Convex optimization
Topology and Manifolds
The Five Miracles of Mirror Descent

Hadar Tal

hadar.tal@mail.huji.ac.il

This paper is a summary of the educational materials and lectures from

- **Optimization for Computer Science** by Professor Tomer Koren, Tel Aviv University
- **Manifolds 1** by Dr. Julian P. Grossmann, "The Bright Side of Mathematics"
- **The Five Miracles of Mirror Descent** by Professor Sebastien Bubeck, Microsoft Research (Claire Boyer's notes)
- **Wikipedia**

Winter 2024

Contents

Chapter 1

Multivariable Calculus

Definition 1.0.1. *Diffrentiability, single variable*
Let $f : (a, b) \to \mathbb{R}$ be a function. We say that f is differentiable at $x_0 \in (a, b)$ if

$$\lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h} \tag{1.1}$$

exists. If f is differentiable at x_0, then $f'(x_0)$ is the derivative of f at x_0.

Definition 1.0.2. *Diffrentiability, single variable (alternative)*
Let $f : (a, b) \to \mathbb{R}$ be a function. We say that f is differentiable at $x_0 \in (a, b)$ if there exists a number m such that:

$$f(x_0 + h) = f(x_0) + m \cdot h + E(h) \text{ where } \lim_{h\to 0} \frac{E(h)}{h} = 0 \tag{1.2}$$

If f is differentiable at x_0, then $f'(x_0) = m$ is the derivative of f at x_0.

Suppose the $S \subseteq \mathbb{R}^n$ and $f : S \to \mathbb{R}$ is a function.

Definition 1.0.3. *Limit, multivariate function*
We say that the limit of f at x_0 is L if for all $\epsilon > 0$, there exists $\delta > 0$ such that for all x such that $||x - x_0|| < \delta$, we have $|f(x) - L| < \epsilon$.

Definition 1.0.4. *Diffrentiability, multivariable*
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function. We say that f is differentiable at x_0 if there exists a vector m $\in \mathbb{R}^n$ such that:

$$\lim_{h\to 0} \frac{f(x_0 + h) - f(x_0) - m \cdot h}{||h||} = 0 \tag{1.3}$$

If f is differentiable at x_0, then m is the gradient of f at x_0, denoted $\nabla f(x_0)$.

Definition 1.0.5. *Diffrentiability, multivariable (alternative)*
We say that f is differentiable at x_0 if there exists a vector $m \in \mathbb{R}^n$ such that:

$$f(x_0 + h) = f(x_0) + m^T \cdot h + E(h) \text{ where } \lim_{h\to 0} \frac{E(h)}{||h||} = 0 \tag{1.4}$$

If f is differentiable at x_0, then m is the gradient of f at x_0, denoted $\nabla f(x_0)$.

Definition 1.0.6. ***Differentiability of a function*** $f : \mathbb{R}^n \to \mathbb{R}^m$
Let $f : \mathbb{R}^n \to \mathbb{R}^m$ *be a function. We say that* f *is differentiable at* x_0 *if there exists a matrix* $A \in \mathbb{R}^{m\times n}$ *such that:*

$$\lim_{h\to 0} \frac{||f(x_0+h) - f(x_0) - A\cdot h||}{||h||} = 0 \tag{1.5}$$

If f *is differentiable at* x_0*, then* A *is the Jacobian matrix of* f *at* x_0*, denoted* $J_f(x_0)$*.*

Definition 1.0.7. *The Jacobian matrix*
The Jacobian matrix of $f : \mathbb{R}^n \to \mathbb{R}^m$ *at* x *is the matrix of partial derivatives of* f *at* x*:*

$$J = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} \nabla^T f_1 \\ \vdots \\ \nabla^T f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Example. *Let* $f : \mathbb{R}^2 \to \mathbb{R}^2$ *be a function defined by* $f(x,y) = (x^2, y^2)$*. Then the Jacobian matrix of* f *is:*

$$J_f(x,y) = \begin{bmatrix} 2x & 0 \\ 0 & 2y \end{bmatrix} \tag{1.6}$$

Definition 1.0.8. *Partial Derivative*
The partial derivative of f *with respect to the* i*-th variable at* x *is:*

$$\frac{\partial f}{\partial x_i}(x) = \lim_{h\to 0} \frac{f(x + h\cdot e_i) - f(x)}{h} \tag{1.7}$$

where e_i *is the* i*-th standard basis vector.*

Theorem 1.0.1. *(Diffrentiability vs. Partial Derivatives)*
If f *is differentiable at* x*, then all partial derivatives of* f *exist at* x *and:*

$$\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_n}(x)\right) \tag{1.8}$$

- If any partial derivative of f does not exist at x, then f is not differentiable at x.
- If all partial derivatives of f exist at x, then f may still not be differentiable at x and the vector $m = \nabla f(x)$ is the only possible vector that satisfies the definition of differentiability.

Definition 1.0.9. *The partial derivative of the function* f_i *with respect to the variable* x_j*, denoted by* $\frac{\partial f_i}{\partial x_j}$*, is defined as the limit*

$$\frac{\partial f_i}{\partial x_j} = \lim_{h\to 0} \frac{f_i(x_1, \ldots, x_j + h, \ldots, x_n) - f_i(x_1, \ldots, x_j, \ldots, x_n)}{h}$$

Definition 1.0.10. *Continuously Differentiable*
We say that f *is continuously differentiable or of class* C^1 *if all partial derivatives of* f *exist and are continuous at every point in* S*.*

Theorem 1.0.2. *If* f *is continuously differentiable, then* f *is differentiable.*

Definition 1.0.11. *The directional derivative*
For a given $x \in S$ and a unit vector $u \in \mathbb{R}^n$, the directional derivative of f at x in the direction of u is:

$$\partial_u f(x) = \lim_{h \to 0} \frac{f(x + h \cdot u) - f(x)}{h} \tag{1.9}$$

Equivalently, $\partial_u f(x) = g'(0)$ where $g(h) = f(x + h \cdot u)$.

Theorem 1.0.3. *If f is differentiable at x, then for all $u \in \mathbb{R}^n$, the directional derivative of f at x in the direction of u exists and is given by:*

$$\partial_u f(x) = \nabla f(x) \cdot u \tag{1.10}$$

Theorem 1.0.4. *Fermat's Theorem*
If f is differentiable at x and x is a local minimum of f, then $\nabla f(x) = 0$.

Definition 1.0.12. ***Differential of a function at a point*** *(in $\mathbb{R}^n$)*
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function of class C^1.
The differential of f at a point $p \in \mathbb{R}^n$ is the linear map $df_p : \mathbb{R}^n \to \mathbb{R}$ defined as

$$df_p(v) = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(p) \cdot v_i$$

for every $v = (v_1, \ldots, v_n) \in \mathbb{R}^n$.

The relation between the differential and the gradient:

$$df_p(v) = \nabla f(p) \cdot v \tag{1.11}$$

Recall that the gradient is the Jacobian matrix of the function $f : \mathbb{R}^n \to \mathbb{R}$, So in general the relation between the differential and the Jacobian matrix is:

$$df_p(v) = J_f(p) \cdot v \tag{1.12}$$

The range of the differential is the set of all linear functions of the form $df_p(v) = J_f(p) \cdot v$ for $v \in \mathbb{R}^n$, and is called the tangent space to the point p of the graph of the function f.

Theorem 1.0.5. *Suppose that $f : S \to \mathbb{R}$ is differentiable at x. Then $\nabla f(x)$ is orthogonal to the level set of f that passes through x.*

Theorem 1.0.6. *The mean value theorem*
If $f : S \to \mathbb{R}$ is differentiable on the open interval between a and b, then there exists $c \in [a, b]$ such that:

$$f(b) - f(a) = \nabla f(c) \cdot (b - a) \tag{1.13}$$

where $[a, b] = a + t(b - a)|t \in [0, 1]$.

Definition 1.0.13. *Second-order partial derivatives*
Suppose that f is a C^1 function. If the partial derivatives of f are differentiable, then the second-order partial derivatives of f are:

$$\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right) \tag{1.14}$$

Equivalently, $\frac{\partial^2 f}{\partial i \partial j} = \partial_j \partial_j f$. If $i = j$ we denote $\frac{\partial^2 f}{\partial x_i^2}$ or $(\partial_i^2 f$

Definition 1.0.14. *The C^2 class*
We say that f is of class C^2 if all second-order partial derivatives of f exist and are continuous.

Theorem 1.0.7. *Clairaut's Theorem*
If f is of class C^2, then $\frac{\partial^2 f}{\partial x_i \partial x_j} = \frac{\partial^2 f}{\partial x_j \partial x_i}$.

Definition 1.0.15. *Hessian Matrix*
The Hessian matrix of f at x is the matrix of second-order partial derivatives of f at x:

$$\nabla^2 f(x) = \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2} & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} & \frac{\partial^2 f}{\partial x_2^2} & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \frac{\partial^2 f}{\partial x_n \partial x_2} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{bmatrix} \tag{1.15}$$

Corollary. *The interpretation of the Hessian matrix*
Let $u \in \mathbb{R}^n$ be a unit vector. then

$$\partial^2_{uu} f(x) = \sum_{i,j=1}^{n} \partial_{ij} f(x) u_i u_j = u^T \nabla^2 f(x) u \tag{1.16}$$

We can generalize the Hessian matrix to functions $f : \mathbb{R}^n \to \mathbb{R}^m$ by defining the Hessian tensor.

Definition 1.0.16. ***Hessian Tensor***
The Hessian tensor of $f : \mathbb{R}^n \to \mathbb{R}^m$ at x is the tensor of second-order partial derivatives of f at x:

$$\mathbf{H}(x) = \begin{bmatrix} \nabla^2 f_1(x), & \nabla^2 f_2(x), & \ldots, & \nabla^2 f_m(x) \end{bmatrix} \tag{1.17}$$

where $\nabla^2 f_i(x)$ is the Hessian matrix of f_i at x.

1.1 Taylor series

Definition 1.1.1. *Taylor Series*
Let $f : \mathbb{R} \to \mathbb{R}$ be a function that is k times differentiable at x_0. Then the Taylor series of f at x_0 is given by:

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \frac{f''(x_0)}{2!}(x - x_0)^2 + \ldots + \frac{f^{(k)}(x_0)}{k!}(x - x_0)^k + R_k(x) \tag{1.18}$$

where $R_k(x) = \frac{f^{(k+1)}(c)}{(k+1)!}(x - x_0)^{k+1}$ for some c between x and x_0.

Definition 1.1.2. *Taylor Series for Multivariable Functions (k=2)*
Let $f : \mathbb{R}^n \to \mathbb{R}$ be a function that is C^2 at x_0. Then for any h such that $x_0 + h \in S$, there exists $\theta \in [0, 1]$ such that:

$$f(x_0 + h) = f(x_0) + \nabla f(x_0) \cdot h + \frac{1}{2} h^T \nabla^2 f(x_0 + \theta h) h \tag{1.19}$$

Chapter 2

Algebraic Structures

2.0.1 Properties

Definition 2.0.1. *Closure (sgirot)*
An operation $*$ on a set G is said to have the property of closure if for every $a, b \in G$, the result $a * b$ is also in G.

Definition 2.0.2. *Commutativity (hilofiot)*
An operation $*$ on a set G is commutative if for every $a, b \in G$, we have $a * b = b * a$.

Definition 2.0.3. *Associativity*
An operation $*$ on a set G is associative if for every $a, b, c \in G$, we have $(a * b) * c = a * (b * c)$.

Definition 2.0.4. *Distributivity*
An operation $*$ on a set G is distributive if for every $a, b, c \in G$, we have $a * (b + c) = a * b + a * c$.

Definition 2.0.5. *Identity (zehot)*
An operation $*$ on a set G has an identity element if there exists an element $e \in G$ such that for every $a \in G$, $a * e = e * a = a$.

Definition 2.0.6. *Inverse (ofchiot)*
An operation $*$ on a set G has inverses if for every $a \in G$, there exists an element $b \in G$ such that $a * b = b * a = e$, where e is the identity element.

2.0.2 Structures

Group

Definition 2.0.7. ***Group*** *(havura)*
A group is a set G along with an operation $*$ such that $\forall a, b, c \in G$ the following properties hold:

1. *$a * b \in G$ (closure)*
2. *$(a * b) * c = a * (b * c)$ (associativity)*
3. *There exists an element $e \in G$ such that $a * e = e * a = a$ (identity)*
4. *For each $a \in G$ there exists $b \in G$ such that $a * b = b * a = e$ (inverse)*

Example. *Examples of groups:*

1. *$(\mathbb{R}, +)$ is a group.*
2. *$(\mathbb{Z}, +)$ is a group.*
3. *Non-zero reals, complex, and rational numbers are groups under multiplication.*

Definition 2.0.8. ***Homomorphism (Groups)***
Let $(G, \cdot)$ and $(H, *)$ be groups. A map $h : G \to H$ is called a homomorphism if it preserves the group operation, meaning that:

$$\forall g_1, g_2 \in G, \quad h(g_1 \cdot g_2) = h(g_1) * h(g_2)$$

Example. *Let $G = (\mathbb{R}, +)$ and $H = (\mathbb{R}_{>0}, \cdot)$.*
The map $h : \mathbb{R} \to \mathbb{R}_{>0}$ defined by $h(x) = e^x$ is a homomorphism.

$$h(x + y) = e^{x+y} = e^x \cdot e^y = h(x) \cdot h(y)$$
$$h(x \cdot y) = e^{x \cdot y} = e^x \cdot e^y = h(x) \cdot h(y)$$

Definition 2.0.9. ***Abelian Group***
An abelian group is a group $(G, *)$ in which the binary operation $*$ is commutative, meaning that for all $a, b \in G$, $a * b = b * a$.

