General Relativity and Differential Geometry

Siddharth Bhat

Monsoon 2019

Contents

Chapter 1

Introduction

I am following the following sources:

- Susskind's General Relativity lectures as part of the Theoretical minimum: (The link is here). Note that the videos do not load. However, one can view the source to access the link to the iframe.

- Susskind's other General Relativity lectures, as part of his modern physics course: (link to playlist here). These seem to be taught at a much gentler pace.

- The book Gravitation by Misner, Thorne and wheeler.

The notes as scribed here are a mix from all of these sources, as well as tangential points I find interesting.

1.1 The equivalence principle

Gravity is in some sense the same thing as acceleration. First, an elementary derivation which formalizes the intuition of the equivalence principle.

We consider an elevator moving upward. Let its distance from the bottom be $L(t)$.

$$z' = z - L(t) \quad t' = t \quad x' = x$$

if $\frac{d^2L(t)}{dt^2} = 0$, then the force is the same in the new coordinate system as that of the old coordinate system

1.2 Galileo's theory of flat space and gravitation

Newton's laws:

$$\vec{F} = m\vec{a}$$
$$\vec{F} = m\frac{d^2x}{dt^2}$$

Galileo's gravitation, under the approximation that the earth is flat: If we pick downwards to be negative direction along the 2 dimension, then his equation can be written as $F_2 = -mg$ where g is a constant.

This is special, because the force is proportional to the mass, which is not the case of things like electromagnetism.

Combining the two equations, we get $m\frac{d^2x}{dt^2} = -mg$, or $\frac{d^2x}{dt^2} = -mg$.

That the acceleration induced by the graviational force is independent of the mass of the object is known as the *equivalence principle*. At this stage, we can say that gravity is equivalent across all objects independent of their mass.

Let's now consider a collection of point masses — A diffuse cloud of particles, and have it fall. Different particles maybe heavy, light, large, small. However, since all of them have the same acceleration, the point cloud looks unchanged as it falls. That is, the object will have no stresses or strains as it falls. We can't tell by looking at our neighbours that there is a force being exterted on us, since all our neighbours are moving along with us! We cannot tell the difference between being in free space versus being in a graviational field.

1.3 Newton's theory of gravity

all bodies exert equal and opposite forces on each other. Given two bodies a and b of masses m and M with distance R, the force on a is $F_a \equiv \frac{GmM}{R^2}\hat{r}$. where $\hat{r}$ is the direction from a to b.

Again, we can prove that the acceleration of an object a does not depend on its own mass.

Now that gravitation depends on distance, we can actually feel something if we are in a gravitational field, since different parts of a given object will have a different force on it, due to the varying distance from (say) the earth.

Gravitational field is defined as the force exerted on a test mass at every point in space.

Gauss' theorem

$\int \boldsymbol{\nabla} \cdot A dx dy dz = \int A_\perp d\sigma$ where σ is the differential unit of surface area of the surface.

Show that the divergence of a field in 3 dimensions will lead to an inverse square law.

1.4 Geometry and curvature

To describe a geometry, all we need is the distance between neighboring points on a blackboard. In general, given a parametrization, we can draw a possibly distorted grid of lines of constant corrdinate. The distance between two points (x, y) and $(x + dx, y + dy)$ will be $ds^2 = g_{11} dx^2 + 2g_{12} d_x d_y + g_{22} d_y^2$.

Chapter 2

Opimisation on Steifel Manifolds

We have the space $O(n) = \{X \in \mathbb{R}^{n \times n} | X^T X = I\}$. To find an element of the tangent space at the point $P \in O(n)$, we parametrize a curve $C(t) : \mathbb{R} \to O(n)$, such that $c(0) = P$. Then, we differentiate the curve and evaluate it at 0. That is, $\frac{dc}{dt}|_{t=0} \in T_P O(n)$.

