Contents

Chapter 1: Topological Manifolds

Topological Manifolds

The study of differential geometry begins with tens of pages of definitions.

Definition 1.1: Topological Manifold

Let M be a topological space. M is a topological manifold of dimension m if it is Hausdorff, second-countable, and locally homeomorphic to $\mathbb{R}^n$.

Definition 1.2: Local homeomorphism

M locally homeomorphic to $\mathbb{R}^n$ if every point $x \in M$ an open set U, equipped with a homeomorphism which sends points in U into an open subset of $\mathbb{R}^n$.

$$\phi : U \to \phi(U)$$

The tuple (U, ϕ) is called a coordinate chart.

Definition 1.3: More on coordinate charts

- A coordinate chart (U, ϕ) is centered at $p \in M$ if $p \in U$ and $\phi(p) = 0 \in \mathbb{R}^n$.
- We call U the coordinate domain, and
- we call ϕ the coordinate map.
- If the choice of (U, ϕ) is unambiguous, then the local coordinates of p are simply the coordinates of $\phi(p)$ in $\mathbb{R}^n$, and
- we sometimes also denote $\phi(U)$ by $\hat{U}$ if it is unambiguous to do so.
- If $\hat{U}$ is an open ball/cube, then U is called a coordinate ball/cube.

The central theme of point-set topology (or even metric topology) is that of passing a topological argument to the basis or to a neighbourhood. Manifolds in particular have a nice basis.

Proposition 1.1: Basis of precompact coordinate balls

Every topological manifold has a countable basis of precompact coordinate balls.

Proposition 1.2: Additional facts about topological manifolds

If M is a topological manifold,

- M is locally compact. (Lee, Proposition 1.12)
- M is paracompact, and every open cover has a refinement that is another countably locally finite open cover whose elements are chosen from an arbitrary (but fixed) basis of M. (Lee, Theorem 1.15)

- M is locally-path connected.
- M is connected iff it is path-connected.
- M is metrizable. (Munkres Chapter 6)

Smooth Manifolds

We wish to perform calculus on manifolds.

Definition 2.1: Smooth function $F : \mathbb{R}^n \to \mathbb{R}^m$

Let $F : \mathbb{R}^n \to \mathbb{R}^m$, replacing $\mathbb{R}^n$ and $\mathbb{R}^m$ with open subsets if necessary. F is smooth if its (scalar-valued) component functions has continuous partial derivatives of all orders. The set of smooth functions from $\mathbb{R}^n$ to $\mathbb{R}^m$ is sometimes denoted by $C^\infty(\mathbb{R}^n, \mathbb{R}^m)$. If $m = 1$, we sometimes write $C^\infty(\mathbb{R}^n)$, similar to a test function on the Schwartz Space.

Definition 2.2: Transition map from ϕ to ψ

Let (U, ϕ) and (V, ψ) be coordinate charts on M. The composite function (whenever $U \cap V \neq \varnothing$)

$$\psi \circ \phi^{-1} : \phi(U \cap V) \to \psi(U \cap V)$$

is called the transition map. Notice $\psi \circ \phi^{-1}$ is by definition a homeomorphism.

Definition 2.3: Smoothly compatiable

Two coordinate charts on M, (U, ϕ) and (V, ψ) are called smoothly compatible if either their domains are disjoint, or their transition map is a diffeomorphism on $\mathbb{R}^m$.

Definition 2.4: Smooth atlas

An atlas $\mathcal{A}$ of M is a collection of charts $\{(U_\alpha, \phi_\alpha)\}$ whose collection of coordinate domains $\{U_\alpha\}$ for an open cover of M.
It is called a smooth atlas if any two charts in the atlas are pairwise smoothly compatible.

Definition 2.5: Smooth manifold

A smooth atlas $\mathcal{A}$ on M is maximal if it is not contained (properly) in any other smooth atlas as a subset. In other words, if (U', ϕ') is a chart on M that is smoothly compatible with all elements in $\mathcal{A}$, then $(U', \phi') \in \mathcal{A}$ already.

This smooth atlas is often very large, it includes all translations of charts, dilations, and composition with diffeomorphisms in $\mathbb{R}^m$, restrictions onto open subsets, etc. A maximal smooth atlas is sometimes called a complete atlas, or a smooth manifold structure.

A smooth manifold is the tuple $(M, \mathcal{A})$, where $\mathcal{A}$ is some smooth atlas. It can happen if M is originally a topological manifold with a huge number of charts, some of which are smoothly compatible with others, that $\mathcal{A}$ is a strict subset, and both $\mathcal{A}_1$ and $\mathcal{A}_2$ are maximal smooth atlases on M, but $\mathcal{A}_1 \neq \mathcal{A}_2$. We often omit $\mathcal{A}$ and write M if the smooth atlas is understood or not of importance.

Definition 2.6: Smooth coordinate terminologies

Let $(M, \mathcal{A})$ be a smooth manifold.

- Any coordinate chart $(U, \phi) \in \mathcal{A}$ is called a smooth chart, similar to def. 1.3
- We call U the *smooth coordinate domain* or *smooth coordinate neighbourhood* of any $p \in U$, and
- we call ϕ the *smooth coordinate map.*
- The terms *smooth coordinate ball* and *smooth coordinate cube* are used similarly.
- A set $B \subseteq M$ is a *regular coordinate ball* if its image is a smooth coordinate ball centered at the origin; and the closure of this ball in $\mathbb{R}^m$ is a subset of the image of another smooth coordinate ball, centered at the origin.

Definition 2.7: Standard smooth structure on $\mathbb{R}^n$

The maximal smooth atlas containing $(\mathbb{R}^n, \mathrm{id}_{\mathbb{R}^n})$ is called the *standard smooth structure on* $\mathbb{R}^n$.

Manifolds with boundary are not as important as regular manifolds for now, but they are worth mentioning.

Definition 2.8: Closed n-dimensional upper half-plane $\mathbb{H}^n \subseteq \mathbb{R}^n$

We define the following symbols for the upper half plane.

- $\mathbb{H}^n = \left\{x \in \mathbb{R}^n,\, x^n \geq 0\right\}$,
- $\mathrm{Int}\,\mathbb{H}^n = \left\{x \in \mathbb{R}^n,\, x^n > 0\right\}$,
- $\partial\mathbb{H}^n = \left\{x \in \mathbb{R}^n,\, x^n = 0\right\}$

Definition 2.9: Manifolds with boundary

A topological space M is called a manifold with boundary if it is Hausdorff, second-countable, and locally homeomorphic to an open subset of $\mathbb{H}^n$ (endowed with the subspace topology from $\mathbb{R}^n$).

A chart (U, ϕ) is an *interior chart* if its coordinate image is disjoint from the 'boundary' of the upper-half plane. This means $\phi(U) \cap \partial\mathbb{H}^n = \varnothing$. Similarly, (V, ψ) is a *boundary chart* if its range contains a point in $\partial\mathbb{H}^n$; so $\psi(V) \cap \partial\mathbb{H}^n \neq \varnothing$.

Similar to def. 1.3 and def. 2.6, we use the terms *coordinate half-ball*, *coordinate half-cube*, *regular coordinate half-ball*.

Let $p \in M$, it is called an *interior point of* M (not to be confused with the topological interior) if it is in the domain of some interior chart, and p is called a *boundary point of* M if there exists a boundary chart that sends p into $\partial\mathbb{H}^n$. The set of interior points and boundary points of M will be denoted by $\operatorname{Int} M$ and ∂M.

Example 2.1: Sphere as a topological manifold

The n-sphere as a topological manifold. Define

$$S^n = \left\{ x \in \mathbb{R}^{n+1},\ |x| = 1 \right\}$$

We claim that $\{U_i^{\pm}\}_{i=1}^{n+1}$ form an open cover, where

$$U_i^+ = \left\{ x \in S^n, x^i > 0 \right\} \quad U_i^- = \left\{ x \in S^n, x^i < 0 \right\}$$

Each $U_i^{\pm}$ is the inverse image of $\pi_i^{-1}((0, +\infty)) \cap S^n$ or $\pi_i^{-1}((0, -\infty)) \cap S^n$, hence open. For every $x \in S^n$, there exists at least some $1 \leq j \leq n+1$ that makes the j-th coordinate of x, $x^j \neq 0$. So

$$S^n = \bigcup_i U_i^{\pm}$$

Denote the unit ball $\left\{ x \in \mathbb{R}^n,\ |x| < 1 \right\}$ in $\mathbb{R}^n$ by $\mathbb{B}^n$.

Chapter 2: Smooth Maps

Smooth Maps

Definition 1.1: Smooth functions $C^\infty(M, \mathbb{R}^k)$

Let $F : M \to \mathbb{R}^k$ be a vector-valued function on a smooth manifold M. We say F is a smooth function if for every $p \in M$, there exists a smooth chart $p \in (U, \phi)$ such that the *coordinate representation of F at p, with respect to* (U, ϕ) is a smooth function from $\mathbb{R}^m$ to $\mathbb{R}^k$, denoted by $\hat{F}$ (in the sense of Definition 2.1).

$$\hat{F} = F \circ \phi^{-1} : \phi(U) \to \mathbb{R}^k \in C^\infty(\phi(U), \mathbb{R}^k)$$

if $k = 1$, then we denote the space of *test functions* on M by $C^\infty(M) = C^\infty(M, \mathbb{R})$

Definition 1.2: Smooth maps between manifolds $C^\infty(N, M)$

Let $F : N \to M$ be a map between smooth manifolds N and M (note we switched the order). F is a smooth map if at every $p \in M$, there exists

- a chart in the smooth atlas of N (the domain), $p \in (U, \phi)$,
- another chart in the smooth atlas of M (the range), $F(U) \subseteq (V, \psi)$,
- such that, the *coordinate representation of F at p with respect to* (U, ϕ)*, and* (V, ψ) is a smooth function from $\mathbb{R}^n$ to $\mathbb{R}^m$, also denoted by $\hat{F}$.

$$\hat{F} = \psi \circ F \circ \phi^{-1} : \phi(U) \to \psi(V) \in C^\infty(\mathbb{R}^n, \mathbb{R}^m) \tag{1}$$

The following propositions summarizes common operations on smooth maps, a few sources of them.

Proposition 1.1: Smooth maps are continuous

If $F : N \to M$ is a smooth map, then F is continuous with respect to the topologies on N and M.

Proof. Let $p \in N$ be fixed, because F is smooth this induces two smooth charts, one in the domain and another in the range; as in def. 1.2. $F(p)$ is a point in M. From eq. (1), $\hat{F}|\phi(U)$ is a smooth hence continuous function. Since $\phi : U \mapsto \phi(U)$ and $\psi : V \mapsto \psi(V)$ are homeomorphisms,

$$F|U = \underbrace{\psi^{-1}}_{\text{continuous}} \circ \underbrace{\hat{F}|\phi(U)}_{\text{smooth}} \circ \underbrace{\phi}_{\text{continuous}} \quad \text{is continuous on} \quad U$$

Let the point p range through all the points in N, so F is continuous at every p, hence on N. ∎

Proposition 1.2: Characterizations of Smooth Maps in $C^\infty(N, M)$

Let N and M be smooth manifolds, and $F : N \to M$. F is a smooth map iff

- For every $p \in N$, there exists smooth charts $p \in (U, \phi)$ and $F(p) \in (V, \psi)$ such that $U \cap F^{-1}(V)$ is an open set in N, and the composite map (the coordinate representation)

$$\psi \circ F \circ \phi^{-1}|(U \cap F^{-1}(V)) : \phi(U \cap F^{-1}(V)) \to \psi(V) \quad \text{is smooth}$$

- F is continuous and there exist smooth atlases $\{(U_\alpha, \phi_\alpha)\} \subseteq \mathcal{A}_N$, and $\{(V_\beta, \psi_\beta)\} \subseteq \mathcal{A}_M$ such that the coordinate representation

$$\psi_\beta \circ F \circ \phi_\alpha^{-1} : \phi(U_\alpha \cap F^{-1}(V_\beta)) \to \psi(V_\beta) \quad \text{is smooth}$$

  whenever it makes sense.

- the restriction of F onto any arbitrary open set U, $F|U : U \mapsto M$ is smooth (in the sense of open submanifold).

Proof. By Proposition 1.1, and the fact that complete atlases are closed under restrictions onto open sets, it is clear that the original definition implies the two. The first definition also clearly implies the original definition, as we can restrict

$$(U, \phi) \mapsto \left(U \cap F^{-1}(V),\, \phi\Big|_{U \cap F^{-1}(V)}\right)$$

since $U \cap F^{-1}(V)$ is open in the domain manifold.

The second definition implies the original one as well, since the smooth atlases are taken from the maximal atlas, we can pass the argument to any smoothly-compatible chart. Atlases must cover both the domain and the range, and coordinate transitions between smoothly compatible charts are diffeomorphisms. If F is smooth on a subcollection of those charts, meaning

$$\psi_\beta \circ F \circ \phi_\alpha^{-1} \in C^\infty(\phi_\alpha(U_\alpha) \cap F^{-1}(V_\beta), \psi_\beta(V_\beta))$$

it is smooth with respect to every pair of (smooth) charts in the two atlases $\mathcal{A}_N$, $\mathcal{A}_M$, as a composition of smooth maps:

$$\psi \circ \underbrace{\psi^{-1} \circ \psi_\beta}_{\text{smooth}} \circ F \circ \underbrace{\phi_\alpha^{-1} \circ \phi}_{\text{smooth}} \circ \phi^{-1}$$

where we can restrict $\phi_\alpha \mapsto \phi_\alpha|(U_\alpha \cap F^{-1}(V_\beta))$ by continuity of F.
I will prove the third and last equivalence later. ■

Proposition 1.3: Sources of smooth maps

Let N, M, P be smooth manifolds, then

- Every constant map is smooth,
- The identity map $\mathrm{id}_M : M \to M$ is smooth,
- The inclusion map $\iota : W \to M$ is smooth, where W is an open submanifold of M.
- The composition of smooth maps is again a smooth map: if $F \in C^\infty(N, M)$ and $G \in C^\infty(M, P)$, then $(G \circ F) \in C^\infty(N, P)$

Diffeomorphisms

Definition 2.1: Diffeomorphism between Manifolds $\mathcal{D}(N, M)$

Let N and M be smooth manifolds, $F : N \to M$ is a diffeomorphism if it is a smooth bijective map with a smooth inverse. We denote the space of diffeomorphisms from N to M by $\mathcal{D}(N, M)$.

Proposition 2.1: Properties of Manifold Diffeomorphisms

Let N, M and P be smooth manifolds, then

- The composition of diffeomorphisms is again a diffeomorphism, that is, if $F \in \mathcal{D}(N, M)$ and $G \in \mathcal{D}(M, P)$, then $(G \circ F) \in \mathcal{D}(N, P)$.
- The open-manifold restriction of a diffeomorphism onto its image is again a diffeomorphism,
- Every diffeomorphism is a homeomorphism and an open map.

Proof. Trivial. ■

Partitions of Unity

See Folland Chapters 4 and 8. Including Urysohn's Lemma, Tietze's Extension Theorem, the usual construction of C_c^∞ bump functions.

Chapter 3: Tangent Spaces

Algebra of Germs on $C^\infty(N)$

The tangent space is a powerful concept that acts almost like the dual in distribution theory.

Definition 1.1: Algebra of Germs at p: $C_p^\infty(N)$

Let N be a smooth manifold and $p \in N$. We define an equivalence relation on the space of test functions on N, $C^\infty(N)$. If $f, g \in C^\infty(N)$, we write $f \sim g$ if $f = g$ for some open neighbourhood about p. We denote this equivalence class by $C_p^\infty(N)$, and it is clear $C^\infty(N)$ is closed under pointwise multiplication by the product rule, and form an algebra; so $C_p^\infty(N)$ is an algebra too.

Tangent spaces of manifolds

Definition 2.1: Vector space of derivations at p: T_pN

Let $\nu : C_p^\infty(N) \to \mathbb{R}$ be a linear functional on the vector space of germs at p. It is called a derivation at p if it satisfies the product rule, if $f, g \in C_p^\infty(N)$,

$$\nu(fg) = g(p)\nu(f) + f(p)\nu(g)$$

then we say

- ν is a tangent vector at p,
- $\nu \in T_pN$,
- ν is an element of the *tangent space at* p.
- ν is a derivation on N at p.

Proposition 2.1: Properties of derivations at p

Let N be a smooth manifold and $p \in N$.

- If $f \in C_p^\infty$ is constant in some neighbourhood of p, then $\nu(f) = 0$ for every $\nu \in T_pN$,
- If $f(p) = g(p) = 0$, then $\nu(fg) = 0$ for tangent vector ν at p.

Tangent spaces of $\mathbb{R}^n$

Proposition 3.1: Basis of $T_p\mathbb{R}^n$

Let $\mathbb{R}^n$ be equipped with the standard smooth structure as in Definition 2.7. The vector space of derivations at $p \in \mathbb{R}^n$ are spanned by the n partial derivatives at p

$$\left.\frac{\partial}{\partial x^j}\right|_p : f \mapsto \left.\frac{\partial}{\partial x^j} f(x)\right|_p, \quad 1 \le j \le n,\ f \in C^\infty(\mathbb{R}^n)$$

Moreover, the n vectors form a basis, and $\dim T_p\mathbb{R}^n = n$.

Definition 3.1: Standard Basis of $T_p\mathbb{R}^n$

The standard basis for the tangent space at $p \in \mathbb{R}^n$ is the n partial derivatives at p.

$$T_p\mathbb{R}^n = \left\{ \left.\frac{\partial}{\partial x^1}\right|_p, \ldots, \left.\frac{\partial}{\partial x^n}\right|_p \right\} \tag{2}$$

Definition 3.2: Coordinate functions x^j on $\mathbb{R}^n$

Let $x^1, \ldots x^n$ be the standard coordinates on $\mathbb{R}^n$. If $1 \leq j \leq n$, x^j is a function (also represented by the same symbol as the standard coordinate) from $\mathbb{R}^n$ to $\mathbb{R}$, which maps each $p = (p^1, \ldots, p^j, \ldots, p^n)$ to its j-th coordinate $p \mapsto p^j$.

This map is linear, hence smooth, has the matrix representation of $\mathcal{M}\{x^j\} = (\delta_{jk})_{1,k} \in \mathbb{R}^{1\times n}$ where δ_{jk} denotes the discrete mass at j.

Furthermore, the coordinate functions behave like the dual basis for the derivations $\partial/\partial x^j\big|_p$ at p

$$\left.\frac{\partial}{\partial x^j}\right|_p (x^k) = \delta_{jk}$$

If $\nu \in T_p\mathbb{R}^n$, and has the basis representation

$$\nu = \sum_j \nu^j \left.\frac{\partial}{\partial x^j}\right|_p = \nu^j \left.\frac{\partial}{\partial x^j}\right|_p$$

then

$$\nu = \sum_j \nu(x^j) \left.\frac{\partial}{\partial x^j}\right|_p = \nu(x^j) \left.\frac{\partial}{\partial x^j}\right|_p \tag{3}$$

Differential of a smooth map $F \in C^\infty(N, M)$

Definition 4.1: Differential of a smooth map $dF_p : T_pN \to T_{F(p)}M$

Let $F \in C^\infty(N, M)$, $p \in N$ and $\nu \in T_pN$ be a tangent vector at p. The differential of a smooth map is a linear map that sends tangent vectors in T_pN to tangent vectors in $T_{F(p)}M$. If $f \in C^\infty(M)$ is a test function on M, then $f \circ F$ is a test function on N, and

$$dF_p(\nu)(f) = \nu\Big(\underbrace{f \circ F}_{C^\infty(N)}\Big)$$

We state the following without proof, as the proof is tedious. It simply involves unboxing the definition of the differential $dF_p : T_pN \to T_pM$, and projecting onto the coordinates.

Proposition 4.1: Properties of the differential

Let N, M and P be smooth manifolds, if $F \in C^\infty(N, M)$, $G \in C^\infty(M, P)$ then

- dF_p is a linear map between T_p and $T_{F(p)}N$,
- $d(G \circ F)_p = dG_{F(p)} \circ dF_p$
- $d(\mathrm{id}_N)_p = \mathrm{id}_{T_pN}$,
- if $F \in \mathcal{D}(N, M)$, then dF_p is a linear isomorphism between T_pN and $T_{F(p)}M$, and

$$(dF_p)^{-1} = d(F^{-1})_{F(p)}$$

Proposition 4.2: Matrix representation of the differential of $F : N \to M$

Let $F \in C^\infty(N, M)$, and $p \in N$ induces two charts $p \in (U, \phi)$ and $F(p) \in (V\,\psi)$. The matrix representation of the differential at p, $dF_p : T_pN \to T_{F(p)}M$ is nothing but the Jacobian matrix of size $m \times n$ of the coordinate representation at p.

$$\mathcal{M}\{dF_p\} = \begin{bmatrix} \frac{\partial \hat{F}^1}{\partial x^1}\Big|_{\phi(p)} & \frac{\partial \hat{F}^1}{\partial x^2}\Big|_{\phi(p)} & \cdots & \cdots & \frac{\partial \hat{F}^1}{\partial x^n}\Big|_{\phi(p)} \\ \frac{\partial \hat{F}^2}{\partial x^1}\Big|_{\phi(p)} & \frac{\partial \hat{F}^2}{\partial x^2}\Big|_{\phi(p)} & \cdots & \cdots & \frac{\partial \hat{F}^2}{\partial x^n}\Big|_{\phi(p)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \hat{F}^m}{\partial x^1}\Big|_{\phi(p)} & \frac{\partial \hat{F}^m}{\partial x^2}\Big|_{\phi(p)} & \cdots & \cdots & \frac{\partial \hat{F}^m}{\partial x^n}\Big|_{\phi(p)} \end{bmatrix} \tag{4}$$

Alternately, if we write $\hat{p} = \phi(p)$ as the $\mathbb{R}^m$ coordinates at p, then

$$\mathcal{M}\{dF_p\} = \begin{bmatrix} \frac{\partial \hat{F}^1}{\partial x^1}\Big|_{\hat{p}} & \frac{\partial \hat{F}^1}{\partial x^2}\Big|_{\hat{p}} & \cdots & \cdots & \frac{\partial \hat{F}^1}{\partial x^n}\Big|_{\hat{p}} \\ \frac{\partial \hat{F}^2}{\partial x^1}\Big|_{\hat{p}} & \frac{\partial \hat{F}^2}{\partial x^2}\Big|_{\hat{p}} & \cdots & \cdots & \frac{\partial \hat{F}^2}{\partial x^n}\Big|_{\hat{p}} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial \hat{F}^m}{\partial x^1}\Big|_{\hat{p}} & \frac{\partial \hat{F}^m}{\partial x^2}\Big|_{\hat{p}} & \cdots & \cdots & \frac{\partial \hat{F}^m}{\partial x^n}\Big|_{\hat{p}} \end{bmatrix} \tag{5}$$

Differential of a smooth map $F \in C^\infty(\mathbb{R}^n, \mathbb{R}^m)$

An important application of this is the following. We begin with the $\mathbb{R}^m \to \mathbb{R}^n$ case. We will see that if p and $F(p)$ are represented by another pair of coordinate charts (smoothly compatible with the previous pair), then the rank of dF_p does not change. So the rank of the differential is an invariant of the choice of coordinate chart.

Definition 5.1: Matrix representation of the differential of $F : \mathbb{R}^m \to \mathbb{R}^n$

Let $F \in C^\infty(\mathbb{R}^m, \mathbb{R}^n)$, and $p \in \mathbb{R}^m$ induces two charts $p \in (U, \mathrm{id}_{\mathbb{R}^m})$ and $F(p) \in (V\ \mathrm{id}_{\mathbb{R}^n})$, where $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$. The matrix representation of the differential at p, $dF_p : T_p\mathbb{R}^m \to T_{F(p)}\mathbb{R}^n$ is nothing but the Jacobian matrix of F at p.

$$\mathcal{M}\{dF_p\} = DF(p) = \begin{bmatrix} \frac{\partial F^1}{\partial x^1}\Big|_p & \frac{\partial F^1}{\partial x^2}\Big|_p & \cdots & \cdots & \frac{\partial F^1}{\partial x^m}\Big|_p \\ \frac{\partial F^2}{\partial x^1}\Big|_p & \frac{\partial F^2}{\partial x^2}\Big|_p & \cdots & \cdots & \frac{\partial F^2}{\partial x^m}\Big|_p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F^n}{\partial x^1}\Big|_p & \frac{\partial F^n}{\partial x^2}\Big|_p & \cdots & \cdots & \frac{\partial F^n}{\partial x^m}\Big|_p \end{bmatrix} \tag{6}$$

Change of Coordinates Matrix

We will go through the section on the Change of Coordinates, and how different coordinate charts change the representation of a derivation at $p \in M$, where M is some smooth manifold.

Definition 6.1: Standard basis of T_pN

From prop. 2.1, since ϕ^{-1} is a diffeomorphism, $d\Big(\phi^{-1}\big|_{\phi(p)}\Big) : T_{\phi(p)}\mathbb{R}^n \to T_pN$ is a linear isomorphism. Hence T_pN is a n-dimensional vector space, and the standard basis vectors of $T_p\mathbb{R}^n$ are denoted by

$$\left\{ \frac{\partial}{\partial x^1}\bigg|_p, \dots, \frac{\partial}{\partial x^n}\bigg|_p \right\} \tag{7}$$

where each basis vector $\frac{\partial}{\partial x^1}\Big|_p \triangleq d\Big(\phi^{-1}\big|_{\phi(p)}\Big)$ is the push-forward derivation (through ϕ^{-1}) of the j-th standard basis vector in $T_{\phi(p)}N$.

Proposition 6.1: Differential of $\psi \circ \phi^{-1} : M \to\ M$

Let M be a smooth manifold, and fix $p \in M$. If $\nu \in T_pM$ is given with respect to the bases

$$\left\{ \frac{\partial}{\partial x^1}\bigg|_p, \dots, \frac{\partial}{\partial x^m}\bigg|_p \right\} \quad \text{and} \quad \left\{ \frac{\partial}{\partial y^1}\bigg|_p, \dots, \frac{\partial}{\partial y^m}\bigg|_p \right\}$$

Defined by

$$\frac{\partial}{\partial x^j}\bigg|_p \triangleq d\left(\phi^{-1}\Big|_{\phi(p)}\right)\left(\frac{\partial}{\partial x^j}\bigg|_{\phi(p)}\right) \quad \text{and} \quad \frac{\partial}{\partial y^j}\bigg|_p \triangleq d\left(\psi^{-1}\Big|_{\psi(p)}\right)\left(\frac{\partial}{\partial y^j}\bigg|_{\psi(p)}\right)$$

and we write ν in terms of the first basis

$$\nu = \nu^j \left.\frac{\partial}{\partial x^j}\right|_p = \sum_{j=1}^{m} \nu^j \left.\frac{\partial}{\partial x^j}\right|_p$$

and the second basis

$$\nu = \nu^j \left.\frac{\partial y^k}{\partial x^j}\right|_{\phi(p)} \left.\frac{\partial}{\partial y^k}\right|_p = \sum_{k=1}^{m}\sum_{j=1}^{m} \nu^j \left.\frac{\partial y^k}{\partial x^j}\right|_{\phi(p)} \left.\frac{\partial}{\partial y^k}\right|_p$$

If $f \in C^\infty(M)$, then

$$\nu(f) = \nu^j \left.\frac{\partial}{\partial x^j}\right|_p f = \nu^j \left.\frac{\partial y^k}{\partial x^j}\right|_{\phi(p)} \left.\frac{\partial}{\partial y^k}\right|_p f$$

Proof. Recall $\left.\frac{\partial}{\partial x^j}\right|_p f \triangleq \left.\frac{\partial}{\partial x^j}\right|_{\phi(p)} f \circ \phi^{-1}$, similarly for $\left.\frac{\partial}{\partial y^j}\right|_p f$. Deriving f and p and by vector space operations on T_pM, the first basis expansion gives

$$\nu^j \left.\frac{\partial}{\partial x^j}\right|_p f = \nu^j \left.\frac{\partial}{\partial x^j}\right|_{\phi(p)} f \circ \phi^{-1} \tag{8}$$

and the second expression reads

$$\nu^j \left.\frac{\partial y^k}{\partial x^j}\right|_{\phi(p)} \left.\frac{\partial}{\partial y^k}\right|_p f = \nu^j \left.\frac{\partial y^k}{\partial x^j}\right|_{\phi(p)} \left.\frac{\partial}{\partial y^k}\right|_{\psi(p)} f \circ \psi^{-1} \tag{9}$$

Since $f \circ \phi^{-1} \in C^\infty(\mathbb{R}^m, \mathbb{R})$, we see the expressions are indeed equal. By the chain rule, if

$$\psi \circ \phi^{-1}(x^1, \dots x^m) = (y^1, \dots y^m)$$

then

$$D(\psi \circ \phi^{-1})(\phi(p)) = \begin{bmatrix} \left.\frac{\partial y^1}{\partial x^1}\right|_{\phi(p)} & \left.\frac{\partial y^1}{\partial x^2}\right|_{\phi(p)} & \cdots & \cdots & \left.\frac{\partial y^1}{\partial x^m}\right|_{\phi(p)} \\ \left.\frac{\partial y^2}{\partial x^1}\right|_{\phi(p)} & \left.\frac{\partial y^2}{\partial x^2}\right|_{\phi(p)} & \cdots & \cdots & \left.\frac{\partial y^2}{\partial x^m}\right|_{\phi(p)} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \left.\frac{\partial y^m}{\partial x^1}\right|_{\phi(p)} & \left.\frac{\partial y^m}{\partial x^2}\right|_{\phi(p)} & \cdots & \cdots & \left.\frac{\partial y^m}{\partial x^m}\right|_{\phi(p)} \end{bmatrix}$$

It follows from Proposition 3.6d) that the matrix $D(\psi \circ \phi^{-1})\big|_{\phi(p)}$ is invertible, as $\psi \circ \phi^{-1}$ is a diffeomorphism. ∎

Proposition 6.2: Rank of a dF_p is invariant under coordinate change

Let F be a smooth map between M and N, at every $p \in M$, $\operatorname{rank} dF_p$ is an invariant over (smoothly compatible) pairs of charts in M and N.

Proof. Let $p \in (U_1, \phi_1) \cap (U_2, \phi_2)$, and $F(p) \in (V_1, \psi_1) \cap (V_2, \psi_2)$. Where all charts are smoothly compatible if it makes sense to talk about it. Both $\phi_2 \circ \phi_1^{-1}$ and $\psi_2 \circ \psi_1^{-1}$ are diffeomorphisms, and the change of basis matrices $D(\phi_2 \circ \phi_1^{-1})\Big|_{\phi_1(p)}$ and $D(\psi_2 \circ \psi_1^{-1})\Big|_{\psi_1(F(p))}$ are invertible by Proposition 3.6d) again, so the ranks dF_p with respect to any of the two charts are equal.

$$\underbrace{D(\psi_2 \circ \psi_1^{-1})\Big|_{\psi_1(F(p))}}_{\text{invertible}} \Big(\mathcal{M}\{dF_p\}\Big) \underbrace{D(\phi_2 \circ \phi_1^{-1})\Big|_{\phi_1(p)}}_{\text{invertible}}$$

■

Chapter 4: Submersions, Immersions and Embeddings

Matrices

The following is of utmost importance. It states that that rank of a matrix, square or otherwise, is an 'open condition'.

Example 1.1: Lee Example 1.28 (Matrices of Full Rank)

Let $A \in \mathcal{M}(m \times n, \mathbb{R})$ be the set of $m \times n$ matrices with real entries. A has rank m iff there exists some $m \times m$ sub-matrix of A, denoted by S st S is invertible. We wish to show the set of rank-m matrices is invertible. Indeed, let

$$F : \mathcal{M}(m \times n, \mathbb{R}) \to \mathbb{R},\ \Delta_{m\times m}(A) = \sum_{\substack{S \text{ is a } m\times m \\ \text{sub-matrix of } A}} |\det\{S\}|$$

Since $S \mapsto \det\{S\}$ is continuous in the entries of S, hence continuous in the entries of A, $\Delta_{m\times m}$ is continuous.
So the set $\left\{A \in \mathcal{M}(m \times n,\ \mathbb{R}),\ \operatorname{rank} A = m\right\} = F^{-1}(\mathbb{R} \setminus \{0\})$ is open.

Estimates in vector calculus

Before proving the inverse function theorem, we will need several Lemmas

Proposition 2.1: Rudin Theorem 9.7

If A and B are in $L(\mathbf{X}, \mathbf{Y})$, then

$$\|BA\| \leq \|B\|\|A\|$$

Proof. Let $\|x\| = 1$, and

$$\|B(Ax)\| \leq \|B\|\|Ax\| \leq \|B\|\|A\|\|x\|$$

this holds for every $\|x\| = 1$, hence

$$\|BA\| \leq \|B\|\|A\|$$

■

Proposition 2.2: Rudin Theorem 9.19

Let f map a convex open set $U \subseteq \mathbb{R}^n$ into $\mathbb{R}^m$, if f is differentiable (pointwise) in U, and there exists some M st its derivative its bounded (in the operator norm)

$$\|Df(x)\| \leq M \quad x \in U$$

then, for every pair of elements x_1, x_2 in U,

$$\|f(x_1) - f(x_2)\| \leq M\|x_1 - x_2\|$$

Proof. This proof 'passes the argument' to the scalar-valued version, in short: if x_1 and x_2 are in U. Define

$$c(t) = (1 - t)x_1 + tx_2$$

as the convex combination of x_1 and x_2. The takeaway intuition here is that it suffices to check on the line joining the two points', to obtain an estimate for $\|f(x_1) - f(x_2)\|$. Indeed, define

$$g(t) = f(c(t)) \text{ is a curve } g : \mathbb{R} \to \mathbb{R}^m$$

Using Theorem 5.19, of which we will state below

Proposition 2.3: Rudin Theorem 5.19

Let $g : [0,1] \to \mathbb{R}^m$, and g be differentiable on $(0,1)$, then there exists some $x \in (0,1)$ with

$$|f(b) - f(a)| \leq (b-a)|f'(x)|$$

Proof. Read from Rudin Theorem 5.19. ■

Since $Dg(t) = Df(c(t)) \circ Dc(t)$ by the Chain Rule, and $Dc(t) = b - a$ by inspection,

$$\|Dg(t)\| = \|Df(c(t)) \circ Dc(t)\| \leq \|Df\|\|Dc\| = \|Df\|(b-a)$$

This holds for every $t \in [0,1]$. Applying Theorem 5.19 gives

$$\underbrace{\|g(1) - g(0)\|}_{\text{curve endpoints}} \leq M\|b - a\|$$

Replacing $\|g(1) - g(0)\| = \|f(x_1) - f(x_2)\|$ and $\|Df\| \leq M$ we get

$$\|f(x_1) - f(x_2)\| \leq M\|x_1 - x_2\|$$

■

Inverse Function Theorem (Rudin)

Proposition 3.1: Rudin Theorem 9.24

Suppose $f \in C^1(\mathbb{R}^n, \mathbb{R}^n)$, and $Df(a)$ is invertible for some $a \in \mathbb{R}^n$, and define $b = f(a)$. Then,

(a) there exist open sets U and V in $\mathbb{R}^n$ such that $a \in U$, $b \in V$, and f is one-to-one on U, and $f(U) = V$.

