Analysis II - HKU

Fei Ye

December 8, 2015

Contents

1 Topology on Euclidian spaces

Denote by $\mathbb{R}$ the set of all real numbers, $\mathbb{C}$ the set of complex numbers, and $\mathbb{R}^n = \{(x_1, \dots, x_n) \mid x_i \in \mathbb{R}\}$ the set of n-tuples of real numbers.

In this section, I will introduce the (metric) topology in $\mathbb{R}^n$, functions on $\mathbb{R}^n$ and continuity. During the discussion, I am going to introduce you some formal definitions. However, we will practically focus on $\mathbb{R}^n$ with the Euclidian metric topology. Recall that a function f on a subset X is a relation $f : X \to \mathbb{R}$ such that for any $x \in X$ there is a unique $y \in \mathbb{R}$ such that $f(x) = y$. A relation $f : D \subset \mathbb{R}^n \to \mathbb{R}^m$ whose a components are functions is usually called a mapping or a vector valued function or simply a function.

Let's recall some basic knowledge about $\mathbb{R}$.

- We have natural algebraic operations: addition and multiplication.
- Given a point $x \in \mathbb{R}$, the distance $d(x, y)$ from x to another point $y \in \mathbb{R}$ is defined as the absolute value $|x - y|$ of their difference.
- A subset $U \in \mathbb{R}$ is open if for any x in U there is a $\delta \geq 0$ such that the interval $(x - \delta, x + \delta) = \{y \in \mathbb{R} \mid |y - x| < \delta\}$ is contained in U. Any interval $(a, b) = \{x \in \mathbb{R} \mid a < x < b\}$ is open.
- A subset $X \subset \mathbb{R}$ is compact if the limit of any convergent sequence $\{x_i\}$ of numbers in X is still in X. Closed intervals are compact.
- A function f on X is continuous at $x_0 \in X$, if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $|f(y) - f(x)| < \varepsilon$ for any $y \in X$ such that $|y - x| \leq \delta$.
- Let $\{x_i\}$ is a sequence of numbers in $\mathbb{R}$. It is convergent if there exists a number $x \in \mathbb{R}$ such that for any $\varepsilon > 0$ there exists an $N \in \mathbb{Z}_+$ such that for any $i > N$, $|x_i - x| < \varepsilon$.

1.1 Vector spaces

On $\mathbb{R}^n$, there are still natural algebraic operations: component-wise addition, i.e. $(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n)$, and scalar multiplication, i.e.

$$\lambda(x_1, \ldots, x_n) = (\lambda x_1, \ldots, \lambda x_n).$$

These operations assign $\mathbb{R}^n$ an algebraic structure called vector space.

Definition 1.1. A *vector space* V over a field k is a set together with two operations, addition and scaler multiplication, satisfying the following eight axioms. An elements in V is called a vector. For any two vectors x, y in V, the addition assigns a third vector $x + y$ in V. For a number $a \in k$ and a vector $x \in V$, the scaler multiplication assigns a vector ax in V.

(a) There exists a vector 0 such that $0 + x = x$ for any vector $x \in V$.

(b) For the identity $1 \in k$, $1x = x$ for any $x \in V$.

(c) For any $x \in V$, there exists a vector $-x$ such that $x + (-x) = 0$.

(d) $x + y = y + x$ for any vectors x and y in V.

(e) $(x + y) + z = x + (y + z) = x + y + z$.

(f) $a(x + y) = ax + ay$.

(g) $(a + b)x = ax + bx$.

(h) $(ab)x = a(bx)$.

We say $\{x_1, \ldots, x_n\}$ is a *basis* of the vector space V over k if the following conditions hold:

(a) For any $x \in V$, there exist numbers $\lambda_1, \ldots, \lambda_n \in k$ such that $x = \sum_{i=1}^n \lambda_i x_i$.

(b) $\sum_{i=1}^n \lambda_i x_i = 0$ if and only if $x_i = 0$ for all i.

The *dimension* $\dim V$ of vector space V over k is defined to be the length of a basis. We say V is a *finitely dimensional* space if it has a finite basis.

Theorem 1.2. *Any two bases of a finitely dimensional vector space have the same length.*

Proof. Let $\{x_1, \ldots, x_m\}$ and $\{y_1, \ldots, y_n\}$ be two bases. Assume that $m > n$. There exist numbers α_{ij}, $i = 1, \ldots n$, $j = 1, \ldots, m$ such that

$$y_i = \sum_{j=1}^{m} \alpha_{ij} x_j.$$

Since $m > n$, there exist numbers a_i, for $i = 1, \ldots, n$ such that $(a_1, \ldots, a_n) \neq 0$ and $\sum_{i=1}^{n} a_i \alpha_{ij} = 0$ for $j = 1, \ldots, m$ Therefore, $\sum_{i=1}^{n} a_i y_i = 0$ which contradicts to the independency of $y_1, \ldots, y_n$. Thus, $m \leq n$ and similarly, $n \leq m$. Hence, $m = n$. Q.E.D

One can check that $\mathbb{C}^n$ is a n-dimensional vector spaces over $\mathbb{C}$ but a $2n$-dimensional vector space over $\mathbb{R}$.

Before discussing distance on $\mathbb{R}^n$, recall that on $\mathbb{R}^n$, we define the *inner product* $\langle x, y\rangle$ of two vector x and y as $\langle x, y\rangle = x_1 y_1 + \cdots + x_n y_n$.

Theorem 1.3. *The inner product on $\mathbb{R}^n$ has the following properties.*

(a) $\langle x, y\rangle = \langle y, x\rangle$, *(symmetry).*

(b) $\langle ax, y\rangle = a\langle x, y\rangle = \langle x, ay\rangle$
$\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$, *(bilinearity).*

(c) $\langle x, x\rangle \geq 0$ *and* $\langle x, x\rangle = 0$ *if and only if* $x = 0$, *(positive definiteness).*

(d) $\langle x, y\rangle \leq \sqrt{\|x\| \cdot \|y\|}$ *and the equality holds if and only if* $x = ay$ *for some* $a \in \mathbb{R}$, *(Cauchy-Schwart inequality).*

Proof. The first three are easy. For Cauchy-Schwart inequality, we use the fact that $\|x - ay\| \geq 0$ for any $a \in \mathbb{R}^n$. Notice that the function

$$f(a) = a^2 \sum_{i=1}^{n} y_i^2 - 2a \sum x_i y_i + \sum_{i=1}^{n} x_i^2 = \|x - ay\| \geq 0.$$

Therefore the minimum $\dfrac{\|x\|\|y\| - \langle x, y\rangle^2}{\|y\|}$ of $f(a)$ is nonnegative which implies that $\sqrt{\|x\|\|y\|} \geq \langle x, y\rangle$ and the equality holds if and only if the minimum of $f(a) = 0$ for some a. By positive definiteness, we know that $x = ay$. for some $a \in \mathbb{R}$. Q.E.D

In the above theorem, we see the first three consequence are independent. In general, for any vector space V over $\mathbb{R}$, a bilinear function $\langle \cdot, \cdot \rangle : V \times V \to k$ is called an inner product if it satisfies the condition symmetry, bilinearity, and positive definiteness.

Using inner product, we define the distance $d(x, y)$ between two points $x, y \in \mathbb{R}^n$ as $d(x, y) = \|x - y\|$, where $\|x\| = \sqrt{\langle x, x \rangle}$ is the of x. One can check that the distance d satisfies the following conditions.

(M1) $d(x, y) = d(y, x)$,

(M2) $d(x, y) \geq 0$ and equality holds if and only if $x = y$.

(M3) $d(x, y) + d(y, z) \geq d(x, z)$.

In general, we call a function $d : X \times X \to \mathbb{R}$ which satisfies (M1)-(M3) a on X and call X a metric space. The metric $d(x, y) = \|x - y\|$ on $\mathbb{R}^n$ is called the and $\mathbb{R}^n$ is called the Euclidian space.

You may ask if the Euclidian metric is the unique metric on $\mathbb{R}^n$. In fact, it is not. There is another metric called the *sup metric*: $d(x, y) = \max_{i=1}^n \{|x_i - y_i|\}$.

We will see that the sup metric and the Euclidian metric on $\mathbb{R}^n$ define exactly the same metric topology.

1.2 Metric topology

Let me first introduce you the definition of topology on sets.

Definition 1.4. A *topology* on a set X is a collection $\mathscr{T} = \{U \mid U \subset X\}$ of subsets covering X, i.e. $\cup_{U \subset \mathscr{T}} U = X$, such that $\mathscr{T}$ contains the empty set, X, the union of any two elements of $\mathscr{T}$, and the intersection of any finitely many elements of $\mathscr{T}$. A member in $\mathscr{T}$ is called an *open* subset. A subset $Z \subset X$ is *closed* if the *complement* $Z^C = X \setminus Z$ is open. A set X with a topology $\mathscr{T}$ is called a *topological space*. An *open covering* $\mathscr{O}$ of a topological space $(X, \mathscr{T})$ is a collection of open subsets $\mathscr{O} = \{U \mid U \text{ is open in } X\}$ such that $X = \cup_{U \in \mathscr{O}} U$. A topological space $(X, \mathscr{T})$ is *compact* if every open covering $\mathscr{O}$ has a finite open subcovering, i.e. a finite subset $\mathscr{F} \subset \mathscr{O}$ such that $X = \cup_{U \in \mathscr{F}} U$. The *closure* $\bar{A}$

of a subset A is the intersection of all closed subset containing A. The *interior* A^{int} of a subset A is the union of all open subsets contained in A. The *exterior* A^{ext} of a subset A is the union of all open subsets contained in the complement A^C. The *boundary* ∂A of A is defined as $\bar{A} \setminus A^{int}$.

Example 1.5. The boundary of *unit disk* $D = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 < 1\}$ is the *unit circle* $C = \{(x, y) \in \mathbb{R}^2 \mid x^2 + y^2 = 1\}$ in $\mathbb{R}^2$.

Lemma 1.6. *A point x is in the boundary ∂X if and only if for any $\varepsilon > 0$ the intersections of $B(x, \varepsilon)$ with X^{int} and X^{ext} are both non-empty.*

Proof. By definition $\partial X = \overline{X} \setminus X^{int} = \overline{X} \cap \overline{X^C}$. Therefore, $x \notin \partial X$ if only if there exists a number ε such that $B(x, \varepsilon) \subset \overline{X}^C \cup \overline{X^C}^C$. Q.E.D

Let $(X, \mathscr{T})$ be a topological space. Any subset $Y \subset X$ admits a *subspace topology* $\mathscr{T}_Y$ given by $\mathscr{T}_Y = \{U \cap Y \mid U \in \mathscr{T}\}$. We call $(Y, \mathscr{T}_Y)$ a *topological subspace*.

Example 1.7. On the real line $\mathbb{R}$, we can define a topology as $\mathscr{T} = \{(a, b) \mid a, b \in \mathbb{R} \cup \{\pm\infty\}\}$, i.e. the collection of open intervals.

Note that one can define a topology by telling what are closed subsets.

For a metric space, there is a natural way to define topology.

Definition 1.8. Let (X, d) is a metric space and $o \in X$ be a point. Given $\varepsilon > 0$, we call the set

$$B(o, \varepsilon) = \{x \in X \mid d(o, x) < \varepsilon\}$$

a *ε-neighborhood* of o. We can define a *metric topology* $\mathscr{T}^d$ on X in the following way. A subset U of X is *open subset* if for each $x \in U$, there exists a number $\varepsilon > 0$ such that $B(x, \varepsilon)$ is contained in U.

Theorem 1.9. *Let (X, d) be a metric space. Then finite intersection and arbitrary union of open sets of X are open. Hence, X is topological space.*

Proof. Let U_1 and U_2 be two open subsets. For any point $x \in U_1 \cap U_2$, there exist ε_1 and ε_2 such that $U(x, \varepsilon_1) \subset U_1$ and $U(x, \varepsilon_2) \subset U_2$. Let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. Then that $B(x, \varepsilon) \subset B(x, \varepsilon_1) \cap B(x, \varepsilon_2) \subset U_1 \cap U_2$. Hence $U_1 \cap U_2$ is open. Take an arbitrary union $U = \cup_{\alpha\in A} U_\alpha$ of open sets. Let $x \in U$. Then $x \in U_\alpha$ for some α. Hence, there exists an ε-neighborhood $B(x, \varepsilon) \subset U_\alpha \subset U$. Hence U is open. Q.E.D

From now on, we assume that $X \subset \mathbb{R}^n$ is equipped with the Euclidian metric topology.

1.3 Continuity

Definition 1.10. Let $X \subset \mathbb{R}^m$ and $Y \subset \mathbb{R}^n$ be two topological subspaces. A function $f : X \to Y$ is *continuous* at a point $x_0 \in X$ for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $d(f(x_0), f(y)) < \varepsilon$ for any $y \in B(x_0, \delta)$. The function f is called a *continuous function* if f is continuous at any point $x \in X$.

As you may expect, the space of continuous functions is closed under usual algebraic operations.

Theorem 1.11. *Let f and g be two continuous functions on a subspace $X \subset \mathbb{R}^n$. Then the functions $f \pm g$, fg and f/g with $g \neq 0$ are continuous.*

Proof. Follows from definition. Q.E.D

As you may suspect that continuity may change as the metric changes. However, as long as two metrics defines the same topology, the continuity remains the same.

Theorem 1.12. *Let X and Y be two subspaces in Euclidian spaces. A function $f : X \to Y$ is continuous if and only if for any open subset $U \subset Y$ the preimage $f^{-1}(U)$ is an open subset of X.*

Proof. We prove the sufficiency. Let x be a point in X. For any $\varepsilon > 0$, the subset $B(f(x), \varepsilon) = \{y \in Y \mid d(f(x), y) < \varepsilon\}$ is open by definition. Since $f^{-1}(B_\varepsilon)$ is open and $x \in f^{-1}(B(f(x)\varepsilon)$, then there exists a $\delta > 0$ such that $B(x, \delta) = \{y \in x \mid d(x, y) < \delta\}$ is in $f^{-1}(B(f(x), \varepsilon)$. Hence, for any $y \in B(x, \delta)$, we have $f(x') \in B(f(x), \varepsilon)$. By definition, f is continuous at x. The necessary can be proved similarly. Q.E.D

Corollary 1.13. *Let $f : X \to Y$ and $g : Y \to Z$ be two continuous functions. Then the composition $g \circ f : X \to Z$ is a continuous function. In particular, for any subspace $S \in X$, $f|_S : S \to Y$ is continuous.*

1.4 Compactness

Now we generalize limits and compactness to $\mathbb{R}^n$. Compactness is very important in analysis.

Definition 1.14. An infinite sequence $\{x_n\}$ of points in $\mathbb{R}^n$ is *convergent* if there exists a point $x \in \mathbb{R}^n$ such that the following hold: for any $\varepsilon > 0$, there is a number $N \in \mathbb{Z}_+$ such that $d(x, x_n) < \varepsilon$ for all $n > N$. We call $x = \lim_{n\to\infty} x_n$ the *limit* of $\{x_n\}$.

Theorem 1.15. *A topological subspace $X \subset \mathbb{R}^n$ is closed if and only if the limit x of any convergent sequence $\{x_n\}$ lying in X is still in X.*

Proof. Assume that X is closed. Then X^C is open. If $x \notin X$, then there is a $\delta > 0$ such that $B(x, \delta) \subset X^C$. But $\{x_n\} \cap B(x, \delta) \neq \emptyset$ by definition of convergence. It is a contradiction.

Let x be a point in $\partial\overline{X}$. Then $x \in \overline{X} \cap \overline{X^C}$ which implies that for each n the intersection $B(x, \frac{1}{n}) \cap X^{int} \neq \emptyset$. Therefore, there is a convergent sequence $\{x_n\}$ lying in X whose limit is x. We conclude that $\overline{X} = X$ and X is closed. Q.E.D

Using convergence of infinite sequences, we can give another characterization of compactness. A space X that every infinite sequence $\{x_n\}$ in X has a subsequence x_{i_n} converging to a point $x \in X$ is called *sequentially compact*. In general, sequentially compact is not the same as compact. However, for metric topology, they are the same. We will show the equivalence in three steps which imply the Heine-Borel Theorem.

Lemma 1.16. *Let X be a compact subspace in $\mathbb{R}^n$. Then any infinite sequence $\{x_n\}$ has a subsequence $\{x_{i_n}\}$ converges to a point $x \in X$.*

Proof. It is enough to show that there exists a point $x \in X$ such that for any $n \in \mathbb{Z}_+$ the set $B(x, \frac{1}{n}) \cap \{x_n\}$ consists of infinitely many points. Assume in the contrary that for any point $y \in X$, there exists a number $n(x)$ such that $B(y, \frac{1}{n(x)}) \cap \{x_n\}$ is finite. Since X is compact, then there will be finitely many points $y_1, \ldots, y_m$ such that $\{x_n\} \subset X \subset \cup_{i=1}^m B(y_i, \frac{1}{n(y_i)})$ which is a contradiction. Q.E.D

We say a subspace $X \subset \mathbb{R}^n$ is *bounded* if there exist a point in b and number $r > 0$ such that $X \subset B(b, r) \subset \mathbb{R}^n$.