We consider $C(t) : \mathbb{R} \to O(n)$, such that $C(0) = P$. Since $C(t) \in O(n)$, we can write $C(t)^T C(t) = I$. This in index notation, is $c^{ik}(t)c^{jk}(t) = \delta^{ij}$. Differentiating with respect to t, we get:

$$
\begin{aligned}
&C^T(t)C(t) = I \\
&c^{ik}(t)c^{jk}(t) = \delta^{ij} \\
&\frac{d(c^{ik}(t)c^{jk}(t))}{dt} = \frac{d(\delta^{ij})}{dt} \\
&\dot{c}^{ik}(t)c^{jk}(t) + c^{ik}(t)\dot{c}^{jk}(t) = 0 \qquad \text{(chain rule)} \\
&\dot{C}^T(t)C(t) + C(t)^T\dot{C}(t) = 0
\end{aligned}
$$

Now, we know that $C(0) = P$, and hence $\dot{C}(0) \in T_P O(n)$. By evaluating the above equation at $t = 0$, we obtain the relation:

$$
\begin{aligned}
&\dot{C}^T(0)C(0) + C(0)^T\dot{C}(0) = 0 \\
&\dot{C}^T(0)P + P^T\dot{C}(0) = 0
\end{aligned}
$$

Hence, we conclude that for all $Z \in T_P O(n)$, $Z^T P + P^T Z = 0$. Indeed, we can characterize $T_P O(n)$ this way and prove the reverse inclusion (how?), to show that:

$$T_P O(n) \equiv \{Z \mid Z^T P + P^T Z = 0\}$$

However, this equation for the Z is "ineffective", in that it does not tell us how to *compute* the set of Zs. We can only *check* if a particular $Z_0 \in_? T_P O(n)$.

We will solve the characterization of Z by first solving a slightly easier problem: $T_I O(n)$, where I is the identity matrix.

$$T_I O(n) \equiv \{Z \mid Z^T I + I^T Z = 0\} = \{Z \mid Z^T = -Z\}$$

We now have a complete enumeration of the *tangent space at the identity*: We know that this consists of all skew-symmetric matrices!

We now wish to transport this structure of $T_I O(n)$ to an arbitrary $T_P O(n)$. Here, I will let you in on a secret about the structure of Lie groups, which we will later prove: The vector space at $T_P O(n)$ is obtained by multiplying by P to $T_I O(n)$.

In this case, it tries to inform us that:

$$T_P O(n) = \{PZ \mid Z^T = -Z\}$$

We distrust this assertion at first, of course, so we can plug this equation back in the characterization of $T_P O(n)$ that we had developed and see what pops out:

$T_P O(n) \equiv \{X \mid X^T P + P^T X = 0\}$
Let $X = PZ$ where $Z^T = -Z$, and check that they satisfy the condition $X^T P + P^T X = 0$
$(PZ)^T P + P^T PZ = Z^T P^T P + P^T PZ = Z^T I + I^T Z = -Z + Z = 0$

Hence, they do satisfy the condition, and we can assert that at least:

$$\{PZ \mid Z^T = -Z\} \subseteq \{X \mid X^T P + P^T X = 0\}$$

.

Chapter 3

Symplectic vector spaces

Let V be a m-dimensional real vector space. Let $\Omega : V \times V \to V$ be a bilinear form on V that is skew-symmetric: $\forall a, b \in V, \Omega(a, b) = -\Omega(b, a)$.

Theorem 1 *Let Ω be a bilinear skew-symmetric map on V. Then there is a basis $u_1, u_2, \ldots u_k$, $e_1, e_2, \ldots e_n$, $f_1, f_2 \ldots f_n$ such that:*

- $\Omega(u_i, v) = 0 \quad \forall v \in V$
- $\Omega(e_i, e_j) = \Omega(f_i, f_j)0$
- $\Omega(e_i, f_j) = \delta_{ij}$

Proof 1 *Generalize Gram-Schmidt process. Let $U = \{u \in V | \Omega(u, v) = 0, \ \forall v \in V\}$ U is a subspace of V. Let $u_1, u_2, \ldots u_k$ be a basis of U.*

Establish e_j, f_j by induction. Let $V = U \oplus W$. Take $e_1 \in W$. Then, there must exist an $f_1 \in W$ such that $\Omega(e_1, f_1) \neq 0$. Otherwise, we could expand the size of U by adding e_1 to U. But we assumed that U contains all such vectors. U and W share no non-zero vectors since $V = U \oplus W$. Define $W_1 \equiv \mathrm{span}(e_1, f_1)$. Now build $W_1^\Omega \equiv \{w \in W | \Omega(w, W_1) = 0\}$.