(b) if g is the inverse of f (which exists, by Part a), defined in V by $g(f(x)) = x$ for every $x \in U$ then $g \in C^1(\mathbb{R}^n, \mathbb{R}^n)$

Proof of Part A. We define $Df(a) = A \in \mathbb{R}^{n \times n}$, so A is invertible, and $\left\|A^{-1}\right\| \neq 0$, where $\|\cdot\|$ denotes the operator norm. Recall all norms on finite-dimensional vector spaces are equivalent, this will be useful later.
Choose $\lambda > 0$ st

$$\lambda = \left\|A^{-1}\right\|^{-1} 2^{-1} \tag{10}$$

By continuity of $Df(x)$ at the point a, let $\lambda > 0$, this induces a $B(\delta, a)$ with $x \in B(\delta, a)$ means

$$\underbrace{\|Df(x) - Df(a)\|}_{\text{operator norm}} < \lambda \tag{11}$$

as $Df : \mathbb{R}^n \to L(\mathbb{R}^n, \mathbb{R}^n)$ takes a point in $\mathbb{R}^n$ and returns a linear map., with $L(\mathbb{R}^n, \mathbb{R}^n)$ endowed with the usual vector space structure. Fix $y \in \mathbb{R}^n$, and define

$$\phi(x) = \underbrace{x + A^{-1}(y}_{\text{offset}} - f(x))$$

this is now a function solely in x, and $\phi(x) = x \iff f(x) = y$ is clear, but such a fixed point is not necessarily unique. We claim that it is unique in $B(\delta, a)$. We will use the contractive mapping principle.

Differentiating $\phi(x)$ reads

$$D\phi(x) = \underbrace{I}_{I=A^{-1}A} - A^{-1}Df(x) = A^{-1}(A - Df(x))$$

Proposition 2.1 tells us the norm of a product is bounded above by the product of the norms. Using eqs. (10) and (11), if $x \in U$ we have

$$\|D\phi(x)\| = \left\|A^{-1}(A - Df(x))\right\| \leq \left\|A^{-1}\right\| \|A - Df(x)\| \leq 2^{-1}$$

The total derivative of ϕ is uniformly bounded in U, applying Proposition 2.2 tells us that ϕ is a contractive mapping

$$\|D\phi(x)\| \leq 2^{-1} \implies \|\phi(x_1) - \phi(x_2)\| \leq 2^{-1}\|x_1 - x_2\|$$

for x_1, x_2 in U.

To show $f|U$ is indeed a bijection, fix $y \in f(U)$ so $y = f(x)$ for some $x \in U$, and there can only be one fixed point stemming from $\phi|U$, with $\phi(z) = z + A^{-1}(y - f(z))$ being the 'fixed point detector'. Write $(f|U)^{-1}(y) = \lim\{(\phi|U)(x_n)\}_n$ and every point in $f(U)$ has a unique inverse.

For the last part of the proof, we wish to show $V = f(U)$ is open. Let $y_0 \in V$ and we can 'hone into' the inverse of y_0 using the same construction as earlier. So $f(x_0) = y_0$ for some unique $x_0 \in U$.

If x_0 is in U, it induces an open ball (see fig. 1) st

$$x_0 \in B(r, x_0) \subseteq \overline{B(r, x_0)} \subseteq U, \quad r > 0$$

We claim the open ball $B(\lambda r, y_0) \subseteq V$. Indeed, suppose $y \in \mathbb{R}^n$ with

$$d(y, y_0) < \lambda r$$

If ϕ is the 'fixed-point detector' with respect to y (the point we are trying to prove that is in $f(U)$), in fact: we will prove $y \in f(\overline{B(r, x_0)}) \subseteq f(U)$.

$$\underbrace{\phi(x_0) - x_0}_{\text{removing the offset from } \phi(x_0)} = A^{-1}(y - f(x_0)) = A^{-1}(y - y_0)$$

using the operator norm on $A^{-1}(y - y_0)$ reads

$$\|\phi(x_0) - x_0\| = \left\|A^{-1}(y - y_0)\right\| \leq \left\|A^{-1}\right\| \|y - y_0\| \leq \left\|A^{-1}\right\| \lambda r = r2^{-1}$$

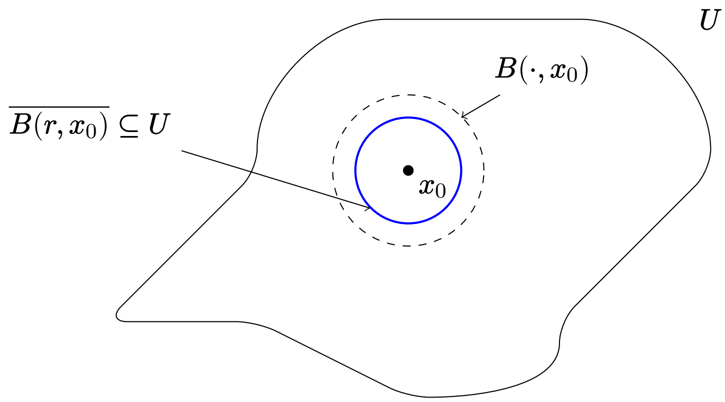

Figure 1: Every point x_0 in an open set U admits an open ball that hides in U

We will drag y into the image of the closed ball as follows: suppose x is another point that lies in the closed ball, ϕ is contractive on $\overline{B} \subseteq U$ regardless of the point y that induces ϕ. But $\overline{B}$ is closed, hence it is complete. So the Cauchy sequence (from the contractive mapping theorem) produces exactly one point in $\overline{B}$. It remains to show that if we start our sequence at some point $x \in \overline{B}$, then $\phi(x) \in \overline{B}$ as well, and a simple induction will produce our contractive sequence.

To this, fix $x \in \overline{B}$, and

$$\begin{aligned} |\phi(x) - x_0| &\leq |\phi(x) - \phi(x_0)| + |\phi(x_0) - x_0| \\ &\leq \overbrace{2^{-1}|x - x_0|}^{\text{contraction on } \overline{B} \subseteq U} + \overbrace{r2^{-1}}^{\text{earlier}} \\ &= r \end{aligned}$$

therefore ϕ contracts to a fixed point $x^* \in \overline{B}$, and $f(x^*) = y$. So $y \in f(\overline{B}) \subseteq f(U)$ as desired. ■

Proof of Part B. The proof is quite long, and we will only focus on the important bits. Rudin uses the technique of approximating smooth functions using first-order terms. He writes

$$\begin{cases} f(x) & = y \\ f(x+h) & = y + k \end{cases} \implies k = f(x+h) - f(x)$$

Furthermore, if $x \in U$, then the derivative $Df(x)$ is invertible, this is from Theorem 9.8, obtains an estimate on the open ball in $GL(n, \mathbb{R})$. Roughly speaking, this open ball 'drags' other matrices into $GL(n, \mathbb{R})$. If A is invertible, and B is a conformable matrix with A, then

$$\underbrace{\|B - A\|}_{\substack{\text{distance} \\ \text{between} \\ A,B}} \left\|A^{-1}\right\| < 1 \implies B \in GL(n, \mathbb{R})$$

If $x \in B(\delta, a)$, then Equation (11) reads

$$\|Df(x) - A\| < \lambda \implies \|Df(x) - A\| \left\|A^{-1}\right\| < 2^{-1} < 1$$

so $Df(x)$ is invertible with inverse T.

And we estimate the deviation $|k|^{-1} \leq \lambda |h|^{-1}$ by using the contraction inequality with y as the basepoint for ϕ. Skipping a few lines ahead (to the confusing part), we see that

$$|h| \leq |h - A^{-1}k| + |A^{-1}k| \leq 2^{-1}|h| + |A^{-1}k|$$

subtracting over, and multiplying across gives a upper bound on $|k|^{-1}$

$$2^{-1}|h| \leq |A^{-1}k| \implies 2^{-1}|h| \leq \left\|A^{-1}\right\| |k| \implies |k|^{-1} \leq \underbrace{\frac{2}{\|A^{-1}\|}}_{\lambda} |h|^{-1}$$

Notice $2\lambda\|A^{-1}\| = 1$, so $2/\|A^{-1}\| = \lambda$. Finally, we 'factor out' $-T$ on the line just before the difference quotient.

$$\begin{aligned}\overbrace{g(y+k) - g(y) - Tk}^{\substack{\text{numerator in}\\ \text{difference quotient}}} &= h - Tk \\ &= -T\Big(\underbrace{f(x+h) - f(x)}_{=k} - \underbrace{Df(x)h}_{=T^{-1}h}\Big)\end{aligned}$$

We see that $T = Dg(y)$, indeed:

$$\begin{aligned}\frac{|g(y+k) - g(y) - Tk|}{|k|} &\leq \frac{\|T\|}{\lambda} \frac{|f(x+h) - f(x) - Df(x)h|}{|h|} \\ &\lesssim \frac{|f(x+h) - f(x) - Df(x)h|}{|h|} \\ = \underbrace{o(h) = o(k)}_{|h| \lesssim |k|} \to 0\end{aligned}$$

Finally, $Df|U : U \to GL(n, \mathbb{R})$ is a continuous mapping. By Theorem 9.8, $(Df|U)^{-1} : U \to GL(n, \mathbb{R})$ is continuous as well. Therefore $g \in C^1(U, U)$, and $f|U$ is a C^1-diffeomorphism. ∎

Remark 3.1

The inverse function theorem is extremely powerful. If a f is a C^1 map from and into $\mathbb{R}^n$, and the total differential of f is full rank (hence invertible, as it is square) at some point $a \in \mathbb{R}^n$, the theorem states three things:

- For points x within a small enough neighbourhood a, the total differential $Df(x)$ is invertible,
- On this same neighbourhood (denoted by U), $f(U)$ is a bijection,
- the inverse of f is a C^1 map. This makes $f|U$ a C^1-diffeomorphism

Inverse Function Theorem on Manifolds

Let F be a smooth map between two smooth manifolds M and N, with dimensions m and n respectively.

Definition 4.1: Rank of a map

The rank of F at $p \in M$ is the rank of the linear map:

$$dF_p : T_pM \to T_{F(p)}N$$

Definition 4.2: Constant rank maps

A smooth map $F \in C^\infty(M, N)$ has constant rank if its differential $dF_p : T_pM \to T_{F(p)}N$ has the same rank at every point $p \in M$.

There are three types of constant rank maps that are of interest.

Definition 4.3: Smooth submersion

F is a smooth submersion if dF_p is a surjection onto $T_{F(p)}N$ at p-everywhere. That is, $\operatorname{rank} dF_p = \dim T_{F(p)}N = \dim N$

Definition 4.4: Smooth immersion

F is a smooth immersion if dF_p is an injection onto $T_{F(p)}N$ at p-everywhere. That is, $\operatorname{rank} dF_p = \dim T_pM = \dim M$

Definition 4.5: Smooth embedding

F is a smooth embedding if it is a smooth immersion, and it is a topological homeomorphism onto its range $F(M) \subseteq N$.

Definition 4.6: Topological embedding

$F \in C(\mathbf{X}, \mathbf{Y})$, where $\mathbf{X}$ and $\mathbf{Y}$ are topological spaces is a *topological embedding* if F is a homeomorphism onto its range.

Inclusion maps are topological embeddings, as the restriction onto the range of the map is the identity.

If $\mathbf{X}$ is compact and F is injective, then F is a topological embedding. If F is injective and an open map, then F is a topological embedding as well.

Definition 4.7: Topological submersion

Definition 4.8: Local diffeomorphism

F is a local diffeomorphism if every $p \in M$ in its domain induces a neighbourhood $U \subseteq M$, $F(U)$ is open in N with $F|U : U \to F(U)$ is a diffeomorphism (in the sense of two open sub-manifolds).

Proposition 4.1: Rank as an open condition

Suppose $F : M \to N$ is a smooth map, and $p \in M$. If dF_p is a surjection (resp. injection), pointwise at p, there exists a neighbourhood U of p where $F|U$ is a smooth submersion (resp. immersion)

Proof. Trivial. See Example 1.1. ■

Proposition 4.1 roughly states that, if the differential of F at some point p is injective or surjective, then there exists a neighbourhood U about p such that $dF|U(p)$ is an injection or surjection. The continuity of the map $dF|U(p) \mapsto \Delta_{m\times m}(dF|U(p))$, induces a neighbourhood in the vector space of matrices about the differential $dF|U(p)$. This vector space is endowed with any of the equivalent norms on $\mathcal{M}(m \times n, \mathbb{R})$, which is equivalent to the entrywise 2-norm. Since all partials of the form $\left.\frac{\partial \hat{F}^k}{\partial x^j}\right|_{\hat{p}}$ are continuous, we take the intersection over all $n \times m$ partials such that $dF|U(p)$ is an injection or surjection. Finally, send this neighbourhood about $\hat{p}$ through to p by using the continuity of ϕ.

Proposition 4.2: Inverse Function Theorem on Manifolds

Let M and N be smooth manifolds, and $F : M \to N$ be a smooth map. Suppose the differential of F is invertible at some point $p \in M$, then there exists connected neighbourhoods U_0 of p, and V_0 of $F(p)$ such that $F|U_0 : U_0 \to V_0$ is a diffeomorphism.

Proof. Trivial. See the regular inverse function theorem Proposition 3.1 on Euclidean space, and pass the argument back to the manifolds using coordinate charts. ■

Proposition 4.3: Rank Theorem for Manifolds

Let $F : M \to N$ be a smooth map with constant rank r, then at every $p \in M$, there exists smooth charts $p \in (U, \phi)$ and $F(U) \subseteq (V, \psi)$, where the coordinate representation of F takes the form

$$\hat{F}(x) = \begin{bmatrix} \mathrm{id}_{r\times r} & 0_{r\times m-r} \\ 0_{n-r\times r} & 0_{n-r\times m-r} \end{bmatrix} x, \qquad \text{or equivalently} \tag{12}$$

$$\hat{F}(x^1, \ldots, x^r, x^{r+1}, \ldots, x^m) = (x^1, \ldots, x^r) \tag{13}$$

Proof. Tedious. However, some techinques are worth remembering:

- Passing the argument to the Euclidean case as usual,
- We are free to shrink the sizes of open cubes and balls, and exploit local compactness,
- Suppose we are given a matrix of size $m \times n$, which has rank r, then we can attach a sub-matrix to make it square and invertible, then rehearse the usual arguments with the Inverse Function Theorems Propositions 3.1 and 4.2 to obtain a neighbourhood small enough that preserves the rank of the square matrix. Then pass the argument back to the smaller sub-matrix.

The last bullet point is worth elaborating, suppose we are given a rectangular matrix, where A is square and invertible. Take $z = (x, y)^T$ with dimensions that make the formulas below make sense.

$$Mz = \begin{bmatrix} A & B \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \implies \begin{bmatrix} M \\ I \end{bmatrix} (z, y)^T = \begin{bmatrix} A & B \\ 0 & I \end{bmatrix} \begin{bmatrix} z \\ y \end{bmatrix} \quad \text{is square and invertible}$$

by Proposition 4.1, we see that there exists a neighbourhood about the square matrix $\begin{bmatrix} M \\ I \end{bmatrix}$ such that it remains invertible, hence a neighbourhood about M that makes A invertible (as a sub-matrix), so the rank of M is preserved. ■

Corollary 4.1: Rank Theorem for Manifolds - Special Cases

Let $F : M \to N$ be a smooth map with constant rank. If F is a smooth immersion, then Equation (12) takes the form:

$$\hat{F}(x) = \begin{bmatrix} \mathrm{id}_{m\times m} \\ 0_{n-m\times m} \end{bmatrix} x, \qquad \text{or equivalently} \tag{14}$$

$$\hat{F}(x^1, \ldots, x^m) = (x^1, \ldots, x^m, 0 \ldots, 0) \tag{15}$$

If F is a smooth submersion,

$$\hat{F}(x) = \begin{bmatrix} \mathrm{id}_{n\times n} & 0_{n\times m-n} \end{bmatrix} x, \qquad \text{or equivalently} \tag{16}$$

$$\hat{F}(x^1, \ldots, x^n, x^{n+1}, \ldots, x^m) = (x^1, \ldots, x^n) \tag{17}$$

More on immersions and embeddings

Proposition 5.1: Characterization of smooth immersions

F is a smooth immersion iff every point $p \in M$ has a neighbourhood $U \subseteq M$ where $F|U : U \to N$ is a smooth embedding.

Proof. We will prove it for when M and N are smooth manifolds, see Lee for the full proof with boundary. It involves extending the argument by composing F with an inclusion map. From Lemma 3.11 (Lee), if $a \in \partial\mathbb{H}^n$, then the differential of the inclusion map $\iota : \mathbb{H}^n \to \mathbb{R}^n$ is a linear isomorphism between tangent spaces.

$$d\iota_a : T_a\mathbb{H}^n \to T_a\mathbb{R}^n, \quad \underbrace{T_a\mathbb{H}^n \cong T_a\mathbb{R}^n}_{\text{isomorphic}}$$

If for every $p \in M$, there exists a neighbourhood U of p with $F|U : U \to N$ a smooth embedding, then $dF|U_p$ has rank m, so dF_p has rank m, and the differential is injective pointwise everywhere. Conversely, if dF_p is a smooth immersion, the Rank Theorem (Proposition 4.3) tells us there exists connected neighbourhoods of p and $F(p)$, where F has coordinate representation in Equation (14) with respect to an appropriate choice of coordinate charts centered at p, so $\hat{F}(\hat{p}) = 0 \in \mathbb{R}^n$. Let $\hat{p} \in \hat{U}$ and $\hat{F}(\hat{p}) \in \hat{V}$, the proof then devolves into a linear-map problem. $\hat{F}$ given by the expression in Equation (14) is clearly injective. Therefore it is bijective onto its range, its inverse is nothing but the map that removes the extra zeroes at the end. Therefore $F|U$ is a smooth embedding. ∎

Definition 5.1: Section of $\pi : M \to N$

If $\pi : M \to N$ is a continuous map, a *section of* π is a continuous right inverse for π, i.e $\sigma : N \to M$, $\sigma \in C(N, M)$, $\pi \circ \sigma = \mathrm{id}_N$.

A *local section for* π is a continuous function σ from an open set $U \subseteq V$ into M with $\pi \circ \sigma = \mathrm{id}_U$.

Proposition 5.2: Characterization of smooth submersion

Let $\pi : M \to N$ be smooth, then π is a smooth submersion iff every point of M is in the image of a smooth local section of π.

Proof. Suppose π is a smooth submersion, and fix $p \in M$, by the Rank Theorem Proposition 4.3, and Equation (16), π has the coordinate representation

$$\hat{\pi}(x^1, \dots, x^n, x^{n+1}, x^m) = (x^1, \dots, x^n)$$

between two open sets $U \subseteq M$ and $V \subseteq N$, (it really does not matter). Now, define

$$\sigma : V \to M, (x^1, \dots, x^n) \mapsto \underbrace{(x^1, \dots, x^n, 0, \dots, 0)}_{\mathbb{R}^m} \in U$$

the charts by assumption are centered, and $\pi \circ \sigma$ is clearly smooth (check coordinatewise), so σ reaches p. Conversely, recall if the composition of maps $(g \circ f)$ is a surjection, then g is a surjection. Now, fix $p \in M$, this induces an open set V containing $\pi(p)$, and a smooth local section $\sigma_V : V \to M$. By Proposition 4.1, *the differential of a composition is equal to the composition of the differentials*

$$\mathrm{id}_{T_qN} = d(\mathrm{id}_U) = d(\pi)_{\sigma(q)} \circ d(\sigma)_q$$

so $d(\pi)_{\sigma q} = d(\pi)_p$ is a surjection and the proof is complete. ∎

Regular values and level sets

If $F : \mathbf{X} \to \mathbf{Y}$, and $c \in \mathbf{Y}$, we call the $F^{-1}(\{y\})$ a *level set at* c, and c the *level value*. We often write $F^{-1}(y)$ in place of $F^{-1}(\{y\})$. If $\mathbf{Y} = \mathbb{R}^k$, then $F^{-1}(0)$ is the *zero set of* F.

Definition 6.1: Critical point of $F \in C^\infty(N, M)$

$p \in M$ is a *critical point* of F if the differential $dF_p : T_pN \to T_{F(p)}M$ fails to be surjective at p, otherwise p is a *regular point* of F.

Definition 6.2: Critical value of $f \in C^\infty(N, \mathbb{R})$

Let f be a test function on N, $c \in \mathbb{R}$ is a critical value if there exists a $p \in f^{-1}(c)$ where df_p is not surjective.

Otherwise c is called a regular value. Caveat: if c is not in the image of f, we call c a regular value as well. It is clear c is a regular value iff $c \notin f(M)$ or for every $p \in f^{-1}(c)$, df_p is a surjection.

Notice $df_p : T_pN \to T_{f(p)}\mathbb{R}$ is not surjective $\iff$ all partials of the coordinate representation of f vanish. Since the matrix representation of df_p has the form

$$\mathcal{M}\{df_p\} = \begin{bmatrix} \left.\frac{\partial f}{\partial x^1}\right|_p & \cdots & \left.\frac{\partial f}{\partial x^m}\right|_p \end{bmatrix}$$

$\operatorname{rank} \mathcal{M}\{df_p\} \neq 1 \iff \left.\frac{\partial f}{\partial x^j}\right|_p = 0$ for $1 \leq j \leq m$.

Definition 6.3: Regular level set of $f \in C^\infty(N, \mathbb{R})$

If c is a regular value of f, then $f^{-1}(c)$ is called a regular level set.

Define $g = f - c$, the partials at $p \in N$ for both f and g agree, so the matrix representations of df_p and dg_p are identical (whereas their ranges might not be),

$$\begin{aligned}
\mathcal{M}\{dg_p\} &= \begin{bmatrix} \left.\frac{\partial g}{\partial x^1}\right|_p & \cdots & \left.\frac{\partial g}{\partial x^m}\right|_p \end{bmatrix} \\
&= \begin{bmatrix} \left.\frac{\partial f-c}{\partial x^1}\right|_p & \cdots & \left.\frac{\partial f-c}{\partial x^m}\right|_p \end{bmatrix} \\
&= \begin{bmatrix} \left.\frac{\partial f}{\partial x^1}\right|_p & \cdots & \left.\frac{\partial f}{\partial x^m}\right|_p \end{bmatrix} \\
&= \mathcal{M}\{df_p\}
\end{aligned}$$

Submanifolds

Definition 7.1: Embedded submanifold of M

A subset S of M is called an *embedded submanifold of M*, when:

- equipped with the subspace topology from M,
- it has a smooth structure that makes it a smooth manifold, and
- the inclusion map $\iota : S \hookrightarrow M$ is a smooth embedding.

Definition 7.2: Immersed submanifold of M

A subset S of M is called an *immersed submanifold of M*, when:

- equipped with an *arbitrary topology*,
- it has a smooth structure that makes it a smooth manifold (with or without boundary), and
- the inclusion map $\iota : S \hookrightarrow M$ is a smooth immersion.

Definition 7.3: Properly embedded submanifold

is an embedded submanifold, and the inclusion map $\iota : S \to M$ is a *proper* map.

Definition 7.4: Weakly embedded submanifold

is an immersed submanifold of M, whenever $F : N \to M$ is a smooth map and the range of F lies in S, then $F : N \to S$ as a map between two manifolds.

Measure zero

Definition 8.1: Measure zero

A subset $E \subseteq \mathbb{R}^n$ has *measure zero* if $\mu(E) = 0$

$$\mu(E) = \inf\left\{\sum_{j\geq 1} m(U_j),\ E \subseteq \bigcup U_j,\ U_j \in \mathcal{M}\right\}$$

where $\mathcal{M}$ denotes the Borel σ-algebra on $\mathbb{R}^n$. Each U_j can be taken to be open/closed balls or cubes, as they are measurable.

We sometimes refer to a measure zero set as *null*. If M is a smooth manifold of dimension m, then $E \subseteq M$ has *measure zero* if for every chart $(U, \varphi) \in \mathcal{A}_M$,

$$\varphi(U \cap E) \subseteq \mathbb{R}^m \quad \text{has measure zero in } \mathbb{R}^m$$

Proposition 8.1: Subadditivity and Monotonicity of null sets

The countable union of null sets is again null. Any measurable subset of a null set is again null. This holds for $\mathbb{R}^n$ and the smooth manifold case.

Proof. See any textbook on the theory of integration. ■

Proposition 8.2: Slice criterion for measure zero

Suppose $A \subseteq \mathbb{R}^n$ is compact (not necessary), and for every $c \in \mathbb{R}$, define the c-section of A

$$A_c = \left\{x \in \mathbb{R}^{n-1},\ (c, x) \in A\right\}$$

If A_c has measure zero in $\mathbb{R}^{n-1}$ for every $c \in \mathbb{R}$, then A has measure zero

Proof. By Tonelli's Theorem, since the Lebesgue measure on $\mathbb{B}_{\mathbb{R}^n}$ is a Radon measure, it is finite on compact sets. Hence

$$\mu(A) = \int f d\mu_n = \int \chi_A d\mu_1 \times \cdots \times \mu_n$$

where $f(c) = \int \chi_{A_c} d\mu_1 \times \cdots \times \mu_{n-1}$ is 0 for all c, hence $\int f d\mu_n$ integrates to zero. ■

Proposition 8.3: Equivalent condition for manifold measure zero

Let $E \subseteq M$, if $\{(U_\alpha, \varphi_\alpha)\}$ is a collection of charts that cover A, and $\varphi_\alpha(E \cap U_\alpha)$ has measure zero for every α, then E has measure zero.

Proposition 8.4: Smooth maps send null sets to null sets

If $F \in C^\infty(M, N)$, and $A \subseteq M$ has measure zero, so does $F(A)$. In particular, this holds for $F \in C^\infty(\mathbb{R}^m, \mathbb{R}^n)$ with the standard smooth structures.

Lemmas for Sard's Theorem

Lemma 9.1: Lemmas on regular/critical points

Let M and N be smooth manifolds, and $F \in C^\infty(M, N)$. Then,

1. The set of regular points of F are open in M.
2. The set of critical points of F are closed in M.
3. The set of critical values of F are precisely the direct image of its critical points. In symbols:

$$\text{Critical Values} = F(\text{critical points}) \tag{18}$$

Lemma 9.2: Local characterization of critical points

Let $p \in M$, and if $p \in (U, \varphi)$, $F(U) \subseteq (V, \psi)$. Then the rank of dF_p is independent of the choice of coordinate charts used, and it is equal to the rank of $d\hat{F}_{\hat{p}}$. Where $\hat{F}$ is the coordinate representation of F with respect to the charts (U, φ) and (V, ψ).

$$\hat{F} : \varphi(U) \to \psi(V) \quad \hat{F} \triangleq \psi \circ F \circ \varphi^{-1}$$

Moreover, if p a critical point of F, it is also a critical point of $F|_{U'} : U' \to V'$. Where U' is an open subset of U, and V' is an open subset of V that contains $F(U')$ entirely.

Lemma 9.3: LCH Lemma for Sard's Theorem

Let $\mathbf{X}$ be a LCH space. Suppose $x \in \mathbf{X}$, and U is an open set containing x. There exists an open precompact neighbourhood about x that *hides in* U. If V denotes this neighbourhood, then

$$x \in V \subseteq \overline{V} \subseteq U$$

Lemma 9.4: Covering Lemmas for Sard's Theorem

The following holds for an arbitrary m.

1. If $U \subseteq \mathbb{R}^m$ is an open set, it can be covered by
   - countably many open balls,
   - countably many open cubes,
   - countably many closed balls,
   - countably many closed cubes,

   all of the above are subsets of U.
2. If $E \subseteq \mathbb{R}^m$ is a closed cube with side length R, then it can be covered by K^m many closed cubes with equal side lengths $R' = RK^{-1}$.

Lemma 9.5: Diffeomorphism invariances for Sard's Theorem

Let M, N, P, Q be smooth manifolds. Let $F \in C^\infty(M, N)$, and $G \in \mathcal{D}(M, M)$,

1. If $p \in M$, define $H = F \circ G$, then
   - $\operatorname{rank} dF_p = \operatorname{rank} dH_p$,
   - p is a critical point of F iff it is a critical point of H,
   - $E \subseteq M$ has measure zero iff $G(E)$ has measure zero,
   - topological properties are preserved (applies for homeomorphisms G). E is compact, open, closed, connected or path-connected iff $G(E)$ is.
2. Let $H \in \mathcal{D}(Q, M)$, if U is an open subset of Q, then $H|_U : U \to H(U)$ is again a diffeomorphism, where U and $H(U)$ are interpreted as open-submanifolds of Q and M.

Sard's Theorem

Proposition 10.1: Sard's Theorem

Let M and N be smooth manifolds with dimensions m and n, then the set of critical values of F has measure-zero in N.

The proof for Sard's Theorem is extremely long. We will use induction on the dimension of M. The base case will be divided into two sub-cases, one where $n = 0$ and $n \geq 1$.

Step 1: Base case: $m = 0$

Let $G \in C^\infty(P, N)$ where P and N are smooth manifolds with dimensions $p = 0$ and $n \geq 1$. The set of critical values of G have measure zero in N.

As we will soon see, the induction hypothesis allows us to use several dimension reduction techniques to prove the induction step. These techniques will depend on the rank theorem.

Step 2: Induction hypothesis

Let $G \in C^\infty(P, N)$ where P and N are smooth manifolds with dimensions $0 \le p \le m-1$ and $n \ge 1$. The set of critical values of G have measure zero in N.

The induction step is extremely fussy. First, we dissect the domain and range of M and N into countably many charts $\{((U_{jk}, \varphi_{jk}), (V_{jk}, \psi_{jk})\}_{j,k\ge 1}$ where

$$F(U_{jk}) \subseteq V_{jk} \tag{19}$$

that also cover M and N.

By Lemma 9.2, if p is a critical point of F and $p \in U_{jk}$, then p is a critical point of

$$F_{jk} : U_{jk} \to V_{jk} \tag{20}$$

Step 3: Construction of charts

We construct a countable sequence of $\{((U_{jk}, \varphi_{jk}), (V_{jk}, \psi_{jk})\}_{j,k\ge 1}$ charts that satisfy Equations (19) and (20).

Proof. Let $\{(U_\alpha, \varphi_\alpha)\}$ be the maximal smooth atlas on M. Each U_α is open in M, and M is second-countable. It follows that there exists a countable open cover (indexed by $j \ge 1$) that covers M, and a countable smooth atlas $\{(U_j, \varphi_j)\}$. Similarly for $\{(V_k, \psi_k)\}$.

Let us write $U_{jk} \triangleq U_j \cap F^{-1}(V_k)$. Note U_{jk} is open, since smoothness implies continuity (Proposition 2.4). And

$$(V_{jk}, \psi_{jk}) \triangleq (V_k, \psi_k) \quad F_{jk} = F|_{U_{jk}}$$

so $F_{jk}(U_{jk}) \subseteq V_{jk}$. Finally, define $\varphi_{jk} = \varphi_j|_{U_{jk}}$. ■

Using Proposition 8.4, smooth maps preserve measure-zero sets. If the critical values of some F_{jk} have measure zero in N_{jk}, then so does ι(critical points of F_{jk}), where is the smooth $\iota : N_{jk} \hookrightarrow N$ is the inclusion map (which is a smooth embedding by Definition 7.1). Thus,

Step 4: Pass the measure-zero condition to each chart

Prove the following

$$\begin{aligned} & F(\text{critical points of } F) \text{ has measure zero} \\ \iff & \text{For } j,k \ge 1, \quad F_{jk}(\text{critical points of } F_{jk}) \text{ null in } V_{jk} \subseteq N \\ \iff & \text{For } j,k \ge 1, \quad \hat{F}_{jk}(\text{critical points of } \hat{F}_{jk}) \text{ null in } \psi_{jk}(V_{jk}) \subseteq \mathbb{R}^n \end{aligned}$$

where $\hat{F}_{jk}$ denotes the coordinate representation of F with respect to the charts (U_{jk}, φ_{jk}) and (V_{jk}, ψ_{jk}).

$$F_{jk} : \varphi_{jk}(U_{jk}) \to \psi_{jk}(V_{jk}), \quad \hat{F}_{jk} \triangleq \psi_{jk} \circ F_{jk} \circ \varphi_{jk} \tag{21}$$

Proof. The first equivalence follows from Lemma 9.2, and the observation that the image of F commutes with unions. The second follows from Proposition 8.4 and lem. 9.5. ■

To simplify the problem even further, we assume U_{jk} and V_{jk} are open subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$, as the core of the 'measure-zero' argument is invariant under diffeomorphisms of the coordinate charts (see Lemma 9.5). Let j, k be fixed, and relabel

$$((U_{jk}, \varphi_{jk}),\ (V_{jk}, \psi_{jk})) \mapsto ((U, \varphi),\ (V, \psi)) \quad \text{and} \quad F_{jk} \mapsto F$$

similarly for $\hat{F}_{jk} \mapsto \hat{F}$.

Step 5: Pass measure-zero condition to $\mathbb{R}^m$ and $\mathbb{R}^n$

Let U and V be open sub-manifolds of M and N, and φ and ψ be global (smooth) charts.

$$F(\text{critical points of } F) = (\psi^{-1} \circ \hat{F} \circ \varphi)(\text{critical points of } F)$$

and

$$\text{critical points of } F = \varphi^{-1}(\text{critical points of } \hat{F})$$

Proof. The first equation follows from the definition of $\hat{F}$, after unboxing the domains and ranges. ■

We again relabel:

$$U \mapsto \varphi(U), \quad V \mapsto \psi(V), \quad F \mapsto \hat{F}$$

and $M \mapsto \mathbb{R}^m$, $N \mapsto \mathbb{R}^n$, and critical points of F. Now $F : U \to V$ is a smooth map between open subsets of $\mathbb{R}^m$ and $\mathbb{R}^n$. Let us state the goal of the induction.

Step 6: Simplified induction goal

Let $U \subseteq \mathbb{R}^m$ and $V \subseteq \mathbb{R}^n$ be open subsets. $F \in C^\infty(U, \mathbb{R}^n)$ is a smooth map. Where $\mathbb{R}^m$ and $\mathbb{R}^n$ are considered to be smooth manifolds with the standard smooth structure. Prove that the set of critical values of F have measure zero.

We define some standard notation for partials.

- $\mathbb{N}_0 \triangleq \{0, 1, \ldots\}$,
- α is a *m-multi-index* or a *multi-index of order m* if $\alpha = (\alpha_1, \ldots, \alpha_m)$ is a m-tuple with entries in $\mathbb{N}_0$. The set of m-multi-indices are denoted by $\mathbb{N}_0^m$.
- The order of a multi-index α, denoted by $|\alpha|$ is its l^1 norm. In symbols:

$$|\alpha| = \sum |\alpha_j|$$

- If $g \in C^\infty(\mathbb{R}^m, \mathbb{R})$, and $\alpha \in \mathbb{N}_0^m$, we define

$$\partial^\alpha g \triangleq \left(\frac{\partial}{\partial x^1}\right)^{\alpha_1} \cdot \cdots \cdot \left(\frac{\partial}{\partial x^m}\right)^{\alpha_m}$$

  so that $\partial^\alpha \triangleq \prod_1^m \left(\frac{\partial}{\partial x^j}\right)^{\alpha_j} = \prod_{j=1}^m \partial_j^{\alpha_j}$.