Ring

Definition 2.0.10. ***Ring*** *(hug)*
A ring is a set R equipped with two binary operations $+$ (addition) and $\times$ (multiplication) satisfying the following three sets of axioms:

1. *R is an **abelian group** under addition, meaning that:*
   - *$(a + b) + c = a + (b + c)$ for all $a, b, c \in R$ (associativity).*
   - *$a + b = b + a$ for all $a, b \in R$ (commutativity).*
   - *There is an element $0 \in R$ such that $a + 0 = a$ for all $a \in R$ (additive identity).*
   - *For each $a \in R$ there exists $-a \in R$ such that $a + (-a) = 0$ (additive inverse).*
2. *R is a **monoid** under multiplication, meaning that:*
   - *$(a \times b) \times c = a \times (b \times c)$ for all $a, b, c \in R$ (associativity).*
   - *There is an element $1 \in R$ such that $a \times 1 = a$ and $1 \times a = a$ for all $a \in R$ (multiplicative identity).*
3. *Multiplication is distributive with respect to addition, meaning that:*
   - *$a \times (b + c) = (a \times b) + (a \times c)$ for all $a, b, c \in R$ (left distributivity).*
   - *$(b + c) \times a = (b \times a) + (c \times a)$ for all $a, b, c \in R$ (right distributivity).*

Example. *Examples of rings:*

1. *$(\mathbb{Z}, +, \times)$ is a ring.*
2. *$(\mathbb{R}, +, \times)$ is a ring.*
3. *The set of odd integers is not a ring because it is not closed under addition.*

Field

Definition 2.0.11. ***Field*** *(sadeh)*
A field is a set F with two operations, addition $+$ and multiplication $\times$, such that:

1. *$(F, +)$ is an **abelian group** with the identity element 0 (additive identity).*
2. *$(F \setminus \{0\}, \times)$ is an **abelian group** with the identity element 1 (multiplicative identity).*

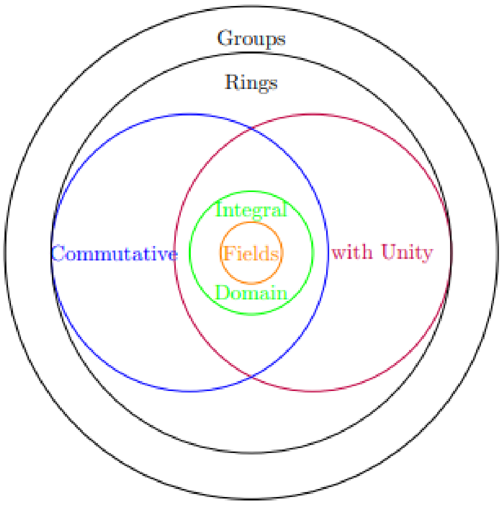


Figure 2.1: Algebraic Structures

3. *Multiplication is distributive with respect to addition, meaning that:*
   - $a \times (b + c) = (a \times b) + (a \times c)$ *for all* $a, b, c \in R$ *(left distributivity).*
   - $(b + c) \times a = (b \times a) + (c \times a)$ *for all* $a, b, c \in R$ *(right distributivity).*

Example. *Examples of fields:*

1. $(\mathbb{R}, +, \times)$ *is a field.*
2. $(\mathbb{Q}, +, \times)$ *is a field.*
3. $(\mathbb{C}, +, \times)$ *is a field.*
4. $(\mathbb{Z}_p, +, \times)$ *for a prime* p *is a field.*

2.0.3 Spaces

Vector Space

A major difference between a field and a vector space is that the operations on a field $\mathbb{F}$ are

- $+ : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$
- $\times : \mathbb{F} \times \mathbb{F} \to \mathbb{F}$

but the operations on a vector space $\mathbb{V}$ over a field $\mathbb{F}$ are

- $+ : \mathbb{V} \times \mathbb{V} \to \mathbb{V}$
- $\cdot : \mathbb{F} \times \mathbb{V} \to \mathbb{V}$

Definition 2.0.12. ***Vector Space***
A vector space over a field F *is a non-empty set* V *together with two operations: vector addition* $+$ *and scalar multiplication* $\cdot$*, satisfying the following axioms for every* $u, v, w \in V$ *and* $a, b \in F$*:*

1. *Associativity of vector addition:* $u + (v + w) = (u + v) + w$
2. *Commutativity of vector addition:* $u + v = v + u$

3. *Identity element of vector addition: There exists an element* $0 \in V$*, called the* ***zero vector****, such that* $v + 0 = v$ *for all* $v \in V$*.*

4. *Inverse elements of vector addition: For every* $v \in V$*, there exists an element* $-v \in V$*, called the* ***additive inverse*** *of* v*, such that* $v + (-v) = 0$*.*

5. *Compatibility of scalar multiplication with field multiplication:* $a(bv) = (ab)v$

6. *Identity element of scalar multiplication:* $1v = v$*, where* 1 *denotes the multiplicative identity in* F*.*

7. *Distributivity of scalar multiplication with respect to vector addition:* $a(u + v) = au + av$

8. *Distributivity of scalar multiplication with respect to field addition:* $(a + b)v = av + bv$

Example. *Examples of vector spaces:*

1. $\mathbb{R}^n$ *is a vector space.*

2. *The space* $M_{m\times n}(\mathbb{F})$ *of* $m \times n$ *matrices over a field* $\mathbb{F}$ *is a vector space.*

3. *The set of all continuous functions over some interval is a vector space.*

4. *The space of all differentiable functions over a certain interval is a vector space.*

Definition 2.0.13. ***Homomorphism (Vector spaces)***
Let V *and* W *be vector spaces over the same field* F*. A map* $T : V \to W$ *is called a* ***homomorphism****, or more specifically, a* ***linear transformation****, if for all vectors* $u, v \in V$ *and any scalar* $c \in F$*, the following conditions hold:*

- ***Additivity:*** $T(u + v) = T(u) + T(v)$*.*

- ***Homogeneity:*** $T(c \cdot u) = c \cdot T(u)$*.*

These conditions ensure that the map T *preserves the vector space structure between* V *and* W*.*

If a homomorphism is bijective, it is called an isomorphism.

Definition 2.0.14. ***Isomorphism (Vector spaces)***
Let V *and* W *be vector spaces over the same field* F*. A map* $T : V \to W$ *is called an* ***isomorphism*** *if it is a bijective* ***linear transformation****, meaning that* T *is both a homomorphism and has an inverse* $T^{-1} : W \to V$ *which is also a homomorphism. For* T *to be an isomorphism, the following conditions must be met:*

- ***Bijectivity:*** T *is one-to-one and onto.*

- ***Additivity:*** $T(u + v) = T(u) + T(v)$ *for all vectors* $u, v \in V$*.*

- ***Homogeneity:*** $T(c \cdot u) = c \cdot T(u)$ *for all vectors* $u \in V$ *and any scalar* $c \in F$*.*

An isomorphism thus establishes a perfect correspondence between the two vector spaces, preserving all vector space operations.

If V and W are isomorphic, we write $V \cong W$.

Definition 2.0.15. ***Complex conjugate***
The complex conjugate of a complex number $z = a + bi$ *is the number* $\overline{z} = a - bi$*.*

Inner Product Space

Definition 2.0.16. ***Inner Product Space***
An inner product space is a vector space V over a field F equipped with an inner product, which is a function that associates each pair of vectors u, v in V with a scalar in F, denoted $\langle u, v\rangle$, and satisfies the following properties for all $u, v, w \in V$ and $a \in F$:

1. *Linearity in the first argument:* $\langle au + v, w\rangle = a\langle u, w\rangle + \langle v, w\rangle$
2. *Conjugate symmetry:* $\langle u, v\rangle = \overline{\langle v, u\rangle}$
3. *Positive-definiteness:* $\langle u, u\rangle \geq 0$ *and* $\langle u, u\rangle = 0$ *if and only if* $u = 0$

Definition 2.0.17. ***Hermetian adjoint***
Let V be an inner product space over a field F. The Hermetian adjoint of a linear operator $T : V \to V$ is the unique linear operator $T^* : V \to V$ such that for all $u, v \in V$, we have $\langle Tu, v\rangle = \langle u, T^*v\rangle$.

A classic exmaple of an inner product space is the **Euclidean space** $\mathbb{E}^n$, which is a vector space equipped with the inner product $\langle u, v\rangle = u^T v$. The geometry of Euclidean space follows the familiar rules of Euclidean geometry, which include notions such as angles, lengths, and the Pythagorean theorem. It is always complete, meaning that every Cauchy sequence in Euclidean space converges to a point within the space. Euclidean space can be thought of as the "standard" n-dimensional space that conforms to our intuitive geometric concepts.

Chapter 3

Convexity and Optimization

3.1 Important subsets of $\mathbb{R}^n$

Definition 3.1.1. *Open set*
A set $S \subseteq \mathbb{R}^n$ is open if for all $x \in S$, there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq S$.

Definition 3.1.2. *Closed set*
A set $S \subseteq \mathbb{R}^n$ is closed if its complement is open.

Definition 3.1.3. *Interior point*
A point $x \in S$ is an interior point of S if there exists $\epsilon > 0$ such that $B(x, \epsilon) \subseteq S$.

Corollary 3.1.1. *Open set characterization*
A set $S \subseteq \mathbb{R}^n$ is open if and only if every point in S is an interior point of S.

Definition 3.1.4. *Boundary point*
A point $x \in S$ is a boundary point of S if for all $\epsilon > 0$, $B(x, \epsilon) \cap S \neq \emptyset$ and $B(x, \epsilon) \cap S^c \neq \emptyset$.

Definition 3.1.5. *Half-space*
A half-space in $\mathbb{R}^n$ is a set of the form $\{x \in \mathbb{R}^n : a^T x \leq b\}$ for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Definition 3.1.6. *Hyperplane*
A hyperplane in $\mathbb{R}^n$ is a set of the form $\{x \in \mathbb{R}^n : a^T x = b\}$ for some $a \in \mathbb{R}^n$ and $b \in \mathbb{R}$.

Definition 3.1.7. *Polyhedron (Polyhedra)*
A polyhedron in $\mathbb{R}^n$ is a set of the form $\{x \in \mathbb{R}^n : Ax \leq b\}$ for some $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Equivalently, a polyhedron is the intersection of finitely many half-spaces.

Definition 3.1.8. *Polytope*
A polytope in $\mathbb{R}^n$ is a bounded polyhedron - i.e., there exists $r > 0$ such that $\forall x \in \{x \in \mathbb{R}^n : Ax \leq b\} \implies ||x|| \leq r$. Equivalently, a polytope is the convex hull of finitely many points.

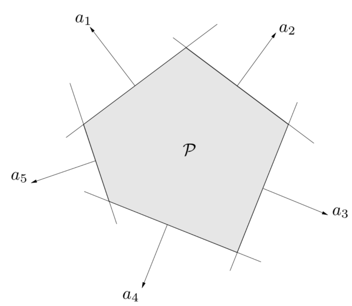


Figure 3.1: Polytope

Definition 3.1.9. *Convex set*
A set $S \subseteq \mathbb{R}^n$ is convex if for all $x, y \in S$ and $\lambda \in [0,1]$, we have $\lambda t + (1-\lambda)y \in S$.

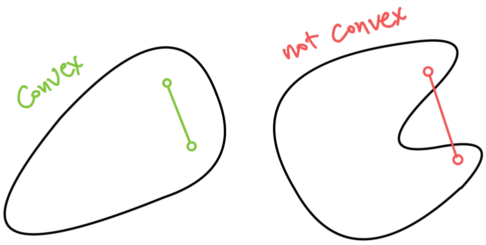


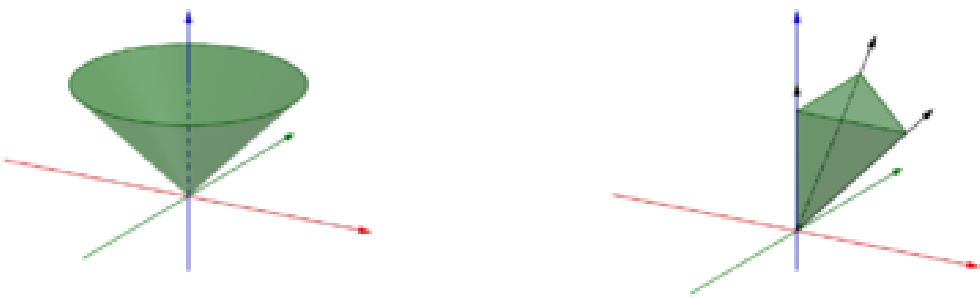
Figure 3.2: Convex set

Definition 3.1.10. *Convex hull*
The convex hull of a set $S \subseteq \mathbb{R}^n$ is the smallest convex set that contains S.

Definition 3.1.11. *Conic combination*
A point $x \in \mathbb{R}^n$ is a conic combination of $y_1, \ldots, y_k \in \mathbb{R}^n$ if there exist $\lambda_1, \ldots, \lambda_k \geq 0$ such that $x = \sum_{i=1}^k \lambda_i y_i$.

Definition 3.1.12. *Conic hull*
The conic hull of a finite set $S \subseteq \mathbb{R}^n$ is the set of all conic combinations of points in S.

Definition 3.1.13. *Convex cone*
A set $S \subseteq \mathbb{R}^n$ is a convex cone if for all $x \in S$ and $\lambda \geq 0$, we have $\lambda x \in S$.



(a) Convex cone that is not a conic hull of finitely many generators.

(b) Convex cone genrated by the conic combination of three black vectors (conic hull).

Definition 3.1.14. *Normal cone*
The normal cone to a set S at a point x is defined as

$$N_S(x) = \{v \in \mathbb{R}^n : \langle v, y - x\rangle \leq 0 \text{ for all } y \in S\} \tag{3.1}$$

Definition 3.1.15. *Tangent cone*
The tangent cone to a set S at a point x is defined as

$$T_S(x) = \{v \in \mathbb{R}^n : \lim_{t \to 0^+} \frac{x + tv - x}{t} \in S\} \tag{3.2}$$

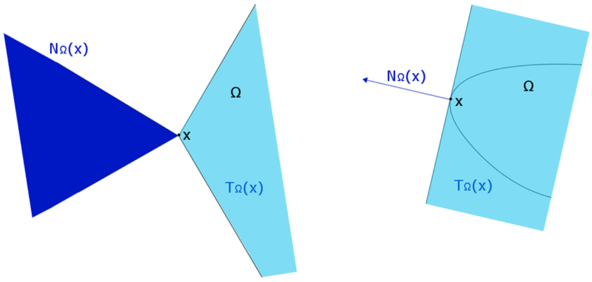


Figure 3.4: Normal and tangent cones

Theorem 3.1.1. *Normal cone of polyhedron*
The normal cone to a polyhedron $S = \{x \in \mathbb{R}^n : \forall j \in [m] \quad a_j \cdot x \leq b_j\}$ *at a point* x *is given by*

$$N_S(x) = \{\sum_j \lambda_j a_j : \lambda_j \geq 0 \text{ and } a_j \cdot x = b_j\} \tag{3.3}$$

3.2 Definitions and Fundamental Theorems

Definition 3.2.1. *(Convex function): A function* $f : S \to \mathbb{R}$ *defined on a convex set* S *is convex if, for all* $x, y \in S$ *and* $\lambda \in [0, 1]$,

$$f(\lambda x + (1 - \lambda)y) \leq \lambda f(x) + (1 - \lambda)f(y).$$

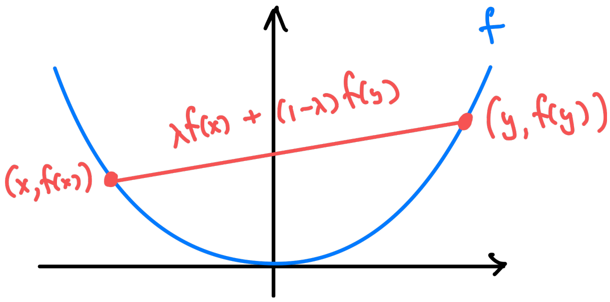


Figure 3.5: Convex function

Theorem 3.2.1. *(Characterization via epigraph): A function* $f : S \to \mathbb{R}$ *is convex if and only if its epigraph* $\{(x, t) \in S \times \mathbb{R} : f(x) \leq t\}$ *is a convex set.*

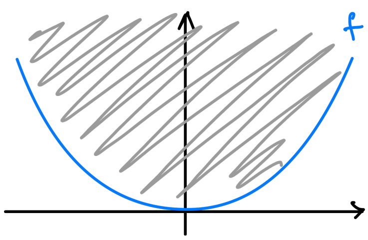


Figure 3.6: Epigraph of a convex function

claim 3.2.1. *(Convexity of sublevel sets): If* $f : S \to \mathbb{R}$ *is convex, then the sublevel set* $S_t = \{x \in S : f(x) \leq t\}$ *is convex for any* $t \in \mathbb{R}$.

