^+(r, \theta) := \begin{cases} a(r, \theta) & 0 \leq r \leq \rho(\theta), \\ 0 & r > \rho(\theta). \end{cases}$$

We then see that

$$\text{the map } r \mapsto \frac{a^+(r, \theta)}{sn_K(r)} \text{ is non-increasing for } r > 0, \text{ for each } \theta \in S_p.$$

With this definition, we see that

$$\mathrm{Vol}\,(B_{\mathbf{g}}(p, r)) = \int_0^r \int_{S^{n-1}} 1 d\mathbf{vol}_{\mathbf{g}} = \int_0^r \int_{S^{n-1}} a^+(t, \theta)^{n-1} dt \wedge d\mathbf{vol}_{\mathbf{g}_{S^{n-1}}}.$$

Equation (6.9) is now replaced by

$$\frac{d}{dr}\left[\frac{\mathrm{Vol}\,(B_{\mathbf{g}}(p, r))}{V_K(r)}\right] = \frac{1}{\mathrm{Vol}\,(S^{n-1})} \frac{1}{\left(\int_0^r sn_K(t)^{n-1} dt\right)^2} \int_0^r sn_K(t)^{n-1} sn_K(r)^{n-1}$$
$$\times \int_{S^{n-1}} \left[\left(\frac{a^+(r, \theta)}{sn_K(r)}\right)^{n-1} - \left(\frac{a^+(t, \theta)}{sn_K(t)}\right)^{n-1}\right] d\theta dt.$$

Since $\frac{a^+(t,\theta)}{sn_K(t)}$ is a non-increasing function of t, we have

$$\frac{d}{dr}\left[\frac{\mathrm{Vol}\,(B_{\mathbf{g}}(p, r))}{V_K(r)}\right] \leq 0.$$

□

Remark 8.7. Where we said "complete, simply-connected space of constant curvature K" in the statement of this theorem, we simply mean metrics $\mathbf{g}_K$ and the corresponding manifolds that we introduced in Section 1[11]. Also,

$$V_K(r) = \int_0^r \int_{S^{n-1}} sn_K(t)^{n-1} d\theta dt.$$

[11]The reason I stated the final version of the Bishop-Gromov result as above (which, admittedly, is inconsistent with the way I have presented material previously) is that it is the way you are likely to find the result stated in textbooks.

Corollary 8.8 (Myers's theorem)**.** *Let M be a complete Riemannian manifold of dimension n, such that the Ricci curvature of M obeys the inequality*

$$\mathbf{Ric}(\mathbf{X}, \mathbf{X}) \geq (n-1)K\mathbf{g}(\mathbf{X}, \mathbf{X}), \qquad \forall \mathbf{X} \in \mathfrak{X}(M)$$

for some constant $K > 0$. Then the diameter of M satisfies $\operatorname{diam}(M) \leq \pi/\sqrt{K}$.

Proof. Assume, to the contrary, that $\operatorname{diam}(M) > \pi/\sqrt{K}$. Then there exist points $p, q \in M$ such that $d(p, q) > \pi/\sqrt{K}$. Let $r(x) := d(p, x)$, for $x \in M$. From the comparison results earlier, we know that

$$\Delta r = (n-1)\frac{\partial_r a}{a} \leq (n-1)\frac{\partial_r sn_K(r)}{sn_K(r)} = (n-1)\sqrt{K}\cot\left(\sqrt{K}r\right), \tag{8.2}$$

as long as x remains away from the cut-locus of p. Let c be a minimising geodesic from p to q, parametrised by arc-length (i.e. $|c'| = 1$). Since c is minimal, we know that $c(\pi/\sqrt{K})$ is not in the cut-locus of p, and hence r is smooth at that point. However, letting $t \to \pi/\sqrt{K}$ from below, we deduce from (8.2) that $\Delta r \to \infty$. This is a contradiction, since the left-hand-side must be finite away from the cut-locus. Hence $\operatorname{diam} M \leq \pi/\sqrt{K}$. □

Lecture 12

Corollary 8.9. *Let M obey the conditions of Myers's Theorem. Then M is compact.*

Proof. By Theorem 8.8, given a point $p \in M$ then $M = \{q \in M : d(p, q) \leq \operatorname{diam} M\}$, and $\operatorname{diam} M \leq \pi/\sqrt{K} < \infty$. Since this is a closed ball of finite radius in a complete manifold, the Hopf-Rinow theorem implies that this ball, and hence M, is compact. □

Rearranging this result a little, we obtain a result that anticipates some of the ideas inherent in the proof of the singularity theorems in Lorentzian geometry:

Corollary 8.10. *Let M be a non-compact Riemannian manifold of dimension n, such that the Ricci curvature of M obeys the inequality*

$$\mathbf{Ric}(\mathbf{X}, \mathbf{X}) \geq (n-1)K\mathbf{g}(\mathbf{X}, \mathbf{X}), \qquad \forall \mathbf{X} \in \mathfrak{X}(M)$$

for some constant $K > 0$. Then M is incomplete.

Part 3. Lorentzian geometry

9. Inner products of Lorentzian signature

In Riemannian geometry, one has an inner product on the tangent space, T_pM, for each $p \in M$. Each tangent space is isomorphic to $\mathbb{R}^n$, and we may choose an orthonormal basis such that the inner product takes the form

$$\langle \mathbf{v}, \mathbf{w} \rangle_R = v^1 w^1 + v^2 w^2 + \cdots + v^n w^n,$$

where $(v^1, \ldots, v^n)$, $(w^1, \ldots, w^n)$ are the components of the vectors $\mathbf{v}$, $\mathbf{w}$ with respect to the chosen basis. We then define a norm, $\| \cdot \|_R$ on $\mathbb{R}^n \cong T_pM$, with the property that

$$\|\mathbf{v}\|_R^2 = \langle \mathbf{v}, \mathbf{v} \rangle_R = \left(v^1\right)^2 + \left(v^2\right)^2 + \cdots + \left(v^n\right)^2.$$

For any $r > 0$, the surface $\{\mathbf{v} \in T_pM : \|\mathbf{v}\|_R = r\}$ is an $(n-1)$-sphere and, in particular, is compact.

On a Lorentzian manifold, we have an inner product of Lorentzian signature on each tangent space. It is conventional to take the dimension of our manifold to be $(n+1)$, so $T_pM \cong \mathbb{R}^{n+1}$, and to choose a basis $\{\mathbf{e}_0, \mathbf{e}_1, \ldots, \mathbf{e}_n\}$ such that inner product takes the form

$$\langle \mathbf{v}, \mathbf{w} \rangle_L := -v^0 w^0 + v^1 w^1 + \cdots + v^n w^n,$$

where $(v^0, v^1, \ldots, v^n)$ are the components of a vector $\mathbf{v} \in T_pM$, etc. We also define the Lorentzian "norm-squared" of a vector by

$$\|\mathbf{v}\|_L^2 = \langle \mathbf{v}, \mathbf{v} \rangle_L = -\left(v^0\right)^2 + \left(v^1\right)^2 + \left(v^2\right)^2 + \cdots + \left(v^n\right)^2.$$

In contrast to the Riemannian case, a non-zero vector $\mathbf{v} \in T_pM$ may have norm-squared that is non-positive. A non-zero vector $\mathbf{v}$ is defined to be

- *space-like* if $\|\mathbf{v}\|_L^2 > 0$;
- *time-like* if $\|\mathbf{v}\|_L^2 < 0$;
- *null* if $\|\mathbf{v}\|_L^2 = 0$.

In physical terms, the 0-direction corresponds to the time direction. Bearing this in mind, we say that a non-space-like vector[12] $\mathbf{v}$ is *future-directed* if $v^0 > 0$ and *past-directed* if $v^0 < 0$.

In Lorentzian geometry, the level sets $\mathbf{v} \in \mathbb{R}^{n+1}$ with $\|\mathbf{v}\|_L^2 = a$, for $a \in \mathbb{R}$, are non-compact. These are of different character, depending on the value of a:

(1) $a = +k^2 > 0$: In this case, we need $\{\mathbf{v} : \sum_{i=1}^n (v^i)^2 = k^2 + (v^0)^2\}$, which is a non-compact hyperboloid. For $n > 1$, this set is connected.
(2) $a = -k^2 < 0$: We require $\{\mathbf{v} : (v^0)^2 = k^2 + \sum_{i=1}^n (v^i)^2\}$. This set consists of two hyperboloids, corresponding to whether v^0 is positive or negative. Each of these hyperboloids is connected, but non-compact.
(3) $a = 0$: We require $\{\mathbf{v} : (v^0)^2 = \sum_{i=1}^n (v^i)^2\}$. If we include the zero vector, then this set is the union of two cones, each with vertex at 0. This set is called the null cone/light cone (of 0). The subset with $v^0 > 0$ is the future null/light cone, and the subset with $v^0 < 0$ is the past null/light cone. The light-cone and its future/past subsets are non-compact.

10. Lorentzian manifolds

A Lorentzian metric on a manifold M is a section, $\mathbf{g}$, of $T_2^0(M)$ that is symmetric, non-degenerate and defines an inner product of Lorentzian signature on each tangent space T_xM, for each $x \in M$. A Lorentzian manifold will mean a pair $(M, \mathbf{g})$ consisting of a manifold M and a Lorentzian metric, $\mathbf{g}$, on M.

Standard material from Riemannian geometry concerning the Levi-Civita connection, the curvature tensor and the Ricci tensor carry over, unchanged, to the case of Lorentzian manifolds. The Riemannian concept of sectional curvature is not, generally, well-defined for Lorentzian metrics, and is infrequently used.

[12] i.e. time-like or null. We will also refer to such vectors as *causal*.

A vector field, $\mathbf{X} \in \mathfrak{X}(M)$, is defined as being *time-like*, *space-like* or *null* if $\mathbf{X}_p \in T_pM$ is time-like, space-like or null with respect to the Lorentzian inner product induced on T_pM, for each $p \in M$. We will generally assume that our manifolds are *time-orientable*, in the sense that it is possible to define, continuously, a division of non-space-like vectors into two classes, referred to as future-directed and past-directed. This is equivalent [O'N, Chapter 5, Lemma 32] to the existence of a non-vanishing time-like vector field on M. Although the existence and uniqueness theorem for geodesics carries over directly, we must now consider geodesics that are

- space-like i.e. $\|c'\|^2 > 0$;
- time-like i.e. $\|c'\|^2 < 0$;
- null i.e. $\|c'\|^2 = 0$

separately. In the case of non-space-like geodesics, we also characterise geodesics as future-directed and past-directed according to whether their tangent vectors are future or past-directed.

As general references on Lorentzian geometry, the book closest to the general approach that we will take is [HE], although there are some points at which we will follow [O'N]. **Lecture 13**

10.1. **Examples.**

Example 10.1 (Minkowski space)**.** $M = \mathbb{R}^4$, and we adopt local coordinates (t, x, y, z). It is conventional to also write these as $x^0 = t$, $x^1 = x$, $x^2 = y$, $x^3 = z$

$$\mathbf{g} = -dt^2 + dx^2 + dy^2 + dz^2. \tag{10.1}$$

Generally, we denote $\mathbb{R}^{n+1}$ with a Lorentzian metric of the above form by $\mathbb{R}^{n,1}$. More generally still, we denote $\mathbb{R}^{p+q}$ with the flat metric of signature (p, q)

$$\mathbf{g} = \sum_{a=1}^{p}(dx^a)^2 - \sum_{i=1}^{q}(dt^i)^2 \tag{10.2}$$

by $\mathbb{R}^{p,q}$.