Let $C \subseteq U$ denote the set of critical points of F, we decompose $F(C)$ into the following

$$F(C) = F(C \setminus C_1) + \bigcup_{j=1}^{k-1} F(C_j \setminus C_{j+1}) + F(C_k) \tag{22}$$

and show each of the three sets in the union have measure zero. (the direct image commutes with unions). Where

$$C_k \triangleq \left\{ x \in C,\ \forall i \leq n,\ \alpha \in \mathbb{N}_0^m,\ |\alpha| \leq k,\ \partial^\alpha F^i(x) = 0 \right\}$$

and F^i are the *component functions* of F, or $F(x) = (F^1(x), \dots,\ F^n(x))$.

Step 7: Properties of C, C_k

C, C_k is closed for every $k \geq 1$, and $U \supseteq C \supseteq C_1 \supseteq C_2 \supseteq \cdots$.

Proof. C is closed by Lemma 9.1. The function

$$h(x) \triangleq \sum_{\substack{\alpha \in \mathbb{N}_0^m \\ |\alpha| \leq k \\ i=1,\dots,n}} |\partial^\alpha F^i(x)|$$

is a continuous map from $U \to \mathbb{R}$. and $h^{-1}(0) \cap C$ is closed by continuity. $C \supseteq C_1$ follows by definition of C_1, and $C_k \supseteq C_{k+1}$ follows from comparing h for the two values of k. ■

Step 8: Case I: $C \setminus C_1$ has measure zero.

Let $a \in C \setminus C_1$, noting that $U \setminus C_1$ is an open set containing a. Since $a \notin C_1$, there exists a first order non-vanishing partial for a component function of F. Suppose

$$\frac{\partial}{\partial x^j} F^i(a) \neq 0 \quad \text{with respect to the standard charts}$$

We wish to swap the columns and rows of the coordinate charts so that $\frac{\partial}{\partial x^1} F^{-1}(a) \neq 0$. More concretely, the transposition of the ith and the 1st row is a permutation; so is the transposition of the jth and the 1st column. By invariance, we are free to relabel

$$F \mapsto P_{i,1} \circ F \circ P_{j,1}$$

Next, we use a convenient change of coordinates, to arrive at the form described below.

Step 9: Change of Coordinates

Perform a change of coordinates, such that if $(u, v^2, \dots, v^m) = (u, v)$, where $v \in \mathbb{R}^{n-1}$, then

$$F(u, v) = (u, F^2(u, v), \dots, F^n(u, v)) \tag{23}$$

and the differential is given by

$$dF_{(u,v)} = \begin{bmatrix} 1 & 0 \\ \left.\frac{\partial F^i}{\partial u}\right|_{(u,v)} & \left.\frac{\partial F^i}{\partial v^j}\right|_{(u,v)} \end{bmatrix} \quad {\scriptstyle i=2,\dots,n \atop j=2,\dots,m} \tag{24}$$

where $dF_{(u,v)}$ is not surjective for each $(u,v) \in C \setminus C_1 \subseteq C$. This change of coordinates is valid on some open, precompact set containing a, denoted by V_a. More precisely

$$a \in V_a \subseteq \overline{V}_a \subseteq B \subseteq U \setminus C_1 \tag{25}$$

Proof. If $(x^1, \dots, x^m)$ are the standard coordinates of $\mathbb{R}^m$, the map

$$T : (x^1, \dots, x^m) \mapsto (F^1(a), x^2, \dots, x^m) \quad \text{is a diffeomorphism}$$

on some open, precompact neighbourhood V_a of a.

$$a \in V_a \subseteq \overline{V_a} \subseteq U$$

The Jacobian of T is given by

$$\text{Jacobian}(T)(x) = \begin{bmatrix} \frac{\partial F^1}{\partial x^1} & \cdots \\ * & Id_{m-1} \end{bmatrix}$$

and it is invertible at a. The continuity of $\frac{\partial F^1}{\partial x^1}$ at a gives us an open ball $B \subseteq U \setminus C_1$ containing a, on which $\frac{\partial F^1}{\partial x^1} \neq 0$. Let $V_a \subseteq \overline{V}_a \subseteq B$ be an open, precompact neighbourhood that hides in this ball as in Lemma 9.3.

Next, we absorb the change of coordinates into F, so

$$F \mapsto (F \circ T^{-1})|_B$$

Verification of the Equation (23) is straightforward, as

$$T^{-1}(u,v) = ((F^1)^{-1}(u,v), v^2, \dots, v^m)$$

$$\begin{aligned}(F \circ T^{-1})(u,v) &= \Big(F^1[(F^1)^{-1}(u,v))], \dots, \dots\Big) \\ &= (u, F^2(u,v), \dots, F^n(u,v))\end{aligned}$$

Computing the Jacobian matrix gives Equation (24), and V_a clearly satisfies Equation (25). ∎

Let a range through all points in $C \setminus C_1$, then $\{C \setminus C_1 \cap V_a\}_{a \in C \setminus C_1}$ is an open cover Rel $C \setminus C_1$, so there exists a countable subcover. We will prove something slightly stronger, we prove that $F(C \setminus C_1 \cap \overline{V_a})$ has measure zero for each a in this countable collection. Proposition 8.1 will then prove $F(C \setminus C_1)$ is null.

Notice $C \setminus C_1 \cap \overline{V_a}$ is a closed subset of $\overline{V_a}$, hence compact. Continuous functions send compact sets to compact sets, so $F(C \setminus C_1 \cap \overline{V_a})$ is compact in $\mathbb{R}^n$. We wish to use the induction hypothesis and 'lower the dimension' of F, or at the very least — pass the argument to a lower dimension.

Step 10: $F(C \setminus C_1 \cap \overline{V_a})$ has measure zero.

In this step, we show every point in $C \setminus C_1$, admits a neighbourhood which is null. So

$$F(C \setminus C_1 \cap V_a) \subseteq F(C \setminus C_1 \cap \overline{V}_a)$$

V_a is open Rel $C \setminus C_1$, and a ranges through all the points in $C \setminus C_1$. $C \setminus C_1$ is second countable in the subspace topology, so $\{V_a\}$ admits a countable subcover, and we can apply subadditivity.

Proof. Let a an element in the countable subcollection described above. Let $c \in \mathbb{R}^1$ be arbitrary, if

$$\{c\} \times \mathbb{R}^{n-1} \cap F(C \setminus C_1 \cap \overline{V}_a) = \varnothing$$

then the c-section E_c of $F(C \setminus C_1 \cap \overline{V}_a)$,

$$E_c \triangleq \{q \in \mathbb{R}^{n-1},\ (c, q) \in F(C \setminus C_1 \cap \overline{V}_a)\}$$

is empty, so E_c is null. Next, if E_c is not empty, we will pass the argument to some $F_c : \underbrace{B_c}_{\subseteq \mathbb{R}^{m-1}} \to \mathbb{R}^{n-1}$ and show E_c is *exactly* (double inclusion) the set of critical points of F_c. Indeed, define

$$B_c = \left\{v,\ (c, v) \in \overline{V}_a\right\} \quad \text{and} \quad F(c, v) = (c, F_c(v))$$

so $F_c(v) = (F^2(c, v), \dots, F^n(c, v))$ for $v \in B_c$. The Jacobian (differential) of F_c resembles that of F. With c fixed,

$$(dF_c)_v = \left[v^j{}_{(u,v)}\right]_{\substack{i=2,\dots,n \\ j=2,\dots,m}}$$

Compare the previous equation with Equation (24), what we have done essentially is *removing the first column and row* of $dF_{(c,v)}$. By elementary linear algebraic arguments about canonical forms, rank $dF_{(c,v)} \neq n \iff \operatorname{rank}(dF_c)_v \neq n-1$. So $(c, v) \in C \setminus C_1 \cap \overline{V}_a$ if and only if v is a critical point of F_c. If (c, x) is a critical value in $C \cap \overline{V_a}$, then DF cannot have full rank so

$$\left[\frac{\partial F^i}{\partial v^j}\right](x) \ \text{ as a function of } x \in \mathbb{R}^{n-1} \text{ cannot have full rank.}$$

Now we wish to use the c-section Lemma in Proposition 8.2 in conjunction with our induction hypothesis. $C \setminus C_1 \cap \overline{V}_a$ is a closed subset of compact $\overline{V}_a$, hence $F(C \setminus C_1 \cap \overline{V}_a)$ is again compact. The c-section

$$\left[F(C \cap \overline{V}_a\right]_c = \left\{x \in \mathbb{R}^{n-1},\ (c, x) \in F(C \cap \overline{V}_a)\right\}$$

so (c, x) is a critical value of $F|_{\overline{V}_a}$. Another key idea is the following:

- Define $F_c(v) = (F^2(c, v), \dots, F^n(c, v))$,
- so $(DF_c)|_v = \left[\frac{\partial F^i}{\partial v^j}\right]$ from slicing the bottom right corner of the matrix in Equation (24).
- $\left[F(C \cap \overline{V}_a)\right]_c = \left\{x \in \mathbb{R}^{n-1},\ (c, x) \in F(C \cap \overline{V}_a)\right\}$

■

This completes Case I.

Whitney Embedding Theorem

Whitney Approximation Theorems

Tubular Neighbourhood Theorem

Transversality

Chapter 9: Integral Curves and Flows

Time invariant flows

Lie Derivatives of Vector Fields

Definition 2.1: Lie Derivatives of Vector Fields

Let M be a smooth manifold and V, W be smooth vector fields on M, and θ be the flow of V. The *Lie derivative of W with respect to V* is a rough vector field $\mathcal{L}_V W$

$$(\mathcal{L}_V W)_p \triangleq \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} d(\theta_t)_{\theta_t(p)}(W_{\theta_t(p)}) \tag{26}$$

The above uses the Frechet derivative on the (normed) vector space T_pM, and the result is an element in T_pM. Equivalently,

$$(\mathcal{L}_V W)_p = \lim_{t\to 0} \frac{d(\theta_{-t})_{\theta_t(p)}(W_{\theta_t(p)}) - W_p}{t} \tag{27}$$

Remark 2.1

If we use the regular the 'velocity of curves' definition, we will see the result is an element in $T_{W_p}(T_pM)$, which is not what we want. However, Proposition 3.13 allows us to identify, through the canonical linear isomorphism $T_pM \cong T_{W_p}(T_pM)$ for a fixed $W_p \in T_pM$.

Proposition 2.1: Existence of $\mathcal{L}_V W$

Let M be a smooth manifold and $V, W \in \mathfrak{X}(M)$, $(\mathcal{L}_V W)_p$ exists for every $p \in M$, and $\mathcal{L}_V W$ defines a smooth vector field.

Proof. Let θ denote the flow of V, and $p \in M$ be arbitrary. Let (x^i) be the local coordinates in some neighbourhood U about p. Since $(0, p)$ is a point in the flow domain of V, there exists an open interval containing 0 J_0, and an open neighbourhood U_0 containing p, such that $U_0 \subseteq U$.
Since every single point $\theta_t(x)$ for $(t, x) \in J_0 \times U_0$ share the same local coordinates as p, the matrix representation of the differential $d(\theta_{-t})_{\theta_t(x)} : T_{\theta_t(x)} : T_{\theta_t(x)}M \to T_xM$ is precisely the Jacobian of the coordinate representation of θ.

$$d(\theta_{-t})_{\theta_t(x)} = \left(\frac{\partial\theta^i}{\partial x^j}\right)(-t, \theta(t, x))$$

is a matrix of smooth functions $\frac{\partial\theta^i}{\partial x^j}$ evaluated at $(-t, \theta(t, x)) \in J_0 \times U_0$. Thinking of T_pM as a normed vector space, it has basis

$$\left(\left.\frac{\partial}{\partial x^1}\right|_p, \dots \left.\frac{\partial}{\partial x^n}\right|_p\right)$$

To show that $(\mathcal{L}_V W)_p$ exists, let $X : J_0 \to T_pM$ be a map (not a curve) starting at W_p, given by $X(t) = d(\theta_{-t})_{\theta_t(p)}(W_{\theta_t(p)})$, computing its Frechet derivative with respect to the above basis,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} X(t) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} \underbrace{\frac{\partial\theta^i}{\partial x^j}}_{\text{smooth matrix}} \underbrace{W^j(\theta(t, p))}_{W\text{ vector}} \underbrace{\left.\frac{\partial}{\partial x^i}\right|_p}_{\text{basis}} \tag{28}$$

$X(t)$ clearly has smooth coefficients in $J_0 \times U_0$, so the Frechet derivative exists. Further, the coefficients in Equation (28) vary smoothly with t and p in $J_0 \times U_0$. So the rough vector field $(\mathcal{L}_V W)$ is smooth around any p in M, and the proof is complete. ∎

Remark 2.2

The above holds for manifold with boundary as well, we require the vector field V to be tangent to ∂M, for the flow to be defined on the boundary. Further, we can view this Lie derivative as the Frechet limit of the *pullback* through the diffeomorphism $\theta_{-t} : M_{-t} \to M_t$. This point of view will be useful later on, when we look at Lie derivatives of (covariant) tensor fields.

Proposition 2.2: Computing Lie derivatives at t_0

Let M be a smooth manifold, $V, W \in \mathfrak{X}(M)$, and the flow of V by θ with domain $\mathcal{D}$. If $(t_0, p) \in \mathcal{D}$,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=t_0} d(\theta_{-t})_{\theta_t(p)}(W_{\theta_t}(p)) = d(\theta_{-t_0})\Big((\mathcal{L}_V W)_{\theta_{t_0}(p)}\Big)$$

Proof. Let $J_0 \times U_0$ is the product of an open interval containing 0 and an open neighbourhood of p, and $X : J_0 \to T_pM$, where

$$X(t) \triangleq d(\theta_{-t})_{\theta_t(p)}(W_{\theta_t}(p))$$

This is a smooth curve, and we can shrink J_0 and U_0 respectively so that θ maps $J_0 \times U_0$ into the same chart $(U, (x^i))$. Writing X in local coordinates, noting that X maps into the *vector space* T_pM. Equation (28) tells us X is indeed a smooth curve in T_pM. The composition of smooth maps is again smooth, define $\tau : J_0 \to \tau(J_0)$, with $t \mapsto t_0 + t$

$$(X \circ \tau)(t) = X(t + t_0) \quad \text{is smooth on } \tau(J_0)$$

and

$$(X \circ \tau)(t) = d(\theta_{-t_0})_{\theta_{t_0}(p)} \circ d(\theta_{-t})_{\theta_t(\theta_{t_0}(p))}(W_{\theta_t(\theta_{t_0}(p))})$$

$$\begin{aligned}
X'(t_0) &= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=t_0} X(t) \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} X(t + t_0) \\
&= \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} d(\theta_{-t_0}) \circ d(\theta_{-t})(W_{\theta_t(\theta_{t_0}(p))}) \\
&= d(\theta_{-t_0})\left(\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} W_{\theta_t(\theta_{t_0}(p))}\right) \\
&= d(\theta_{-t_0})(\mathcal{L}_V W)_{\theta_{t_0}(p)}
\end{aligned}$$

One way to justify the second last equality is to view the whole computation using Frechet derivatives. ∎

Flowouts from Boundary

This will be a draft of Problem 9.11. We wish to modify the proof for Theorem 9.20 (Flowout Theorem). We need to modify the following aspects

- M is now a smooth manifold with boundary,
- The original embedded submanifold of the flow $\mathcal{D}$, is $\mathcal{O} = (\mathbb{R} \times S) \cap \mathcal{D}$. We need to change it to

  $$\mathcal{O} = ([0, +\infty) \times \partial M) \cap \mathcal{D}$$

  where $\mathcal{D}$ is the flow of the smooth vector field N, which is everywhere inward pointing on the boundary of M.
- The induced smooth embedding $\Phi : P_\delta \to M$ (and thus diffeomorphism, since ∂M has codimension 1), is defined on

  $$P_\delta = \left\{(t, p),\ p \in \partial M,\ 0 \leq t < \delta(p)\right\}$$

  where $\delta : \partial M \to \mathbb{R}^+$ is a smooth, strictly positive function.
- For each $p \in \partial M$, the map $t \mapsto \Phi(t, p)$ is an integral curve of N starting at p.

Time varying flows

Chapter 12: Tensors

Tensor Products

Definition 1.1: Covariant Tensor

Let V be a finite-dimensional vector space. Covariant tensors are multi-linear maps $\alpha : V \times \cdots \times V \to \mathbb{R}$. If α sends k-copies of V into $\mathbb{R}$, it is called a k-*tensor*, or a *covariant tensor of rank* k. It is clear that k-tensors form a vector space over $\mathbb{R}$ for any $k \geq 0$. The convention is that 0-tensors are real numbers as they have 0 arguments in V.

$$T^k(V^*) = \left\{ \alpha : \underbrace{V \times \cdots \times V}_{k \text{ copies}} \to \mathbb{R},\ \alpha \text{ is multi-linear.} \right\}$$

Definition 1.2: Contravariant Tensor

Contravariant tensors are multi-linear maps $\alpha : V^* \times \cdots \times V^* \to \mathbb{R}$, similar to covariant tensors. Denoted by

$$T^k(V) = \left\{ \alpha : \underbrace{V^* \times \cdots \times V^*}_{k \text{ copies}} \to \mathbb{R},\ \alpha \text{ is multi-linear.} \right\}$$

Definition 1.3: Mixed Tensors

A (k, l)-mixed tensor, or just (k, l)-tensor is a multi-linear map that takes k copies of V^*, and l copies of V into $\mathbb{R}$. In symbols,

$$T^{(k,l)}(V) = \left\{ \alpha : \underbrace{V^* \times \cdots \times V^*}_{k \text{ copies}} \times \underbrace{V \times \cdots \times V}_{l \text{ copies}} \to \mathbb{R},\ \alpha \text{ is multi-linear.} \right\}$$

The space of k-covariant, k-contravariant, and (k, l)-tensors form a vector space over $\mathbb{R}$. We will rarely use contra-variant tensors, so we will use the word *tensor* to refer to covariant tensors only.

Definition 1.4: Tensor Product

Let F and G be k and l-tensors on V respectively, we define a bi-linear map,

$$F \otimes G : \underbrace{V \times \cdots \times V}_{k+l \text{ copies}} \to \mathbb{R}$$

by

$$(F \otimes G)(v_{\underline{k}}, w_{\underline{l}}) = \underbrace{F(v_{\underline{k}})G(w_{\underline{l}})}_{\substack{\text{scalar} \\ \text{multiplication} \\ \text{in } \mathbb{R}}}$$

Tensor product is clearly associative, and commutative (although we will not use this fact).

Proposition 1.1: Basis of $T^k(V^*)$

Let V be a n-dimensional vector space with basis (E_i), and dual basis (ε^i). $T^K(V^*)$ has basis

$$\mathbb{B} = \left\{ \bigotimes \varepsilon^{i_{\underline{k}}},\ i_{\underline{k}} \in \{\underline{n}\} \right\} = \left\{ \varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_k},\ 1 \leq i_j \leq n,\ 1 \leq j \leq k \right\}$$

This means, if A is a k-tensor, and for every multi-index $I = (i_{\underline{k}})$, define

$$A_I = A_{i_1,\dots i_k} = A_{i_{\underline{k}}} = A(E_{i_{\underline{k}}}) = A(E_{i_1}, \cdots, E_{i_k}) = A(E_I)$$

are precisely the *basis coefficients* of F with respect the basis $\mathbb{B}$,

$$F = F_{i_1,\dots,i_k} \varepsilon^{i_1} \otimes \cdots \otimes \varepsilon^{i_k} \tag{29}$$

Tensor Fields on Manifolds

Similar to the tangent bundle $TM = \coprod_{p\in M} T_pM = \bigcup_{p\in M}\{p\} \times T_pM$, is endowed with a unique smooth structure that makes the canonical projection $\pi : TM \to M$, $\pi(p, v) = p$ a smooth embedding. We can use the algebraic machinery we have built up from Chapter 10 to construct abstract tensor bundles on M. For completeness, we will state the definitions for covariant and mixed bundles as well.

Definition 2.1: Tensor Bundles

Let M be a smooth manifold with or without boundary, define

- Covariant k-bundle over M by
$$T^kT^*M = \coprod_{p\in M} T^k(T_p^*M)$$
- Contravariant k-bundle over M by
$$T^kTM = \coprod_{p\in M} T^k(T_pM)$$
- (k, l)-mixed bundle over M by
$$T^{(k,l)}TM = \coprod_{p\in M} T^{(k,l)}(T_pM)$$

Definition 2.2: Tensor Field on M

A *k-covariant tensor field*, or just a *k-tensor field* over a smooth manifold (with or without boundary) M, is a smooth section of the vector bundle as defined in Definition 2.1. As with vector fields $\mathfrak{X}(M)$, they form a R-module over $C^\infty(M)$ by pointwise multiplication.

We denote the space of k-tensor fields over M by $\mathcal{T}^k(M)$. Moreover, if $A \in \mathcal{T}^k(M)$, we can write A in local coordinates (x^i) by

$$A = A_{i_1,\dots,i_k} dx^{i_1} \otimes \cdots \otimes dx^{i_k}$$

with summation convention in effect, since $(dx^1, \dots, dx^n)$, when evaluated pointwise in local coordinates, form a dual basis of T_pM. Where each dx^i is the Chapter 11 differential of the i-th coordinate function.

Technicalities for Tensor Fields

We will skip the technicalities for now. Recall k-tensors on a vector space act on a k-tuple of vectors, the same is true for k-tensor fields over a manifold M. If $A \in \mathcal{T}^k(M)$, and $(X_1, \dots, X_k) \subseteq \mathfrak{X}(M)$, then

$$A(X_{\underline{k}}) = A(X_1, \dots, X_k) \quad \text{such that} \quad A(X_{\underline{k}})(p) = A_p(X_{\underline{k}}|_p) = A_p(X_1|_p, \dots, X_k|_p) \in \mathbb{R}$$

We should expect, that if A is a rough k-tensor field over M, it is smooth if and only if for every k-tuple of smooth vector fields, the resulting function (as defined pointwise above), is $C^\infty(M)$. See Lee Proposition 12.19 for more details.

Pullbacks of Tensor Fields

Definition 4.1: Pointwise pullback $dF_p^* : T^k(T_{F(p)}N) \to T^k(T_pM)$

Let M and N be smooth manifolds with or without boundary. Let $F : M \to N$ be a smooth map. For every $p \in M$, if α is a k-tensor on $T_{F(p)}N$, (so that $\alpha \in T^k(T^*_{F(p)}N)$), the *pointwise pullback* of α through dF_p is a k-tensor on T_pM. Denoted by $dF_p^*(\alpha)$, if $(v_{\underline{k}})$ are tangent vectors in T_pM, then

$$dF_p^*(\alpha)(v_{\underline{k}}) = \alpha\Big(dF_p(v_{\underline{k}})\Big) \tag{30}$$

Definition 4.2: Tensor Pullback $F^* : T^kN \to T^kM$

Let $F : M \to N$ be a smooth map, between smooth manifolds with or without boundary, if A is a k-tensor field on N, we define

$$(F^*A)_p(v_{\underline{k}}) = dF_p^*(A_{F(p)}) = A_{F(p)}(dF_p(v_{\underline{k}})) \tag{31}$$

the result is a k-tensor field (which is smooth, by the next Proposition).

Proposition 4.1: Proposition 12.25, Properties of Tensor Pullbacks

Let $F : M \to N$ and $G : N \to P$ be smooth maps between smooth manifold with or without boundary, and A and B are k-tensor fields on N, then

(i) For every $f \in C^\infty(N)$,

$$F^*(fB) = \underbrace{(f \circ F)}_{C^\infty(M)} F^*B$$

(ii) F^* commutes with tensor products,

$$F^*(A \otimes B) = F^*(A) \otimes F^*(B)$$

(iii) F^* is linear over $\mathbb{R}$,

$$F^*(aA + bB) = aF^*(A) + bF^*(B)$$

(iv) $F^*(B)$ is smooth,

(v) The tensor pullback satisfies the following co-functoral properties

- The (tensor) pullback of the composition is the (tensor) pullback of the *pre*-composition,

$$(G \circ F)^* = F^* \circ G^*$$

- The tensor pullback of the identity map on M is the identity map on tensor fields over M,

$$\mathrm{id}_M^* = \mathrm{id}_{\mathcal{T}^k(M)}$$

Proposition 4.2: Tensor Field Pullback in Coordinates

Let M and N be smooth manifolds with or without boundary, and $F : M \to N$ be a smooth map. A is a smooth covariant k-tensor field on N, and $p \in M$ induces local coordinates $(U, (x^i))$ about p, $(V, (y^j))$ about $F(p)$. If A has the following coordinate representation about $F(p)$,

$$A = \underbrace{A_{i_1,\ldots,i_k}}_{\substack{\text{smooth}\\ \text{function in N}}} dy^{i_1} \otimes \cdots \otimes dy^{i_k} = A_{i_{\underline{k}}} \Big(\bigotimes dy^{i_{\underline{k}}} \Big) \tag{32}$$

then F^*A is given by

$$\begin{aligned} F_p^* A &= (A_{i_1,\ldots,i_k} \circ F) d(y^{i_1} \circ F) \otimes \cdots \otimes d(y^{i_k} \circ F) && (33) \\ &= (A_{i_1,\ldots,i_k} \circ F) d(F^{i_1}) \otimes \cdots \otimes d(F^{i_k}) && (34) \\ &= (A_{i_{\underline{k}}} \circ F) \Big(\bigotimes d(F^{i_{\underline{k}}}) \Big) && (35) \end{aligned}$$

in local coordinates.

Lie Derivatives of Tensor Fields

Definition 5.1: Lie Derivative of Tensor Fields

Let M be a smooth manifold and $V \in \mathfrak{X}(M)$, and let θ be the flow of V. If $A \in \mathcal{T}^k(M)$ is a smooth k-tensor field, we define the *Lie derivative of A with respect to V*, denoted by $\mathcal{L}_V A$,

$$(\mathcal{L}_V A)_p \triangleq \frac{\mathrm{d}}{\mathrm{d}t}\bigg|_{t=0} (\theta_t^* A)_p = \lim_{t \to 0} \frac{d(\theta_t)_p^*(A_{\theta_t(p)}) - A_p}{t} \tag{36}$$

Similar to Definition 2.1, we view $T^k(T_p^*M)$ as a vector space, and $\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=0}(\theta_t^*A)_p$ is the Frechet Derivative on $T^k(T_p^*M)$.

Proposition 5.1: Existence of $\mathcal{L}_V A$

Let M be a smooth manifold, and $V \in \mathfrak{X}(M)$ with integral flow θ defined on $\mathcal{D} \subseteq \mathbb{R} \times M$. If A is a smooth covariant k-tensor field on M, then $(\mathcal{L}_V A)_p$ exists for every $p \in M$, and defines a smooth k-tensor field.

Proof. Fix $p \in M$, following ***Proposition*** 2.1 closely, since $(0, p) \in \mathcal{D}$ induces an open interval J_0 containing 0 and an open neighbourhood U_0 containing p. We write A in local coordinates for $(t, x) \in J_0 \times U_0$. Shrinking U_0 if necessary, we it suffices to assume θ maps $J_0 \times U_0$ into the same chart $(U, (x^i))$.

If $(t, x) \in J_0 \times U_0$, $A_{\theta_t(x)} \in T^k(T_{\theta_t(x)}M)$ has coordinate representation

$$A_{\theta_t(x)} = \underbrace{A_{i_1,\dots,i_k}(\theta(t,x))}_{\text{smooth function}} \underbrace{dx^{i_1}_{\theta_t(x)} \otimes \cdots \otimes dx^{i_k}_{\theta_t(x)}}_{\text{basis vector}}$$

By Equation (33), the pullback $d(\theta_t^*)_p(A_{\theta_t(p)})$ precomposes $A_{i_{\underline{k}}}$ with $\theta_t(p)$. Since $\theta(t, x) \in (U, (x^i))$, $\theta^j(t, x)$ is precisely the jth coordinate function for all $(t, x) \in J_0 \times U_0$.

$$d(\theta_t^*)_p(A_{\theta_t(p)}) = (A_{i_{\underline{k}}} \circ \theta^{(p)})(t)\left(\bigotimes dx_p^{i_{\underline{k}}}\right) \in T^k(T_p^*M)$$

$(A_{i_{\underline{k}}} \circ \theta^{(p)})(t)$ is clearly smooth in both arguments. So $(\mathcal{L}_V A)_p$ exists, and is smooth about every $p \in M$. ■

Proposition 5.2: Computing Lie derivatives at t_0 (Tensor Fields)

Let M be a smooth manifold, $V \in \mathfrak{X}(M)$, and the flow of V by θ with domain $\mathcal{D}$. Let A be a smooth covariant k-tensor field on M. If $(t_0, p) \in \mathcal{D}$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|_{t=t_0} d(\theta_{-t})_{\theta_t(p)}(A_{\theta_t}(p)) = d(\theta_{-t_0})\Big((\mathcal{L}_V A)_{\theta_{t_0}(p)}\Big)$$

Proof. The proof is the same as in Proposition 2.2. ■

Chapter 13: Riemannian Metrics

Chapter 14: Differential Forms

Multi-linear algebra

Definition 1.1: Wedge Product

Let $\omega \in \Lambda^k(V^*)$, $\eta \in \Lambda^l(V^*)$ the wedge product of ω and η is defined

$$\omega \wedge \eta \triangleq \frac{(k+l)!}{k!l!} \operatorname{Alt}(\omega \otimes \eta) \tag{37}$$

with $\operatorname{Alt}(\omega\otimes\eta) = \frac{1}{(k+l)!}\sum_{\sigma\in S_{k+l}}(\operatorname{sgn}\sigma)^\sigma(\omega\otimes\eta)$, and if $(v_1 \dots v_k, v_{k+1}, \dots v_{k+l})$ are vectors in V, then

$$\operatorname{Alt}(\omega\otimes\eta)(v_1 \dots v_k, v_{k+1}, \dots v_{k+l}) \triangleq \frac{1}{(k+l)!}\sum_{\sigma\in S_{k+l}}(\operatorname{sgn}\sigma)\omega(v_{\sigma(1)},\dots,v_{\sigma(k)})\eta(v_{\sigma(k+1)},\dots,v_{\sigma(k+l)}) \tag{38}$$

To simplify the expressions, we will use the following shorthand,

$$(v_1, \dots v_k) = (v_{\underline{k}}) \tag{39}$$
$$(v_{k+1}, \dots v_{k+l}) = (v_{k+1,\underline{k+l}}), \text{ alternatively} \tag{40}$$
$$(v_{k+1}, \dots v_{k+l}) = (v_{k+\underline{l}}) \tag{41}$$

and Equation (38) becomes

$$\operatorname{Alt}(\omega\otimes\eta)(v_{\underline{k+l}}) = \frac{1}{(k+l)!}\sum_{\sigma\in S_{k+l}}(\operatorname{sgn}\sigma)\omega(v_{\underline{\sigma(k)}})\eta(v_{\sigma(k),\underline{\sigma(k+l)}}) \tag{42}$$

Definition 1.2: Elementary k-covectors

Let V be a finite dimensional vector space with basis (E_i), and if (ε^i) is a dual basis for V^*, and if I is a k-index, where $I = (i_{\underline{k}})$, then

$$\varepsilon^I(v_{\underline{k}}) = \det\left(\begin{bmatrix} \varepsilon^{i_1}(v_1) & \cdots & \varepsilon^{i_1}(v_k) \\ \vdots & \ddots & \dots \\ \varepsilon^{i_k}(v_1) & \cdots & \varepsilon^{i_k}(v_k)\end{bmatrix}\right) = \det\left(\begin{bmatrix} v_1^{i_1} & \cdots & v_k^{i_1} \\ \vdots & \ddots & \dots \\ v_1^{i_k} & \cdots & v_k^{i_k}\end{bmatrix}\right) = \det\Big((\varepsilon^{i_j})v_l\Big)_{jl}$$

where $\varepsilon^{i_j}(v_l)$ is the projection of v_l onto the i_j basis vector E_{i_j}. The space of all ε^I where I is a k-index are called *elementary k-covectors*. It is clear that ε^I are alternating k-covectors.

Definition 1.3: Concatenating multi-indices

Let I and J be multi-indices, $I = (i_{\underline{k}})$ and $J = (j_{\underline{l}})$, then

$$IJ = (i_{\underline{k}}, j_{\underline{l}})$$

is a $k+l$ multi-index obtained by concatenating I and J.

Proposition 1.1: Proposition 14.8

Let V be a n-dimensional vector space with basis (E_i), and if (ε^i) is a dual basis for V^*, then $\Lambda^k(V^*)$ has the basis:

$$\overline{\mathcal{E}} = \left\{\varepsilon^I,\ I \text{ is an increasing } k\text{-index}\right\}$$

In particular, this means the vector space (algebra) of alternating k-covectors have dimension $\binom{n}{k}$. If $k = n$, then it is spanned by

$$\varepsilon^{(1,\ldots,n)} = \varepsilon^{(\underline{n})}$$

and all alternating n-covectors are of the form $A\varepsilon^{(\underline{n})}$, where $A \in \mathbb{R}$. This roughly means there is, up to a scalar multiple of $\varepsilon^{\underline{n}}$, only one (oriented) way of measuring volume in a basis-independent, and dimension-independent manner.

Proposition 1.2: Lemma 14.10

Let I and J be multi-indices, $I = (i_{\underline{k}})$ and $J = (j_{\underline{l}})$, then

$$\varepsilon^I \wedge \varepsilon^J = \varepsilon^{IJ} \tag{43}$$

Proposition 1.3: Lemma 14.11, Properties of the Wedge Product

The wedge product satisfies the following properties

- It is bi-linear over $\mathbb{R}$,
- It is associative,
- It is anti-commutative, for $\omega \in \Lambda^k(V^*)$ and $\eta \in \Lambda^l(V^*)$, then

  $$\omega \wedge \eta = (-1)^{kl}\eta \wedge \omega$$

- Formula for elementary covectors, let $\varepsilon^{i_{\underline{k}}}$ be covectors (covectors are assumed to have rank one), then

  $$\varepsilon^I \triangleq \varepsilon^{i_1} \wedge \cdots \wedge \varepsilon^{i_k}$$

  is an alternating k-covector.
- Determinant Law

  $$\omega^1 \wedge \cdots \wedge \omega^k(v_{\underline{k}}) = \det\Big(\omega^j(v_i)\Big)_{ji}$$

  where $(A_i^j)_{ji}$ denotes the matrix with entries A_i^j in the jth row and ith column.

One of the better ways of introducing the wedge product is through the 'determinant' convention, and working backwards. That way, interior multiplication will come naturally.

Definition 1.4: Interior multiplication

Let V be a finite dimensional vector space. If $x \in V$ is a vector, it induces a linear map $i_x : \Lambda^k(V^*) \to \Lambda^{k-1}(V^*)$ such that for every $\omega \in \Lambda^k(V^*)$, and $v_{\underline{k-1}} \in V$,

$$i_x(\omega)(v_{\underline{k-1}}) \triangleq \omega(x, v_{\underline{k-1}})$$

by fixing x in the first argument of ω. The result is an alternating $k-1$ covector, and it is clearly linear in ω over $\mathbb{R}$. We also write

$$i_x(\omega) = x \lrcorner \omega$$

Definition 1.5: Graded algebra

An algebra A is said to be *graded* if it has a direct sum decomposition $A = \bigoplus_{k\in\mathbb{Z}} A^k$ such that the algebra-product $(\cdot,\cdot) : A \times A \to A$ satisfies

$$(a, b) \in A^{k+l} \quad \text{for every } a \in A^k,\ b \in A^l$$

A graded algebra is said to be *anti-commutative*, if the algebra product satisfies

$$(a, b) = (-1)^{kl}(b, a)$$

We see that $\bigoplus_{k\in\mathbb{Z}} \Lambda^k(V^*)$ is a graded algebra, if we define $\Lambda^k(V^*) = \{0\}$ for $k \leq -1$, or $k \geq n+1$.