3.3 Inequalities and Characterizations

Theorem 3.3.1. *(Jensen's inequality): If f is a convex function, then for any $x_1, x_2, \ldots, x_n \in S$ and any non-negative weights α_i such that $\sum_{i=1}^n \alpha_i = 1$,*

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right) \leq \sum_{i=1}^{n} \alpha_i f(x_i).$$

Theorem 3.3.2. *(First-order characterization, aka "the gradient inequality"): If f is a differentiable convex function on an open set S, then for all $x, y \in S$,*

$$f(y) \geq f(x) + \nabla f(x)^\top (y - x).$$

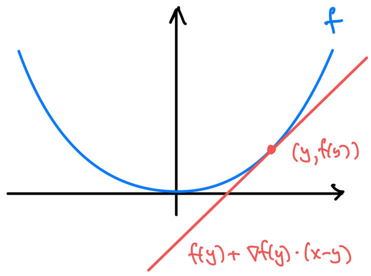


Figure 3.7: First-order characterization of convexity

Definition 3.3.1. *Bergman divergence (distance)*
The Bergman divergence between two points $x, y \in \mathbb{R}^n$ is defined as

$$D_f(x, y) = f(x) - f(y) - \nabla f(y)^\top (x - y) \tag{3.4}$$

Theorem 3.3.3. *(Jensen's inequality, generalized for expectation): If f is a convex function and X is a random variable over S, then*

$$f(\mathbb{E}[X]) \leq \mathbb{E}[f(X)].$$

Theorem 3.3.4. *(Second-order characterization of convexity): A twice differentiable function f is convex on an open set S if and only if the Hessian matrix of f is positive semidefinite at every point in S.*

3.4 Optimization and Projection

Definition 3.4.1. *(Convex optimization): The problem of minimizing a convex function over a convex set.*

Theorem 3.4.1. *(Optimality conditions, unconstrained): If f is convex and differentiable, x^* is a local minimum of f $\Leftrightarrow$ x^* is a global minimum of f $\Leftrightarrow$ $\nabla f(x^*) = 0$.*

Theorem 3.4.2. *(Optimality conditions, constrained): If f is differentiable and C is a convex set, x^* is a local minimum of f on C if and only if $\langle \nabla f(x^*), x - x^* \rangle \geq 0$ for all $x \in C$.*

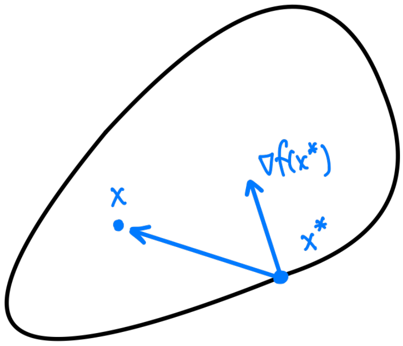


Figure 3.8: Optimality conditions, constrained

Corollary 3.4.1. *Optimality conditions, constrained (alternative)*
If f is differentiable and C is a convex set, then x^* is a local minimum of f on C if and only if $-\nabla f(x^*) \in N_C(x^*)$.

Definition 3.4.2. *(Projection): The projection of a point x onto a convex set S is defined as* $\Pi_S(x) = \arg\min_{y \in S} \|y - x\|$.

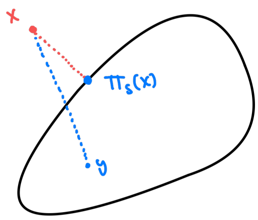


Figure 3.9: Projection

Theorem 3.4.3. *Generalized cosine theorem*
Let $S \subseteq \mathbb{R}^d$ be convex and $x \in \mathbb{R}^d$. Then the projection $\Pi_S[x]$ is unique and satisfies:

$$\|x - \Pi_S[x]\|^2 + \|\Pi_S[x] - y\|^2 \leq \|x - y\|^2, \quad \forall y \in S. \tag{3.5}$$

In particular:

$$\|\Pi_S[x] - y\| \leq \|x - y\|, \quad \forall y \in S. \tag{3.6}$$

3.5 Properties of Convex Functions

Definition 3.5.1. *L - Lipschitz continuous*
A function $f : S \to \mathbb{R}$ is L-Lipschitz continuous if for all $x, y \in S$,

$$|f(x) - f(y)| \leq L||x - y|| \tag{3.7}$$

Theorem 3.5.1. *Convexity and Lipschitz continuity*
If f is convex, differentiable and L-Lipschitz continuous, then $||\nabla f(x)|| \leq L$ for all $x \in S$.

Definition 3.5.2. *Smooth function*
A differentiable function f is β-smooth over $S \subseteq domf$ if for all $x, y \in S$:

$$-\frac{\beta}{2}\|y-x\|^2 \leq f(y) - f(x) - \nabla f(x) \cdot (y-x) \leq \frac{\beta}{2}\|y-x\|^2.$$

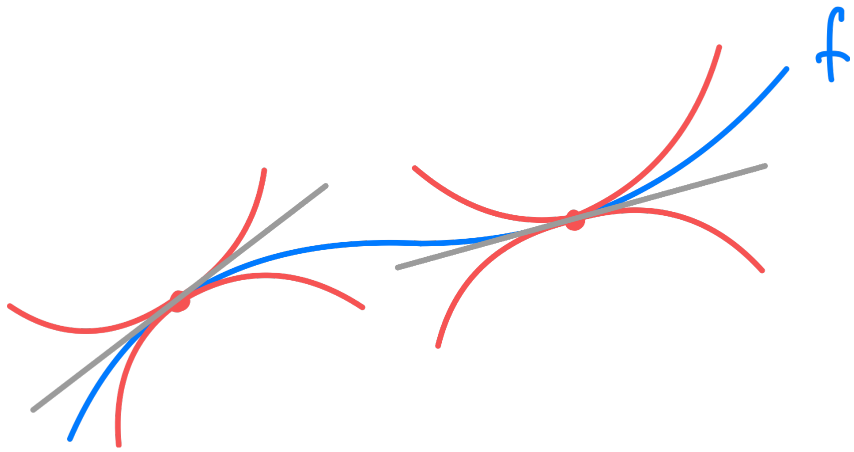


Figure 3.10: Smooth function

Theorem 3.5.2. *Lipschitz gradient interpretation*
Let f be differentiable and let $S \subseteq domf$ be convex and closed. Suppose that

$$\|\nabla f(x) - \nabla f(y)\| \leq \beta\|x-y\|, \quad \forall x, y \in S.$$

Then f is β-smooth over S.

Theorem 3.5.3. *Second-order characterization of smoothness*
Let f be C^2 and let $S \subseteq domf$ be convex and closed. Then f is β-smooth over S if and only if

$$-\beta I \preceq \nabla^2 f(x) \preceq \beta I, \quad \forall x \in S.$$

Lemma 3.5.1. *The Descent Lemma*
Let $f : \mathbb{R}^d \to \mathbb{R}$ be β-smooth, and let $x \in \mathbb{R}^d$.

- For $\eta \leq \frac{1}{\beta}$, $x^+ = x - \eta\nabla f(x)$, we have

$$f(x^+) - f(x) \leq -\frac{\eta}{2}\|\nabla f(x)\|^2.$$

- For $x^* \in \arg\min_x f(x)$, we have

$$\frac{1}{2\beta}\|\nabla f(x)\|^2 \leq f(x) - f(x^*).$$

Basic Facts:

- An affine function $f : \mathbb{R}^d \to \mathbb{R}, f(x) = a^\top x + b$, is 0-smooth.
- A quadratic function $f : \mathbb{R}^d \to \mathbb{R}, f(x) = \frac{1}{2}x^\top Ax + b^\top x + c$, is $\lambda_{\max}(A)$-smooth.

- A linear combination of smooth functions is smooth with an appropriate parameter.
- A convex combination of β-smooth functions is β-smooth.

Definition 3.5.3. *Strong convexity*
A function f is α-strongly convex (for $\alpha \geq 0$) over a convex and closed set $S \subseteq domf$ if for any $x \in S$, there exists $g_x \in \partial f(x)$ such that:

$$\forall y \in S, \quad f(y) \geq f(x) + g_x \cdot (y-x) + \frac{\alpha}{2}\|y-x\|^2.$$

In particular, a differentiable f is α-strongly convex over S if for any $x \in S$,

$$\forall y \in S, \quad f(y) \geq f(x) + \nabla f(x) \cdot (y-x) + \frac{\alpha}{2}\|y-x\|^2.$$

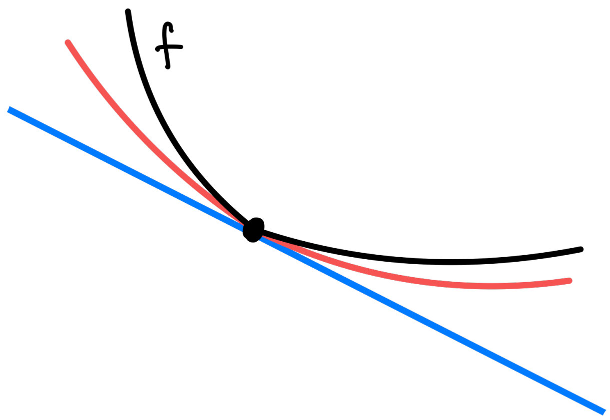


Figure 3.11: Strongly convex function

Theorem 3.5.4. *Strong convexity, second-order characterization*
Let f be C^2 and let $S \subseteq domf$ be convex and closed. Then f is α-strongly convex over S if and only if

$$\forall x \in S, \quad \nabla^2 f(x) \succeq \alpha I.$$

Theorem 3.5.5. *Usage of strong convexity*
If a differentiable f is α-strongly convex over a convex and closed $S \subseteq domf$ with a minimum at $x^* \in S$, then

$$\forall x \in S, \quad \frac{\alpha}{2}\|x-x^*\|^2 \leq f(x) - f(x^*) \leq \frac{1}{2\alpha}\|\nabla f(x)\|^2.$$

In particular, the minimum of a strongly convex function is unique.

3.6 Important Inequalities

Theorem 3.6.1. $1 + x \leq e^x$
For all $x \in \mathbb{R}$, we have $1 + x \leq e^x$.

Proof. Let $f(x) = e^x - 1 - x$. Then $f'(x) = e^x - 1$ and $f''(x) = e^x > 0$. Thus, f is convex and $f(0) = 0$. Therefore, $f(x) \geq 0$ for all $x \in \mathbb{R}$. $\square$

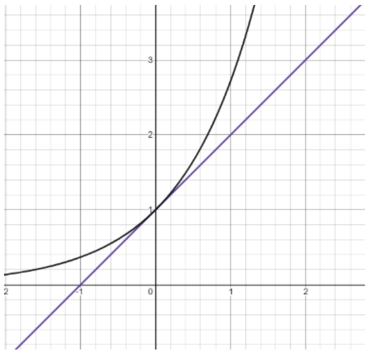


Figure 3.12: $1 + x \leq e^x$

Chapter 4

Topology and manifolds

4.1 Metric space and complete metric space

Definition 4.1.1. ***Metric Space***
A metric space is a set X equipped with a metric, which is a function that defines a distance between each pair of elements in X, satisfying the following properties for all $x, y, z \in X$:

1. *Non-negativity: $d(x, y) \geq 0$ and $d(x, y) = 0$ if and only if $x = y$*
2. *Symmetry: $d(x, y) = d(y, x)$*
3. *Triangle inequality: $d(x, y) + d(y, z) \geq d(x, z)$*

Inner Product Space vs Metric Space

- An inner product space is a vector space equipped with an inner product, which is a function that associates each pair of vectors with a scalar.
- A metric space is a set equipped with a metric, which is a function that defines a distance between each pair of elements in the set.
- Every inner product space is a metric space, but not every metric space is an inner product space.

  Example. *Each inner product space must satisfy the Parallelogram Law, which states that for all u, v in the space, $2\|u\|^2 + 2\|v\|^2 = \|u + v\|^2 + \|u - v\|^2$. A clasic example of a metric space that is not an inner product space is the space of continuous functions on the interval $[0, 1]$ with the metric $d(f, g) = \max_{x \in [0,1]} |f(x) - g(x)|$.*

Definition 4.1.2. ***Open ϵ-ball***
Let (X, d) be a metric space and $x \in X$.
The open ε-ball centered at x is the set $B(x, \varepsilon) = \{y \in X : d(x, y) < \varepsilon\}$.

Definition 4.1.3. ***Open set*** *(metric space)*
A subset U of a metric space (X, d) is open if for every point $x \in U$, there exists an open ε-ball centered at x that is contained in U.

Definition 4.1.4. ***Convergence of a sequence***
A sequence (x_n) in a metric space (X, d) is said to converge *to a limit $a \in X$ if, for every positive real number $\varepsilon > 0$, there exists a positive integer N such that for all positive integers $n \geq N$, $x_n \in B_\varepsilon(a)$.*

Definition 4.1.5. ***Cauchy Sequence***
In a metric space (X, d), a sequence $\{x_1, x_2, x_3, \ldots\}$ is said to be Cauchy *if, for every positive real number $\varepsilon > 0$, there exists a positive integer N such that for all positive integers $m, n > N$, the distance*

$$d(x_m, x_n) < \varepsilon.$$

(a) The plot of a Cauchy sequence (x_n) shown in blue, as x_n versus n. If the space is complete, then the sequence has a limit.

(b) A sequence that is not Cauchy. The elements of the sequence do not get arbitrarily close to each other as the sequence progresses.

Figure 4.1: Cauchy Sequence

Definition 4.1.6. ***Complete Metric Space***
A metric space is complete if every Cauchy sequence in the space converges to a limit in the space.

Hilbert Space

Definition 4.1.7. ***Hilbert Space***
A Hilbert space *is a complete inner product space. That is, it is an inner product space $\mathcal{H}$ that is also a complete metric space with respect to the metric induced by its inner product. The metric is given by* $d(x, y) = \sqrt{\langle x - y, x - y\rangle}$ *for all $x, y \in \mathcal{H}$.*

A Hilbert space generalizes the notion of Euclidean space to an infinite-dimensional context. In a Hilbert space, one can still use concepts like angle, orthogonality, and projection, which are crucial in many areas of mathematics and physics. Completeness is a key feature of Hilbert spaces, meaning that every Cauchy sequence in the space converges to a limit within the space.

4.2 Topological Space

Definition 4.2.1. ***Topology***
A topology $\mathcal{T}$ on a set X is a collection of subsets of X that satisfy the following properties:

1. $\emptyset, X \in \mathcal{T}$.
2. *The intersection of any finite number of sets is in $\mathcal{T}$ -*
   $U_1, U_2, \ldots, U_n \in \mathcal{T}$ *implies* $\bigcap_{i=1}^{n} U_i \in \mathcal{T}$.
3. *The union of any number of sets (finite or infinite) is in $\mathcal{T}$ -*
   $U_\alpha \in \mathcal{T}$ *for all α in some index set A implies* $\bigcup_{\alpha \in A} U_\alpha \in \mathcal{T}$.