If we adopt spherical coordinates, (r, θ, ϕ) in the spatial (i.e. x, y, z) directions, then the metric (10.1) takes the form

$$\mathbf{g} = -dt^2 + dr^2 + r^2\mathbf{g}_{S^2},$$

where $\mathbf{g}_{S^2}$ denotes the standard round metric on the unit 2-sphere. More generally, the $(n+1)$-dimensional Minkowski metric takes the form

$$\mathbf{g} = -dt^2 + dr^2 + r^2\mathbf{g}_{S^{n-1}}. \tag{10.3}$$

Example 10.2 (de Sitter space)**.** The $(n+1)$-dimensional de Sitter metric may be looked on as the metric induced on the hyperboloid $-t^2 + r^2 = 1$, in $\mathbb{R}^{n+1,1}$ with the metric (10.3). Noting that r cannot be zero on the surface $-t^2 + r^2 = 1$, we may adopt new coordinates (T, R) with

$$r = R\cosh T, \qquad t = R\sinh T,$$

on a neighbourhood of this surface, which is the set $R = 1$. In terms of which the Minkowski metric takes the form

$$\mathbf{g} = dR^2 - R^2 dT^2 + R^2\cosh^2 T\mathbf{g}_{S^n}$$

on this neighbourhood. The de Sitter metric is then the metric induced on the hyper-surface $R = 1$:

$$\mathbf{g}_{dS} = -dT^2 + \cosh^2 T\mathbf{g}_{S^n}, \qquad T \in \mathbb{R}.$$

This metric is of constant curvature, in the sense that the $(0, 4)$ version of the curvature tensor takes the form

$$\mathbf{R}(\mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{X}) = K\left(\mathbf{g}(\mathbf{U}, \mathbf{W})\mathbf{g}(\mathbf{V}, \mathbf{X}) - \mathbf{g}(\mathbf{U}, \mathbf{X})\mathbf{g}(\mathbf{V}, \mathbf{W})\right), \qquad \forall \mathbf{U}, \mathbf{V}, \mathbf{W}, \mathbf{X} \in \mathfrak{X}(M)$$

with $K = 1$ in the current case. The Ricci tensor of $\mathbf{g}$ is then

$$\mathbf{Ric} = Kn\mathbf{g}.$$

Example 10.3 (Anti-de Sitter space)**.** The corresponding metric with constant negative curvature, $K = -1$, is the anti-de Sitter metric. The $(n+1)$-dimensional anti-de Sitter metric may constructed as the induced metric on a hyper-surface in $\mathbb{R}^{n+2}$ with the flat metric of signature $(n, 2)$. In particular, we write

$$\begin{aligned}\mathbf{g} &= -du^2 - dv^2 + \sum_{i=1}^{n} \left(dx^i\right)^2, \qquad u, v, x^i \in \mathbb{R}\\ &= -\left(dT^2 + T^2 dt^2\right) + dR^2 + R^2 \mathbf{g}_{S^{n-1}}, \qquad T, R \in [0, \infty), \qquad t \in [0, 2\pi),\\ &= -dT^2 + dR^2 - T^2 dt^2 + R^2 \mathbf{g}_{S^{n-1}}, \qquad T, R \in [0, \infty), \qquad t \in [0, 2\pi),\end{aligned}$$

where $u = T\cos t$, $v = T\sin t$, $R = |\mathbf{x}|$. The anti-de Sitter metric is then the metric induced on the component of the submanifold $-T^2 + R^2 = -1$. Rewriting the metric above with $T = \rho\cosh r$, $R = \rho\sinh r$ with $\rho \in [0, \infty)$, $r \in \mathbb{R}$, we have

$$\mathbf{g} = -d\rho^2 + \rho^2 dr^2 - \rho^2\cosh^2 r\, dt^2 + \rho^2 \sinh^2 r\, \mathbf{g}_{S^{n-1}}.$$

The metric induced on the surface $\rho = 1$ is then

$$\mathbf{g}_{AdS} = -\cosh^2 r\, dt^2 + dr^2 + \sinh^2 r\, \mathbf{g}_{S^{n-1}}.$$

Although the construction from $\mathbb{R}^{n+2}$ leads to the coordinate t having period 2π, this leads to the above metric having closed time-like curves. We avoid this problem by "de-identifying" the time variable t, and taking the above metric to be defined for

$$t \in \mathbb{R}, \qquad r \in \mathbb{R}.$$

In this case, the topology of the manifold on which the anti-de-Sitter metric lies is simply $\mathbb{R}^{n+1}$.

A feature of the anti-de Sitter metric that will be of interest to us is that it is geodesically complete, but there exist points in it that are not connected by any geodesic. Consider, for simplicity, the case $n = 1$, in which case we have the two-dimensional Lorentzian metric

$$\mathbf{g}_{AdS} = -\cosh^2 r\, dt^2 + dr^2, \qquad (t, r) \in \mathbb{R}^2.$$

As indicated above, this may be viewed as the induced metric on the hyperboloid $\Sigma := \{(u, v, x) \in \mathbb{R}^{2,1} : -u^2 - v^2 + x^2 = -1\}$ in $\mathbb{R}^{1,2}$ with the metric $\mathbf{g} = -du^2 - dv^2 + dx^2$. The geodesics of the anti-de Sitter metric are then the intersection of the hyperboloid Σ with planes in $\mathbb{R}^{1,2}$ that pass through the origin. In particular, letting $\mathbf{x} := (u, v, x) \in \mathbb{R}^{1,2}$ be the position vector in $\mathbb{R}^{1,2}$ then a plane through the origin is the set of points $\mathbf{x}$ that satisfy an equation of the form

$$\mathbf{g}(\mathbf{a}, \mathbf{x}) = -au - bv + cx = 0,$$

where $\mathbf{a} = (a, b, c)$ is a (constant) vector in $\mathbb{R}^{1,2}$. A geodesic on Σ (viewed simply as a set, rather than a parametrised curve $c : \mathbb{R} \to \Sigma$) is then the intersection

$$\{\mathbf{x} \in \mathbb{R}^{1,2} : \mathbf{g}(\mathbf{x}, \mathbf{x}) = -1, \mathbf{g}(\mathbf{a}, \mathbf{x}) = 0\}.$$

In terms of the coordinates introduced above, this implies that we need

$$\begin{aligned}& au + bv = cx\\ \Longrightarrow\quad & aT\cos t + bT\sin t = cR\\ \Longrightarrow\quad & a\rho\cosh r\cos t + b\rho\cosh r\sin t = c\rho\sinh r.\end{aligned}$$

Since $\Sigma = \{\rho = 1\}$, the geodesics on Σ are lines of the form

$$a\cosh r\cos t + b\cosh r\sin t = c\sinh r$$

i.e.

$$a\cos t + b\sin t = c\tanh r,$$

where a, b, c are real constants. If we consider geodesics through the point $t = r = 0$, then we deduce that we must have $a = 0$, and therefore

Lecture 14

$$b\sin t = c\tanh r.$$

$\boxed{\text{Case 1: } c = 0}$ In this case, the geodesic equations, $\nabla_{\mathbf{v}}\mathbf{v} = 0$ imply that $\frac{dr}{ds} = \text{const}$, where s is an affine parameter along the geodesic. Parametrising by arclength implies that the constant is ± 1, in which case the geodesic curves take the form

$$t(s) = 0, \qquad r(s) = \pm s.$$

As such, the geodesics simply travel along the r axis in the positive and negative directions.

$\boxed{\text{Case 2: } c \neq 0}$ For $c \neq 0$, we then deduce that the geodesic takes the form

$$r(t) = \tanh^{-1}(\lambda \sin t), \tag{10.4}$$

where $\lambda := b/c \in \mathbb{R}$. If we now consider t as a parameter on the geodesic (i.e. view the geodesic as a map $c : t \mapsto (t, r(t))$ for $t \in I$, where I is some sub-interval of $\mathbb{R}$), then we find that

$$\left\| \frac{dx}{dt} \right\|^2 \equiv \left(\frac{dr}{dt} \right)^2 - \cosh^2 r = \cdots = \frac{\lambda^2 - 1}{(1 - \lambda^2 \sin^2 t)^2}.$$

(Note that the right-hand-side of this expression is not constant, implying that t is not an affine parameter along the geodesics. In particular, if we let $\mathbf{v} = \frac{dx}{dt}$, then $\nabla_{\mathbf{v}}\mathbf{v}$ is equal to a non-zero multiple of $\mathbf{v}$, rather than 0. We can, however, find a reparametrisation of the curve as $s \mapsto (t(s), r(x))$ for which $\widetilde{\mathbf{v}} = \frac{dx}{ds}$ satisfies $\nabla_{\widetilde{\mathbf{v}}}\widetilde{\mathbf{v}} = 0$.) There are now three separate cases to consider

$\boxed{\text{Case 2a: } \lambda > 1}$ In this case, the geodesics have $\left\| \frac{dx}{dt} \right\|^2 > 0$, and so are space-like. From (10.4), we see that $r \to \infty$ as $t \to \arcsin \lambda^{-1}$. As such, the space-like geodesics reach space-like infinity at a finite value of t. The affine distance along such a curve, $\int \left\| \frac{dx}{dt} \right\| dt$, is, however, infinite.

$\boxed{\text{Case 2b: } \lambda = 1}$ In this case, the geodesics have $\left\| \frac{dx}{dt} \right\|^2 = 0$, and so are null. From (10.4), we see that $r \to \infty$ as $t \to \frac{\pi}{2}$.

$\boxed{\text{Case 2a: } \lambda < 1}$ In this case, the geodesics have $\left\| \frac{dx}{dt} \right\|^2 < 0$, and so are time-like. From (10.4), we see that $r(t + 2\pi) = r(t)$ and $r(t + \pi) = -r(t)$. Therefore all of the time-like geodesics starting at $t = r = 0$ pass through the points $t = \pi, r = 0$ and $t = 2\pi, r = 0$.

A consequence of this discussion is that, if we consider any point $t = \pi, r \neq 0$, then there exists no geodesic between the point $t = 0, r = 0$ and that point. More generally, if we consider any points contained in the region enclosed by the null geodesics that pass through the point $t = \pi, r = 0$ (i.e. points on any space-like geodesic through that point) then there is no geodesic connecting the point $t = 0, r = 0$ to that point.

Since anti-de Sitter space is geodesically complete, this implies that the corollary of the Hopf-Rinow theorem that states that, for Riemannian manifolds, completeness implies the existence of minimising geodesics between points, is strictly false in the Lorentzian case.

Definition 10.4. For sets $A, U \subseteq M$ we define the *chronological future of A relative to U* to be the set

$$I^+(A, U) := \{q \in U : \exists p \in A \text{ and a future-directed time-like curve from } p \text{ to } q \text{ in } U\}$$

and the *causal future of A relative to U*

$$J^+(A, U) := (A \cap U) \cup \{q \in U : \exists p \in A \text{ and a future-directed causal curve from } p \text{ to } q \text{ in } U\}.$$

When $U = M$, we write $I^+(A)$ and $J^+(A)$, respectively.
Similarly, we define $I^-(A, U), J^-(A, U)$.

Remarks 10.5.

(1) By a curve we mean one of non-zero extent, so we can have $S \not\subseteq I^+(S, U)$.
(2) A case that will be of particular interest to us is when the set A consists of a point $p \in M$. In this case, we use the notation $I^\pm(p, U)$, $J^\pm(p, U)$, $I^\pm(p)$, $J^\pm(p)$.
(3) We will see later that the set $I^\pm(A)$ is open. The set $J^\pm(A)$ need not be closed, though.

(4) In the case of anti-de Sitter space, if we take the point $p := \{t = r = 0\}$ and $q := \{t = \pi, r = 0\}$, then the set $J^+(p) \cap J^-(q)$ is non-compact. The fact that there exist points in M that are time-like separated that cannot be joined by a time-like geodesic turns out to be related to this fact.

Lecture 15

Example 10.6 (The Schwarzschild solution)**.** In physics, this is the metric that describes a non-rotating, un-charged black hole. It depends on a parameter $m \in \mathbb{R}$ which, asymptotically, can be identified with the Newtonian mass. Therefore, as far as physics is concerned, it is best to assume that $m \geq 0$, although the solution does make sense for $m < 0$.

Let (t, r, θ, ϕ) be coordinates on $\mathbb{R}^4$, with (r, θ, ϕ) the usual spherical polar coordinates on $\mathbb{R}^3$. The exterior Schwarzschild solution is defined on the subset $r > 2m$ of $\mathbb{R}^4$, which is topologically $\mathbb{R}^2 \times S^2$. In this region, the metric takes the form

$$\mathbf{g} = -\left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right),$$

This metric becomes badly-behaved as $r \to 2m$, but this is simply due to the fact that (t, r) are not a good choice of coordinates on this region. We may rewrite the metric in the form

$$\mathbf{g} = -\left(1 - \frac{2m}{r}\right)\left[dt - \left(1 - \frac{2m}{r}\right)^{-1} dr\right]\left[dt + \left(1 - \frac{2m}{r}\right)^{-1} dr\right] + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right),$$

and define a new coordinate, $u \in \mathbb{R}$, by

$$du = dt + \left(1 - \frac{2m}{r}\right)^{-1} dr = \cdots = d[t + r + \log|r - 2m|],$$

so

$$u = t + r + \log|r - 2m| + \text{const}.$$

Rewriting the metric in terms of the coordinates (u, r, θ, ϕ), it then takes the form

$$\begin{aligned}
\mathbf{g} &= -\left(1 - \frac{2m}{r}\right)\left[du - 2\left(1 - \frac{2m}{r}\right)^{-1} dr\right] du + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right), \\
&= -\left(1 - \frac{2m}{r}\right) du^2 + 2dr du + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right). \qquad (10.5)
\end{aligned}$$

We may use this expression to define a new metric, $\mathbf{g}'$, given by the above coordinate expression on the manifold, M', defined by

$$u \in \mathbb{R}, \qquad r \in (0, \infty), \qquad (\theta, \phi) \in S^2.$$

On restriction to the original manifold M, where $r > 2m$, the metric $\mathbf{g}'|_M$ is isometric to the original metric $\mathbf{g}$.

Definition 10.7. In general, a Lorentzian manifold $(M', \mathbf{g}')$ is said to be an *extension* of a Lorentzian manifold $(M, \mathbf{g})$ if there exists an isometric embedding $i : M \to M'$. A Lorentzian manifold $(M, \mathbf{g})$ for which any extension $(M', \mathbf{g}')$ satisfies $i(M) = M'$ is said to be *inextendible*.

Remark 10.8. It turns out that the extension, $(M', \mathbf{g}')$ of the external Schwarzschild solution is not inextendible. See [HE, Chapter 5.5] for the full extension of the external Schwarzschild solution.

Calculating the norm-squared of the curvature tensor

$$|\mathbf{R}|^2 := \sum_{i,j,k,l} \mathbf{R}(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k, \mathbf{e}_l)\mathbf{R}(\mathbf{e}_i, \mathbf{e}_j, \mathbf{e}_k, \mathbf{e}_l),$$

where $\mathbf{e}_i$ is an orthonormal basis:

$$\mathbf{g}'(\mathbf{e}_i, \mathbf{e}_j) = \begin{cases} -1 & i = j = 0, \\ 1 & i = j = 1, 2, 3, \\ 0 & \text{otherwise}, \end{cases}$$

then we find that

$$|\mathbf{R}|^2 = \text{constant} \times \frac{m^2}{r^6}.$$

Therefore, the curvature of the Schwarzschild solution is singular at $r = 0$. As such, we cannot extend the metric in a smooth way to include the set $r = 0$.