Proposition 1.4: Lemma 14.13, Properties of interior multiplication

Let V be a finite dimensional vector space and fix $x \in V$,

(i) $i_x \circ i_x \equiv 0$, where we interpret i_v as a linear map on the entire graded algebra.

(ii) i_x satisfies some kind of product rule. For every $\omega \in \Lambda^k(V^*)$ and $\eta \in \Lambda^l(V^*)$,

$$i_x(\omega \wedge \eta) = (i_x\omega) \wedge \eta + (-1)^k \omega \wedge (i_x\eta)$$

Differential Forms on Manifolds

Definition 2.1: Differential k-form on M

Let M be a smooth manifold with or without boundary, and $p \in M$ be fixed, $\lambda^k(T_p^*M)$ is a linear subspace of $T^k(T_p^*M)$ of dimension $\binom{n}{k}$, and

$$\Lambda^k T^*M \triangleq \coprod_{p\in M} \Lambda^k(T_p^*M)$$

is a smooth vector bundle of rank $\binom{n}{k}$ over M.

A differential k-form on M is a k-tensor field (or a smooth global section of the k-tensor bundle T^kT^*M) that is *pointwise alternating*, or equivalently: it is a smooth global section of the vector bundle Λ^kT^*M. The space of differential k-forms on M is denoted by

$$\Omega^k(M) = \Gamma(\Lambda^kT^*M)$$

Definition 2.2: Graded algebra of all k-forms

$\Omega^k(M)$ is a vector space over $\mathbb{R}$, and a left $C^\infty(M)$ module. If we stick with scalars over $\mathbb{R}$, and define the wedge product of two tensor fields ω, $\eta \in \Omega^k(M)$ (recall these are smooth global sections of some vector bundle),

$$(\omega \wedge \eta)_p \triangleq \omega_p \wedge \eta_p$$

We can inherit all the properties of the pointwise wedge product (see Proposition 1.3), namely

- Bi-linearity over $\mathbb{R}$,
- Associativity,
- Anti-commutativity

With this, the direct sum of all $\Omega^k(M)$ forms an associative, anticommutative graded algebra over $\mathbb{R}$, with the tensor wedge product as the algebra product.

$$\Omega^*(M) = \bigoplus_{k\geq 0} \Omega^k(M)$$

Definition 2.3: $\Omega^k(M)$ in local coordinates

Let $\omega \in \Omega^k(M)$, and (x^i) be the local coordinates on some open subset U of M. If

$$\left(\frac{\partial}{\partial x^1}, \dots, \frac{\partial}{\partial x^n}\right) \quad \text{is a coordinate frame on } TU \triangleq \pi^{-1}(U) \tag{44}$$

where $\pi : TM \to M$ is the canonical projection of the tangent bundle. Then,

$$\left(dx^1, \dots, dx^n\right) \quad \text{is the dual co-frame to the frame in Equation (44).}$$

by Proposition 1.1, we can write ω uniquely as a linear combination (over $C^\infty(M)$) of k-tensors which are *increasing elementary k-covectors* pointwise.

$$\omega = {\sum_I}' \omega_I dx^{i_1} \wedge \cdots \wedge dx^{i_k} = {\sum_I}' \omega_I dx^I$$

where ${\sum}'_I$ a sum over all increasing k-indices, and dx^I is the shorthand for the wedge product above, and ω_I is a number obtained by evaluating ω at the following

$$\omega_I \triangleq \omega\left(\frac{\partial}{\partial x^{i_1}}, \dots, \frac{\partial}{\partial x^{i_k}}\right), \quad \text{where} \quad I = (i_1, \dots i_k)$$

Since differential k-forms are (smooth) k-tensor fields, the pullback $F^* : \Omega^k(N) \to \Omega^k(M)$ is well defined, in particular,

Proposition 2.1: Lemma 14.16, Properties of k-form pullback

Let $F : M \to N$ be a smooth map, then

- The k-form pullback is linear over $\mathbb{R}$,
- The k-form pullback commutes with wedge products, or

$$F^*(\omega \wedge \eta) = F^*\omega \wedge F^*\eta$$

- In any smooth chart, if (y^i) are the local coordinates of N. For any k-form on N $\omega = \sum'_I \omega_I dy^{i_1} \wedge \dots \wedge dy^{i_k} = \sum'_I \omega_I dy^I$,

$$F^*\left(\sum{}' \omega_I dy^{i_1} \wedge \dots \wedge dy^{i_k}\right) = \sum_I{}' (\omega_I \circ F) d(y^{i_1} \circ F) \wedge \dots \wedge d(y^{i_k} \circ F)$$

Proof. Linearity of the k-form pullback over $\mathbb{R}$ is an immediate consequence of the linearity of the tensor field pullback. To prove the second claim, notice

$$F^*(A \otimes B) = F^*(A) \otimes F^*(B)$$

and

$$\begin{aligned} F^*(\operatorname{Alt}(A \otimes B)) &= F^*\left(\frac{1}{(k+l)!} \sum_{\sigma \in S_{k+l}} (\operatorname{sgn}\sigma)\sigma(A \otimes B)\right) \\ &= \frac{1}{(k+l)!} \sum_{\sigma \in S_{k+l}} (\operatorname{sgn}\sigma) F^*(\sigma(A \otimes B)) \end{aligned}$$

The tensor pullback commutes with permutations, let C be a k-tensor field on N, and $\sigma \in S_k$, then

$$(\sigma F^*(C))_p(v_{\underline{k}}) = (\sigma C)(dF_p(v_{\underline{k}})) = C(dF_p(\underline{v_{\sigma(k)}})) = F^*(\sigma C)_p(v_{\underline{k}})$$

Hence,

$$\begin{aligned} F^*(\operatorname{Alt}(A \otimes B)) &= \frac{1}{(k+l)!} \sum_{\sigma \in S_{k+l}} (\operatorname{sgn}\sigma)\sigma[F^*(A \otimes B)] \\ &= \frac{1}{(k+l)!} \sum_{\sigma \in S_{k+l}} (\operatorname{sgn}\sigma)\sigma[F^*(A) \otimes F^*(B)] \\ &= \operatorname{Alt}(F^*(A) \otimes F^*(B)) \end{aligned}$$

Since the wedge product differs by the alternating product by a constant, F^* commutes with the wedge product. The general case follows upon induction, and associativity of the wedge product.

For the third claim, use linearity, and the pullback of the wedge product is the wedge product of the pullback, noting that each y^{i_j} is a smooth function on N, so $F^*(dy^{i_j}) = d(y^{i_j} \circ F)$ by Proposition 11.25b in Lee, we state this Proposition for completeness:

> Let $F : M \to N$ be a smooth map, and u is a smooth function on N, then
> $$F^* du = d(u \circ F)$$

■

Proposition 2.2: Corollary 14.21, Tensor Pullback of Top-Degree Forms

Let $F : M \to N$ be a smooth map between manifolds with or without boundary *of the same rank* n. If (x^i) and (y^j) are local coordinates on open subsets U and V of M and N respectively, for every $u \in C(V, \mathbb{R})$,
$$F^*(u dy^1 \wedge \cdots \wedge dy^n) = (u \circ F)(\det DF) dx^1 \wedge \cdots \wedge dx^n$$
on an open set $U \cap F^{-1}(V)$, where DF is the Jacobian matrix of the coordinate representation of F.

Proof. The proof uses the fact that both manifolds are of dimension n, and any n-form on U is necessarily of the form
$$g dx^1 \wedge \cdots \wedge dx^n \tag{45}$$
where g is some continuous, real valued function on U. Solving for g pointwise yields the result. Suppose $p \in U \cap F^{-1(V)}$, and evaluating eq. (45) with the coordinate derivations $\left(\frac{\partial}{\partial x^{\underline{n}}}\Big|_p\right)$ reads

$$\begin{aligned}
\left(g dx^1 \wedge \cdots \wedge dx^n\right)_p \left(\frac{\partial}{\partial x^{\underline{n}}}\bigg|_p\right) &= g(p)(dx^1 \wedge \cdots \wedge dx^n)_p \left(\frac{\partial}{\partial x^{\underline{n}}}\bigg|_p\right) && \text{(pw. eval. of tensor field)} \\
&= g(p) dx^1_p \wedge \cdots \wedge dx^n_p \left(\frac{\partial}{\partial x^{\underline{n}}}\bigg|_p\right) && \text{(wedge pw. is pw. wedge)} \\
&= g(p) \det\left(dx^i_p\left(\frac{\partial}{\partial x^j}\right)\right)_{ij} && \text{(det. law from Proposition 1.3)} \\
&= g(p) \det\left(\delta^i_j\right)_{ij} \\
&= g(p)
\end{aligned}$$

Let $F^i = y^i \circ F$ be the ith coordinate component of F (which is a smooth function on M). So that dF^i and $\frac{\partial}{\partial x^j}$ form a covector-vector pair if evaluated pointwise. Apply the determinant law from Proposition 1.3 for every $p \in U$ reads

$$(dF^1_p \wedge \cdots \wedge dF^n_p)\left(\frac{\partial}{\partial x^1}\bigg|_p, \ldots, \frac{\partial}{\partial x^n}\bigg|_p\right) = \det\left(\left[dF^i_p\left(\frac{\partial}{\partial x^j}\bigg|_p\right)\right]_{ij}\right) \tag{46}$$

and
$$dF^1 \wedge \cdots \wedge dF^n\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right) = \det\left(dF^i\left(\frac{\partial}{\partial x^j}\right)\right)$$

Using the tensor pullback formula from Chapter 12, and by Lemma 14.16,

$$\begin{aligned} F^*(udy^1 \wedge \cdots \wedge dy^n) &= (u \circ F)d(y^1 \circ F) \wedge \cdots \wedge d(y^n \circ F) \\ &= (u \circ F)dF^1 \wedge \cdots \wedge dF^n \end{aligned}$$

Evaluating the last expression using the coordinate frame on $\left(\frac{\partial}{\partial x^1}, \ldots, \frac{\partial}{\partial x^n}\right)$ at some $p \in U \cap F^{-1}(V) \subseteq U$,

$$\begin{aligned} F^*(udy^1 \wedge \cdots \wedge dy^n)_p\left(\frac{\partial}{\partial x^{\underline{n}}}\bigg|_p\right) &= \underbrace{(u \circ F)(p)}_{\text{scalar}}(dF^1 \wedge \cdots \wedge dF^n)_p\left(\frac{\partial}{\partial x^{\underline{n}}}\bigg|_p\right) && \text{(pw. eval. of tensor field)} \\ &= (u \circ F)(p)(dF^1_p \wedge \cdots \wedge dF^n_p)\left(\frac{\partial}{\partial x^{\underline{n}}}\bigg| p\right) && \text{(wedge pw. is pw. wedge)} \\ &= (u \circ F)(p) \det\left(dF^i\left(\frac{\partial}{\partial x^j}\right)\right)_p && \text{(Equation (46))} \end{aligned}$$

Equating the two expressions we see that

$$g(p) = (u \circ F)(p) \det\left(dF^i\left(\frac{\partial}{\partial x^j}\right)\right)_p = (u \circ F)(p) \det(DF)_p$$

■

Exterior Derivatives

Definition 3.1: Exterior Derivative on Euclidean Space

Let U be an open subset of $\mathbb{R}^n$ or $\mathbb{H}^n$, and (x^i) be the local coordinates in U. If $\omega \in \Omega^k(U)$,

$$\omega = \sum_J{}' \omega_J dx^J$$

and we define the *exterior derivative of* ω by

$$d\omega = d\left(\sum_J{}' \omega_J dx^J\right) = \sum_J{}' d\omega_J \wedge dx^J$$

where $d\omega_J$ is the Chapter 11 differential of the smooth function ω_J, and $d\omega_J$ is a covector field on U. If $p \in u$ and $v \in T_pU$,

$$d(\omega_J)_p(v) \triangleq v(\omega_J)$$

Since $\Omega^1(U) = \mathfrak{X}(U)$, the wedge product $d\omega_J \wedge dx^J = d\omega_J \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_k}$ is alternating.

With this, $d : \Omega^k(\mathbb{R}^n) \to \Omega^{k+1}(\mathbb{R}^n)$ (resp. $\mathbb{H}^n$) is a linear map over $\mathbb{R}$. We can rewrite $d\omega_J$ in terms of the coframe on U, suppose

$$d\omega_J = \frac{\partial \omega_J}{\partial x^i} dx^i \implies d\omega_J \wedge dx^J = \frac{\partial \omega_J}{\partial x^i} dx^i \wedge dx^J = \sum_{i=1}^n \frac{\partial \omega_J}{\partial x^i} dx^i \wedge dx^J$$

Summing over all increasing k-indices J, we get

$$d\omega = d\left(\sum_J{}' \omega_j dx^J\right) = \sum_J{}' \sum_{i=1}^n \frac{\partial \omega_J}{\partial x^i} dx^i \wedge dx^{j_1} \wedge \cdots \wedge dx^{j_k}$$

Chapter 22: Symplectic Manifold

Symplectic Tensors

Definition 1.1: Billinear forms

Let V be a vector space, a *billinear form* $\omega : V \times V \to \mathbb{R}$ is a 2-tensor on V.

Definition 1.2: Characterization of billinear forms

Let ω be a billinear form on V, it is

- *symmetric* if
$$\omega(x, y) = \omega(y, x)$$
- *skew-symmetric* or *anti-symmetric* if
$$\omega(x, y) = (-1)\omega(y, x)$$
- *alternating* if
$$\omega(x, x) = 0$$

If V is a vector space over the field F and $\mathrm{char}(F) \neq 2$, then the last two conditions are equivalent. Moreover,

- V is called an *orthogonal geometry* if ω is symmetric.
- V is called a *symplectic geometry* if ω is alternating.

Definition 1.3: Metric vector space

A vector space (not necessarily finite dimensional) is called a *metric vector space* if it is a orthogonal or symplectic geometry.

Matrices and billinear forms

Definition 2.1: Matrix of billinear form

If $B = (b_1, \ldots, b_n)$ is an ordered basis for V, we define the *matrix representation of* ω by

$$\mathcal{M}(\omega) = (a_{ij}) = (\omega(b_i, b_j))$$

Proposition 2.1: Matrix induces a billinear form

Let $A = (a_{ij})$ be a matrix on V with respect to some basis $B = (b_{\underline{n}})$ it is clear that A induces a billinear form, on V through $A(x, y) = [x]_B^T A[y]_B$, where $[\cdot]_B$ denotes the canonical isomorphism $V \cong \mathbb{R}^n$ with respect to the basis B.

$$[x]_B^T A [y]_B = \begin{bmatrix} x^1 & \dots & x^n \end{bmatrix} A \begin{bmatrix} y^1 \\ \vdots \\ y^n \end{bmatrix}$$

for $x = x^i b_i$ and $y = y^j b_j$.

Moreover,

$$A[x]_B = \begin{bmatrix} A(b_1, x) \\ \vdots \\ A(b_n, x) \end{bmatrix}$$ is a *column* vector whose entries are given by applying x on the second coordinate

and

$$[x]_B^T A = \begin{bmatrix} A(x, b_1) & \cdots & A(x, b_n) \end{bmatrix}$$ is a *row* vector whose entries are given by applying x on the first coordinate

Let A_B be the matrix representation of ω with respect to the B, if C is another basis on V, then how do we compute A_C? The answer is simple, recall for any vector $x \in V$, $x = x_B^i b_i$ and $x = x_C^j c_j$, then

$$[x]_B = M_{C,B}[x]_C \quad \text{for some matrix of an automorphism } M_{C,B}$$

$\omega(x, y) = [x]_B^T A_B [y]_B = ([x]_C^T M_{C,B}^T) A_B (M_{C,B}[y]_C) = [x]_C^T A_C [y]_C$, then

$$M_{C,B}^T A_B M_{C,B} = A_C \tag{47}$$

We can describe this relation between the two matrices A_B and A_C by the following

Definition 2.2: Congruent matrices

Two matrices M and N are said to be *congruent*, if there exists an invertible matrix P for which

$$P^T M P = N$$

Congruence is an equivalence relation on the space of matrices, and the equivalence classes over congruence are called *congruence classes*.

Proposition 2.2: Characterization of matrices using congruence

Let A_1 and A_2 be matrix representations of two billinear forms with respect to the basis B.

$$A_1 = (A_1(b_i, b_j))_{ij} \quad A_2 = (A_2(b_i, b_j))_{ij}$$

They induce the same billinear form if and only if they are congruent.

Definition 2.3: Alternate matrices

Let M be a matrix with F-coefficients, it is *alternate* if it is skew symmetric and is *hollow*; meaning it has 0s on the main diagonal. If $F = \mathbb{R}$ or $\mathrm{char}(F) \neq 2$, then alternate matrices are and are precisely the skew-symmetric matrices.

Orthogonality

For this section, (V, ω) will denote a metric vector space, not necessarily finite-dimensional unless we are using matrix representations.

Definition 3.1: Orthogonal complements

A vector $x \in V$ is orthogonal to another vector $y \in V$, written $x \perp y$, if $\omega(x, y) = 0$.

If V is an orthogonal or symplectic geometry then $\perp$ is a symmetric relation. If E is a subset of V, we denote the *orthogonal complement of E* by

$$E^\perp \triangleq \Big\{ v \in V,\, v \perp E \Big\}$$

Definition 3.2: Characterization of metric vector spaces

- A nonzero vector $x \in V$ is *isotropic*, or *null* if $\omega(x, x) = 0$
- V is *isotropic* if it contains at least one isotropic vector.
- V is *anisotropic* or *nonisotropic* if for every $x \in V$, $\omega(x, x) = 0 \implies x = 0$,
- V is *totally isotropic* (that is, symplectic if $\mathrm{char}(F) \neq 2$) if $\omega(x, x) = 0$ for every vector $x \in V$.

The first bullet point above is about vectors in V, while the others are properties of V.

- A vector $x \in V$ is called *degenerate* if $x \perp V$, that is,

$$\forall y \in V,\, \omega(x, y) = 0$$

- The *radical* of V, denoted by $\mathrm{rad}(V)$ is the set of all degenerate vectors in V,

$$\mathrm{rad}(V) \triangleq V^\perp$$

- V is *singular* or *degenerate* if $\mathrm{rad}(V) \neq \{0\}$,
- V is *non-singular* or *non-degenerate* if $\mathrm{rad}(V) = \{0\}$,
- V is *totally singular*, if $\mathrm{rad}(V) = V$.

To summarize,

- V is isotropic if there exists a non-zero isotropic vector, meaning $\omega(x, x) = 0$, for some $x \neq 0$,
- V is degenerate if there exists a degenerate vector, $x \perp V$.

Proposition 3.1: Matrix invariants under congruence

Non-singularity, symmetry, and skew-symmetry are invariants under congruence.

Proof. ■

Proposition 3.2: Characterization of non-degeneracy

V is non-degenerate if and only if every matrix representation A of ω is non-singular.

Proof. Suppose V is non-degenerate, then let $B = (b_{\underline{n}})$ be a basis for V, if A is the matrix representation of ω with respect to B, let x be a non-zero vector in V, so $x \notin \operatorname{rad}(V)$

$$b_i^T A[x]_B = \omega(b_i, x) \neq 0 \implies A[x]_B \neq 0$$

so A is non-singular. If A' is another matrix representation with respect to another basis C, by Equation (47) A' is non-singular as well.

Conversely, if every matrix representation of ω is non-singular, let x be a non-zero vector in V, then $A[x]_B \neq 0$ is a non-zero vector so there exists some basis component $(A[x]_B)^j$ that is non zero, and

$$[b_j]_B^T A[x]_B = \omega(b_j, x) \neq 0$$

therefore V is non-degenerate. ■

Proposition 3.3: Characterisation of billinear forms from matrix representations

Let ω be a billinear form on V, if $\mathcal{M}(\omega)$ the induced matrix representation relative to any basis. Assume V is a vector space over $\mathbb{R}$, then

- it is symmetric iff $\mathcal{M}(\omega)$ symmetric as a matrix,
- it is skew-symmetric, iff alternating iff $\mathcal{M}(\omega)$ is skew-symmetric as a matrix.

Corollary 3.1: Characterisation of non-singular symplectic form

Let (V, ω) be a finite dimensional vector space over $\mathbb{R}$, equipped with a billinear form ω. (V, ω) is a non-singular symplectic vector space iff the matrix representation of ω with respect to every basis is non-singular and skew-symmetric.

Riesz Representation Theorems

Proposition 4.1

Let (V, ω) be a nonsingular metric vector space, the map $x \mapsto x \lrcorner \omega \in V^*$ defined by

$$x \lrcorner \omega = \omega(x, \cdot), \quad \text{and} \quad (x \lrcorner \omega)(y) = \omega(x, y), \quad \forall y \in V$$

is a linear isomorphism from V to V^*.

Isometries

Definition 5.1: Isometry between MVS

Let (V, ω) and (W, η) be metric vector spaces. An *isometry* $\tau \in L(V, W)$ is a linear isomorphism that preserves the billinear form.

$$\omega(u, v) = \eta(\tau u, \tau v)$$

Definition 5.2: Orthogonal, symplectic groups

Let V be a nonsingular metric vector space. If V is an orthogonal (resp. symplectic) geometry, the set of all isometries on V is called the *orthogonal (resp. symplectic) group on* V. It is a group under composition, and is denoted by $\mathcal{O}(V)$ (resp. $\mathrm{Sp}(V)$).

Hyperbolic spaces, nonsingular completions

Canonical Forms

Symplectic Manifolds

Darboux's Theorem

Proposition 9.1: Lie Derivatives of Tensor Fields (along time-varying vector fields)

Let M be a smooth manifold. Suppose $V : J \times M \to TM$ is a smooth time-varying vector field on M. Denote the time-varying flow of V by $\psi : \mathcal{E} \to M$. Let $A \in \mathcal{T}^k(M)$ be a smooth time-invariant covariant k-tensor field on M. For every $(t_1, t_0, p) \in \mathcal{E}$,

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=t_1} (\psi^*_{t,t_0} A)_p = (\psi^*_{t_1,t_0}(\mathcal{L}V_{t_1} A))_p \tag{48}$$

Chapter Hofer book

Definition 0.1: Symplectic vector space

Let V be a finite dimensional vector space over $\mathbb{R}$. It is a *symplectic vector space* if it admits a non-singular, antisymmetric billinear form $\omega : V \times V \to \mathbb{R}$.

$$\omega(u, v) = -\omega(v, u)$$

for $u, v \in V$. By the previous section on Riesz Representation, the linear map

$$\hat{\omega} : V \to V^*, \quad v \mapsto \omega(v, \cdot)$$

is a linear isomorphism of V onto its dual vector V^*.

We define the *standard symplectic vector space* $(\mathbb{R}^{2n}, \omega_0)$, where $n \in \mathbb{N}^+$, where

$$\omega_0(u, v) =< Ju, v > \quad J \triangleq \begin{bmatrix} 0 & I_n \\ -I_n & 0 \end{bmatrix}$$

where $< \cdot, \cdot >$ denotes the standard inner product on $\mathbb{R}^{2n}$.

$$\omega_0(u, v) =< Ju, v >=< u, J^T v >= u^T J^T v \tag{49}$$

$J^T = -J$ by Corollary 3.1.

We will mainly deal with non-singular symplectic forms because of Riesz isomorphism.

Definition 0.2: Symplectic linear map

Let (V, ω) be a symplectic vector space. A linear map $F \in \mathrm{Hom}(V)$ is *symplectic* if it preserves symplectic form ω. For every $u \in V$,

$$< u, v >=< Au, Av > \triangleq A^* \omega(u, v)$$

where $A^* : \mathcal{T}^*(V) \to \mathcal{T}^*(V)$ denotes the tensor pullback by precomposing any tensor S by A

$$\forall S \in \mathcal{T}^k(V), \quad A^* S(v_{\underline{k}}) \triangleq S(Av_{\underline{k}})$$

The set of linear symplectic maps on a $2n$-dimensional vector space form a group under composition. It is a Lie Group denoted by $\mathrm{Sp}(n)$.

Proposition 0.2: Symplectic Maps are Area-preserving

Let $(\mathbb{R}^{2n}, \omega_0)$ denote the standard symplectic space. If $\varphi \in \mathrm{Sp}(n)$, then $\det \varphi = 1$.

Proof. See page 4. ■

$$
\begin{aligned}
(\Lambda_{s-|\alpha|}\partial^\alpha f)\hat{} &= (1+|\zeta|^2)^{s/2-|\alpha|/2} \cdot (\partial^\alpha f)\hat{} && (50)\\
&= (1+|\zeta|^2)^{s/2-|\alpha|/2} \cdot (2\pi i\zeta)^\alpha \cdot \hat{f} && (51)\\
&= (2\pi i)^{|\alpha|}(1+|\zeta|^2)^{(s-|\alpha|)/2} \cdot |\zeta|^{|\alpha|} \cdot \hat{f} && (52)\\
&\leq |\alpha|(1+|\zeta|^2)^{s/2}\hat{f} && (53)
\end{aligned}
$$

Chapter 0: Banach Spaces

Introduction

This section is quite incomplete, and all over the place. I have been meaning to put all the notation/terminology I am going to use in this section. Please skip to the Chapter 1 for now.

Banach Spaces

A *Banach space* is a normed vector space that is Cauchy-complete under the usual metric induced by its norm.
If E and F are Banach spaces over $\mathbb{R}$. We will denote the norms on E, and F by single lines, so

$$|x| = \|x\|_E \quad \text{and} \quad |y| = \|y\|_F \quad \forall x \in E,\ y \in F$$

$\mathcal{L}(E, F)$ will denote the space of linear maps between E and F. In the category of Banach spaces, the space of morphisms are called *toplinear morphisms* - or *CLM*s (*continuous linear maps*); which we will denote by $L(E, F)$ for toplinear morphisms bewteen E and F.

We use $\|\cdot\|_{L(E,F)}$ or $\|\cdot\|$ to denote the operator norm, depending on how much emphasis we wish to place on $L(E, F)$. Recall,

$$\begin{aligned}\|\varphi\|_{L(E,F)} &= \inf\Big\{A \geq 0,\ |\varphi(x)| \leq A|x|\ \forall x \in E\Big\} \\ &= \sup\Big\{|\varphi(x)|,\ x \in E,\ |x| = 1\Big\}\end{aligned}$$

By the open mapping theorem: any continuous surjective linear map is an open map. Hence invertible elements in $L(E, F)$ are naturally called *toplinear isomorphisms*. If $\varphi \in L(E, F)$ such that φ preserves the norm between the Banach Spaces, that is for every $x \in E$, $|x| = |\varphi(x)|$ then we call φ an *isometry*, or a *Banach space isomorphism*. If E_1 and E_2 are Banach spaces, we will use the usual *product norm* $(x_1, x_2) \mapsto \max(|x_1|, |x_2|)$.

- We say a map F is *between* the spaces X and Y if $F : X \to Y$.
- $\mathcal{L}(V^K, W)$ denotes the space of k-linear maps from V to W that are not necessarily continuous.

Proposition 2.1: Hahn Banach Theorem (Geometric Form)

Let E be a Banach space, A and B are closed disjoint subsets of E. Assuming one of the two is compact, then there exists a clf λ which *strictly separates* A and B.

$$A \subseteq [\lambda \leq \alpha - \varepsilon] \quad \text{and} \quad B \subseteq [\lambda \geq \alpha + \varepsilon] \tag{54}$$

for $\alpha \in \mathbb{R}$ and $\varepsilon > 0$.

Definition 2.1: Product of Banach Spaces

Let $E_1, \ldots, E_k$ be Banach spaces over $\mathbb{R}$. The Cartesian product of $(E_1, \ldots E_k)$ is denoted by $\prod_i^k E_i$.

It is again a Banach space with the norm

$$(x_1, \ldots, x_k) \mapsto |(x_1, \ldots, x_k)| = \sup_{1 \le i \le k} |x_i| \tag{55}$$

Vector Spaces

Let V be any vector space over $\mathbb{R}$ or $\mathbb{C}$, and $\{v_\alpha\} \subseteq V$, the symbol $\sum^{\wedge} v_\alpha$ refers to a partially specified object which is any **finite** linear combination of the elements of $\{v_\alpha\}$. If the cardinality of $\{v_\alpha\}$ is finite,

$$\sum^{\wedge} v_\alpha = \sum^{\wedge} v_{\underline{k}} \text{ for some } k \geq 1. \tag{56}$$

where eq. (56) should be interpreted as eq. (57)

$$\sum^{\wedge} v_{\underline{k}} = \sum_{i=\underline{k}} c^i v_i \tag{57}$$

for some $c^i \in \mathbb{R}$ or $\mathbb{C}$ where $i = \underline{k}$.
Composition of maps: If $f : E \to F$ and $g : F \to G$ are maps between Banach spaces, we write gf to mean $g \circ f$.

Chapter 1: Multilinear maps

Bilinear maps

Definition 1.1: Bilinear map

A map $\varphi : E_1 \times E_2 \to F$, where F is also a Banach space, is said to be *bilinear* if

$$\varphi(x, \cdot) : E_2 \to F \quad \text{and} \quad \varphi(\cdot, y) : E_1 \to F$$

are linear for every $x \in E_1$ and $y \in E_2$.

Proposition 1.1: Continuity criterion of a bilinear map

Let E_1, E_2, F be Banach spaces, a bilinear map $m : E_1 \times E_2 \to F$ is continuous if and only if there exists a $C \geq 0$, where

$$|m(x, y)| \leq C|x||y| \tag{58}$$

Proof. Suppose such a C exists, fix a convergent sequence $(x_n, y_n) \to (x, y)$ in $E_1 \times E_2 = E$. Because the projection maps are continuous, this means $x_n \to x$ and $y_n \to y$. Using inspiration from the proof where $x_n y_n \to xy$, where

$$x_n(y_n - y) + (x_n - x)y = x_n y_n - xy \quad x, y, x_n, y_n \in \mathbb{R}$$

Using the inspiration, and replacing multiplication in $\mathbb{R}$ with the bilinear map m, we have:

$$\begin{aligned} m(x_n,\, y_n - y) + m(x_n - x,\, y) &= m(x_n,\, y_n) - m(x,\, y) \\ |m(x_n,\, y_n) - m(x,\, y)| &\leq C[|x_n| \cdot |y_n - y| + |x_n - x| \cdot |y|] \to 0 \end{aligned}$$

Conversely, if m is continuous, then it is continuous at the origin $(0, 0) = 0$. There exists a δ where $|(x, y)| \leq \delta$ implies $|m(x, y)| \leq 1$. Now, if $x, y \neq 0$ are elements in E, we normalize so that (x, y) has length δ

$$|(x|x|^{-1}\delta, y|y|^{-1}\delta)| = \delta|(x|x|^{-1}, y|y|^{-1})| = \delta$$

So that $|m(x|x|^{-1}\delta, y|y|^{-1}\delta)| \leq 1$, using bilinearity of m:

$$|m(x, y)| \leq \delta^{-2}|x| \cdot |y|$$

Setting $\delta^{-2} = C$ finishes the proof (notice if eithe x or y is 0, then m is trivially 0 and the inequality holds). ■

Proposition 1.2: $L(E_1, E_2; F)$ is isomorphic to $L(E_1, L(E_2, F))$

For each bilinear map $\omega \in L(E_1, E_2; F)$, there exists a unique map $\varphi_\omega \in L(E_1, L(E_2, F))$ such that $|\omega| = |\varphi_\omega|$; such that for every $(x, y) \in E_1 \times E_2$, $\omega(x, y) = \varphi(x)(y)$.

Proof. Let $\varphi_\omega : E_1 \to L(E_2, F)$ be the unique map such that $\varphi_\omega(x)(y) = \omega(x, y)$. Proposition 1.1 shows that $\varphi_\omega(x)$ is a continuous linear map into F at each x, and $|\varphi_\omega(x)| \leq |\omega||x|$. This holds for an arbitrary x, and $\varphi_\omega(\cdot)$ is clearly linear, hence $|\varphi_\omega| \leq |\omega|$. Reversing the roles of ω and φ shows proves the other estimate.

The rule as outlined above is linear in ω; and it is not hard to see $\varphi : L(E_1, E_2; F) \to L(E_1, L(E_2, F))$ is an injection. By the open mapping theorem, the proposition is proven if φ is a surjection. Fix $\theta \in L(E_1, L(E_2, F))$, define a map $\omega : E_1 \times E_2 \to F$ such that $\omega(x, \cdot) = \theta(x)(\cdot)$. So that ω is linear in its second argument. To show ω is linear in its first: fix a linear combination $A = \sum^{\wedge} x$ in E_1, and $y \in E_2$.

$$\omega(A, y) = \theta(\sum^{\wedge} x)(y) = \sum^{\wedge} \theta(x)(y) = \sum^{\wedge} \omega(x, y)$$

Continuity follows from Equation (58), and $\varphi_\omega = \theta$ as needed. ∎

Notation

We will use the following notation to simplify computations with multilinear maps. Let E and F be sets, and $v_1, \dots, v_k \in E$. $f : E \to F$.

- Listing individual elements: $v_{\underline{k}}$ means $v_1, \dots, v_k$ as separate elements.
- Creating a k-list: $(v_{\underline{k}}) = (v_1, \dots, v_k) \in \prod E_{j \le k}$ if $v_i \in E_i$ for $i = \underline{k}$.
- Double indices: $(v_{\underline{n_k}}) = (v_{n_{\underline{k}}}) = (v_{n_1}, \dots, v_{n_k})$, and

$$(v_{\underline{n_k}}) \neq (v_{n_{(1,\dots,k)}})$$

- Closest bracket convention:

$$(v_{(n_{\underline{k}})}) = (v_{(n_1,\dots,n_k)}) \quad \text{and} \quad (v_{n_{(\underline{k})}}) = (v_{n_{(1,\dots,k)}})$$

- Underlining 0 means it is iterated 0 times:

$$(v_{\underline{0}}, a, b, c) = (a, b, c)$$

- Skipping an index:

$$(v_{\underline{i-1}}, v_{i+\underline{k-i}}) = (v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_k)$$

  for $i = \underline{k}$.
- Applying f to a particular index:

$$(v_{\underline{i-1}}, f(v_i), v_{i+\underline{k-i}}) = (v_1, \dots, v_{i-1}, f(v_i), v_{i+1}, \dots, v_k)$$

  Of course, if $i = 1$, then the above expression reads $(f(v_1), v_2 \dots, v_k)$ by the $\underline{0}$ interpretation.
- In any list using this 'underline' notation, we can find the size of a list by summing over all the underlined terms, and the number of terms with no underline.
- If $\wedge : E \times E \to F$ is any associative binary operation,

$$\wedge(v_{\underline{k}}) = v_1 \wedge \cdots \wedge v_k$$

Remark 2.1: Preview of exterior calculus

We can write the formula for the determinant of a $\mathbb{R}^{k\times k}$ matrix in this notation. Suppose $a_i \in \mathbb{R}$, and $b_i \in \mathbb{R}^{k-1}$ for $i = \underline{k}$.

$$M = \begin{bmatrix} a_1 & \cdots & a_k \\ | & & | \\ b_1 & \cdots & b_k \\ | & & | \end{bmatrix}$$

The determinant of M is a linear combination of determinants of $k-1$-sized matrices, given in terms of the columns of b

$$\det(M) = \sum_{i=\underline{k}} (-1)^{i-1} a_i \det\Big(b_{\underline{i-1}}, b_{i+\underline{k-i}}\Big)$$

k-linear maps

Definition 3.1: k-linear maps

Let $E_{\underline{k}}$, F be Banach spaces. A map $\varphi : \prod E_{\underline{k}}$ is k-linear if for every $i = \underline{k}$, $v_i \in E_i$,

$$\varphi(\cdot^{\underline{i-1}}, v_i, \cdot^{\underline{k-i}}) : \prod(E_{\underline{i-1}}, E_{i+\underline{k-i}}) \to F \quad \text{is } (k-1)\text{-linear}$$

A k-linear *symmetric* map between Banach spaces E, F is a map $A \in \mathcal{L}(E^k, F)$ such that for every k-permutation $\theta \in S_{\underline{k}}$,

$$A(v_{\underline{k}}) = A(v_{\underline{\theta(k)}})$$

The following theorem should give confidence to the notation we have adopted to use.