Definition 4.2.2. ***Topological Space***
A topological space is a pair $(X, \mathcal{T})$ consisting of a set X and a topology $\mathcal{T}$ on X.

A metric space (X, d) induces a topology on X by defining the open sets to be the unions of open ε-balls centered at each point in X.

Example. *Examples of topological spaces:*

1. $X = 1, 2, 3$ *with the topology* $T = \{\varnothing, \{1\}, \{1, 2\}, \{1, 2, 3\}\}$.

2. *Given any set X, **the discrete topology** on X is the power set of X - $\mathcal{T} = P(X)$ and is the largest possible topology on X.*

3. ***The indiscrete topology** on X is $\mathcal{T} = \{\emptyset, X\}$ and is the smallest possible topology on X - $\forall \mathcal{T}$, $\{\emptyset, X\} \subseteq \mathcal{T} \subseteq P(X)$.*

4. *The real line $\mathbb{R}$ with the standard topology.*

5. *The set of integers $\mathbb{Z}$ with the discrete topology.*

6. *The set of real numbers $\mathbb{R}$ with the lower limit topology.*

Definition 4.2.3. ***Open Set (topological space)***
A subset U of a topological space $(X, \mathcal{T})$ is open if $U \in \mathcal{T}$.

Definition 4.2.4. ***Metrizible Space***
A topological space $(X, \mathcal{T})$ is metrizable if there exists a metric d on X such that the topology induced by d is equal to $\mathcal{T}$.

4.3 Interior, exterior, boundary, and accumulation points

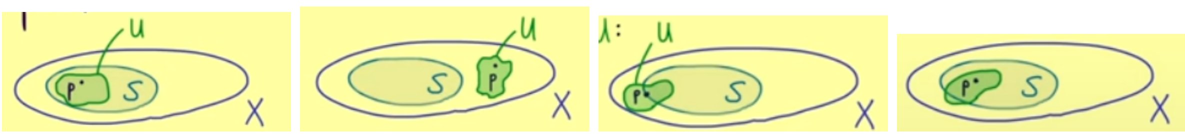
Let $(X, \mathcal{T})$ be a topological space , and $S \subseteq X$ (it can be an open set but it not necessarily has to be).

Definition 4.3.1. ***Interior point***
A point $p \in S$ is an interior point of S if there exists an open set U such that $p \in U \subseteq S$.

Definition 4.3.2. ***Exterior point***
A point $p \in X$ is an exterior point of S if there exists an open set U such that $p \in U \subseteq X \setminus S$.

Definition 4.3.3. ***Boundary point***
A point $p \in X$ is a boundary point of S if for every open set U containing p, both $U \cap S$ and $U \cap (X \setminus S)$ are non-empty.

Definition 4.3.4. ***Accumulation (limit) point***
A point $p \in X$ is an accumulation point of S if for every open set U containing p, the set $(U \setminus \{p\}) \cap S$ is non-empty.



(a) Interior point (b) Exterior point (c) Boundary point (d) Accumulation point

Figure 4.2: Interior, exterior, and boundary points of a set

Definition 4.3.5. ***Interior of a set***
The interior of a set S, denoted by $int(S)$ or S°, is the set of all interior points of S.

$$S^\circ = \{p \in S : p \text{ is an interior point of } S\}$$

Definition 4.3.6. ***Exterior of a set***
The exterior of a set S, denoted by $ext(S)$, is the set of all exterior points of S.

$$ext(S) = \{p \in X : p \text{ is an exterior point of } S\}$$

Definition 4.3.7. ***Boundary of a set***
The boundary of a set S, denoted by $bd(S)$ or ∂S, is the set of all boundary points of S.

$$\partial(S) = \{p \in X : p \textit{ is a boundary point of } S\}$$

The boundary of a set S consists of all points that are either in the interior of S or in the exterior of S.

Definition 4.3.8. ***Derived set*** *(derivative of a set)*
The derived set of a set S, denoted by S', is the set of all accumulation points of S.

$$S' = \{p \in X : p \textit{ is an accumulation point of } S\}$$

Definition 4.3.9. ***Closure of a set***
The closure of a set S, denoted by $\overline{S}$, is the set of all points in X that are either in S or are accumulation points of S.

$$\overline{S} = S \cup S'$$

Example. $X = \mathbb{R}$, $\mathbb{T} = \{\emptyset, \mathbb{R}\} \cup \{(a, \infty) : a \in \mathbb{R}\}$ *is a topology on* $\mathbb{R}$.
$S = (0, 1)$ *is not an open set, and do not have any interior points.*
$X \setminus S = (-\infty, 0] \cup [1, \infty) \Rightarrow ext(S) = (1, \infty) \Rightarrow \partial S = [-\infty, 1)$.*)*

Definition 4.3.10. ***Closed set***
Let $(X, \mathcal{T})$ be a topological space. A set $S \subseteq X$ is closed if any of the following equivalent conditions hold:

- *A set S is closed if it contains all of its boundary points.*
- *A set S is closed if its complement $X \setminus S$ is open.*
- *A set S is closed if it is equal to its closure $\overline{S}$.*

4.4 Hausdorff space

Definition 4.4.1. ***Open neighbothood of a point***
An open neighborhood of a point a in a topological space $(X, \mathcal{T})$ is an open set containing a.

Definition 4.4.2. ***Convergence of a sequence***
A sequence (x_n) in a topological space $(X, \mathcal{T})$ is said to converge to a limit $a \in X$ if for every open neighborhood U of a, there exists a positive integer N such that for all positive integers $n \geq N$, $x_n \in U$.

Note that there can be multiple limits of a sequence in a topological space, and the limit need not be unique.

Example. $X = \mathbb{R}$, $\mathbb{T} = \{\emptyset, \mathbb{R}\} \cup \{(a, \infty) : a \in \mathbb{R}\}$ *is a topology on* $\mathbb{R}$.
$(a_n)_{n\in\mathbb{N}} = (\frac{1}{n})_{n\in\mathbb{N}}$ *converges to 0, but also to any* $a \in (0, \infty)$.

- *It converges to 0 - every open neighborhood of 0 looks like (a, ∞) for some $a < 0$ so $\frac{1}{n} \in (a, \infty)$.*
- *It converges to -1 - every open neighborhood of -1 looks like (a, ∞) for some $a < -1$ so $\frac{1}{n} \in (a, \infty)$.*
- *The same argument can be made for any $a \in (-\infty, 0]$.*

Definition 4.4.3. ***Hausdorff Space***
A topological space $(X, \mathcal{T})$ is a Hausdorff space if for every pair of distinct points $x, y \in X$, there exist open sets U and V such that $x \in U$, $y \in V$, and $U \cap V = \emptyset$.

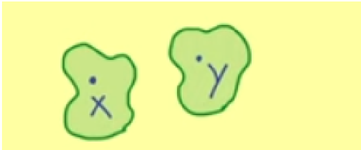


Figure 4.3: Hausdorff space

In a Hausdorff space, every pair of distinct points can be separated by disjoint open sets.

4.5 Quotient space

Definition 4.5.1. ***Projective space***
The projective space $\mathbf{P}^n(\mathbb{R})$ is the set of all lines passing through the origin in $\mathbb{R}^{n+1}$. It can be thought of as the space of all one-dimensional subspaces of $\mathbb{R}^{n+1}$.

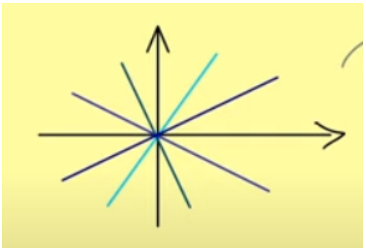


Figure 4.4: The projective space $\mathbf{P}^1$ - all the lines passing through the origin in $\mathbb{R}^2$

The directions define a set, and we want to create a topological space for this set.

Definition 4.5.2. ***Equivalence relation***
An equivalence relation on a set X is a relation $\sim$ that satisfies the following properties for all $x, y, z \in X$:

1. *Reflexivity: $x \sim x$*
2. *Symmetry: $x \sim y$ implies $y \sim x$*
3. *Transitivity: $x \sim y$ and $y \sim z$ implies $x \sim z$*

Definition 4.5.3. ***Equivalence class***
Let X be a set and $\sim$ be an equivalence relation on X.
The equivalence class of x with respect to $\sim$ is

$$[x]_\sim = \{y \in X : y \sim x\}$$

Definition 4.5.4. ***Quotient set***
Let X be a set and $\sim$ be an equivalence relation on X.
The quotient set $X/\sim$ is defined as

$$X/\sim \quad = \{[x]_\sim : x \in X\}$$

Definition 4.5.5. ***Quotient map (Canonical projection)***
Let X be a set and $\sim$ be an equivalence relation on X.
The canonical projection is a function:

$$q : X \to X/\sim \quad \text{such that} \quad q(x) = [x]_\sim$$

Definition 4.5.6. ***Quotient Topology***
Let $X, \mathcal{T}$ be a topological space, $\sim$ be an equivalence relation on X and $q : X \to X/\sim$ be the canonical projection.
The quotient topology $\hat{\mathcal{T}}$ on $X/\sim$ is defined as

$$\hat{\mathcal{T}} = \{U \subseteq X/\sim \quad : q^{-1}(U) \in \mathcal{T}\}$$

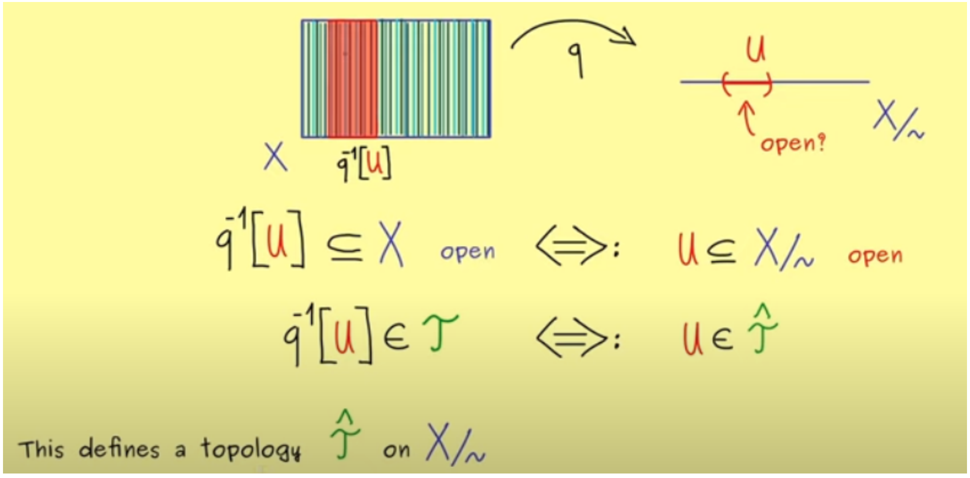


Figure 4.5: Quotient Topology

Example. *A classic example of a quotient space is the* ***mobius strip****, which is constructed by taking a rectangular strip of paper, giving one end a half-twist, and then gluing the two ends together. The mobius strip is a non-orientable surface, meaning that it does not have a consistent notion of "left" and "right" across the entire surface. The mobius strip is a quotient space of a square, where the opposite edges of the square are identified in a specific way. Formally,*

$$X = [0, 1] \times (-1, 1) \quad \text{and} \quad \sim \text{ is the equivalence relation that identifies the points}$$
$$(0, y) \sim (1, -y) \quad \text{for all} \quad y \in (-1, 1).$$

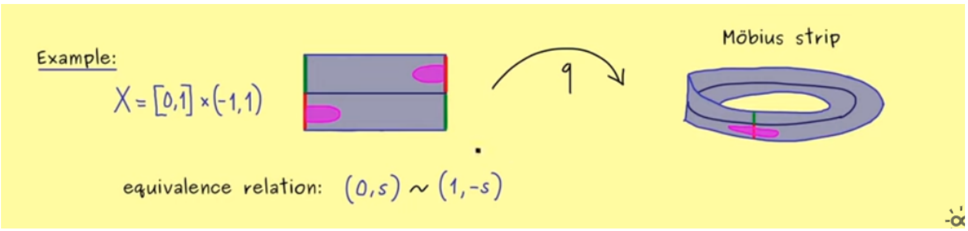


Figure 4.6: Mobius Strip

4.6 Projective Space

So far we have seen the transition from topological space to quotient space:

$$(X, \mathcal{T}) \to (X/\sim, \hat{\mathcal{T}})$$

Recall that the projective space $\mathbf{P}^n(\mathbb{R})$ is the set of all lines passing through the origin in $\mathbb{R}^{n+1}$.

Definition 4.6.1. ***Sphere***
The sphere S^n is the set of all points in $\mathbb{R}^{n+1}$ that are at a fixed distance of 1 from the origin.

$$S^n = \{x \in \mathbb{R}^{n+1} : \|x\|_2 = 1\}$$

Definition 4.6.2. ***The Projective Space as quotient space***
The projective space $\mathbf{P}^n(\mathbb{R})$ can be constructed as a quotient space of the sphere S^n by identifying antipodal points.

$$\mathbf{P}^n(\mathbb{R}) = S^n / \sim$$

where the equivalence relation $\sim$ is defined as

$$x \sim y \quad \textit{if} \quad x = y \quad \textit{or} \quad x = -y$$

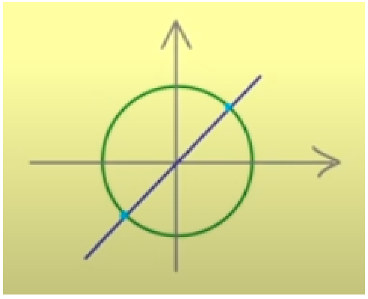


Figure 4.7: Projective Space as a quotient space

The sphere S^n is a hausdorff space, and the projective space $\mathbf{P}^n(\mathbb{R})$ is also a hausdorff space. But in general, the quotient space of a hausdorff space is not necessarily a hausdorff space. Take $[x]_\sim, [y]_\sim \in \mathbf{P}^n(\mathbb{R})$ such that $[x]_\sim \neq [y]_\sim \Rightarrow x \neq y$ and $x \neq -y$.