We need the folloing facts:

- $W_1^\Omega \cap W_1 = \{0\}$
- $W = W \oplus W_1^\Omega$

So, recurse on W_1.

Finally, $V = U \oplus W_1 \oplus W_1 \cdots \oplus W_n$.

Remark 1 $\dim(U) = k$ *is invariant for (V, Ω) since U was defined in a coordinate free way. But, remember that $k + 2n = m$, and hence n is an invariant of (V, Ω). n is called as the* rank *of Ω.*

Ω written in the basis of $\{u_i, e_i, f_i\}$ gives $\Omega = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & I \\ 0 & -I & 0 \end{bmatrix}$

3.0.1 Symplectic maps

Let V be an m dimensional real vector space over $\mathbb{R}$. Let $\Omega : V \times V \to \mathbb{R}$ be a skew-symmetric bilinear map. We define $\tilde{\Omega} : V \to V^*$, $\tilde{\Omega}(v)(u) \equiv \Omega(v, u)$. The problem is that this map has a non-trivial kernel $U = \{u_i\}$ in general, so we cannot use it like a metric to identify the two spaces.

Definition 1 *A skew-symmetric bilinear map $\Omega : V \times V \to \mathbb{R}$ is Symplectic iff $\tilde{\Omega}$ is bijective. ie, $U = \{0\}$. (V, Ω) is then called a Symplectic vector space.*

3.1 Properties of a Symplectic map

- $\tilde{\Omega}$ is an identification between V and V^*.
- Since each (e_i, f_i) come in pairs, the dimension of the vector space V is divisible by 2. ie, $m = 2n$.
- We have a basis (e_i, f_i) called the Symplectic basis for V.
- Ω in matrix form with respect to the Symplectic basis is $\Omega = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$

3.1.1 Subspaces of a Symplectic vector space

Definition 2 *A subspace W of V is a Symplectic subspace if $\Omega|_W$ is non-degenrate. For example, take the subspace $\mathrm{span}(e_1, f_1)$.*

Definition 3 *A subspace W of V is an Isotropic subspace if $\Omega|_W = 0$. For example, take the subspace $\mathrm{span}(e_1, e_2)$.*

3.1.2 Morphisms: Symplectomorphism

Definition 4 *A map between (V, Ω) and (V', Ω') is a linear isomorphism $\phi : V \to V'$ such that $\phi^*\Omega' = \Omega$. That is, $\Omega'(\phi(v), \phi(w)) = \Omega(v, w)$. ($\phi^*$ is the pullback). The map ϕ is called a Symplectomorphism and the spaces (V, Ω) and (V', Ω') are said to be Symplectomorphic*

(Question: why do we need to define this in terms of a pullback? Why not pushforward? ie, $\Omega(v, w) = \Omega'(\phi(v), \phi(w))$?

3.1.3 Prototypical example of Symplectic space

Let $V = \mathbb{R}^{2n}$. Let $e_j = (0_1, \ldots, 1_j, \ldots 0_n; 0_{n+1} \ldots 0_{2n})$. Let $f_j = (0_1, \ldots 0_n; 0_{n+1}, \ldots 1_{n+j}, \ldots 0_{2n})$. take $\{e_j, f_j\}$ as a basis for V. $\Omega_0 = \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix}$

Let $\phi : V \to V$ be a Symplectomorphism of V. Then, we can show that $M_\phi^T \Omega_0 M_\phi = \Omega_0$. ($M_\phi$ is the matrix of ϕ associated to the $\{e_j, f_k\}$ basis).

3.1.4 Symplectic Manifolds

We are generalizing our symplectic vector space. We are postulating a space that locally looks like a symplectic vector space.

Let $\omega : T_p M \times T_p M \to \mathbb{R}$ be a 2-form on a manifold. This is bilinear and skew-symmetric (by definition of begin a differential form).

We say that ω is closed iff $d\omega = 0$.