Proposition 3.1: Continuity criterion of k-linear maps

Let $E_{\underline{k}}$ and F be Banach spaces, a k-linear map $\varphi : \prod E_{\underline{k}} \to F$ is continuous iff there exists a $C > 0$, such that for every $x_i \in E_i$, $i = \underline{k}$

$$\Big|\varphi(x_{\underline{k}})\Big| \le C \prod \Big|x_{\underline{k}}\Big|$$

Proof. Suppose φ is continuous, then it is continuous at the origin. Picking $\varepsilon = 1$ induces a $\delta > 0$ such that for $\Big|(x_{\underline{k}})\Big| \le \delta$, $\Big|\varphi(x_{\underline{k}})\Big| \le 1$. The usual trick of normalizing an arbitrary vector $(x_{\underline{k}}) \in \prod E_{\underline{k}}$ does the job:

$$\Big|\varphi(x_k \cdot \Big|x_{\underline{k}}\Big|^{-1} \cdot \delta)\Big| \le 1 \implies \Big|\varphi(x_{\underline{k}})\Big| \le \delta^{-k} \prod \Big|x_{\underline{k}}\Big|$$

Conversely, fix a sequence (indexed by n, in k elements in the product space $\prod E_{\underline{k}}$), so

$$(x_n^{\underline{k}}) \to (x^{\underline{k}}) \quad \text{as } n \to +\infty \tag{59}$$

To proceed any further, we need to prove an important equation that decomposes a difference in φ.

$$\varphi(b^{\underline{k}}) - \varphi(a^{\underline{k}}) = \sum_{i=\underline{k}} \varphi(b^{\underline{i-1}}, \Delta_i, a^{i+\underline{k-i}}) \tag{60}$$

where $(b^{\underline{k}})$ and $(a^{\underline{k}})$ are elements in $\prod E_{\underline{k}}$, and $\Delta_i = b^i - a^i$ for $i = \underline{k}$. The proof is in the following note, which is in more detail than usual - to help the reader ease into the new notation.

Note 3.1

We proceed by induction, and eq. (60) follows by setting $m = k$ in

$$\varphi(a^{\underline{k}}) = \varphi(b^{\underline{m}}, a^{m+\underline{k-m}}) - \sum_{i=\underline{m}} \varphi(b^{\underline{i-1}}, \Delta_i, a^{i+\underline{k-i}}) \tag{61}$$

Base case: set $m = 1$, by definition of k-linearity (def. 3.1) of φ. Since $a^1 = b^1 - \Delta_1$,

$$\varphi(a^{\underline{k}}) = \varphi(b^1 - \Delta_1, a^{1+\underline{k-1}}) = \varphi(b^1, a^{1+\underline{k-1}}) - \varphi(\Delta_1, a^{1+\underline{k-1}})$$

Induction hypothesis: suppose eq. (61) holds for a fixed m. Since $a^{m+1} = b^{m+1} - \Delta_{m+1}$,

$$\begin{aligned}
\varphi(a^{\underline{k}}) &= \varphi(b^{\underline{m}}, a^{m+\underline{k-m}}) - \sum_{i=\underline{m}} \varphi(b^{\underline{i-1}}, \Delta_i, a^{i+\underline{k-i}}) \\
&= \varphi(b^{\underline{m}}, a^{m+1}, a^{(m+1)+\underline{k-(m+1)}}) - \sum_{i=\underline{m}} \varphi(b^{\underline{i-1}}, \Delta_i, a^{i+\underline{k-i}}) \\
&= \varphi(b^{\underline{m+1}}, a^{(m+1)+\underline{k-(m+1)}}) - \varphi(b^{\underline{m+1}}, \Delta_{m+1}, a^{(m+1)+\underline{k-(m+1)}}) - \sum_{i=\underline{m}} \varphi(b^{\underline{i-1}}, \Delta_i, a^{i+\underline{k-i}})
\end{aligned}$$

and this proves eq. (60)

We substitute $a^i = x^i$, and $b^i = x_n^i$ for $i = \underline{k}$, and eq. (60) becomes eq. (62)

$$\varphi(x_n^{\underline{k}}) - \varphi(x^{\underline{k}}) = \sum_{i=\underline{k}} \varphi(x_n^{\underline{i-1}}, x_n^i - x^i, x^{i+\underline{k-i}}) \tag{62}$$

Then the triangle inequality reads

$$\begin{aligned}
\left|\varphi(x_n^{\underline{k}}) - \varphi(x^{\underline{k}})\right| &\leq \sum_{i=\underline{k}} \left|\varphi(x_n^{\underline{i-1}}, x_n^i - x^i, x^{i+\underline{k-i}})\right| \\
&\leq \sum_{i=\underline{k}} |\varphi| \cdot \Pi\left(x_n{}^{\underline{i-1}}, \Delta_i, x^{i+\underline{k-i}}\right) \\
&\leq \sum_{i=\underline{k}} |\varphi| \cdot \left|x_n^i - x^i\right| \, \Pi \left(x_n^{\underline{i-1}}, x^{i+\underline{k-i}}\right) \\
&\lesssim_n |\varphi| \sup_{i=\underline{k}} |x_n^i - x^i| \to 0
\end{aligned}$$

where we identify the product $\Pi\left(v^{\underline{k}}\right)$ with the product of their norms $\Pi\left(\left|v^{\underline{k}}\right|\right)$. ∎

Remark 3.1: Currying isomorphism

The k-linear variant of prop. 1.2 holds. We will use but not prove this fact.

Remark 3.2: k-linear maps from the same space

We denote the space of k-linear maps from E into F by $L(E_{\underline{k}}; F) = L(E^k, F) = L^k(E, F)$. *Tensors* on E are k-linear maps from the product space of E into $\mathbb{R}$, by replacing F with $\mathbb{R}$.

Chapter 2: Differentiation

The derivative

Definition 1.1: Open sets and neighbourhoods

If U is an open subset of a topological space X, we denote this by $U \mathring{\subseteq} X$. If U is a *neighbourhood* of a point $p \in X$, we write $p \mathring{\in} U$.

We do not require neighbourhoods to be open sets; rather, we say U is a neighbourhood of p when the interior of U contains p.

Definition 1.2: Little o

A real-valued function in a real variable defined for all t sufficiently small is said to be $o(t)$ if $\lim_{t\to 0} o(t)/t = 0$. A map $\psi : U \to F$ where $U \mathring{\subseteq} E$ contains 0 in E, is said to be $o(h)$ if $|\psi(h)|/|h| \to 0$ as $h \to 0$ in E.

Definition 1.3: Differentiability

Let $f : E \to F$ be a map, replacing E and F by their open subsets if necessary. We say f is *differentiable* at $x \in E$ when there exists a **continuous linear map on** E: $\lambda \in L(E, F)$ such that

$$f(x+h) = f(x) + \lambda h + o(h) \quad \text{for sufficiently small } h \tag{63}$$

The role $o(h)$ plays here is a map from $U \to F$, where U is some neighbourhood of 0.

Proposition 1.1: Basic properties of the derivative

If f is differentiable at x, then the λ in eq. (63) is unique. We write $f'(x) = Df(x) = \lambda$ as in ??. Furthermore, if $f'(x)$ and $g'(x)$ exist, then $(f+g)'(x) = f'(x) + g'(x)$ as linear maps, similar for scalar multiplication.

Proof. Suppose $\lambda_i \in L(E, F)$ are both derivatives of f at x. Then,

$$\begin{cases} f(x+h) = f(x) + \lambda_1(h) + o(h) \\ f(x+h) = f(x) + \lambda_2(h) + o(h) \end{cases}$$

And $(\lambda_1 - \lambda_2)(h) = o(h) = \varphi(h) \cdot |h|$, where $\varphi(h) \to 0$ as $h \to 0$. Using the operator norm, we see that

$$\|\lambda_1 - \lambda_2\|_{L(E,F)} \leq |\varphi(h)| \to 0$$

This proves uniqueness. Suppose f and g are differentiable at x, denote $\lambda_f = f'(x)$ (resp. $g'(x)$). The definition of def. 1.3 reads

$$\begin{aligned} f(x+h) + g(x+h) &= (f(x) + g(x)) + (\lambda_f(h) + \lambda_g(h)) + o(h) + o(h) \\ (f+g)(x+h) &= (f+g)(x) + (\lambda_f + \lambda_g)(h) + o(h) \end{aligned} \tag{64}$$

since eq. (64) satisfies eq. (63), the proof is complete. ∎

Proposition 1.2: Chain rule

Let E, F, G be Banach spaces. If $f \in C^1(E, F)$, $g \in C^1(F, G)$, for every $x \in E$,

$$(g \circ f)'(x) = g'(f(x)) \circ f'(x) \tag{65}$$

Proof. Since f is differentiable at x, $f(x+h) = f(x) + f'(x)(h) + o_1(h)$, (resp. for g, $o_2(h)$). Set $k(h) = f(x+h) - f(x)$, and

$$\begin{aligned} g(f(x+h)) &= g(f(x)) + g'(f(x))(k(h)) + o_2(k(h)) \\ &= g(f(x)) + g'(f(x))(f'(x)(h) + o_1(h)) + o_2(k(h)) \\ (g \circ f)(x+h) &= (g \circ f)(x) + g'(f(x)) \circ f'(x)(h) + g'(f(x))(o_1(h)) + o_2(k(h)) \\ (g \circ f)(x+h) &= (g \circ f)(x) + g'(f(x)) \circ f'(x)(h) + o(h) \end{aligned}$$

because $|A(o_1(h))| \leq |A||o_1(h)|$ for all $A \in L(E, F)$; and $o(k(h)) = o(h)$ for every continuous $k : E \to F$ such that $k(h) \to 0$ as $h \to 0$. ■

Proposition 1.3: Derivatives of CLMs

If $\lambda \in L(E, F)$, then $\lambda \in C^1(E, F)$ and $D\lambda(x) = \lambda$ for every $x \in E$. Furthermore, if $f \in C^1(E, F)$, and $\nu \in L(F, G)$, then the composition $\nu \circ f$ is in $C^1(E, G)$, and $(\nu \circ f)'(x) = \nu \circ f'(x)$ for every $x \in E$.

Proof. See $\lambda(x+h) = \lambda(x) + \lambda(h) + 0$ at every $x \in E$. Using the chain rule (prop. 1.2) proves the second claim. ■

Proposition 1.4: Product rule in k variables

Let $m : \prod F_{\underline{k}} \to G$ be a k-linear map between Banach spaces $F_{\underline{k}}$ and G. Suppose $f_i \in C^1(E, F_i)$ with $i = \underline{k}$, writing

$$m(f_{\underline{k}})(x) = m(f_{\underline{k}}(x)) \tag{66}$$

then $m(f_{\underline{k}})$ is in $C^1(E, G)$ and for every $y \in E$,

$$Dm(f_{\underline{k}})(x)(y) = \sum_{i=\underline{k}} m(f_{\underline{i-1}}(x), Df_i(x)(y), f_{i+\underline{k-i}}(x)) \tag{67}$$

Proof. Let x be fixed. Equation (67) is proven if we show eq. (68)

$$m(f_{\underline{k}})(x+h) = m(f_{\underline{k}})(x) + \left(\sum_{i=\underline{k}} m(f_{\underline{i-1}}(x), Df_i(x)(h), f_{i+\underline{k-i}}(x)) \right) + o(h) \tag{68}$$

and for sufficiently small h we have

$$f_i(x+h) - f_i(x) = Df_i(x)(h) + o(h^i) \tag{69}$$

We will use the difference formula in eq. (61), with the following substitutions

$$f_i(x+h) = b^i \qquad f_i(x) = a^i \tag{70}$$

$$Df_i(x)(h) = c^i \qquad o(h^i) = \varepsilon^i \tag{71}$$

$$f_i(x+h) - f_i(x) = c^i + \varepsilon^i \qquad \Delta^i = o(h^i) + c^i \tag{72}$$

With these substitutions, the equation we want to prove (eq. (67)) becomes eq. (73)

$$m(b^{\underline{k}}) - m(a^{\underline{k}}) = \left(\sum_{i=\underline{k}} m(a^{\underline{i-1}}, c^i, a^{i+\underline{k-i}})\right) + o(h) \tag{73}$$

Starting from eq. (61),

$$m(b^{\underline{k}}) - m(a^{\underline{k}}) = \sum_{i=\underline{k}} m(b^{\underline{i-1}}, \Delta^i, a^{i+\underline{k-i}})$$

We can expand each term, if $i = \underline{k}$,

$$m(b^{\underline{i-1}}, \Delta^i, a^{i+\underline{k-i}}) = m(b^{\underline{i-1}}, c^i, a^{i+\underline{k-i}}) + m(b^{\underline{i-1}}, o(h^i), a^{i+\underline{k-i}}) \tag{74}$$

Let us study the first term in eq. (74), and with i held fixed, define

$$m_i(z^{\underline{i-1}}) = m(z^{\underline{i-1}}, c_i, a^{i+\underline{k-i}}) \tag{75}$$

Expanding the first term within eq. (74), and because m_i as defined in eq. (75) is $i-1$-linear (because it is a k-linear map with $k-(i-1)$ variables held constant); we use eq. (61) again.

$$m_i(b^{\underline{i-1}}) = \left(\sum_{i=\underline{k}} m_i(b^{\underline{j}}, \Delta^j, a^{j+\underline{(i-1)-j}})\right) + m_i(a^{\underline{i-1}}) \tag{76}$$

Unboxing the last term in eq. (76) using the definition of m_i reads

$$m(b^{\underline{i-1}}, \Delta^i, a^{i+\underline{k-i}}) = m(a^{\underline{i-1}}, c^i, a^{i+\underline{k-i}}) + \sum_{j=\underline{i-1}} m_i(b^{\underline{j}}, \Delta^j, a^{j+\underline{(i-1)-j}}) \tag{77}$$

We wish to remove all of the b^is. Since $\Delta^i = c^i + \varepsilon^i$ (eq. (72)), we have

$$\begin{aligned}
m(b^{\underline{k}}) - m(a^{\underline{k}}) &= \sum_{i=\underline{k}} m(b^{\underline{i-1}}, c^i, a^{i+\underline{k-i}}) + m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}}) \\
&= \left(\sum_{i=\underline{k}} m_i(b^{\underline{i-1}})\right) + \sum_{i=\underline{k}} m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}}) \\
&= \left(\sum_{i=\underline{k}} m_i(a^{\underline{i-1}}) + \sum_{j=\underline{i-1}} m_i(b^{\underline{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}})\right) + \sum_{i=\underline{k}} m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}}) \\
&= \left(\sum_{i=\underline{k}} m_i(a^{\underline{i-1}})\right) + \sum_{\substack{i=\underline{k} \\ j=\underline{i-1}}} m_i(b^{\underline{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}}) + \sum_{i=\underline{k}} m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}})
\end{aligned} \tag{78}$$

The last term within eq. (78) is $o(h)$, since it is a linear combination of $o(h^i)$s.

$$\left|\sum_{i=\underline{k}} m(b^{\underline{i-1}}, \varepsilon^i, a^{i+\underline{k-i}})\right| \lesssim_{m,a,b} |o(h)| \tag{79}$$

Each summand in the second last term in eq. (78) is $o(h)$ as well, as

$$\begin{aligned}\left|m_i(\underline{b^{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}})\right| &\leq |m_i| \, \textcircled{\scriptsize{II}} \, (\underline{b^{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}}) \\ &\leq |m| \cdot \left(\textcircled{\scriptsize{II}}(c^i, a^{i+\underline{k-i}})\right)\left(\textcircled{\scriptsize{II}}(\underline{b^{j-1}}, \Delta^j, a^{j+\underline{(i-1)-j}})\right) \\ &\lesssim_{m,a,b} \sup_{\substack{i=\underline{k} \\ j=\underline{i-1}}} \left|c^i\right| \cdot \left|\Delta^j\right| \\ &\lesssim_{m,a,b} \sup_{\substack{i=\underline{k} \\ j=\underline{i-1}}} |Df_i(x)(h)| \cdot |f_j(x+h) - f_j(x)| \\ &\lesssim_{m,a,b} |Df_i(x)||h| \sup_{\substack{i=\underline{k} \\ j=\underline{i-1}}} \left|\Delta^j\right| \\ &\lesssim_{m,a,b} |o(h)| \end{aligned} \tag{80}$$

for the second last estimate we used $\Delta^j \to 0$. Therefore the second term in eq. (78) is $o(h)$, and eq. (68) is proven. Therefore $m(f_{\underline{k}})$ is differentiable at x. Continuity of $Dm(f_{\underline{k}})$ follows from the fact that

$$Dm(f_{\underline{k}})(x) = \sum_{i=\underline{k}} m(f_{\underline{i-1}}(x), Df_i(x)(\cdot), f_{i+\underline{k-i}}(x)) \tag{81}$$

and each of the summands eq. (81) can be broken down as the product of the compositions shown in eqs. (82) and (83)

$$x \mapsto (f_{\underline{i-1}}(x), f_{i+\underline{k-i}}(x)) \mapsto m(f_{\underline{i-1}}(x), \cdot, f_{i+\underline{k-i}}(x)) \tag{82}$$

$$x \mapsto Df_i(x)(\cdot) \tag{83}$$

which are continuous from E to $L(E, F)$. ■

Chapter 3: Integration

Introduction

This chapter will be on the integration of *regulated* mappings, the space of which are precisely the uniform closure of rectangle functions. from a compact interval. We will go through some of the elementary results, and prove the Fundamental Theorem.

Integration of step mappings

Definition 2.1: Partition on $[a, b]$

Let $I = [a, b]$ be a compact interval. An N-*partition* P on I is a list of $N + 1$ elements in $[a, b]$, which are assumed to be well ordered as in $p_0 \le p_1 \le \cdots \le p_N$.

$$P = (a = p_0, p_1, \ldots, p_N = b) \quad \text{or} \quad P = (p_0, p_{\underline{N}}) \tag{84}$$

The space of partitions on I will be denoted by I_p.

As per usual, we have *common refinements of partitions*, given two partitions P and Q on the same compact interval $I = [a, b]$, where P is defined as in eq. (84), and $Q = (q_0, q_{\underline{N}})$ similarly. The common refinement of P and Q is another partition R on I which contains all of the elements in $P \cup Q$.

- Given a partition P of size N represented as $P = (p_0, p_{\underline{N}})$, the cells of P are indexed using their rightmost points.
- The interval (p_{i-1}, p_i) is denoted as $\text{cell}(p_i)$, and
- the *length* of the ith cell: $|\text{cell}\, p_i| = |p_i - p_{i-1}|$.
- If we want to sequence the cells of P based on their right endpoints, it is expressed as $\text{cell}(P) = \Big(\text{cell}(p_{\underline{N}})\Big)$.
- Note that these cells do not form a disjoint union of I.

Remark 2.1: Assume all intervals are compact

For the rest of this chapter, we assume all intervals are compact and of the form $I = [a, b]$. If P, Q, R are partitions, their elements will be represented by p_i, (resp. r_i, q_i).

Definition 2.2: Step mapping

A step mapping on $I = [a, b]$ is a vector space of maps from I to a Banach space E over $\mathbb{R}$. It is equipped with the supremum norm, and its elements are denoted by Σ,

$$\Sigma = \Big\{ f : [a, b] \to E, \text{ there exists a } N\text{-partition } P \in I_p,\ \{v_{\underline{N}}\} \subseteq E \text{ such that } f|_{(p_{i-1}, p_i)} = v_i\ \forall i = \underline{N} \Big\} \tag{85}$$

I $f \in \Sigma$, we denote its norm by $\|f\|_u = \sup_{x \in I} |f(x)|$.

Definition 2.3: Integration on Σ

If $f \in \Sigma$ and is of the form inside the set-builder notation in eq. (85), we define the integral of f by

$$\int_a^b f = \sum_{i=\underline{N}} (p_i - p_{i-1}) v_i \tag{86}$$

Remark 2.2: Distinguishing between intervals I, J

If I and J are compact intervals, we distinguish the step mappings from I and J by Σ_I and Σ_J.

We now state some definition and properties of eq. (86) which we will not prove.

Proposition 2.1: Properties of the integral on Σ

Let I and J be intervals, $f, f_{\underline{k}} \in \Sigma_I$, and $g \in \Sigma_J$.

- The integral is linear, that is
$$\int \sum^{\wedge} f_{\underline{k}} = \sum^{\wedge} \int f_{\underline{k}} \tag{87}$$
- The integral over $[b, a]$ is *defined* to be the negative of eq. (86):
$$\int_a^b f = -\int_b^a f \tag{88}$$
- The integral is domain-additive, if $b = c$, then
$$\int_a^b f + \int_c^d g = \int_a^d (f + g) \tag{89}$$
where we identify $(f + g)$ to be the step mapping in $\Sigma_{[a,d]}$ whose restriction I (resp. J) agree with f (resp. g).

Product of step mappings

Let $E_{\underline{k}}$ be Banach spaces, and $I = [a, b]$ a fixed compact interval. Let E refer to the product space $\prod E_{\underline{k}}$, which is equipped with the supremum norm as outlined in def. 2.1

$$\Sigma_i = \left\{ f_i : I \to E_i,\ f_i \text{ is a step mapping.} \right\}$$

There are two ways of defining the space of step-mappings from I into E eqs. (90) and (91). Using a combinatorical argument with common refinements, it is not hard to see the two are subsets of each other.

$$\Sigma_E^1 = \left\{ f : I \to E,\ \mathrm{proj}_i\, f \in \Sigma_i\ \forall i = \underline{k} \right\} \tag{90}$$

$$\Sigma_E^2 = \left\{ f : I \to E,\ f \text{ is a step mapping.} \right\} \tag{91}$$

And since the product space E is toplinearly isomorphic to its external direct sum, $E_1 \times \cdots E_k$, the integral over $\Sigma_E = \Sigma_E^1 = \Sigma_E^2$ is defined to be

$$\int_a^b f = \left(\int_a^b \mathrm{proj}_{\underline{k}} f\right) = \left(\int_a^b \mathrm{proj}_1 f, \ldots, \int_a^b \mathrm{proj}_k f\right) \tag{92}$$

Regulated mappings

Definition 4.1: Regulated mappings

Let I be a compact interval. A mapping from I into E is *regulated* if it is the uniform limit of step mappings. We denote the space of regulated mappings by $\overline{\Sigma_I}$ or $\overline{\Sigma}$.

Proposition 4.1: Continuity implies a regulated mapping

Every continuous function $f : I \to E$ is the uniform limit of step mappings in $\Sigma_I = \Sigma$.

Proof. Let $f \in C(I, E)$, the continuity of f is uniform; given $\varepsilon > 0$ there exists $\delta > 0$ where $|y - x| < \delta$ implies $|f(y) - f(x)| < \varepsilon$. δ induces a smallest integer $n \geq 1$ such that $p_n = a + n\delta > b$. Define $p_0 = a$ and $p_i = a + i\delta$, relabelling $p_n = b$, we see that $P = (p_0, p_{\underline{n}})$ is a partition.

We construct a step mapping by sampling values of f. Set $g|_{\mathrm{cell}(p_i)} = f(p_i)$, $g(a) = f(a)$, $g(p_i) = f(p_i)$. Defining the endpoints is necessary, and g still remains a member of Σ_I by eq. (85). Each $x \in I \setminus P$ belongs in some cell(p_i), of which $|p_i - x| < \delta$, and $g(x) = f(p_i)$ implies $|g(x) - f(x)| < \delta$. If x is in P, then $g(x) = f(x)$, and $\|f - g\|_u \leq +\varepsilon$. ■

Proposition 4.2: Integration of regulated mappings

Let $f : I \to E$ be continuous, if $\{f_n\} \subseteq \Sigma$ converges uniformly to f, then $\{\int_a^b f_n\}$ is Cauchy in E, whose limit we *define* to be $\int_a^b f$ — the integral of f. Furthermore,

1. For any regulated mapping $f : I \to E$,

$$\left|\int_a^b f\right| \leq \int_a^b |f| \leq (b - a)\|f\|_u \tag{93}$$

2. The integral on $\overline{\Sigma}$ (resp. $\overline{\Sigma_I}$, $\overline{\Sigma_J}$) satisfies all of the properties in prop. 2.1.

Proof. Let f be a step mapping on E, we wish to show eq. (93) holds. If f is induced by some n-partition P,

$$\int_a^b f = \sum_{i=\underline{n}} |\mathrm{cell}(p_i)| f(p_i) \leq \sum_{i=\underline{n}} |\mathrm{cell}(p_i)||f(p_i)| = \int_a^b |f| \tag{94}$$

The integral in eq. (94) should be interpreted as a Riemann integral on $\mathbb{R}$, and eq. (95) is immediate:

$$\int_a^b |f| \leq |b - a|\|f\|_u \tag{95}$$

Next, let $\{f_n\}_{n\geq 1}$ be a sequence of step mappings in I which converges uniformly to $f \in \overline{\Sigma}$. Equation (95) tells us the sequence of integrals is uniformly Cauchy, as

$$\left|\int_a^b f_m - \int_a^b f_n\right| \leq |b-a| \|f_m - f_n\|_u \tag{96}$$

Hence $\int_a^b f$ is well defined, eq. (93) and the properties listed in prop. 2.1 follow upon taking limits. ■

Proposition 4.3: Integration and clms

Let E and F be Banach spaces, and $\lambda \in L(E, F)$. For a fixed interval I, denote the space of step mappings from I to E (resp. F) by Σ_E (resp. Σ_F), and regulated mappings similarly. If $\{f_n\} \subseteq \Sigma_E$ converges uniformly to $f \in \overline{\Sigma_E}$, then $\{\lambda f_n\} \to \lambda f$ uniformly in $\overline{\Sigma_F}$. Moreover,

$$\lambda\left(\int_a^b f\right) = \int_a^b \lambda f \tag{97}$$

Proof. The map λ is Lipschitz between E and F, and it descends into a map between the vector spaces Σ_E and Σ_F by composition. If f is a step mapping, and $f|_{\mathrm{cell}(p_i)} = v_i$ for $i = \underline{k}$; the composition of f with λ is again a step mapping $\lambda f|_{\mathrm{cell}\, p_i} = \lambda v_i$.

It is not hard to see $\|\lambda f\|_u =\leq |\lambda| \|f\|_u$, and

- λ is Lipschitz between E and F,
- λ, when viewed as a map between Σ_E and Σ_F, is Lipschitz.

Computing the integral of $\lambda f \in \Sigma_F$,

$$\int_a^b \lambda f = \sum_{i=\underline{k}} |\mathrm{cell}(p_i)| \lambda v_i = \lambda\left(\sum_{i=\underline{k}} |\mathrm{cell}(p_i)| v_i\right) = \lambda \int_a^b f$$

proves eq. (97) for step mappings, and the general case follows from continuity. ■

Fundamental Theorem of Calculus

Proposition 5.1

Let I be a compact interval, and $f : I \to E$ be regulated. Defining $\varphi : I \to E$ as the *integral of f with basepoint a*

$$\varphi(t) = \int_a^t f \tag{98}$$

Then φ is differentiable where f is continuous, and if $t_0 \in I$ is such a point:

$$(D\varphi)(t_0) = f(t_0) \tag{99}$$

Remark 5.1: Identifications

The left hand side in eq. (99) should be thought of as a clm in $L(\mathbb{R}, E)$. We identify the point $f(t_0)$ as the map $t \mapsto t \cdot f(t_0)$.

Proof. Suppose f is continuous at t_0. For all h sufficiently small, set $\varepsilon(h) = \sup_{|t-t_0|\le h,\, t\in I} |f(t) - f(t_0)|$ as the modulus of continuity; where $\varepsilon(h) \to 0$ as $h \to 0$. Applying the well-known technique of estimating the integrand $f(t) = [f(t) - f(t_0)] + f(t_0)$, we have

$$\begin{aligned} \varphi(t_0 + h) - \varphi(t_0) &= \int_{t_0}^{t_0+h} f(t)dt \\ &= f(t_0) \cdot h + \int_{t_0}^{t_0+h} [f(t) - f(t_0)]dt \end{aligned} \tag{100}$$

The last term within eq. (100) is $o(h)$, and the proof is complete. ∎

Mean value theorems

If $\lambda \in L(E, F)$, and $x \in E$, we write $\lambda \dot{x} = x \dot{\lambda}$. If $t \in \mathbb{R}$, and we want to think of x as the map $t \mapsto tx$, we will write $t \cdot x = x \cdot t = tx$ to emphasize the role that x plays. The duality pairing between $L(E, F) \times E \to F$ is bilinear and continuous. For any regulated mapping $\alpha : I \to L(E, F)$,

$$\int_a^b \alpha(t) \cdot x dt = \left(\int_a^b \alpha(t)dt \right) \cdot x \tag{101}$$

Furthermore, if $f \in C^1(I, E)$, we use the notation $f'(t)$ to refer to $Df(t)$; and we identify $f'(t)$ with an element in E; while $Df(t)$ should be thought of as a mapping in $L(\mathbb{R}, E)$.

Lemma 6.1: Constant curves

If $\alpha \in C^1(I, E)$, $\alpha' = 0$, iff α is constant.

Proof. Suppose α' vanishes, and assume for contradiction there exists points $t_0 < t_1$ in I such that $\alpha(t_0) \neq \alpha(t_1)$. Hahn Banach gives us a clf $\lambda \in L(E, \mathbb{R})$ that strictly separates the two points. See prop. 2.1 for a refresher. The ordinary derivative of $\lambda \circ f$ is 0 everywhere which implies $\lambda \circ f$ is constant. The converse is trivial. ∎

Lemma 6.2: FTC 2

Let $f \in C^1(I, E)$, then

$$f(b) - f(a) = \int_a^b f'(t)dt \tag{102}$$

where the integrand in eq. (102) is — rigourously speaking — a map $\mathbb{R} \to L(\mathbb{R}, E)$, but we treat $f'(t) \in E$.

Proof. Throughout this proof, we will treat $f' : \mathbb{R} \to E$. Because f' is continuous everywhere, it is regulated. Define $\varphi(t) = \int_a^t f'(t)dt$, by eq. (99):

$$\varphi'(t) - f'(t) \equiv 0$$

By lem. 6.1, it suffices to show $(\varphi - f)(t) = f(a)$ at any point $t \in [a, b]$. Take $t = a$, and $(\varphi(a) - f(a)) = 0$, so that

$$\varphi(t) = f(t) + f(a)$$

and eq. (102) follows. ■

Remark 6.1: Usefulness of FTC 2

lem. 6.2 is most useful when $[a, b] = [0, 1]$, and the f is a curve interpolating between a C^1 function evaluated two different points, as in prop. 6.1.

Proposition 6.1: MVT 1

Let $U \overset{\circ}{\subseteq} E$ and $x \in U$, $y \in E$. If the line segment $L = \{x + ty, 0 \leq t \leq 1\}$ is also contained in U (draw a picture), then eq. (103) holds.

$$f(x + y) = f(x) + \int_0^1 Df(x + ty)ydt = \left(\int_0^1 Df(x + ty)dt\right) \cdot y \tag{103}$$

Proof. The curve $g(t) = f(x + ty)$ is composed of $f \circ l(t)$, for $l(t) = x + ty$. It has derivative

$$g'(t) = Df(x + ty) \circ l'(t) = Df(x + ty) \circ (y \in L(\mathbb{R}, E))$$

By lem. 6.2, $g(1) - g(0) = \int_0^1 Df(x+ty) \cdot ydt$. Given $g(1) - g(0) = f(x+y) - f(x)$, the proof is complete. ■

Chapter 4: Higher order derivatives

Introduction

We start with the definition of $C^p(E,F)$. Let E and F be Banach Spaces, if $p \geq 1$ is an integer, we define the class C^p to be the set of maps which are p times differentiable, and $D^p f \in C(E,X)$, where

$$X = L(E, L(E, L(E, \cdots F))p \text{ times} \overset{\mathcal{L}}{\rightleftharpoons} L(E^p, F)$$

Sometimes we replace E with an open subset $U \subseteq E$ if necessary, and we write $f \in C(U,F)$ if $D^p \in C(U,X)$. Note, even if $f \in C^1(U,F)$, Df is still a map from U into $L(E,F)$.
We will prove two major results in this section.

- The structure of the derivative $D^p f$, in particular, if $f \in C^p(E,F)$, then $D^p f(x)$ is a *symmetric multilinear map* in p arguments.
- Taylor's Theorem

The second derivative

Proposition 2.1: Product rule in 2 variables

Let E_1, E_2 and F be Banach spaces, if $\omega : E_1 \times E_2 \to F$ is bilinear and continuous, then ω is differentiable, and for every $(x_1, x_2) \in E_1 \times E_2$, $(v_1, v_2) \in E_1 \times E_2$,

$$D\omega(x_1, x_2)(v_1, v_2) = \omega(x_1, v_2) + \omega(v_1, x_2)$$

Furthermore, $D^2\omega(x,y) = D\omega \in L(E^2, F)$, and $D^3\omega = 0$.

Proof. By the definition of ω, using the familiar interpolation method

$$\omega(x_1 + h_2, x_2 + h_2) = \omega(x_1, x_2) + \omega(x_1, h_2) + \omega(h_1, x_2) + \omega(h_1, h_2)$$

by continuity of ω, the last term (which we wish to make $o(h)$):

$$|\omega(h_1, h_2)| \leq \|\omega\| \cdot |(h_1, h_2)|^2$$

so that $\omega(h_1, h_2) = o(h)$, and $D\omega(x_1, x_2)$ exists and is continuous, and is given by the *linear map* $\omega(x_1, \cdot) + \omega(\cdot, x_2)$. The rest of the proof follows, if it is not immediately obvious then read the following note.

Note 2.1

Write $E = E_1 \times E_2$ for convenience. The linear map $A = D\omega(x_1, x_2)$ takes arguments E into F, consider the projections π_1 and π_2, and $v \in E_1 \times E_2$, then

$$A(v) = \omega(x_1, \pi_1 v) + \omega(\pi_2 v, x_2)$$

We can view $A(x) = D\omega(x_1, x_2) \in L(E,F)$. It is clear that A is linear in x, if we fix $v \in E$,

$$A(x + y, v) = \omega(\pi_1(x + y), \pi_2 v) + \omega(\pi_1 v, \pi_2(x + y)) = A(x, v) + A(y, v)$$

and similarly for scalar multiplication. Hence $DA(x) = A \in L(E, L(E,F))$ and $D^2A(x) = D^3\omega = 0$.