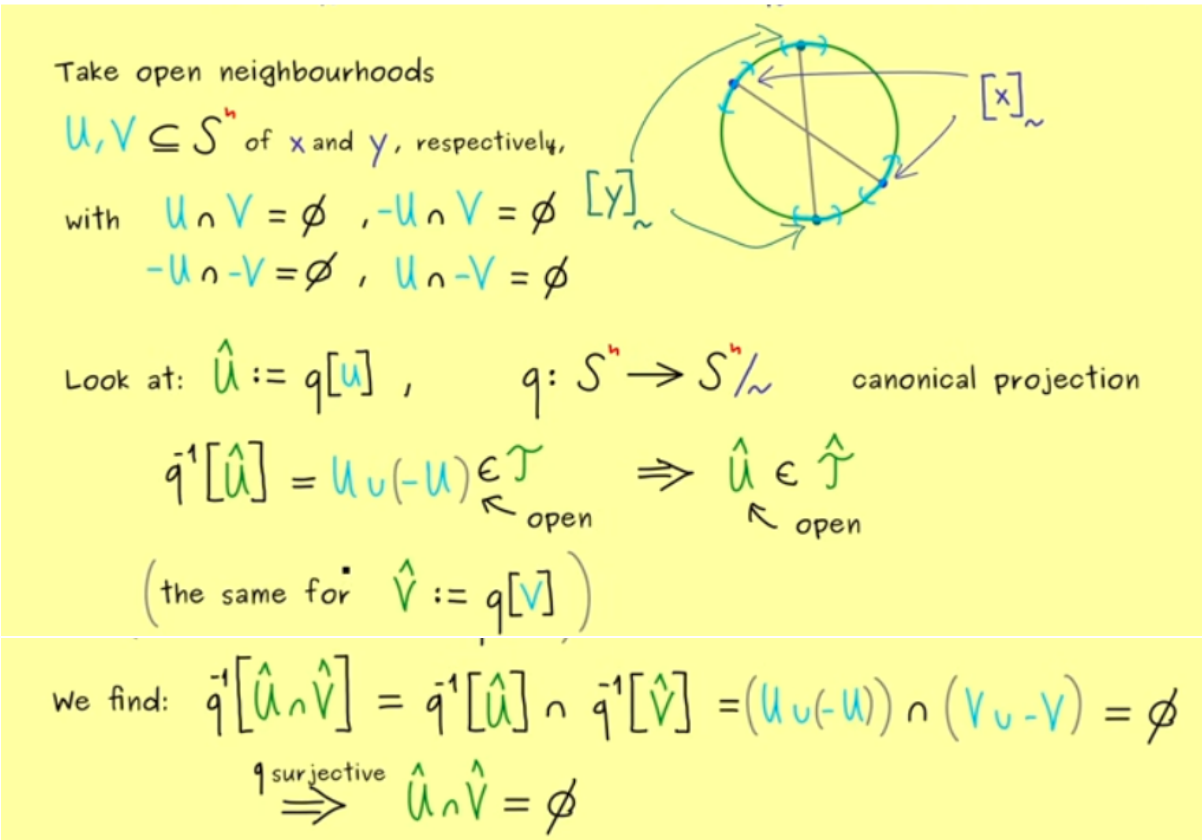


Figure 4.8: Projective Space as a hausdorff space

4.7 First Countable Space

Definition 4.7.1. ***First Countable Space***
A topological space $(X,\mathcal{T})$ is said to be first countable *if at every point x in X, there exists a countable basis for the neighborhoods of x. That is:*

$$\begin{array}{l} \forall x \in X, \quad \exists \{U_i\}_{i=1}^{\infty} \quad \textit{such that} \quad \forall i \in \mathbb{N} \quad x \in U_i \\ \textit{and} \quad \forall V \in \mathcal{T} \quad \textit{such that} \quad x \in V, \quad \exists i \quad \textit{such that} \quad U_i \subseteq V \end{array}$$

The countable set of open sets $\{U_i\}_{i=1}^{\infty}$ is called a ***local basis*** *at x.*

Examples. *of first countable spaces:*

- ***Metric Spaces****: Every metric space is first countable. The local basis at any point x in a metric space can be chosen as the collection of open balls centered at x with radii $1/n$, where n is a positive integer.*
- ***Subspaces of Metric Spaces****: Any subspace of a metric space inherits the first countability. If Y is a subspace of a metric space X, then the local bases in Y can be constructed using the intersections of open balls in X with Y.*
- ***Euclidean Spaces*** $\mathbb{R}^n$*: The Euclidean space with its standard topology, which is generated by the open balls, is first countable.*
- ***The Space of Rational Numbers*** $\mathbb{Q}$*: As a subspace of the real line $\mathbb{R}$, which is a metric space, the space of rational numbers $\mathbb{Q}$ is first countable.*

4.8 Second Countable Space

Definition 4.8.1. ***Basis***
A basis for a topological space $(X,\mathcal{T})$ is a collection $\mathcal{B} \subseteq \mathcal{T}$ of open sets such that every open set in $\mathcal{T}$ can be written as a union of sets in $\mathcal{B}$.

Examples. *of basis:*

- *$\mathcal{B} = \mathcal{T}$ is a basis for $\mathcal{T}$.*
- *If $\mathcal{T}$ is discrete, then $\mathcal{B} = \{\{x\} : x \in X\}$ is a basis for $\mathcal{T}$.*
- *Let $(X,\mathcal{T})$ be a topological space induced by a metric space (X,d).*
  Then $\mathcal{B} = \{B_\varepsilon(x) : x \in X, \varepsilon > 0\}$ is a basis for $\mathcal{T}$.
- *$\mathbb{R}^n$ with the standard topology (defined by the Euclidean metric) has a basis of open balls. Then $\mathcal{B} = \{B_\varepsilon(x) : x \in \mathbb{Q}^n, \varepsilon \in \mathbb{Q}, \varepsilon > 0\}$ is a basis for $\mathcal{T}$. Eventhought the space is uncountable, the basis has a countable number of elements.*

Definition 4.8.2. ***Second Countable Space***
A topological space $(X,\mathcal{T})$ is second countable if there exists a countable basis for $\mathcal{T}$.

Examples. *of second countable spaces:*

- *Any finite or countable discrete space.*
- *The real line $\mathbb{R}$ with the standard topology.*
- *The Euclidean space $\mathbb{R}^n$ with the standard topology.*

- *The space of continuous functions $C([0,1])$ with the topology of uniform convergence.*

Examples. *of spaces that are not second countable:*

- *Consider an uncountable set (like $\mathbb{R}$) with the discrete metric (the distance between each pair of distinct points is 1, and the distance between a point and itself is 0). In this metric space, every singleton set is open, therefore to have a base for the topology, we need to include every singleton, which is an uncountable number.*

Definition 4.8.3. ***The space of continuous functions with the topology of uniform convergence***
Let $C([0,1])$ be the space of continuous functions on the interval $[0,1]$.
The topology of uniform convergence on $C([0,1])$ is defined by the basis

$$\mathcal{B} = \{B_\varepsilon(f) : f \in C([0,1]), \varepsilon > 0\}$$

where $B_\varepsilon(f) = \{g \in C([0,1]) : \|f - g\|_\infty < \varepsilon\}$.

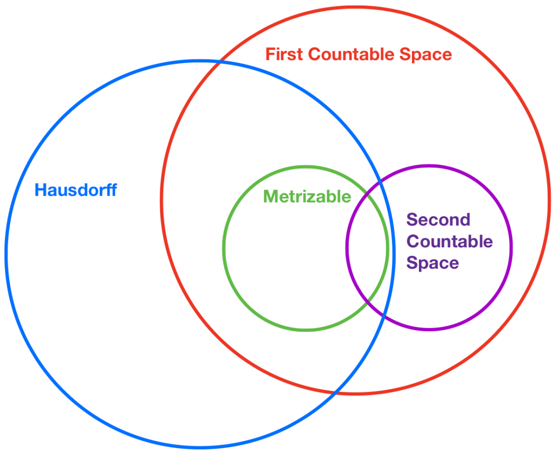


Figure 4.9: Venn diagram of topological spaces - Metrizable, Hausdorff, First Countable, Second Countable

4.9 Continuity

Recall the definition of continuity of a function in R^n:

Definition. ***Continuity of a function***
A function $f : \mathbb{R}^n \to \mathbb{R}$ is continuous at a point $x \in \mathbb{R}^n$ if for every $\varepsilon > 0$, there exists $\delta > 0$ such that for all $y \in \mathbb{R}^n$ with $\|x - y\|_2 < \delta$, we have $\|f(x) - f(y)\|_2 < \varepsilon$.

or equivalently, in term of sequences:

Definition. ***Continuity of a function (sequence definition)***
A function $f : \mathbb{R}^n \to \mathbb{R}$ is continuous at a point $x \in \mathbb{R}^n$ if for every sequence (x_k) in $\mathbb{R}^n$ such that $x_k \to x$, we have $f(x_k) \to f(x)$.

Now, we extend the definition of continuity to topological spaces.

Definition 4.9.1. ***Continuity of a function at a point (topological space)***
Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces.
A function $f : X \to Y$ is continuous at a point $x \in X$ if for every neighborhood V of $f(x)$ in Y, there exists a neighborhood U of x in X such that $f(U) \subseteq V$.

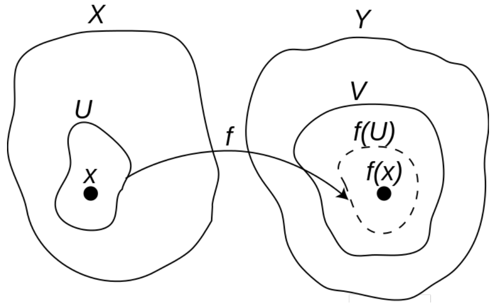


Figure 4.10: Continuity of a function at a point

Definition 4.9.2. ***Continuity of a function (topological space)***
Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces.
A function $f : X \to Y$ is continuous if for every open set $U \in \mathcal{T}_Y$, $f^{-1}(U) \in \mathcal{T}_X$ (the preimage of an open set is open).

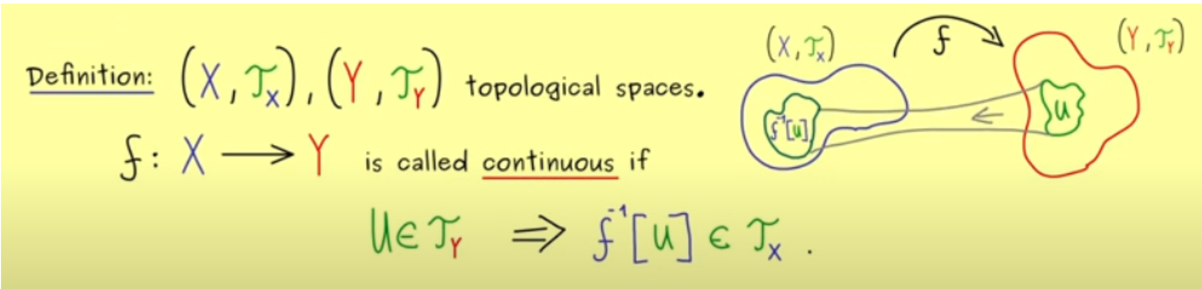


Figure 4.11: Continuity of a function

In term of sequences:

Definition 4.9.3. ***Sequentially continuous function (topological space)***
Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces.
A function $f : X \to Y$ is sequentially continuous if for every $x \in X$ and every sequence $(x_n)_{n\in\mathbb{N}} \subseteq X$ such that $x_n \to x$, $(f(x_n))_{n\in\mathbb{N}} \subseteq Y$ convergent with $(f(x_n))_{n\in\mathbb{N}} \to f(x)$.

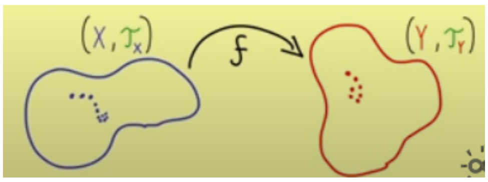


Figure 4.12: Sequentially continuous function

Theorem 4.9.1. ***Continuity vs Sequential Continuity***
Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be topological spaces and $f : X \to Y$ be a function.

$$f \text{ is continuous} \Rightarrow f \text{ is sequentially continuous}$$
$$f \text{ is continuous} \not\Leftarrow f \text{ is sequentially continuous}$$

Theorem 4.9.2. ***Continuity vs Sequential Continuity in First Countable Spaces***
Let $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ be First Countable spaces and $f : X \to Y$ be a function. Then

$$f \text{ is continuous} \Leftrightarrow f \text{ is sequentially continuous}$$

Examples. *of continuous maps*

- *The indiscrete topology on X and Y - every function is continuous.*
  e.g., $\forall f : X \to Y$, $\mathcal{T}_X = \{\emptyset, X\}$ and $\mathcal{T}_Y = \{\emptyset, Y\}$, then f is continuous since the preimage of any open set in Y is either $\emptyset$ or X.

- *The discrete topology on X and Y - every function is continuous.*
  e.g., $\forall f : X \to Y$, $\mathcal{T}_X = P(X)$ and $\mathcal{T}_Y = P(Y)$, then f is continuous since the preimage of any open set in Y must be a subset of X.

- *The quotient map $q : X \to X/\sim$ is continuous.*
  Let $(X, \mathcal{T})$ be a topological space and $\sim$ be an equivalence relation on X.
  Then the quotient map $q : X \to X/\sim$ is continuous.

  Proof. Let $U \in \mathcal{T}_{X/\sim}$ be an open set in $X/\sim$.
  Then $q^{-1}(U) \in \mathcal{T}_X$ is an open set in X by definition of the quotient topology. The same argument holds for the other direction. □

Definition 4.9.4. ***Homeomorphism*** *(NOT homomorphism)*
A function $f : X \to Y$ between topological spaces $(X, \mathcal{T}_X)$ and $(Y, \mathcal{T}_Y)$ is a homeomorphism if

1. *f is bijective (both injective and surjective - one-to-one and onto).*
2. *$f : X \to Y$ is continuous.*
3. *$f^{-1} : Y \to X$ is continuous.*

Not all continuous functions are homeomorphisms. For example, a quotient map is continuous but not necessarily a homeomorphism, if more than one point is mapped to the same equivalence class.

4.10 Compactness

We know that any closed and bounded subset of $\mathbb{R}$ is compact, for example, the closed interval $[a, b] \in \mathbb{R}$ is compact.

Definition 4.10.1. ***Compact set***
Let $(X, \mathcal{T})$ be a topological space, and $A \subseteq X$.
A is called compact if for every open cover $\{U_i\}_{i \in I}$ of A, there exists a finite subcover of A.

$$A \subseteq \bigcup_{i \in I} U_i \quad \Rightarrow \quad A \subseteq \bigcup_{i \in I_0} U_i \quad \textit{for some finite} \quad I_0 \subseteq I$$

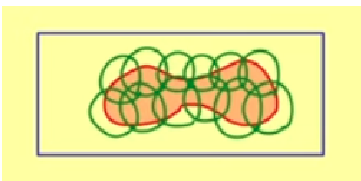


Figure 4.13: Compact set - every open cover has a finite subcover

We are already familiar with the Heine-Borel theorem for compact sets in $\mathbb{R}^n$ with the standard topology.

Theorem 4.10.1. ***Heine-Borel Theorem***
A subset $A \subseteq \mathbb{R}^n$ is compact $\Leftrightarrow$ A is closed and bounded.

We don't have this theorem for general topological spaces, simply because the term "bounded" is not defined in a general topological spaces. If we had a metric we could define boundedness in terms of the metric, but even then the Heine-Borel theorem would not hold in general.

Theorem 4.10.2. ***Compactness in hausdorff spaces***
Let $(X, \mathcal{T})$ be a hausdorff space and $A \subseteq X$.
If A is compact, then A is closed.