Definition 5 *The two form ω is a symplectic form if it is closed and ω_p is symplectic for all $p \in M$. That is, we need ω_p.*

Definition 6 *A symplectic manifold is a pair (M, ω) where ω is a symplectic form on M.*

3.1.5 Prototypical example of Symplectic manifold

Let $M = \mathbb{R}^{2n}$ with coordinates $X_1, \dots X_n, Y_1, \dots Y_n$. $\omega = \sum_i dx_i \wedge dy_i$. ω is symplectic since it has constant coefficients for each $dx_i \wedge dy_i$.

Symplectic basis for the tangent space $T_p\mathbb{R}^{2n} \equiv \{\partial_{X_i}, \partial_{Y_i}\}$.

3.1.6 2 sphere as a symplectic manifold

Let $M = S^2$ as an embedded manifold in $\mathbb{R}^3$. $S^2 \equiv \{v \in \mathbb{R}^3, |v| = 1\}$. $T_pS^2 \equiv \{w \in R^3 | p \cdot w = 0\}$. This works because p is normal to the plane spanned by T_pS^2. We define $\omega_p(u, v) \equiv p \cdot (u \times v)$. Clearly, this is a 2-form since it is bilinear and anti-symmetric. ω is also closed, since there cannot be any degree 3 forms on a 2D manifold! Hence, $d\omega = 0$.

We need to check that it is non-degenrate. For any point $p \in S^2$, $v \in T_pS^2$, we can take $u = p \times v$. Now, $\omega_p(u, v) = p \cdot (u \times v)$ will be a non-degenrate area of a parallelopiped. The parallelopiped is non-degenrate as v, p are perpendicular by definition. u is perpendicular to both v and p by construction (cross-product). (TODO: draw picture).

3.1.7 Mapping between Symplectic Manifolds

Let (M, Ω) and (M', Ω') are Symplectomorphic if $\phi : M \to M'$ is diffeomorphic, and such that $\phi^*\Omega' = \Omega$.

3.1.8 Symplectic manifolds are locally like $\mathbb{R}^{2n}$

Theorem 2 *(Darboux): Let (M, ω) be a $2n$ dimensional symplectic manifold. Let $p \in M$. Then, there is a coordinate chart called as the Darboux chart U with basis $X_1, \dots X_n, Y_1 \dots Y_n$ such that the two-form $\omega = \sum_i dx_i \wedge dy_i$.*

3.2 The cotangent bundle and Symplectic forms

Let X be an n-dimensional manifold. Let $M \equiv T^*X$ be its cotangent bundle. M is also a manifold.

Define coordinate charts on M, $(T^*U : X_1, X_2, \dots X_n, \xi_1, \xi_2, \dots \xi_n)$, where U is a chart on X, $X_1, \dots X_n$ are coordinate functions for U, and $\xi_1, \dots xi_n$ are coordinates for $T^*X_{x_0}$ where $x_0 \in X$, defined by $\xi_{x_0} = \sum_i \xi_i dx_{i\,x_0}$.

We need to show that the transition functions are smooth.

Let $p \in U \cap U'$. We need to express coordinates in one chart as a smooth function of coordinates in another chart. $(U, X_1, \dots X_n)$, $(U', X'_1, \dots, X'_n)$ are coordinates, and let $\xi \in T^*_p M$. $(p, \xi) \in T^*X \equiv M$.

$$\xi = \sum_i \xi_i dx_i = \sum_{i,k} \xi_i \frac{\partial x_i}{\partial x'_k} dx'_k = \sum_k \xi'_k dx'_k$$

where the $\xi'_k = \xi_i \frac{\partial x_i}{\partial x'_k}$ are smooth since the ξ_i are smooth, and that the transition maps derivaties are smooth since the charts are smooth.

3.2.1 Canonical Symplectic structure of contangent bundle

The physicist intuition for the following is that when we consider a particle moving on a manifold X. To fully specify the state of the particle, we need both position of the particle $x \in X$, and also momentum $p \in T^*X$ (why does it belong to contangent bundle? Intuition: given a velocity, it returns the momentum along it??? We can also supposedly look at how momentum transforms). This data is called "phase space" by physicists. Mathematically, this is the cotangent bundle.