Our next result is the following, which states that if $f : U \to F$ where $U \mathring{\subseteq} E$, and $Df, DDf = D^2 f$ exists and are continuous maps from U into $L(E, F)$ and $L(E, L(E, F)$ respectively, then $D^2 f(x)$ is a *symmetric bilinear map*. The proof is non-trivial, and relies on computing the 'Lie Bracket':

$$D^2 f(x)(v, w) - D^2 f(x)(w, v)$$

Which we will prove is equal to 0 for every $x \in U$, and $v, w \in E$.

Proposition 2.2: Second derivative is symmetric

Let $f \in C^2(U, F)$, where $U \mathring{\subseteq} E$ with the possibility that $U = E$. For every point $x \in U$, the *second derivative* $D^2 f(x)$ is bilinear and symmetric.

Proof. Fix $x \in U \hookleftarrow B(r) + x \mathring{\subseteq} U$. We restrict our attention to vectors $v, w \in E$ where $|v|, |w| < r2^{-1}$ for now, so that the

$$\left\{x, x + w, x + v, x + v + w\right\} \subseteq U$$

We will denote the following quantity by Δ

$$\Delta = f(x + w + v) - f(x + w) - f(x + v) + f(x)$$

By rearranging terms, we see that Δ can be approximated in two ways:

- Postponing the discussion about the the domain of y, set $g(y) = f(y + v) - f(y)$ is C^2, and
$$\Delta = g(x + w) - g(x) \tag{104}$$
- Again, for y sufficiently close to x, define $h(y) = f(y + w) - f(y)$, and
$$\Delta = h(x + v) - h(x) \tag{105}$$
- To find the domain for y, an easy argument using the Triangle inequality gives us $g, h \in C^2(B(r2^{-1}) + x, F)$,
- Leaving the computations of h as an exercise, we compute Dg, recall the shift map $y \mapsto y + v$ commutes with D, and
$$Dg(y) = D(\tau_{-w} f)(y) - Df(y) = Df(y + w) - Df(y) \tag{106}$$

Using MVT twice, once on Equation (104) (the line segment $x + tw$, $0 \le t \le 1$ is contained in the domain of g), and another time on Equation (106) (with $y = x + tw$ in the integrand). We obtain:

$$\begin{aligned}
\Delta &= g(x + w) - g(x) \\
&= \int_0^1 Dg(x + tw) \cdot w dt \\
&= \int_0^1 \int_0^1 D^2 f(x + tw + sv) \cdot v ds\, dt \cdot w \\
&= \int_0^1 \int_0^1 D^2 f(x + tw + sv) ds\, dt \cdot v \cdot w
\end{aligned}$$

We can rewrite the application of v then w by $\cdot(v, w)$, and using the approximation $D^2 f(x + tw + sv) \cdot (v, w) = D^2 f(x) \cdot (v, w) + \delta_1(tw, sv)$. Integrating over s, t gives

$$\Delta = D^2 f(x) \cdot (v, w) + \int_0^1 \int_0^1 \delta_1(tw, sv) ds\, dt$$

Note 2.2

The error term δ_1 in the integrand is given by

$$\delta_1(tw, sv) = D^2 f(x + tw + sv)(v, w) - D^2 f(x)(v, w)$$

for v, w sufficiently small and $0 \leq s, t \leq 1$.

A similar argument for h shows that $\Delta = D^2 f(x) \cdot (w, v) + \int_0^1 \int_0^1 \delta_2(tw, sv) ds\, dt$. Combining the two together, the following holds for all v, w sufficiently small:

$$D^2 f(x) \cdot (v, w) - D^2 f(x) \cdot (w, v) = \int_0^1 \int_0^1 \delta_1(tw, sv) ds\, dt - \int_0^1 \int_0^1 \delta_2(tw, sv) ds\, dt \tag{107}$$

To show the right hand side is 0, we will need the following note.

Note 2.3

We wish to show the RHS of Equation (107) is 0. We begin by controlling the RHS and show that it is super-bilinear; meaning it shrinks after than the product $|v||w|$. Then, we will prove a lemma which will show the only bilinear map that satisfies this property is the 0 map.

- For $j = 1, 2$, relabel $\delta = \delta_j$ for convenience. We can use the L^1 inequality, to obtain the estimate

$$\left| \int_0^1 \int_0^1 \delta(tw, sv) ds\, dt \right| \leq \int_0^1 \int_0^1 |\delta(tw, sv)| ds\, dt \tag{108}$$

- $\delta(tw, sv)$ is controlled by $|D^2 f(x + tw + sv) - D^2 f(x)||v||w|$. Take $y = tw + sv$, then $|y| \leq |tw| + |sv|$. Hence,

$$|\delta_j| \leq \left| D^2 f(x + tw + sv) - D^2 f(x) \right| |v||w| \tag{109}$$

- Let A denote the span of w, v for scalars $s, t \in [0, 1]$. In symbols,

$$A = \left\{ tw + sv,\ s, t \in [0, 1] \right\}$$

  A is clearly compact, and the continuity of $D^2 f$ means

$$R(v, w, \delta) = \sup_{y \in A} \left| D^2 f(x + y) - D^2 f(x) \right| \quad \text{is finite,} \quad \text{and} \quad \lim_{(v,w) \to 0} R(v, w, \delta) = 0 \tag{110}$$

  See remark 2.1 for a generalization of this argument.

- Relabel $R(v, w)$ to be the maximum across $R(v, w, \delta_1)$ and $R(v, w, \delta_2)$.

- Combining Equations (108) to (110), we obtain the following bound on Equation (107)

$$
\begin{aligned}
\left|D^2 f(x)\cdot(v,w) - D^2 f(x)\cdot(w,v)\right| &\leq \left|\iint \delta_1(tw,sv)ds\, dt - \iint \delta_2(tw,sv)ds\, dt\right| \\
&\leq \iint |\delta_1| ds\, dt + \iint |\delta_2| ds\, dt \\
&\leq |v||w|R(v,w) \qquad (111)
\end{aligned}
$$

The following Lemma gives a useful criterion to check when a multilinear map is identically 0.

Lemma 2.1

Let E be a Banach space, and $k \geq 1$ be an integer. If $\lambda \in L(E^k, F)$ and there exists another map $\theta : E^k \to F$ (defined perhaps on an open neighbourhood of the origin), such that

$$|\lambda(u_{\underline{k}})| \leq |\theta(u_{\underline{k}})| \cdot \prod |u_{\underline{k}}|$$

for all $(u_{\underline{k}})$ sufficiently small. And $\lim_{(u_{\underline{k}})\to 0} \theta(u_{\underline{k}}) = 0$, then, $\lambda = 0$.

Proof. Fix arbitrary $(u_{\underline{k}}) \in E^k$, for $s > 0$ sufficiently small, the left hand side of the equation reads

$$|s|^k |\lambda(u_{\underline{k}})| \leq |\theta(su_{\underline{k}})| \cdot |s|^k \prod |u_{\underline{k}}|$$

The rest of the argument is Archimedean: divide by $|s|^k$ and send $s \to 0$ (while paying attention to the term with θ): perhaps after relabelling $v_s = su_{\underline{k}}$ for sufficiently small s, then $|\theta(v_s)| \to 0$ as $s \to 0$. ∎

∎

Remark 2.1: Compact linear combinations

Generalization of the "compact linear combination" argument used above. Let $(t_{\underline{k}}) \subseteq \mathbb{C}^k$ or $\mathbb{R}^k$, and vectors $v_{\underline{k}} \in E$. Suppose further $(t_{\underline{k}}) \subseteq A$ is compact in $\mathbb{C}^k$ or $\mathbb{R}^k$. It is clear that if $y = t_i v^i \in E$, where the summation convention is in effect. Then,

$$|y| \lesssim_A \left|(v^{\underline{k}})\right|_{E^k}$$

Now, fix a continuous function $f \in C(E,F)$, we can approximate the maximum error over all such y

$$\sup_{y\in B} |f(x+y) - f(x)| < \varepsilon \quad \forall |y| \lesssim_A |(v^{\underline{k}})| < \delta$$

where

$$B = \left\{ \sum t_i v^i,\ (t_{\underline{k}}) \subseteq A,\ (v^{\underline{k}}) \in E^k \right\}$$

The p-th derivatives

If f is p times differentiable, and $f, Df, D^2 f, \ldots, D^p f$ are all continuous, then we say $f \in C^p(E,F)$ (replacing E with an open subset of E if necessary).

Proposition 3.1

If $f \in C^p(E,F)$, then $D^p f(x)$ is symmetric for every $x \in E$. (Replace E with an open set if necessary).

Proof. The main proof proceeds as follows. We will use induction on p, with $p = 2$ serving as the base case. Our induction hypothesis is that for every $f \in C^{p-1}(E,F)$, for every permutation $\beta \in S_{p-1}$, at every point $x \in E$, for every possible choice of $p-1$ vectors $(v_2, \ldots, v_p) = (v_{1+\underline{p-1}})$,

$$D^{p-1} f(x)(v_{1+\underline{p-1}}) = D^{p-1} f(x)(v_{1+\beta(\underline{p-1})})$$

To prove the assertion for p, it suffices to show $D^p f(x)(v_{\underline{p}})$ is invariant under transpositions of indices; since the transpositions generate S_p. Furthermore, the transpositions in S_p are generated by

- the transposition $(1,2,\ldots) \mapsto (2,1,\ldots)$ where the omitted indices are held fixed, and
- the transpositions which leave the first index fixed:

  $$(1, 1+\underline{p-1}) \mapsto (1, 1+\beta(\underline{p-1}))$$

  where $\beta \in S_{p-1}$

so it suffices to prove invariance under those two types of transpositions. Let $g = D^{p-2} f$, so $g \in C^2(E, L(E^{p-2}, F))$. Because the application of vectors (currying) on a multilinear map $A \in L(E^p, F)$ is associative, illustrated as follows:

$$(A \cdot v_1) \cdot v_2 = A \cdot (v_1, v_2) = A(v_1, v_2, \cdot) \in L(E^{p-2}, F)$$

Then, let $\lambda : L(E^{p-2}, F) \to F$ be the evaluation map at $(v_3, \ldots, v_p) = (v_{2+\underline{p-2}})$. Using the base case on $D^{p-2} f = g \in C^2(E, L(E^{p-2}, F))$,

$$(D^2 g)(x)(v_1, v_2) = (D^2 g)(x)(v_2, v_1) \implies \lambda\Big((D^2 g)(x)(v_1, v_2)\Big) = \lambda\Big((D^2 g)(x)(v_2, v_1)\Big)$$

But λ is the map that *applies* the rest of the vectors, and

$$(D^2 g)(x)(v_1, v_2) \cdot (v_{2+\underline{p-2}}) = (D^2 g)(x)(v_2, v_1) \cdot (v_{2+\underline{p-2}}) \tag{112}$$

Since D commutes with continuous linear maps (and λ is continuous because $(v_{2+\underline{p-2}})$ is fixed),

$$\lambda(D^2(D^{p-2} f)) = D(\lambda(D(D^{p-2} f)) = D(D\lambda \circ D^{p-2} f) = D^2(\lambda \circ D^{p-2} f) \tag{113}$$

Substituting Equation (112) for the rightmost hand side of Equation (113) gives the result.

Note 3.1

There are no magic 'identifications' being made here. To be perfectly clear, for each $x \in E$, $g(x)$ is an element in $L(E^{p-2}, F)$, and $(D^2 g)(x) \in L(E^2, L(E^{p-2}, F))$. Evaluating g at a point x gives a bilinear map that takes values in the Banach space $L(E^{p-2}, F)$.

For the second case, beginning from the induction hypothesis. If θ is a p-permutation that leaves the first coordinate unchanged, then there exists a unique $p-1$-permutation $\beta \in S_{p-1}$ such that

$$\begin{aligned}(\,\theta(\underline{p})\,) &= \big(1, \theta(1+\underline{p-1})\big) \\ &= \big(1, 1+\beta(\underline{p-1})\big)\end{aligned} \tag{114}$$

Using a similar argument as the first case, set $g = D^{p-1}f$ and $\lambda, \lambda' \in L(E^{p-1}, F)$ to be the evaluation maps of $(v_1, v_{1+\underline{p-1}}) = (v_{\underline{p}})$ and $(v_1, v_{1+\beta(\underline{p-1})})$ respectively. Rehearsing the same proof as before:

$$\begin{aligned}(D^p f)(x)(v_{\underline{p}}) &= D\big(\lambda D^{p-1} f\big)(x)(v_1) && \text{Equation (113)} \\ &= D\big(\lambda' D^{p-1} f\big)(x)(v_1) && \text{ind. hyp.} \\ &= (D^p f)(x)(v_{\theta(\underline{p})}) && \text{Equation (113)}\end{aligned}$$

This proves the induction step, and the proof is complete. ∎

Before stating and proving Taylor's Theorem, an important remark on the 'postcomposition' of linear maps. Summarized in the following note.

Note 3.2

Let $f \in C^p(E, F)$, and $\lambda \in L^p(F, G)$. λ induces a map between $L(E^p, F)$ and $L(E^p, G)$ by post-composing any multi-linear map $A \in L(E^p, F)$ by λ. Denoting this map by λ_*,

$$\lambda_* : L(E^p, F) \to L(E^p, G)$$

It is clear λ_* is linear and continuous. And its action on A, evaluated at $(v_{\underline{p}}) \in E^p$ is given by

$$\lambda_*(A) \in L(E^p, G) \quad \big(\lambda_*(A)\big)(v_{\underline{p}}) = \lambda\big(A(v_{\underline{p}})\big) = (\lambda \circ A)(v_{\underline{p}})$$

Now, recall that for $p = 1$

$$\big[D(\lambda \circ f)\big](x) = \lambda\big[(Df)(x)\big]$$

To simplify the notation, we want to 'move' the evaluation x outside of the brackets, and somehow write $x \mapsto \lambda\big[(Df)(x)\big]$ as one map between E and $L(E, G)$. We further *identify* λ as this map, so that

$$\big[D(\lambda \circ f)\big](x) = \lambda = \big(\lambda \circ Df\big)(x)$$

Dropping the x from the expression, for $p \geq 2$ *assuming a similar formula holds*, then we write $\big[D^p(\lambda \circ f)\big] = \lambda_* \circ D^p f$. We make a final identification, of $\lambda = \lambda_*$ (thereby conflating the two different maps, the first is a map from E to F, the second is a map from $L(E^p, F)$ into $L(E^p, G)$).

Proposition 3.2

If $p \geq 2$, $f \in C^p(E, F)$, $\lambda \in L(F, G)$, then

$$D^p(\lambda \circ f) = \lambda \circ D^p f$$

Where we have identified λ as the same map that acts on $L(E^p, F)$ to produce another map in $L(E^p, G)$, and suppressed the point x.

Proof. Use induction on p. ■

Proposition 3.3: Taylor's Formula

Let $f \in C^p(U, F)$, where $U \mathring{\subseteq} E$. For $x \in U$ and $y \in E$ such that $L = \{x + ty,\, 0 \leq t \leq 1\}$ is in contained in U, then

$$f(x+y) = f(x) + \left(\sum_{i=\underline{p-1}} \frac{D^i f(x) \cdot (y^{(i)})}{(p-1)!} \right) + R_p \tag{115}$$

where $\cdot(y^{(i)})$ denotes the application of y, consecutively for i times. The remainder R_p is given by eq. (116)

$$R_p = \int_0^1 \frac{(1-t)^{p-1}}{(p-1)!} D^p f(x+ty) dt \cdot (y^{(p)}) \tag{116}$$

Furthermore, we include the pth term in the series using eq. (117)

$$f(x+y) = f(x) + \sum_{i=\underline{p}} \frac{D^i f(x) \cdot (y^{(i)})}{i!} + \theta(y) \tag{117}$$

where θ is defined for small y, and $o(|y|^p)$.

$$|\theta(y)| \leq \sup_{0 \leq t \leq 1} \frac{|D^p f(x+ty) - D^p f(x)|}{p!} |y|^p \tag{118}$$

Chapter A: Review of Topology

Set Operations

This section is meant for reference.

Proposition 1.1: Direct and Inverse Images of Maps

Let $f : \mathbf{X} \to \mathbf{Y}$, where $\mathbf{X}$ and $\mathbf{Y}$ are sets. If $A \subseteq \mathbf{X}$, $B \subseteq \mathbf{Y}$, and $\{E_\alpha\}$ is an indexed collection of subsets of $\mathbf{X}$, $\{G_\beta\}$ is an indexed collection of subsets of $\mathbf{Y}$, then

Direct images

$$f\left(\bigcap E_\alpha\right) \subseteq \bigcap f(E_\alpha) \quad \text{equality if injective} \tag{119}$$

$$f\left(\bigcup E_\alpha\right) = \bigcup f(E_\alpha) \tag{120}$$

Estimates

$$f\left(f^{-1}(B)\right) \subseteq B \quad \text{equality if surjective} \tag{121}$$

$$A \subseteq f^{-1}(f(A)) \quad \text{equality if injective} \tag{122}$$

Inverse images

$$f^{-1}\left(\bigcup G_\beta\right) = \bigcup f^{-1}(G_\beta) \tag{123}$$

$$f^{-1}\left(\bigcap G_\beta\right) = \bigcap f^{-1}(G_\beta) \tag{124}$$

$$f^{-1}(B^c) = \left(f^{-1}(B)\right)^c \tag{125}$$

Proposition 1.2: Composition of Maps

Let $h = g \circ f$, we assume this composition is well defined.

- If h is a surjection, then g is a surjection,
- If h is an injection, then f is an injection.

Proof. Take the contrapositive. ■

Proposition 1.3: Left and Right inverses

Let $F : \mathbf{X} \to \mathbf{Y}$,

- F is surjective if and only if there exists right inverse $G : \mathbf{Y} \to \mathbf{X}$,

$$F \circ G = \mathrm{id}_{\mathbf{Y}}$$

  if $A \subseteq \mathbf{X}$,

$$G^{-1}(A) \subseteq F(A)$$

- F is injective if and only if there exists a left inverse $H : F(\mathbf{X}) \to \mathbf{X}$

$$H \circ F = \mathrm{id}_{\mathbf{X}}$$

and if $B \subseteq \mathbf{Y}$,

$$F^{-1}(B) \subseteq H(B)$$

Topological Spaces

This section will roughly follow Munkres text on General Topology, in particular we hope to cover Chapters 2, 3, 4 and 9. The rest of the Chapters should be covered proper by the subsequent section.

Definition 2.1: Topology

Let $\mathbf{X}$ be a non-empty set. A topology $\mathcal{T}$ on $\mathbf{X}$, sometimes denoted by $\mathcal{T}_{\mathbf{X}}$ is a family of subsets of $\mathbf{X}$,

- $\{\varnothing, \mathbf{X}\} \subseteq \mathcal{T}$,
- If U_1 and U_2 are elements of $\mathcal{T}$, so is their intersection.
- If $\{U_\alpha\}$ is an arbitrary family of sets in $\mathcal{T}$, their union is also contained in $\mathcal{T}$ as an element.

We call the elements of $\mathcal{T}$ open sets. The complements of elements in $\mathcal{T}$ are closed sets.

Basis of a Topology

Definition 3.1: Basis of a topology

A basis $\mathbb{B}$ is a family of subsets of $\mathbf{X}$, that satisfies:

- Every $x \in \mathbf{X}$ belongs (as an element) in some $V \in \mathbb{B}$.
- If B_1 and B_2 are basis elements, such that their intersection is non-empty. Then every $x \in B_1 \cap B_2$ induces a $B_3 \in \mathbb{B}$ with
$$x \in B_3 \subseteq B_1 \cap B_2$$

This roughly means a basis is 'finitely' fine at every point in x.

If $\mathbb{B}$ is a basis, it 'generates' a topology $\mathcal{T}$ through

$$\mathcal{T} = \left\{U \subseteq \mathbf{X},\ \forall x \in U,\ x \in B \subseteq U \text{ for some } B \in \mathbb{B}\right\} \tag{126}$$

Notice this is equivalent to $\mathcal{T}$ is the collection of all unions of basis elements in $\mathbb{B}$.

Proposition 3.1

Let $\mathbb{B}$ be a basis as defined in Definition 3.1, then $\mathcal{T}$ as defined in Equation (126) is a valid topology on $\mathbf{X}$. And every member of $\mathcal{T}$ is and is precisely the union of elements in $\mathbb{B}$.

Proof. Every point in $\mathbf{X}$ belongs in some basis element, so $\mathbf{X} \in \mathcal{T}$, so does $\varnothing$. Next, if U_1 and U_2 are in $\mathcal{T}$, then

$$\begin{cases} x \in U_1 \leadsto x \in B_1 \subseteq U_1 \\ x \in U_2 \leadsto x \in B_2 \subseteq U_2 \end{cases} \implies x \in B_3 \subseteq B_1 \cap B_2 \subseteq U_1 \cap U_2$$

for some $B_3 \in \mathbb{B}$, so $\mathcal{T}$ is closed under finite intersections (perhaps after a standard induction argument).

If $\{U_\alpha\} \subseteq \mathcal{T}$, and x belongs in the union of all U_α, then $x \in B_\alpha \subseteq U_\alpha$, which is a subset of the entire union. So the union over U_α is again contained in $\mathcal{T}$, and $\mathcal{T}$ is a topology on $\mathbf{X}$.

It is worth noting that $\mathbb{B} \subseteq \mathcal{T}$. Finally, if $U \in \mathcal{T}$,

$$U = \bigcup_{x \in U} B_x$$

where B_x is the basis element taken to satisfy $x \in B_x \subseteq U$. Every point in U is included in some B_x, and hence is included in the union. For the reverse inclusion, notice the union of subsets of U is again a subset of U.

Now, if $E \subseteq \mathbf{X}$ is the union of basis elements in $\mathbb{B}$, if E is non-empty, then every point $x \in E$ belongs in some B_x. Recycling the previous argument, and we see that E is open in $\mathcal{T}$. If E is empty, we define the 'union' of no sets as the empty set. So $\mathcal{T}$ is precisely the collection of all unions of basis elements $\mathbb{B}$. ■

We are now in a position to compare the relative 'fineness' of topologies.

Definition 3.2: Fineness of topologies

If $\mathcal{T}'$ and $\mathcal{T}$ are both topologies on some non-empty set $\mathbf{X}$. We say $\mathcal{T}'$ is finer than $\mathcal{T}$, or $\mathcal{T}$ is coarser than $\mathcal{T}'$ if

$$\mathcal{T}' \supseteq \mathcal{T}$$

Proposition 3.2

If $\mathbb{B}$ and $\mathbb{B}'$ are bases for $\mathcal{T}'$ and $\mathcal{T}$, the following are equivalent:

- $\mathcal{T}'$ is finer than $\mathcal{T}$,
- If B is an arbitrary basis element in $\mathbb{B}$, then every point $x \in B$ induces a basis element in $\mathbb{B}'$ with

$$x \in \mathbb{B}' \subseteq \mathbb{B}$$

Proof. Suppose $\mathcal{T}'$ is finer than $\mathcal{T}$. Notice $\mathbb{B} \subseteq \mathcal{T}'$ as well. By Equation (126), each $x \in B$ induces a $B' \in \mathbb{B}'$

$$x \in B' \subseteq B$$

Conversely, fix any open set $U \in \mathcal{T}$, and for each $x \in U$,

$$x \in B' \subseteq B \subseteq U$$

Applying Definition 3.1 tells us U is open in $\mathcal{T}'$. ■

The last of the big three 'generating' definitions for topologies will be the sub-basis. It simply means the first condition (but not necessarily) the second, is satisfied in Definition 3.1

Definition 3.3: Sub-basis of a topology

A sub-basis $S \in \mathbb{P}(\mathbf{X})$ is a family of subsets of $\mathbf{X}$ that satisfies one property. Any point x in $\mathbf{X}$ belongs to at least one member of S.

A sub-basis can be upgraded to a basis by collecting all of its finite intersections.

Proposition 3.3

Let S be a sub-basis of $\mathbf{X}$, then the collection of all finite intersections of S forms a basis $\mathbb{B}$ of $\mathbf{X}$.

Proof. Every point in $\mathbf{X}$ lies in some element of S, hence in some element of $\mathbb{B}$. The second basis property is immediate, since $\mathbb{B}$ is closed under finite intersections. ■

Product Topology

We will start with products of a finite collection of topological spaces.

Definition 4.1: Finite Product of Topological Spaces

Let $(\mathbf{X}, \mathcal{T}_\mathbf{X})$ and $(\mathbf{Y}, \mathcal{T}_\mathbf{Y})$ be topological spaces. The product topology (denoted by $\mathcal{T}_{\mathbf{X}\times\mathbf{Y}}$) on $X \times Y$ is defined as the topology generated by the basis

$$\mathbb{B}_{\mathbf{X}\times\mathbf{Y}} = \Big\{ U \times V,\ (U, V) \in \mathcal{T}_\mathbf{X} \times \mathcal{T}_\mathbf{Y} \Big\} \tag{127}$$

Since bases are easier to describe than topologies, we have the following statement concerning the basis of the product topology.

Proposition 4.1

If $\mathbb{B}_\mathbf{X}$ and $\mathbb{B}_\mathbf{Y}$ are bases for $\mathcal{T}_\mathbf{X}$ and $\mathcal{T}_\mathbf{Y}$, then the product topology (as described in Definition 4.1) is also generated by

$$\mathcal{M} = \Big\{ U \times V,\ (U, V) \in \mathbb{B}_\mathbf{X} \times \mathbb{B}_\mathbf{Y} \Big\} \tag{128}$$

Proof. We will introduce (and use) the technique of 'double inclusion' by proving that the topologies generated are both finer than the other. Let us denote the topology generated by $\mathcal{M}$ in Equation (128) by $\mathcal{T}_\mathcal{M}$.

Since $\mathbb{B}_\mathbf{X} \times \mathbb{B}_\mathbf{Y} \subseteq \mathcal{T}_\mathbf{X} \times \mathcal{T}_\mathbf{Y}$, if $U \times V \in \mathcal{M}$ as in Equation (128), then we can pick the same 'open rectangle' again. We trivially have

$$x \in \underbrace{U \times V}_{\text{member of } \mathcal{T}_\mathbf{X}\times\mathcal{T}_\mathbf{Y}} \subseteq U \times V$$

and by Proposition 3.2, $\mathcal{T}_{\mathbf{X}\times\mathbf{Y}}$ is finer than $\mathcal{T}_\mathcal{M}$.

Fix any set $U \times V \in \mathbb{B}_{\mathbf{X}\times\mathbf{Y}}$, and if $(p, q) \in U \times V$, each coordinate induces basis elements from $\mathbb{B}_\mathbf{X}$ and By, more precisely:

$$\begin{cases} p \in U \implies p \in \text{Basis element of } \mathbb{B}_\mathbf{X} \subseteq U \\ q \in V \implies q \in \text{Basis element of } \mathbb{B}_\mathbf{Y} \subseteq V \end{cases} \implies (p, q) \in \underbrace{\quad}_{\text{in } \mathbb{B}_\mathbf{X}} \times \underbrace{\quad}_{\text{in } \mathbb{B}_\mathbf{Y}} \subseteq U \times V$$

by Proposition 3.2, $\mathcal{T}_{\mathcal{M}}$ is finer than $\mathcal{T}_{\mathbf{X}\times\mathbf{Y}}$ and $\mathcal{T}_{\mathbf{X}\times\mathbf{Y}} = \mathcal{T}_{\mathcal{M}}$. ■

Continuity

Definition 5.1: Continuous maps $C(\mathbf{X},\ \mathbf{Y})$

Let f be a map from $\mathbf{X}$ to $\mathbf{Y}$. It is called *continuous* if $f^{-1}(U)$ is open in $\mathbf{X}$ for every open set U in $\mathbf{Y}$. We denote the set of continuous functions from $\mathbf{X}$ to $\mathbf{Y}$ by $C(\mathbf{X},\ \mathbf{Y})$.

Proposition 5.1: Continuity preserving operations

The composition of continuous functions is again continuous, and the product of continuous functions is again continuous.

Proof. Suppose $f \in C(\mathbf{X},\ \mathbf{Y})$ and $g \in C(\mathbf{Y},\ \mathbf{Z})$. Fix an open set $U \subseteq \mathbf{Z}$. Then $g^{-1}(U)$ is open in $\mathbf{Y}$, hence

$$f^{-1}(g^{-1}(U)) = (g \circ f)^{-1}(U) \quad \text{is open in } \mathbf{X}$$

Next, let $\{f_\alpha\}_{\alpha\in A}$ be a collection of continuous functions, where each $f_\alpha \in C(\mathbf{X}_\alpha,\ \mathbf{Y}_\alpha)$. Let us write

$$\mathbf{X} \triangleq \prod \mathbf{X}_\alpha \quad \text{and} \quad \mathbf{Y} \triangleq \prod \mathbf{Y}_\alpha$$

and the projection maps:

$$\pi_\alpha^{\mathbf{X}} : \mathbf{X} \to \mathbf{X}_\alpha, \quad \text{For every } x \in \mathbf{X},\ \pi_\alpha^{\mathbf{X}}(x) = x(\alpha) \in \mathbf{X}_\alpha$$

similarly for $\pi_\alpha^{\mathbf{Y}} : \mathbf{Y} \to \mathbf{Y}_\alpha$. The product function $F = \prod f_\alpha$, is defined through its behaviour 'on each coordinate'

$$\pi_\alpha^{\mathbf{Y}} \circ F = f_\alpha \circ \pi_\alpha^{\mathbf{X}} \tag{129}$$

A function $F : \mathbf{X} \to \prod \mathbf{Y}_\alpha$ is continuous iff $\pi_\alpha^{\mathbf{Y}} \circ F$ is continuous for each $\alpha \in A$. By Equation (129), it is clear that each $\pi_\alpha^{\mathbf{Y}} \circ F$ is continuous, since the right member is the composition of two continuous functions, which is again continuous by the first part of this proof, therefore F is continuous. ■

Definition 5.2: Open/Closed Maps

Let $f : \mathbf{X} \to \mathbf{Y}$ be a map (not necessarily continuous), it is called *open* (resp. *closed*) if for every open (resp. closed) set $E \subseteq \mathbf{X}$, $f(E)$ is open (resp. closed).

Clearly, the composition of open (resp. closed) maps is again open (resp. closed).

Quotient Topology

Product Topology

The Cartesian Product of an arbitrary family of topological spaces, if equipped with the product topology, preserves a lot of the structure. If $\{X_\alpha\}_{\alpha\in A}$ is a family of topological spaces which are ______, then $\prod X_\alpha$ is ______. Replace ______ with:

1. Hausdorff, (Folland)
2. Regular,
3. Connected, (Munkres chp23, exercise 10)
4. First countable, if A is countable,
5. Second countable, if A is countable,
6. Compact (Tynchonoff's Theorem, Folland)

Proposition 7.1: Product of Closed sets again Closed

The product of closed sets is again closed. More concretely, if $\{E_\alpha\}_{\alpha \in A}$ is a family of sets such that $E_\alpha \subseteq \mathbf{X}_\alpha$, then

$$\prod \overline{E_\alpha} = \overline{\prod E_\alpha}$$

Connectedness

Definition 8.1: Connectedness

A topological space $\mathbf{X}$ is connected if U and V are disjoint open subsets whose union is $\mathbf{X}$, then at least one of U or V is empty.

See Folland Exercise 4.10 for more properties.

Definition 8.2: Path-connectedness

A topological space $\mathbf{X}$ is path-connected if for any two pair of points x, $y \in \mathbf{X}$. There exists a continuous function $f : [a, b] \to \mathbf{X}$, with $f(a) = x$ and $f(b) = y$.

Definition 8.3: Connected component

The connected components of $\mathbf{X}$ is the family of equivalence classes on $\mathbf{X}$, where $x \sim y$ if there is a connected subspace of $\mathbf{X}$ that contains both of them.

Proposition 8.1

Continuous functions map connected spaces to connected spaces (in the subspace topology).

Proof. Let $\mathbf{X}$ and $\mathbf{Y}$ be topological spaces and $f : \mathbf{X} \to \mathbf{Y}$ be continuous. If $f(\mathbf{X})$ is disconnected, then we can find U and V, open and disjoint in $\mathcal{T}_{f(\mathbf{X})}$ such that

$$U \cup V = f(\mathbf{X}) \implies f^{-1}(U) \cup f^{-1}(V) = \mathbf{X}$$

where $f^{-1}(f(\mathbf{X})) = \mathbf{X}$. Both $f^{-1}(U)$ and $f^{-1}(V)$ are open, non-empty, and are pairwise disjoint. So $\mathbf{X}$ is separated. ■

Proposition 8.2

Let $(\mathbf{X}_\alpha, \mathcal{T}_\alpha)$ be a family of connected topological spaces indexed by $\alpha \in A$. Then $\prod_{\alpha\in A} \mathbf{X}_\alpha$ is disconnected in the product topology.

Proof. We will attempt the contrapositive. Suppose $\prod_{\alpha\in A} \mathbf{X}_\alpha$ is disconnected, then ■

Interiors and closures

Definition 9.1: Interior of a set

A^o is defined to be the largest open subset of A,

$$A^o = \bigcup_{\substack{U \text{ open}, \\ U\subseteq A}} U$$

Corollary 9.1

The union of subsets of A is again a subset of A, therefore Corollary 9.1 implies $A^o \subseteq A$ for any $A \subseteq \mathbf{X}$.

Definition 9.2: Closure of a set

and $\overline{A}$ is the smallest closed superset of A,

$$\overline{A} = \bigcap_{\substack{K \text{ closed}, \\ A\subseteq K}} K$$

Proposition 9.1

The complement of the closure is the interior of the complement, or equivalently: $(\overline{A})^c = A^{co}$

Proof. Taking complements, and the substitution $U = K^c$ reads

$$\begin{aligned}(\overline{A})^c &= \left[\bigcap_{\substack{K \text{ closed}, \\ A\subseteq K}} K\right]^c \\ &= \bigcup_{\substack{K \text{ closed}, \\ K^c\subseteq A^c}} K^c \\ &= \bigcap_{\substack{U \text{ open}, \\ U\subseteq A^c}} U \\ &= A^{co}\end{aligned}$$

■

Remark 9.1

Personally, I remember this as pushing the complement inside and flipping the bar to a c!

Neighbourhoods

The concept of a neighbourhood allows us to characterize the interior of a set 'locally'.

Definition 10.1: Neighbourhood (not necessarily open)

A neighbourhood of $x \in \mathbf{X}$ is a set $U \subseteq \mathbf{X}$ where $x \in U^o$. The set of neighbourhoods for a point $x \in \mathbf{X}$ will sometimes be denoted by $\mathcal{N}(()x)$.

Proposition 10.1: Characterization of the interior

If $W = \left\{ x \in \mathbf{X}, \text{ there exists a neighbourhood } U \text{ of } x, U \subseteq A \right\}$, then $W = A^o$.

Proof. If $x \in A^o$, then A is a neighbourhood of x, and $A \subseteq A$, so $x \in W$. Conversely, if x is a member of W, it has a neighbourhood $U \subseteq A$ (not necessarily open). By monotonicity of the interior,

$$x \in U^o \subseteq A^o$$

and $x \in A^o$. ∎

It is easy to see that A is open $\iff A^o = A \iff A$ is a neighbourhood of itself.