Proof. Let A be a compact set in a hausdorff space X.
We want to show that A is closed.
Let $b \in X \setminus A$. For every $a \in A$, since X is hausdorff, there exist open sets U_a and V_a such that $a \in U_a$, $b \in V_a$, and $U_a \cap V_a = \emptyset$.
Then $\{U_a\}_{a \in A}$ is an open cover of A.
Since A is compact, there exists a finite subcover $\{U_{a_1}, U_{a_2}, \ldots, U_{a_n}\}$ of A.
Let $V = \bigcap_{i=1}^{n} V_{a_i}$.
Then V is an open set containing b and $V \cap A = \emptyset$, which implies that b is an interior point of $X \setminus A$. Therefore, A is closed. $\square$

4.11 Locally Euclidean Spaces

Definition 4.11.1. ***n-dimensional (topological) Manifold***
A topological space $(M, \mathcal{T})$ is called a manifold of dimension n if it satisfies the following conditions:

1. *$(M, \mathcal{T})$ is hausdorff space.*
2. *$(M, \mathcal{T})$ is second countable.*
3. *$(M, \mathcal{T})$ is locally euclidean of dimension n.*

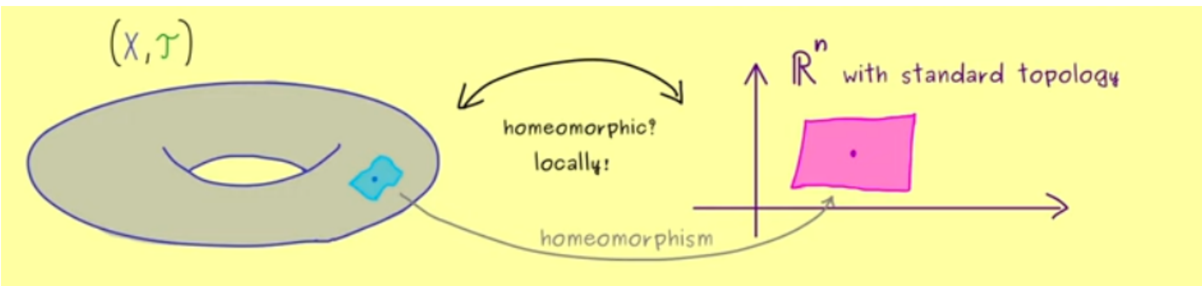


Figure 4.14: Locally Euclidean Topological Space

Definition 4.11.2. ***Locally Euclidean Space***
A topological space $(M, \mathcal{T})$ is locally euclidean of dimension n if
$\forall x \in M$, there exists an open neighborhood $U \in \mathcal{T}$ and a homeomorphism $h : U \to V$ where V is an open subset of $\mathbb{R}^n$.
Such an homeomorphism $h : U \to V$ is called a ***chart*** *of $(M, \mathcal{T})$ at x.*

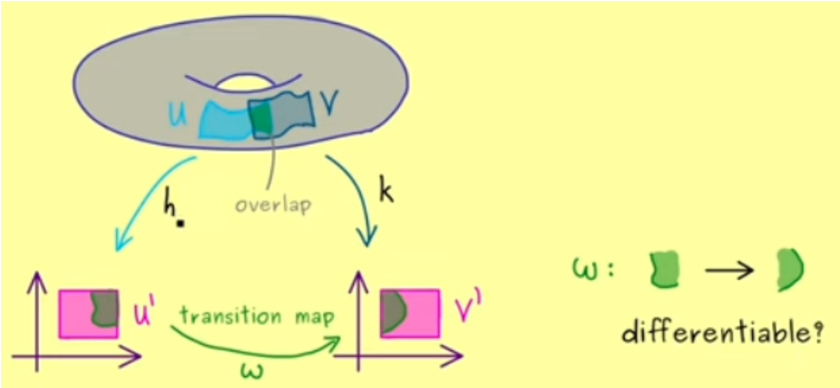


Figure 4.15: Charts and Atlases of a Manifold

4.12 Examples of Manifolds

Definition 4.12.1. ***Atlas***
An atlas of a manifold M is a collection of charts that cover M.
Formally, an atlas $\mathcal{A}$ of a manifold M is a collection of charts $\{(U_i, h_i)\}_{i\in I}$ such that

$$M = \bigcup_{i\in I} U_i$$

Examples. *of Manifolds*

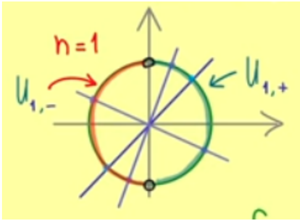
- *$M, \mathcal{T}$ a discrete topological space with countably many points.*
  Every point in M is a manifold of dimension 0.
- *$M \subseteq \mathbb{R}^n$ with the standard topology, $(M, \mathcal{T})$ is a manifold of dimension n.*
  h is the identity map and M is locally euclidean of dimension n (in fact, it is globally euclidean).
- *$S^2 = \{x \in \mathbb{R}^3 : \|x\|_2 = 1\}$ is a manifold of dimension 2.*
  We have seen that S^2 is a hausdorff space.
  The atlas of S^2 consists of two charts: the upper hemisphere and the lower hemisphere.
  The upper hemisphere is the set $U_{3,-} = \{x \in S^2 : x_3 > 0\}$ and the lower hemisphere is the set $U'_{3,-} = \{x \in S^2 : x_3 < 0\}$.
  The homeomorphism $h_{3,-} : U_{3,-} \to \mathbb{R}^2$ is defined as $h_{3,-}(x_1, x_2, x_3) = (x_1, x_2)$.
  The inverse homeomorphism $h_{3,-}^{-1} : \mathbb{R}^2 \to U_1$ is defined as $h_{3,-}^{-1}(x_1, x_2) = (x_1, x_2, \sqrt{1 - x_1^2 - x_2^2})$.
  The same argument holds for the lower hemisphere.
  So $(U_{i,\pm}, h_{i,\pm})_{i\in\{1,2,3\}}$ is an atlas of S^2.

4.13 Projective Space is a Manifold

As we have seen that the sphere $S^n = \{x \in \mathbb{R}^{n+1} : \|x\|_2 = 1\}$, is an n-dimensional manifold with atlas $(U_{i,\pm}, h_{i,\pm})_{i\in\{1,\dots,n+1\}}$ where $U_{i,\pm} = \{x \in \mathbb{R}^{n+1} : \pm x_i > 0\}$ and $h_{i,\pm}(x_1, \dots, x_{n+1}) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{n+1})$. We have also seen that the projective space $\mathbf{P}^n(\mathbb{R}) = S^n/\sim$ with quotient topology and equivalence relation $x \sim y \Leftrightarrow x = y$ or $x = -y$.
We will show that the projective space $\mathbf{P}^n(\mathbb{R})$ is a manifold of dimension n.

We will define $V_i = \{[x]_\sim \in \mathbf{P}^n : x_i \neq 0\}$. Note that every equivilance class $[x]_\sim$ contains exactly 2 points from the sphere S^n. $q^{-1}[V_i] = U_{i,+} \cup U_{i,-}$ is an open set in the sphere S^n and $q : S^n \to \mathbf{P}^n$ is a quotient map, so V_i is an open set in the projective space $\mathbf{P}^n$.

In order to show that $\mathbf{P}^n(\mathbb{R})$ is a manifold of dimension n, we will present homeomorphisms $h_i : V_i \to \mathbb{R}^n$ for each $i \in \{1, \ldots, n+1\}$ such that (V_i, h_i) is a chart of $\mathbf{P}^n(\mathbb{R})$.



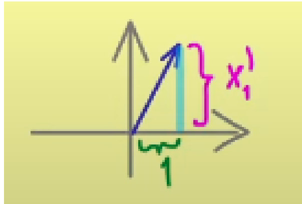


Figure 4.16: Projective Space of dimension 1 as a quotient space of the circle

For n = 1:
$V_1 = \{[x]_\sim \in \mathbf{P}^1 : x_1 \neq 0\}$ is an open set in $\mathbf{P}^1$.
$q^{-1}[V_1] = U_{1,+} \cup U_{1,-}$
We need homeomorphisms $h_1 : V_1 \to V_1'$ where V_1' is an open subset of $\mathbb{R}$.
We can define $h_1 : V_1 \to V_1'$ as $h_1([x]_\sim) = \frac{x_2}{x_1}$ (the slope) for $x_1 \neq 0$.
The invese homeomorphism $h_1^{-1} : V_1' \to V_1$ is defined as $h_1^{-1}(x1) = [\binom{1}{x1'} \cdot \frac{1}{\sqrt{1^2+(x1')^2}}]_\sim$.
$V_2 = \{[x]_\sim \in \mathbf{P}^1 : x_2 \neq 0\}$ and we will do the same with the change of coordinates.

For $n \in \mathbb{N}$:
$V_i = \{[x]_\sim \in \mathbf{P}^n : x_i \neq 0\}$ is an open set in $\mathbf{P}^n$.
$q^{-1}[V_i] = U_{i,+} \cup U_{i,-}$
We need homeomorphisms $h_i : V_i \to V_i'$ where V_i' is an open subset of $\mathbb{R}^n$.
We can define $h_i : V_i \to V_i'$ as $h_i([x]_\sim) = \frac{1}{x_i}(x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_{n+1})$ for $x_i \neq 0$.
In the inverse we apply 1 to the i-th coordinate and normalize the vector.

4.14 Smooth structures (differential structures)

Definition 4.14.1. ***Transition map***
Let $(M, \mathcal{T})$ be an n-manifold and (U, h) and (V, k) be two charts of M.
The transition map $\omega : h(U \cap V) \to k(U \cap V)$ is called the transition map between the charts (U, h) and (V, k).

Note that the transition map is a transoformation of subset of the R^n space, it does not have to be a linear transformation.
We have seen in definition 1.0.14 that a function $f : \mathbb{R}^n \to \mathbb{R}$ is of class C^2 if all partial derivatives of f up to order 2 exist and are continuous. We can extend this definition to manifolds.

Definition 4.14.2. ***C^k-Diffeomorphism (Vector Space)***
A function $\omega : \mathbb{R}^n \to \mathbb{R}^n$ C^k-diffeomorphism ($k \in \{0, 1, \cdots\} \cup \{\infty\}$) if the following conditions hold:

1. *ω is k times continuously differentiable (partial derivatives up to order k exist and are continuous).*
2. *ω is bijective (both injective and surjective - one-to-one and onto).*
3. *The inverse function $\omega^{-1} \in C^k$.*

If $k = \infty$, then ω is called a smooth diffeomorphism (or simply a diffeomorphism) - all partial derivatives of ω exist and are continuous.
If $k = 0$, then ω is called a homeomorphism - ω is continuous, bijactive, and the inverse function ω^{-1} is continuous (see definition 4.9.4). By the composition of charts, the transition map is a always a homeomorphism.

Definition 4.14.3. ***C^k-smoothly compatible charts***
Two charts (U,h) and (V,k) of a manifold M are called C^k-smoothly compatible if the transition map $\omega : h(U \cap V) \to k(U \cap V)$ is a C^k-diffeomorphism.

(If $k = \infty$, then the charts are called smoothly compatible.)

What happens if any two charts of a manifold are C^k-smoothly compatible?

Definition 4.14.4. ***C^k-atlas***
An atlas $\mathcal{A} = \{(U_i, h_i)\}_{i \in I}$ of a manifold M is called a C^k-atlas if every pair of charts in $\mathcal{A}$ are C^k-smoothly compatible.

Definition 4.14.5. ***Maximal C^k-atlas (C^k-smooth structure)***
A maximal C^k-atlas of a manifold M is $\mathcal{A}$ is:

1. *$\mathcal{A}$ is a C^k-atlas of M.*

2. *For any other C^k-atlas $\mathcal{B}$ of M, $\mathcal{A} \nsubseteq \mathcal{B}$.*

Definition 4.14.6. ***n-dimensional C^k-smooth manifold***
A manifold M is called a C^k-smooth manifold if:

1. *M is n-dimensional topological manifold.*

2. *There exists a maximal C^k-atlas of M.*

4.15 Examples of smooth manifolds

Example. ***The n-dimensional Euclidean sphere $S^n \subseteq \mathbb{R}^{n+1}$***
We have seen that the sphere $S^n = \{x \in \mathbb{R}^{n+1} : \|x\|_2 = 1\}$, is an n-dimensional manifold with atlas $(U_{i,\pm}, h_{i,\pm})_{i \in \{1,\dots,n+1\}}$ where $U_{i,\pm} = \{x \in \mathbb{R}^{n+1} : \pm x_i > 0\}$ and $h_{i,\pm}(x_1, \dots, x_{n+1}) = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_{n+1})$.

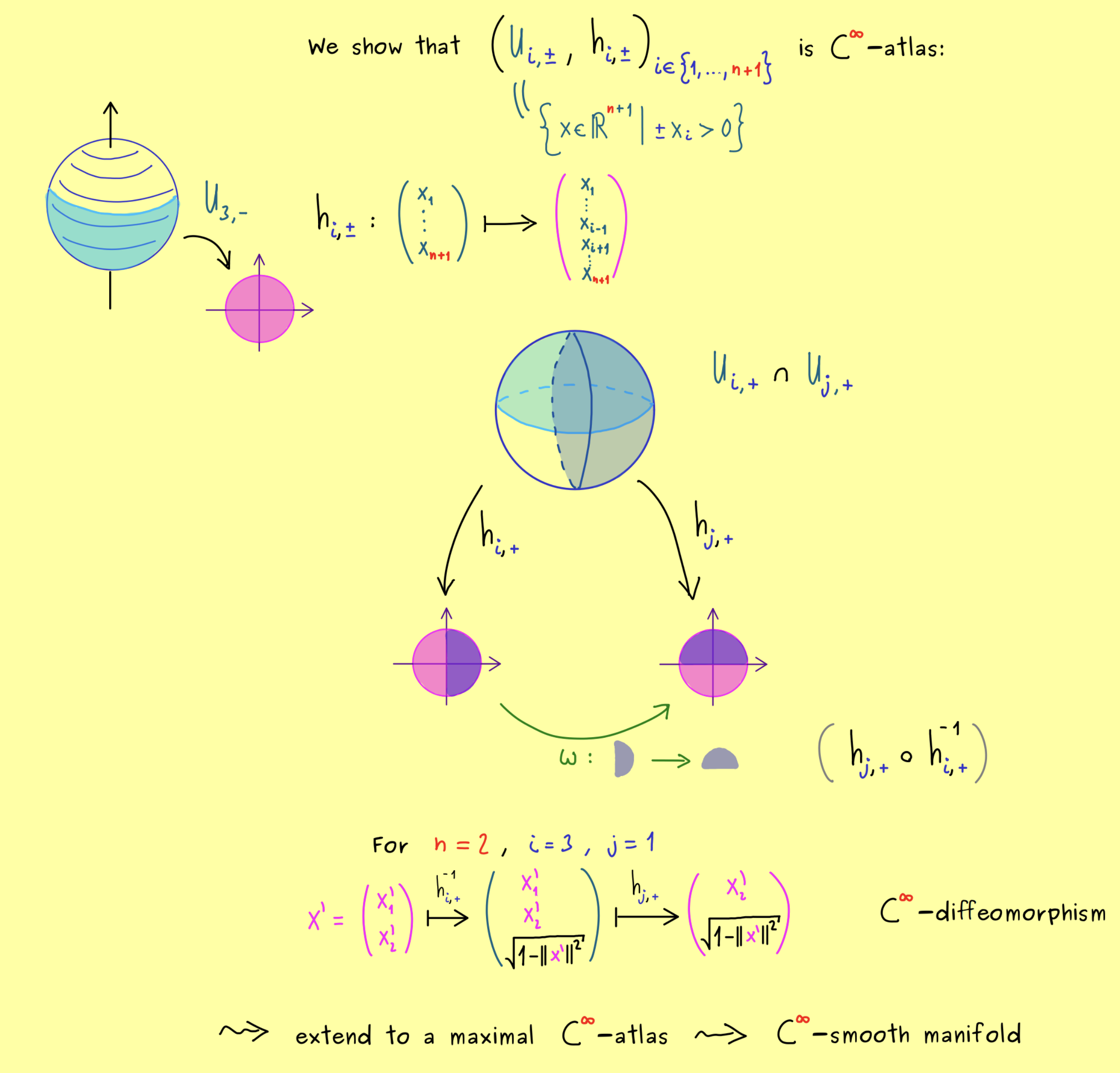


Figure 4.17: Sphere S^n is an n-dimensional C^∞ smooth manifold

Example. $\mathbb{R}^n$ ***is a*** C^∞ ***smooth manifold***
The Euclidean space $\mathbb{R}^n$ *is a* C^∞ *smooth manifold with the atlas* $\{(\mathbb{R}^n, id)\}$*. The is the standard smooth structure of* $\mathbb{R}^n$*.*

Example. ***Consider*** $f \in C^1$
Let $f : \mathbb{R} \to \mathbb{R}$ *be a function of class* C^1*.*
Let $G_f = \{(x, f(x)) \in \mathbb{R}^2 : x \in \mathbb{R}\}$ *be the graph of* f*.*
Then G_f *is a* C^1 *smooth manifold with one chart:* $h : G_f \to \mathbb{R}$ *defined as* $h(x, f(x)) = x$*.*

In fact any graph of a function of class C^k is a C^k smooth manifold. We can also observe that G_f is a 1-dimensional C^1-smooth manifold that is embedded in another smooth manifold, $\mathbb{R}^2$.