In order to determine whether the Schwarzschild solution is geodesically complete or not, we consider time-like geodesics that are ingoing (i.e. $\frac{dr}{ds} \le 0$, where s is an affine parameter along the geodesic) and radial (i.e. $\frac{d\theta}{ds} = \frac{d\phi}{ds} = 0$, where (θ, ϕ) are coordinates on S^2). The geodesic equations then imply that

$$\left(1 - \frac{2m}{r}\right) = E,$$

$$-1 = -\frac{E^2}{\left(1 - \frac{2m}{r}\right)} + \left(1 - \frac{2m}{r}\right)^{-1} \left(\frac{dr}{ds}\right)^2,$$

where E is a constant. The second equation implies that

$$\left(\frac{dr}{ds}\right)^2 = (E^2 - 1) + \frac{2m}{r}.$$

In the case $E^2 - 1 > 0$, this implies that

$$\frac{dr}{ds} = -\left((E^2 - 1) + \frac{2m}{r}\right)^{1/2}.$$

Therefore, letting $r(s = 0) := r_0 > 0$, we have

$$s = -\int_{r_0}^{r(s)} \frac{\sqrt{r}dr}{\left((E^2 - 1) + \frac{2m}{r}\right)^{1/2}}.$$

Since the denominator is greater than or equal to $\left(E^2 - 1\right)^{1/2}$ and hence bounded away from zero, and the numerator is bounded, it follows that $r(s_0) = 0$, where

$$s_0 = -\int_{r_0}^{0} \frac{\sqrt{r}dr}{\left((E^2 - 1) + \frac{2m}{r}\right)^{1/2}} < \infty.$$

As such, the radially ingoing geodesic $c : [0, s_0) \to M'$ cannot be extended (continuously) to the interval $[0, s_0]$, and is therefore incomplete.

As such, the Schwarzschild solution contains incomplete time-like geodesics and, hence, is geodesically incomplete.

For more information concerning global properties of several important exact solutions of the Einstein equations, see [HE, Chapter 5].

11. Geodesics

Lecture 16

Proposition 11.1. *Let $p \in M$ and $\mathbf{v} \in T_pM$. Then there exists a unique geodesic $\gamma_{\mathbf{v}}$ in M such that*

1. *the initial velocity of $\gamma_{\mathbf{v}}$ is $\mathbf{v}$; i.e. $\frac{d}{dt}\gamma_{\mathbf{v}}(t)\big|_{t=0} = \mathbf{v}$;*
2. *the domain $I_{\mathbf{v}}$ of $\gamma_{\mathbf{v}}$ is the largest possible. Hence if $\alpha : J \to M$ is a geodesic with initial velocity $\mathbf{v}$ then $J \subseteq I_{\mathbf{v}}$ and $\alpha = \gamma_{\mathbf{v}}|_J$.*

The geodesic $\gamma_{\mathbf{v}}$ is said to be maximal *or* geodesically inextendible.

Definition 11.2. Let $p \in M$. We define $\mathcal{D}_p \subseteq T_pM$ to be the set of vectors $\mathbf{v} \in T_pM$ such that the inextendible geodesic $\gamma_{\mathbf{v}}$ is defined at least on the interval $[0, 1]$. The exponential map of M at p is the map

$$\exp_p : \mathcal{D}_p \to M, \quad \mathbf{v} \mapsto \gamma_{\mathbf{v}}(1).$$

Gauss Lemma. *Let $p \in M$, $\mathbf{x} \in \mathcal{D}_p \setminus \{0\}$. Let $\mathbf{v_x}, \mathbf{w_x} \in T_\mathbf{x}(T_pM)$ with $\mathbf{v_x}$ radial. Then*

$$\mathbf{g}_p(\mathbf{v_x}, \mathbf{w_x}) = \mathbf{g}_{\exp_p \mathbf{x}}\left(\left(D_\mathbf{x} \exp_p\right)(\mathbf{v_x}), \left(D_\mathbf{x} \exp_p\right)(\mathbf{w_x})\right).$$

Proposition 11.3. *Let $(M, \mathbf{g})$ be a Lorentzian manifold. For each $p \in M$ there exists a neighbourhood, U_p, of $0 \in T_pM$ and a neighbourhood, V_p, of p in M such that the map $\exp_p\big|_{U_p} : U_p \to V_p$ is a diffeomorphism.*

Definition 11.4. A subset S of a vector space is *star-shaped about* 0 if $v \in S$ implies $tv \in S$ for all $t \in [0,1]$. If U_p and V_p are as above and U_p is star-shaped, then V_p is called a *normal neighbourhood* of p. An open set $C \subseteq M$ is *convex* if C is a normal neighbourhood of each of its points.

We wish to consider smooth causal (i.e. time-like and null) curves in a convex normal neighbourhood V_p of M with $\exp_p : U_p \to V_p$ as above. (Most of our considerations may be generalised to the case of piece-wise smooth curves where, if there is a discontinuity in the tangent vector $\mathbf{t}$ at time t, then we require that

$$\mathbf{g}(\mathbf{t}(t-), \mathbf{t}(t+)) < 0,$$

so that the tangent vector lies in the same half of the null cone (i.e. both are future-directed, or both are past-directed).)

The following result is, essentially, a corollary to the Gauss Lemma:

Lemma 11.5. *Let V_p be a convex normal neighbourhood of $p \in M$. The time-like geodesics through p are orthogonal to the surfaces of constant σ (with $\sigma < 0$) where, for $q \in V_p$, $\sigma(q) := \mathbf{g}_p(\exp_p^{-1} q, \exp_p^{-1} q)$.*

Sketch of Proof. This is similar to the proof of the Gauss Lemma. Let $\mathbf{X}(s)$, $s \in (-\epsilon, \epsilon)$ be a smooth curve in T_pM with $\mathbf{g}(\mathbf{X}(s), \mathbf{X}(s)) = -1$, and $\alpha(t,s) := \exp_p(t\mathbf{X}(s))$. Then $\sigma(\alpha(t,s)) = -t^2$. We need to show that $\mathbf{g}(\mathbf{T}, \mathbf{V}) = 0$. However,

$$\mathbf{T}\mathbf{g}(\mathbf{T}, \mathbf{V}) = \nabla_\mathbf{T}\mathbf{g}(\mathbf{T}, \mathbf{V}) = \mathbf{g}(\mathbf{T}, \nabla_\mathbf{T}\mathbf{V}) = \frac{1}{2}\nabla_\mathbf{V}\mathbf{g}(\mathbf{T}, \mathbf{T}) = 0.$$

Remark 11.6. This implies that the surfaces $\sigma =$ constant, where constant < 0, are space-like. A hyper-surface with normal vector $\mathbf{n}$ is

- *space-like* if its normal is time-like (i.e. $\mathbf{g}(\mathbf{n}, \mathbf{n}) < 0$);
- *time-like* if its normal is spac-like (i.e. $\mathbf{g}(\mathbf{n}, \mathbf{n}) > 0$);
- *null* if its normal is null (i.e. $\mathbf{g}(\mathbf{n}, \mathbf{n}) = 0$).

Example 11.7. Consider the surfaces $r =$ constant in the Schwarzschild solution in the coordinate system (u, r, θ, ϕ). The normal one-form to the surface $r =$ constant can be taken as $\boldsymbol{\sigma} = dr$, for which the corresponding normal vector field is $\mathbf{n} = \left(1 - \frac{2m}{r}\right)\partial_r + \partial_u$ (i.e. $\mathbf{g}'(\partial_u, \cdot) = dr$). Then $\mathbf{g}(\mathbf{n}, \mathbf{n}) = \langle dr, \mathbf{n}\rangle = \left(1 - \frac{2m}{r}\right)$. Therefore the surface $r = r_0$ is time-like for $r_0 > 2m$, space-like for $r_0 < 2m$ and null for $r_0 = 2m$.

For a convex normal neighbourhood $V_p \subseteq M$ the intrinsic causality is the same as that of Minkowski space. In particular, we have

Proposition 11.8. *Let V_p be a convex normal neighbourhood of $p \in M$. Then the points that can be reached from p by time-like curves in V_p are those of the form $\exp_p \mathbf{v}$ where $\mathbf{v} \in T_pM$ obeys $\mathbf{g}(\mathbf{v}, \mathbf{v}) < 0$.*

Sketch of Proof. Let $C_p \subset T_pM$ denote the set of time-like vectors at p. (i.e. the interior of the null cone in T_pM with vertex $0 \in T_pM$.) Let $q \in V_p$ be a point that can be reached from p by a time-like curve, γ, with $\gamma(0) = p$. Let $\overline{\gamma} := \exp_p^{-1} \circ \gamma$ be the corresponding curve in T_pM. Then $\gamma'(0) \in T_pM$, $\overline{\gamma}'(0) \in T_0(T_pM)$. Identifying the tangent spaces $T_x(T_pM)$ with T_pM in the usual way and using the fact that $D_0 \exp_p$ is then the identity map, we have

$$\gamma'(0) = \overline{\gamma}'(0).$$

Lecture 17 Therefore, $\overline{\gamma}'(0)$ is a time-like vector at $0 \in T_pM$. Hence the curve $\overline{\gamma}$ enters the set C_p.

Note that $(\exp_p C_p) \cap V_p$ is the subset of V_p consisting of points r for which $\sigma(r) < 0$. Moreover, the surfaces $\sigma =$ constant are spacelike. Since γ is time-like, it therefore cannot be tangent to the sets $\sigma =$ constant, so σ must decrease monotonically along γ. Therefore $\sigma(q) < 0$, and hence $q \in \exp_p C_p$. Therefore there exists $\mathbf{v} \in C_p$ such that $q = \exp_p \mathbf{v}$.

A limiting argument gives the following, stronger result:

Proposition 11.9. *Let V_p be a convex normal neighbourhood of $p \in M$. Then the points that can be reached from p by time-like (resp. causal) curves in V_p are those of the form $\exp_p \mathbf{v}$ where $\mathbf{v} \in T_pM$ obeys $\mathbf{g}(\mathbf{v}, \mathbf{v}) < 0$ (resp. $\mathbf{g}(\mathbf{v}, \mathbf{v}) \leq 0$).*

Sketch of Proof. The only case not covered by the previous result is where $q \in V_p$ can be reached from p by a causal curve, but no time-like curve. In this case, we perturb the causal curve (λ, say), and define an approximating family of time-like curves λ_i with endpoints p and q_i, where $q_i \to q$ as $i \to \infty$. Using continuous dependence of solutions of the geodesic equations on initial conditions (or just continuity of the exponential map), it then follows that $q \in \overline{\exp C_p} = \exp \overline{C_p}$. Since $\overline{C_p} = \{\mathbf{v} \in T_pM : \mathbf{g}(\mathbf{v}, \mathbf{v}) \leq 0\}$, we are done.

Remark 11.10. In particular, if $q \in V_p$ can be reached from p by a causal curve but not by any time-like curve, then we must have $q = \exp_p \mathbf{v}$ where $\mathbf{v} \in T_pM$ is null (i.e. $\mathbf{g}(\mathbf{v}, \mathbf{v}) = 0$). In this case $q = \exp_p \mathbf{v}$ lies on a null geodesic $t \mapsto \exp_p(t\mathbf{v})$ through p. To summarise, any point in V_p that can be reached from p by a causal curve but not by any time-like curve must lie on a null geodesic through p.

12. Lengths of causal curves

Definition 12.1. Given a causal curve, $c : [a, b] \to M$, from p to q, we define the *length* (or *arc-length*) of c to be

$$L[c] := \int_a^b \left(-\mathbf{g}(c'(t), c'(t))\right)^{1/2} dt. \tag{12.1}$$

(If the curve is only piece-wise smooth, we take the integral over the differentiable sections of the curve.)

Remark 12.2. More generally, given a piecewise smooth curve $c : [a, b] \to M$, we can define the arc-length of c to be

$$L[c] := \int_a^b |c'(t)|\, dt.$$

where $|c'(t)| := |\mathbf{g}(c'(t), c'(t))|$. If we restrict to curves that are either (piecewise) causal (i.e. $\mathbf{g}(c'(t), c'(t)) \leq 0$ for all $t \in [a, b]$) or (piecewise) spacelike (i.e. $\mathbf{g}(c'(t), c'(t)) > 0$ for all $t \in [a, b]$), then we can define $|c'(t)| := \left(\epsilon \mathbf{g}(c'(t), c'(t))\right)^{1/2}$, where $\epsilon = \operatorname{sign} \mathbf{g}(c'(t), c'(t))$.

Since we are interested only in the causal case, we simply work with the the definition (12.1).

Example 12.3. In flat Minkowski space, $M := \mathbb{R}^{n,1}$, let $p = \{t = 0, x^1 = \cdots = x^n = 0\}$, $q = \{t = 1, x^1 = \cdots = x^n = 0\}$. Then there exists a timelike geodesic $c : [0, 1] \to M : s \to (s, 0, \ldots, 0)$ from p to q of length

$$L[c] = \int_0^1 \sqrt{-1 \cdot -1} dt = 1.$$

If we allow piecewise smooth curves, then there exists a null curve also joining p to q of the form

$$\tilde{c}(s) = \begin{cases} (s, s, 0, \ldots, 0) & s \in [0, \frac{1}{2}], \\ (s, 1 - s, 0, \ldots, 0) & s \in (\frac{1}{2}, 1]. \end{cases}$$

In this case, since the curve is null, its length is

$$L[\tilde{c}] = \int_0^{1/2} \sqrt{-1 \cdot 0} dt + \int_{1/2}^1 \sqrt{-1 \cdot 0} = 0 < L[c].$$

Therefore, unlike in the Riemannian case, geodesics are generally not curves of minimal length in Lorentzian geometry. The following proposition shows that essentially the opposite is true.

Proposition 12.4. *Let $p \in M$ and q lie in a convex normal neighbourhood $V_p \subseteq M$ of p. Then, if p and q can be joined by a timelike curve in V_p, the longest such curve is the (unique) timelike geodesic from p to q in V_p.*

Proof. Let $q \in V_p$ be such that there exists a timelike curve from p to q. By the previous result, $q = \exp_p \mathbf{v}$ with $\mathbf{v} \in T_pM$ timelike, and the length of this curve is

$$L[\gamma_{\mathbf{v}}] = \int_0^1 \sqrt{-\mathbf{g}(\gamma_{\mathbf{v}}', \gamma_{\mathbf{v}}')}dt = \sqrt{-\mathbf{g}(\exp_p^{-1}(q), \exp_p^{-1}(q))} = \sqrt{-\sigma(q)}.$$

Now let $\lambda : [\alpha, \beta] \to V_p; s \mapsto \lambda(s)$ be any timelike curve from p to q that lies in V_p, with length

$$L[\lambda] = \int_\alpha^\beta \sqrt{-\mathbf{g}(\lambda'(s), \lambda'(s))}ds$$

Given $\mathbf{X} : [\alpha, \beta] \to T_pM$ with $\mathbf{g}(\mathbf{X}(s), \mathbf{X}(s)) = -1$, we define

$$\alpha(t, s) := \exp_p (t\mathbf{X}(s)) .$$

Then, given the curve λ, there exists such an $\mathbf{X}$ and a function $s \mapsto f(s)$ such that

$$\lambda(s) = \alpha(f(s), s).$$

We then have

$$\lambda'(s) = f'(s)\mathbf{T} + \mathbf{V}.$$

Since $\mathbf{T}$ is tangent to time-like geodesics, and $\mathbf{V}$ is tangent to surfaces $\sigma =$ constant, it follows from Lemma 11.5 that $\mathbf{g}(\mathbf{T}, \mathbf{V}) = 0$. Moreover, the surfaces $\sigma =$ constant are space-like, so $\mathbf{g}(\mathbf{V}, \mathbf{V}) \geq 0$, with equality if and only if $\mathbf{V} = 0$. Finally, $\mathbf{T}$ is tangent to (affinely-parametrised) geodesics, therefore $\mathbf{g}(\mathbf{T}, \mathbf{T}) = \mathbf{g}_p(\mathbf{X}(s), \mathbf{X}(s)) = -1$. Putting these relations together, we therefore find that

$$\begin{aligned}
-\mathbf{g}(\lambda', \lambda') &= -\mathbf{g}(f'(s)\mathbf{T} + \mathbf{V}, f'(s)\mathbf{T} + \mathbf{V}) \\
&= -f'(s)^2\mathbf{g}(\mathbf{T}, \mathbf{T}) - \mathbf{g}(\mathbf{V}, \mathbf{V}) \\
&= -\mathbf{g}(\mathbf{V}, \mathbf{V}) + f'(s)^2 \\
&\leq f'(s)^2,
\end{aligned}$$

with equality if and only if $\mathbf{V} = 0$. Hence

$$L[\lambda] = \int_\alpha^\beta \sqrt{-\mathbf{g}(\lambda'(s), \lambda'(s))}ds \leq \int_\alpha^\beta f'(s)ds = f(\beta) - f(\alpha).$$

However,

$$\sigma(\lambda(s)) = \sigma(\exp_p(f(s)\mathbf{X}(s))) = \|f(s)\mathbf{X}(s)\|^2 = -f(s)^2,$$

so $f(s) = \sqrt{-\sigma(\lambda(s))}$. Hence $f(\alpha) = 0$ and $f(\beta) = \sqrt{-\sigma(q)}$, so

$$L[\lambda] \leq \sqrt{-\sigma(q)} = L[\gamma_{\mathbf{v}}].$$

As remarked above, for equality we require $\mathbf{V} = 0$. In this case, λ is also a geodesic. Since $\exp_p : U_p \to V_p$ is a diffeomorphism, the geodesic from p to q lying in V_p is unique (i.e. it is $\gamma_{\mathbf{v}}$), and therefore, up to reparametrisation, λ must coincide with $\gamma_{\mathbf{v}}$. □

Lecture 18

Corollary 12.5. *Let $p \in M$ and q lie in a convex normal neighbourhood $V_p \subseteq M$ of p. Then, if p and q can be joined by a causal curve in V_p, the longest such curve is the (unique) causal geodesic from p to q in V_p.*

Proof. The previous result covers the case when p and q can be joined by a timelike curve. If p and q can be joined by a causal curve, but not any timelike curve, then q lies on a (unique) null geodesic, c, through p. Then c is the only causal curve from p to q, so $L[c] = 0$ is maximal. □

13. Variation of arc-length

Since the initial results here are, up to some minus signs, the same as in the Riemannian case, we simply give a brief summary of how these results are adapted to the Lorentzian setting.

The notation is as before: given a smooth time-like curve $c : [a,b] \to M$ we define the (one-parameter) smooth variation $\alpha : [a,b] \times (-\epsilon, \epsilon) \to M$ where we now assume that the curves $c_s : [a,b] \to M; t \to c_s(t) := \alpha(t,s)$ are time-like. We define the vector fields along α:

$$\mathbf{T} := (D\alpha)\left(\frac{\partial}{\partial t}\right), \qquad \mathbf{V} := (D\alpha)\left(\frac{\partial}{\partial s}\right),$$

in terms of which we now require that $\mathbf{g}(\mathbf{T},\mathbf{T}) < 0$, for all $(t,s) \in [a,b] \times (-\epsilon,\epsilon)$.

13.1. First variation of arc-length.

Lemma 13.1 (First variation formula (Lorentzian))**.** *Let $c : [a,b] \to M$ be a smooth, timelike curve segment, parametrised by arc-length (i.e. $|c'| \equiv \sqrt{-\mathbf{g}(c'(t), c'(t))} = 1$, for all $t \in [a,b]$). Let $\alpha : [a,b] \times (-\epsilon,\epsilon) \to M$ be a smooth variation of α as above. Then*

$$\frac{d}{ds} L[c_s]\bigg|_{s=0} = -\,\mathbf{g}(\mathbf{t},\mathbf{v})\Big|_a^b + \int_a^b \mathbf{g}(\nabla_{\mathbf{t}}\mathbf{t}, \mathbf{v})\, dt.$$

Proof. As in Riemannian case, but remember the minus sign in $|\mathbf{T}|$. □

Corollary 13.2. *A smooth time-like curve $c : [a,b] \to M$ of constant speed $|c'| = 1$ is a critical point of the arc-length functional under fixed end-point variations if and only if it is a geodesic.*

We may also consider a timelike curve $c : [a,b] \to M$ from a spacelike three-surface S to a point p. In this case, we consider variations with the property that $\alpha(a,s) \in S, \alpha(b,s) = p$ for all $s \in (-\epsilon,\epsilon)$. Therefore the variation vector field $\mathbf{v}(a)$ lies in $T_{c(a)}S$.

Corollary 13.3. *Let S be a space-like submanifold of M without boundary and $p \in M \setminus S$ such that there exists a time-like curve from S to p. A necessary condition for a curve $c : [a,b] \to M$ with $c(a) \in S$, $c(b) = p$ to be the longest time-like curve from S to p is that c is a time-like geodesic, and $c'(a)$ is orthogonal to $T_{c(a)}S$.*

Proof. The proof is essentially the same as that of Proposition 3.7 in the Riemannian case. In the Lorentzian case, let $\mathbf{x} \in T_{c(a)}S$ with $\mathbf{g}_{c(a)}(c'(a), \mathbf{x}) > 0$. The corresponding variation has

$$\frac{d}{ds} L[c_s]\bigg|_{s=0} = \mathbf{g}(c'(a), \mathbf{x}) > 0.$$

Hence, for sufficiently small $s > 0$, we have $L[c_s] > L[c]$, so c cannot be maximising. □

13.2. Second variation of arc-length.

As before, for a two-parameter smooth variation of c, $\alpha : [a,b] \times (-\delta,\delta) \times (-\epsilon,\epsilon) \to M$, we define the vector fields along α:

$$\mathbf{T} := D\alpha\left(\frac{\partial}{\partial t}\right), \qquad \mathbf{V} := D\alpha\left(\frac{\partial}{\partial r}\right), \qquad \mathbf{W} := D\alpha\left(\frac{\partial}{\partial s}\right).$$

We still assume that the variation is time-like i.e. $\mathbf{g}(\mathbf{T},\mathbf{T}) < 0$, for all $(t,r,s) \in [a,b] \times (-\delta,\delta) \times (-\epsilon,\epsilon)$.

Lemma 13.4. *Let $c : [a,b] \to M$ be a time-like geodesic parametrised by arc-length (i.e. $|c'| \equiv \sqrt{-\mathbf{g}(c'(t), c'(t))} = 1$, for all $t \in [a,b]$). If α is a variation of c as above then*

$$\begin{aligned}
\frac{\partial^2 L}{\partial r \partial s}\bigg|_{r=s=0} &= -\,\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{w}}\mathbf{v})|_a^b - \int_a^b \Big[\mathbf{g}((\nabla_{\mathbf{t}}\mathbf{v})^{\perp}, (\nabla_{\mathbf{t}}\mathbf{w})^{\perp}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{w},\mathbf{t})\mathbf{v})\Big]\, dt, \\
&= -\,\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{v}}\mathbf{w})|_a^b - \int_a^b \Big[\mathbf{g}((\nabla_{\mathbf{t}}\mathbf{v})^{\perp}, (\nabla_{\mathbf{t}}\mathbf{w})^{\perp}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{v},\mathbf{t})\mathbf{w})\Big]\, dt,
\end{aligned}$$

where, given any vector field $\mathbf{X}$ *along* c*, we define*

$$\mathbf{X}^{\perp} := \mathbf{X} - \frac{\mathbf{g}(\mathbf{X}, \mathbf{t})}{\mathbf{g}(\mathbf{t}, \mathbf{t})}\mathbf{t} = \mathbf{X} + \mathbf{g}(\mathbf{X}, \mathbf{t})\mathbf{t},$$

the component of $\mathbf{X}$ *orthogonal to* $\mathbf{t} \equiv c'$.

Remark 13.5. Note that

$$(\nabla_{\mathbf{t}}\mathbf{v})^{\perp} = \nabla_{\mathbf{t}}\mathbf{v} - \frac{\mathbf{g}(\nabla_{\mathbf{t}}\mathbf{v}, \mathbf{t})}{\mathbf{g}(\mathbf{t}, \mathbf{t})}\mathbf{t} = \nabla_{\mathbf{t}}\left(\mathbf{v} - \frac{\mathbf{g}(\mathbf{v}, \mathbf{t})}{\mathbf{g}(\mathbf{t}, \mathbf{t})}\mathbf{t}\right) = \nabla_{\mathbf{t}}\left(\mathbf{v}^{\perp}\right).$$

Therefore the integrand in the second variation depends only on $\mathbf{v}^{\perp}$ and $\mathbf{w}^{\perp}$, the components of $\mathbf{v}$ and $\mathbf{w}$ orthogonal to $\mathbf{t}$.

As in the Riemannian case, we define the index form by

$$I(\mathbf{v}, \mathbf{w}) := -\int_a^b \left[\mathbf{g}(\nabla_{\mathbf{t}}\mathbf{v}, \nabla_{\mathbf{t}}\mathbf{w}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{v}, \mathbf{t})\mathbf{w})\right] dt = \int_a^b \left[\mathbf{g}(\mathbf{v}, \nabla_{\mathbf{t}}\nabla_{\mathbf{t}}\mathbf{w} + \mathbf{R}(\mathbf{w}, \mathbf{t})\mathbf{t})\right] dt$$

for $\mathbf{v}, \mathbf{w} \in \Gamma_0(c, TM) := \left\{\mathbf{X} \in \Gamma(c, TM) \,\middle|\, \mathbf{X}(a) = 0, \mathbf{X}(b) = 0, \mathbf{g}_{c(t)}(\mathbf{X}(t), c'(t)) = 0, \forall t \in [a, b]\right\}$. As in the Riemannian case, the bilinear form $I : \Gamma_0(c, TM) \times \Gamma_0(c, TM) \to \mathbb{R}$ is degenerate if and only if there exists a Jacobi field along c that vanishes at the endpoints of c.

Definition 13.6. A timelike geodesic, c, from p to q is *maximal* if the index form is negative semi-definite (i.e. $I(\mathbf{v}, \mathbf{v}) \leq 0$, for all $\mathbf{v} \in \Gamma_0(c, TM)$).

Remarks 13.7.

(1) Equivalently, c is maximal if $\frac{d^2}{ds^2}L[c_s]\Big|_{s=0} \leq 0$ for any variation, α, of c.
(2) A geodesic, c, is not maximal if there exists a small variation of c that yields a longer curve between p and q.

Proposition 13.8. *A timelike geodesic curve,* c*, from* p *to* q *is maximal if and only if there is no point conjugate to* p *along* c *between* p *and* q*.*

Proof. In the Riemannian case, we showed that if there exists a point along c, before q, that is conjugate to p along c, then the index form at q is not positive semi-definite (cf. Proposition 7.14). Adapting the same proof to the Lorentzian case, we deduce that if $c : [a, b] \to M$ with $c(a) = p$, $c(b) = q$ and there exists $t_0 \in (a, b)$ such that $c(t_0)$ conjugate to p along c, then the index form at q is not negative semi-definite. Taking the contra-positive, if the index form at q is negative semi-definite (i.e. c is maximal) then there exists no point along c conjugate to p before q.

In the reverse direction, we first prove that if there are no points conjugate to p up to and including q, then the index form at q is negative definite.

Therefore, assume that any (non-trivial) Jacobi field along c that is zero at $c(a)$ is non-zero at $c(t)$, for $t \leq b$. Let $\{\mathbf{e}_1, \ldots, \mathbf{e}_n\}$ be a basis for the orthogonal complement, $c'(a)^{\perp}$, of $c'(a)$ in $T_{c(a)}M$. We then define $\mathbf{E}_i$ to be the Jacobi fields along c with initial conditions

$$\mathbf{E}_i(a) = 0, \qquad \mathbf{E}_i'(a) = \mathbf{e}_i.$$

($'$ will denote $\nabla_{\mathbf{t}}$, where $\mathbf{t} := c'$ is the tangent vector to c.) Since these are (non-trivial) Jacobi fields along c that vanish at p, they are then non-zero at $c(t)$ for all $t \in (a, b]$.

Claim 1: $\mathbf{g}(\mathbf{t}, \mathbf{E}_i) = 0$ along c.

Proof. Taking derivatives, we have

$$\nabla_{\mathbf{t}}^2\mathbf{g}(\mathbf{t}, \mathbf{E}_i) = \mathbf{g}(\mathbf{t}, \nabla_{\mathbf{t}}^2\mathbf{E}_i) = -\mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{E}_i, \mathbf{t})\mathbf{t}) = 0.$$

Therefore $\nabla_{\mathbf{t}}\mathbf{g}(\mathbf{t}, \mathbf{E}_i) = \mathbf{g}(\mathbf{t}, \nabla_{\mathbf{t}}\mathbf{E}_i) = \text{constant} = 0$, since $\mathbf{g}(\mathbf{t}, \nabla_{\mathbf{t}}\mathbf{E}_i) = 0$ at $c(a)$. Therefore $\mathbf{g}(\mathbf{t}, \mathbf{E}_i)$ is constant. Since $\mathbf{E}_i(a) = 0$, we have $\mathbf{g}(\mathbf{t}, \mathbf{E}_i) = 0$ along c □

Hence the $\mathbf{E}_i$ are orthogonal to c' along c. As such $\mathbf{E}_i$ form a basis for vector fields along c that vanish at p and are orthogonal to c'. Given any vector field, $\mathbf{V} \in \Gamma_0(c, TM)$, there exist functions f^i (with $f^i(q) = 0$) with the property that

$$\mathbf{V} = \sum_{i=1}^{n} f^i \mathbf{E}_i.$$

To calculate $I(\mathbf{V}, \mathbf{V})$, we need to calculate

$$\mathbf{g}(\nabla_{\mathbf{t}}\mathbf{V}, \nabla_{\mathbf{t}}\mathbf{V}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{V}, \mathbf{t})\mathbf{V}).$$

We write

$$\nabla_{\mathbf{t}}\mathbf{V} = \mathbf{A} + \mathbf{B},$$

where

$$\mathbf{A} := \sum \left(f^i\right)' \mathbf{E}_i, \qquad \mathbf{B} := \sum f^i \mathbf{E}_i'.$$

Claim 2:

$$\mathbf{g}(\nabla_{\mathbf{t}}\mathbf{E}_i, \mathbf{E}_j) = \mathbf{g}(\mathbf{E}_i, \nabla_{\mathbf{t}}\mathbf{E}_j)$$

along c.

Proof.

$$\nabla_{\mathbf{t}}\left(\mathbf{g}(\nabla_{\mathbf{t}}\mathbf{E}_i, \mathbf{E}_j) - \mathbf{g}(\mathbf{E}_i, \nabla_{\mathbf{t}}\mathbf{E}_j)\right) = -\mathbf{g}(\mathbf{R}(\mathbf{E}_i, \mathbf{t})\mathbf{t}, \mathbf{E}_j) + \mathbf{g}(\mathbf{R}(\mathbf{E}_j, \mathbf{t})\mathbf{t}, \mathbf{E}_i) = 0.$$

Therefore $\mathbf{g}(\nabla_{\mathbf{t}}\mathbf{E}_i, \mathbf{E}_j) - \mathbf{g}(\mathbf{E}_i, \nabla_{\mathbf{t}}\mathbf{E}_j)$ is constant along c. Since $\mathbf{E}_i$ vanishes at p, the result follows. ☐

Then

$$\begin{aligned}
\mathbf{g}(\mathbf{B}, \mathbf{B}) &= \sum_{i,j} f^i f^j \mathbf{g}(\mathbf{E}_i', \mathbf{E}_j') \\
&= \sum_{i,j} f^i f^j \left(\mathbf{g}(\mathbf{E}_i', \mathbf{E}_j)' - \mathbf{g}(\mathbf{E}_i'', \mathbf{E}_j)\right) \\
&= \left(\sum_{i,j} f^i f^j \mathbf{g}(\mathbf{E}_i', \mathbf{E}_j)\right)' - \sum_{i,j} \left(f^i f^j\right)' \mathbf{g}(\mathbf{E}_i', \mathbf{E}_j) + \sum_{i,j} f^i f^j \mathbf{g}(\mathbf{R}(\mathbf{E}_i, \mathbf{t})\mathbf{t}, \mathbf{E}_j) \\
&= (\mathbf{g}(\mathbf{B}, \mathbf{V}))' - \sum_{i,j} \left((f^i)' f^j \mathbf{g}(\mathbf{E}_i, \mathbf{E}_j') + f^i (f^j)' \mathbf{g}(\mathbf{E}_i', \mathbf{E}_j)\right) + \mathbf{g}(\mathbf{R}(\mathbf{V}, \mathbf{t})\mathbf{t}, \mathbf{V}) \\
&= (\mathbf{g}(\mathbf{B}, \mathbf{V}))' - 2\mathbf{g}(\mathbf{A}, \mathbf{B}) + \mathbf{g}(\mathbf{R}(\mathbf{V}, \mathbf{t})\mathbf{t}, \mathbf{V}),
\end{aligned}$$

where we have used Claim 2 in the fourth equality. **Lecture 19**

Hence

$$\begin{aligned}
\mathbf{g}(\nabla_{\mathbf{t}}\mathbf{V}, \nabla_{\mathbf{t}}\mathbf{V}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{V}, \mathbf{t})\mathbf{V}) &= \mathbf{g}(\mathbf{A}, \mathbf{A}) + 2\mathbf{g}(\mathbf{A}, \mathbf{B}) + \mathbf{g}(\mathbf{B}, \mathbf{B}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{V}, \mathbf{t})\mathbf{V}) \\
&= \mathbf{g}(\mathbf{A}, \mathbf{A}) + (\mathbf{g}(\mathbf{B}, \mathbf{V}))'.
\end{aligned}$$

We therefore have

$$I(\mathbf{V}, \mathbf{V}) = -\left.\mathbf{g}(\mathbf{B}, \mathbf{V})\right|_a^b - \int \mathbf{g}(\mathbf{A}, \mathbf{A})dt = -\int \mathbf{g}(\mathbf{A}, \mathbf{A})dt,$$

since $\mathbf{V}(p) = \mathbf{V}(q) = 0$. Since $\mathbf{A}$ is space-like, it follows that $I(\mathbf{V}, \mathbf{V}) \leq 0$. For equality, we require $\mathbf{A} = 0$, which implies f^i are constant, hence zero. Therefore $I(\mathbf{V}, \mathbf{V}) \leq 0$ with equality if and only if $\mathbf{V} = 0$. Therefore I is negative definite.

Finally, if q is the first conjugate point to p along c, then the above implies that the index form on the interval $[a, t]$ is negative definite for all $t \in (a, b)$. Given $\mathbf{V} \in \Gamma_0(c, TM)$ and a sequence $\epsilon_i > 0$ with $\epsilon_i \to 0$, we construct the vector field $\mathbf{V}_i$ along the curve c by defining

$$\mathbf{V}_i(t) := \begin{cases} \frac{(b-\epsilon_i)-t}{b-t}\mathbf{V}(t) & t \in [a, b-\epsilon_i], \\ 0 & t \in (b-\epsilon_i, b]. \end{cases}$$

Then $I(\mathbf{V}_i, \mathbf{V}_i) < 0$ and converge to $I(\mathbf{V}, \mathbf{V})$. Hence $I(\mathbf{V}, \mathbf{V}) \leq 0$ for all $\mathbf{V} \in \Gamma_0(c, TM)$, so I is negative semi-definite. □

Let $S \subset M$ be a space-like hyper-surface (i.e. submanifold of codimension 1.) We will denote its unit (time-like) normal vector field by $\mathbf{n} \in \Gamma(S, TM)$, with $\mathbf{g}(\mathbf{n}, \mathbf{n}) = -1$. Given vector fields tangent to $\mathbf{S}$ (i.e. orthogonal to $\mathbf{n}$), we define the second-fundamental form/extrinsic curvature[13]

$$\chi(\mathbf{v}, \mathbf{w}) := \mathbf{g}(\nabla_{\mathbf{v}} \mathbf{n}, \mathbf{w}).$$

Since S is a submanifold, it follows that $[\mathbf{v}, \mathbf{w}]$ is tangent to S which, in turn, implies that χ is symmetric. We also define the trace of χ:

$$\kappa := \operatorname{tr} \chi = \sum_{i=1}^{n} \chi(\mathbf{e}_i, \mathbf{e}_i),$$

where $\{\mathbf{e}_i\}$ is any orthonormal basis for $T_p S$.

Given that a geodesic c from space-like hypersurface S to point q is a critical point of the arc-length functional if (and only if) if it is normal to S (i.e. $c'(a)$ is orthogonal to $T_{c(a)} S$), we only consider the second variation for such geodesics.

Lemma 13.9. *Let $c : [a, b] \to M$ be a time-like geodesic, parametrised by arc-length, normal to spacelike three-surface S. If α is a variation of c with $\alpha(a, r, s) \in S$, $\alpha(b, r, s) = c(b)$ for all $(r, s) \in (-\delta, \delta) \times (-\epsilon, \epsilon)$ and the variation vectors $\mathbf{v}$, $\mathbf{w}$ orthogonal to $\mathbf{t} \equiv c'$, then*

$$\left. \frac{\partial^2 L}{\partial r \partial s} \right|_{r=s=0} = -\chi(\mathbf{v}(a), \mathbf{w}(a)) - \int_a^b \left[\mathbf{g}(\nabla_{\mathbf{t}} \mathbf{v}, \nabla_{\mathbf{t}} \mathbf{w}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{v}, \mathbf{t}) \mathbf{w}) \right] dt,$$

where χ is the second fundamental form of S.

Proof. We simply need to calculate the endpoint term in the second variation formula. We have

$$\begin{aligned} - \left. \mathbf{g}(\mathbf{t}, \nabla_{\mathbf{v}} \mathbf{w}) \right|_a^b = \left. \mathbf{g}(\mathbf{t}, \nabla_{\mathbf{v}} \mathbf{w}) \right|_a &= \nabla_{\mathbf{v}(a)} \mathbf{g}(\mathbf{t}(a), \mathbf{w}(a)) - \mathbf{g}(\nabla_{\mathbf{v}(a)} \mathbf{t}(a), \mathbf{w}(a)) = 0 - \chi(\mathbf{v}(a), \mathbf{w}(a)) \\ &= -\chi(\mathbf{v}(a), \mathbf{w}(a)), \end{aligned}$$

as required. □

Definition 13.10. We define the index form by

$$I_S(\mathbf{v}, \mathbf{w}) := -\chi(\mathbf{v}(a), \mathbf{w}(a)) - \int_a^b \left[\mathbf{g}(\nabla_{\mathbf{t}} \mathbf{v}, \nabla_{\mathbf{t}} \mathbf{w}) + \mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{v}, \mathbf{t}) \mathbf{w}) \right] dt$$

for $\mathbf{v}, \mathbf{w} \in \Gamma_0^S(c, TM) := \left\{ \mathbf{X} \in \Gamma(c, TM) \,\middle|\, \mathbf{X}(b) = 0, \mathbf{g}_{c(t)}(\mathbf{X}(t), c'(t)) = 0, \forall t \in [a, b] \right\}$.

Definition 13.11. A timelike geodesic curve from S to p normal to S is *maximal* if $I_S(\mathbf{v}, \mathbf{v}) \leq 0$, for all $\mathbf{v} \in \Gamma_0^S(c, TM)$. (Equivalently, $\left. \frac{d^2 L}{ds^2} \right|_{s=0} \leq 0$ for all variations with $c_s(a) \in S$, $c(b) = q$.)