Now, we will see the canonical symplectic structure on the cotangent bundle.

3.2.2 Tautological and Canonical forms

Let $(U, x_1, \dots x_n)$ be a coordinate chart of X and $(T^*U, x_1, \dots x_n, \xi_1, \dots \xi_n)$ be the corresponding chart of $M \equiv T^*X$. We will now define a 2-form on M on the chart T^*U via:

$$w = \sum_j dx_j \wedge d\xi_j$$

Theorem 3 *This definition is coordinate independent*

Proof 2 *Consider a one-form on* $T^*U \subset M$. *Let* $\alpha = \sum_j \xi_j dx_j$ *This is called as the* tautological form. α *is a one-form on* $M \equiv T^*X$, *when* T^*X *is treated as a manifold. This is also a point on* T^*X. *Note that* $d\alpha = -\omega$. *But* α *is intrinsically defined.* $\alpha = \sum_j \xi_j dx_j = \sum_j \xi'_j dx'_j$.

TODO: I don't understand why this gives us coordinate independence.

Coordinate-free definition

Let $M \equiv T^*X$. There is a canonical map $\pi : M \to X$, $\pi(x, \xi) = x$. We are going to pullback T^*X along $T^*M = T^*(T^*X)$ $(d\pi)^* : T^*X \to T^*M$. Let $p \in M; p = (x, \xi)$. $\xi \in T^*_x X$. We define α pointwise. $\forall p \in M, \alpha_p \equiv d\pi^*_p \xi_{\pi(p)}$. Equivalently, let $\nu_p \in T_p M \equiv T_p(T^*X)$. Now, $\alpha_p(\nu_p) \equiv \xi(d\pi_p(\nu_p))$.

TODO: draw example (https://www.youtube.com/watch?v=hAX7ZCMM2kQ, 43:52)

Example 1

Let $X = \mathbb{R}$. Now, $M = \mathbb{R} \times \mathbb{R}$. $(x, y) \in M$ (position-momentum). $\pi : M \to X; \pi((x, y)) = x$.

$$\alpha_{(x,y)} \equiv y dx \qquad \omega = -d\alpha = -(\partial_y y \cdot dy \wedge dx) = dx \wedge dy$$
$$\alpha_{(x,y)}(v_x, v_y) = y dx(v_x \partial_x + v_y \partial_y) = y v_x$$

Example 1

Let $X = S^1$. Now, $M = S^1 \times \mathbb{R}$. $(x, y) \in M$ (position-momentum). $\pi : M \to X; \pi((\theta, v)) = \theta$.

$$\alpha_{(x,y)} \equiv y d\theta$$
$$\omega_{(x,y)} \equiv d\theta \wedge dy$$

3.2.3 Naturality of Tautological form

Let X_1, X_2 be n-dimensional manifolds with T^*X_1 and T^*X_2 as cotangent bundles. Let $f : X_1 \to X_2$ is a diffeomorphism. We will show that the tautological manifolds also match.

We show that there is a diffeomorphism $f_\sharp : T^*X_1 \to T^*X_2$ which lifts f. ie, if $f_\sharp((x, \xi)) = (x', \xi')$, where $x' = f(x)$, $\xi = df^* \xi'$.

Theorem 4 *The lift $f_\sharp$ of $f : X_1 \to X_2$ pulls the tautological form on $M_2 \equiv T^*X_2$ onto the tautological form on $M_1 \equiv T^*X_1$.*

Proof 3 *Pointwise, we wish to show that $(df_\sharp^\star)_{p_1}(\alpha_2)_{p_2} = (\alpha_1)_{p_1}$ where $p_2 = f^\sharp p_1$.*

$p_2 = f^\sharp p_1 \implies p2 = (x_2, \xi_2), x_2 = f(x_1) \wedge df^\star(\xi_1) = \xi_2$.

TODO!

Chapter 4

Vector fields and flows

TODO: integrate with notes on computer

Theorem 5 *For any point $x \in M$, there exists a differentiable function $\sigma : \mathbb{R} \times M \to M$ such that $\sigma(0, x) = x$, $\sigma(t, \sigma(s, x)) = \sigma(t + s, x)$, and the map $t \mapsto \sigma(t, x)$ satisfies nice properties (which ones?)*

Proof 4

Definition 7 *Let $\sigma : \mathbb{R} \times M \to M$ be a flow. Write $\sigma_t(p) \equiv \sigma(t, p)$. The map σ_t is an **isotopy** if each $\sigma_t : M \to M$ is a diffeomorphism and $\sigma_0 \equiv$ identity.*

Conversely, on a compact *manifold M, there is a one-to-one correspondence between isotopies and time-dependent vector fields, given by the equation:*

$$(\partial_x \sigma_t)(x_0) = v_t(\sigma_t(x_0))$$

Note that the situation can get complicated even on compact manifolds. Eg. a vector field on a torus with constant irrational slope — The space foliates with 1D subspaces. (TODO: add picture)

Definition 8 *When $X = v_t$ is independent of t, the isotopy is said to be the **exponential map of the flow of** X. It is denoted by $\sigma^\mu(t, x) \equiv \exp(tX)x^\mu$. $\{\exp(tX) : M \to M \mid t \in \mathbb{R}\}$ is a unique, smooth family of diffeomorphisms, satisfying:*

- $\exp(0X) = \mathrm{id}$
- $\partial_t \exp(tX) = X \circ \exp(tX)$

Let us justify naming this object $\exp$:

$$\begin{aligned}
\sigma^\mu(0 + t, x) &= \text{taylor series around } t = 0 \\
&= x^\mu + t(\partial_t \sigma^\mu)(0, x) + \frac{t^2}{2!}(\partial_t \partial_t \sigma^\mu)(0, x) + \ldots \\
&= e^{t\partial_t} \sigma^\mu(t, x)|_{t=0} = e^{tX} \sigma^\mu(0, x)
\end{aligned}$$

Definition 9 *The flow σ_t satisfies:*

- $\sigma(0, x) = \exp(0X)$
- $\partial_t \sigma(t, x) = X(e^{tX}(x))$
- $\sigma(t, \sigma(s, x)) = \sigma(t, e^{sX}x) = e^{tX}(e^{sX}x) = e^{(t+s)X}x = \sigma(t+s, x)$

Chapter 5

Lie derivative: $\mathfrak{L}_X Y$

5.1 The definition of the lie derivative

The problem with manifolds is that to compare values at $x, y \in M$, it's unclear how to compare objects. We simply cannot, since there is no structure available to do this. So, we construct the lie derivative. Given two vector fields X, Y. Let $\sigma(s, x)$ and $\tau(t, x)$ be the flows generated by X, Y respectively. Hence, $(\partial_s \sigma^\mu)|_{(s,p)} = X(\sigma(s,p))$, $(\partial_t \tau^\mu)|_{(t,p)} = X(\tau(t,p))$.

The derivative of Y along the integral curve σ generated by X is:

- map $Y(\sigma_\epsilon(x)) : T_{\sigma_\epsilon(x)}$ to $T_x M$, by $\sigma_{-\epsilon\star} : T_{\sigma_\epsilon(x)} M \to T_x M$.
- Take the difference at x, between $(\sigma_{-\epsilon\star}(Y(\sigma_\epsilon(x)))$ and $Y(x)$.
- Let $\epsilon \to 0$: $\mathfrak{L}_X Y \equiv \lim_{\epsilon \to 0} \frac{1}{\epsilon} \sigma_{-\epsilon\star}(Y(\sigma_\epsilon(x))) - Y(x)$.

5.2 Coordinate definition of Lie derivative

Let (U, ϕ) be a chart with coordinates X^μ. We define $e_\mu \equiv \partial_{X_\mu}$ We write all our objects in terms of this chart.

- $\mathbf{X} \equiv \mathbf{X}^\mu e_\mu$
- $\mathbf{Y} \equiv \mathbf{Y}^\mu e_\mu$
- $\sigma^i_\epsilon(x) = x^i + \epsilon \mathbf{X}^i$
- $\sigma^i_{-\epsilon}(x) = x^i - \epsilon \mathbf{X}^i$
- $\mathbf{Y}^j(\sigma_\epsilon(x)) = \mathbf{Y}^j(x + \epsilon \mathbf{X}) = Y^j + \epsilon \mathbf{X}^j \partial_k \mathbf{Y}^k$
- $\sigma^i_{-\epsilon\ \star}(\nu) = \nu^j \partial_j \sigma^i_{-\epsilon} = \nu^j \partial_j (x^i + \epsilon \mathbf{X}^i) = \nu^j (\delta^i_j + \epsilon \partial_j \mathbf{X}^i)$
- $\sigma^i_{-\epsilon\ \star}(Y(\sigma_\epsilon(x)) = \mathbf{Y}(\sigma_\epsilon(x))^j (\delta^i_j + \epsilon \partial_j \mathbf{X}^i) = (\mathbf{Y}^j + \epsilon \mathbf{X}^j \partial_k \mathbf{Y}^k)(\delta^i_j + \epsilon \partial_j \mathbf{X}^i) = \mathbf{Y}^j \delta^i_j + \epsilon(\mathbf{X}^j \partial_k \mathbf{Y}^k \delta^i_j + \mathbf{Y}^j \partial_j \mathbf{X}^i) + \epsilon^2(\dots) = \mathbf{Y}^i + \epsilon(\mathbf{X}^i \partial_k \mathbf{Y}^k + \mathbf{Y}^k \partial_k \mathbf{X}^i)$

TODO: check the above derivation, seems incorrect somehow.

Example 1 *Manifold* M, *chart* ϕ, *coordinates* X^1, X^2. $P \equiv -X^2\partial_{X^1} + X^1\partial_{X^2}, Q \equiv (X^1)^2\partial_{X^1} + X^2\partial_{X^2}$. *The lie derivative is computed as:*

$\mathfrak{L}_X Y = \{(-X^2\partial_{X^1} + X^1\partial_{X^2})(X^1)^2 - ((X^1)^2\partial_{X^1} + X^2\partial_{X^2})(-X^2)\}e_1 + (\dots)e_2 \qquad = \{(-X^2(2X^1) + 0) - (0 + X^2(-1))\}e_1$

5.2.1 The lie bracket

Definition 10 *Let* $X^\mu\partial_\mu, Y^\mu\partial_\mu \in \mathfrak{X}(M)$. *We define the lie derivative as* $[X,Y] \equiv [X,Y](f) = X(Y(f)) - Y(X(f))$.

Now this might have second-order derivatives. However, for it to be a vector field, it is only allowed to have first order derivatives. Let's prove that the lie bracket only contains first-order derivatives.

$$\begin{aligned}
[X,Y]f &= \sum_{\mu,\nu} X^\mu\partial_\mu(Y^\nu\partial_\nu f) - Y^\mu\partial_\mu(X^\nu\partial_\nu f) \\
&= \sum_{\mu,\nu} X^\mu(Y^\nu\partial_\mu\partial_\nu f + \partial_\nu f\partial_\mu Y^\nu) - Y^\mu(X^\nu\partial_\mu\partial_\nu f + \partial_\nu f\partial_\mu X^\nu) \\
&= (X^\mu Y^\nu\partial_\mu\partial_\nu f - Y^\mu X^\nu\partial_\mu\partial_\nu f + \partial_\nu f(X^\mu\partial_\mu Y^\nu - Y^\mu\partial_\mu X^\nu) = 0 + \mathfrak{L}_X Y f
\end{aligned}$$

The terms are zero since $X^\mu Y^\nu\partial_\mu\partial_\nu f - Y^\mu X^\nu\partial_\mu\partial_\nu f$ vanishes: $\partial_\mu\partial_\nu f = \partial_\mu\partial_\nu f$ by smoothness. Then the sum $X^\mu Y^\nu - Y^\mu X^\nu$ disappears due to the double summation.

5.2.2 Properties of the lie bracket

Lemma 1 *The lie bracket is bilinear:* $[X, cY + dY'] = c[X,Y] + d[X,Y']$.

Proof 5 *TODO*

Lemma 2 *The lie bracket is skew-symmetric:* $[X,Y] = -[Y,X]$.

Proof 6 *TODO*

Lemma 3 *The lie bracket satisfies the Jacobi identity:* $[[X,Y],Z] + [[Y,Z],X] + [[X,Z],Y] = 0$.

Proof 7 *TODO*

Let us define $(fX)(p) \equiv f(p)X^\mu(p)e_\mu$.

Lemma 4 $\mathfrak{L}_{fX} Y = f[X,Y] - Y[f]X$

Proof 8 *TODO*

Lemma 5 $\mathfrak{L}_Y(fX) = f[X,Y] + X[f]Y$

Proof 9 *TODO*

Lemma 6 $f_\star[X,Y] = [f_\star X, f_\star Y]$

Proof 10 *TODO*

5.2.3 Lie bracket as failure of flows to commute

(TODO: draw picture)

Lemma 7 $\tau(\delta, \sigma(\epsilon, x)) - \sigma(\epsilon, \tau(\delta, x)) = \epsilon[X, Y] + O(\epsilon^2)$.

Proof 11 *TODO*

5.2.4 Lie derivatives for one forms

$\mathfrak{L}_X \omega \equiv \lim_{\epsilon \to 0} \frac{1}{\epsilon} \left[\sigma_{-\epsilon}{}^{\star} \omega|_{\sigma_\epsilon(x)} - \omega(x) \right]$

Chapter 6

Optimisation on manifolds

6.0.1 Sketch of optimisation on manifolds

We now consider manifold optimisation techniques on embedded riemannian manfiolds M, equipped with the metric $g : (p : M) \to T_pM \times T_pM \to \mathbb{R}$. The metric at a point $g(p)$ provides an inner product structure on the point T_pM for a $p \in M$.

where we are optimising a cost function $c : M \to \mathbb{R}$. We presume that we have a diffeomorphism $E : M \to \mathbb{R}^n$ which preserves the metric structure. We will elucidate this notion of preserving the metric structure once we formally define the mapping between tangent spaces. This allows us to treat M as a subspace of $\mathbb{R}^n$.

For any object X defined with respect to the manifold, we define a new object $\overline{X}$, which is the embedded version of X in $\mathbb{R}^n$.

We define $\overline{M} \subset \mathbb{R}^n; \overline{M} \equiv \mathit{image}(E)$. We define $\overline{c} : \overline{M} \subseteq \mathbb{R}^n \to \mathbb{R}; \overline{c} \equiv c \circ E^{-1}$

We then needh two operators, that allows us to project onto the tangent space and the normal space. The tangent space at a point $x_0 \in M$, $\overline{T_{x_0}M} \equiv \mathit{span}(\partial_i E|_{E(x_0)})$. We get an induced mapping of tangent spaces $dE : T_{x_0}M$ and $T_{x_0}\overline{M}$.

we consider the gradient $\overline{\boldsymbol{\nabla}}c : (p : \overline{M}) \to \overline{T_pM}; \overline{\boldsymbol{\nabla}}c \equiv dE\overline{d}c$

The normal space, $\overline{N_{x_0}M}$ is the orthogonal complement of the tangent space, defined as $\overline{N_{x_0}M} \equiv \left\{ v \in \mathbb{R}^n \mid \langle v | \overline{T_{x_0}M} \rangle = 0 \right\}$. It is often very easy to derive the projection onto the normal space, from whose orthogonal complement we derive the projection of the tangent space.

The final piece that we require is a retraction $R : \mathbb{R}^n \to \overline{M} \subseteq \mathbb{R}^n$. This allows us to project elements of the ambient space that are not on the manifold. The retraction must obey the property $R(p \in \overline{M}) = p$. (TODO: is this correct? Do we need $R(\overline{M}) = \overline{M}$ or is this pointwise?) (what are the other conditions on the retraction? smoothness?)

Given all of this machinery, the algorithm is indeed quite simple.

- $x : \overline{M}$ is the current point on the manifold
- $g = \boldsymbol{\nabla}c(x) \in T_x\mathbb{R}^n$ is the gradient with respect to $\mathbb{R}^n$ on the manifold
- $\overline{g} = P_{T_x} g \in T_xM$ is the projection of the gradient onto the tangent space
- $x' : \mathbb{R} \equiv x + \eta\overline{g}$

- $\bar{x}' : \overline{M} \equiv R(x')$