4.16 Submanifolds

Definition 4.16.1. ***Submanifold***
Let M *be a an n-dimensional smooth manifold.* $M_0 \subseteq M$ *is called a k-dimensional submanifold of* M *if for every* $p \in M_0$*, there exists a chart* (U, h) *of* M *such that* $h(U \cap M_0) = h(U) \cap (\mathbb{R}^k \times \{0\}^{n-k})$*.* (U, h) *is called a* ***submanifold chart*** *for* M_0*.*

Note: M_0 is also a manifold: $(U, h) submanifoldchart \to (\hat{U}, \hat{h})$ chart, $\hat{U} = U \cap M_0$, $\hat{h} = h|_{\hat{U}}$.

4.17 Regular value theorem in $\mathbb{R}^n$

Definition 4.17.1. ***Critical point***
Let $f : U \to \mathbb{R}^m$, $U \subseteq \mathbb{R}^n$ open, C^1-function.
A point $x \in U$ is called a critical point of f if the Jacobian matrix of f at x has rank less than m.

Definition 4.17.2. ***Regular value***
Let $f : U \to \mathbb{R}^m$, $U \subseteq \mathbb{R}^n$ open, C^1-function.
A point $y \in f[U]$ is called a regular value of f if for every $x \in f^{-1}(y)$, x is not a critical point of f.

Theorem 4.17.1. ***Regular value theorem in*** $\mathbb{R}^n$
Let $f : U \to \mathbb{R}^m$, $U \subseteq \mathbb{R}^n$ open, C^∞-function.
If c is a regular value of f, then $f^{-1}[\{c\}]$ is a submanifold of $\mathbb{R}^n$ of dimension $n - m$.

For example if $f : \mathbb{R}^n \to \mathbb{R}$, $f(x_1, \ldots, x_n) = x_1^2 + \cdots + x_n^2$, then 0 is a regular value of f.
let $0 \neq r \in \mathbb{R}$, then r is a regular value of f, for example $r = 1$. Then the sphere S^{n-1} is a submanifold of $\mathbb{R}^n$ of dimension $n - 1$.

4.18 Smooth maps

Until now we have defined smooth manifolds (and C^k-smooth manifolds).
Recall that

- an n-dimensional manifold is a topological space that is locally euclidean of dimension n and hausdorff and second countable.
- a smooth manifold is a topological manifold with a maximal C^∞-atlas.
- a C^∞-atlas is a collection of charts that cover the manifold and the transition maps between any two charts are C^∞-diffeomorphisms.
- C^k-diffeomorphism is a function that is k times continuously differentiable, bijective, and the inverse function is also C^k.

Definition 4.18.1. ***k-times differentiable map at a point***
Let $f : M \to N$ be a map between two smooth manifolds.
f is called k-times differentiable at a point $p \in M$ if for every chart (U, h) of M and every chart (W, k) of N, with $p \in U$ and $f(p) \in V$, the map $k \circ f \circ h^{-1}$ is k-times differentiable at $h(p)$.

Definition 4.18.2. *(C^∞)* ***Smooth map***
Let $f : M \to N$ be a map between two smooth manifolds.
f is called smooth if f is k-times differentiable at every point $p \in M$ for every $k \in \mathbb{N}$.

We often denote the set of smooth maps between two smooth manifolds M and N as $C^\infty(M, N)$.

4.19 Examples of smooth maps

- $S^2 \to \mathbb{R}^3$ where i is the identity matrix.
- the quotient map $q : S^2 \to P^2(\mathbb{R}) = S^2/\sim$

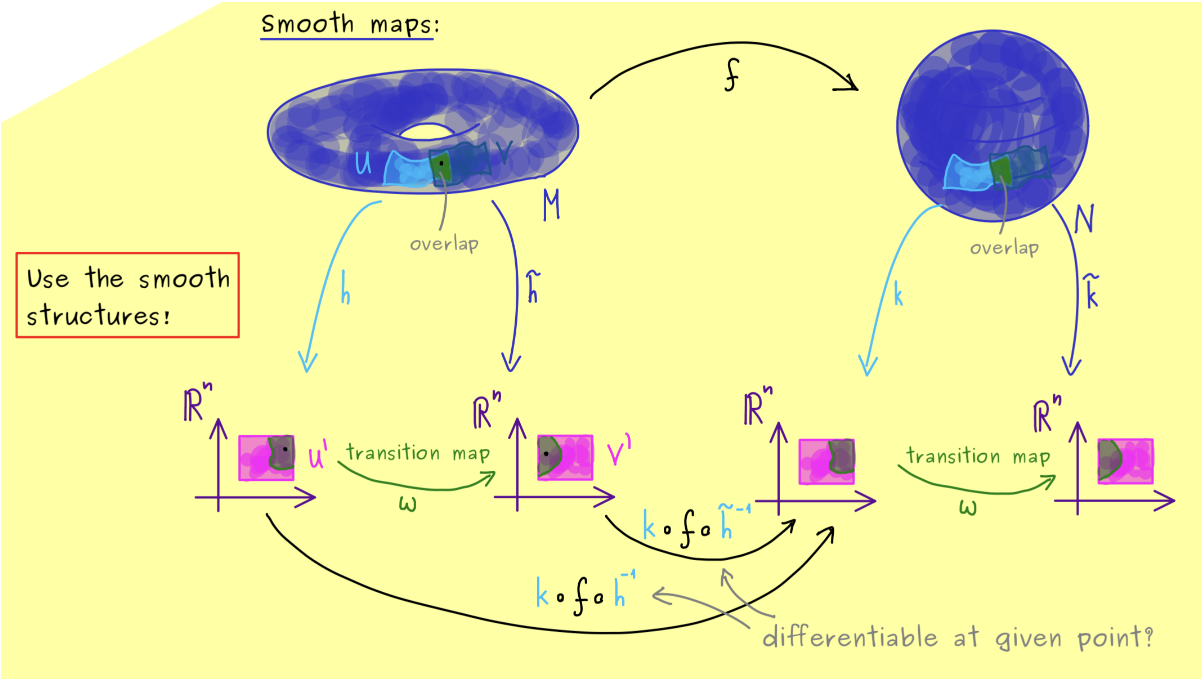


Figure 4.18: Smooth map between two manifolds

4.20 Regular value theorem (abstract version)

Definition 4.20.1. ***Critical point***
Let $f : M \to N$ be a smooth map between two smooth manifolds of dimensions m and n $(m \geq n)$.
Let $p \in M$, (U, h) a chart of M at p and (V, k) be a chart of N at $f(p)$.
A point $p \in M$ is called a critical point of f if

$$rank(f_p) := rank(J_{k \circ f \circ h^{-1}}(h(p))) < n$$

Theorem 4.20.1. ***Regular value theorem (abstract version)***
Let $f : M \to N$ be a smooth map between two smooth manifolds of dimensions n and m $(m \geq n)$.
Let $q \in N$ be a regular value of f.
Then $f^{-1}[\{q\}]$ is a submanifold of M of dimension $m - n$.

4.21 Tangent space for submanifolds

Definition 4.21.1. ***Local parametrization of a submanifold***
Let M be an n-dimensional smooth manifold and $M_0 \subseteq M$ be a k-dimensional submanifold of M.
Let $p \in M_0$ and (U, h) be a submanifold chart for M_0 at p, where $h : U \to U'$, $\quad U' \subseteq \mathbb{R}^k$ is a homeomorphism.
The function $\varphi : \mathbb{R}^k \cap U' \to M \cap U$ is called a local parametrization of M_0 at p.

Example. *The circle Let $S^1 = \{x \in \mathbb{R}^2 : \|x\|_2 = 1\}$ be the unit circle in $\mathbb{R}^2$.*
Let $M_0 = \{x \in S^1 : x_1 > 0\}$ be the right half of the circle.
Let $p = (1, 0) \in M_0$ and (U, h) be a submanifold chart for M_0 at p, where $h : U \to U'$ is a homeomorphism.
The function $\varphi : (0, 2\pi) \to M \cap U$ defined as $\varphi(t) = (\cos(t), \sin(t))$ is a local parametrization of M_0 at p.

Definition 4.21.2. ***The Tangent Space of a submanifold***
Let $(M \subseteq \mathbb{R}^n)$ be a k-dimensional submanifold of $\mathbb{R}^n$.
Let $p \in M$ and $\varphi : U' \to U$ be a local parametrization of M at p.
Let $\tilde{p} = \varphi^{-1}(p)$.
The tangent space of M at p is defined as

$$T_p^{sub} M = d\varphi_{\tilde{p}}[\mathbb{R}^k] = \{J_\varphi(\varphi^{-1}(p)) \cdot v : v \in \mathbb{R}^k\}$$

Note thet $T_p^{sub} M \subseteq \mathbb{R}^n$ is a k-dimensional subspace. If $M = \mathbb{R}^n$, then $\phi = \phi^{-1} = id$, $J_\phi(\phi^{-1}(p)) = I$ and $T_p^{sub}\mathbb{R}^n = \mathbb{R}^n$.

4.22 Tangent curves

Definition 4.22.1. ***Smooth curve***
Let M be an n-dimensional smooth manifold.
A smooth curve in M is a smooth map $\gamma : I \to M$ where I is an open interval in $\mathbb{R}$.

create a secti
curves in R^n
calculus chap

Definition 4.22.2. ***Differentiable curve in*** $\mathbb{R}^n$

A differentiable curve in $\mathbb{R}^n$ is a function $\gamma : I \to \mathbb{R}^n$, where I is an interval in $\mathbb{R}$, such that each of the component functions of γ is differentiable at every point in I.
In other words, a curve $\gamma(t) = (\gamma_1(t), \gamma_2(t), \ldots, \gamma_n(t))$ is differentiable if each $\gamma_i(t)$ has a derivative $\gamma_i'(t)$ for all $t \in I$.
Formally, a curve $\gamma : (a, b) \to \mathbb{R}^n$ is differentiable at a point $t_0 \in (a, b)$ if the limit

$$\lim_{t \to t_0} \frac{\gamma(t) - \gamma(t_0)}{t - t_0}$$

exists, and is a vector in $\mathbb{R}^n$. This vector is the tangent vector to the curve at the point $\gamma(t_0)$, and the collection of all such tangent vectors as t varies in (a, b) forms a vector field along the curve, provided the curve is differentiable on the interval.

Definition 4.22.3. ***Differentiable curve on a manifold*** M
A differentiable curve on a manifold M is a function $\gamma : I \to M$, where I is an interval in $\mathbb{R}$. The curve γ is differentiable at a point $t_0 \in I$ if for every chart (U, ϕ) of M with $\gamma(t_0) \in U$, the composition $\phi \circ \gamma : I \to \mathbb{R}^n$ is differentiable at t_0. A curve is said to be differentiable on I if it is differentiable at every point in I.

Definition 4.22.4. ***Tangent curve***
Let M be an n-dimensional smooth manifold and $\gamma : I \to M$ be a smooth curve in M.
Let $p = \gamma(t_0)$ for some $t_0 \in I$.
The tangent curve of γ at t_0 is defined as

$$\gamma'(t_0) = d\gamma_{t_0}[\mathbb{R}]$$

The derivative $\gamma'(t_0)$ at t_0 is then defined to be a vector in the tangent space $T_{\gamma(t_0)}M$.

Definition 4.22.5. ***The Tangent Space of a submanifold***
Let $M \subseteq \mathbb{R}^n$ be a k dimensional submanifold and $p \in M$. The tangent space of M at p is:

$$T_p^{sub} M = \{\gamma'(0) \quad | \quad \gamma : (-\epsilon, \epsilon) \to M \quad \textit{differentiable and} \quad \gamma(0) = p\}$$

Theorem 4.22.1. ***The 2 definitions for the tangent space of submanifold are equivalent***
Let $M \subseteq \mathbb{R}^n$ be a k-dimensional submanifold and $p \in M$.
Let $\varphi : U' \to U$ be a local parametrization of M at p.
Let $\tilde{p} = \varphi^{-1}(p)$. Then

$$T_p^{sub}M = \{J_\varphi(\varphi^{-1}(p)) \cdot x : x \in \mathbb{R}^k\} = \{\gamma'(0) \quad | \quad \gamma : (-\epsilon, \epsilon) \to M \quad \textit{differentiable and} \quad \gamma(0) = p\}$$

Proof. We will show double inclusion between the groups:

Let $\gamma : (-\epsilon, \epsilon) \to M$ be a smooth curve in M with $\tilde{\gamma}(0) = p$.
Let $\tilde{\gamma} : (-\epsilon, \epsilon) \to U' \subseteq \mathbb{R}^k$ be the curve defined as $\tilde{\gamma}(t) = \varphi^{-1}(\gamma(t))$.

$\subseteq$: $v \in \{J_\varphi(\varphi^{-1}(p)) \cdot x : x \in \mathbb{R}^k\}$
$v = J_\varphi(\tilde{\gamma}(0)) \cdot \tilde{\gamma}'(0)$ with $\tilde{\gamma}(t) = \tilde{p} + t(\tilde{\gamma}'(0))$
Then $v = \frac{d}{dt}(\varphi \circ \tilde{\gamma}) \mid_{t=0} = \frac{d}{dt}(\gamma) \mid_{t=0} = \gamma'(0)$

$\supseteq$: $v \in \{\gamma'(0) \quad | \quad \gamma : (-\epsilon, \epsilon) \to M \quad \text{differentiable and} \quad \gamma(0) = p\}$
There exists an open set U such that $\forall y \in (-\epsilon, \epsilon) \quad \gamma(y) \in U$.
Let $\tilde{\gamma} : (-\epsilon, \epsilon) \to U'$ be defined as $\tilde{\gamma}(t) = \varphi^{-1}(\gamma(t))$.
Then $\tilde{\gamma}'(0) = \frac{d}{dt}(\varphi^{-1} \circ \gamma) \mid_{t=0} = J_\varphi^{-1}(\gamma(0)) \cdot \gamma'(0) \in T_p^{sub}M$ □

4.23 Tangent space (definition via tangent curves)

Now we will define the tangent space of an abstract smooth manifold (that is not embedded in $\mathbb{R}^n$).