Given a geodesic, $c : [a, b] \to M$, with $c(a) \in S$, $c(b) = q$, we wish to consider variations, α, of c with initial endpoints in S. As such, we restrict to variations with $\mathbf{V}(a, 0, 0)$ tangent to S. Moreover, we wish to have $\mathbf{T}(a, r, s) = \mathbf{n}(\alpha(a, r, s))$ the unit normal to S. Since $\nabla_{\mathbf{T}} \mathbf{V} = \nabla_{\mathbf{V}} \mathbf{T}$, this implies that we require

$$\left. \nabla_{\mathbf{T}} \mathbf{V} \right|_{r=s=0, t=a} = \mathbf{v}'(a) = \left. \nabla_{\mathbf{V}} \mathbf{T} \right|_{r=s=0, t=a} = \nabla_{\mathbf{v}(a)} \mathbf{n}.$$

Hence, given any tangential vector $\mathbf{w} \in T_{c(a)} S$, we have

$$\mathbf{g}(\mathbf{w}, \mathbf{v}'(a)) = \mathbf{g}(\mathbf{w}, \nabla_{\mathbf{v}(a)} \mathbf{n}) = \chi(\mathbf{w}, \mathbf{v}(a)). \tag{13.1}$$

As such, if we are considering variations through geodesics normal to S, then the natural boundary conditions to impose are $\mathbf{v}(a) \in T_{c(a)} S$ and (13.1).

---

[13]In more general contexts, one defines the second fundamental form of a submanifold $S \subset M$ to be the map $II : \mathfrak{X}(S) \times \mathfrak{X}(S) \to \mathfrak{X}(S)^{\perp}$ defined by

$$\mathbf{II}(\mathbf{v}, \mathbf{w}) = \text{normal component of } \nabla_{\mathbf{v}} \mathbf{w}.$$

In the current context, this carries the same information as our object χ.

Definition 13.12. A S-Jacobi field along normal geodesic c is a Jacobi field along c, $\mathbf{J}$, with

1. $\mathbf{J}(a)$ is tangent to S;
2. $\mathbf{g}(\mathbf{w}, \mathbf{J}'(a)) = \chi(\mathbf{w}, \mathbf{J}(a))$ for all $\mathbf{w} \in T_{c(a)}S$.

A point $c(t)$, for $t \in (a, b]$, is conjugate to S along c (or a *focal point of S along c*) if there exists a non-zero S-Jacobi field, $\mathbf{J}$, along c with $\mathbf{J}(t) = 0$.

Proposition 13.13. *A timelike normal geodesic, c, from S to q is maximal if and only if there is no point conjugate to S along c before q.*

Proof. As for 13.8. □

Lecture 20

Proposition 13.14. *Let $(M, \mathbf{g})$ be a Lorentzian manifold such that $\mathbf{Ric}(\mathbf{V}, \mathbf{V}) \geq 0$ for all time-like vector fields $\mathbf{V}$. Let S be a spacelike hypersurface in M and $p \in S$. Let c be the geodesic through p with $c(0) = p$, $c'(0) = \mathbf{n}(p)$. Then, if $\kappa(p) < 0$, there will be a point conjugate to S along c within a distance $b(p) := -n/\kappa(p)$ of S, provided that the geodesic c can be extended that far.*

Remark 13.15. From now on, we will often explicitly mention that results hold along geodesics with the proviso that the geodesic "can be extended that far". We are aiming to prove that various Lorentzian manifolds are geodesically incomplete. The logic of the proofs is to assume that (usually time-like or causal) geodesics can be extended arbitrarily far (i.e. that the manifold is geodesically complete in some sense) and show that, when combined with various geometrical conditions (e.g. curvature restrictions), this assumption leads to a logical contradiction. As such, we will assume that the relevant geodesics may be extended to the point where results such as Proposition 13.14 become relevant.

If, on the other hand, we were to assume from the outset that there exist geodesics that *cannot* be extended to the required value of the affine parameter, then our manifold would then, by definition, be incomplete and there would be nothing to prove.

Proof of Proposition 13.14. We show that, under the stated conditions, the index form is not negative semi-definite. Let $\mathbf{e}_i$, $i = 1, \ldots, n$ be an orthonormal basis for T_pS and parallel transport these along c to give vector fields $\mathbf{E}_i$ along c. Then let $f(t) := 1 - t/b(p)$ for $t \in [0, b(p)]$. Then $f\mathbf{E}_i \in \Gamma_0^S(c, TM)$ and

$$\begin{aligned}-I_S(f\mathbf{E}_i, f\mathbf{E}_i) &= \chi(f(0)\mathbf{E}_i(0), f(0)\mathbf{E}_i(0)) + \int_0^{b(p)} \left[\mathbf{g}(\nabla_{\mathbf{t}}(f\mathbf{E}_i), \nabla_{\mathbf{t}}(f\mathbf{E}_i)) + \mathbf{g}(\mathbf{t}, \mathbf{R}(f\mathbf{E}_i, \mathbf{t})f\mathbf{E}_i)\right] dt \\ &= \chi(\mathbf{e}_i, \mathbf{e}_i) + \int_0^{b(p)} \left[f'(t)^2 + f(t)^2\mathbf{g}(\mathbf{t}, \mathbf{R}(\mathbf{E}_i, \mathbf{t})\mathbf{E}_i)\right] dt.\end{aligned}$$

Summing over $i = 1, \ldots, n$ gives

$$\begin{aligned}\sum_{i=1}^n -I_S(f\mathbf{E}_i, f\mathbf{E}_i) &= \kappa(p) + \int_0^{b(p)} \left[nf'(t)^2 - f(t)^2\mathbf{Ric}(\mathbf{t}, \mathbf{t})\right] dt. \\ &= \kappa(p) + \int_0^{b(p)} n\frac{1}{b(p)^2} dt - \int_0^{b(p)} f(t)^2\mathbf{Ric}(\mathbf{t}, \mathbf{t})\, dt \\ &\leq \kappa(p) + \frac{n}{b(p)} = \kappa(p) - \kappa(p) = 0.\end{aligned}$$

Therefore

$$\sum_{i=1}^n I_S(f\mathbf{E}_i, f\mathbf{E}_i) \geq 0,$$

so there exists an i such that $I_S(f\mathbf{E}_i, f\mathbf{E}_i) \geq 0$. Therefore there exists $\mathbf{v} \in \Gamma_0^S(c, TM)$ with $I_S(\mathbf{v}, \mathbf{v}) \geq 0$ (i.e. $\mathbf{v} := f\mathbf{E}_i$), so I is not negative definite. Therefore c is not maximal, so there must be a conjugate point to S along c at, or before, $c(b(p))$. □

Corollary 13.16. *Let S be a compact spacelike hypersurface in M without boundary. If $\mathbf{Ric}(\mathbf{V}, \mathbf{V}) \geq 0$ for all time-like vectors $\mathbf{V}$ and $\kappa(p) < 0$ for all $p \in S$. Then there exists $q \in S$ such that $\max_{p \in S} \kappa(p) = \kappa(q) < 0$. Defining $b := -n/\kappa(q)$ then, along each future-directed normal geodesic* *to S, there will exist a point conjugate to S along that geodesic within a distance b of S, provided that the geodesics can be extended that far.*

Proof. The existence of q follows from the compactness of S. From the previous proposition, given $p \in S$ there exists a conjugate point along the corresponding normal geodesic within distance $b(p)$ of S. Since

$$b(p) = -\frac{n}{\kappa(p)} = \frac{n}{|\kappa(p)|} \leq \frac{n}{|\kappa(q)|} = -\frac{n}{\kappa(q)} \equiv b,$$

it follows that, given any $p \in S$, there exists a conjugate point along the corresponding normal geodesic within distance b of S. □

Example 13.17 (Misner space)**.** Consider flat $\mathbb{R}^{1,1}$ with the metric $\mathbf{g} = -dt^2 + dx^2$. We consider the interior of the past null cone of 0, with coordinates (T, θ) such that $t = T\cosh\theta$, $x = T\sinh\theta$ with $T < 0$, $\theta \in \mathbb{R}$. If we now identify points (T, θ) with $(T, \theta + 2n\pi)$, then the space-like hyper-surfaces $T = \text{constant} < 0$ are compactified to become topologically S^1. The metric on the resulting space is simply

$$\mathbf{g}_{Misner} = -dT^2 + T^2 d\theta^2, \qquad T < 0, \qquad \theta \in [0, 2\pi).$$

Lecture 21

The surfaces $S_a := \{T = a < 0\}$ have unit future-directed time-like normal

$$\mathbf{n} = \partial_T.$$

To calculate the extrinsic curvature of S_a, we require

$$\chi(\partial_\theta, \partial_\theta) \equiv \mathbf{g}(\partial_\theta, \nabla_{\partial_\theta}\partial_T) = T^2\langle d\theta, \nabla_{\partial_\theta}\partial_T\rangle.$$

In local coordinates (T, θ), we therefore need to calculate

$$\langle d\theta, \nabla_{\partial_\theta}\mathbf{n}\rangle \equiv \langle d\theta, \sum_{a=r,\theta}\left(\partial_\theta n^a + \sum_{b=r,\theta}\Gamma^a{}_{\theta b}n^b\right)\partial_a\rangle = \Gamma^\theta{}_{\theta T} = \frac{1}{2}g^{\theta\theta}\partial_T g_{\theta\theta} = \frac{1}{T}.$$

Hence

$$\chi(\partial_\theta, \partial_\theta) = T.$$

Taking the trace,

$$\kappa = g^{\theta\theta}\chi(\partial_\theta, \partial_\theta) = \frac{1}{T}.$$

Note that, since $T < 0$, if follows that $\kappa < 0$. Note that, in this case, $b(p) = 1/|T(p)|$ and that the normal geodesics from the surface S_a all converge to the point $T = 0$, which is at parameter value $|a|$ from S_a. In the current example the metric is flat, so $\mathbf{Ric} = 0$, so this is consistent with Corollary 13.16.

14. Causality

14.1. **Causality relations.** Recall from previously:

Definition 14.1. For sets $A, U \subseteq M$ we define the *chronological future of A relative to U* to be the set

$$I^+(A, U) := \{q \in U : \exists p \in A \text{ and a future-pointing time-like curve from } p \text{ to } q \text{ in } U\}$$

and the *causal future of A relative to U*

$$J^+(A, U) := (A \cap U) \cup \{q \in U : \exists p \in A \text{ and a future-pointing causal curve from } p \text{ to } q \text{ in } U\}.$$

When $U = M$, we write $I^+(A)$ and $J^+(A)$, respectively. Similarly, we define $I^-(A, U)$, $J^-(A, U)$.

Remark 14.2. By a curve we mean one of non-zero extent, so we can have $S \not\subseteq I^+(S, U)$.

Definition 14.3. The *time separation* $d : M \times M \to [0, \infty]$ is defined by

$$d(p,q) = \begin{cases} 0 & q \notin J^+(p) \\ \sup\{L[\gamma]\} & q \in J^+(p), \end{cases}$$

for $p, q \in M$, where the supremum is taken over future-directed piecewise causal curves from p to q. Similarly, if A, B are subsets of M then we define

$$d(A,B) = \sup\{d(p,q) : p \in A, q \in B\}.$$

Remark 14.4. Note that $d(p,q) \neq d(q,p)$ in general.

Definition 14.5. Let $c : I \to M$ be a future-directed, causal curve (We will assume that all curves are piece-wise smooth.) defined on domain $I \subseteq \mathbb{R}$. Let $a := \inf I$, $b := \sup I$, where we allow $a = -\infty$ and/or $b = +\infty$. A point $p \in M$ is a *future endpoint* of c if for all sequences $\{b_i\}$ with $b_i \to p$ then $c(b_i) \to c(p)$. A causal curve is *future-inextendible* (in a set S) if it has no future endpoint (in S). More generally, given a subset $S \subseteq M$, a causal curve is *future-inextendible in S* if it has no future endpoint in S.

We define the corresponding notions for past endpoints, etc, by considering sequences $\{a_i\}$ with $a_i \to a$.

Intuitively speaking, a causal curve without a future endpoint either extends indefinitely into the future, or must reach some finite parameter value at which it does not admit a continuous extension within M. The latter occurs, for example, if we consider flat $\mathbb{R}^{n,1}$ and remove a point.

Definition 14.6. Let S be a closed subset of M. The *future Cauchy development* of S, $D^+(S)$, is the set of points $p \in M$ such that every past-inextendible causal curve through p intersects S. Similarly, we define $D^-(S)$, and $D(S) := D^+(S) \cup D^-(S)$.

Lecture 22

Definition 14.7. A set S is *achronal* if $S \cap I^+(S)$ is empty (i.e. no two points of S have time-like separation). A set S is *acausal* if $S \cap J^+(S)$ is empty (i.e. no two points of S have causal separation).

A closed achronal set S is a *future Cauchy hypersurface* if $J^+(S) \subseteq D^+(S)$.

Proposition 14.8. *Let S be a closed, achronal, spacelike hypersurface in M. Given any $q \in D^+(S)$, there exists a maximising geodesic from S to q.*

Proof. Later. □

The reason that we are interested in such a result is because it allows us to deduce the following rudimentary singularity theorem:

Corollary 14.9. *Let $(M, \mathbf{g})$ be a Lorentzian manifold such that $\mathbf{Ric}(\mathbf{V}, \mathbf{V}) \geq 0$ for all time-like vector fields $\mathbf{V}$. If there exists a compact, achronal, spacelike future Cauchy surface without boundary, S, with $\kappa \leq \kappa_0 < 0$ on S, then M is time-like geodesically incomplete.*

Proof. Let c be a normal, future-directed geodesic from S. Then, from Corollary 13.16, we know that there exists a conjugate point to S along c at $c(t_0)$ where $t_0 \leq b := -n/\kappa_0$. Letting $t_1 > b$, then there are two possibilities. Firstly, it is possible that c cannot be extended to this value of t. In this case, by definition, M is geodesically incomplete and we are finished. The second possibility is that there exists $q := c(t_1) \in M$. Since $c(t_0)$ is conjugate to S along c, and $t_0 < t_1$, it follows that c cannot be a maximising geodesic from S to q. Since S is a future Cauchy surface and $q \in J^+(S)$, we deduce that $q \in D^+(S)$ and therefore, from 14.8, we know that there exists a maximising geodesic, $\tilde{c}$, from S to q. From first variation of arclength, we know that $\tilde{c}$ must be normal to S. However, since $\tilde{c}$ is maximising, we must have $L[\tilde{c}] \geq L[c] = t_1 > b$, therefore we again deduce from Corollary 13.16 that there must be a point conjugate to S along $\tilde{c}$ prior to q. Therefore $\tilde{c}$ cannot be maximising up to q. This contradicts the definition of $\tilde{c}$. □

Remark 14.10. Note that, in the previous proof, we have actually shown the stronger result that every time-like geodesic normal to S has length less than or equal to b.

The conditions in the above theorem are rather strong and, with some work, can be significantly relaxed. The most general result along these lines is the following.

Theorem 14.11. *Let $(M, \mathbf{g})$ be a (four-dimensional) Lorentzian manifold such that the following conditions hold:*

(1) $\mathbf{Ric}(\mathbf{v}, \mathbf{v}) \geq 0$ *for every causal vector* $\mathbf{v}$*;*
(2) *there exists a compact space-like three-surface S without edge;*
(3) *the unit normals to S are everywhere converging on S (i.e. $\kappa < 0$ on S).*

Then the space-time $(M, \mathbf{g})$ is time-like geodesically incomplete.

Remark 14.12. The conditions on the surface S (i.e. condition (2)) are much weaker than in 14.9. In this case, however, only the existence of an incomplete time-like geodesic is proved, rather than the stronger conclusion mentioned in Remark 14.10.

For the rest of the course, we will concentrate on outlining the proof of Proposition 14.8, which implies the weaker result 14.9.

End of examinable material

Appendix A. Proof of Proposition 14.8 (Non-examinable)

Lecture 23

The following is an outline of the main arguments leading to the proof of the following result.

Proposition A.1. *Let S be a closed, achronal, spacelike hypersurface in M (without boundary). Given any $q \in D^+(S)$, there exists a maximising geodesic from S to q.*

For full details of the proof, see either Chapter 14 of [O'N] or Chapter 6 of [HE].

Definition A.2. The *strong causality condition* holds at $p \in M$ if given any neighbourhood, U, of p there exists a neighbourhood $V \subseteq U$ of p such that any causal curve $c : [a,b] \to M$ with $c(a), c(b) \in V$ lies in U i.e. $c(t) \in U$ for all $t \in [a,b]$.

A set $N \subseteq M$ is *globally hyperbolic* if the strong causality condition holds on N and if, for any $p, q \in N$, the set $J^+(p) \cap J^-(q)$ is compact and contained in N.

Remarks A.3.

(1) Intuitively speaking, this means that a causal curve that starts arbitrarily close to p (i.e. in the set V) and leaves a fixed neighbourhood of p (i.e. U) cannot later return arbitrarily close to p (i.e. reenter the set V).
(2) Global hyperbolicity, along with suitable convexity conditions, implies that every neighbourhood of p contains a neighbourhood of p that no (future-inextendible) causal curve intersects more than once. (See [O'N, §14, Exercise 10(b)].) This is the definition of strong causality used by Hawking and Ellis [HE].

Lemma A.4. *The map $d : M \times M \to [0, \infty]$ is lower semi-continuous.*

Proof. The only non-trivial case is when $d(p,q) > 0$. Therefore let $q \in I^+(p)$ where, for the moment, we assume $0 < d(p,q) < \infty$. We want to show that, given any $\epsilon > 0$, there exists a neighbourhood U of p and a neighbourhood V of q such that $d(p', q') > d(p,q) - \epsilon$ for all $p' \in U$, $q' \in V$. Let γ be a time-like curve from p to q with $L[\gamma] > d(p,q) - \frac{\epsilon}{3}$. Let $\mathcal{C}$ be a convex normal neighbourhood of q and q_1 a point of $\gamma \cap \mathcal{C}$ prior to q. There then exists a neighbourhood, V, of q such that if $q' \in V$ then the geodesic from q_1 to q' is causal and has length $> L[\gamma|_{[q_1,q]}] - \frac{\epsilon}{3}$. Similarly, we construct a neighbourhood U of p such that, given any $p' \in U$ and $q' \in V$, they can be joined by a causal curve of length $> L[\gamma|_{[p,q]}] - \frac{2\epsilon}{3} > d(p,q) - \epsilon$, as required.

Finally, if $q \in I^+(p)$ and $d(p,q) = \infty$, then the above construction shows that, given any $N > 0$, there are neighbourhoods U and V such that $d(p', q') > N$ for all $p' \in U$ and $q' \in V$. □

Lemma A.5. *Let $N \subseteq M$ be a globally hyperbolic open set. Then $d(p,q) < \infty$ for all $p, q \in N$, and the map $d : N \times N \to [0, \infty)$ is continuous.*

Proof. Given $p, q \in N$ with $q \in I^+(p)$, we first wish to prove that $d(p,q) < \infty$. To do this, note that strong causality and the fact that $J^+(p) \cap J^-(q)$ is compact implies that we can cover $J^+(p) \cap J^-(q)$ by a finite number, $k < \infty$, of convex normal neighbourhoods such that each set contains no causal curve of length greater than some bound $\epsilon > 0$. Since any causal curve from p to q can enter each such neighbourhood at most once, it must have finite length. In particular, $d(p,q) \leq k\epsilon < \infty$.

Lecture 24

We prove that d is upper semi-continuous by contradiction. In particular, assume that d is not upper semi-continuous at $(p,q) \in N \times N$. Then there exists an $\epsilon > 0$ and sequences $p_n \to p$, $q_n \to q$ with the property that $d(p_n, q_n) \geq d(p,q) + \epsilon$ for all n. Since $d(p_n, q_n) > 0$, there exists a causal curve c_n from p_n to q_n with $L[c_n] > d(p_n, q_n) - \frac{1}{n}$. Since N is open, there exist points $p_- \in I^-(p) \cap N$ and $q_+ \in I^+(q) \cap N$ and, without loss of generality, we may assume that $p_n \in I^+(p_-)$ and $q_n \in I^-(q_+)$. In this case, the curves c_n are contained in the set $J^+(p_-) \cap J^-(q_+)$. Since N is globally hyperbolic, this set is compact. In the next paragraph, we show that these conditions imply the existence of a causal curve, c, from p to q of length $L[c] \geq d(p,q) + \epsilon$. This contradicts the definition of $d(p,q)$, giving the required result.

Finally, we sketch the proof of the fact that given causal curves, c_n, contained in a compact subset, $K := J^+(p_-) \cap J^-(q_+)$, of a globally hyperbolic set N with endpoints p_n, q_n that converge to points p, q, then we can construct a causal curve (in fact, a broken geodesic) from p to q with length $\geq d(p,q)+\epsilon$. Beginning at the point $p_0 := p$, let B_0 be a convex normal neighbourhood of p contained in K. There are two possibilities:

1). If $q \in B_0$, then we take the (unique) geodesic, c, from p to q in B_0. In this case, $L[c] \geq \lim_{n\to\infty} L[c_n] > \lim_{n\to\infty} d(p_n, q_n) \geq d(p,q)+\epsilon$, so we are done.

2). If $q \notin B_0$, then for sufficiently large n, $q_n \notin B_0$. Therefore, for such n, the curves c_n must leave the set B_0. Let $c_n(s_n)$ be the first points at which c_n intersect ∂B_0. Since ∂B_0 is compact, there exist a subsequence of $\{c_n\}$, which we again denote by $\{c_n^1\}$, with the property that $c_n^1(s_n^1)$ converge to a point $p_1 \in \partial B_0$. Note that $p_1 \in J^+(p)$ but, by construction, $p_1 \neq p$. Now let B_1 be a convex normal neighbourhood of p_1 and repeat the above procedure.

The above process produces a sequence of points $\{p_1, p_2, \dots\}$ in K. Since each consecutive pair of points lie in a convex neighbourhood, we may form the broken geodesic curve, c, that we get by joining together the geodesics between the consecutive points in the sequence. If the above procedure does not terminate, then the curve c will be an future-inextendible causal curve. As such, it must leave the compact set K after some finite time. This, however, implies that $p_k \notin K$ for sufficiently large k, contradicting the fact that c_n are contained in K. Therefore, the above sequence must be finite $\{p_1, \dots, p_k\}$ and ends at $p_k \equiv q$. Forming the broken geodesic that we get by joining the points of the sequence $\{p_0, \dots, p_k\}$ gives a broken geodesic from p to q, and a subsequence $\{\tilde{c}_n\}$ of $\{c_n\}$, with points $\tilde{c}_n(s_{n,i})$ that converge to the p_i as $n \to \infty$, for each $i = 1, \dots, k$. The length $L[\tilde{c}_n]$ will be less than or equal to the length, L_n, of the corresponding broken geodesic that connects the points $\{\tilde{c}_n(0), \tilde{c}_n(s_{n,1}), \dots, \tilde{c}_n(s_{n,k})\}$. Since $L_n \to L[c]$ as $n \to \infty$, we deduce that $L[c] \geq \lim_{n\to\infty} L[\tilde{c}_n] \geq d(p,q)+\epsilon$. □

Remark A.6. The sequence of points $\{p_i\}$ constructed from the sequence of future-directed causal curves $\{c_n\}$ is called a *limit sequence for* $\{c_n\}$. Given any sequence of future-directed causal curves with $c_n(0) \to p$ and such that there exists a neighbourhood of p that almost all of the c_n leave, then $\{c_n\}$ has a limit sequence starting at p.

The broken geodesic, c, constructed from the limit sequence is a *quasi-limit* of the sequence $\{c_n\}$. In the case where the set of points $\{p_i\}$ is infinite, then c is a future inextendible causal curve. The existence of such a curve may often violate various causality conditions and, hence, allows us to deduce that, for a quasi-limit of a sequence $\{c_n\}$, the set $\{p_i\}$ will be finite.

Proposition A.7. *Let $N \subseteq M$ be a globally hyperbolic open set, and $p, q \in N$ with $q \in J^+(p)$. Then there exists a causal geodesic from p to q the length of which is equal to $d(p,q)$.*

Proof. Let $p, q \in N$, with $q \in J^+(p)$. Let U be a convex normal neighbourhood of p contained in N. (Such a U exists, since N is, by assumption, open.) If $q \in U$ the result is automatic from the fact that U is a convex normal neighbourhood, so assume that $q \notin U$. Now consider the map $\partial U \cap (J^+(p) \cap J^-(q)) \to [0,\infty)$ given by $r \mapsto d(p,r)+d(r,q)$. Note that, since p, q and r are elements of the globally hyperbolic set N, it follows from Lemma A.5 that this map is continuous. Also, since $\partial U \cap (J^+(p) \cap J^-(q))$ is a closed subset of a compact set (i.e. $J^+(p) \cap J^-(q)$) it is therefore compact. Hence there will exist $x \in \partial U \cap (J^+(p) \cap J^-(q))$ at which the above function achieves its maximal value.

Claim 1: $d(p,x)+d(x,q) = d(p,q)$.

Proof of Claim 1: Let $\gamma : [0,b] \to N$ be any continuous, future-directed, causal curve from p to q, with $\gamma(0) = p$, $\gamma(b) = q$. Since M is assumed Hausdorff, there exists $a \in (0,b)$ such that $\gamma(a) \in \partial U$. Therefore

$$L[\gamma] = L[\gamma|_{[0,a]}] + L[\gamma|_{[a,b]}] \leq d(p,\gamma(a)) + d(\gamma(a),q) \leq d(p,x)+d(x,q),$$

where the last inequality follows from the definition of the point $x \in \partial U$. Therefore, any continuous, future-directed, causal curve from p to q has length $L[\gamma] \leq d(p,x)+d(x,q)$. Therefore $d(p,q) = \sup_\gamma L[\gamma] \leq d(p,x)+d(x,q)$. The triangle inequality, however, yields $d(p,q) \geq d(p,x)+d(x,q)$. Therefore $d(p,q) = d(p,x)+d(x,q)$, as required.

Given p and $x \in \partial U$, construct the future-directed, causal geodesic from p through x, which we denote by c. We assume that $p = c(0)$, $x = c(a)$ for some $a > 0$. Assuming that M is geodesically complete, we can assume that $c : [0, \infty) \to M$ is defined.

Claim 2: For all $y \in J^-(q) \cap c$ we have $d(p, y) + d(y, q) = d(p, q)$.

Proof of Claim 2: Assume the contrary. The map $t \mapsto d(p, c(t)) + d(c(t), q)$ is continuous, so the set $\{t \in [0, \infty) : d(p, c(t)) + d(c(t), q) = d(p, q) \text{ is true}\}$ is a closed subset of $[0, \infty)$. This set is non-empty, since 0 and a are elements of it. Therefore, there will exist a $T > 0$ such that the claim is true for $t \leq T$, but false for $t > T$. Now construct a convex normal neighbourhood of $c(T)$, V, contained in N. As above, we construct a point $m \in \partial V \cap J^+(c(T)) \cap J^-(q)$ such that $d(c(T), m) + d(m, q) = d(c(T), q)$. We then construct the geodesic λ from $c(T)$ passing through m. We then have

$$d(p, c(T)) + d(c(T), m) + d(m, q) = d(p, c(T)) + d(c(T), q) = d(p, q) \geq d(p, m) + d(m, q).$$

Therefore

$$d(p, c(T)) + d(c(T), m) \geq d(p, m).$$

Along with the triangle inequality this implies that $d(p, c(T)) + d(c(T), m) = d(p, m)$. Therefore $\lambda \circ c$ is a curve that maximises the distance from p to m. From the first variation formula, this implies that $\lambda \circ c$ must be a smooth geodesic. Therefore λ is a continuation of the geodesic curve c. This contradicts the definition of T as the final point on c at which our required equality holds.

Conclusion of proof of Proposition: Finally, since the set $J^+(p) \cap J^-(q)$ is compact, the curve c must leave the set $J^-(q)$ at some point y. If $y \neq q$ then y must lie on a past-directed null geodesic, λ, from q to y. In this case, $\lambda \circ c$ gives a causal curve from p to q of length $d(p, q)$. (Note, also, that $d(p, y) = d(p, q)$.) However, this curve is not a causal geodesic, and hence, from the first variation formula, can be varied to give a longer causal curve from p to q. This contradicts the fact that $d(p, y) = d(p, q)$. As such, the assumption that $y \neq q$ must be false i.e. $y = q$. Therefore, c is a causal geodesic from p to q with $L[c] = d(p, q)$. □

Remark A.8. Except for the last part, which is Lorentzian in nature, the above proof is just a modification of the usual proof of the fact that any two points on a complete Riemannian manifold can be joined by a minimising geodesic.

Remark A.9. More generally, Proposition A.7 holds for N any globally hyperbolic set, not necessarily open. (See, e.g., [HE, Proposition 6.7.1].)

Lemma A.10. *Let A be an achronal set. Then* $\text{int}(D(A))$ *is globally hyperbolic.*

Sketch of Proof. [14] To begin, we show that strong causality holds on $\text{int}(D(A))$. We do this in two stages.

1). Let $p \in \text{int}(D(A))$ and assume that there exists a closed causal curve, c, through p. Then, traversing c over and over again defines an inextendible causal curve, $\tilde{c}$, through p. Such a curve necessarily intersects $I^+(A)$ and $I^-(A)$. As such, there exist points $q \in c \cap I^-(A)$, $r \in c \cap I^+(A)$. Since $r \in J^-(q)$ (as there exists a past-directed causal curve from q to r i.e. the relevant part of $\tilde{c}$) it follows that $r \in I^-(A)$. Since $r \in I^+(A)$, this implies the existence of a time-like curve that intersects A twice. This contradicts the fact that A is achronal. Therefore there can exist no closed causal curve through any point $p \in \text{int}(D(A))$.

2). To show that $\text{int}(D(A))$ obeys strong causality, we assume, on the contrary, that there exists $p \in \text{int}(A)$ and future-directed causal curves $c_n : [0, 1] \to M$ such that $c_n(0) \to p$ and $c_n(1) \to p$, but every c_n leaves some fixed neighbourhood, U, of p. We then construct a quasi-limit of the $\{c_n\}$ with vertices $\{p_i\}$. If this set is finite, then it must end at $\lim_{n\to\infty} c_n(1) = p$. We have then constructed a closed causal curve through p, which violates the result of Part 1). If the set $\{p_i\}$ is infinite then the corresponding quasi-limit, c, is a future-inextendible causal curve. As such, c must enter $I^+(A)$ at some point and remain there. Hence there exists an i such that $p_i \in I^+(A)$

[14]For full details, see [O'N, §14, Theorem 38]

and a subsequence $\{c_m\}$ of $\{c_n\}$ and an $s \in [0,1]$ such that $c_m(s) \to p_i$. (WLOG, may assume that $c_m(s) \in I^+(A)$.) Since $p_i \neq p$, we now consider the curves $\{c_m|[s,1]\}$. From these, we may construct a past-directed quasi-limit with vertices $\{\overline{p}_i\}$ starting from p (since we know that $\lim_{n\to\infty} c_n(1) = p$). If $\{\overline{p}_i\}$ is finite, then it must end at p_i (since $\lim_{m\to\infty} c_m(s) = p_i$). Putting the quasi limits together then gives a causal curve from p to itself, which again violates the result of Part 1). However, if $\{\overline{p}_i\}$ is infinite, then the quasi-limit, $\overline{c}$, is a past-inextendible causal curve, and hence must intersect $I^-(A)$. Hence some $c_m|[s,1]$ must intersect $I^-(A)$ i.e. there exists $\overline{s} > s$ with $c_m(\overline{s}) \in I^-(A)$. Since $c_m(s) \in I^+(A)$, we may therefore construct a time-like curve that intersects A more than once, violating achronality of A.

Let $p, q \in \mathrm{int}(D(A))$. To prove compactness of $J^+(p) \cap J^-(q)$, first assume that $q \in J^+(p)$ (otherwise the set is empty, in any case). If $q = p$ then Part 1) above implies that $J^+(p) \cap J^-(p) = \{p\}$, so we are done. Therefore, let $q \in J^+(p) \setminus \{p\}$. Let $\{x_n\}$ be a sequence of points in $J^+(p) \cap J^-(q)$. We may then construct future-directed causal curves, $\{c_n\}$, that pass through p, x_n and q. We then construct a limit-sequence $\{p_i\}$ from the $\{c_n\}$. By similar methods to those used in Part 2), one can show that this sequence being infinite contradicts achronality of A, so the sequence $\{p_i\}$ must be finite. It then follows that there exists intervals $[s_{m,i}, s_{m,i+1}]$, where $c_m(s_{m,i}) \to p_i$, $c_m(s_{m,i+1}) \to p_{i+1}$ such that the point x_m lies on the segment $c_m|[s_{m,i}, s_{m,i+1}]$ for infinitely many m. Passing to this subsequence, we may assume that the x_m lie (for sufficiently large m) in a convex open neighbourhood $\mathcal{C}$ with compact closure. Therefore $\{x_m\}$ will have a subsequence converging to a point x which will satisfy $x \in J^+(p_i)$ and $x \in J^-(p_{i+1})$. Hence $x \in J^+(p) \cap J^-(q)$.

Finally, given $p, q \in \mathrm{int}(D(A))$, we need to show that $J^+(p) \cap J^-(q)$ is contained in $\mathrm{int}(D(A))$. Again, we may assume $q \in J^+(p) \setminus \{p\}$. The main ideas are contained in the case where $p, q \in I^+(A)$. We then let $q_+ \in I^+(q) \cap D(A) \subseteq D^+(A)$. Then, $J^+(p) \cap J^-(q)$ is contained in the open set $I^-(q_+) \cap I^+(A)$, so it is enough to show that this set is contained in $\mathrm{int}(D(A))$. Let γ be a past-directed time-like curve from q_+ to a point $y \in I^-(q_+) \cap I^+(A)$. If γ intersects A then y would lie in $I^-(A)$ (since γ would be a past-directed time-like curve from A to y) and $I^+(A)$, and would therefore violate achronality of A. Hence γ cannot intersect A. Therefore, given any past-directed inextendible causal curve, c, from y, the curve $c \circ \gamma$ is a past-directed inextendible causal curve from q_+. Since $q_+ \in D^+(A)$, it follows that $c \circ \gamma$ must intersect A. Since γ does not intersect A, it follows that c must intersect A. Hence $y \in D^+(S)$. Hence $I^-(q_+) \cap I^+(A) \subseteq D^+(A)$. Since $I^-(q_+) \cap I^+(A)$ is open, it follows that $I^-(q_+) \cap I^+(A) \subseteq \mathrm{int}(D(A))$. The proof for arbitrary $p, q \in \mathrm{int}(D(A))$ involves no new ideas. □

In a similar way, one can prove the following results[15]

Lemma A.11. *Let A be an achronal set and $p \in \mathrm{int}(D(A)) \setminus I^-(A)$, then the set $J^-(p) \cap D^+(A)$ is compact.*

Lemma A.12. *Let S be a closed, achronal, space-like hypersurface in M (without boundary)*[16]*. Then $D(S)$ is globally hyperbolic.*

Proof of Proposition A.1. By Lemma A.12, $D(S)$ is a globally hyperbolic open set. By Lemma A.11, given $q \in \mathrm{int}(D(A)) \setminus I^-(A)$, the set $J^-(q) \cap D^+(S)$ is compact. Hence, since S is closed, the set $(J^-(q) \cap D^+(S)) \cap S$ is compact i.e. $J^-(q) \cap S$ is compact. By Lemma A.5, the map $r \mapsto d(r, q)$ is continuous on $J^-(q) \cap S$, and therefore takes a maximum at some point $p \in S$. By Proposition A.7, there exists a geodesic, c, from p to q of length $d(p, q) = d(S, q)$. □

[15]For proofs, see [O'N, §14, Lemma 40] and [HE, §6, Proposition 6.6.7, Part (2)], respectively.

[16]Such an S is necessarily acausal

References

[CE] J. Cheeger, D. G. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland Publishing Co., Amsterdam, 1975. 5

[Cha] I. Chavel, *Riemannian geometry, A modern introduction*, Cambridge Studies in Advanced Mathematics, vol. 98, Cambridge University Press, Cambridge, 2006. 2, 5, 23

[GHL] S. Gallot, D. Hulin, J. Lafontaine, *Riemannian geometry*, Springer-Verlag, Berlin, 2004. 2, 5, 25

[Gr] M. Gromov, *Metric structures for Riemannian and non-Riemannian spaces*, Progress in Mathematics, vol. 152, Birkhäuser Boston Inc., Boston, MA, 1999. 17

[HE] S. W. Hawking, G. F. R. Ellis, *The large scale structure of space-time*, Cambridge University Press, London, 1973. 29, 32, 33, 45, 47, 48

[HY] J. G. Hocking, G. S. Young, *Topology*, second ed., Dover Publications Inc., New York, 1988. 25

[K] S. Kobayashi, *On conjugate and cut loci*, Studies in Global Geometry and Analysis, Math. Assoc. Amer. (distributed by Prentice-Hall, Englewood Cliffs, N.J.), 1967, pp. 96–122. 25

[O'N] B. O'Neill, *Semi-Riemannian geometry*, Pure and Applied Mathematics, vol. 103, Academic Press Inc. 2, 29, 45, 47, 48

[P] P. Petersen, *Riemannian geometry*, Graduate Texts in Mathematics, vol. 171, Springer-Verlag, New York, 1998. 5

[W] F. W. Warner, *Foundations of differentiable manifolds and Lie groups*, Graduate Texts in Mathematics, vol. 94, Springer-Verlag, New York, 1983. 2

Index

acausal set, 43
achronal set, 43
arc-length, 35

Bishop-Gromov Comparison Theorem, 17, 26

causal future of A relative to U, 31, 42
chronological future of A relative to U, 31, 42
conjugate point, 21
  order, 21
connection
  affine, 2
    curvature, 2
    torsion, 2
  Levi-Civita, 3
  metric, 3
  on a vector bundle, 2
  torsion-free, 3
convex, 34
covariant derivative, 2
curvature
  of connection on vector bundle, 2
  scalar, 4
  sectional, 6
curve, 6
  arc-length, 6
  length, 6
  regular, 6
  tangent vector, 6
cut point, 23
cut-locus, 25
cut-locus of p in T_pM, 25

distance function, 7, 12

exponential map, 9

first variation formula, 7
future Cauchy development, $D^+(S)$, 43
future Cauchy hypersurface, 43
future endpoint, 43
future-inextendible curve, 43
future-inextendible curve in S, 43

geodesic, 8
  maximal, 38
  minimising, 22
  parametrised by arc-length, 8
geodesically inextendible, 9, 33
globally hyperbolic, 45
gradient
  of a function, 10

Hessian, 11

index form, 20
injectivity radius
  of $(M, \mathbf{g})$ at p, 25

Jacobi equation, 20
Jacobi field, 20

Koszul formula, 3

Laplacian, 12
Lorentzian manifold
  time-orientable, 29

maximal, 9, 33
metric
  Lorentzian, 3
  Riemannian, 3
  semi-Riemannian, 3
Myers's theorem, 27

normal coordinates, 9
normal neighbourhood, 34

pull-back
  tensor field, 4

Riccati equation, 14
Ricci curvature tensor, 3
Ricci scalar, 4
Ricci tensor, 3

scalar curvature, 4
second variation formula, 19
sectional curvature, 6
smooth variation of c, 7
  two parameter, 18
star-shaped about 0, 34
strong causality condition, 45

time separation, 43

variation vector field, 7
vector
  null, 28
  space-like, 28
  time-like, 28
    future-directed, 28
    past-directed, 28
vector field
  along a map, 7
  null, 29
  space-like, 29
  time-like, 29
vector fields
  Lie bracket, 2