- The first equivalence follows from:
$$E \subseteq \mathbf{X} \implies E^o \subseteq E$$
and if A is an open set, it is an open subset of itself, by Corollary 9.1 $A \subseteq A^o$. If $A^o = A$, then it suffices to show that A^o is open. Which it is, since it is the arbitrary union of open sets.
- To prove the second equivalence: suppose $A^o = A$, then each $x \in A$ has a neighbourhood contained (as a subset) in A, namely A itself. (This statement is hard to parse, the reader is encouraged to really work through this and be honest).
$$x \in A^o \subseteq A \implies A \subseteq A^o$$
so A is a neighbourhood of itself. Conversely, if $A \subseteq A^o$, then $A = A^o$, since the reverse inclusion follows immediately from Corollary 9.1.

Adherent points

Similar to the neighbourhood, the concept of an adherent point of a set allows us to speak of the closure in more concrete terms. The following definition is key in understanding the relationship between the closure, interior, and the boundary.

Definition 11.1: Adherent point of a set

Let $A \subseteq \mathbf{X}$, $x \in \mathbf{X}$ is an adherent point of A if every neighbourhood U of x intersects A. In symbols,

$$U \cap A \neq \varnothing, \quad \forall U \in \mathcal{N}(x)$$

Proposition 11.1: Characterization of the closure

Let $A \subseteq X$, and let W be the set of adherent points of A, then $\overline{A} = W$

Proof. Suppose $x \notin W$, then there exists a neighbourhood U of x where

$$U \cap A = \varnothing \iff U \subseteq A^c$$

this is exactly the definition of the interior of A^c, so $x \in A^{co}$ and recall (from Proposition 9.1) that $(\overline{A})^c = A^{co}$, so $x \notin \overline{A}$. For the reverse inclusion, read the proof backwards, by flipping $\forall \to \exists$ within the set, and we see that

$$W^c = A^{co} = (\overline{A})^c$$

■

Dense and nowhere dense subsets

Definition 12.1: Dense subset

A subset of a topological space $E \subseteq \mathbf{X}$ is dense if $\overline{E} = \mathbf{X}$.

Definition 12.2: Nowhere dense subset

A subset of a topological space $E \subseteq \mathbf{X}$ is nowhere dense if $\overline{E}^o = \varnothing$.
This means E is dense in none of the (non-trivial) open subspaces of $\mathbf{X}$.

Proposition 12.1

E is dense in $\mathbf{X}$ iff for every non-empty, open set $U \subseteq \mathbf{X}$, $U \cap E \neq \varnothing$.

Proof of Proposition 12.1. Suppose E is dense, then $\overline{E} = \mathbf{X}$. Every point of $\mathbf{X}$ is an adherent point of E. Let $U \subseteq \mathbf{X}$ be a non-empty open set. If $x \in U$ then U is a neighbourhood of x, thus U intersects E. Conversely, suppose every non-empty open set U intersects E. Fix any point $x \in \mathbf{X}$, and any neighbourhood U of x. U has a non-empty interior (because it must contain x). But U^o is a non-empty open set, therefore $\varnothing \neq U^o \cap E \subseteq U \cap E$ ■

Proposition 12.2

Let $f : \mathbf{X} \to \mathbf{X}$ be a homeomorphism. E is nowhere dense iff $f(E)$ is nowhere dense.

Proof. Since f^{-1} is a homeomorphism, suppose $\overline{f^{-1}(E)}^o \neq \varnothing$, there exists a non-empty, open subset $U \subseteq \mathbf{X}$ with

$$\overline{f^{-1}(E)} \cap U = U$$

The direct image yields

$$f\Big((\overline{f^{-1}(E)}) \cap U\Big) = f(U)$$

since f is a bijection (injectivity is necessary here), it commutes with intersections.

$$f(\overline{f^{-1}(E)}) \cap f(U) = f\Big((\overline{f^{-1}(E)}) \cap U\Big) = f(U) \tag{130}$$

and f is continuous, so $f(\overline{A}) \subseteq \overline{f(A)}$ for any $A \subseteq \mathbf{X}$. For the reverse inclusion, f is a closed map, so $f(\overline{A})$ is a closed superset of $f(A)$ so

$$f(\overline{A}) = \overline{f(A)}$$

Take $A = f^{-1}(E)$, and $f(\overline{f^{-1}(E)}) = \overline{f(f^{-1}(E))} = \overline{E}$. From eq. (130), we see that

$$\overline{E} \cap f(U) = f(U)$$

$f(U)$ is a non-empty open subset of $\mathbf{X}$, since f is an open map, so E is not no-where dense. The reverse implication can be proven by replacing f with f^{-1}. ■

Urysohn's Lemma

Proposition 13.1: Folland Theorem 4.14

Suppose that A and B are disjoint closed subsets of the normal space X, and let $\Delta = \{k2^{-n} : n \geq 1 \text{ and } 0 < k < 2^n\}$ be the set of dyadic rationals in $(0,1)$. There is a family $\{U_r : r \in \Delta\}$ of open sets such that

1. $A \subseteq U_r \subseteq B^c$ for every $r \in \Delta$,
2. $\overline{U_r} \subseteq U_s$ for $r < s$, and
3. For every $r < s$, $\overline{U}_r \subseteq U_s$

Proof. The goal of this proof is to show that for every $r \in \Delta$, there exists a open U_r that satisfies the above. As usual for these types of proofs we will proceed by induction. We can divide the problem by 'layers' (as I will hereinafter explain).

Let us suppose that for some $N \geq 1$ that all previous U_r in previous layers have been constructed properly, meaning if $r = k/2^n$, then for every $1 \leq n \leq N-1$, we have

$$r = \frac{k}{2^n},\ 1 \leq n \leq N-1,\ 1 \leq k \leq 2^{n-1}$$

And by 'constructed properly', we mean that for each U_r,

- $A \subseteq U_r \subseteq B^c$ and
- $U_r \in \mathcal{T}_X$

Then for this fixed layer $N \geq 1$, we only have to construct the $U_{k/2^N}$ for every odd k, this is because if k is an even number, then $k = 2j$ and $r = 2j/2^N = j/2^{N-1}$ and for this particular U_r is already constructed. So for every odd $k = 2j+1$, the sets of the form $U_{(k-1)/2^N}$ and $U_{(k+1)/2^N}$ are already defined, and satisfy

$$A \subseteq \overline{U}_{(k-1)/2^N} \subseteq U_{(k+1)/2^N} \subseteq B^c$$

For every $k-1 \neq 0$ and $k+1 \neq 1$. (We will consider these cases later). We claim that for every pair of open sets, $E_1, E_2 \in \mathcal{T}_X$, then there exists some open set $G \in \mathcal{T}_X$ such that if $(E_1, E_2) \in H \subseteq (\mathcal{T}_X \times \mathcal{T}_X)$ where H is defined as the set

$$H = \left\{(E_1, E_2) \subseteq (\mathcal{T}_X \times \mathcal{T}_X) : \overline{E_1} \cap E_2^c = \varnothing\right\}$$

Then there exists some $G = \mathcal{J}(E_1, E_2) \in \mathcal{T}_X$ such that

$$E_1 \subseteq \overline{E_1} \subseteq G \subseteq \overline{G} \subseteq E_2$$

Now consider any any $(E_1, E_2) \in H$, then this pair induces a pair of disjoint sets $\overline{E_1}$ and E_2^c since

$$\overline{E_1} \subseteq E_2 \implies \overline{E_1} \cap E_2^c = \varnothing$$

And by normality, there exists disjoint open sets G_1, G_2 such that

- $\overline{E_1} \subseteq G_1 \in \mathcal{T}_X$
- $E_2^c \subseteq G_2 \in \mathcal{T}_X$
- $G_1 \cap G_2 = \varnothing \implies G_1 \subseteq G_2^c \subseteq E_2$
- Since G_2^c is a closed set that contains G_1 as a subset, $\overline{G_1} \subseteq G_2^c \subseteq E_2$

It is at this point that we will make no further mention of G_2 (so we may discard the notion of G_2 in our minds). Let us now replace G with G_1 then it is an easy task to verify that $G = G_1 = \mathcal{J}(E_1, E_2)$ has the required properties.

Now define for every odd k, since $(U_{(k-1)/2^N}, U_{(k+1)/2^N}) \in H$ (we note in passing that $\mathcal{J}$ is not a function as the set G may not be unique).

$$U_{k/2^N} = \mathcal{J}\left(U_{(k-1)/2^N}, U_{(k+1)/2^N}\right)$$

Then, if $U_{(k-1)/2^N}$ and $U_{(k+1)/2^N}$ is 'well constructed' we have

$$A \subseteq \overline{U}_{(k-1)/2^N} \subseteq U_{(k+1)/2^N} \subseteq B^c$$

Therefore $U_{k/2^N} = \mathcal{J}(U_{(k-1)/2^N}, U_{(k+1)/2^N})$ sits 'right inbetween' the two sets so that

- $A \subseteq \overline{U}_{(k-1)/2^N} \subseteq U_{k/2^N}$ and
- $\overline{U}_{k/2^N} \subseteq U_{(k+1)/2^N} \subseteq B^c$

Combining the above two estimates will give us a 'well constructed' $U_{k/2^N}$ for every $k-1 \neq 0$ and $k+1 \neq 1$. Now let us deal with the remaining pathological cases.

If $k-1$ so happens to be 0, then no $r \in \Delta$ satisfies $r = 0/2^N$, and we substitute

$$\overline{U}_0 = A, \qquad \text{or alternatively, } U_0 = A^o$$

Then $U_0 \in \mathcal{T}_X$, $\overline{U}_0 = A \subseteq B^c$. It is at this point that we must mention that $0, 1 \notin \Delta$, so U_0 and U_1 do not have to obey the rules we have laid out for $U_{r \in \Delta}$.

Now if $k+1$ is equal to 2^N (this makes $r = (k+1)/2^N = 1$) we define

$$U_1 = B^c \in \mathcal{T}_X$$

With this, for every $0 \leq m \leq 2^N - 1, U_{m/2^N}$ must staisfy

$$\overline{U}_{m/2^N} \subseteq B^c = U_1$$

And the pair $(U_{(k-1)/2^N}, U_{(k+1)/2^N}) \in H$ (even for when $N = 1$, since $A = \overline{U}_0 \subseteq U_1 = B^c$) and a corresponding $U_{k/2^N} = \mathcal{J}(\cdot, \cdot)$ such that

- $A \subseteq \overline{U}_{(k-1)/2^N} \subseteq U_{k/2^N}$
- $\overline{U}_{(k+1)/2^N} \subseteq B^c$

Now as a final step, we complete the base case for when $N = 1$. We would only have to construct for $k = 1$, since

$$U_{1/2} = \mathcal{J}(U_0, U_1) = \mathcal{J}(A, B^c)$$

Apply the induction step, and the proof is complete, at long last. ∎

Proposition 13.2: Folland Theorem 4.15: Urysohn's Lemma

Urysohn's Lemma. Let X be a normal space, if A and B are disjoint closed subsets of X, then there exists a $f \in C(X, [0, 1])$ such that $f = 0$ on A and $f = 1$ on B.

Proof. Let $r \in \Delta$ be as in Lemma 4.14, and set U_r accordingly except for $U_1 = X$. Define

$$f(x) = \inf\{k : x \in U_k\}$$

Let us also write $W = \{k : x \in U_k\}$, Then for every $x \in A$ we have $f(x) = 0$, since by the construction of the 'onion' function in Lemma 4.14, for each $r \in \Delta \cap (0, 1)$,

$$x \in A \subseteq U_r \implies f(x) \leq r$$

Since $r > 0$ is arbitrary, and $0 \in W$, we can use a classic ε argument. If $f(x) > 0$ then there exists some $0 < r < f(x)$ by density of the dyadic rationals on the line, if $f(x) < 0$ then this implies that there exists some $f(x) < r < 0$ such that $x \in U_r$, but no $r \in \Delta$ can be negative, hence $f(x) = 0$.

Now, for every $x \in B$, since A and B are disjoint, and $A \subseteq U_r \subseteq B^c$, then for every $x \in B$ means that x is not a member of any U_r, but we set $U_1 = X$. Since none of the $r \in (0, 1)$ is a member of the set we are taking the infimum, and $x \in U_1 = X$. The ε argument follows: suppose for every $\varepsilon > 0$, $(1-\varepsilon) \notin W$, and $1 \in W$, then $f(x) = 1$.

Since $x \in U_1 = X$, for every $x \in X$, $f(x) \leq 1$, and $f(x)$ cannnot be negative as $r > 0$ for every $r \in \Delta$. So $0 \leq f(x) \leq 1$. Now we have to show that this $f(x)$ is continuous. The remainder of the proof is divided into two parts. We would like to show that the inverse images of the half lines are open in X. So $f^{-1}((-\infty, \alpha)) \in \mathcal{T}$ and $f^{-1}((\alpha, +\infty)) \in \mathcal{T}$.

Suppose that $f(x) < \alpha$, so $\inf W < \alpha$, and using the density of Δ, there exists an r, $f(x) < r < \alpha$ such that $x \in U_r$ such that $x \in \bigcup_{r<\alpha} U_r$. So $f^{-1}((-\infty, \alpha)) \subseteq \bigcup_{r<\alpha} U_r$.

Fix an element $x \in \bigcup_{r<\alpha} U_r$, this induces an r such that $\inf W \leq r < \alpha$ therefore $f(x) < \alpha$, and $\bigcup_{r<\alpha} U_r \subseteq f^{-1}((-\infty, \alpha))$.
For the second case, suppose that $f(x) > \alpha$, then $\inf W > \alpha$, and there exists an r (by density) such that $\inf W > r > \alpha$ such that for every $k \in W$, $k \neq r$. Therefore $x \notin U_r$, but by density again, and using the property of the onion function: for every $s < r$ we get $\overline{U_s} \subseteq U_r$, taking complements (which reverses the estimate) — we have $x \notin \overline{U_s}$, but $\left(\overline{U_s}\right)^c$ is open in X. It immediately follows that

$$x \in f^{-1}((\alpha, +\infty)) \implies x \in (U_r)^c \subseteq \left(\overline{U_s}\right)^c \subseteq \bigcup_{s>\alpha} \left(\overline{U_s}\right)^c$$

So $f^{-1}((\alpha, +\infty))$ is a subset of $\bigcup_{s>\alpha} \left(\overline{U_s}\right)^c$. To show the reverse, fix an element x in the union, then this induces some $x \in \left(\overline{U_s}\right)^c \subseteq (U_s)^c$. Then for this $s > \alpha$, $(-\infty, s)$ contains no elements of W. This is because for every $p < s$ implies that $(U_s)^c \subseteq (U_p)^c$, so $p \notin W$. Our chosen s is a lower bound for W, and $\alpha < s \leq \inf W = f(x)$.

Since all of the inverse images from the generating set of $(\mathbb{R}, \mathcal{T}_{\mathbb{R}})$ are open in X, using Theorem 4.9 finishes the proof. ■

Notes on the construction of the countable 'onion' sequence within a normal space **X**.

If **X** is a normal space, and A and B are disjoint closed subsets, then we can easily find an open U with

$$A \subseteq U \subseteq \overline{U} \subseteq B^c \tag{131}$$

We say that U hides in B^c if the closure of U is contained in B^c. Define $\Delta_n = \left\{k2^{-n},\ 1 < k < 2^n\right\}$, so that $\Delta_n \subseteq (0,1)$ for all $n \geq 1$. Notice

$$\Delta_1 \supseteq \cdots \supseteq \Delta_n \supseteq \Delta_{n+1}$$

and the even indices for Δ_{n+1} are contained in Δ_n. Suppose Δ_n is well defined, it suffices to choose the odd indices for Δ_{n+1}. If $r = j2^{-(n+1)}$, where j is odd, then r sits in between precisely two elements in $\Delta_n \cup \{0, 1\}$. If r sits between an endpoint, then define $\overline{U_0} = A$, and $B^c = U_1$. And denote the closest left and neighbours by s, t respectively. If $s < r < t$, it is clear that $\overline{U_s}$ and U_t^c are disjoint closed sets.

Use the 'normal space' construction to obtain an superset of $\overline{U_s}$ that hides in U_t, denote this open set by U_r, and similar to Equation (131)

$$\overline{U_s} \subseteq U_r \subseteq \overline{U_r} \subseteq U_t$$

Now that the construction of this sequence is complete, we wish to prove Urysohn's Lemma. Let A and B be disjoint closed sets. And define

$$f(x) = \inf\left\{r \in \Delta \cup \{1\},\, x \in U_r\right\}$$

where $U_1 = \mathbf{X}$. So that $0 \leq f(x) \leq 1$ is immediate. If $x \in A$, then x is in all U_r, and by density of $\Delta \subseteq (0,1)$, we have $f(x) = 0$. Conversely, if $x \in B$ then $x \notin U_r$ for all $r \in \Delta$, if E denotes the indices in Δ where $x \in U_s$ when $s \in E$,

$$(-\infty,\, r) \subseteq E^c \iff E \subseteq [r,\, +\infty) \iff \inf(E) \geq r \tag{132}$$

Send $r \to 1$ and $f(x) = 1$. Thus $f|_A = 0$ and $f|_B = 1$.

To show continuity, it suffices to show that the inverse images of the open half $\left\{(x > \alpha),\, (x < \alpha)\right\}_{\alpha \in \mathbb{R}}$ lines are indeed open in $\mathbf{X}$. Let α be fixed. And if $x \in \{f < \alpha\}$, we can 'wiggle' the infimum towards the right (towards α), and using density of Δ within $(0,1)$, there exists a $r \in E$ that satisfies $f(x) < r < \alpha$. This is equivalent to

$$x \in \bigcup_{r<\alpha} U_r$$

If there exists an $r < \alpha$ st x belongs to U_r as an element, then $f(x) \leq r < \alpha$.

If $f(x) > \alpha$, then $(-\infty,\, \alpha) \subseteq E^c$, by Equation (132). Suppose $\alpha < 1$, otherwise $\{f > \alpha\} = \varnothing$. Wiggle $f(x)$ to the left and obtain an $r \in \Delta$, $\alpha < r < f(x)$ with $x \notin U_r$. By density again, take any $s < r$ by a small amount (st $s > \alpha$, $s \in \Delta$), and

$$\overline{U}_s \subseteq U_r \iff U_r^c \subseteq \overline{U}_s$$

so that $x \in \overline{U}_s^c$ for some $s > \alpha$. This is equivalent to

$$x \in \bigcup_{s>\alpha} \overline{U}_s^c$$

Conversely, if $x \notin \overline{U}_s^c$ for some $s > \alpha$, since $\{U_r\}$ (thus $\{\overline{U}_r\}$) is increasing, and $x \notin U_r$ for every $r \leq s$. Hence,

$$(-\infty,\, s] \subseteq E^c \iff E \subseteq (s,\, +\infty) \iff f(x) \geq s > \alpha$$

Compactness

Compactness is one of the most important concepts in topology and analysis.

Definition 14.1: Compact topological space

A topological space $\mathbf{X}$ is compact if every open covering $\{U_\alpha\}$ contains a finite subcover. That is, if $\{U_\alpha\}$ is an arbitrary collection of open sets, then

$$\mathbf{X} = \bigcup U_{\alpha \in A} \implies \bigcup_{j \leq n} U_{\alpha_j}$$

Definition 14.2: Compact set

$E \subseteq \mathbf{X}$ is compact if it is compact in the subspace topology.

Definition 14.3: Precompact set

$E \subseteq \mathbf{X}$ is precompact if its closure is compact (as a subset).

Definition 14.4: Paracompact space

A topological space $\mathbf{X}$ is paracompact if every open covering of $\mathbf{X}$ has a locally finite open refinement that covers $\mathbf{X}$.

Definition 14.5: Locally finite collection of sets

Let $\mathcal{A}$ be a collection of subsets of $\mathbf{X}$. It is called locally finite, if at every point $p \in \mathbf{X}$, we can find a neighbourhood U of p (not necessarily open), that intersects only finitely many members of $\mathcal{A}$. In symbols,

$$U \cap E = \varnothing \quad \text{for all but finitely many} \quad E \in \mathcal{A}$$

We do not require $\mathcal{A}$ to be a cover of $\mathbf{X}$, nor do we require $\mathcal{A}$ to be a collection of open sets.

Definition 14.6: Countably locally finite

A collection $\mathbb{B}$ is countably locally finite if it is the countable union of locally finite families.

$$\mathbb{B} = \overbrace{\bigcup_{\mathbb{N}}}^{\substack{\text{countable}\\\text{union}}} \mathbb{B}_n, \quad \text{where each } \mathbb{B}_n \text{ is a locally finite collection}$$

Definition 14.7: Refinement

If $\mathcal{A}$ is a collection of sets, $\mathbb{B}$ is a refinement of $\mathcal{A}$ if every element $B \in \mathbb{B}$, induces an element $A \in \mathcal{A}$, such that $B \subseteq A$.

Remark 14.1: Intuition for refinements

If $\mathbb{B}$ is a refinement of $\mathcal{A}$, we can use the 'absolute continuity' muscle. For each element in $\mathbb{B}$ is dominated by some element (through subset inclusion) in $\mathcal{A}$. Recall, if ν and μ are non-negative measures, then $\nu \ll \mu$ if for every measurable set $E \in \mathcal{M}$, $\mu(E) = 0 \implies \nu(E) = 0$.

A refinement of a family of sets is another family of sets, whose elements are dominated by some other element in the un-refined family. *Refining families makes them 'smaller', cover less area.*

> **Proposition 14.1**
>
> Compact Hausdorff spaces are normal, compact subsets of Hausdorff spaces are closed, and closed subsets of compact sets are again compact.

Properties of Compact Spaces

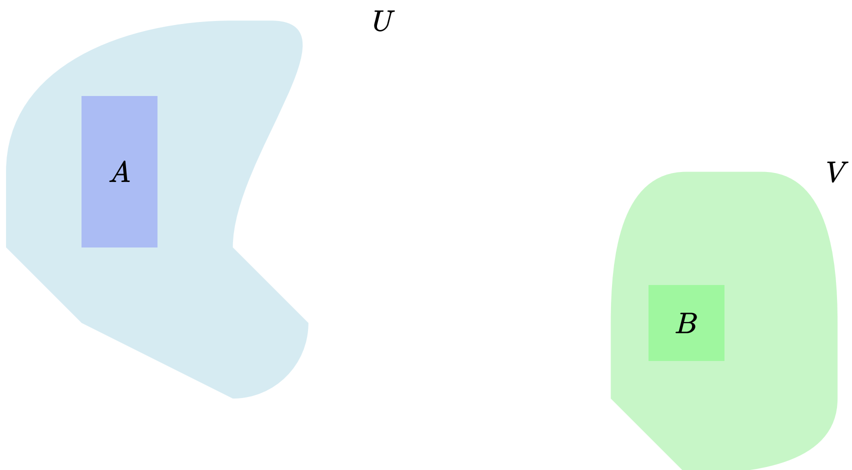

Figure 2: Closed sets A and B within open sets U and V, respectively.

> **Proposition 15.1**
>
> Let $\mathbf{X}$ and $\mathbf{Y}$ be topological spaces.
>
> (a) If $F \in C(\mathbf{X}, \mathbf{Y})$, and $\mathbf{X}$ is compact, then $F(\mathbf{X})$ is compact.
>
> (b) If $\mathbf{X}$ is compact and $F \in C(\mathbf{X}, \mathbb{R})$, then $F(\mathbf{X})$ is bounded, and F attains its supremum and infimum on $\mathbf{X}$.
>
> (c) A finite union of compact subspaces of $\mathbf{X}$ is again compact.
>
> (d) If $\mathbf{X}$ is Hausdorff, and A, B are disjoint, compact subspaces of $\mathbf{X}$, there exists open U and V, (see fig. 2).
>
> (e) Every closed subset of a compact space is compact.
>
> (f) Every compact subset of a Hausdorff space is closed.
>
> (g) Every compact subset of a metric space is bounded.
>
> (h) Every finite product of compact spaces is compact.
>
> (i) Every quotient of a compact space is compact.

Proof of Proposition 15.1 Part A. Let $f \in C(\mathbf{X}, \mathbf{Y})$ with $\mathbf{X}$ compact. Fix an open cover of $f(\mathbf{X})$ in the relative topology,

$$\{U_\alpha \cap f(\mathbf{X})\}_{\alpha \in A} \text{ covers } \mathbf{X},\ U_\alpha \text{ open in } \mathbf{Y}$$

So that $\bigcup f^{-1}(U_\alpha) = \bigcup f^{-1}(U_\alpha \cap f(\mathbf{X})) = \mathbf{X}$. Since $\{f^{-1}(U_\alpha)\}_{\alpha \in A}$ is an open cover for $\mathbf{X}$, this induces a finite subcollection of indices $\{\alpha_1, \dots, \alpha_n\}$ with

$$\bigcup_{j=1}^{n} f^{-1}(U_{\alpha_j}) = \bigcup_{j=1}^{n} f^{-1}(U_{\alpha_j} \cap f(\mathbf{X}))$$

The direct image commutes with unions, therefore

$$f(\mathbf{X}) = f\left(\bigcup_{j=1}^{n} f^{-1}(U_\alpha \cap f(\mathbf{X}))\right) = \bigcup_{j=1}^{n} f\left(f^{-1}(U_{\alpha_j})\right) = \bigcup_{j=1}^{n} U_{\alpha_j}$$

■

Proof of Proposition 15.1 Part B. Let $\mathbf{X}$ be compact, and $f \in C(\mathbf{X}, \mathbb{R})$, so that $f(\mathbf{X}) \subseteq \mathbb{R}$ is compact. Compact subsets are closed and bounded in $\mathbb{R}$, let $A = \sup f(\mathbf{X})$ and $B = \inf f(\mathbf{X})$. Both A and B are accumulation points of $f(\mathbf{X})$, so $A = f(x)$ and $B = f(y)$ for some x, y in $\mathbf{X}$. ■

Proof of Proposition 15.1 Part C. Let $\mathbf{X}$ be a topological space, and $K_1, \dots K_n$ be compact subspaces. Denote $K = \bigcup_{j=1}^{n} K_j$. Let $\{U_\alpha \cap K\}_{\alpha \in A}$ be an open cover for K, where U_α is open in $\mathbf{X}$. We can pass the argument to each individual K_j as follows. Let $1 \leq j \leq n$, then $\{U_\alpha \cap K_j\}_{\alpha \in A}$ is an oepn cover for K_j, so there exists as finite subcollection of indices $I_j \subseteq A$, (a finite subset of A) whose open sets cover K_j. Repeat this process for each j and

$$I = \bigcup_{j=1}^{n} I_j \text{ is a finite subset of } A$$

with $K_j \subseteq \bigcup_{\alpha \in I_j} (U_\alpha \cap K_j) \subseteq \bigcup_{\alpha \in I_j} (U_\alpha \cap K)$. Taking the union over all K_j reads

$$K = \bigcup_{j=1}^{n} K_j \subseteq \bigcup_{j=1}^{n} \bigcup_{\alpha \in I_j} (U_\alpha \cap K) = \bigcup_{\alpha \in I} U_\alpha \cap I$$

■

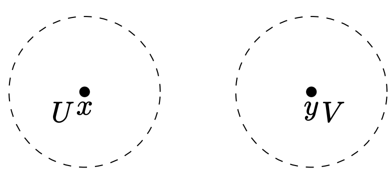

Figure 3: In a Hausdorff space, any two distinct points x and y can be separated by disjoint open neighbourhoods U and V.

Proof of Proposition 15.1 Part D. Let $\mathbf{X}$ be Hausdorff. We first prove that compact subspaces of $\mathbf{X}$ are closed. Indeed, if K is compact in $\mathbf{X}$, fix any $x \in K^c$. Let y range through the elements of K, then $x \neq y$ induces a pair of disjoint open sets U_y and V_y, such that

- $x \in U_y$
- $y \in V_y$
- $U_y \cap V_y = \varnothing$
- See fig. 3

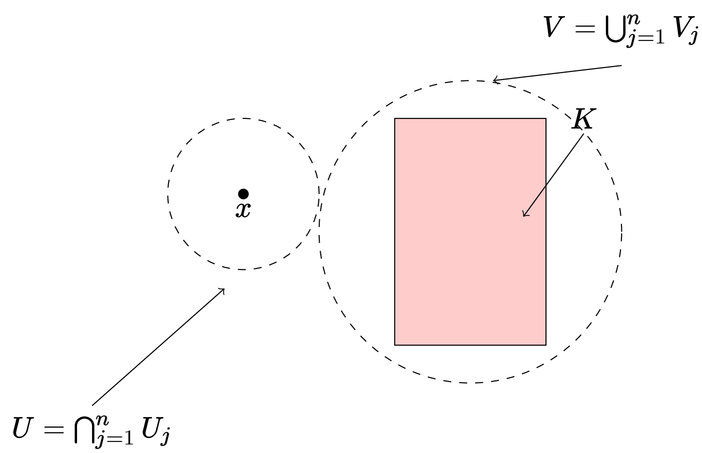

Figure 4: Compact sets are closed in Hausdorff spaces

Let V_y range through all possible $y \in K$, So that $\{V_y\}_{y\in K}$ is an open cover. There exists a finite subcollection of 'anchor points' of K, $y_1, \ldots, y_n$ that corresponds with $\{V_{y_j}\}_{j=1}^{n}$.
A finite intersection of open sets is again open, so

$$U = \bigcap_{j=1}^{n} U_{y_j} \text{ is open}$$

Define $V = \bigcup_{j=1}^{n} V_{y_j}$, so $V \subseteq K$ and $U \cap V = \varnothing$ and $x \in U \subseteq K^c$ (see fig. 4). Therefore K is closed.

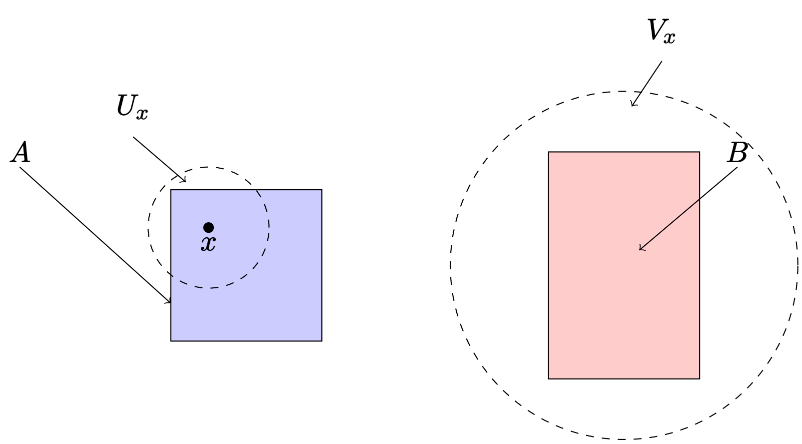

Figure 5: Closed sets A and B, point x in A, and disjoint neighbourhoods U around x and V around B.

Finally, if A and B are disjoint compact sets, then each $x \in A \subseteq B^c$ induces neighbourhoods $x \in U_x$, and $B \subseteq V_x$ (see fig. 5), let x range through all the elements of A. By compactness of A, this produces a finite subcover, and

$$U = \bigcup_{j=1}^{n} U_{x_j} \quad V = \bigcap_{j=1}^{n} V_{x_j}$$

are disjoint open sets that contain A and B respectively.

■

Proof of Proposition 15.1 Part E. Let $K \subseteq \mathbf{X}$ be a closed set of a compact space. Let $\{U_\alpha \cap K\}$ be an open cover for K, where each U_α is open in $\mathbf{X}$. We can append an extra set K^c which is open in $\mathbf{X}$. The collection

$$W = \{U_\alpha\} \cup \{K^c\} \text{ covers } \mathbf{X}$$

so there exists a finite subcollection of $W_1, \ldots, W_n$ that cover $\mathbf{X}$ (since $\mathbf{X}$ is compact by itself). Remove K^c from this finite subcollection if it exists, and take the intersection with K for each element W_j, and

$$\{W_1 \cap K, \ldots, W_n \cap K\} = \{U_1 \cap K, \ldots, U_n \cap K\} \text{ covers } K$$

so K is compact. ■

Proof of Proposition 15.1 Part F. Proven in Part D. ■

Proof of Proposition 15.1 Part G. let $K \subseteq \mathbf{X}$ be a compact subset of the metric space $(\mathbf{X}, d)$. Compact subsets of $\mathbf{X}$ are totally bounded, and hence bounded. ■

Proof of Proposition 15.1 Part H. See Tynchonoff's Theorem in Folland Chapter 4. ■

Proof of Proposition 15.1 Part I. Let $\mathbf{X}$ and $\mathbf{Y}$ be topological spaces and $\pi : \mathbf{X} \to \mathbf{Y}$ be a quotient map. So that $\mathbf{Y}$ is endowed with the quotient topology. So that π is a surjective continuous map. and $\pi(\mathbf{X}) = \mathbf{Y}$. Apply Part A, and we see that $\mathbf{Y}$ is compact. ■

Locally Compact Hausdorff Spaces

Compactness is an intrinsic topological property (in the subspace topology). We see from Proposition 4.25 that compact Hausdorff spaces are normal, which gives a sufficient condition for us to approximate and extend any continuous function; and allows us to extend certain 'local' properties to 'global' properties.

If given a Hausdorff space, not necessarily compact, the natural question is to ask 1) whether a topological space has 'enough' compact subsets to work with, and 2) whether we can embed a given topological space in a larger one to force it to be compact.

Definition 16.1: LCH space

Let $\mathbf{X}$ be a Hausdorff space. We call $\mathbf{X}$ a LCH space if every point $p \in \mathbf{X}$ admits a compact neighbourhood. That is, a compact set K whose interior contains p.

We note in passing that the above definition differs slightly from the usual 'local' definitions.

Definition 16.2: Locally connected

Let $\mathbf{X}$ be a topological space, it is locally connected if for every $x \in \mathbf{X}$, and open neighbourhood U containing x, there exists a connected, open neighbourhood V of x such that $x \in V \subseteq U$.

Definition 16.3: Locally path-connected

Let $\mathbf{X}$ be a topological space, it is locally connected if for every $x \in \mathbf{X}$, and open neighbourhood U containing x, there exists a path-connected, open neighbourhood V of x such that $x \in V \subseteq U$.

Definition 16.4: Local homeomorphism

$\mathbf{X}$ locally homeomorphic to $\mathbb{R}^n$ if every point $x \in \mathbf{X}$ belongs to a coordinate chart (U, ϕ), where U is an open neighbourhood of x and ϕ is a homeomorphism from $U \to \phi(U) \subseteq \mathbb{R}^n$.

Definition 16.5: Local diffeomorphism

Let M be a smooth manifold and $F \in C^\infty(M, N)$. F is a local diffeomorphism if every $p \in M$ in its domain induces a neighbourhood $U \subseteq M$ with $F|U : U \to F(U)$ is a diffeomorphism (in the sense of two open sub-manifolds).

Chapter B: Abstract Algebra

Groups

Definition 1.1: Semigroups, Monoids

A non-empty set G equipped with an associative binary operation $G \times G \to G$ is called a semigroup. For every $a, b, c \in G$, we have

$$a(bc) = (ab)c \tag{133}$$

A *monoid* is a semigroup G which contains a *two-sided identity* element $e \in G$ such that $ae = ea$ for all $a \in G$. (not necessarily unique)

Monoids admit unique two-sided identities.