Definition 4.23.1. ***The set of smooth curves that pass throught p*** ($C_p(M)$)
Let M be an k-dimensional smooth manifold and $p \in M$.

$$C_p(M) = \{\gamma : (-\epsilon, \epsilon) \to M \quad | \quad \gamma \quad \textit{differentiable and} \quad \gamma(0) = p\}$$

Let γ, α be two smooth curves in $C_p(M)$ and (U, h) be a chart of M at p.
We define the equivalence relation $\gamma \sim \alpha \iff (h \circ \gamma)'(0) = (h \circ \alpha)'(0)$.
The equivalence classes of $C_p(M)$ are denoted as $[\gamma]_\sim$ and each equivalence class repsents a tangent vector at p.

Definition 4.23.2. ***Tangent space of a smooth manifold***
Let M be an k-dimensional smooth manifold and $p \in M$.
The tangent space of M at p is defined as

$$T_pM = C_p(M)/\sim \quad = \{[\gamma]_\sim \quad | \quad \gamma \in C_p(M)\}$$

Result:

- For a submanifold $M \subseteq \mathbb{R}^n$, $T_p^{sub}M = T_pM \quad (\gamma'(0) \iff [\gamma]_\sim)$.
- Let $v, w \in T_pM$ and $\lambda \in \mathbb{R}$. Then T_pM is a vector space with the operations:
  - $v + w := h_*(v) + h_*(w)$ with $h_* : [\gamma]_\sim \to (h \circ \gamma)'(0)$
    (h_* sends an equivalence class of curves to their common tangent vector in $\mathbb{R}^k$).
  - $\lambda \cdot v := h_*^{-1}(\lambda \cdot h_*(v))$

4.24 Coordinate bases

Let M be an n-dimensional smooth manifold and (U, h) be a chart of M at p.
Then h is a homeomorphism between U and an open set $U' \subseteq \mathbb{R}^n$.
Let $\varphi = h^{-1} : U' \to U$ be the inverse of h (it exists and smooth (homeomorphism) by definition).
Let $\tilde{p} = h(p)$. We have seen that the tangent space of M at p is $T_pM = C_p(M)/\sim$.
The tangent space T_pM can also be mapped to $\mathbb{R}^n$ by the map $h_* : T_pM \to \mathbb{R}^n$ defined as

$$h_*([\gamma]_\sim) = (h \circ \gamma)'(0)$$

In fact, as n is the dimension of M, n is also the dimension of T_pM that means h_* is linear and bijective $\Rightarrow h_*$ is an isomorphism.

Definition 4.24.1. ***Coordinate basis (standard basis with respect to** (U, h)**)***
For (U, h) and $p \in U$ we define the tangent vector $\partial_j := \varphi_*(e_j)$ where $(e_1, \dots, e_n)$ is the standard basis of $\mathbb{R}^n$. The set $\{\partial_1, \dots, \partial_n\}$ is called the coordinate basis of T_pM with respect to (U, h).

Remark: For n-dimensional submanifolds $M \subseteq \mathbb{R}^N$, $T_p^{sub}M = T_pM \cong \mathbb{R}^n$ and $\varphi_*(e_j) = J_\varphi(\tilde{p}) \cdot e_j$, and $(\partial_1, \dots, \partial_n)$ is essentially $(\frac{\partial \varphi}{\partial x_1}(\tilde{p}), \dots, \frac{\partial \varphi}{\partial x_n}(\tilde{p}))$.
We have already defined what a smooth map $f : M \to N$ is,
next we will define the differential of a smooth map $df_p : T_pM \to T_pN$.

4.25 Differential (definition)

Definition 4.25.1. ***Tangent Bandle***
Let M be an n-dimensional smooth manifold and $p \in M$.
The tangent bandle of M is defined as

$$TM := \sqcup_{p \in M} T_pM := \cup_{p \in M} \{p\} \times T_pM$$

Note:

- $\sqcup$ is the disjoint union.
- Note that if M is a submanifold of $\mathbb{R}^n$, then the union of the tangent spaces of M is not necessarily disjoint.
- The tangent bandle TM is a smooth manifold of dimension $2n$.

Recall, a map $f : M \to N$ is called C^k-smooth if

$\forall p \in M, \quad \exists$ a chart (U, h) of M at p and a chart (V, k) of N at $f(p)$ such that
$k \circ f \circ h^{-1}$ is k times differentiable at $h(p)$.

For each point $p \in M$, we define the differential of f at p as a linear map $df_p : T_pM \to T_{f(p)}N$, denoted as $df_p([\gamma])$ for $[\gamma] \in T_pM$.

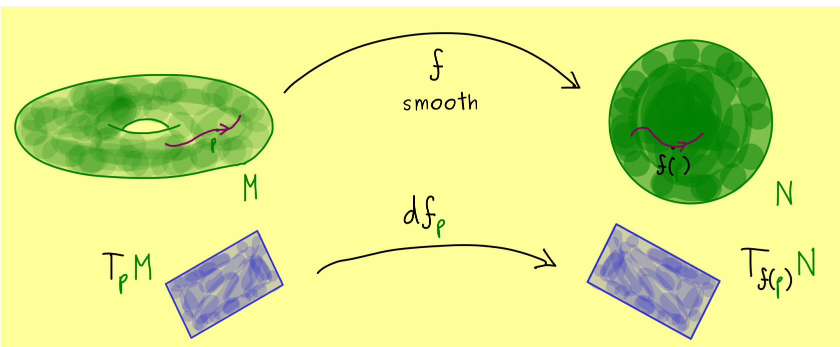


Figure 4.19: Differential of a smooth map

Definition 4.25.2. ***Differential of f at point p***
Let $f : M \to N$ be a smooth map between two smooth manifolds.
Let $p \in M$ and $[\gamma] \in T_pM$. The differential of f at p is defined as

$$df_p([\gamma]) = [f \circ \gamma] \in T_{f(p)}N$$

The linear map given by the differential is a linear approximation of the map f at the point p.

Definition 4.25.3. ***Differential***

$$df : M \to TN \quad \textit{defined as} \quad df(p) = df_p$$

Example for submanifolds:
Let M, N be submanifolds of $\mathbb{R}^n$, $p \in M$ and $f : M \to N$ be a smooth map.
In this case, $T_{f(p)}N = T^{sub}_{f(p)}N$.
$df_p([\gamma]_\sim) = [f \circ \gamma_p]_\sim \overset{\text{bijection}}{=} (f \circ \gamma_p)'(0)$

Example for $f : \mathbb{R}^n \to \mathbb{R}$ a smmoth map:
$df_p([\gamma]_\sim) = (f \circ \gamma)'(0) = J_f(\gamma(0)) \cdot \gamma'(0) = \nabla f(p) \cdot \gamma'(0)$
and we get that it's exactly the directional derivative of f at p along $[\gamma]$.

Chapter 5

The First Miracle: Robustness

Let f be a convex function, and let x^* be a minimizer of f.

5.1 Gradient Descent

Definition 5.1.1. *Gradient Descent*

$$x_{t+1} = x_t - \eta \nabla f(x_t) \tag{5.1}$$

It holds that:

$$f(x^*) \geq f(x_t) + \nabla f(x_t) \cdot (x^* - x_t) \tag{5.2}$$

$$0 \leq f(x_t) - f(x^*) \leq \nabla f(x_t) \cdot (x_t - x^*) \tag{5.3}$$

5.1.1 Analysis of the Gradient Descent Algorithm

$$\begin{aligned}
&\|a\|^2 = \|b\|^2 + \|a-b\|^2 \\
&\|b\|^2 = \|a\|^2 - \|a-b\|^2 = \|a\|^2 - (\|a\|^2 - 2a \cdot b + \|b\|^2) = 2a \cdot b - \|b\|^2
\end{aligned}$$

Then we have:

$$\begin{aligned}
\|x^* - x_t\|^2 - \|x^* - x_{t+1}\|^2 &= -2\eta(x^* - x_t) \cdot \nabla f(x_t) - \eta^2 \|\nabla f(x_t)\|^2 \\
&= 2\eta(x_t - x^*) \cdot \nabla f(x_t) - \eta^2 \|\nabla f(x_t)\|^2 \\
&\geq 2(f(x_t) - f(x^*)) - \eta^2 L^2
\end{aligned}$$

Where the last inequality follows from the convexity and the Lipschitz continuity of f.

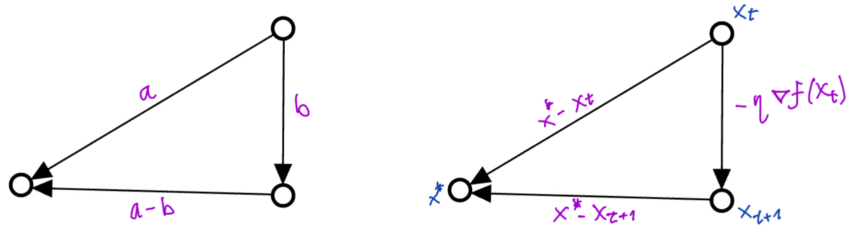


Figure 5.1: Gradient Descent

Then if we sum the above inequality from $t = 1$ to T, we get:

$$\sum_{t=1}^{T}(f(x_t) - f(x^*)) \leq \frac{\|x_1 - x^*\|^2}{2\eta} + \frac{\eta L^2}{2}T$$

In fact, this is a specific case of the Fundamental Inequality of Optimization.

Theorem 5.1.1. *Fundamental Inequality of Optimization (unconstrained version)*
Suppose $x_{t+1} = x_t - \eta g_t$ *for all* t*, where* $g_1, \ldots, g_T \in \mathbb{R}^d$ *are arbitrary vectors. Then for all* $x^* \in \mathbb{R}^d$ *it holds that*

$$\sum_{t=1}^{T} g_t \cdot (x_t - x^*) \leq \frac{\|x_1 - x^*\|^2}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\|g_t\|^2.$$

Proof. Fundamental Inequality of Optimization
The proof tracks $\|x_t - x^*\|^2$ as a "potential". First write

$$\|x_{t+1} - x^*\|^2 = \|(x_t - x^*) - \eta g_t\|^2 = \|x_t - x^*\|^2 - 2\eta g_t \cdot (x_t - x^*) + \eta^2\|g_t\|^2,$$

that is,

$$\|x_t - x^*\|^2 - \|x_{t+1} - x^*\|^2 = 2\eta g_t \cdot (x_t - x^*) - \eta^2\|g_t\|^2.$$

Summing over $t = 1, \ldots, T$ and telescoping terms, we obtain

$$\|x_1 - x^*\|^2 - \|x_{T+1} - x^*\|^2 = 2\eta\sum_{t=1}^{T} g_t \cdot (x_t - x^*) - \eta^2\sum_{t=1}^{T}\|g_t\|^2.$$

Organizing terms, we conclude:

$$\sum_{t=1}^{T} g_t \cdot (x_t - x^*) \leq \frac{\|x_1 - x^*\|^2 - \|x_{T+1} - x^*\|^2}{2\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\|g_t\|^2.$$

$\square$

osing eta

Chapter 6

The Second Miracle: Potential Based

6.1 Experts Problem

At each time step, the player picks an action $I_t \in [n]$ (we have n experts) and the adversary picks a loss vector $l_t \in 0, 1^n$. The player incurs loss $l_t(I_t)$ and the goal is to minimize the regret:

$$\text{Regret}_T(i) = \sum_{t=1}^{T} (l_t(I_t) - l_t(i)) \tag{6.1}$$

We consider the case where in each time step the player chooses an action from a distribution $\vec{p}$ over the n experts (a vector from the simplex):

$$\vec{p} \in \triangle_n = \{\vec{p} \in \mathbb{R}^n_+ : p_i \geq 0, \sum_{i=1}^{n} p_i = 1\}$$

Approach 1: Gradient Descent

We can use gradient descent on $f_t(\vec{p_t}) = \vec{l_t} \cdot \vec{p}$, where $\vec{l_t}$ is the loss vector at time t. It holds that $\nabla f_t(\vec{p_t}) = \vec{l_t}$. We can use the analysis of the gradient descent algorithm for gradient descent of convex functions varying in time.
Let $q \in \triangle_n$ be any distribution. Then we have:

$$\begin{aligned} f_t(q) \geq f_t(\vec{p_t}) + \nabla f_t(q) \cdot (q - \vec{p_t}) \implies \\ f_t(\vec{p_t}) - f_t(q) \leq \nabla f_t(q) \cdot (\vec{p_t} - q) \end{aligned}$$

Then:

$$\begin{aligned} \|q - p_t\|^2 - \|q - p_{t+1}\|^2 = -2\eta(q - p_t) \cdot \nabla f_t(p_t) - \eta^2 \|\nabla f_t(p_t)\|^2 \implies \\ f_t(\vec{p_t}) - f_t(q) \leq \nabla f_t(q) \cdot (\vec{p_t} - q) = \frac{1}{2\eta} \left(\|q - \vec{p_t}\|^2 - \|q - \vec{p}_{t+1}\|^2 \right) + \frac{\eta}{2} \|\nabla f_t(\vec{p_t})\|^2 \implies \\ \sum_{t=1}^{T} (f_t(\vec{p_t}) - f_t(q)) \leq \frac{1}{2\eta} \left(\|q - \vec{p}_1\|^2 - \|q - \vec{p}_{T+1}\|^2 \right) + \frac{\eta}{2} \sum_{t=1}^{T} \|\nabla f_t(\vec{p_t})\|^2 \\ \leq \frac{1}{2\eta} \|q - \vec{p}_1\|^2 + \frac{\eta}{2} \sum_{t=1}^{T} \|\nabla f_t(\vec{p_t})\|^2 \\ \leq \frac{1}{\eta} + \frac{\eta}{2} Tn = \mathbf{O}(\sqrt{Tn}) \end{aligned}$$

We have used the facts that:

- Both q and $\vec{p}_1$ are distributions, so $\|q - \vec{p}_1\|^2 \leq 2$.
- $\|\nabla f_t(\vec{p}_t)\|^2 \leq n$ (as the loss vector is in $0, 1^n$).

we can see that in this case, the rate of convergence DO depend on the dimension of the problem, in contrast to the non-varying case. The fact that the rate of convergence DO NOT depend on the dimension of the problem in GD is one of the reasons why GD is so useful in practice.

2: M

Approach 2: Multiplicative Weights Update (MWU)

6.2 Mirror Descent

Endow K with a Riemannian structure: $< \cdot, \cdot >_x$ for each $x \in K$. Before:

$$x_{t+1} = x_t - \eta \nabla f(x_t) \to f(x + dx) \approx f(x) + \nabla f(x) \cdot dx \tag{6.2}$$

Chapter 7

The Third Miracle:

Chapter 8

The Fourth Miracle:

Chapter 9

The Fifth Miracle: