THE GEOMETRY OF GENERAL RELATIVITY

GUSTAV HOLZEGEL,
TRANSCRIBED BY IAN TOBASCO

ABSTRACT. These notes were taken in the first week of the 2011 Study Analysis and Geometry summer program at Princeton University.

Contents

Date: July 6 to 9, 2011.

1. An Introduction to Differential Geometry

First we discuss manifolds, and the properties they inherit from $\mathbb{R}^n$. A manifold looks locally like $\mathbb{R}^n$, in that it is locally homeomorphic to $\mathbb{R}^n$. One ex is the two-sphere, which is locally like $\mathbb{R}^2$. We want to perform calculus on manifolds, so we linearize. One can think of this as attaching a linear space to each point of the manifold. The inhabitants of these linear spaces are vectors. Given a vector field (smooth choice of vector at each point) we can associate a flow field. This will lead to the notion of Lie Derivative. Following this, we will introduce an additional structure to manifolds: connections. This will allow us to compare vectors in different tangent spaces. A connection gives rise to the notion of parallel transport from point to point. As we'll see, a connection will also lead to the notions of geodesic flow, curvature, and torsion.

1.1. Manifolds and Tangent Spaces.

1.1.1. *Manifolds.*

Definition 1.1. An *n-dimensional smooth manifold* M is a topological space (Hausdorff[1]) together with a collection of open sets $U_\alpha \subset M$ and homeomorphisms $\phi_\alpha : U_\alpha \to \phi_\alpha(U_\alpha) \subset \mathbb{R}^n$ such that the following holds:

(1) U_α cover M, i.e., $M = \cup_\alpha U_\alpha$
(2) For $p \in U_\alpha \cap U_\beta$ the map $\phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \to \phi_\beta(U_\alpha \cap U_\beta)$ is smooth as a map from $\mathbb{R}^n$ to $\mathbb{R}^n$
(3) The collection (U_α, ϕ_α) is maximal with respect to (1) and (2).

Definition 1.2. The maps ϕ_α in the previous definition are called *transition maps.* A pair (U_α, ϕ_α) is called a *coordinate chart* for M. A collection (U_α, ϕ_α) satisfying (1) and (2) is called an *atlas* for M.

Example 1.3. A first ex is stereographic projection of the two-sphere. There is another way to generate charts for the two-sphere. Using six charts we can cover the sphere by projecting hemispheres.

Now we can talk about differentiable maps between manifolds.

Definition 1.4. A map $f : M \to N$ between two manifolds is called *differentiable* at $p \in M$ if given a chart (V, ψ) around $f(p) \in N$ there exists a chart (U, ϕ) around p with $f(U) \subset V$ and so that the map $\tilde{f} = \psi \circ f \circ \phi^{-1} : \phi(U) \to \psi(V)$ is differentiable at $\phi(p)$.

Exercise 1.5. Show that this definition does not depend on choice of V.

1.1.2. *Tangent Spaces.*

Consider a curve $\gamma(t) : I \to \mathbb{R}^n$, $\gamma(t) = \left(x^1(t), \dots, x^n(t)\right)$. We already have a notion of tangent vector to the curve at a point, say $\gamma(0) = p$.

[1] We'll usually assume paracompactness as well.

But now we'll think of the tangent vector as a directional derivative. Consider

$$\begin{aligned}
\left.\frac{d}{dt}(f\circ\gamma)\right|_{t=0} &= \left.\frac{d}{dt} f\left(x^1(t),\ldots,x^n(t)\right)\right|_{t=0}\\
&= \sum_{i=1}^{n}\left.\frac{\partial f}{\partial x^i}\right|_p \cdot \frac{dx^i}{dt}(0)\\
&= \left[\sum_{i=1}^{n}\frac{dx^i}{dt}(0)\frac{\partial}{\partial x^i}\right] f.
\end{aligned}$$

In this way we can think of the tangent vector as operating on f in the form of a derivative. The next definition generalizes this to manifolds.

Definition 1.6. Given a differentiable curve γ through p and $f \in \mathcal{F}_p(M)$ (the space of functions differentiable at p) we define the *tangent vector to the curve* γ at p (denoted γ_p') to be the map $\gamma_p' : \mathcal{F}_p(M) \to \mathbb{R}$, $f \mapsto \frac{d}{dt}(f\circ\gamma)\big|_{t=0}$.

Definition 1.7. The set of tangent vectors (obtained from curves via the previous definition) is called the *tangent space* at p and is denoted T_pM.

Let's explore this definition in a coordinate chart. Pick a chart ϕ around p and call the coordinates $x^1,\ldots,x^n$. So write $f\circ\phi^{-1} = f\left(x^1,\ldots,x^n\right)$. Given a curve γ, we can represent in coordinates as $\phi\circ\gamma = \left(x^1(t),\ldots,x^n(t)\right)$. Then the definition becomes

$$\begin{aligned}
\left.\frac{d}{dt}(f\circ\gamma)\right|_{t=0} &= \sum_{i=1}^{n}\left.\frac{\partial\left(f\circ\phi^{-1}\right)}{\partial x^i}\right|_{\phi(p)}\left.\frac{d(\phi\circ\gamma)^i}{dt}\right|_{t=0}\\
&= \sum_{i=1}^{n}\left(x^i\right)'(0)\left[\frac{\partial}{\partial x^i}\right]_p f
\end{aligned}$$

where

$$\left[\frac{\partial}{\partial x^i}\right]_p f = \left.\frac{\partial\left(f\circ\phi^{-1}\right)}{\partial x^i}\right|_{\phi(p)}.$$

In particular, if γ_k is the curve such that $\phi\circ\gamma_k$ is the kth coordinate axis in $\mathbb{R}^n$, then

$$\left(\gamma_k'\right)_p f = \left[\frac{\partial}{\partial x^k}\right]_p f.$$

Exercise 1.8. Show that T_pM is a vector space of dimension n. Hints:

- $v, w \in T_pM \longrightarrow \gamma_1, \gamma_2$
- $v + w$, natural candidate curve $\phi\circ\tilde{\gamma} = \phi\circ\gamma_1 + \phi\circ\gamma_2 - \phi(p)$
- Show that $\left[\frac{\partial}{\partial x^k}\right]_p$ is a basis.

Proposition 1.9. *Let $f : M \to N$ be a differentiable map between manifolds of dimension m and n. For every $p \in M$ and every $v \in T_pM$ we choose a differentiable curve through p, $\gamma : (-\epsilon,\epsilon) \to M$, $\gamma(0) = p$, and $\gamma_p' = v$. Define a curve $\beta = f\circ\gamma$ which is a differentiable curve through $f(p)$ with tangent $\beta'_{f(p)}$. Then the map $df_p : T_pM \to T_{f(p)}N$, $\gamma_p' \mapsto \beta'_{f(p)}$ is linear and does not depend on the choice of γ.*

Proof. Expressed in coordinates x and y, we can write

$$\psi \circ f \circ \phi^{-1} = \left(y^1\left(x^1,\ldots,x^m\right),\ldots,y^n\left(x^1,\ldots,x^m\right)\right).$$

Also $\phi \circ \gamma = \left(x^1(t),\ldots,x^m(t)\right)$. Then

$$\begin{aligned}\frac{d}{dt}(\psi\circ\beta)\Big|_{t=0} &= \frac{d}{dt}\left(\psi\circ f\circ\phi^{-1}\circ\phi\circ\gamma\right)\Big|_{t=0} \\ &= \left(\sum_{i=1}^m \frac{\partial y^1}{\partial x^i}\left(x^i\right)'(0),\ldots,\sum_{i=1}^m \frac{\partial y^n}{\partial x^i}\left(x^i\right)'(0)\right)\end{aligned}$$

and so

$$\frac{d}{dt}(\psi\circ\beta)^k\Big|_{t=0} = \frac{\partial y^k}{\partial x^i}\underbrace{\left(x^i\right)'(0)}_{\gamma_p'}.$$

This is a linear transformation written in bases $\left[\frac{\partial}{\partial x^i}\right]$ and $\left[\frac{\partial}{\partial y^i}\right]$. And the expression does not depend on the choice of β. □

Exercise 1.10. Given smooth $f : M \to N$, $v \in T_pM$, and a function $g : N \to \mathbb{R}$ differentiable at $f(p)$, show that

$$\left(df_p v\right) g\big|_{f(p)} = v\left(g\circ f\right)\big|_p .$$

Definition 1.11. If $f : M \to N$ is differentiable, bijective, and has differentiable inverse, then we say that f is a *diffeomorphism* between the manifolds M and N.

Sometimes it is useful to speak of local diffeomorphisms, which is a diffeomorphism in a neighborhood.

Theorem 1.12. *(Inverse Function Theorem) If $f : M \to N$ is a differentiable map whose differential at p is an isomorphism, then M and N are locally diffeomorphic.*

1.2. **Tensors on Manifolds.** Recall to a vector space V we can associate a *dual space* V^*, the space of linear functionals on V.

Definition 1.13. The *dual space* T_p^*M of T_pM is the vector space of linear maps $g : T_pM \to \mathbb{R}$. Its elements are called *one-forms* or *covectors*.

Choosing $N = \mathbb{R}$ and $g = \text{id}$ in the result of exercise 1.10 yields

Proposition 1.14. *The differential of $f : M \to \mathbb{R}$ lives in the dual space T_p^*M and acts on a vector $X_p \in T_p$ via the rule*

$$df_p\left(X_p\right) = X_p\left(f\right).$$

In particular, taking f to be the ith coordinate function x^i yields

$$dx^i\left[\frac{\partial}{\partial x^k}\right]_p = \delta^i_k,$$

so the dx^i form a basis (the dual basis) of T_p^*M.

What happens if we change charts? Write $\lambda = \lambda\left(\left[\frac{\partial}{\partial x^k}\right]_p\right) dx^k$.[2] If we choose λ to be $y^i(x)$, then we see

$$dy^i = \frac{\partial y^i}{\partial x^k}dx^k = A^i_k dx^k.$$

[2] We'll use Einstein notation (implied summation) from here on.

Using the property of the dual basis, we find

$$\begin{aligned}\left[\frac{\partial}{\partial x^i}\right]_p &= \frac{\partial x^k}{\partial y^i}\left[\frac{\partial}{\partial x^k}\right]_p \\ &= \left(A^{-1}\right)_i^k \left[\frac{\partial}{\partial x^k}\right]_p .\end{aligned}$$

For a general basis $e_1, \dots, e_n$ of T_pM and its dual basis $\omega^1, \dots, \omega^n$ of T_p^*M ($\omega^i(e_j) = \delta_j^i$) we have

$$\begin{aligned}\hat{\omega}^i &= A_k^i \omega^k, \\ \hat{e}_i &= \left(A^{-1}\right)_i^k e_k.\end{aligned}$$

To see the transformation law for the components note that

$$\begin{aligned}\lambda &= \hat{\lambda}_i \hat{\omega}^i \\ &= \hat{\lambda}_i A_j^i \omega^j \\ &= \lambda_i \omega^i.\end{aligned}$$

Definition 1.15. A $(1,2)$-*tensor* S on T_pM is a map $S : T_pM \times T_pM \times T_p^*M \to \mathbb{R}$ which is linear in each argument.

The general (p,q)-tensor is defined similarly. A tensor is determined by its action on a basis:

$$\begin{aligned}T(X,Y,\lambda) &= T\left(X^a e_a, Y^b e_b, \lambda_c \omega^c\right) \\ &= X^a Y^b \lambda_c T\left(e_a, e_b, \omega^c\right) \\ &= X^a Y^b \lambda_c T_{ab}^c.\end{aligned}$$

From this one can derive the transformation law for components:

$$\begin{aligned}T\left(\hat{e}_a, \hat{e}_b, \hat{\omega}^c\right) &= T\left(\left(A^{-1}\right)_a^d e_d, \left(A^{-1}\right)_b^e e_e, A_f^c \omega^f\right) \\ &= \left(A^{-1}\right)_a^d \left(A^{-1}\right)_b^e A_f^c T\left(e_d, e_e, \omega^f\right).\end{aligned}$$

Note. In general relativity one sees the "contraction" operation, which has the rule: If T_{ab}^c is a $(1,2)$-tensor, then T_{ab}^a is a $(0,1)$-tensor. One has to show this is indeed a tensor, i.e., it is linear in the remaining component.

Last time we introduced manifolds and tangent spaces and reduced the calculus of manifolds to calculus on Euclidean space. Given a map $f : M \to N$ we developed a notion of a differential mapping $df_p : T_pM \to T_{f(p)}\mathrm{N}$. This map is also called the "push-forward" since it pushes vectors from T_pM to $T_{f(p)}N$. Finally we discussed tensors and corresponding change of coordinate formulas.

Exercise 1.16. Prove that if the components of a tensor all agree in one basis, then they all agree in all others.

1.3. **Vector Fields and Derivations.** First we collect together the tangent spaces into a global object.

Definition 1.17. Given a manifold M, its *tangent bundle* is the set $TM = \cup_{p\in M} T_pM$.

Proposition 1.18. *TM is a manifold of dimension $2n$.*

Now we can discuss the notion of a vector field.

1.3.1. *Vector Fields.*

Definition 1.19. A *vector field* X is a (smooth) map $X : M \to TM$ such that $X(p) = X_p$ is a vector in T_pM.

Recall we defined a tangent vector as an operator on differentiable functions at a point. Now we can do this globally: we can think of a vector field as a map from $\mathcal{F}(M) \to \mathcal{F}(M)$. So we can consider iterates of vector fields, i.e., $XY(f)$ or $YX(f)$. Note, however, that either of these alone will not in general be a vector field. Indeed,

$$\begin{aligned} XYf &= X_p^i \left[\frac{\partial}{\partial x^i}\right]_p \left(Y^m \frac{\partial\left(f \circ \phi^{-1}\right)}{\partial x^m}\right) \\ &= X_p^i Y_p^m \frac{\partial\left(f \circ \phi^{-1}\right)}{\partial x^i \partial x^m} + X_p^i \left[\frac{\partial}{\partial x^i}\right]_p Y^m \cdot \underbrace{\frac{\partial\left(f \circ \phi^{-1}\right)}{\partial x^m}}_{\left[\frac{\partial}{\partial x^m}\right]_p f}, \end{aligned}$$

and the presence of the second derivative term prevents this from being a vector field. However,

$$\left(X_pY - Y_pX\right)(f) = \left(X_p^i \left[\frac{\partial}{\partial x^i}\right]_p Y^m - Y_p^i \left[\frac{\partial}{\partial x^i}\right]_p X^m\right) \left[\frac{\partial}{\partial x^m}\right]_p f$$

is a vector field.

Definition 1.20. The *commutator* of X and Y is the vector field given by the expression above, and is written as $[X, Y]$.

1.3.2. *The Lie Derivative.* To understand the geometric significance of the commutator, we must first understand flows/integral curves.

Definition 1.21. Given a smooth vector field X, we call $\gamma : I \to M$ an *integral curve* of the vector field X if at each point along γ the tangent vector to γ is X_p.

Theorem 1.22. *Given a smooth vector field X on M, there exists locally an integral curve through each point $p \in M$.*

In components, γ is an integral curve if it satisfies the ode

$$\frac{d}{dt}\left(x^i \circ \gamma(t)\right) = X^i\left(x \circ \gamma(t)\right)$$

subject to initial conditions $x^i \circ \gamma(0) = x_p^i$. One can prove this ode is locally solvable, and that the solution has continuous dependence on the data. This gives rise to the notion of local flow, $\phi_t : (-\epsilon, \epsilon) \times U \to M$. Here, U is a neighborhood of p.

Exercise 1.23. Suppose X is a vector field non-vanishing at p. Then locally near p one can find coordinates such that $X = \frac{\partial}{\partial x^1}$. Hint: choose a hypersurface S through p which is transversal to X_p. Then find a diffeomorphism $\{-\epsilon, \epsilon\} \times X \to U$ which in the given coordinates sends $\left(x^1, x^2, \ldots, x^n\right) \mapsto \left(x^1 + t, x^2, \ldots, x^n\right)$.

Theorem 1.24. *Let X, Y be smooth vector fields on M, $p \in M$ and ϕ_t the local flow of X near p. Then*

$$[X,Y]_p = \lim_{t\to 0} \frac{1}{t}\left[Y_{\phi_t(p)} - d\phi_t Y_p\right].$$

Proof. Use the chart such that $X = \frac{\partial}{\partial x^1}$. Then we have

$$\begin{aligned}\lim_{t\to 0}\frac{1}{t}\left[Y^i\left(x^1+t,x^2,\ldots,x^n\right) - Y^i\left(x^1,x^2,\ldots,x^n\right)\right] &= \left.\frac{\partial Y^i}{\partial x^1}\right|_p \\ &= X_p\left(Y^i\right) \\ &= X_p\left(Y^i\right) - \underbrace{Y_p\left(X^i\right)}_{=0} \\ &= [X,Y]_p^i\end{aligned}$$

which completes the proof. □

Definition 1.25. The *Lie derivative* of Y along X is the vector field satisfying

$$\mathcal{L}_X Y(p) = \lim_{t\to 0}\frac{1}{t}\left[Y_{\phi_t(p)} - d\phi_t Y_p\right].$$

As we defined vector fields, we also have the notion of tensor fields. These are multi-linear maps which are linear with respect to $\mathcal{F}(M)$, in that

$$T(fX+gY,Z) = fT(X,Z) + gT(Y,Z).$$

This implies that tensors are local objects, i.e., the value of $T(X,Y)$ at a point p depends only on the values of X and Y at p. Note, however, that this is false for the Lie derivative, which depends on derivatives as well.

1.3.3. *Connections and Covariant Derivatives.* As it turns out, Lie derivatives are not general enough to relate vectors in different tangent spaces. So we have the notion of a connection.

Definition 1.26. A *connection* ∇ on M is a map $\mathcal{X}(M)\times\mathcal{X}(M) \ni (X,Y) \mapsto \nabla_X Y \in \mathcal{X}(M)$ ($\mathcal{X}(M)$ is the set of smooth vector fields on M) such that for $f : M \to \mathbb{R}$ a function we have the following properties:

(1) $\nabla_X(fY) = f\nabla_X Y + (Xf)Y$
(2) $\nabla_{X+fY} Z = \nabla_X Z + f\nabla_Y Z$
(3) $\nabla_X f = X(f)$.

Definition 1.27. The vector field $\nabla_X Y$ is called the *covariant derivative* of Y with respect to X. The $(1,1)$-tensor ∇Y is called the *covariant derivative* of Y.

Remark. That ∇Y is a $(1,1)$-tensor follows from property (2) in the definition above.

Example 1.28. Do connections exist? For instance define, given $X = X^i\partial_i$,

$$\nabla_Y X = \left(Y^i\partial_i X^k\right)\partial_k.$$

In $\mathbb{R}^2$, ∂_x and ∂_y would satisfy

$$\nabla_{\partial_x}\partial_x = \nabla_{\partial_y}\partial_y = \nabla_{\partial_x}\partial_y = \nabla_{\partial_y}\partial_x = 0.$$

But note that in polar coordinates on $\mathbb{R}^2$,

$$\partial_r = \frac{1}{r}\left(x\partial_x + y\partial_y\right)$$
$$\partial_\phi = -y\partial_x + x\partial_y$$

so the connection must transform in a non-obvious way.

Definition 1.29. We say that Y is *parallely transported* along X if $\nabla_X Y = 0$.

Let's work out the covariant derivative in a (moving orthonormal) basis. Write $\nabla_{e_a} e_b = \Gamma^c_{ab} e_c$ (the Γ^c_{ab} are the "connection coefficients"), then

$$\begin{aligned}\nabla_X Y &= \nabla_{X^a e_a}\left(Y^b e_b\right)\\ &= X^a \nabla_{e_a}\left(Y^b e_b\right)\\ &= X\left(Y^b\right) e_b + X^a Y^b \Gamma^b_{ac} e_b.\end{aligned}$$

So specifying the connection coefficients specifies a covariant derivative. The components satisfy

$$\left(\nabla_X Y\right)^b = X\left(Y^b\right) + X^a Y^c \Gamma^b_{ac}.$$

Example 1.30. For $\mathbb{R}^2$ and $e_1 = \partial_x$, $e_2 = \partial_y$ we have $\Gamma = 0$. For polar coordinates we have $e_r = \partial_r$ and $e_\phi = \frac{1}{r}\partial_\phi$, so $\nabla_{e_\phi} e_r \neq 0$. Intuitively, this is because e_1 and e_2 are parallely transported along themselves, whereas the basis e_r, e_ϕ changes as it is paralelly transported in the ϕ-direction.

Note. We said that if the components of a tensor are zero in one basis, then they will be zero in all bases. But the connection symbols change from basis to basis! The previous ex shows that Γ^c_{ab} cannot be a tensor.

Exercise 1.31. Show that if $\hat{e}_a = B^b_a e_b$, then

$$\hat{\Gamma}^a_{bc} = \left(B^{-1}\right)^a_f B^g_b B^h_c \Gamma^f_{gh} + \left(B^{-1}\right)^a_f B^h_c e_h \left(B^f_b\right).$$

Hint: we know $\nabla_{\hat{e}_a} \hat{e}_b = \hat{\Gamma}^c_{ba} \hat{e}_c$. Now substitute $\hat{e}_a = B^b_a e_b$ and use the properties of the connection to expand.

Remark. The difference of two connections is a tensor:

$$D\left(X, Y\right) = \nabla_X Y - \tilde{\nabla}_X Y.$$

Now we'll generalize the covariant derivative to covectors. If ω is a one-form, then $\nabla_X \omega$ should be a one-form as well. For the Leibniz rule to hold, we must have

$$\nabla_X \omega\left(Y\right) = \nabla_X\left(\omega\left(Y\right)\right) - \omega\left(\nabla_X Y\right).$$

In this way we can define the covariant derivative of a covector.

Exercise 1.32. Show that

$$\left(\nabla_a \omega\right)_b = \partial_a \omega_b - \Gamma^c_{ab}\omega_c.$$

Here, $\nabla_a \omega$ means $\nabla_{e_a}\omega$.

1.4. **Geodesic Flow.** We'll take a closer look at the equation of parallel transport. Suppose we have a curve γ with tangent vector $X = X^a \partial_a = \partial_t$. Then in components the condition on Y for Y to be parallely transported is

$$X^a \nabla_a Y^c + \Gamma^c_{ab} X^a Y^b = 0$$

or just

$$\frac{d}{dt} Y^c + \Gamma^c_{ab} X^a Y^b = 0.$$

This is a linear ode, and can be solved for Y^c subject to initial conditions.

Definition 1.33. Let X be a vector field such that $\nabla_X X = 0$. Then the integral curves of X are called *geodesics*.

Theorem 1.34. *There is precisely one geodesic through each point p whose tangent vector at p is equal to X_p.*

Proof. We'll use an ode theorem. We have

$$\frac{d}{dt} X^c + \Gamma^c_{ab} X^a X^b = 0$$

and since

$$X^c = \frac{d}{dt} x^c (t) = \frac{d}{dt} \left(x^c \circ \gamma (t) \right),$$

we in fact have

$$\frac{d^2}{dt^2} x^c (t) + \Gamma^c_{ab} \dot{x}^a \dot{x}^b = 0. \tag{1.1}$$

Once we specify initial conditions an ode theorem on existence and uniqueness gives the result. □

Remark. Equation (2.1) is referred to as the "geodesic equation."

1.5. **Torsion and Curvature.** Having specified a connection, one can introduce some geometrically important tensors. Here are two now.

Definition 1.35. The *torsion tensor* is a $(1, 2)$-tensor field T defined by

$$T (X, Y) = \nabla_X Y - \nabla_Y X - [X, Y].$$

One must check this is indeed a tensor field; i.e., one should check this is linear in each argument. As T is anti-symmetric, one only needs to check it is linear in X. This is a calculation.

Note. We will work with torsion-free connections (symmetric connnections). These are called as such as $T = 0$ implies the anti-symmetric part of Γ^c_{ab} is zero. (That is, $\Gamma^c_{ab} - \Gamma^c_{ba} = 0$.)

Definition 1.36. The *Riemann curvature tensor* is a $(1, 3)$-tensor field defined by

$$R (X, Y) Z = \nabla_Y \nabla_X Z - \nabla_Y \nabla_X Z + \nabla_{[X,Y]} Z.$$

In a coordinate basis, this is

$$R^a_{bcd} = \Gamma^a_{bc,d} - \Gamma^a_{bd,c} - \Gamma^c_{pc} \Gamma^p_{bd} + \Gamma^a_{pd} \Gamma^p_{bc}.$$

This tensor measures the failure of covariant derivatives to commute. Imagine parallel transport of a vector along a triangle on the surface of a two-sphere. The resulting vector will be different than the initial vector, and indeed $R^a_{bcd} \neq 0$.

2. Local/Global Riemmanian Geometry

So far we have discussed the differential calculus of manifolds. We introduced additional structure via a connection, which gave rise to the notions of curvature and torsion. In this section, we'll endow our manifolds with a metric, $g(v,w)$, a symmetric bilinear form on each tangent space. In general, g can be positive definite (Riemannian geometry) or indefinite of signature $(-,+,\dots,+)$ (Lorentzian geometry). Eventually, we'll try to understand how these geometries are similar/different. But, as we'll see now, there is a unique connection for which torsion vanishes, the so-called Levi-Civita connection; this will be the connection we work with for the rest of the course.

2.1. **Metrics and the Levi-Civita Connection.**

Definition 2.1. A *metric tensor* g at a point p is a symmetric $(0,2)$-tensor. It assigns a *magnitude* $d(X) = \sqrt{g(X,X)}$ to each vector X, and we'll say that X and Y are *orthogonal* if $g(X,Y) = 0$.

Definition 2.2. The *length* of a curve $\gamma(t) : [t_1, t_2] \to M$ is defined as

$$l(\gamma) = \int_{t_1}^{t_2} \sqrt{g(X,X)}\, dt.$$

In a coordinate basis we can write

$$g = g_{ab} dx^a \otimes dx^b.$$

Definition 2.3. A smooth assignment of a metric to each tangent space is called a *metric* on M.

Definition 2.4. We say (M,g) is a *Riemannian manifold* if g is positive definite. We say (M,g) is *Lorentzian* if g has signature $(-,+,\dots,+)$.

Note. Positive definiteness implies non-degeneracy, i.e., that $g(X,Y) = 0$ for all Y implies $X = 0$. This gives an isomorphism $T_pM \ni v \mapsto g(v,\cdot) \in T_p^*M$. This allows us to observe the convention of raising and lowering indicies. If v^a are the components of a vector, then $v_b = g_{ab}v^a$ is the metrically equivalent co-vector. Also, $\left(g^{-1}\right)^{ab} g_{bc} = \delta^b_c$, but most books write g^{ab} to mean the inverse of g_{ab}.

Definition 2.5. The *gradient* of $f : M \to \mathbb{R}$ is the vector metrically equivalent to df, i.e., the vector satisfying

$$g(\operatorname{grad} f, \cdot) = df.$$

Now here is a miracle which simplifies much of our work.

Theorem 2.6. *For a pseudo-Riemannian manifold (M,g) there is a unique torsion-free connection which is metric compatible, i.e., which satisfies $\nabla g = 0$.*

Proof. See exercise 2.15 the course notes. □

We'll use the connection described above implicitly throughout the rest of the course: it is called the *Levi-Civita connection.* Now we are ready to discuss local Riemannian geometry.

2.2. **The Exponential Map.** Recall the geodesic equation

$$\ddot{x}^a(t) + \Gamma^a_{bc}\dot{x}^b(t)\dot{x}^c(t) = 0. \tag{2.1}$$

In flat Euclidean space the geodesics are straight lines (in cartesian coordinates $\Gamma^a_{bc} = 0$). Any two points are connected by a unique geodesic, which minimizes length. We ask two questions:

- Can we always connect two points by a geodesic?
- Do geodesics minimize the length?

Both are false in general. The first fails for the punctured plane. The second fails on the sphere. But perhaps we can come up with conditions under which these are true.

Write $x^a(t) = x^a \circ \gamma(t)$ and $\frac{d}{dt}x^a(t) = \frac{d}{dt}(x^a \circ \gamma(t)) = X^a$. With these we can think of equation (2.1) as a first order system:

$$\frac{d}{dt}x^a(t) = X^a(t)$$
$$\frac{d}{dt}X^a(t) = -\Gamma^a_{bc}(x)X^bX^c,$$

subject to initial conditions $x^a(0) = x^a_p$, $X^a(0) = X^a_p$. Now given $q \in M$ and a neighborhood V of q, define the set

$$\mathcal{U}_{V,\epsilon} = \left\{(p,v) \;\middle|\; p \in V, v \in T_pM, ||v|| < \epsilon\right\}.$$

Applying the usual ode theorem for existence and uniqueness (and continuous dependance on the data) immediately gives the following result:

Proposition 2.7. *Given $q \in M$ we can find a neighborhood V and an ϵ such that the map $\gamma : (-\delta,\delta) \times \mathcal{U}_{V,\epsilon} \to M$, $(t,p,v) \mapsto \gamma(t,p,v) =$the unique geodesic passing through p with tangent vector v at $t = 0$ is smooth.*

Lemma 2.8. *If $\gamma(t,p,v)$ is defined for $(-\delta,\delta)$ then $\gamma(t,p,av)$ is defined for $(-\delta/a,\delta/a)$ and $\gamma(t,p,av) = \gamma(at,p,v)$.*

Proof. $\gamma(t,p,v)$ is defined for $(-\delta,\delta)$ implies that $h(t) = \gamma(at,p,v)$ is defined for $(-\delta/a,\delta/a)$. Now check

$$\nabla_{h'(t)}h'(t) = a^2\nabla_{\gamma'}\gamma' = 0.$$

Since $h(0) = p$ and $h'(0) = av$, uniqueness gives the result. □

Remark. The point here is that we can always rescale the interval of time on which γ is defined to $(-1,1)$.

Definition 2.9. Let $q \in M$ and $\mathcal{U}_{V,\epsilon}$ be as above. The map $\exp : \mathcal{U}_{V,\epsilon} \to M$ given by

$$\exp(q,v) = \gamma(1,q,v) = \gamma\left(||v||, q, \frac{v}{||v||}\right)$$

is called the *exponential map.*

We'll be interested in the restriction to a point p, i..e, the map $\exp_p : T_pM \to M$ given by

$$\exp_p(v) = \gamma(1,p,v) = \exp(p,v).$$

Proposition 2.10. *The exponential map is a local diffeomorphism at p.*

Proof. Note that $d\left(\exp_p\right)_0 : T_0T_pM \to T_pM$. Usually we make the identification $T_0T_pM \simeq T_pM$. Let's compute this differential:

$$\begin{aligned} d\left(\exp_p\right)_0 v &= \frac{d}{dt}\left(\exp_p(tv)\right)\Big|_{t=0} \\ &= \frac{d}{dt}\gamma(1,p,tv)\Big|_{t=0} \\ &= \frac{d}{dt}\gamma(t,p,v)\Big|_{t=0} \\ &= v. \end{aligned}$$

So the $d\left(\exp_p\right)_0$ is the identify, and hence the claim. □

Definition 2.11. A neighborhood U in M which is the diffeomorphic image of $\exp_p$ of a neighborhood in T_pM is called a *normal neighborhood*.

Remark. One can refine this to a *totally normal neighborhood*, i.e., a neighborhood which is a normal neighborhood of each of its points. This will be useful later.

Now we can describe the notion of normal coordinates. Given q in a normal neighborhood we can write $q = \exp_p(v)$ for unique $v \in T_pM$. Now choose an orthonormal basis e_a in T_pM and write $v = x^a(q)\,e_a$. The $x^a(q)$ are called "normal coordinates" of q. Using the exponential map we can solve for the $x^a(q)$:

$$\begin{aligned} q &= \exp_p\left(x^a(q)\,e_a\right) \\ \implies x^a(q) &= \omega^a \exp_p^{-1}(q) \end{aligned}$$

where ω^a is the dual basis to e_a.

Exercise 2.12. Show that in normal coordinates $g_{ij}(p) = \delta_{ij}$ and $\Gamma^i_{jk}(p) = 0$. Hint: in these coordinates, $x^a(\gamma(t)) = tv^a$ where $v = \gamma'$.

2.3. **Minimizing Properties of Geodesics.** Our goals are now to show

(1) If $p, q \in M$ and there is a curve γ which minimizes the length between p and q, then γ must be a geodesic.
(2) For a sufficiently small neighborhood, one can connect any two points by a geodesic. Moreover, this geodesic minimizes the length.

The solution of the first involves a variational principle. In the exercises we'll consider variations of the length functional

$$l(\gamma) = \int_a^b \sqrt{g(\gamma',\gamma')}\,dt,$$

and we'll see that in order to extremize this functional γ must satisfy the geodesic equation.

We discuss (2) now. The key lemma is the so-called "Gauss lemma."

Lemma 2.13 (Gauss)**.** *Let $p \in M$ and $v \in T_pM$ be so that $\exp_p(v) = q$ is defined. Then for any $w \in T_pM$ we have*

$$g\left(d\left(\exp_p\right)_v v, d\left(\exp_p\right)_v w\right) = g(v,w).$$

Proof. Consider the parametrized surface $f(t,s) = \exp_p(t(v+sw))$. We compute

$$\frac{\partial f}{\partial t}(t,0) = d\left(\exp_p\right)_{tv} v,$$
$$\frac{\partial f}{\partial s}(t,0) = d\left(\exp_p\right)_{tv} tw.$$

Note that $d\left(\exp_p\right)_{tv} v$ is the tangent vector to the radial geodesic. Recall that the norm of the tangent vector does not change along a geodesic. So for $w = v$ the statement is clear; in fact it is clear for $w = \alpha v$ for any α. So we can assume $w \perp v$.

Observe that

$$g\left(d\left(\exp_p\right)_v v, d\left(\exp_p\right)_v w\right) = g\left(\frac{\partial f}{\partial t}(1,0), \frac{\partial f}{\partial s}(1,0)\right).$$

We must show this is zero. We have that

$$\lim_{t\to 0} g\left(\frac{\partial f}{\partial t}(1,0), \frac{\partial f}{\partial s}(1,0)\right) = 0,$$

but $g\left(\frac{\partial f}{\partial t}, \frac{\partial f}{\partial s}\right)$ is constant in t! □

Exercise 2.14. Finish the proof by showing $\frac{\partial}{\partial t} g\left(\frac{\partial f}{\partial t}, \frac{\partial f}{\partial s}\right) = 0$.

Let's try to build intuition about the lemma. Recall the map $\exp_p$ gives rise to normal coordinates in the manifold. The following will be useful:

T_pM	$\xrightarrow{\exp_p}$	M
$\tilde{\tau} : T_pM \to \mathbb{R}$		$\tau : M \to \mathbb{R}$
$\tilde{\tau}^2(x) = \left(x^1\right)^2 + \cdots + (x^n)^2$		$\tau^2 = \tilde{\tau}^2 \circ \exp_p^{-1}$
$\left(\tilde{\tau}^2\right)^{-1}(c)$ are hyperspheres		$\left(\tau^2\right)^{-1}(c)$ are distorted hyperspheres
$\operatorname{grad}\tilde{\tau}^2 = 2\tilde{P} = 2x^i\partial_i$		$P = d\left(\exp_p\right)_x \tilde{P}$, tangent to radial geodesics
$g\left(\operatorname{grad}\tilde{\tau}^2, \tilde{w}\right) = \tilde{w}\left(\tilde{\tau}^2\right) = 0$		$\operatorname{grad}\tau^2 = 2P$
($\tilde{w}$ is tangent to the sphere)		(For proof see below.)
$\tilde{U} = \frac{\tilde{P}}{r}$		$U = \frac{P}{r}$, unit radial vector

Proof of $\operatorname{grad}\tau^2 = 2P$: Choose $v = d\exp_p \tilde{v}$ and compute

$$\begin{aligned} g\left(\operatorname{grad}\tau^2, v\right) &= v\left(\tau^2\right) = d\exp_p \tilde{v}\left(\tau^2\right) \\ &= \tilde{v}\left(\tau^2 \circ \exp_p\right) = \tilde{v}\left(\tilde{\tau}^2\right) = g\left(\operatorname{grad}\tilde{\tau}^2, \tilde{v}\right) \\ &= g\left(\tilde{P}, \tilde{v}\right) = g(P, v). \end{aligned}$$

Now we can resolve (2).

Proposition 2.15. *Let $p \in M$ and B a normal ball around p. Let $\gamma : [0,1] \to B$ be the radial geodesic connecting p and q, i.e., $\gamma(0) = p$ and $\gamma(1) = q$. ($\gamma(t) = \exp_p(tv) \implies q = \exp_p(v)$.) Then any competing (piecewise differentiable) curve $\alpha : [0,1] \to M$ joining p to q has longer length unless α is a monotone reparameterization of γ, in which case equality holds.*

Proof. The length of α is

$$l(\alpha) = \int_0^1 \sqrt{g(\alpha'(t), \alpha'(t))}\, dt.$$

For γ we have

$$\begin{aligned} l(\gamma) &= \int_0^1 \sqrt{g(\gamma', \gamma')}\, dt \\ &= \int_0^1 \sqrt{g(v, v)}\, dt \\ &= \sqrt{g(v, v)} \\ &= r(q), \end{aligned}$$

the radial distance. Look at the competitor curve. Decompose α' as

$$\alpha' = g(\alpha', U)\, U + N$$

where $U \perp N$ (N is tangent to the geodesic ball). Compute

$$\begin{aligned} \sqrt{g(\alpha', \alpha')} &= \sqrt{(g(\alpha', U))^2 + g(N, N)} \\ &\geq \sqrt{(g(\alpha', U))^2} \\ &\geq g(\alpha' U), \end{aligned}$$

so we can write

$$\begin{aligned} \int_0^1 \sqrt{g(\alpha', \alpha')}\, dt &\geq \int_0^1 g(\alpha', U)\ dt \\ &= \int_0^1 g(\alpha', \operatorname{grad} \tau)\ dt \\ &= \int_0^1 \frac{d}{dt}(r \circ \alpha)\ dt \\ &= r(q). \end{aligned}$$

When equality holds, show γ and α have the same image in M, and also that $\frac{dr}{dt} > 0$. These will show that α is a monotonic reparametrization of γ. □

Remark. The proof above assume α sits in the normal neighborhood. But if it leaves, then part of it sits in the normal neighborhood and the proof carries over there. And the length of the peice outside is non-zero, so the result still holds.

The characterization of geodesics as locally length-minimizing does not extend as a global result in general. Indeed, consider the geodesic paths connecting two points on the two-sphere. Next, we'll see exactly when this characterization extends to a Riemannian manifold in the large.

2.4. **The Hopf-Rinow Theorem.** Let (M, g) be a connected Riemannian manifold. Define

$$\Omega_{p,q} = \{\text{piecewise differentiable curves connecting } p \text{and } q\}$$

and then define a distance

$$d(p, q) = \inf_{\gamma \in \Omega_{p,q}} \{l(\gamma)\}.$$

Exercise 2.16. Check this is a distance. Hint: for $d(p,q)=0 \implies p=q$ use the Hausdorff criterion.

The distance d turns (M,d) into a metric space, and induces a topology on M via balls.

Proposition 2.17. *The topology induced on M by d is the same as the manifold topology.*

Corollary 2.18. *The function $p \mapsto d(p,\cdot)$ is continuous.*

Definition 2.19. We say M is *geodesically complete* if $\exp_p$ is defined for all $v \in T_pM$ at each $p \in M$.

Theorem 2.20. *Let (M,g) be a Riemannian manifold and $p \in M$. Then the following are equivalent:*

1. *exp_p is defined for all $v \in T_pM$*
2. *The closed and bounded sets of M are compact*
3. *M is complete as a metric space*
4. *M is geodesically complete.*

In addition, any of (1)-(4) implies

1. *For any $q \in M$ there exists a geodesic connecting p and q whose length realizes the distance $d(p,q)$.*

Proof. First show (1) $\implies$ (5). Fix p and q and call $d(p,q)=r$. Pick a normal ball around p, $B_\delta(p)$ with boundary $S_\delta(p)$. This boundary will be compact, so the map $S_\delta(p) \to \mathbb{R}$, $\tilde{p} \mapsto d(\tilde{p},q)$ acheives a minimum. Say this minimum is at p', then there exists $v \in T_pM$ so that $p' = \exp_p(\delta v)$ for $||v||=1$. We want to show that in fact $q = \exp_p(rv)$.

To do this we'll use an open/closed argument. Consider the set

$$A = \left\{ s \in [0,r] \;\middle|\; d\left(q, \exp_p(sv)\right) = r-s \right\}.$$

It is not empty, since $0 \in A$. It is also closed, for d and exp are continuous. We will show now that if $s_0 \in A$ then $s_0+\delta' \in A$ for sufficiently small δ'. (It follows that $[0,r) \subset A$ and so by closedness $A=[0,r]$.) Choose a normal ball around x_0, $B_{r'}(x_0)$ with boundary $S_{\delta'}(x_0)$. Say d is minimized on $S_{\delta'}(x_0)$ at the point x_0'. We have

$$d(x_0,q) = \delta' + d(x_0',q),$$

and since $x_0 = \exp_p(s_0,v)$ the left hand side is $r-s_0$, and thus

$$d(x_0',q) = r - s_0 - \delta'.$$

But also

$$d(p,q) \le d(p,x_0') + d(x_0',q)$$

and thus

$$d(p,x_0') \ge r - (r - s_0 - \delta') = s_0 + \delta'.$$

There is a curve which acheives equality, the (possibly broken) geodesic which goes first to x_0 and then from x_0 to x_0'. But since this curve extremizes, it must not be broken (if it were, find a totally normal ball about the break point and produce a shorter curve). By uniqueness we get

$$x_0 = \exp_p((s_0+\delta')v)$$

and thus $s_0 + \delta' \in A$. This completes the argument.

Now we'll do (1)+(5) $\implies$ (2). Let U be closed and bounded. By (5) and boundedness, we have that

$$U \subset \exp_p \left(\overline{B_R(0)} \right)$$

for sufficiently large R. Since U is a closed subset of a compact set, it must be compact.

The rest will be an exercise. $\square$

Exercise 2.21. Show (2) $\implies$ (3) and (3) $\implies$ (4) to finish the proof.

3. Lorentzian Geometry and Special Relativity

We are working towards general relativity and the Einstein equations. As we'll see, a solution to these equations takes the form of a Lorentzian manifold and its $(-,+,+,+)$ metric. Thus we'll consider now some basic results in Lorentzian geometry, through the lens of special relativity.

The domain of special relativity is $\mathbb{R}^{3+1}$ equipped with the Minkowski metric

$$\eta = \eta_{\mu\nu} dx^\mu \otimes dx^\nu,$$

where

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & & & \\ & 1 & & \\ & & 1 & \\ & & & 1 \end{pmatrix}.$$

Note. Sometimes this is presented as

$$\eta = -dt^2 + dx^2 + dy^2 + dz^2.$$

The first question we ask is: What are the isometries? So we look at the Killing fields, i.e., $\mathcal{L}_K \eta = 0$. From the exercises in the course notes, we know

$$\nabla_a K_b + \nabla_b K_a = 0,$$
$$\nabla_a \nabla_b K_c = R^d_{abc} K_d.$$

But Minkowski space is flat, so this is

$$\partial_a K_b + \partial_b K_a = 0,$$
$$\partial_a \partial_b K_c = 0.$$

These equations admit the solution

$$K_c = a_{cb} x^b + b_c$$

for antisymmetric a_{cb}. So the Killing fields are the translations $\partial_t, \partial_x, \partial_y, \partial_z$; the rotations $(-x\partial_y + y\partial_x)$ + two others; and the Lorentz boosts $(x\partial_t + t\partial_x)$ + two others. These transformations generate the Poincare group.

This was not how special relativity was described classically. In those days, one would search for variable transformations which preserved solutions to Maxwell's equations – these are the Poincare transformations. In the language of differential geometry, one would consider transformations which preserve the wave equation

$$\eta^{\mu\nu} \partial_\mu \partial_\nu \psi = 0.$$

On a generic manifold this equation depends on the coordinate frame; the completely covariant presentation is

$$\Box_g \psi = g^{\mu\nu} \nabla_\mu \nabla_\nu \psi = 0.$$

We'll come back to this equation again later in the course.

Consider the $1+1$ picture, so $g = -dt^2 + dx^2$. We have $g(\partial_t, \partial_t) = -1$, $g(\partial_x, \partial_x) = +1$, but $g(\partial_t \pm \partial_x, \partial_t \pm \partial_x) = 0$. So we say ∂_t is "timelike," ∂_x is "spacelike," and $\partial_t \pm \partial_x$ is a "null-vector." Note that $g(v,v) = -1$ is a hyperboloid. This is quite different from Euclidean geometry!

Definition 3.1. Let (M, g) be a Lorentzian manifold and $v \in T_pM$ for some $p \in M$. We will call v

- *spacelike* if $g(v, v) > 0$
- *timelike* if $g(v, v) < 0$
- *null* if $g(v, v) = 0$ (and $v \neq 0$).

Also we'll call v *causal* if it is either timelike or null.

Remark. This definition will be inherited by curves: a curve will be spacelike if its tangent is always spacelike, etc.

Exercise 3.2. Let X, Y be non-zero in T_pM satisfying $g(X, Y) = 0$. Prove the following statements:

a) If X is timelike, Y is spacelike.
b) If X is null, Y is spacelike or null.
c) If X is spacelike, everything is possible.
Hint: for (a) choose a basis such that $e_0 = X$.

Lemma 3.3. *Let $X, Y \in T_pM$ be causal. Define $||X|| = \sqrt{-g(X, X)}$. Then*

- $|g(X, Y)| \geq ||X||||Y||$
- $|g(X + Y, X + Y)| \geq ||X|| + ||Y||$ *if X and Y point in the same half of the light-cone.*

Proof. For the first statement, it suffices to assume $||X|| = 1$. Then write $e_0 = X$ then complete to an orthonormal basis e_i. Decompose Y as $Y = Y^0 e_0 + Y^1 e_1 + \cdots + Y^n e_n$, then we must verify

$$|Y^0| \geq 1 \cdot \sqrt{|Y^0|^2 - (\cdots)^2}.$$

The second follows easily. □

Exercise 3.4. Fill in the details above.

Remark. The "reverse triangle inequality" above explains the twin-paradox of special relativity. It also demonstrates that time-like curves cannot be length minimizing.

4. General Relativity

General relvativity was Einstein's answer to Newtonian mechanics. In this section we'll first see the equations and an important solution. Then we'll discuss some of the associated mathematical questions.

4.1. **The Einstein Equations.**

What are the Lorentzian manifolds of (physical) interest? These are the one which satisfy the Einstein equations:

$$R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} = 8\pi T_{\mu\nu}$$

on a Lorentzian manifold (M, g). Here $R_{\mu\nu} = R^{s}_{\mu s v}$ is the Ricci-tensor, $R = g^{\mu\nu} R_{\mu\nu}$ is the Ricci-scalar, and $T_{\mu\nu}$ is the energy momentum tensor of matter. This equation replaces Newton's equation,

$$\triangle \phi = -4\pi\rho$$

for the gravitational potential ϕ. Here ρ is the matter distribution. Upon specifying ϕ one has the force

$$F = -\nabla \phi.$$

The Einstein equations are horribly more complicated than Newton's equations. Indeed, if $\rho = 0$ and if we expect ϕ to vanish far from the origin, then we must have $\phi \equiv 0$. This is not true in Einstein's theory! Indeed, Minkowski space is one solution where $T_{\mu\nu} = 0$.

How does one study the Einstein equations? Formally we might write

$$\begin{aligned} R &= \partial\Gamma + \Gamma\Gamma \\ &= \partial\partial g + (\partial g)(\partial g). \end{aligned}$$

The leading order part is $g^{\mu\nu}\partial_\mu\partial_\nu g_{\alpha\beta}$, but only in harmonic coordinates. Recall the wave equation:

$$\Box\psi = \left(-\partial_t^2 + \partial_x^2 + \partial_y^2 + \partial_z^2\right)\psi = 0.$$

This can be written as

$$\eta^{\mu\nu}\partial_\mu\partial_\nu\psi = 0,$$

which looks quite similar. To study the wave equation one considers the Cauchy problem. The similarity suggests that to study the Einstein equation one should set up initial data and find solutions forward in time. However this is quite complicated, and took several decades to formulate. Next we'll discuss another approach, via Killing fields (symmetry).

4.2. **Schwarzschild's Solution.**

Schwarzschild arrived at his famous solution by modelling a star inside of a vacuum, and in particular examining space-time outside of the star. His solution, which solves the vacuum Einstein equations

$$R_{\mu\nu} = 0,$$

is the metric

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

on $(-\infty, \infty) \times (2M, \infty) \times S^2$. Note that M is a parameter and is the mass of the star in question.

Observe first that this is a static metric: ∂_t is Killing and

$$g(\partial_t, \partial_t) = -\left(1 - \frac{2M}{r}\right) = -(1-\mu) < 0$$

for $r > 2M$. So this is a timelike Killing field. Also, notice that for large r this approaches the Minkowski metric in polar coordinates,

$$g = -dt^2 + dr^2 + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right).$$

This makes sense, for we are only studying an isolated star. More interesting, however, is that there is a singularity at $r = 2M$ and $r = 0$. Schwarzchild only considered this solution on $r \geq r_0 > 2M$, for his matter model told him the star must have radius $r_0 > 2M$. We'll discuss now $r = 2M$ via geodesics and a coordinate transformation. Throughout, we'll time orient by declaring ∂_t to be future pointing. We ask the basic question: Can the domain of the metric be extended?

Example 4.1. Let

$$g = -\frac{1}{t^4}dt^2 + dx^2 \tag{4.1}$$

on $-\infty < x < \infty$, $0 < t < \infty$. Consider the transformation $t \mapsto t' = 1/t$. Then we have

$$g = -dt'^2 + dx^2.$$

Clearly this makes sense for $-\infty < t' < \infty$. Note that (4.1) is geodesically incomplete as $t \to \infty$. But by extending we can complete the geodesics.

With this in mind, let's consider geodesics in Schwarzschild space. Let $u^\mu = \frac{dx^\mu}{d\tau}$ denote the tangent vector of a geodesic, with τ an affine parameter. Then

$$g_{\mu\nu}u^\mu u^\nu = -(1-\mu)\dot{t}^2 + \frac{1}{1-\mu}\dot{r}^2 + r^2\left(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2\right) = \begin{cases} -1 & \text{for timelike} \\ 0 & \text{for null} \end{cases}.$$

It suffices to study $\theta = \pi/2$ (from the equation for $\ddot{\theta}$ and rotational symmetry). Recall that $u^\mu K_\mu$ is constant along geodesics. This leads to the conserved quantities

$$E = -g_{\mu\nu}u^\mu(\partial_t)^\nu = (1-\mu)\dot{t},$$
$$L = -g_{\mu\nu}u^\mu(\partial_\phi)^\nu = r^2\sin^2\theta\dot{\phi} = r^2\dot{\phi}$$

given the earlier reduction. Thus we have the simpler equation

$$-\frac{E^2}{1-\mu} + \frac{\dot{r}^2}{1-\mu} + \frac{L^2}{r^2} = K = \begin{cases} -1 & \text{for timelike} \\ 0 & \text{for null} \end{cases}.$$

Note that for positive E, the geodesic moves toward the future.

We can make some observations. First, write

$$\dot{r}^2 + \frac{L^2}{r^2}(1-\mu) = K(1-\mu) + E^2$$

and in the timelike case,

$$\dot{r}^2 + \left(1 + \frac{L^2}{r^2}\right)(1-\mu) = E^2.$$

Expanding this gives

$$\dot{r}^2 + \left(1 - \frac{2M}{r}\right) + \frac{L^2}{r^2} - \frac{2ML^2}{r^3} = E^2.$$

The first three terms on the left describe Keplerian orbits in Newton's theory. The fourth term is a higher order correction to Newton's theory. Testing for this correction experimentally was one of the first verifications of the theory.

Next we'll focus on radial ($L = 0$) null geodesics ($K = 0$). Then

$$\dot{r}^2 = E^2.$$

If we consider

$$\dot{r} = -E,$$

corresponding to moving inwards, then we have

$$r = -E\tau + r_0.$$

Note that the affine length to reach $r = 2M$ from τ_0 is finite! But note that

$$\frac{dt}{dr} = \frac{\dot{t}}{\dot{r}} = -\frac{1}{1-\mu},$$

so we cannot reach $r = 2M$ in finite time. This suggests we can make a coordinate transformation to extend the metric.

So we choose new coordinates: $v = t + r^*$ with

$$r^* = r - 2M \log\left(r - 2M\right).$$

Note that

$$\frac{dr^*}{d\tau} = \frac{1}{1-\mu},$$

and $dv = 0$ along ingoing null geodesics. In the new coordinates,

$$\begin{aligned} g &= -\left(1-\mu\right)\left(dv - \frac{dr}{1-\mu}\right)^2 + \frac{dr^2}{1-\mu} + r^2 d\omega_2^2 \\ &= -\left(1-\mu\right) dv^2 + 2dvdr + r^2 d\omega_2^2. \end{aligned}$$

This metric can be defined for all $r > 0$; the extension contains an isometric copy of the restricted space. Note that

$$v = \text{const. surfaces are null hypersurfaces,}$$

$$r = \text{const. surfaces are} \begin{cases} \text{spacelike} & \text{for } r < 2M \\ \text{null} & \text{for } r = 2M \\ \text{timelike} & \text{for } r > 2M \end{cases}.$$

One can similarly extend the metric via outward moving geodesics. What is the maximal extension? Define the set

$$\mathcal{U} = \left\{(T, R) \in \mathbb{R}^2 \,\middle|\, T^2 - R^2 < 1\right\}.$$

Then on $\mathcal{U} \times S^2$ we can put a metric

$$g = \frac{32M^3}{r} e^{-\frac{r}{2M}} \left(-dT^2 + dR^2\right) + r^2 d\omega_2^2$$

where

$$T^2 - R^2 = \left(1 - \frac{r}{2M}\right) e^{\frac{r}{2M}}.$$

Remark. There is another family of black hole solutions (containing the Schwarzschild family) describing rotating black holes (Kerr).

Remark. Stability of the global picture is an open problem.

5. Some Results in Stability Theory

We want to study the stability of the (fully non-linear, tensorial) Einstein equation. But this is very diffucult! So for the rest of the course, we'll study the linear, scalar wave equation

$$\Box_g \psi = 0 \tag{5.1}$$

for g being the Schwarzschild metric. We eventually hope to ask if the black hole picture persists under perturbations. So in the simplified (linear, scalar) picture we seek to show uniform boundedness and decay for ψ on the black hole exterior.[3] If we can understand this, then perhaps we can ask questions about

$$\Box_g \psi = (\partial \psi)^2 .$$

And then, we'd perturb g. At the end of the day, we would like to move to the full non-linear, tensorial equation. This is where current research is heading; but for now, we'll consider only the properties of equation 5.1.

Exercise 5.1. Show

$$\Box_g \psi = g^{ab} \nabla_a \nabla_b \psi = \frac{1}{\sqrt{g}} \partial_a \left(\sqrt{g} g^{ab} \partial_b \psi \right)$$

where $\sqrt{g} = \sqrt{-\det g}$.

5.1. **Estimates in Minkowski Space.** Let's return to Minkowski space, and study the wave equation there:

$$\Box_g \psi = \left(-\partial_t^2 + \partial_x^2 + \partial_y^2 + \partial_z^2 \right) \psi = 0.$$

As usual, we multiply by ∂_t and integrate

$$\int_B \frac{1}{2} \partial_t \Big((\partial_t \psi)^2 + (\partial_x \psi)^2 + \cdots + (\partial_z \psi)^2 - \nabla \left(\nabla_\psi \partial_t \psi \right) \Big) \, dt dx dy dz,$$

leading to the conservation law

$$\int_{\Sigma_T} (\partial_t \psi)^2 + (\nabla \psi)^2 = \int_{\Sigma_0} (\partial_t \psi)^2 + (\nabla \psi)^2$$

with Σ_T the time-evolution of the initial hypersurface Σ_0 through time $t = T$.

There is a more geometric formulation of energy conservation, via Killing fields. Write the energy-momentum tensor as

$$T_{\mu\nu} = \nabla_\mu \psi \nabla_\nu \psi - \frac{1}{2} g_{\mu\nu} \left(g^{\alpha\beta} \nabla_\alpha \psi \nabla_\beta \psi \right),$$

then

$$\nabla^\mu T_{\mu\nu} = 0$$

for ψ satisfying $\Box_g \psi = 0$. So for an arbitrary vector field X we gain the important identity

$$\begin{aligned} \nabla^\mu \left(T_{\mu\nu} X^\nu \right) &= T_{\mu\nu} \nabla^\mu X^\nu \\ &= \frac{1}{2} T_{\mu\nu} \left(\nabla^\mu X^\nu + \nabla^\nu X^\mu \right) \\ &= \frac{1}{2} T_{\mu\nu} \pi^{\mu\nu} \end{aligned}$$

[3]We study the exterior for this is what we can observe from outside the black hole.

where $\pi^{\mu\nu} = (\mathcal{L}_X g)^{\mu\nu}$ is the *deformation tensor*. If we call $\mathcal{J}_\mu = T_{\mu\nu}X^\nu$ the *energy current* and $K^N = T_{\mu\nu}\pi^{\mu\nu}$ the *space-time term*, the identity then reads

$$\nabla^\mu \mathcal{J}_\mu = \frac{1}{2}K^N.$$

Integrating and applying Stoke's theorem yields the geometric conservation law

$$\int_{\partial B} \mathcal{J}_\mu n^\mu_{\partial B} = \frac{1}{2}\int_B K^N.$$

This relation is particularly useful if $\mathcal{J}_\mu n^\mu_{\partial B}$ has one sign.

Exercise 5.2. Check that

$$\mathcal{J}_\mu n^\mu = T_{\mu\nu}X^\nu n^\mu \geq 0$$

if X and n are furture directed causal.

Suppose in particular we have two null hypersurfaces N_u, N_v intersecting each other and the hypersurface Σ in Minkowski space. Then if we set $X = \partial_t$ (Killing), we find

$$\int_\Sigma (\partial_t\psi)^2 + (\nabla\psi)^2 = \int_{N_u} |D\psi|^2 + \int_{N_v} |D\psi|^2.$$

But one can prove each term on the right is non-negative, as N_u, N_v are null. So if two solutions agree on Σ, then the energy of their difference there is zero; definiteness of the right hand terms prove that the solutions must also agree on N_u, N_v! (And on the entire region bounded by the three hypersurfaces, by a similar argument.) This is one way to prove uniqueness of solutions geometrically.

Different vector fields X lead to different conservation laws. Here is a conformal Killing vector field of Minkowski, due to Morowitz:

$$K = \left(t^2 + r^2\right)\partial_t + 2tr\partial_r.$$

It's possible to show $\int T_{\mu\nu}\pi^{\mu\nu}$ vanishes (modulo boundary terms). The induced conservation law is

$$\int_{\Sigma_T} \left(t^2 + r^2\right)|D\psi|^2 \leq c\int_{\Sigma_0} \left(t^2 + r^2 + 1\right)|D\psi|^2.$$

At the initial hypersurface, $t = 0$, so in fact

$$\int_{\Sigma_T} \left(t^2 + r^2\right)|D\psi|^2 \leq c\int_{\Sigma_0} \left(r^2 + 1\right)|D\psi|^2.$$

Now if the right is finite, then the left must be finite for all time t. So $|D\psi|^2$ has to vanish in some way as $t \to \infty$, which can be a very useful estimate.

One other possible type of estimate is the "integrated decay estimate." The idea is to choose a vector field so that

$$T_{\mu\nu}\pi^{\mu\nu} \geq |D\psi|^2$$

and so that the boundary terms satisfy

$$\int_\Sigma T_{\mu\nu}X^\mu n^\nu \leq \int_\Sigma |D\psi|^2.$$

Then integrating in time gives

$$\int_0^T \int_{\Sigma_t} |D\psi|^2\, dt \leq c\int_{\Sigma_0} |D\psi|^2.$$

Here is one last result which allows us to gain control of derivatives.

Proposition 5.3. *(Commutation) If X is a Killing field and $\Box_g \psi = 0$, then $\Box_g (K\psi) = 0$.*

5.2. **Estimates in Schwarzchild Space.** The metric is now

$$g = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\omega_2^2,$$

and once we take $t^* = t + 2M \log(r - 2M)$ we have

$$g = -\left(1 - \frac{2M}{r}\right)(dt^*)^2 + \frac{4M}{r} dt^* dr + \left(1 + \frac{2M}{r}\right) dr^2 + r^2 d\omega_2^2.$$

Some useful notation:

- Denote spacelike slices of constant t^* as Σ_{t^*}.
- On the event horizon, $\mathcal{H}^+(\tau_1, \tau_2) = \mathcal{H}^+ \cap \mathcal{J}^+(\Sigma_{\tau_1}) \cap \mathcal{J}^-(\Sigma_{\tau_2})$.
- Call $\mathcal{R}(\tau_1, \tau_2) = \cup_{\tau_1 < t < \tau_2} \Sigma_t$.

The vector field $T = \partial_{t^*}$ is Killing and timelike, since $g(\partial_{t^*}, \partial_{t^*}) = -(1 - \mu)$. Via this Killing field (applying the important identity) we have the conservation law

$$\int_{\Sigma_{t^*}} (\partial_{t^*}\psi)^2 + (1-\mu)(\partial_r \psi)^2 + \underbrace{||\nabla\psi||^2}_{\substack{\text{angular}\\ \text{derivatives}}} + \int_{\mathcal{H}(t^*,0)} (\partial_{t^*}\psi)^2$$
$$= \int_{\Sigma_0} (\partial_{t^*}\psi)^2 + (1-\mu)(\partial_r \psi)^2 + \underbrace{||\nabla\psi||^2}_{\substack{\text{angular}\\ \text{derivatives}}}.$$

We can get pointwise boundedness of ψ away from $r = 2M$ via commutation with ∂_{t^*} and elliptic estimates (and Sobolev embedding). But how do we control ψ on the event horizon?

Proposition 5.4. *There is a timelike vectorfield N such that*

$$K^N \geq b\mathcal{J}_\mu^N n_\Sigma^\mu \text{ for } r \leq \frac{5}{2}M,$$
$$K^N \leq B\mathcal{J}_\mu^T n_\Sigma^\mu \text{ for } \frac{5}{2}M \leq r \leq 3M,$$
$$K^N = 0 \text{ for } r \geq 3M.$$

Here b is some small constant, B some large constant.

Remark. In the regoin $r \geq 3M$ we can just set $N = T$.

Applying the important identity with N yields

$$\int_{\Sigma_{\tau_2}} \mathcal{J}_\mu^N n_\Sigma^\mu + \int_{\mathcal{H}(\tau_1,\tau_2)} \mathcal{J}_\mu^N n_{\mathcal{H}}^\mu + \int_{\mathcal{R}(\tau_1,\tau_2)\cap\left\{r \leq \frac{5}{2}M\right\}} K^N$$
$$= \int_{\mathcal{R}(\tau_1,\tau_2)\cap\left\{\frac{5}{2}M \leq r \leq 3M\right\}} \left(-K^N\right) + \int_{\Sigma_{\tau_1}} \mathcal{J}_\mu^N n_{\Sigma_{\tau_1}}^\mu.$$

After adding a multiple of $\int_{\mathcal{R}\cap\left\{r\geq\frac{5}{2}M\right\}}\mathcal{J}^T_\mu n^\mu_\Sigma$ to this we arrive at the estimate

$$\begin{aligned}&\int_{\Sigma_{\tau_2}}\mathcal{J}^N_\mu n^\mu_\Sigma+\int_{\mathcal{H}(\tau_1,\tau_2)}\mathcal{J}^N_\mu n^\mu_{\mathcal{H}}+b\int_{\mathcal{R}(\tau_1,\tau_2)}\mathcal{J}^N_\mu n^\mu_\Sigma\\&\leq B\int_{\mathcal{R}(\tau_1,\tau_2)}\mathcal{J}^T_\mu n^\mu_\Sigma+\int_{\Sigma_{\tau_1}}\mathcal{J}^N_\mu n^\mu_{\Sigma_{\tau_1}}.\end{aligned}$$

Now we estimate the first term on the right by

$$\begin{aligned}B\int_{\mathcal{R}(\tau_1,\tau_2)}\mathcal{J}^T_\mu n^\mu_\Sigma&\leq B\int_{\tau_1}^{\tau_2}\int_{\Sigma_t}\mathcal{J}^T_\mu n^\mu_\Sigma\,dt\\&\leq BD\left(\tau_2-\tau_1\right)\end{aligned}$$

where we've taken a supremum (controlled by the initial energy) out of the integral. Calling

$$f\left(\tau\right)=\int_{\Sigma_\tau}\mathcal{J}^N_\mu n^\mu_\Sigma,$$

we have arrived at

$$f\left(\tau_2\right)+\int_{\tau_1}^{\tau_2}f\left(\tau\right)\,d\tau\leq BD\left(\tau_2-\tau_1\right)+f\left(\tau_1\right).$$

Exercise 5.5. Show that this implies

$$f\leq B\left(D+f\left(0\right)\right)$$

and hence

$$\int_{\Sigma_\tau}\mathcal{J}^N_\mu n^\mu_{\Sigma_\tau}\leq B\int_{\Sigma_0}\mathcal{J}^N_\mu n^\mu_\Sigma.$$

Upon commuting with ∂_{t^*} one can conclude pointwise boundedness of ψ. But this is only part of stability, the other part is decay.

True stability involves long-time decay given small initial data. This turns out to be much more difficult. Recall we derived

$$\frac{1}{2}\dot{r}^2+\frac{L^2}{2r^2}\left(1-\mu\right)=\frac{E^2}{2}.$$

We can think of the second term on the left as a potential. Note that $r=3M$ are null geodesics, which orbit the black hole (neither decay into or leave) for all time. So energy concentrates on these orbits for long time; this is an obstacle to proving decay. This is called "trapping." But we can still get the integrated decay estimates

$$\begin{aligned}\int_{\mathcal{R}(\tau_1,\tau_2)}|D\phi|^2\left(r-3M\right)^2&\leq\int_{\Sigma_{\tau_1}}|D\phi|^2,\\\int_{\mathcal{R}(\tau_1,\tau_2)}|D\phi|^2&\leq\int_{\Sigma_{\tau_1}}|D\phi|^2+|D^2\phi|^2.\end{aligned}$$