Lemma 1.1: Monoids: unique identity

Let e and i be two-sided identities for a monoid G, then

Proof.

$$e = ei = i$$

■

Definition 1.2: Group

A semigroup G is a group if every element $a \in G$ admits a two-sided inverse a^{-1}. (not necessarily unique)

$$aa^{-1} = a^{-1}a = e$$

Proposition 1.1: Properties of Groups (Hungerford: Theorem 1.2)

Let G be a group with identity e, which is unique by lem. 1.1. Then

(i) $c \in G$ and $cc = c$ implies $c = e$.

(ii) Left/Right cancellation:

$$\begin{cases} ab = ac \implies b = c \\ ba = ca \implies b = c \end{cases}$$

(iii) If $a \in G$, its two-sided inverse is unique.

(iv) Let $a \in G$, then the inverse of its two-sided inverse (uniqueness guaranteed by iii), is a itself; or $(a^{-1})^{-1} = a$.

(v) If $a, b \in G$, then the following equations in x, y admit unique solutions

$$\begin{cases} ax = b \\ ya = b \end{cases}$$

Proof of Proposition 1.1.

Proof of Part (i):

$$cc = c \implies (cc)c^{-1} = cc^{-1} \implies c(cc^{-1}) = e \implies ce = c = e$$

Proof of Part (ii): First claim:

$$\begin{aligned} ab = ac &\implies a^{-1}(ab) = a^{-1}(ac) \\ &\implies (a^{-1}a)b = (a^{-1}a)c \implies eb = ec \implies b = c \end{aligned}$$

Second claim is the same, just cancel from the right using $aa^{-1} = e$ and associativity.

Proof of Part (iii): Suppose b and c are two-sided inverse for a, it follows from Part ii that

$$ab = ac \implies b = c = a^{-1}$$

Proof of Part (iv): From Part iii, the two-sided inverses of group elements exist and are unique, and $a^{-1}a = aa^{-1}$ so a is an inverse for a^{-1}, and it is the only inverse.

Proof of Part (v): First equation: write $ax = b = a(a^{-1}b)$, left-cancelling reads $x = a^{-1}b$, uniqueness follows from Part ii. Second equation is similar.

■

Lemma 1.2: Group: equality lemma

For any pair of elements $a, b \in G$, $a = b \iff ab^{-1} = e$.

Proof. ($\implies$): $a = b \implies ab^{-1} = bb^{-1} = e$. ($\impliedby$): $ab^{-1} = e \implies a(b^{-1}b) = eb \implies a = eb = b$. ■

Proposition 1.2: Semigroup: upgrade to group I (Hungerford Proposition 1.3)

Let G be a semigroup, G is also a group iff both of the conditions below hold

- Existence of a left-identity: there exists $e \in G$ for every $a \in G$, $ea = a$.
- Existence of left-inverses: for every $a \in G$, there exists a $a^{-1} \in G$ with $a^{-1}a = e$, where e is any left-identity element.

Proof. ($\impliedby$) is trivial. Suppose both conditions hold, notice the proof for Proposition 1.1 Part (i) we only used left-cancellation. $cc = c \implies e$. To prove a^{-1} is also a right-inverse for a, we can force it as follows:

$$(aa^{-1})(aa^{-1}) = a(a^{-1}a)a^{-1} = aea^{-1} = e \implies aa^{-1} = e$$

and a^{-1} is also a right-inverse, so every element $a \in G$ admits a two-sided inverse denoted by a^{-1}. To show e is also a right-identity for any arbitrary element $a \in G$,

$$\begin{aligned} ae &= a(a^{-1}a) && \text{left inverse} \\ &= (aa^{-1})a && \text{associativity} \\ &= ea && \text{right inverse} \\ &= a && \text{left identity} \end{aligned}$$

■

Proposition 1.3: Semigroup: upgrade to group II (Hungerford Proposition 1.4)

Let G be a semigroup, G is a group iff for every pair of elements $a, b \in G$, the equations in x and y

$$\begin{cases} ax = b \\ ya = b \end{cases} \tag{134}$$

have solutions (not necessarily unique).

Proof. If G is a group, the existence of the solutions to eq. (134) follow from Proposition 1.1. We will attempt the contrapositive. Suppose G has no left identity, for every $e \in G$ we can always find an element $a \in G$ such that $ea \neq a$, but this is precisely the (first) equation for $a = a$ and $b = a$.

Now suppose G has a left identity element (not necessarily unique). Fix $e \in G$ as any left-identity, and suppose there is an element $a \in G$ with no left inverse, so for every $b \in G$, $ba \neq e$. But b is precisely the solution to the (second) equation with parameters $a = a$ and $b = e$. The negation of Proposition 1.2 is precisely the negation of Proposition 1.3, and the proof is complete. ∎

Proposition 1.4: Hungerford Theorem 1.5

Let $R/\sim$ be an equivalence relation on a group G, such that it 'preserves' the group multiplication. More precisely,

$$\begin{cases} a_1 \sim a_2 \\ b_1 \sim b_2 \end{cases} \implies a_1 b_1 \sim a_2 b_2$$

Then the set G/R of all equivalence classes of G under R is a monoid under the binary operation defined by

$$(\overline{a})(\overline{b}) = \overline{ab} \qquad \text{reads: the product of two classes is the class containing the product of any pair of elements from the two classes} \tag{135}$$

where $\overline{a}$ denotes the equivalence class containing a. If G is a group, so is G/R, if G is an abelian group, so is G/R.

Proof. First, notice the binary operation in Equation (135) is well defined. It is independent of the equivalence class representatives chosen, as we have restriction on R that 'forces' the operation on G/R to be well defined. Indeed, let $\overline{a}$ and $\overline{b}$ be elements of G/R, if $a_1, a_2 \in \overline{a}$, and $b_1, b - 2 \in \overline{b}$, by definition of R:

$$a_1 \sim a_2 \quad \text{and} \quad b_1 \sim b_2$$

by Equation (135), $a_1 b_1 \sim a_1 b_2 \implies \overline{a_1 b_1} = \overline{a_2 b_2}$.

Associativity is proven similarly, fix $\overline{a}, \overline{b}, \overline{c} \in G/R$, we pass the argument to any of the representatives of the three classes, so

$$(\overline{a}\overline{b})\overline{c} \overset{\Delta}{=} \overline{ab}\overline{c} = \overline{(ab)c} = \overline{a(bc)} \overset{\Delta}{=} \overline{a}\overline{bc} = \overline{a}(\overline{b}\overline{c})$$

Pass the argument to the representatives, let e denote the identity element in G, it is easily shown that $\overline{e}$ is the identity element in G/R, similarly for two-sided inverses and commutativity of the binary operation. ∎

Homomorphisms

Definition 2.1: Homomorphism

Let G and H be semigroups, $f : G \to H$ is a semi-group *homomorphism* if for all $a, b \in G$,

$$f(ab) = f(a)f(b) \tag{136}$$

Definition 2.2: Monomorphism

Injective homomorphism.

Definition 2.3: Epimorhpism

Surjective homomorphism.

Definition 2.4: Isomorphism

Bijective homomorphism.

Definition 2.5: Endomorphism

Homomorphism for which the domain and codmain (not the range) are equal; i.e $H = G$.

Definition 2.6: Automorphism

Bijective endomorphism.

Definition 2.7: Kernel of a homomorphism

The kernel of $f \in \mathrm{Hom}(G, H)$ is defined

$$\mathrm{Ker}\, f = \left\{ a \in G,\ f(a) = e \in H \right\} \tag{137}$$

as the set of elements in G that get sent to the identity of H.

Proposition 2.1: Hungerford Theorem 2.3

Let G and H be groups and let $f \in \mathrm{Hom}(G, H)$. Denote the identity elements of G and H by e_G and e_H

(i) $f(e_G) = e_H$,

(ii) $f(a^{-1}) = (f(a))^{-1}$ for every $a \in G$.

(iii) f is a monomorphism iff $\ker f = \{e_G\}$,

(iv) f is an isomorphism iff there exists a homomorphism $f^{-1} : H \to G$ that is also a two-sided inverse for f. In symbols:

$$f \circ f^{-1} = \mathrm{id}_H \quad \text{and} \quad f^{-1} \circ f = \mathrm{id}_G \tag{138}$$

Proof of Proposition 2.1.

Proof of Part (i): We will use Proposition 1.1 (i). Since $f(e_G) = f(e_G e_G) = f(e_G)f(e_G)$ in H, we see that $f(e_G) = H$ and $e_G \in \operatorname{Ker} f$

Proof of Part (ii): Let $a \in G$ be arbitrary, using Part (i), we can 'pass the multiplication' between $f(a)$ and $f(a^{-1})$ into G,

$$f(a)f(a^{-1}) = f(e_G) = e_H \implies f(a^{-1}) = (f(a))^{-1}$$

Proof of Part (iii): Suppose $\ker f = e_G$. Let $a, b \in G$ such that $f(a) = f(b)$. The equality lemma Lemma 1.2 tells us $(f(a))^{-1} = f(b)$ and $b = a^{-1}$, so $a = b$ by the Lemma again; f is injective.

Conversely, suppose f is injective, Part (i) tell us $\{e_G\} \subseteq \ker f$. Suppose $a \in \ker f \subseteq G$, but $e_G \in \ker f$, so $f(a) = f(e_G) = e_H$ forces $a = e_G$, and $\ker f = \{e_G\}$.

Proof of Part (iv): ($\impliedby$) is trivial since the existence of a (functional) two-sided inverse is equivalent to bijectivity. Suppose f is an isomorphism, and define f^{-1} as its two-sided (functional) inverse, it suffices to show that $f^{-1} \in \operatorname{Hom}(H, G)$. Fix $f(a)$ and $f(b)$ as arbitrary elements in H. We can do this because f is a bijection, so every element in H has a unique 'representative' in G.

$$f^{-1}(f(a))\, f^{-1}(f(b)) = ab = f^{-1}(f(ab)) = f^{-1}(f(a)f(b))$$

■

Definition 2.8: Subgroup $H < G$

Let G be a group. If H is a non-empty subset of G that is closed under group operations, then it is a *subgroup of* G. We write $H < G$.

The *trivial subgroup* of G is $\{e\}$ and consists of one element. H is called a *proper subgroup* if $H \neq G$ and $H \neq \{e\}$.

Proposition 2.2: Hungerford Exercise 9

Let $f \in \operatorname{Hom}(G, H)$, $A < G$ and $B < H$,

(i) $\operatorname{Ker} f$ is a subgroup of G,

(ii) $f(A)$ is a subgroup of H,

(iii) $f^{-1}(B)$ is a subgroup of G.

Subgroups

Proposition 3.1: Subgroup criteria (Hungerford Theorem 2.5)

A non-subset $H \subseteq G$ is a subgroup iff for any a, b in H, $ab^{-1} \in H$.

Proof. ($\Longleftarrow$): Choose $a = b$, then $aa^{-1} = e \in H$ acts as the two-sided identity in H, and $ea^{-1} = a^{-1} \in H$ for every $a \in H$. So H is a subgroup by Proposition 1.2. ($\Longrightarrow$): If H is a subgroup, then Proposition 1.2 tells us $b^{-1} \in H$ for every $b \in H$, hence $ab^{-1} \in H$ for elements $a, b \in H$. ■

Corollary 3.1: Hungerford Corollary 2.6

If G is a group and $\{H_i,\, i \in I\}$ is a nonempty family of subgroups, then their intersection $H \triangleq \cap_{i\in I} H_i$ is again a subgroup in G.

Proof. Let $a, b \in H$, then $ab^{-1} \in H_i$ for every $i \in I$, hence $ab^{-1} \in \cap_{i\in I} H_i = H$, and H is a subgroup by Proposition 3.1. ■

Definition 3.1: Subgroup generated by $A \subseteq G$

Let A be a subset of G, the *subgroup generated by* A is the smallest subgroup $H < G$ that contains A as a subset, denoted by $\langle A \rangle$.

Proposition 3.1 gives us an explicit formula for H,

$$H = \bigcap_{\substack{H_i < G \\ A \subseteq H_i}} H_i$$

If A is finite, and $H = \langle A \rangle$ is said to be *finitely generated.* We also write

$$H = \langle a_1, \ldots, a_n \rangle = \Big\langle \{a_1, \ldots, a_n\} \Big\rangle$$

If A consists of one element, $\{a\} = A$, then $\langle a \rangle = \langle \{a\} \rangle$ is called the *cyclic group generated by* a.

Proposition 3.2: Hungerford Theorem 2.8

If G is a group and $A \subseteq G$ is a non-empty subset, the subgroup generated by A is precisely the collection of all finite products (powers included), or

$$\langle A \rangle = \left\{ a_1^{n_1} a_2^{n_2} \cdots a_t^{n_t},\ a_i \in A,\ n_i \in \mathbb{Z} \right\}$$

If $A = \{a\}$, then $\langle A \rangle = \left\{a^k,\ k \in \mathbb{Z}\right\}$.

Proof. Let $W = \left\{a_1^{n_1} a_2^{n_2} \cdots a_t^{n_t},\ a_i \in A,\ n_i \in \mathbb{Z}\right\}$. We first show W is a subgroup of G. Indeed, if $n_1 = n_2 = \cdots = n_t = 0$, then $a_1^{n_1} a_2^{n_2} \cdots a_t^{n_t} = e$. And for any $b \in W$,

$$b = a_1^{n_1} a_2^{n_2} \cdots a_t^{n_t} \implies b^{-1} = a_t^{-n_t} \cdots a_2^{-n_2} a_1^{-n_1}$$

where each $-n_i \in \mathbb{Z}$, so $b^{-1} \in H$ as well. $\langle A \rangle$ is the smallest subgroup containing A, so $\langle A \rangle \subseteq W$. Conversely, fix an element $b \in W$, so b has the form

$$b = a_1^{n_1} a_2^{n_2} \cdots a_t^{n_t},\ a_i \in A,\ n_i \in \mathbb{Z}$$

and a simple induction will show each $a_i^{n_i} \in \langle A \rangle$ for $1 \le i \le t$, so $b \in \langle A \rangle$ and $\langle A \rangle = W$. ∎

Definition 3.2: Lattice of subgroups

Let $\{H_i\}_{i \in I}$ be a collection of subgroups of G, then

$$\mathrm{glb}\{H_i\} = \bigcap_{i \in I} H_i, \quad \mathrm{lub} = \left\langle \bigcup_{i \in H} H_i \right\rangle$$

The collection of subgroups of G is a *complete lattice*.

Cyclic groups

The proof for the following is straight-forward, the book separates the case into $H = \langle 0 \rangle$ and $H \neq \langle 0 \rangle$, then H contains a non-zero element $h \neq 0$, then $|h| \in \mathbb{N}^+$ is an element in H as well, so H contains a least positive element by invoking the Well Ordering Property. For the second half of the proof, we force $r = 0$ by the Division Algorithm.

Definition 4.1: Order of a subgroup $H < G$

The order of a subgroup H is its cardinality $|H|$.

Definition 4.2: Order of an element $a \in G$

The order of an element $a \in G$ is the order of $|\langle a \rangle|$.

Proposition 4.1: Hungerford Theorem 3.1

Let $\mathbb{Z}$ be equipped with its usual addition operation. Then every subgroup $H < \mathbb{Z}$ is cyclic, either $H = \langle 0 \rangle$ or $H = \langle m \rangle$. With m being the least positive integer in H. If $H \neq \langle 0 \rangle$, then H is infinite.

Proposition 4.2: Hungerford Theorem 3.2

Every infinite cyclic group $G = \langle a \rangle$ is isomorphic to the additive group $\mathbb{Z}$, and every finite cyclic group of order $0 < m < +\infty$ is isomorphic to the additive group $\mathbb{Z}_m$.

Proof of Proposition 4.2. Let $f : \mathbb{Z} \to G$ and $f(k) = a^k$. By Proposition 3.2,

$$G = \left\{ a^k,\ k \in \mathbb{Z} \right\}$$

f is clearly surjective and $f \in \mathrm{Hom}(\mathbb{Z}, G)$ is an easy exercise to verify. The proof splits into two parts

(i) Ker f is trivial: By Proposition 2.1, f is an isomorphism from $\mathbb{Z}$ to G, and $G \cong \mathbb{Z} \implies |G| = |\mathbb{Z}|$.

(ii) Ker f is not trivial: Arguing as in Proposition 4.2, Ker f is a non-trivial subgroup of $\mathbb{Z}$, so it contains a least positive element m, and Ker $f = \langle m \rangle$. m is an element of ker f, so

$$f(m) = a^m = e \implies f(jm) = a^{jm} = \prod_{l=1}^{j} a^m = e,\ j \in \mathbb{Z} \tag{139}$$

Now suppose r and s are integers with $f(r) = f(s)$,

$$\begin{aligned} a^r = a^s &\iff a^{r-s} = e \\ &\iff r - s \in \mathrm{Ker}\, f \\ &\iff r - s \in \langle m \rangle \\ &\iff r \equiv s \quad \mathrm{mod}\ m \\ &\iff \overline{r} = \overline{s} \end{aligned}$$

where $\overline{r}$ denotes the $\mathbb{Z}_m$ equivalence class of r. Let β be a map from $\mathbb{Z}_m \to G$, such that

$$\beta(\overline{k}) = f(k) = a^k$$

This is well defined, since β is an invariant on each equivalence class, if $r - s$ differ by a multiple of m, then Equation (139) states that $\beta(\overline{r}) = \beta(\overline{s})$. G is finite, as

$$G = \langle a \rangle = \left\{ a^k,\ k \in \mathbb{Z},\ m < k < m \right\}$$

and the kernel of β is trivial, it is an isomorphism and $\mathbb{Z}_m \cong G$.

■

Cosets

Normal Subgroups

Isomorphism Theorems

Chapter C: Algebraic Topology

Homotopy

This section will follow Munkres Chapters 9 and 13 closely. Possibly other chapters as well.

Definition 1.1: Path

A *path* is a continuous function from the unit interval $f : [0, 1] \to \mathbf{X}$. We say f is a *path form x_0 to x_1* if $f(0) = x_0$ and $f(1) = x_1$.

We denote the set of paths from x_0 to x_1 by $\mathrm{Path}(x_0, x_1)$. If $f \in \mathrm{Path}(x_0, x_1)$, we sometimes denote the *reversal of f* by $\overline{f} \in \mathrm{Path}(x_1, x_0)$, where $\overline{f}(s) \triangleq f(1 - s)$.

Definition 1.2: Loop

A *loop* at $x_0 \in \mathbf{X}$ is a path that begins and ends at x_0, and $\mathrm{Loop}(x_0) \triangleq \mathrm{Path}(x_0, x_0)$. The constant path (or loop) at x_0 is denoted by $e_{x_0} : [0, 1] \to \mathbf{X}$.

$$e_{x_0}(s) = x_0, \quad \forall s \in [0, 1]$$

Definition 1.3: Homotopy of $C(\mathbf{X},\ \mathbf{Y})$

Let f, and g continuous functions from $\mathbf{X}$ to $\mathbf{Y}$. f and g are homotopic, denoted by $f \approx g$ if there exists a continuous function $F \in C(\mathbf{X} \times I,\ \mathbf{Y})$ where

$$F(x, 0) = f(x) \quad \text{and} \quad F(x, 1) = g(x) \tag{140}$$

where $I = [0, 1]$.

The function F is called the *homotopy between f and g*.

If $f \simeq h$, where h is the constant function, we say f is *nulhomotopic*.

Definition 1.4: Path Homotopy of $\mathrm{Path}(x_0, x_1)$

Two paths f_0, $f_1 \in \mathrm{Path}(x_0, x_1)$ are said to be *path homotopic*, if there exists a continuous function $F \in C(I \times I,\ \mathbf{X})$, with

- F is a *homotopy between f_0 and f_1* (in the sense of Definition 1.3). For every $s \in [0, 1]$,

$$F(s, 0) = f_0(s) \quad \text{and} \quad F(s, 1) = f_1(s) \tag{141}$$

- F leaves the endpoints fixed. For every $t \in [0, 1]$, then

$$F(0, t) = x_0 \quad \text{and} \quad F(1, t) = x_1 \tag{142}$$

If f_0 and f_1 are path-homotopic, we write $f_0 \simeq_p f_1$.

- The function $F \in C(I \times I,\ \mathbf{X})$ is called the *path homotopy between f_0 and f_1*.
- If $f \in \mathrm{Loop}(x_0)$ is path homotopic to the constant path e_{x_0}, then f is *nulhomotopic*.
- The relation $\simeq_p$ is defined for paths that have the same initial and final points. So it is a relation *on* $\mathrm{Path}(x_0, x_1)$.

Proposition 1.1: Munkres Lemma 51.1

The relations $\simeq$ and $\simeq_p$ are equivalence relations on $C(\mathbf{X}, \mathbf{Y})$ and $\mathrm{Path}(x_0, x_1)$ respectively.

Proof. ($f \simeq f$): Let $f \in C(\mathbf{X}, \mathbf{Y})$. Define

$$F : \mathbf{X} \times I \to \mathbf{Y} \quad \text{For every } t \in [0,1], F(x,t) = f(x)$$

F is continuous, since $F = \pi_{\mathbf{X}} \circ (f \times \mathrm{id}_{[0,1]})$, where $f \times \mathrm{id}_{[0,1]}$ is the product of two continuous functions, which is again continuous by Chapter A. Moreover, $F(x,0) = f(x) = F(x,1)$, so F is a homotopy between f and itself.

($f \simeq g \implies g \simeq f$): Let F be the homotopy between f and g. Let G be the 'reversal' in the second coordinate of F, meaning

$$G(x,t) = F(x, 1-t) \text{ is continuous, since } G = F \circ (\mathrm{id}_{\mathbf{X}} \times c)$$

where $c : I \to I$ that maps $t \mapsto 1 - t$ is continuous, so $\mathrm{id}_{\mathbf{X}} \times c$ is continuous; hence G is continuous. Notice for every $x \in \mathbf{X}$,

$$G(x,0) = F(x,1) = g(x) \quad \text{and} \quad G(x,1) = F(x,0) = f(x)$$

therefore G is a homotopy between g and f.

($f \simeq g,\ g \simeq h \implies f \simeq h$): Let F be the homotopy between f and g, and G be the homotopy between g and h. Define a function $H : \mathbf{X} \times I \to \mathbf{Y}$ that morphs f into g on $[0, 2^{-1}]$, then g into h on $[2^{-1}, 1]$

$$H(x,t) = \begin{cases} F(x, 2t - \lfloor 2t \rfloor) & \text{for } 0 \le t \le 2^{-1} \\ G(x, 2t - \lfloor 2t \rfloor) & \text{for } 2^{-1} \le t \le 1 \end{cases} \tag{143}$$

where $\lfloor \cdot \rfloor$ denotes the *floor function*.

- H is well defined on the overlap $\mathbf{X} \times 2^{-1}$, since $F(x,1) = G(x,0) = g(x)$ at every $x \in \mathbf{X}$.
- If $t = 0$, then $H(x,1) = F(x,0) = f(x)$, and $t = 1$ gives $H(x,1) = G(x,1) = h(x)$.
- Since $H|_{\mathbf{X} \times [0,2^{-1}]}$ and $H|_{\mathbf{X} \times [2^{-1},1]}$ are continuous functions, and they agree on the overlap, H is continuous by the pasting Lemma, and defines a homotopy between f and h.

Now consider paths f, g, h in $\mathrm{Path}(x_0, x_1)$, ($f \simeq_p f$) is trivial. So is symmetry of $\simeq_p$, as the reversal in the second coordinate (see above) of the path homotopy between f and g is path homotopy between g and f.

Suppose $f \simeq_p g$, and $g \simeq_p g$. Let F, and G be the path homotopies between f, g and g, h. Write H as in Equation (143), it is a continuous function on $I \times I \to \mathbf{X}$, that satisfies

$$H(s,0) = F(s,0) = f(s) \quad \text{and} \quad H(s,1) = G(s,1) = h(s) \text{ for every } s \in [0,1]$$

If $s = 0$, it is easy to see from Equation (143) that for every $t \in [0,1]$,

$$H(0,t) = \begin{cases} F(0, 2t - \lfloor 2t \rfloor) = x_0 & \text{for } 0 \leq t \leq 2^{-1} \\ G(0, 2t - \lfloor 2t \rfloor) = x_0 & \text{for } 2^{-1} \leq t \leq 1 \end{cases} = x_0 \quad \text{and}$$
$$H(1,t) = \begin{cases} F(1, 2t - \lfloor 2t \rfloor) = x_1 & \text{for } 0 \leq t \leq 2^{-1} \\ G(1, 2t - \lfloor 2t \rfloor) = x_1 & \text{for } 2^{-1} \leq t \leq 1 \end{cases} = x_1$$

So the endpoints remain fixed throughout the deformation in t, and H is a path homotopy between f and h. This proves transitivity. ■

Path and PathClass Products

Definition 2.1: Product of Paths $f * g$

Let $f \in \text{Path}(x_0, x_1)$ and $g \in \text{Path}(x_1, x_2)$, the product of f and g, denoted by $f * g$ is another path from x_0 to x_2. For $s \in [0,1]$,

$$(f * g)(s) \triangleq \begin{cases} f(2s - \lfloor 2s \rfloor) & \text{for } 0 \leq s \leq 2^{-1} \\ g(2s - \lfloor 2s \rfloor) & \text{for } 2^{-1} \leq s \leq 1 \end{cases} \tag{144}$$

Notice the similarities between Equations (143) and (144),

Proposition 2.1: Properties of the Path Product

Let $f \in \text{Path}(x_0, x_1)$ and $g \in \text{Path}(x_0, x_1)$, let $k \in C(\mathbf{X},\ \mathbf{Y})$, then

(i) Invariant under left-multiplication: $f \simeq_p g \implies k \circ f \simeq_p k \circ g$, where $k \circ f$ and $k \circ g$ are elements Paths from $k(x_0)$ to $k(x_1)$, and if F be a path homotopy between f and g, then $k \circ F$ is a path homotopy between $k \circ f$ and $k \circ g$.

(ii) If we redefine $f \in \text{Path}(x_0, x_1)$, $g \in \text{Path}(x_1, x_2)$, and k be as above, then

$$k \circ (f * g) = (k \circ f) * (k \circ g)$$

Proof.

Proof of Part (i): It is clear that $k \circ f$ and $k \circ g$ are elements of $\text{Path}(k(x_0), k(x_1))$, and see Part (ii) for the proof of $k \circ f \simeq_p k \circ g$.

Proof of Part (ii): Let F be the path homotopy between f and g. The composition $(k \circ F)$ is in $C(\mathbf{X} \times I,\ \mathbf{Y})$. Equation (141) reads

$$(k \circ F)(s,0) = k(F(s,0)) = (k \circ f)(s) \text{ and}$$
$$(k \circ F)(s,1) = k(F(s,1)) = (k \circ g)(s) \text{ for every } s \in [0,1]$$

and Equation (142) gives

$$(k \circ F)(0,t) = k(F(0,t)) = k(x_0) \text{ and}$$
$$(k \circ F)(1,t) = k(F(1,t)) = k(x_1) \text{ for every } t \in [0,1]$$

therefore $k \circ F$ is a path homotopy between the paths $k \circ f$ and $k \circ g$.

∎

Definition 2.2: Path Homotopy class $[f]$

Let $f \in \mathrm{Path}(x_0, x_1)$, we define the *path homotopy class of f* as

$$[f] \triangleq \left\{ g \in \mathrm{Path}(x_0, x_1),\ g \simeq_p f \right\}$$

Definition 2.3: Product of PathClasses $[f] * [g]$

Let $* : \mathrm{PathClass}(x_0, x_1) \times \mathrm{PathClass}(x_1, x_2) \to \mathrm{PathClass}(x_0, x_2)$ be a binary operation, where

$$[f] * [g] \triangleq [f * g] \text{ is well defined.}$$

for arbitrary $[f] \in \mathrm{PathClass}(x_0, x_1)$ and $[g] \in \mathrm{PathClass}(x_1, x_2)$. This means it is independent of the representative chosen. More formally, if $f \simeq_p f' \in \mathrm{Path}(x_0, x_1)$, and $g \simeq_p g' \in \mathrm{Path}(x_1, x_2)$, then $f * g \simeq_p f' * g'$.

Proposition 2.2: Properties of the PathClass product

Let $[f]$, $[g]$ and $[h]$ be PathClasses from and to the points x_0, x_1, x_2. Then

1. Associativity: $([f] * [g]) * [h] = [f] * ([g] * [h])$,
2. Left and Right identities: if $[f] \in \mathrm{PathClass}(x_0, x_1)$, e_{x_0}, e_{x_1} denote the constant paths on x_0 and x_1 (the initial and final points of any $f \in [f]$), then

$$[e_{x_0}] * [f] = [f] \quad \text{and} \quad [f] * [e_{x_1}] = [f]$$

3. Left and Right inverses: let $[\overline{f}]$ be the PathClass containing the reversal of f (see Definition 1.1) for the definition, then

$$[\overline{f}] * [f] = [e_{x_1}] \quad \text{and} \quad [f] * [\overline{f}] = [e_{x_0}]$$

4. Generalized Associativity: if $\{[f_j]\}_{j \le n}$ is a sequence of PathClasses, such that $[f_j] \in \mathrm{PathClass}(x_{j-1}, x_j)$, then

$$\prod [f_j] \triangleq [f_1] * [f_2] * \cdots * [f_n] \text{ is a well-defined object}$$

meaning we can place the brackets wherever we want.

Proof. We will give an outline for the proof of Generalized Associativity, the rest are trivial. Let $\{[f_j]\}$ be defined as above. If $\{a_j\}_{j=0}^n$, and $\{b_j\}_{j=0}^n$ are 'cell partitions' of the unit interval (in the sense of the

Riemann integral), meaning

$$0 = a_0 < a_1 < \cdots < a_n = 1, \quad \text{and} \quad 0 = b_0 < b_1 < \cdots < b_n = 1$$

We agree to define the following

- the lengths of each cell $l_{a_j} \triangleq a_j - a_{j-1}$ and $l_{b_j} \triangleq b_j - b_{j-1}$, and
- the cells themselves are denoted by $\text{cell}(a_j) = [a_{j-1},\, a_j]$, $\text{cell}(b_j) = [b_{j-1},\, b_j]$,
- $p \in \text{Path}(0,1)$, where p is given explicitly by

$$p(s) = \sum_{j=1}^{n} \chi_{\text{cell}(a_j)\setminus\{a_{j-1}\}} \left(\frac{l_{b_j}}{l_{a_j}}(s - a_j) + b_j \right)$$

It is clear p is continuous, and for $j = 1, \ldots, n$,

$$p|_{\text{cell}(a_j)} \text{ is the positive linear map from } \text{cell}(a_j) \text{ to } \text{cell}(b_j)$$

Using the same line of argumentation as in the proof for associativity, we see that any two 'ways' of bracketing the expression has no impact on the path-homotopy class. ■

Fundamental Group

Definition 3.1: Fundamental group $\pi_1(\mathbf{X}, x_0)$

Let $x_0 \in \mathbf{X}$, the *fundamental group of* $\mathbf{X}$ *relative to (base point)* x_0 is denoted by $\pi_1(\mathbf{X}, x_0) = \text{PathClass}(x_0, x_0)$.

Definition 3.2: Isomorphism induced by $\text{Path}(x_0, x_1)$

Suppose $\alpha \in \text{Path}(x_0, x_1)$, we define a map $\hat{\alpha} : \pi_1(\mathbf{X}, x_0) \to \pi_1(\mathbf{X}, x_1)$, with

$$\hat{\alpha}([f]) = [\overline{\alpha}] * [f] * [\alpha]$$

where $\overline{\alpha}$ is the reversal of α. We call $\hat{\alpha}$ the *isomorphism induced by* α (Munkres Theorem 52.1).

Isomorphism proof. Let $[f]$ and $[g]$ be elements of $\pi_1(\mathbf{X}, x_0)$, then

$$\begin{aligned} \hat{\alpha}([f] * [g]) &= ([\overline{\alpha}] * [f] * [\alpha]) * ([\overline{\alpha}] * [g] * [\alpha]) \\ &= [\overline{\alpha}] * ([f] * [g]) * [\alpha] \\ &= \hat{\alpha}([f]) * \hat{\alpha}([g]) \end{aligned}$$

and $\hat{\alpha}$ is a homomorphism. We claim inverse of $\hat{\alpha}$ is $\hat{\overline{\alpha}}$. Fix $[f] \in \pi_1(\mathbf{X}, x_0)$, $[g] \in \pi_1(\mathbf{X}, x_1)$, then

$$(\hat{\overline{\alpha}} \circ \hat{\alpha})([f]) = [\alpha] * ([\overline{\alpha}] * [f] * [\alpha]) * [\overline{\alpha}] = [f]$$

so $\hat{\overline{\alpha}}$ is the left-inverse for $\hat{\alpha}$. A similar argument shows it is the right inverse as well with $(\hat{\alpha} \circ \hat{\overline{\alpha}})([g]) = [g]$. Therefore $\pi_1(\mathbf{X}, x_0)$ is group isomorphic to $\pi_1(\mathbf{X}, x_1)$. ■

Homomorphisms

Definition 4.1: Homomorphism induced by a continuous map

Let $h \in C(\mathbf{X},\ \mathbf{Y})$, and $y_0 = h(x_0)$, it induces a map between loops at x_0 and y_0.

$$h_* : \mathrm{Loop}(x_0) \to \mathrm{Loop}(y_0), f \mapsto h \circ f$$

It is a also a group homomorphism between fundamental groups. We use the same symbol for the two maps, relying on context to distinguish between the two.

$$h_* : \pi_1(\mathbf{X}, x_0) \to \pi_1(\mathbf{Y}, y_0), [f] \mapsto [h \circ f]$$

is well defined because of Proposition 2.2, it is a homomorphism (again by Proposition 2.2) because h 'distributes' over $*$

$$h \circ (f * g) = (h \circ f) * (h \circ g)$$

Remark 4.1: Functorial properties of the h_*

If $x_0 \in \mathbf{X}$, the tuple $(x_0, \mathbf{X})$ is an object in the category of *pointed topological spaces*, and the map h_* is a *covariant functor* from the category of pointed topological spaces to the category of groups.

Follows from Munkres Theorem 52.4, if the expressions below make sense,

$$(g \circ f)_* = g_* \circ f_* \quad \text{and} \quad h_* \circ (g \circ f)_* = (h \circ g)_* \circ f_*$$

And the identity map $i : \mathbf{X} \to \mathbf{X}$ gets 'sent' to the identity homomorphism in $\mathrm{Hom}(\pi_1(\mathbf{X}, x_0), \pi_1(\mathbf{X}, x_0))$. And if h is a homeomorphism between $\mathbf{X}$ and $\mathbf{Y}$, then h_* is an isomorphism at every point.

Simply connected space

Definition 5.1: Simply connected space

A topological space $\mathbf{X}$ is *simply connected* if it is path-connected, and $\pi_1(\mathbf{X}, x_0) = \{[e_{x_0}]\}$ for some $x_0 \in \mathbf{X}$. Notice this implies every fundamental group of $\mathbf{X}$ is trivial.

Proposition 5.1: Properties of simply connected spaces

If $\mathbf{X}$ is a simply connected space, then $\mathrm{PathClass}(x_0, x_1)$ consists of one element. That is to say, if f and g are Paths from x_0 to x_1, then $f \simeq_p g$.

Covering maps

Definition 6.1: Covering maps and spaces