Lemma 1.17. *Let X be a subspace in $\mathbb{R}^n$. If every infinite sequence $\{x_n\}$ in X has a subsequence x_{i_n} that converges to a point $x \in X$. Then X is closed and bounded.*

Proof. The closedness is follows from the fact that a convergent sequence has the same limit as any of its convergent subsequence.

If X is not bounded. Then for each n there is a point x_n such that $x_n \notin B(b, n)$. But by our assumption, the sequence $\{x_n\}$ has a subsequence $\{x_{i_n}\}$ converges to a point x. Hence $d(b, x_{i_n})$ converges to $d(b, x)$ which is a contradiction. Q.E.D

Lemma 1.18. *A closed and bounded subspace $X \in \mathbb{R}^n$ is compact.*

Proof. It is enough to show that the n-cell $C(a, b) = [a_1, b_1] \times [a_n, b_n]$ is compact, where $a = (a_1, \ldots, a_n)$ and $b = (b_1, \ldots, b_n)$. In fact, any closed subset of a compact set is compact (see Exercise **??**).

We first prove that a closed interval $[a, b]$ is compact. Let $\mathscr{O}$ is an open cover of $[a, b]$. Set

$$A = \{x \in [a, b] \mid [a, x] \text{ is covered by finitely many open subset of } \mathscr{O}\}.$$

Then A is a bounded subset of $[a, b]$. We would like to show that $b \in A$.

Let β be the least upper bound of A, i.e. $\beta \geq x$ for any $x \in [a, b]$ and for any $\varepsilon > 0$, there exists a $x \in (a, b)$ such that $\beta < x + \varepsilon$. We are going to show that $\beta \in [a, b]$ and $b = \beta$. By the choice of β, there is a open subset $U \in \mathscr{O}$ such that $(\beta - r, \beta + r) \in U$ and $(\beta - r, \beta + r) \cap A \neq \emptyset$. Let $x \in (\beta - r, \beta + r) \cap A$, then $[a, x]$ is covered by finitely many open subsets, say $U_1, \ldots, U_t$. Thus $[a, \beta]$ is covered by $U_1, \ldots U_t$ and U which implies that $\beta \in A$.

To show that $b = \beta$, assume instead that $\beta < b$. Then there exists a point $x' \in [\beta, b]$ such that $x' > \beta$ and $[\beta, x'] \subset (\beta - r, \beta + r) \subset U$ for some r. Then $[a, x']$ is covered by finitely many open subsets. Hence, $x' \in A$ which contradicts to that β is the least upper bound of A.

The compactness of C(a, b) follows from Corollary 1.20. Q.E.D

Theorem 1.19. *Let $X \subset \mathbb{R}^m$ be a compact subset and $y \in \mathbb{R}^n$ be a point. Let Ø be an open covering of $X \times \{y\}$. Then there is a open neighborhood $x \in U\mathbb{R}^n$ such that $X \times U$ is covered by finitely many members Ø.*

Proof. For each $(x, y) \in X \times \{y\}$, there is an open subset W_x in $\mathscr{O}$ containing an open neighborhood $V_x \otimes U_x$ of (x, y). By compactness of X, there are finitely many such $V_x \times U_x \subset W_x$, say $V_1 \times U_1, \ldots, V_r \times U_r$, covering $X \times \{y\}$. Set $U = \cap_{i=1}^r U_i$. Then $X \times U$ is covered by finitely many $W \in \mathscr{O}$. Q.E.D

Corollary 1.20. *Let $A \subset \mathbb{R}^m$ and $B \in \mathbb{R}^n$ be two compact subsets. Then $A \times B \in \mathbb{R}^m \times \mathbb{R}^n$ is compact.*

Proof. Let $\mathscr{O}$ be an open covering of $A \times B$. Then $\mathscr{O}$ is an open covering of $A \times \{x\}$ for any $x \in B$. By Theorem 1.19, there is an open neighborhood $U_x \subset \mathbb{R}^n$ such that $A \times U_x$ is covered by finitely many elements in $\mathscr{O}$. Since B is compact, there are finitely many elements in $\{U_x \mid x \in B\}$ covering B. Hence, there are finitely many elements in $\mathscr{O}$ covering $A \times B$. Q.E.D

Theorem 1.21 (Heine-Borel)**.** *A subspace $X \subset \mathbb{R}^n$ is compact if and only if X is closed and bounded.*

Proof. It follows from Lemma 1.16,1.17 and 1.18. Q.E.D

In fact, one can prove directly, without passing through sequential compact, that a closed and bounded subspace is compact.

In general, the image of a closed (open) subset might not be closed (open). However, the image of compact subset is compact which illustrates the importance of compactness.

Theorem 1.22. *Let $f : X \to Y$ be a continuous function of metric spaces. Then $f(X)$ is compact if X is compact.*

Proof. Let $\mathscr{T}_Y = \{U\}$ be open over of Y. Denote by $V_U = f^{-1}(U \cap f(X))$. Then $\{V_U \mid U \in \mathscr{T}_Y\}$ is an open cover of X. Since X is compact, then there are finitely many open subset $V_{U_1} \dots V_{U_m}$ covering X. Hence $U_1, \dots, U_m$ covers $f(X)$. Q.E.D

As an application of Theorem 1.22, you see that every continuous function $f : \mathbb{R}^n \to \mathbb{R}$ has extremal values over a compact subset.

Here is another example illustrating the importance of compactness.

Definition 1.23. A function $f : X \to \mathbb{R}^m$ is *uniformly continuous* if for any $\varepsilon > 0$ there exists a $\delta > 0$ such that $d(f(x), f(y)) < \varepsilon$ for any $x, y \in X$ whose distance $d(x, y) < \delta$.

It is clear that uniform continuous implies continuous. However, the converse may not be true without compactness.

Example 1.24. Consider the function $f : \mathbb{R} \setminus \{0\} \to \mathbb{R}$ given by $f(x) = \frac{1}{x}$. Then f is continuous but not uniformly continuous. In fact, for any $\delta > 0$, there is a positive integer m such that $\frac{1}{m} < \delta$. Then $d(\frac{1}{m}, \frac{1}{m+1}) < \delta$. But $|f(\frac{1}{m+1} - f(\frac{1}{m})| = 1$.

Theorem 1.25 (Uniform continuity)**.** *Let $X \subset \mathbb{R}^n$ be a compact subspace and $f : X \to \mathbb{R}^m$ be a continuous function. Then f is uniformly continuous.*

Proof. Given a number $\varepsilon > 0$, by continuity of f, for each $x \in X$ there is a positive number $2\delta(x)$ such that $d(f(x), f(y)) < \frac{\varepsilon}{2}$ for any y in $B(x, 2\delta(x))$.

Note that the open balls $B(x, \delta(x))$ covers X. Therefore, by compactness X is compact, there are finitely many points $x_1, \dots, x_m$ such that $B(x_i, \delta(x_i))$, $i = 1, \dots, m$, covers X. Let $\delta = \min\{\delta(x_i)\}$. Then for any x, y, such that $d(x, y) < \delta$, there is a point x_i such that $x \in B(x_i, \delta(x_i))$ $d(x_i, y) < d(x_i, x) + d(x, y) \le 2\delta(x_i)$. Therefore, $y \in B(x_i, 2\delta(x_i))$ which implies shat $d(f(x_i), f(y)) < \frac{\varepsilon}{2}$. Therefore, for any x, y such that $d(x, y) < \delta$, we have

$$d(f(x), f(y)) < d(f(x_i), f(y)) + d(f(x_i), f(x)) < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Q.E.D

1.5 Connectedness

Another important topological concept is connectedness.

Definition 1.26. Two subsets A and B of $\mathbb{R}^n$ are separated t if $A \cap \overline{B} = \emptyset$ and $\overline{A} \cap B = \emptyset$. A metric space X is connected if X cannot be covered by two separated nonempty open sets U_1 and U_2.

Theorem 1.27. *The closed interval $[a, b] \in \mathbb{R}$ is connected.*

Proof. Let U_1 U_2 be separated nonempty open subsets of $\mathbb{R}$ and $[a, b] \subset U_1 \cup U_2$. If there are two points $x, y \in [a, b]$ such that $x \in U_1$ and $y \in U_2$. We may assume that $x < y$. Let $a = \partial U_1 \cap [x, y]$ and $b = \inf \partial U_2 \cap [x, y]$. Then $b > a$ and $c = \frac{b+a}{2} \in [x, y] \subset [a, b]$ but $c \notin U_1 \cup U_2$. Q.E.D

Theorem 1.28 (Intermediate-value theorem)**.** *Let X be a connected and $f : X \to Y$ be a continuous function. Then $f(X)$ is connected.*

In particular, if $Y = \mathbb{R}$ and if $f(x_0) < f(x_1)$ for some points $x_0, x_1 \in X$, then for any $f(x_0) < r < f(x_1)$ there is a point $x \in X$ such that $f(x) = r$.

Proof. Follows from the definitions. Q.E.D

1.6 Completeness

Another concept closely related to compact is completeness.

Definition 1.29. An infinite sequence $\{x_n\}$ is called a *Cauchy sequence* if for any $\varepsilon > 0$, there exist $N > 0$ such that $d(x_m, x_n) < \varepsilon$ for any $m, n > N$. A metric space X is *complete* if every Cauchy sequence has a limit point in X.

Being complete is weaker than being compact.

Theorem 1.30. *A sequentially compact subspace X in $\mathbb{R}^n$ is complete.*

Proof. By Lemma 1.16, every infinite sequence in X has a subsequence converges to a point. But a Cauchy sequence has a convergent subsequence must be convergent, moreover, they have the same limit point. Therefore X is complete. Q.E.D

Example 1.31. The interval $[a, +\infty)$ is complete but not compact.

2 Differentiations

In this section, we will discuss the partial derivatives, differential, differentiability, inverse function theorem, implicit function theorem, free extremum problems, constrained extremum problem and method of Lagrange multipliers.

To motivate some definitions, we will frequently identify a function $f : X \to Y$ with its graph $\Gamma_f = \{(x, f(x)) \mid x \in X\} \subset X \times Y$.

Given a single variable continuous function $f : \mathbb{R} \to \mathbb{R}$, the derivative $f'(x_0)$ of f at the point x_0 is the limit

$$\lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0} = \lim_{t \to 0} \frac{f(x_0 + t) - f(x_0)}{t}.$$

The geometric meaning of $f'(x_0)$ is the slope of the tangent line $y = f'(x_0)(x - x_0) + f(x_0)$.

Let $f : \mathbb{R}^m \to \mathbb{R}^n$ be a continuous function. To generalize the derivative, geometrically, there are two obvious choices, the slope of a tangent line or the “slope" of tangent plane, i.e. the space spanned by all tangent lines. But what is the “slope" of a plane?

Algebraically, the slope of a line defines a linear relation between x and y. More precisely, if the line passing through origin, then we simply have $y = sx$. A higher dimensional analogue of slope would be a linear transformation between “x-plane" and “y-plane". More precisely, for a linear subspace V of dimension at most m in $\mathbb{R}^m \times \mathbb{R}^n$, we can find a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ such that $V = \Gamma_T = \{(x, T \cdot x) \mid x \in \mathbb{R}^n\}$. In fact, there is a linear transformation $s : \mathbb{R}^m \to V$ such that the composition of $s \circ p$ with the linear projection $p : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ is the identity on V. Compose with the projection $\mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$, we get the linear transformation T. If $\dim V = m$, then the linear transformation T is unique.

2.1 Derivatives

According to the two choices, we can make two definitions.

Definition 2.1. Let X be a subspace in $\mathbb{R}^m$ and $f : X \to \mathbb{R}^n$ be a function. Suppose that X contains a neighborhood of $x_0 \in X$. Given a nonzero vector $v \in \mathbb{R}^m$, define

$$D_v(x_0) = \lim_{t\to 0} \frac{f(x_0 + tv) - f(x_0)}{t}$$

provided that the limit exists. We call $D_v(x_0)$ the *directional derivative* of f at x_0 in the direction u. Let $e_i = (0, \dots, 0, \overset{i\text{-th}}{1}, 0, \dots, 0)$. We call $D_i f = D_{e_i}$ the i-th *partial derivative*.

Definition 2.2. Let X be a subspace in $\mathbb{R}^m$ and $f : X \to \mathbb{R}^n$ be a function. Given a point $x_0 \in X$. Suppose a neighborhood of x_0 is in X. We say f is *differentiable* at x_0 if there is a $n \times m$ matrix M such that

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

The matrix M is called the *derivative* of f at x_0 and denoted as $Df(x_0)$.

In the above definition, you may worry that M is not unique. In fact, it is unique. To see that, let $u = \frac{h}{\|h\|}$ and M' be another matrix such that

$$\lim_{h\to 0} \frac{\|f(x_0 + h) - f(x_0) - M' \cdot h\|}{\|h\|} = 0.$$

Then, taking the difference of the two limits, we have $(M' - M)u = 0$ for any $h \to 0$. Hence u is arbitrary which implies that $M' - M = 0$.

Are “directional derivative" and “derivative" equivalent? The answer is negative. One can easily show that the existence of derivative implies that of directional derivative. But the converse is not true in general.

Theorem 2.3. *Let X be a subspace in $\mathbb{R}^m$ and $f : X \to \mathbb{R}^n$ be a function differentiable at x_0. Then for any vector $u \in \mathbb{R}^n$, the directional derivative $f'_v(x_0)$ exists and $f'_v(x_0) = Df(x_0) \cdot v$.*

Proof. For any vector v, let $h = tv$, where $t \in \mathbb{R}$. Since f is differentiable at x_0, then

$$\lim_{t\to 0} \frac{\|f(x_0 + tv) - f(x_0) - tDf(x_0) \cdot v\|}{t\|v\|} = 0$$

which implies that $\lim_{t\to 0} \frac{f(x_0 + tv) - f(x_0)}{t} = Df(x_0) \cdot v$ by triangle inequality. Q.E.D

Corollary 2.4. *Let X be a subspace in $\mathbb{R}^m$ and $f : X \to \mathbb{R}^n$ be a function differentiable at x_0. Then*

$$Df(x_0) = (D_1 f(x_0), \dots, D_n f(x_0))^T,$$

where $\cdot^T$ means the transpose.

Proof. By definition $Df(x_0)$ is a $n \times m$ matrix. Then its i-th entry is $Df(x_0) \cdot e_i = D_i f(x_0)$. Q.E.D

Let X be a subspace in $\mathbb{R}^m$ and $f : X \to \mathbb{R}^n$ be a function. Denote by π_i the projection from $\mathbb{R}^m \to \mathbb{R}$ given by $\pi_i(x) = x_i$, where $x = (x_1, \dots, x_m)$. Let $f_i : X \to \mathbb{R}$ be the function $\pi \circ f$ called the i-th component function of f.

Corollary 2.5. *Let X be a subspace in $\mathbb{R}^m$ and $f : X \to \mathbb{R}^n$ be a function differentiable at x_0. Let f_i be i-th component function of f so that*

$$f = \left(f_1 \cdots f_n\right)^T.$$

Then

$$Df(x_0) = \left(Df_1(x_0) \cdots Df_n(x_0)\right)^T = \left(D_i f_j(x_0)\right).$$

The $n \times m$ matrix $\left(D_j f_i(x_0)\right)$ is called the *Jacobian matrix* of f.

Example 2.6. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function defined by

$$f(x, y) = \begin{cases} \frac{x^2 y}{x^4 + y^2} & (x, y) \neq (0, 0) \\ 0 & (x, y = (0, 0)) \end{cases}.$$

For any nonzero vector $v = (x_0, y_0) \in \mathbb{R}^m$, consider directional derivatives at $(0, 0)$. We see that

$$\lim_{t \to 0} \frac{f(tv) - f(0)}{t} = \lim_{t \to 0} \frac{t^3 a^2 b}{t^3(t^2 a^4 + b^2)} = \begin{cases} \dfrac{a^2}{b} & b \neq 0 \\ 0 & b = 0 \end{cases}$$

However, f is not differentiable at $(0, 0)$. In fact, f is not even continuous at $(0, 0)$. A differentiable function should be continuous.

Theorem 2.7. *Let $f : X \to \mathbb{R}^n$ be a function differentiable at x_0. Then f is continuous at x_0.*

Proof. Since f is differentiable at x_0, then

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - M \cdot h\|}{\|h\|} = 0.$$

By triangular inequality, we have

$$\|f(x_0 + h) - f(x_0)\| \leq \|f(x_0 + h) - f(x_0) - M \cdot h\| + \|M \cdot h\|.$$

To show that f is continuous, it suffices to show that

$$\lim_{h \to 0} \frac{\|M \cdot h\|}{\|h\|}$$

is finite. Applying triangular inequality repeatedly, we see that

$$\frac{\|M \cdot h\|}{\|h\|} \leq \sup_{\|u\|=1} \{\|M \cdot u\|\} \leq \max\{\|M_i\| \mid i = 1, \ldots, n\}$$

where M_i is the i-th row vector of M. Q.E.D

In Example 2.6, we see that both the function and directional derivatives are not continuous. In fact, discontinuity is the only obstruction from directional derivative to derivative. To show that we will need a mean value theorem. Recall the mean value theorem in single variable analysis.

Theorem 2.8 (Mean value theorem I)**.** *Let $f : [a, b] \to \mathbb{R}$ be a continuous function and differentiable at each point in (a, b). Then there exists a point $c \in (a, b)$ such that $f(b) - f(a) = (b - a) f'(c)$.* *Q.E.D*

This theorem has its generalization to multivariable functions. We first generalize the chain rule to multivariable functions.

2.2 Chain rule I and its applications

Theorem 2.9 (Chain rule (strong))**.** *Let $X \subset \mathbb{R}^m$, $Y \subset \mathbb{R}^n$ be subspaces, and $f : X \to \mathbb{R}^n$ and $g : Y \to \mathbb{R}^l$ be functions. Assume that f is differentiable at x_0, $f(x_0) \in Z$ and g is differentiable at $f(x_0)$. Then the composition $g \circ f$ is differentiable at x_0 and*

$$D(g \circ f)(x_0) = Dg(f(x_0)) \cdot Df(x_0).$$

Proof. Let $y_0 = f(x_0)$, $\lambda = Df(x_0)$ and $\mu = Dg(y_0)$. Define

$$\begin{aligned} \varphi(x) &= f(x) - f(x_0) - \lambda \cdot (x - x_0), \\ \psi(y) &= g(y) - g(y_0) - \mu \cdot (y - y_0), \\ \rho(x) &= g \circ f(x) - g \circ f(x_0) - (\mu \cdot \lambda) \cdot (x - x_0). \end{aligned} \tag{1}$$

Then we have

$$\lim_{x \to x_0} \frac{\|\varphi(x)\|}{\|x - x_0\|} = 0, \tag{2}$$

$$\lim_{y \to y_0} \frac{\|\psi(y)\|}{\|y - y_0\|} = 0. \tag{3}$$

We will show that

$$\lim_{x \to x_0} \frac{\|\rho(x)\|}{\|x - x_0\|} = 0.$$

By equalities in (1), we have

$$\begin{aligned} \rho(x) &= g \circ f(x) - g \circ f(x_0) - (\mu \cdot \lambda) \cdot (x - x_0) \\ &= \psi(f(x)) + \mu(f(x) - f(x_0)) - (\mu \cdot \lambda) \cdot (x - x_0) \\ &= \psi(f(x)) + \mu \cdot (\varphi(x) - \lambda \cdot (x - x_0)) - (\mu \cdot \lambda) \cdot (x - x_0) \\ &= \psi(f(x)) + \mu \cdot \varphi(x) \end{aligned}$$

By Exercise **??**, we know that $\|\mu \cdot \varphi(x)\| \leq \|\mu\|_{op} \|\varphi(x)\|$. Then by (2), we know that

$$\lim_{x \to x_0} \frac{\|\mu \cdot \varphi(x)\|}{\|x - x_0\|} = 0.$$

By the limit (3), we know that for any $\varepsilon > 0$, there exists a number $\delta > 0$ such that

$$\|\psi(f(x))\| < \varepsilon\|f(x) - f(x_0)\| \text{ whenever } \|f(x) - f(x_0)\| < \delta.$$

By since f is continuous at x_0, there exists a number $\delta' > 0$ such that

$$\|f(x) - f(x_0)\| < \delta \text{ whenever } \|x - x_0\| < \delta'.$$

Therefore,

$$\|\psi(f(x))\| < \varepsilon\|(f(x) - f(x_0))\| \leq \varepsilon\|\varphi(x)\| + \varepsilon\|\lambda\|_{op}\|(x - x_0)\|.$$

It then follows that

$$\lim_{x \to x_0} \frac{\|\rho(x)\|}{\|x - x_0\|} = 0.$$

Q.E.D

The chain rule has many applications.

Corollary 2.10. *Let X be a subspace in $\mathbb{R}^m$ and $f, g : X \to \mathbb{R}$ are differentiable at $x_0 \in X$. Then*

$$D(f + g)(x_0) = Df(x_0) + Dg(x_0)$$
$$D(fg) = g(x_0)Df(x_0) + f(x_0)Dg(x_0).$$

Moreover, if $g(x_0) \neq 0$, then

$$D\left(\frac{f}{g}\right)(x_0) = \frac{g(x_0)D(f)(x_0) - f(x_0)D(g)(x_0)}{g(x_0)^2}.$$

Q.E.D

As a corollary of chain rule, we can generalize the single variable mean value theorem to a multivariable version.

Theorem 2.11 (Mean value theorem II)**.** *Let X be a subspace in $\mathbb{R}^m$ and $f : X \to \mathbb{R}$ be a function. Assume that X contains a line segment*

$$L = \{x_0 + tv \mid x_0 \in X, v \in \mathbb{R}^m, t \in [0, 1]\}$$

and f is differentiable along the interior of L. Then there exits a number $c \in [0, 1]$ such that

$$f(x_0 + v) - f(x_0) = Df(x_0 + cv) \cdot v.$$

Proof. Let $l : \mathbb{R} \to \mathbb{R}^m$ be the linear function $l(t) = x_0 + tv$. Apply the mean value them and change rule to $f \circ l$, we see that there is a number $c \in [0, 1]$ such that

$$f(x_0 + v) - f(x_0) = f \circ l(1) - f \circ l(0) = D(f \circ l)(c) = Df(x_0 + cv) \cdot v.$$

Q.E.D

Using Mean value theorem II, one can proof a more general mean value theorem.

Theorem 2.12 (Mean value theorem III)**.** *Let X be a subspace in $\mathbb{R}^m$, $x_0 \in X$ and $B_r(x_0) = \{x \in X \mid \|x - x_0\| < r\}$. Assume that $f : X \to \mathbb{R}$ is a function differentiable at each point $x \in B_r(x_0)$. Then for any point $x_0 + v$ with $\|v\| < r$, there are points $y_1, \ldots, y_m$ such that*

$$f(x + v) - f(x) = \sum_{i=1}^{m} v_i D_i f(z_i)$$

and $\|x_0 - y_i\| < \|v\|$ for $i = 1, \ldots, m$, where v_i is the i-th entry of v. Q.E.D

Apply chain rule, we can get a formula of derivative of inverse function.

Theorem 2.13. *Let X be an open subspace in $\mathbb{R}^n$, $f : X \to \mathbb{R}^n$ be a function, U be a open neighborhood of $y_0 = f(x_0)$ in $\mathbb{R}^n$. Assume that g is a function $U \to \mathbb{R}^n$ such that $g(y_0) = x_0$ and $g \circ f(x) = x$ for any x in a neighborhood of x_0 in X. If f is differentiable at x_0 and g is differentiable at y_0, then*

$$Dg(y_0) = Df(x_0)^{-1}.$$

Proof. Apply chain rule to $g \circ f(x) = i(x)$, where i is the identity function, we have $Dg(y_0) \cdot Df(x_0) = I_n$, where I_n is the $n \times n$ identity matrix. Therefore, $Dg(y_0) = Df(x_0)^{-1}$. Q.E.D

Apply Theorem 2.9 and Theorem 2.15, we get a weaker version of chain rule.

Theorem 2.14 (Chain rule (weak))**.** *Let $g_1, \ldots, g_n : \mathbb{R}^m \to \mathbb{R}$ be functions continuously differentiable at x_0 and $f : \mathbb{R}^n \to \mathbb{R}$ be a function differentiable at the point $(g_1(x_0), \ldots, g_n(x_0))$. Define $g : \mathbb{R}^m \to \mathbb{R}^n$ to be $g(x) = (g_1(x), \ldots, g_n(x))$. Then*

$$D(f \circ g) = \sum_{j=1}^{n} D_j f(g(x_0)) D_i g_j(x_0).$$

Proof. By Corollary 2.16, we know that g is differentiable at x_0. By Theorem 2.9, we concluded that $D(f \circ g) = \sum_{j=1}^{n} D_j f(g(x_0)) D_i g_j(x_0)$. Q.E.D

2.3 Continuously differentiable functions

Using the mean value theorem III, we can show a sufficient condition for differentiability.

Theorem 2.15. *Let X be a subspace in $\mathbb{R}^m$, $f : X \to \mathbb{R}$ be a function and x_0 be a point with a small neighborhood in X. Then f is differentiable at x_0 if the partial derivatives $D_i f(x)$ exists at each point $x \in X$ and continuous at x_0.*

Proof. For any vector $h = (h_1, \ldots, h_m) \in \mathbb{R}^m$, we set $x_k = x_0 + \sum_{i=1}^{k} h_i e_i$. Then

$$f(x_0 + h) - f(x_0) - \sum_{i=1}^{m} (h_i D_i f(x_0)) = \sum_{k=1}^{m} (f(x_k) - f(x_{k-1}) - h_k D_k f(x_0)).$$

It suffices to show that

$$\lim_{h \to 0} \frac{f(x_k) - f(x_{k-1}) - h_k D_k f(x_0)}{h_k} = 0.$$

It will imply that $Df(x_0) = (D_1 f(x_0), \ldots, D_m f(x_0))$.

Let $\phi_{k-1}(t) = f(x_{k-1} + te_k)$. Then $\phi'_{k-1}(t) = D_k f(x_{k-1} + te_k)$. By mean value theorem for functions of single variables, there exists a point $p_{k-1} = x_{k-1} + c_k e_k$ for $c_k \in [0, h_k]$ such that $f(x_k) - f(x_{k-1}) = h_k D_k f(p_{k-1})$.

Since $D_k f(x)$ is continuous at x_0. Then

$$\lim_{h \to 0} (D_k f(p_{k-1}) - D_k f(x_0)) = 0$$

for $k = 1, \ldots, m$. Therefore,

$$\lim_{h \to 0} \frac{f(x_k) - f(x_{k-1}) - h_k D_i f(x_0)}{h_k} = \lim_{h \to 0} (D_k f(x_0) - D_k(p_{k-1})) = 0.$$

Q.E.D

Corollary 2.16. *Let X be a subspace in $\mathbb{R}^m$, $f : X \to \mathbb{R}^n$ be a function and x_0 be a point with a small neighborhood in X. Then f is differentiable at x_0 if the Jacobian matrix $\left(D_j f_i\right)$ exists and continuous in a neighborhood of x_0.* *Q.E.D*

Example 2.17. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be a function defined by $f(x_1, x_2) = x_1 x_2$. The partial derivatives are $D_1 f = x_2$ and $D_2 f = x_1$. Hence f is differentiable.

A function whose partial derivatives exist and continuous is said to be *continuously differentiable* or *of class C^1*. Given a C^1 function f, we can consider the partial derivatives $D_j D_i f$ of partial derivatives $D_i f$ can called $D_j D_i f$ the *second-order partial derivatives*. Iteratively, one can define *order r partial derivatives* $(D_{i_r} \cdots D_{i_1}) f$. A function is *of class C^r* if the partial derivatives of order less than or equal to r exists and continuous. We say a function is of class C^∞ (or smooth) if the partial derivatives of all orders are continuous.

Remark 2.18. Although it is quite rare, but there are differentiable functions which are not continuously differentiable.

Example 2.19. Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f(x) = \begin{cases} x^2 \sin \frac{1}{x} & x \neq 0, \\ 0 & x = 0. \end{cases}$$

Then the derivative

$$f'(x) = \begin{cases} 2x \sin \frac{1}{x} - \cos \frac{1}{x} & x \neq 0 \\ 0 & x = 0. \end{cases}$$

However, $f'(x)$ is not continuous.

In geometry, there is another way to distinguish different geometric structures by studying functions of different classes.

Theorem 2.20. *Let X be an open subspace in $\mathbb{R}^m$ and $f : \mathbb{R}^m \to \mathbb{R}$ be a function of class C^2. Then*

$$D_i D_j f(x) = D_j D_i f(x).$$

Proof. For any h, k in $\mathbb{R}$, we define

$$\delta(h, k) = f(x + he_i + ke_j) - f(x + he_i) - f(x + ke_j) + f(x).$$

Let $\varphi(s) = f(x+s+ke_j) - f(x+s)$ and $\psi(t) = f(x+he_i+t) - f(x+t)$. By mean value theorem II, there exists a_1, a_2, b_1 and b_2 in [0, 1] such that

$$\delta(h,k) = \varphi(he_i) - \varphi(0) = D_i\varphi(a_1he_i)h = D_jD_if(x + a_1he_i + a_2ke_j)hk$$

and

$$\delta(h,k) = \psi(ke_j) - \varphi(0) = D_j\psi(b_1ke_j)k = D_iD_jf(x + b_2he_i + b_1ke_j)hk.$$

Then

$$D_i\varphi(a_1he_i)h = D_iD_jf(x+a_1he_i+a_2ke_j) = \frac{\delta(h,k)}{hk} = D_jD_if(x+b_2he_i+b_1ke_j).$$

Since D_iD_jf and D_jD_if are continuous, letting h and k goes to zero, we get $D_iD_jf(x) = D_jD_if(x)$. Q.E.D

It should be noticed that without the assumption that f is C^2. The partial derivatives may not commute.

Example 2.21. Let $f : \mathbb{R}^2 \to R$ be the function define by

$$f(x,y) = \begin{cases} \frac{xy(x^2-y^2)}{x^2+y^2} & (x,y) \neq (0,0) \\ 0 & \text{otherwise.} \end{cases}$$

Then $\frac{\partial^2 f}{\partial x \partial y}(0,0) \neq \frac{\partial^2 f}{\partial y \partial x}(0,0)$.

2.4 Inverse function theorem

The formula of differentiable inverse function f^{-1} of a differentiable function f suggests that if the Jacobian matrix $J(f)$ must be non-singular. Surprisingly, the converse is also true.

The key is to show that the function is one-to-one over a small open neighborhood of a point where the Jacobian matrix is non-singular.

We start with a fixed point lemma.

Lemma 2.22 (Fixed point lemma)**.** *Let X be a complete metric space and $f : X \to X$ be a continuous map. Assume that there is a number $c < 1$ such that $d(f(x), f(y)) < cd(x, y)$ for all $x, y \in X$. Then there exists a unique $z \in X$ such that $f(z) = z$.*

Proof. Pick an arbitrary point $x_0 \in X$. Define an infinite sequence $\{x_n\}$ recursively by $x_{n+1} = f(x_n)$ for $n = 0, 1, 2, \ldots$. Then $d(x_{n+1}, x_n) < c^n d(x_1, x_0)$ for $n = 1, 2, \ldots$. By triangle inequality, for any $n < m$, we have

$$d(x_n, x_m) < \sum_{i=n+1}^{m} d(x_i, x_{i-1}) \leq \sum_{i=n}^{m-1} c^i d(x_1, x_0) \leq \frac{c^n}{1-c} d(x_1, x_0).$$

By completeness, the sequence $\{x_n\}$ converges to a point x. By continuity of f, we see that

$$f(x) = \lim_{n\to\infty} f(x_n) = \lim_{n\to\infty} x_{n+1} = x.$$

Q.E.D

Theorem 2.23 (Inverse function theorem)**.** *Let X be an open subspace in $\mathbb{R}^n$, $f : X \to \mathbb{R}^n$ be a continuous differential function. Assume that $Df(x_0)$ is nonsingular. Then there exists small open neighborhoods U of x_0 and W of $f(x_0)$ and a continuous differential function g such that $g : W \to U$ is the inverse of f.*

Let's see what does the theorem claim and discuss possible ideas of proof.

Properly choose small open neighborhood such that the image is well bounded. Choose a small open neighborhood of $f(x_0)$, need to show that any point in the neighborhood has a preimage. Show that the preimage consists of a unique points. Show that the function is continuous. Show that the function is differentiable.

Need to show that a small open neighborhood W of $f(x_0)$ is in the image of $f(U)$. For any y in W and any $f(x)$ in $f(U)$. The function $y - f(x) = 0$ should have a solution. First, we may assume that $\|y - f(x_0)\| < r$. Consider $y - f(x)$ as a function of x. Method one: Show that $\|y - f(x)\|^2$ has a max or min, so that take derivative to reduce to a linear problem. Method two: take linear approximation of $f(x)$: $f(x) = f(x_0) + J(x - x_0) + \eta$. By definition $\lim \frac{\|\eta\|}{\|x-x_0\|} = 0$. So $\|y - f(x) - (y - f(x_0) - J(x - x_0))\| = \|\eta\| \|x - x_0\|$. By Newton's iteration method, we know that the sequence $\{x_{n+1} = x_n + J^{-1}(y - f(x_n))\}$

converges to a point x which would be the preimage of y. Keep iterating, the limit will be a fixed point of the map $\phi : W \to W$ defined by $\phi(x) = x + J^{-1}(y - f(x))$. The following 2-dimensional picture shows the idea of the method.

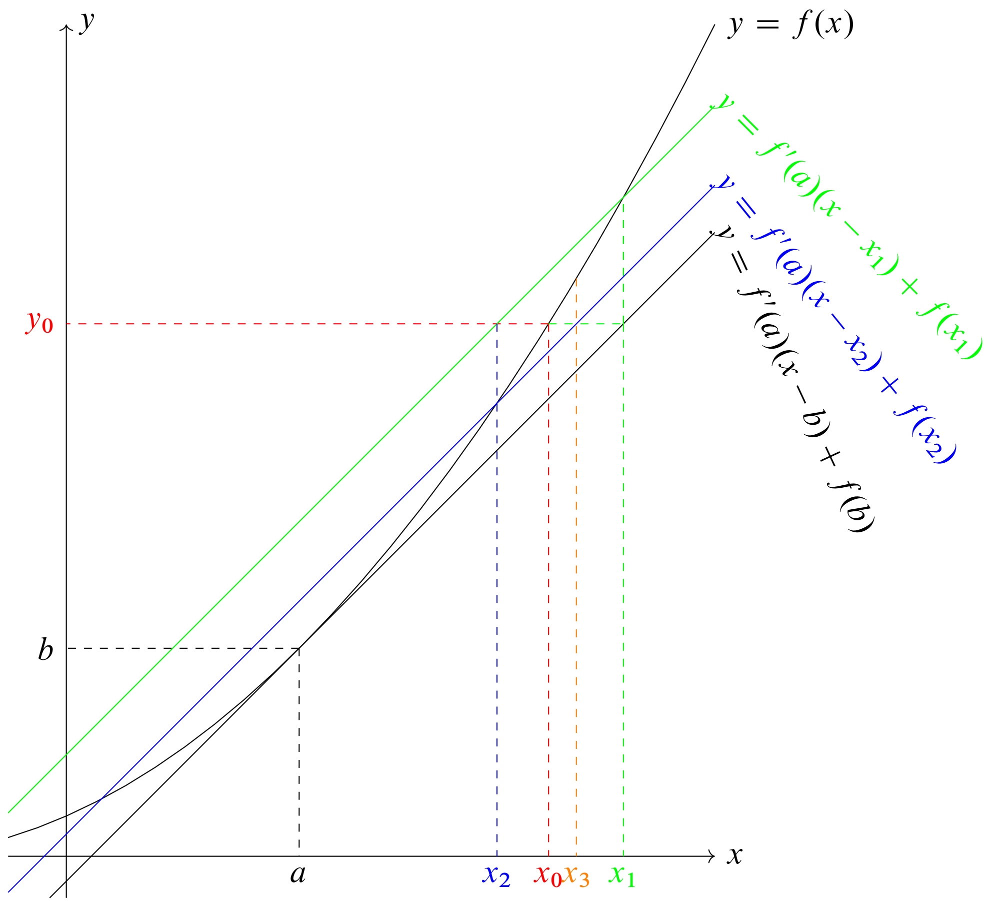


Figure 1: Iteration method for solving $y_0 = f(x)$.

Proof. Step 1: f is locally injective.

Let $J = Df(x_0)$. Choose λ so that $2\lambda\|J^{-1}\| = 1$. Since Df is continuous at x_0, there is a open ball $U \subset X$ centered at x_0 such that $\|Df(x) - J\| < \lambda$ for any $x \in U$. For each $y \in \mathbb{R}^n$, we define a function $\phi : U \to \mathbb{R}^n$ by $\phi(z) = z + J^{-1}(y - f(z))$ for any $z \in X$. We will show that ϕ has a unique fixed point x in a neighborhood of x_0 which implies that f is a locally one-to-one function.

Note that $D\phi(x) = I - J^{-1}Df(x) = J^{-1}(J - Df(x))$. Then over U, we have $\|D\phi(x)\| < \frac{1}{2}$. Therefore,

$$\|\phi(x_1) - \phi(x_2)\| \leq \frac{1}{2}\|x_1 - x_2\| \tag{4}$$

for any $x_1, x_2 \in U$. This implies that ϕ can have at most one fixed point in U. Therefore, there is at most one x in U such that $f(x) = y$. This proves that f is injective.

Step 2: f is locally surjective

Now let $V = f(U)$. We will show that V is an open neighborhood of $f(x_0)$, i.e. for $q \in V$, there is a sufficiently small open ball $B(q, \varepsilon)$ in $\mathbb{R}^n$ which is contained in V.

Let $q = f(p) \in V$ and r be a number such that $\overline{B(p,r)} \subset U$. We claim that $B(q, \lambda r) \subset V$. Note this claim also proves that V is open.

Fix $y \in B(q, \lambda r)$, i.e. $\|y - q\| < \lambda r$. We want to show that $\phi(z) = z + J^{-1}(y - f(z))$ has a fixed point $x \in \overline{B(p,r)}$. It will then follows that $y = f(x) \in V$ and $B(q, \lambda r) \subset V$.

Note that we have

$$\|\phi(p) - p\| = \|J^{-1}(y - q)\| < \|J^{-1}\|\lambda r = \frac{r}{2}.$$

For any $x \in \overline{B(p,r)} \subset U$, by equation (4), we know that

$$\|\phi(x) - \phi(p)\| \leq \frac{1}{2}\|x - p\|.$$

Therefore,

$$\|\phi(x) - p\| \leq \|\phi(x) - \phi(p)\| + \|\phi(p) - p\| < r$$

which implies that $\phi(x) \in B(p, r)$.

Since $\overline{B(p,r)} \subset U$, then the equation (4) holds for any $x_1, x_2 \in \overline{B(p,r)}$. Since $\overline{B(p,r)}$ is closed and bounded, then it is complete. Apply the fixed point theorem, we conclude that ϕ has a unique fixed point x in $\overline{B(p,r)}$. Therefore $y = f(x) \in V$.

Step 2: g is continuous. Since f is one-to-one from $U \to V$, then it has a inverse function $g : V \to U$. For each open subset W in U, we have $g^{-1}(W) = f(W)$. By the argument in step 2, since W is open, hence $f(W) = g^{-1}(W)$ is open.

Step3: The inverse function g is differentiable. For any $a \in U$, we let $b = f(a)$, $J = Df(x)$. Since f is differentiable at x, we may write

$$f(x) = b + J(x-a) + \eta(x, a)$$

with

$$\lim_{x \to a} \frac{\|\eta\|}{\|x-a\|} = 0.$$

Let $x = g(y)$, we have

$$(g(y) - g(b)) = J^{-1}(y-b) - J^{-1}\eta(g(y), g(b)).$$

Since $\|J^{-1}\eta(g(y), g(a))\| \leq \|J^{-1}\|_{op}\|\eta(g(y), g(a))\|$, it suffices to show that

$$\lim_{y \to b} \frac{\|\eta(g(y), g(b))\|}{\|y-b\|} = 0.$$

By our assumption, we know that

$$\|x-a\| - \|J^{-1}(y-b)\| \leq \|(x-a) - J^{-1}(y-b)\| = \|\phi(x) - \phi(a)\| \leq \frac{1}{2}\|x-a\|.$$

Then

$$\frac{\|x-a\|}{\|y-b\|} \leq 2\|J^{-1}\|_{op} \leq \lambda^{-1}$$

Notice that g is continuous. Then

$$\frac{\|\eta(g(y), g(b))\|}{\|y-b\|} = \frac{\|\eta(x, a)\|}{\|x-a\|}\frac{\|g(y) - g(b)\|}{\|y-b\|}$$

goes to 0 as y approaches to b. Therefore, g is differentiable.

Step4: g is continuously differentiable Let $I : GL(n) \to GL(n)$ be the function send a invertible matrix to its inverse. Then I is continuous. Moreover, I is of class C^{∞}. Note that

$$Dg(y) = Df(g(y))^{-1}$$

equals the composite of three functions

$$V \xrightarrow{g} U \xrightarrow{Df} GL(n) \xrightarrow{I} GL(n).$$

Hence, Dg is continuous. Q.E.D

2.5 Implicit function theorem

Theorem 2.24 (Implicit function theorem)**.** *Let* $f = (f^1, \dots, f^n) : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ *be a function continuously differentiable in an open neighborhood of a point* $(x_0, y_0) \in \mathbb{R}^m \times \mathbb{R}^n$ *and* $f(x_0, y_0) = 0$*. Let*

$$M = \left(\frac{\partial f^j}{\partial y_i}(x_0, y_0)\right), \quad 1 \le i, j \le n$$

be the $n \times n$ *matrix of partial derivatives, where* $\frac{\partial f^j}{\partial y_i} = D_{m+i} f^j$*. Assume that* $\det M \neq 0$*. Then there is an open neighborhood* $A \subset \mathbb{R}^m$ *of* a *and an open neighborhood* $B \subset \mathbb{R}^n$ *of* b *with the following property: for each* $x \in A$*, there is a unique* $g(x) \in B$ *such that* $f(x, g(x)) = 0$ *and* g *is continuously differentiable.*

We would like to use inverse function theorem. For this purpose, we need to create a new function between spaces of "same dimension". A choice woulde be $F(x, y) = (x, f(x, y))$.

Proof. Let $F : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^m \times \mathbb{R}^n$ be the function defined by $F(x, y) = (x, f(x, y))$. Then

$$DF(x_0, y_0) = \begin{pmatrix} I_m & 0 \\ \frac{\partial f}{\partial x}(x_0, y_0) & M \end{pmatrix}$$

where $\frac{\partial f}{\partial x} = \left(D_i f^j\right)$ for $1 \le i \le m$ and $1 \le j \le n$. Therefore, $\det DF(x_0, y_0) \neq 0$. By inverse function theorem, there exist open neighborhoods U of (x_0, y_0), V of $(x_0, f(x_0, y_0))$ and a continuously differentiable function $G : V \to U$ such that G is the inverse function of F. Note that $G(F(x, y)) = (x, y)$. Then G must be in the form $G(x, y) = (x, h(x, y))$ for some continuously differentiable function $h : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$. Denote by $\pi : \mathbb{R}^m \times \mathbb{R}^n \to \mathbb{R}^n$ be the projection, i.e. $\pi(x, y) = y$. Then $f(x, y) = \pi \circ F(x, y)$. Therefore,

$$f(x, h(x, y)) = \pi \circ F(x, h(x, y)) = \pi \circ F \circ G(x, y) = y.$$

Let $g(x) = h(x, 0)$. Then it is continuously differentiable and $f(x, g(x)) = 0$.

The uniqueness of g is easy to check. Q.E.D

Example 2.25. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function defined by $f(x, y) = x^2 + y^2 - 1$. If $x^2 + y^2 - 1 = 0$, then we can solve for y to get $y = g(x) = \pm\sqrt{1 - x^2}$. Differentiating of $f(x, g(x)) = 0$, we get $2x + 2g(x)g'(x) = 0$, equivalently, $g'(x) = -\frac{x}{g(x)} = (\pm\sqrt{1 - x^2})'$.

The implicit function theorem has many important generalizations and applications. The following generalization is very useful in differential geometry.

Theorem 2.26 (Rank theorem)**.** *Let $f : \mathbb{R}^m \to \mathbb{R}^p$ be a function continuously differentiable in an open neighborhood of $a \in \mathbb{R}^m$, where $p \leq m$. If $f(a) = 0$ and the $p \times m$ matrix $Df(a)$ has rank p, then there is an open neighborhood U of a and a differentiable function $g : U \to \mathbb{R}^m$ such that $f \circ g(x^1, \ldots, x^m) = (x^{m-p+1}, \ldots, x^m)$.*

Proof. By linear transformation, we may assume that the last p columns of $Df(a)$ is an invertible $p \times p$ matrix M. Write $\mathbb{R}^m = \mathbb{R}^{m-p} \times \mathbb{R}^p$ and $a = (b, c) \in \mathbb{R}^{m-p} \times \mathbb{R}^p$. Since $f(b, c) = 0$ and $\det M \neq 0$. Same as in the proof of implicit function theorem, we know, by inverse function theorem, that there is an open neighborhood $U \subset \mathbb{R}^{m-p} \times \mathbb{R}^p$ of a and a continuously differentiable function $h : U \to \mathbb{R}^p$ such that

$$f((x^1, \ldots, x^{m-p}), h((x^1, \ldots, x^m))) = (x^{m-p+1}, \ldots, x^m).$$

Define a function

$$g(x^1, \ldots, x^m) = ((x^1, \ldots, x^{m-p}), h(x^1, \ldots, x^m)).$$

Then $f \circ g(x^1, \ldots, x^m) = (x^{m-p+1}, \ldots, x^m)$. Q.E.D

2.6 The method of Lagrange multiplier

The Lagrange multiplier method is a strategy for finding extremum constrained to equations. The key idea behind this method is that the implicit function theorem can convert constrained extremum to unconstrained extremum locally.

Theorem 2.27 (Lagrange multiplier)**.** *Let $f, g : X \subset \mathbb{R}^m \to \mathbb{R}$ be continuously differentiable functions. Assume that f has an extremum at $x_0 \in X$ such that $g(x_0) = 0$ and $Dg(x_0) \neq 0$. Then there is a number $\lambda \in \mathbb{R}$ such that $Df(x_0) = \lambda Dg(x_0)$.*

Proof. Write $x_0 = (a_1, \dots, a_m)$, $z_0 = (a_1, \dots, a_{m-1})$, $x = (z_1, \dots, z_m)$ and $z = (z_1, \dots, z_{m-1})$. Since $Dg(x_0) \neq 0$, we may assume that $D_m g(x_0) \neq 0$. Then by implicit function, there exist an open neighborhood $U \subset \mathbb{R}^{m-1}$ of $(a_1, \dots, a_{m-1})$, an open neighborhood $V \in \mathbb{R}$ of a_m, and a unique continuously differentiable function $h : U \to V$ such that

$$g((z^1, \dots, z^{m-1}), h(z^1, \dots, z^{m-1}) = 0.$$

Then f has an extremum at x_0 subject to $g(x_0) = 0$ if and only if $f(z, h(z))$ has an extremum at z_0 which implies that $Df(z_0, h(z_0)) = 0$. Since $g(z, h(z)) = 0$ for $z \in U$, then $Dg(z) = 0$ for $z \in U$, in particular, $Dg(z_0) = 0$. By chain rule, we have

$$0 = D_i f(z_0, h(z_0)) = D_i f(x_0) + D_m f(x_0) D_i h(z_0), \quad i = 1, \dots, m-1$$

and

$$0 = D_i g(z_0, h(z_0)) = D_i g(x_0) + D_m g(x_0) D_i h(z_0), \quad i = 1, \dots, m-1.$$

Denote

$$M = \begin{pmatrix} 1 & 0 & \cdots & 0 & D_1 h(z_0) \\ 0 & 1 & \cdots & 0 & D_2 h(z_0) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & D_{m-1} h(z_0) \end{pmatrix}_{(m-1)\times m}.$$

Writing into matrix equation form, we see that $Df(x_0)$ and $Dg(x_0)$ are both solutions of the following equation

$$MX = 0$$

Note that the rank of the matrix M is $m-1$. Therefore $Df(x_0)$ and $Dg(x_0)$ are linearly equivalent, i.e. there exists a $\lambda \in \mathbb{R}$ such that $Df(x_0) = \lambda Dg(x_0)$. Q.E.D

We see that those theorems, inverse function theorem, implicit function theorem and rank theorem are all locally stated, i.e. works in an open neighborhood. However, under additional condition, the statements can be extended globally.

Example 2.28. Let $f : \mathbb{R} \to \mathbb{R}$ be a continuously differential function such that $f'(x) \neq 0$ for any $x \in \mathbb{R}$ and $\|f(x) - f(y)\| > c\|x - y\|$ for some $c > 0$ and any $x, y \in \mathbb{R}$. Then there is a function $g : \mathbb{R} \to \mathbb{R}$ such that $g \circ f(x) = x$ for any $x \in \mathbb{R}$. In fact, the condition that $\|f(x) - f(y)\| > c\|x - y\|$ for some $c > 0$ and any $x, y \in \mathbb{R}$ implies the image of f is closed. Since $\mathbb{R}$ is connected and the image of f is open (by inverse function theorem) and closed, then $f : \mathbb{R} \to \mathbb{R}$ is a homeomorphism, i.e. there exists $g : \mathbb{R} \to \mathbb{R}$ such that $g \circ f(x) = x$ for any $x \in R$.

Another remark is that the requirements are in general sufficient but not necessary for a single conclusion.

Example 2.29. Let $f : \mathbb{R}^2 \to R$ be the function $f(x, y) = x^2 - y^3$. Although $\frac{\partial f}{\partial y}(0) = 0$, the equation $f(x, y) = 0$ still has a unique solution $g(x) = x^{\frac{2}{3}}$ for y, i.e. $y = g(x)$. However, the function $g'(x)$ is not continuous at $x = 0$. In fact, the equation $f(x, y) = 0$ has a cusp singularity at 0.

Example 2.30. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function $f(x, y) = y^2 - x^4$. Then $\frac{\partial f}{\partial y}(0) = 0$. However, the equation $f(x, y) = 0$ has two solutions $y = g_-(x) = -x^2$ and $y = g_+(x) = x^2$ for y near 0. Moreover, the solutions are continuously differentiable. In fact, the $y^2 - x^4 = 0$ has a tacnode singularity at 0.

3 Integrations

In this section, we will generalize integrals of single-variable functions to multivariable functions and study its properties and applications.

3.1 Integrable functions

As a generalization of intervals in $\mathbb{R}$, we define an *n-cell* to be the set

$$Q = [a_1, b_1] \times \cdots \times [a_n, b_n] = \{(x_1, \ldots, x_n) \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i\}$$

in $\mathbb{R}^n$. The product $\mathrm{vol}(Q) = \prod_{i=1}^{n}(b_i - a_i)$ is called the *volume* of Q. The interval $[a_i, b_i]$ is called the *i-th component interval* of Q. The maximum of the numbers $b_i - a_i$ is called the *width* of Q. The width of a 1-cell $[a, b]$ is also called the length of $[a, b]$.

Definition 3.1. Given an interval $[a, b] \subset \mathbb{R}$, a *partition of an interval* $[a, b]$ is a finite collection $P = \{t_0, t_1 \ldots, t_k\}$ of points in $[a, b]$ such that $a \leq t_0 < t_1 < \cdots < t_k = b$. A *partition of an n-cell* Q is a collection $P = (P_1, \ldots, P_n)$, where P_i is a partition of the i-th component interval.

Given a partition P of Q, if P_i divides the interval $[a_i, b_i]$ into N_i subintervals $[t_{i,j-1}, t_{i,j}]$ $0 \leq j \leq N_i$, then P divides Q into $N_1 N_2 \cdots N_n$ n-cells. These cubes are called the sub-cells of Q associated to the partition P or simple the sub-cells of P.

We denote by $\mathscr{P}_Q$ the set of all partition of the n-cell Q and $\mathscr{S}_P$ the set of subcells of Q associated to the partition P. For each subcell $R \in \mathscr{S}_P$, we denote by w_R the width of R. The maximum $\mu_P = \max\{w_R \mid R \in \mathscr{S}_P\}$ is called the *mesh* of the partition P.

Definition 3.2 (Darboux sums)**.** Let Q be a n-cell, P be a partition of Q and $f : Q \to \mathbb{R}$ be a bounded function. For a sub-cell R of P, we set

$$m(f, R) = \inf\{f(x) \mid x \in R\} \quad \text{and} \quad M(f, R) = \sup\{f(x) \mid x \in R\}.$$

The *lower Darboux sum* and *upper Darboux sum* are defined respectively as

$$L(f, P) = \sum_{R \in \mathscr{S}_P} m(f, R)\mathrm{vol}(R) \quad \text{and} \quad U(f, P) = \sum_{R \in \mathscr{S}_P} M(f, R)\mathrm{vol}(R).$$

Let $P = (P_1, \ldots, P_n)$ be a partition of an n-cell Q. A partition P' of Q obtained by adjoining additional points to some or all of the partitions $P_1, \ldots, P_n$ is called a *refinement* of P. Let P and P'' be two partitions of the n-cell Q. The partition $P' = (P_1 \cup P"_1, \ldots, P_n \cup P"_n)$ is called a common refinement of the partitions P and $P"$ of Q.

Comparing Darboux sums over a partition P and a refinement P', we have the following elementary results.

Lemma 3.3. *Let P be a partition of an n-cell Q and $f : Q \to \mathbb{R}$ be a bounded function. If P' is a refinement of P, then*

$$L(f, P) \leq L(f, P') \quad \text{and} \quad U(f, P') \leq U(f, P).$$

Corollary 3.4. *Let P and P' be two partitions of an n-cell Q and $f : Q \to \mathbb{R}$ be a bounded function. Then $L(f, P') \leq U(f, P)$.*

Denote by

$$\underline{\int_Q} f = \sup_{P \in \mathscr{P}_Q} \{L(f, P)\} \quad \text{and} \quad \overline{\int_Q} f = \inf_{P \in \mathscr{P}_Q} \{U(f, P)\}$$

By Corollary 3.4, we see the

$$\underline{\int_Q} f \leq \overline{\int_Q} f.$$

Definition 3.5 (Riemman-Darboux integral)**.** Let Q be an n-cell. A function $f : Q \to \mathbb{R}$ is called *integrable* on Q if f is bounded and

$$\underline{\int_Q} f = \overline{\int_Q} f.$$

We denote by

$$\int_Q f = \underline{\int_Q} f = \overline{\int_Q} f$$

and call $\int_Q f$ the *integral* of f over Q.

Remark 3.6. In general, the integral defined in Definition 3.5 is called Darboux integral. However, it is also commonly called Riemann integral which is defined as the limit

$$\lim_{\mu_P \to 0} \sum_{R_i \in \mathscr{S}_P} f(t_i)\mathrm{vol}(R_i),$$

where $t_i \in R_i$. The reason is that Darboux integral and Riemann integral are equivalent under our setting that $\mathrm{vol}(R) = \prod_{i=1}^{n}(b_i - a_1)$, where $R = (a_1, b_1) \times \cdots \times (a_n, b_n)$.

One should be aware that there are also so called Riemann-Stieltjes integral and Darboux-Stieltjes integral which are defined similar but replace $\mathrm{vol}(R)$ by "parametrized volume". Those two generalized definitions may not be equivalent in general.

Riemann-Darboux integral is successful in the sense that it gives expected answer or useful results for many problems. However, it has its limitation such as it does not commute well with limits.

Lebesgue integral extends the integral to a much larger class of functions and provides the conceptual insight and full power to handle problems that Riemann-Darboux integral cannot handle. Intuitively, Lebesgue's idea is to partition of the "$f(x)$-axis" instead of x-axis. The philosophy is that one should be able to approximate the values of a function by simple function whose integral can be written in terms of measure.

By Lemma 3.3, we know that refining a partition will increase the lower Darboux sum but decrease the upper Darboux sum. Therefore, if f is integrable over a n-cell Q, for any $\varepsilon > 0$, we should be able find a partitions P such that $U(f, P) - L(f, P) < \varepsilon$. Indeed, this observation leads to the following theorem.

Theorem 3.7 (Riemann theorem)**.** *Let Q be an n-cell. A bounded function $f : Q \to \mathbb{R}$ is integrable if and only if for any $\varepsilon > 0$, there exists a partition P such that $U(f, P) - L(f, P) < \varepsilon$.*

Proof. If the condition holds, then it is clear that

$$\sup_{P \in \mathscr{P}_Q} L(f, P) = \inf_{P' \in \mathscr{P}_Q} U(f, P').$$

Conversely, if f is integrable, then for any $\varepsilon > 0$, there exists partitions P' and P" such that

$U(f, P') - L(f, P") < \varepsilon$. Let P be the common refinement of P' and $P"$. By Corollary 3.4, we see that $U(f, P) - L(f, P) < \varepsilon$. Q.E.D

Remark 3.8. For a integrable function f over a n-cell Q, in fact, one can show that for any $\varepsilon > 0$, there exists a $\delta > 0$ such that $U(f, P) - L(f, P) < \varepsilon$ for any portion $P \in \mathscr{P}_Q$ with mesh $\mu_P < \delta$. Moreover, this conclusion implies the following equivalent definition of integrable function: a bounded function $f : Q \to \mathbb{R}$ is integrable with integral $\int_Q f = A$ if and only if for any $\varepsilon > 0$, there is a $\delta > 0$ such that for any partition $P \in \mathscr{P}_Q$, if the mesh μ_P is less than δ, then

$$|\sum_{R \in \mathscr{S}_P} f(x_R)\text{vol}(R) - A| < \varepsilon,$$

where x_R is any point in the subcell R.

When is a function integrable? Let's see some examples first.

Example 3.9. Let $f : Q = [0, 1]^n \to \mathbb{R}$ be a constant function, i.e. $f(x) = c$, where c is a number. Then for any partition P, we have $m(f, R) = M(f, R)$ for any $R \in \mathscr{S}_P$. Hence $\int_Q f = c$.

More generally, a continuous function is integrable.

Example 3.10. Let $f : \mathbb{R}^m \to \mathbb{R}$ be a continuous function and Q be a n-cell. Then f is integrable over Q. Since Q is compact, then f is bounded. Moreover, it is uniformly continuous, i.e. for any $\varepsilon > 0$, there is a $\delta > 0$ such that $|f(x) - f(y)| < \frac{\varepsilon}{\text{vol}(Q)}$ whenever $x, y \in R$, where R is an subcell with width $w_R < \delta$. Then for any partition P with mesh $\mu_P < \delta$, we have

$$U(f, P) - L(f, P) < \frac{\varepsilon}{\text{vol}(Q)}\text{vol}(Q) < \varepsilon.$$

Conversely, if the function is not continuous, then it may not be integrable.

Example 3.11 (A non-Riemann-Darboux integrable function). Let $f : R = [0, 1]^2 \to \mathbb{R}$ be defined by

$$f(x, y) = \begin{cases} 0 & x \text{ is a rational number,} \\ 1 & x \text{ is an irrational number.} \end{cases}$$

For any partition P, then every subcell R will contain points (x, y) with x rational and points (x, y) with x irrational. Hence, $m(f, R) = 0$ and $M(f, R) = 1$. Then $L(f, P) = 0$

and $U(f,P)=1$. Therefore, the function f is not integrable.

However, if the set discontinuities is “very small", then the function is still integrable.

Example 3.12. Let $f:[0,1]^2\to\mathbb{R}$ be a function defined by

$$f(x,y)=\begin{cases}0 & x\neq y,\\ 1 & x=y.\end{cases}$$

Then f is discontinue at any point in the diagonal. However, f is integrable over $[0,1]^2$. In fact, for any $n\in\mathbb{N}$, the diagonal can be covered by squares Δ_i centered at $(\frac{i}{2n},\frac{i}{2n})$ with width $\frac{1}{n}$ for $i=1,2n-1$. Take a partition P contains Δ. Then $U(f,P)-L(f,P)=\frac{2n-1}{n^2}<\frac{2}{n}$.

In Example 3.12, the reason that f is integrable is that the diagonal, being the set of discontinuities, can be covered by squares that the sum of their areas is “zero".

In fact, we will see that removing a set of “measure zero" won’t change the integral.

Definition 3.13. A subset $A\subset\mathbb{R}^n$ is said to have *measure zero* in $\mathbb{R}^n$ if for any $\varepsilon>0$, there is a covering of $Q_1,Q_2,\ldots$ of A by countably many n-cells such that

$$\sum_{i=1}^{\infty}\text{vol}(Q_i)<\varepsilon.$$

Remark 3.14. Note that in Example 3.12, the diagonal has measure zero in R^2. But considering as a line segment, it is not measure zero. In fact, “measure zero" is a relative concept.

Remark 3.15. In the definition, “measure zero" is indeed in the sense of “Lebesgue measure".

From Definition 3.13, we can derive the following properties of sets of measure zero. You may give a proof by yourself or refer to Munkres’s book “Analysis on Manifolds".

Theorem 3.16.

(a) Let B be a subset of A. If A has measure zero in $\mathbb{R}^n$, then B has measure zero in $\mathbb{R}^n$.

(b) *Let A be the union of countably many subset $A_1, A_2, \ldots$. If A_i has measure zero in $\mathbb{R}^n$ for each i, then A has measure zero.*

(c) *A subset A has measure zero in $\mathbb{R}^n$ if and only if for every $\varepsilon > 0$, there is a countable covering of A by open n-cells $Q_1^{int}, Q_2^{int}, \ldots$ such that*

$$\sum_{i=1}^{\infty} \text{vol}(Q_i) < \varepsilon.$$

(d) *For any n-cell $Q \subset \mathbb{R}^n$, the boundary ∂Q has measure zero in $\mathbb{R}^n$.*

It is not hard to imagine that a function $f : Q \to \mathbb{R}^n$ if the set of discontinuities has measure zero in $\mathbb{R}^n$. Not surprisingly, the converse also holds.

Theorem 3.17 (Lebesgue's criterion for Riemann integrability). *Let Q be an n-cell in $\mathbb{R}^n$, $f : Q \to \mathbb{R}$ be a bounded function and D be the set of points in Q where f fails to be continuous. Then f is integrable if and only if D has measure zero in $\mathbb{R}^n$.*

Proof. We first proof the sufficiency. Assume that D has measure zero. Since f is bounded, by the assumption, there is a $M > 0$ such that $|f(x)| < M$ for any $x \in Q$. For any $\varepsilon > 0$, we will show that there is a partition P of Q such that $U(f, P) - L(f, P) < \varepsilon$. Since D has measure zero, then for any $\varepsilon' > 0$, there exist countably many open n-cells $Q_1^{int}, Q_2^{int}, \ldots$ such that

$$\sum_{i=1}^{\infty} \text{vol}(Q_i) < \varepsilon'.$$

Since f is continuous away from D, then for any $\varepsilon'' > 0$, at each point $a \in Q \setminus D$, there is a n-cell Q_a such that $|f(x) - f(a)| < \varepsilon$" whenever $x \in Q_a^{int} \cap Q$. Then for any $x, y \in Q_a^{int} \cap Q$, we have $|f(x) - f(y)| < 2\varepsilon''$.

Since Q is closed and bounded, hence compact, then there are finitely many subcells among Q_i^{int} and Q_a^{int} that cover Q. We take end points of those finitely many subcells to form a partition P. Under this partition P, by grouping the subcells in $\mathscr{S}_P$ into two sets $\mathscr{D}$ consisting of subcells $R \in \mathscr{S}_P$ contained in a Q_i^{int}, and $\mathscr{C}$ consisting of subcells $R \in \mathscr{S}_P$ contained in a Q_a^{int}, we see that

$$U(f, P) - L(f, P) < 2M \sum_{R \in \mathscr{D}} \text{vol}(R) + 2\varepsilon'' \sum_{R \in \mathscr{C}} \text{vol}(R) < 2M\varepsilon' + 2\varepsilon''\text{vol}(Q).$$

Then for any $\varepsilon > 0$, let $\varepsilon' = \varepsilon'' < \frac{\varepsilon}{2M+2\text{vol}(Q)}$, we have a portition P such that $U(f, P) - L(f, P) < \varepsilon$.

Now we will prove the necessity. Assume that f is integrable.

We first characterize continuity at a point a using the so-called oscillation of f at a defined as follows. For any $\delta > 0$, denote by $A_\delta = \{x \in Q \mid \|x - a\| < \delta\}$. Let

$$M_\delta(f; a) = \sup_{x \in A_\delta}\{f(x)\} \quad \text{and} \quad m_\delta(f; a) = \inf_{x \in A_\delta}\{f(x)\}.$$

The *oscillation* of a function f at a point a is defined as

$$\nu(f, a) = \inf_{\delta > 0}\{M_\delta(f; a) - m_\delta(f; a)\}.$$

Note that $\nu(f, a)$ is nonnegative and f is continuous at a if and only if $\nu(f, a) = 0$. Equivalent, f is discontinuous at a if and only if $\nu(f, a) \geq \varepsilon$ for some $\varepsilon > 0$.

For any integer $m \in \mathbb{N}$, let

$$D_m = \{a \in Q \mid \nu(f, a) \geq \frac{1}{m}\}.$$

Then D equals the union of D_m. By Theorem 3.16.3.16.(b), it is enough to show that D_m has measure zero for each m.

For any given $\varepsilon > 0$, we can choose a partition P of Q such that $U(f, P) - L(f, P) < \frac{\varepsilon}{2m}$. Let D'_m consist of points in D_m that belong to ∂R for some $R \in \mathscr{S}_P$. Let D''_m be the complement of D'_m in D_m. Then D''_m is covered by finitely many open subcells $R_1^{int}, \ldots, R_k^{int}$. For each R_i^{int} and $a \in R_i^{int} \cap D''_m$, there is a $\delta > 0$ such that $A_\delta \in R_i$ and

$$\frac{1}{m} \leq \nu(f, a) \leq M(f, a) - m(f, a) \leq M(f, R_i) - m(f, R_i).$$

Then we have

$$\frac{1}{m}\sum_{i=1}^{k}\text{vol}(R_i) \leq U(f, P) - L(f, P) < \frac{\varepsilon}{2m}$$

which means that the subcells covering D''_m has total volume $\sum_{i=1}^{k}\text{vol}(R_i) < \frac{\varepsilon}{2}$. Note that the boundary of an n-cell in has measure zero in $\mathbb{R}^n$. Then D'_m can be covered by n-cells whose total volume is less than $\frac{\varepsilon}{2}$. The set D_m has a covering of subcells whose total volume is less than ε. Q.E.D

So far we only discussed integrals of functions of several variables over n-cells. We now generalize the definition to other (bounded) sets.

Definition 3.18. Let D be a bounded subset in $\mathbb{R}^n$ and $f : D \to \mathbb{R}$ be a bounded function. Define $f_D : \mathbb{R}^n \to \mathbb{R}$ by

$$f_D(x) = \begin{cases} 0 & x \notin D, \\ f(x) & x \in D. \end{cases}$$

Let Q be an n-cell containing D. We define the integral of f over D as $\int_D f = \int_Q f_D$ provided that f_D is integrable over Q.

By Theorem 3.17, we know that f_D is integrable over Q if and only if the set of discontinuities of f_D have measure zero. According to the definition, we see that the set of discontinuities f_D consists of discontinuities of f in D and the points in the boundary such that $\lim_{x\to b} f(x) = 0$ fails to hold, where $b \in \partial D$ and $x \in D$. More precisely, as a Corollary of Theorem 3.17. We have the following characterization of integrable functions on a bounded subset.

Theorem 3.19. *Let D be a bounded subset in $\mathbb{R}^n$ and $f : D \to \mathbb{R}$ be a bounded function. Denote by I the points in D where f fails to be continuous and B the set of points in the boundary of D where the condition*

$$\lim_{\substack{x\to b \\ b\in\partial D}} f(x) = 0$$

fails to hold. Then $\int_D f$ exists if and only if I and B have measure zero in $\mathbb{R}^n$. In particular, if f is continuous on D and the boundary ∂D has measure zero, then $\int_D f$ exists.

Example 3.20. Let $D \subset \mathbb{R}^n$ be a bounded subset, we define a function χ_D on $\mathbb{R}^n$, called the characteristic function of D, by

$$\chi_D(x) = \begin{cases} 0 & x \notin D, \\ 1 & x \in D. \end{cases}$$

If the boundary ∂D has measure zero, then χ_D is integrable. In fact, if x is in the interior D^{int} or exterior D^{ext} of D, then χ_D is continuous at x. The only discontinuities of χ_D are points in the boundary ∂D.

For a bounded subset $D \subset \mathbb{R}^n$, the integral $\int_D 1$ exists if and only if the boundary ∂D has measure zero. This integral $\int_D 1$ is called the *volume* of D. In this sense, we call a bounded set D is *rectifiable* if its boundary has measure zero.

Remark 3.21. The terminology rectifiable is different from a more general one called Lebesgue measurable.

Note that similar to integrals of single variable functions, the integrals of multivariable functions also have the properties such as positivity, linearity, monotonicity, and additivity.

Theorem 3.22. *Let D be a bounded subset in $\mathbb{R}^n$ and $f, g : D \to \mathbb{R}$ be bounded functions. Then*

(a) (Positivity) If f is nonnegative and integrable over D, then

$$\int_D f \geq 0.$$

(b) (Linearity) If f and g are integrable over D and $a, b \in \mathbb{R}$ are numbers, then $af + bg$ is integrable over D and

$$\int_D (af + bg) = a \int_D f + b \int_D g.$$

(c) (Monotonicity) Let $E \subset D$ be a subset. If f is non-negative on D and integrable over D and E, then

$$\int_E f \leq \int_D f.$$

(d) (Additivity) If $D = D_1 \cup D_2$ and f is integrable over D_1 and D_2, then f is integrable over D and $D_1 \cap D_2$. Moreover,

$$\int_D f = \int_{D_1} f + \int_{D_2} f - \int_{D_1 \cap D_2} f.$$

Sketch of proof. The conclusions follow almost straight forward from definition of integral by taking care of the two function $m(f, R) = \min\{f(x) \mid x \in R\}$ and $M(f, R) = \max\{f(x) \mid x \in R\}$ in special case first then general cases.

To prove linearity, if $a, b \geq 0$, then we have $am(f, R) + bm(g, R) \leq m(af + bg, R)$ and similarly $M(af + bg, R) \leq aM(f, R) + bM(g, R)$. If $a = -1$, then $m(R, -f) = -M(R, f)$. Then taking common refinements will show that the statement holds.

To prove additivity, first consider the special case that $f \geq 0$. In the special case, the function $f_D = \max\{f_{D_1}, f_{D_2}\}$ and $f_{D_1 \cap D_2} = \min\{f_{D_1}, f_{D_2}\}$. In general, we set $f_+(x) = \max\{f(x), 0\}$ and $f_-(x) = \min\{f(x), 0\}$, then both are integrable functions and $f(x) = f_+(x) - f_-(x)$. The first conclusion will follow from linearity. The formula of integrals follows from the fact that

$$f_D = f_{D_1} + f_{D_2} - f_{D_1 \cap D_2}.$$

Q.E.D

3.2 Fubini's Theorem

We can not only develop definition, concepts and study their properties without seeing their application. The first question after defining integrals (generalizing single-variable integrals) is how to calculate the integrals of an integrable function. In single-variable case, we have the fundamental theorem of calculus, i.e. if $f(x) = F'(x)$, then $\int_a^b f \mathrm{d}x = F(b) - F(a)$, where $F(x)$ is called an antiderivative of f. So the difficulty to integral a function in practice is to find an antiderivative.

In higher dimension, what should we do? If we look at the definition of integral, then the definition suggests that we probably can reduce the problem to lower dimension by considering integral along slices first and then taking sum of the integral with thickening. This idea leads to the iterated integral method, known as Fubini's theorem. What if I want to generalize antiderivatives to multivariable? What should the antiderivative be? Formally, the "antiderivative" for a multivariable function $f(x_1, \ldots x_n)$ should be a "function" F such that $dF = f \mathrm{d}x_1 \mathrm{d}x_2 \cdots \mathrm{d}x_n$. If such an "antiderivative" exists for f, we should expect that $\int_D f \mathrm{d}x_1 \cdots \mathrm{d}x_n = \int_{\partial D} F$. This idea leads to Stoke's theorem.

To realize the second idea, there are more problems to be answered such as what should the "function" F look like?

The first idea leads to the following conclusion.

Theorem 3.23 (Fubini's theorem)**.** *Let $Q = A \times B$, where A is an m-cell in $\mathbb{R}^m$ and B is an n-cell in $\mathbb{R}^n$. For each $x \in A$, let $g_x : B \to R$ be the function defined by $g_x(y) = f(x, y)$ and let*

$$\mathscr{L}(x) = \underline{\int_B} g_x = \underline{\int_B} f(x, y)\mathrm{d}y \quad \textit{and} \quad \mathscr{U}(x) = \overline{\int_B} g_x = \overline{\int_B} f(x, y)\mathrm{d}y.$$

If f is integrable over Q, then $\mathscr{L}(x)$ and $\mathscr{U}(x)$ are integrable over A and

$$\int_Q f = \int_A \mathscr{L} = \int_A \left(\underline{\int_B} f(x, y)\mathrm{d}y \right) \mathrm{d}x,$$

$$\int_Q f = \int_A \mathscr{U} = \int_A \left(\overline{\int_B} f(x, y)\mathrm{d}y \right) \mathrm{d}x.$$

Proof. Let P_A and P_B be partitions of A and B. Then they give a partition P of Q whose $(m+n)$-cells are $R_A \times R_B$, where $R_A \in \mathscr{S}_{P_A}$ and $R_B \in \mathscr{S}_{P_B}$ are subcells. Conversely, for any partition P of Q, we have partitions P_A and P_B by restricting P on A and B respectively.

Considering the lower Darboux sum $L(f, P)$, we have

$$L(f, P) = \sum_{R \in \mathscr{S}_P} m(f, R)\mathrm{vol}(R) = \sum_{R_A \in \mathscr{S}_{P_A}} \left(\sum_{R_B \in \mathscr{S}_{P_B}} m(f, R_A \times R_B)\mathrm{vol}(R_B) \right) \mathrm{vol}(R_A).$$

If $x \in R_A$, then $m(g_x, R_B) \geq m(f, R_A \times R_B)$. Therefore, for any $x \in R_A$, we have

$$\sum_{R_B \in \mathscr{S}_{P_B}} m(f, R_A \times R_B)\mathrm{vol}(R_B) \leq \sum_{R_B \in \mathscr{S}_{P_B}} m(g_x, R_B)\mathrm{vol}(R_B) \leq \underline{\int_B} g_x = \mathscr{L}(x).$$

Therefore,

$$L(f, P) \leq L(\mathscr{L}, P_A).$$

Simimlarly, considering the upper Darboux sum $U(f, P)$, we have

$$U(\mathscr{U}, P_A) \leq U(f, P).$$

Then we have

$$L(f, P) \leq L(\mathscr{L}, P_A) \leq U(\mathscr{L}, P_A) \leq U(\mathscr{U}, P_A) \leq U(f, P)$$

and

$$L(f, P) \leq L(\mathscr{L}, P_A) \leq L(\mathscr{U}, P_A) \leq U(\mathscr{U}, P_A) \leq U(f, P).$$

Since f is integrable, then

$$\sup_{P \in \mathscr{P}_Q} \{L(f, P)\} = \inf_{P \in \mathscr{P}_Q} \{U(f, P)\}.$$

Consequently, we have

$$\int_Q f = \int_A \mathscr{L} \quad \text{and} \quad \int_Q f = \int_A \mathscr{U}.$$

Q.E.D

Remark 3.24. If f is integrable over $Q = A \times B$, then an similar proof shows that

$$\int_Q f = \int_B \left(\underline{\int_A} f \, \mathrm{d}x \right) \mathrm{d}y = \int_B \left(\overline{\int_A} f \, \mathrm{d}x \right) \mathrm{d}y.$$

In some book, Fubini's theorem is simply stated as

$$\int_Q f = \int_A \left(\int_B f \, \mathrm{d}y \right) \mathrm{d}x = \int_B \left(\int_A f \, \mathrm{d}x \right) \mathrm{d}y.$$

In that case, you should understand the integral $\int_B f \, \mathrm{d}y$ as a function $h(x)$ such that

$$\mathscr{L}(x) \leq h(x) \leq \mathscr{U}(x).$$

In fact, in Fubini's theorem, we can not claim that g_x is a integrable function for any $x \in A$. The integral $\int_B g_x$ may fail to exist for some $x \in A$. But $\int_B g_x$ can still by considered as an integral function.

Example 3.25. Let $f : [0, 1]^2 \to \mathbb{R}$ be the function defined by

$$f(x, y) = \begin{cases} 1 & \text{if } x \text{ is irrational, or } y \text{ is irrational, or } x = 0, 1, \\ 1 - 1/q & \text{if } x = p/q, \text{ where } p \text{ and } q \text{ are coprime, and } y \text{ is rational.} \end{cases}$$

Then it is easy to see that $\int_{[0,1]^2} f = 1$. But we see that $\int_0^1 f(x, y)\mathrm{d}y = 1$ when x is irrational and $\int_0^1 f(x, y)\mathrm{d}y$ does not exist when x is rational. However, we see that

$$\mathscr{U}(x) = \inf_{P \in \mathscr{P}_{[0,1]}} \{U(g_x, P)\} = \sum_{I \in \mathscr{S}_P} \mathrm{vol}(I) = 1$$

and

$$\mathscr{L}(x) = \sup_{P \in \mathscr{P}_{[0,1]}} \{L(g_x, P)\} = \begin{cases} 1 & \text{if } x \text{ is irrational,} \\ 1 - \frac{1}{q} & \text{if } x = p/q, \text{ where } p \text{ and } q \text{ are coprime.} \end{cases}$$

Moreover, both $\mathscr{U}(x)$ and $\mathscr{L}(x)$ are integrable.

In Fubini's theorem, if g_x is integrable for any $x \in A$, then the integral

$$\int_B g_x \mathrm{d}y = \mathscr{L}(x) = \mathscr{U}(x)$$

is a well-defined integrable function.

Corollary 3.26. *Let $Q = A \times B$, where A is an m-cell in $\mathbb{R}^m$ and B is an n-cell in $\mathbb{R}^n$. For each $x \in A$, let $g_x : B \to R$ be the function defined by $g_x(y) = f(x, y)$. If $\int_Q f$ exits and $\int_A g_x$ exists for each $x \in A$, then*

$$\int_Q f = \int_A \left(\int_B f(x, y)\mathrm{d}y \right) \mathrm{d}x.$$

One should be careful. Even if g_x is integrable for each $x \in A$, it does not mean that $h_y(x) = f(x, y)$ exists for every $y \in B$. For example, in Example 3.25, $\int_A f(x, y)\mathrm{d}x$ exists and $\int_A f(x, y)\mathrm{d}x = 1$ because set of rational numbers has measure zero in $\mathbb{R}$. However, $\int_B f(x, y)\mathrm{d}y$ does not exist when x is rational.

When the function f is a continuous, we can freely apply Fubini's theorem.

Corollary 3.27. *Let $Q = I_1 \times \cdots \times I_n$ be a n-cell in $\mathbb{R}^n$ and $f : Q \to \mathbb{R}$ be a continuous function, then*

$$\int_Q f = \int_{I_{i_n}} \left(\cdots \left(\int_{I_{i_1}} f \mathrm{d}x_{i_1} \right) \cdots \right) \mathrm{d}x_{i_n}.$$

Remark 3.28. In calculus, the following notation for integral is commonly used

$$\int_{a_n}^{b_n} \cdots \int_{a_1}^{b_1} f \, dx_1 \cdots dx_n.$$

However, it should be considered as the Riemann-Darboux integral. It is usually called multiple integral. When the function f is continuous, then by Fubini's theorem, its value is the same as the iterated integral

$$\int_{a_n}^{b_n} \left(\cdots \left(\int_{a_1}^{b_1} f \, dx_1 \right) \cdots \right) dx_n.$$

For a bounded set $C \subset A \times B$, we can also apply Fubini's theorem to evaluate the integral $\int_C f$.

Example 3.29. Let $C = [-1,1] \times [-1,1] \setminus \{(x,y) \mid x^2 + y^2 < 1\}$. Then

$$\int_C 1 = \int_{[-1,1]^2} \chi_C = \int_{-1}^{1} \left(\int_{-1}^{1} \chi_C(x,y) dy \right) dx.$$

Note that for each $x \in [-1,1]$ we have

$$\chi_C(x,y) \begin{cases} 1 & \text{if } y > \sqrt{1-x^2} \text{ or } y < -\sqrt{1-x^2} \\ 0 & \text{otherwise.} \end{cases}$$

Then

$$\int_{-1}^{1} \chi_C dy = \int_{C \cap \{x\} \times [-1,1]} 1 = \int_{-1}^{-\sqrt{1-x^2}} dy + \int_{\sqrt{1-x^2}}^{1} dy = 2(1 - \sqrt{1-x^2}).$$

Therefore,

$$\int_C 1 = \int_{-1}^{1} 2(1 - \sqrt{1-x^2}) dx = 2 \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} (1 - \cos x) d \sin x = 4 - \pi.$$

However, in general, the main difficulty is to determine $C \cap \{x\} \times B$ for $x \in A$ and $C \cap A \times \{y\}$. We will need more techniques.

3.3 Partition of unity

Sometimes, to evaluate the integral $\int_A f$, we may break the set A into bounded measurable subsets C_N over which $\int_{C_N} f$ may be easier to evaluate. Or alternatively, evaluate the integral $\int_A f\chi_{C_N}$ for good choices of C_N. More generally, we can first find a good covering $\{C_N\}$ of A and a piecewise defined linear function $\phi(x) = \phi_i(x)$ on C_i such that $\sum \phi_i = 1$, then take the sum of integrals $\int_{C_i} f\phi_i$. This approach involve a concept, called a “partition of unity", which has many advantages, especially for theoretic purpose.

The existence of partition of unity is the key feature that distinguish smooth structure from real-analytic or complex structure. It plays an very important role of passing local information to global, such as defining integral on smooth manifolds.

We first introduce a notation, the support of a function.

Definition 3.30. Let $\phi : \mathbb{R}^n \to R$ be a function. The *support* $\text{supp}(\phi)$ of ϕ is defined to be the closure of the set $\{x \mid \phi(x) \neq 0\}$.

Note that, equivalently, $x \notin \text{supp}(\phi)$ if and only if there exists a neighborhood U_x of x such that $\phi(y) = 0$ for any $y \in U_x$.

Theorem 3.31 (Partition of unity)**.** *Let $D \subset \mathbb{R}^n$ be a subset and $\mathscr{O} = \{U_i\}$ be an open covering of D. Then there exists a countable collection Φ of C^∞ functions $\phi_i : \mathbb{R}^n \to \mathbb{R}$ such that the following properties hold.*

(a) $\phi_i(x) \geq 0$ *for any* $x \in \mathbb{R}^n$*,*

(b) each point x of D has a neighborhood that intersects only finitely many of the support sets $\text{supp}(\phi_i)$*,*

(c) $\sum_{i=1}^{\infty} \phi_i(x) = 1$ *for any* $x \in D$*,*

(d) for each i, the support $\text{supp}(\phi_i)$ *is compact and contained in an element* $U_j \in \mathscr{O}$*.*

Definition 3.32. A countable collection Φ of functions $\{\phi_i\}$ satisfying the properties (3.31.(a)-3.31.(c)) is called a *partition of unity* for D. If a partition of unity for D satisfies (3.31.(d)) is said to have *compact supports*.

The existence of functions supported locally in an open set is easy to prove.

Lemma 3.33. *Let $Q = [a_1, b_1] \times \cdots \times [a_n, b_n]$ be an n-cell in $\mathbb{R}^n$. There exists a C^∞ function $\phi : \mathbb{R}^n \to \mathbb{R}$ such that $\phi(x) > 0$ for $x \in Q^{int}$ and $\phi(x) = 0$ otherwise.*

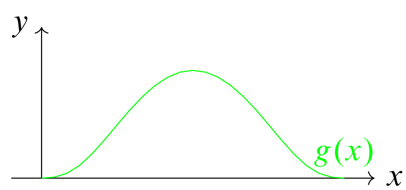
Proof. Let $f : \mathbb{R} \to \mathbb{R}$ be the function defined by

$$f = \begin{cases} e^{-\frac{1}{x}} & \text{if } \ x > 0, \\ 0 & x \leq 0. \end{cases}$$

Then $f(x) > 0$ for $x > 0$ and f is of class C^∞. Now define

$$g(x) = f(x) f(1-x).$$

Then $g(x)$ is of class C^∞ and g is positive for $0 < x < 1$ and zero otherwise.



Therefore,

$$\phi(x) = g(\frac{x_1 - a_1}{b_1 - a_1}) \cdots g(\frac{x_n - a_n}{b_n - a_n})$$

is a function such that $\phi(x) > 0$ for $x \in Q^{int}$ and zero otherwise. Q.E.D

More generally, we have the following result.

Corollary 3.34. *Let $C \subset \mathbb{R}^n$ be a compact subset. Then there exist a compact set D containing C and a function $\phi(x)$ of C^∞ class such that $\phi(x) > 0$ for $x \in C$ and $\phi(x) = 0$ for any $x \notin D$.*

Proof. Since C is compact, it can be covered by finitely many open sets $Q_1^{int}, \ldots, Q_m^{int}$, where Q_i are n-cells. By the above lemma, for each n-cell Q_i, we have a function $\phi_i(x)$ such that $\phi_i(x) > 0$ for any $x \in Q_i^{int}$ and $\phi_i(x) = 0$ otherwise. Taking the sum $\phi(x) = \sum_{i=1}^m \phi_i(x)$ and the union $D = \cup_{i=1}^n Q_i$, we see that ϕ is of class C^∞ such that $\phi(x) > 0$ for $x \in C$ and $\phi(x) = 0$ for $x \notin D$. Q.E.D

Remark 3.35. By the following lemma, in Corollary 3.34, if C is contained in an open subset U, then we can choose D such that $D \subset U$ and $C \subset D^{int}$.

Lemma 3.36. *Let C be a compact set and U be an open set in $\mathbb{R}^n$. Assume that $C \subset U$. Then there exists a compact set D such that $C \subset D^{int}$ and $D \subset U$.*

Proof. Denote by d the distance between C and the complement U^c, i.e. $d = \min\{d(x, y) \mid x \in C, y \in U^c\}$. Since C is contained in U. Then $d > 0$. For each $x \in C$, denote by B_x the open set $\{y \in \mathbb{R}^n \mid d(x, y) < \frac{d}{2}\}$. Since C is compact, there are finitely many such open sets, say $B_1, \ldots, B_m$, covering C. Let D be the union of closures of $B_1, \ldots, B_m$. Then D is compact, $D \subset U$ and $C \subset D^{int}$. Q.E.D

To prove the theorem of partition of unity, we first prove the theorem under additional condition that D is compact.

Lemma 3.37. *A partition of unity exists when D is compact in $\mathbb{R}^n$.*

Proof. Since D is compact, then a finite subcollection $\{U_1, \ldots, U_m\}$ of $\mathscr{O}$ will cover D. By Corollary 3.34, it suffices to construct compact sets D_i for each i such that $D_i \subset U_i$ and whose interior D_i^{int} covers D. The construction can be done inductively using Lemma 3.36. Let

$$C_i = \begin{cases} D \setminus (\bigcup_{i=2}^{m} U_i) & i = 1, \\ D \setminus (D_1^{int} \cup \cdots D_k^{int} \cup U_{k+2} \cup \cdots \cup U_m) & 2 \leq i \leq m-1, \\ D \setminus (D_1^{int} \cup \cdots D_{m-1}^{int}) & i = m. \end{cases}$$

Since D is compact, then each C_i is a compact set and $C_i \subset U_i$. By Lemma 3.36, for each i, there is a compact set D_i contained in U_i such that C_i is contained in the interior D_i^{int}. Moreover, by the construction of C_i, the open sets D_i^{int} covers D. By Corollary 3.34, for each i, there is a function ϕ_i of class C^∞ such that $\phi_i(x) > 0$ for any $x \in D_i$ and $\phi_i(x) = 0$ for any $x \notin E_i$, where E_i is a closed subset containing D_i.

Let ψ_i be the function defined by

$$\psi_i(x) = \frac{\phi_i(x)}{\phi_1(x) + \cdots + \phi_m(x)}.$$

Then ψ_i are the desired functions. Q.E.D

Remark 3.38. A partition of unity $\Phi = \{\phi_i\}$ such that the set $\mathrm{supp}(\phi_1) \subset U_i$ for each i is said to be *subordinated to the collection* $\mathscr{O}$. If D is compact with a finite covering $\mathscr{O}$, then there is a partition of unity subordinated to $\mathscr{O}$ with compact support. In general, if D is paracompact, i.e. every open covering has a refinement which is locally finite, then there exist a partition of unity subordinated to the covering. However, the partition of unity may not have compact support.

Proof of partition of unity theorem. We proof the theorem by reduction to easier cases.

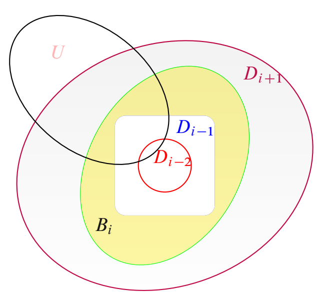
Case 1: D is compact. The existence of partition of unity subordinate to $\mathscr{O}$ follows from Lemma 3.37.

Case 2: $D = D_1 \cup D_2 \cup \cdots$, where each D_i is compact and $D_i \subset D_{i+1}^{int}$.

For each i, let

$$\mathscr{O}_i = \{U \cap (D_{i+1}^{int} \setminus D_{i-2}) \mid U \in \mathscr{O}\}.$$

Then $\mathscr{O}_i$ is an open cover of the compact set $B_i = D_i - D_{i-1}^{int}$.



By case 1, there is a partition of unity Φ_i for B_i subordinate to $\mathscr{O}_i$. For each $x \in D$, the sum

$$\sigma(x) = \sum_i (\sum_{\phi \in \Phi_i} \phi(x))$$

is a finite sum in some open neighborhood of x. Because if $x \in D_i$, we have $\phi(x) = 0$ for any $\phi \in \Phi_j$ with $j \geq i + 2$. For each $\phi \in \Phi_i$, define $\phi'(x) = \frac{\phi(x)}{\sigma(x)}$. The the collection of all ϕ' is the desired partition of unity.

Case 3: assume that D is open.

Let $D_i = \{x \in D \mid \|x\| \leq i \text{ and } d(x, \partial D) \geq \frac{1}{i}\}$. Then we are reduced to case 2.

Case 4: assume that D is arbitrary.

Let B be the union of all $U \in \mathscr{O}$. By case 3, there is a partition of unity of B. Hence there is also a partition of unity of D. Q.E.D

Example 3.39. Let $f : \mathbb{R} \to \mathbb{R}$ be a function defined by

$$f(x) = \begin{cases} \frac{1+\cos x}{2} & \text{for } -\pi \leq x \leq \pi, \\ 0 & \text{otherwise.} \end{cases}$$

Then f is of class C^1. For each integer $k \geq 0$, let $\phi_{2k+1}(x) = f(x - k\pi)$ and $\phi_{2k}(x) = f(x + k\pi)$. Then $\{\phi_i\}$ is a partition of unity of $\mathbb{R}$.

Using partition of unity with compact support, we can extend the definition of integral to arbitrary set in $\mathbb{R}^n$.

Definition 3.40. Let D be an open subset in $\mathbb{R}^n$ and $f : D \to \mathbb{R}$ be a continuous function. Let $\{\phi_i\}$ be a partition of unity on D having compact support. We say f is integrable over D if the series

$$\sum_{i=1}^{\infty} \left(\int_D \phi_i |f|\right)$$

converges.

In the case that $\sum_{i=1}^{\infty}(\phi_i |f|)$ converges, we define the integral

$$\int_D f = \sum_{i=1}^{\infty}(\phi_i f)$$

called the *extended integral*.

Since ϕ_i are nonnegative, then that $\sum_{i=1}^{\infty}(\int_D \phi_i |f|)$ converges implies that $\sum_{i=1}^{\infty}(\int_D \phi_i f)$ converges uniformly and hence itself is converges. It then makes sense to define the integral as the value of convergence. However, we have to show that the definition is independent on the partition of unity.

Theorem 3.41. *Let D be an open subset in $\mathbb{R}^n$ and $f : D \to \mathbb{R}$ be a continuous function. Let $\Phi = \{\phi_i\}$ and $\Psi = \{\psi_j\}$ be two partitions of unity on D having compact supports.*

(a) If $\sum_{\phi_i \in \Phi}(\int_D \phi_i |f|)$ *converges, then* $\sum_{\psi_j \in \Psi}(\int_D \psi_j |f|)$ *also converges and*

$$\sum_{\phi_i \in \Phi}(\int_D \phi_i |f|) = \sum_{\psi_j \in \Psi}(\int_D \psi_j |f|).$$

(b) If D *and* f *are bounded, then* f *is integrable in the extended sense.*

(c) If D *and* f *are bounded and the boundary* ∂D *has measure zero, then the extended integral* $\int_D f$ *is the same as the old one.*

The first statement comes essentially from refining partitions of unity. The second one is easy since D is bounded. To prove the third one, we observe that there are compact sets contained in D such that the volume of difference can be as small as possible. This observation is also used to define so-called improper integrals.

Lemma 3.42. *Let* D *be a rectifiable set in* $\mathbb{R}^n$*. Then for any* $\varepsilon > 0$*, there is a compact rectifiable set* C *contained in* D *such that* $\int_{D \setminus C} 1 < \varepsilon$*.*

Proof. Let $\overline{D}$ be the closure of D in $\mathbb{R}^n$. We see that the boundary $\partial D = \overline{D} \setminus D^{int}$ is closed. Since D is rectifiable, then the boundary ∂D is bounded. Therefore, the boundary ∂D is compact and has measure zero. By definitions of compactness and having measure zero, we know that for any $\varepsilon > 0$, there are finitely many n-cells $Q_1^{int}, \ldots, Q_m^{int}$ covering ∂D such that $\sum_{i=1}^m \mathrm{vol}(Q_1) < \varepsilon$. Let

$$C = \overline{D} \setminus (\bigcup_{i=1}^m Q_i^{int}).$$

Then C is a closed and bounded set contained in D. Therefore, it is compact and

$$\int_{D \setminus C} \leq \sum_{i=1}^m \mathrm{vol}(Q_1) < \varepsilon.$$

Q.E.D

Proof of Theorem 3.41. (a) By definition, $\phi_i f = 0$ except on some compact subset $C \subset \mathrm{supp}(\phi_i)$. By definition again, we know that only finitely many ψ_i are non-zero on C and we can write

$$\sum_{\phi_i\in\Phi}\int_D \phi_i f = \sum_{\phi_i\in\Phi}\sum_{\psi_j\in\Psi}\int_D \psi_j\phi_i f = \sum_{\psi_j\in\Psi}\sum_{\phi_i\in\Phi}\int_D \phi_i\psi_j f = \sum_{\psi_j\in\Psi}\int_D \psi_j f.$$

Replacing f by $|f|$ from those equalities, we see that the convergence of $\sum_{\phi_i\in\Phi}\int_D \phi_i|f|$ implies the converges of

$$\sum_{\phi_i\in\Phi}\sum_{\psi_j\in\Psi}\int_D \psi_j\phi_i|f|$$

and hence the convergence of

$$\sum_{\psi_j\in\Psi}\sum_{\phi_i\in\Phi}\int_D \phi_i\psi_j|f| = \sum_j\int_D \psi_j|f|.$$

Moreover, $\sum_{\psi_j\in\Psi}\int_D \psi_j f$ exists and

$$\sum_{\phi_i\in\Phi}\int_D \phi_i f = \sum_{\psi_j\in\Psi}\int_D \psi_j f.$$

(b) Since D is bounded, then there is an n-cell Q containing D. Let M be a number such that $f(x)\leq M$ for any $x\in D$ and $\{\phi_1,\ldots,\phi_m\}$ is a finite partition of unity for Q. Then

$$\sum_{i=1}^m\int_D \phi_i|f| \leq M\sum_{i=1}^m\int_D \phi_i \leq M\int_D\sum_{i=1}^m \phi_i \leq M\operatorname{vol}(Q).$$

Therefore, f is integrable in the extended sense.

(c) By Lemma 3.42, for any $\varepsilon>0$, there is a compact set C such that $\int_{D\setminus C}1<\varepsilon$. Let M be a number such that $|f(x)|\leq M$ for any $x\in D$. Since C is compact, then the functions ϕ_i which are nonzero on C form a finite subcollection $\mathscr{F}\subset\Phi$. Therefore,

$$\begin{aligned}|\int_D f - \sum_{\phi_i\in\mathscr{F}}\int_D \phi_i f| &\leq \int_D|f-\sum_{\phi_i\in\mathscr{F}}\phi_i f| \leq M\int_D(1-\sum_{\phi_i\in\mathscr{F}}\phi_i)\\ &= M\int_D\sum_{\phi_i\in\Phi\setminus\mathscr{F}}\phi_i \leq M\int_{D\setminus C}1 < M\varepsilon.\end{aligned}$$

Q.E.D

3.4 Change of variables

The definition of integral over general bounded sets that we have discussed in previous section is theoretically very useful, practically not easy to apply. In single variable function, we have the substitution method. In this section, we generalize this method to the change of variable formula.

Theorem 3.43 (Change of variables)**.** *Let $D \subset \mathbb{R}^n$ be an open set and $\varphi : D \to \mathbb{R}^n$ be a one-to-one function of class C^1 such that $\det(D\phi)(x) \neq 0$ for each $x \in D$. Let $f : \phi(D) \to \mathbb{R}$ be a function. If f is integrable over $\phi(D)$, then $(f \circ \phi) \cdot |\det(D\phi)|$ is integrable and*

$$\int_{\phi(D)} f = \int_D (f \circ \phi) \cdot |\det(D\phi)|.$$

The heuristic reason the theorem is true is based on the following observation: for a sufficiently small n-cell Q, the volume $\mathrm{vol}(Q)$ changes to

$$\mathrm{vol}(\phi(Q)) \approx |\det(D\phi)(x)| \mathrm{vol}(Q)$$

for some $x \in Q$. More precisely,

$$\mathrm{vol}(\phi(Q)) = \int_{\phi(Q)} 1 = \int_Q |\det(D\phi)|.$$

The proof of this theorem is quite involved. The usual proof goes by reduction successfully to simpler cases and uses the approximation of integrals by finite sums. In the appendix, you will see another beautiful proof of change of variables originally due to Peter Lax.

Proof of Theorem 3.43.
Step 1: It suffices to show the theorem hold for the function $f = 1$.

If the theorem holds for $f = 1$, then it holds for constant functions. Now for a general integral function, we prove the theorem by definition. Let Q be an n-cell containing $\phi(D)$

and P be a partition of Q. Then

$$
\begin{aligned}
L(f,P) &= \sum_{R\in\mathscr{S}_P} m_R(f)\text{vol}(R^{int}) = \sum_{R\in\mathscr{S}_P}\int_{R^{int}} m_R(f) \\
&= \sum_{R\in\mathscr{S}_P}\int_{\phi^{-1}(R^{int})} m_R(f)\circ\phi|\det(D\phi)| \\
&\leq \sum_{R\in\mathscr{S}_P}\int_{\phi^{-1}(R^{int})} (f\circ\phi)\cdot|\det(D\phi)| \\
&\leq \int_{\phi^{-1}(Q)} (f\circ\phi)\cdot|\det(D\phi)|.
\end{aligned}
$$

Therefore,

$$\int_Q f \leq \int_{\phi^{-1}(Q)} (f\circ\phi)\cdot|\det(D\phi)|.$$

Similarly, we can prove that

$$\int_{\phi^{-1}(Q)} (f\circ\phi)\cdot|\det(D\phi)| \leq \int_Q f.$$

Consequently, we have

$$\int_{\phi(D)} f = \int_D (f\circ\phi)\cdot|\det(D\phi)|.$$

Step 2: if the theorem is true for all functions with compact supports, then it is true. Here we will use partition of unity.

Let $\mathscr{O} = \{U_i\}$ be an open cover of D. Then $\phi(\mathscr{O}) = \{\phi(U_i)\}$ is an open cover of $\phi(D)$. Take a partition of unity $\mathscr{A} = \{p_i\}$ for $\phi(\mathscr{O})$ with compact supports. Then $\mathscr{A}_\phi = \{p_i\circ\phi\}$ is a partition of unity for $\mathscr{O}$ with compact supports because ϕ^{-1} exists and is a one-to-one function of class C^1. Then

$$
\begin{aligned}
\int_{\phi(D)} f &= \sum_{p_i\in\mathscr{A}}\int_{\phi(D)} p_i f = \sum_{p_i\in\mathscr{A}}\int_{\phi(U_i)} p_i f \\
&= \sum_{p_i\in\mathscr{A}}\int_{U_i} ((p_i f)\circ\phi)|\det D\phi| = \int_D (f\circ\phi)\cdot|\det(D\phi)|.
\end{aligned}
$$

Step 3: if the theorem is true for $\phi : D \to \mathbb{R}^n$ and $\psi : E \to \mathbb{R}^n$ such that $\phi(D) \subset E$, then the theorem is true for $\psi \circ \phi$.

Let $f : \psi(E) \to \mathbb{R}$ be a function integrable over $\psi(\phi(D))$. Then $(f \circ \psi)|\det(D\psi)|$ is integrable over $\phi(D)$. Hence $((f \circ \psi) \cdot |\det(D\psi)|) \circ \phi)$ is integrable over D and

$$\int_D f = \int_{(\psi\circ\phi)(D)} (f \circ \psi \circ \phi) \cdot (|\det(D\psi)| \circ \phi) \cdot |\det(D\phi)|.$$

It then suffices to show that

$$|D(\psi \circ \phi)| = |D\psi(\phi)| \cdot |D\phi|.$$

While, this equality follows exactly from the chain rule:

$$D(\psi \circ \phi) = D\psi(\phi) \cdot D\phi.$$

Step 4: if ϕ is a linear map, i.e. there exists a matrix M such that $\phi(x) = M \cdot x$, then the theorem is true. By step 1, it suffices to show for each open n-cell U that

$$\int_{\phi(U)} 1 = \int_U |\det(D\phi)|.$$

By step 3, we may assume that M is an elementary matrix. In that case, the claim is easy to prove.

Step 5: The theorem is true for single variable functions. This is easy to prove.

Step 6: Now we prove the theorem by induction for general ϕ. We may assume that the theorem is true in dimension $n - 1$.

By step 1, it suffices to localize at the point $x_0 \in D$. Moreover, by step 3 and step 4, we may assume that $D\phi(x_0) = I$ is an identity matrix.

Define $\psi : D \to \mathbb{R}$ by $\psi(x) = (\phi_1(x), \ldots \phi_{n-1}(x), x_n)$. Then $D(\psi)(x_0) = I$. Therefore, by inverse function theorem, in the open neighborhood U of x_0, the function ψ is one-to-one and $\det(D\psi) \neq 0$. Now define $\rho : \psi(U) \to \mathbb{R}^n$ by $\rho(x) = (x_1, \ldots, x_{n-1}, \phi_n(\psi^{-1}(x)))$. Then $\phi = \rho \circ \psi$. It follows that $(D\rho)(\psi(x_0)) = I$. By inverse function theorem again, there is an open neighborhood V of $\psi(x_0)$ such that ρ is one-to-one and $\det(D\rho) \neq 0$ on V. By step 3 again, we only need to prove the claim for ψ and ρ. The idea is to apply Fubini's theorem.

We give the proof for ψ, the proof for ρ is similar.

Let $W = \psi^{-1}(V)$ and $Q = E \times [a_n, b_n]$ be a n-cell in W. Then by Fubini's theorem,

$$\int_{\psi(Q)1} = \int_{[a_n,b_n]} (\int_{\psi(E\times\{x_n\})} 1 \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1}) \mathrm{d}x_n.$$

Let $\psi_{x_n} : E \to \mathbb{R}^{n-1}$ be the function define by

$$\psi_{x_n}(x_1, \ldots, x_{n-1}) = (\phi_1(x), \ldots, \phi_{n-1}(x)).$$

Then h_{x_n} is one-to-one and $\det(Dh_{x_n}) = \det(Dh) \neq 0$ over E. Moreover,

$$\int_{\psi(E\times\{x_n\})} 1 \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} \mathrm{d}x_n = \int_{\psi(E)} 1 \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1}.$$

Apply the theorem in dimension $n-1$, we see that

$$\int_{\psi(W)} 1 = \int_{[a_n,b_n]} (\int_E |\det(D\psi_{x_n})| \mathrm{d}x_1 \cdots \mathrm{d}x_{n-1}) \mathrm{d}x_n = \int_W |\det(D\psi)|.$$

Q.E.D

There are many generalizations and variants of the theorem of change of variables. In fact, the condition that $\det(\phi) \neq 0$ may be dropped from the hypotheses using Sard's theorem.

Theorem 3.44 (Sard's Theorem). *Let $\phi : D \to \mathbb{R}^n$ be a continuously differentiable map with D open. Let $B = \{x \in D \mid \det(D\phi)(x) = 0\}$. Then $g(B)$ has measure zero.*

Roughly speaking, the reason $g(B)$ has measure zero in $\mathbb{R}^n$ is that B is a dimension $n-1$ subset of D.

A Lax's theorem of change of variables

We present in this section the idea of Lax to prove change of variables. We first prove Theorem A.1 and then prove the standard version using partition of unity. The idea to prove Theorem A.1 is a generalization of a proof of single variable formula using fundamental theorem of calculus. For a single variable function f, assume that it has an antiderivative g. Then $\int_a^b f\,\mathrm{d}x = g(b) - g(a)$. Now if there is a one-to-one continuously differentiable may $\phi : [c, d] \to [a, b]$, then $g \circ \phi$ is an antiderivative of $f \circ \phi |\mathrm{d}\phi|$.

We follow Taylor's proof.

Theorem A.1 (Lax 1999)**.** *Let $D \subset \mathbb{R}^n$ be an open set and $\varphi : D \to \mathbb{R}^n$ be a one-to-one function of class C^1 such that $\phi(x) = x$ for each $x \in D \setminus B^{int}$, where B^{int} is an open n-ball contained in D. Let $f : \phi(D) \to \mathbb{R}$ be a function. Then*

$$\int_{\phi(D)} f = \int_D (f \circ \phi) \cdot |\det(D\phi)|.$$

Proof. By smooth approximation, we may assume that ϕ is of class C^2 and f is of class C^1. Pick up a number $a > 0$ such that the function $f_a(x) := f(x - ae_1)$ is supported in B^c, where $e_1 = (1, 0, \ldots, 0)$ and B^c is the complement of B. By taking a sufficiently large, we may assume that $\operatorname{supp}(f_a)$ dose not intersect with $\phi(B)$. Let $g = \int_0^a f(x - te_1)\mathrm{d}t$. Then

$$F(x) := f(x) - f(x - ae_1) = \frac{\partial g}{\partial x_1}.$$

We then have the following equality

$$\alpha := F\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n = \mathrm{d}(g\mathrm{d}x_2 \wedge \cdots \wedge \mathrm{d}x_n).$$

Note that

$$\begin{aligned}\phi^*\alpha &= F(\phi)\det(D\phi)\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n \\ &= f(\phi)\det(D\phi)\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n - f_a\mathrm{d}x_1 \wedge \cdots \wedge \mathrm{d}x_n.\end{aligned}$$

Since f_a is supported in B^c, then we have

$$\int_D f(\phi)\det(D\phi) = \int_{\phi(D)} f_a + \int_D \mathrm{d}(\phi^*\beta) = \int_D \mathrm{d}(\phi^*\beta),$$

where $\beta = g\mathrm{d}x_2 \wedge \cdots \wedge \mathrm{d}x_n$. It then suffices to show that $\int_D \mathrm{d}(\phi^*\beta) = \int_{\phi(D)} f$. By stokes theorem, we have

$$\int_D \mathrm{d}(\phi^*\beta) = \int_{\partial D} (\phi^*\beta).$$

Using integration by parts and that $f_a = 0$ on ∂D, we see that

$$\int_{\partial D} (\phi^*\beta) = \int_D g(\phi) - \int_D g(\phi)\mathrm{d}(\mathrm{d}\phi_2 \wedge \cdots \wedge \mathrm{d}\phi_n) = \int_{\phi(D)} f.$$

Q.E.D

Theorem A.2. *Let $D \subset \mathbb{R}^n$ be an open set and $\varphi : D \to \mathbb{R}^n$ be a one-to-one function of class C^1 such that* $\det(D\phi)(x) \neq 0$ *for each $x \in D$. Let $f : \phi(D) \to \mathbb{R}$ be a function with compact support in D. Then*

$$\int_{\phi(D)} f = \int_D (f \circ \phi) \cdot |\det(D\phi)|.$$

Proof. Denote by S the support of f. Let Q be an n-cell contained in D and whose interior Q^{int} contains S. Then there is a portion of unit $\{\alpha, \beta\}$ subordinate to $\{Q^{int}, D \setminus Q\}$. Let $\psi(x) = \alpha\phi + x\beta$. Then $\psi(x) = x$ outside some n-ball contains Q^{int} and $\psi(x) = \phi(x)$ for x in S. Then apply Theorem A.1, we see that

$$\int_{\phi(D)} f = \int_{\phi(S)} f = \int_{\psi(S)} f = \int_{\psi(S)} (f \circ \psi) \cdot |\det(D\psi)| = \int_{\phi(S)} (f \circ \phi) \cdot |\det(D\phi)|.$$

Q.E.D

4 Differential forms

Differential forms $\sum_I f_I \mathrm{d}x_I$ play a very important role in analysis and differential geometry. A heuristic reason of introducing differentials is the practical truth of successfully using infinitesimal changes in classical analysis, differential geometry and physics.

Given a differentiable function $f : \mathbb{R}^m \to \mathbb{R}$, we know that the infinitesimal increment $\Delta f(a) = f(x) - f(a)$ of f around a point $a \in \mathbb{R}^m$ is approximately

$$\Delta f(a) \approx Df(a) \cdot \Delta_a(x) = \sum_{i=1}^{m} D_i f(a)(x_i - a_i),$$

where $\Delta_a(x) = x - a = (x_1 - a_1, \ldots, x_m - a_m)$ is a vector in $\mathbb{R}^m$ representing the directional increment of the variable x . Consider the simplest case that $m = 1$, we notice that the expression $Df(a)\Delta_a(x)$ also appears in the definition of (Riemann) integral of the single variable function Df. The integral of f is usually written as $\int Df \,\mathrm{d}x$.

The notation $\mathrm{d}x$ introduced by Leibniz in 17th-century was then commonly adapted to describe an infinitely small change. Differential forms appeared later in mid-18th-century as the integrand and extensively used since then. However, before Elie Cartan defined them in 1899, no real attempt to define the forms themselves. For histories of differential forms, I suggest you read the papers by Victor J. Katz.

The history of differential forms starts from 1-forms. A *differential 1-form* on $\mathbb{R}^m$ is an expression

$$f_1 \mathrm{d}x_1 + \cdots + f_m \mathrm{d}x_m,$$

where f_i are continuous functions on $\mathbb{R}^m$. The *differential* $\mathrm{d}f$ of a differentiable function $f : \mathbb{R}^m \to R$ is defined as

$$\mathrm{d}f = \sum_{i=1}^{m} D_i f \,\mathrm{d}x_i.$$

Let ω be a differential 1-form. From here, you may ask many questions, for example

- What's the geometric meaning of ω?

- Is is true that $\omega = \mathrm{d}f$ for any ω?

For the first question, we look back the motivations. We intend to think of a 1-form as an infinitesimal increment of the function f as well as the integrand (along a curve). As an infinitesimal increment, a 1-form bears the meaning of "vector" and changes with respect to the increment of x. As an integrand, we may simply consider a 1-form as the "infinitesimal volume". You might think of differential 1-forms as tangent vectors. As an infinitesimal increment, $\mathrm{d}f$ changes along a vector v_p initiated at the point p. In this sense, we may view $\mathrm{d}f$ and hence a 1-form as functions on the tangent vector space. It turns out this point of view makes perfect sense also for "infinitesimal volume". We will see that later.

Once it was noticed that a differential 1-form $\omega = f_1\mathrm{d}x_1 + \cdots + f_m\mathrm{d}x_m$ should be considered as a function on $\mathbb{R}^m \times T_{\mathbb{R}^m}$, it becomes clear how to define the function ω. The function ω should be linear with respect to direction and the output with respect to a tangent direction should be the directional derivative. It then make sense to define

$$\mathrm{d}x_i(\frac{\partial}{\partial x_j}) = \delta_{ij} := \begin{cases} 1 & i = j, \\ 0 & i \neq j, \end{cases}$$

where $\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_1}$ is a basis of the tangent space $T_{\mathbb{R}^m}$. More generally, for any tangent vector $v_p = (k_1, \ldots, k_m) = \sum_{i=1}^m k_i \frac{\partial}{\partial x_j}$ at a point $p \in \mathbb{R}^m$, we have

$$\omega(p, v_p) := \omega(v_p) = \sum_{i=1}^{m} k_i f_i(p) = (f_1(p), \ldots, f_m(p)) \begin{pmatrix} k_1 \\ \vdots \\ k_m \end{pmatrix}.$$

Now for the second question, it is easy to find a 1-form ω which is not the differential of a function. For example, there is not differentiable function $f(x, y)$ such that $\mathrm{d}f = y\mathrm{d}x - x^3\mathrm{d}y$. Otherwise, we must have $\frac{\partial f}{\partial x} = y$ and $\frac{\partial f}{\partial y} = -x^3$. However, we would have $1 = \frac{\partial f}{\partial y \partial x} = \frac{\partial f}{\partial x \partial y} = -2x^2$, which is impossible. Since not every 1-form is the differential of a function, then you may ask under what condition, we may have $\omega = \mathrm{d}f$. The answer is that $\omega = \mathrm{d}f$ if and only if $\mathrm{d}\omega = 0$. It comes another questions what is $\mathrm{d}\omega$? It is true that $\int \mathrm{d}\omega = \omega$? This is an intention of generalizing of the fundamental theorem of calculus.

In this chapter, we will generalize the simplest 1-form $f\,\mathrm{d}x$, answer the above mentioned question, and introduce some applications.

4.1 Multilinear algebra

When generalizing 1-forms to k-forms, we want to keep as many features that 1-forms have as possible. For instance, geometrically, a k-forms should be interpreted as an integrand and the k-th order approximation. Algebraically, it should be considered as a function which is linear along the tangent spaces and bears some reasonable algebraic operations such as summation, multiplication. With those motivations, Elie Cartan adapted some viewpoints of Grassman and eventually gave the formal definition of k-forms. For the purpose of using k-forms, we will learn the formal algebraic definition first and understand their geometric meaning.

For simplicity, our vector spaces will always be over the field of real numbers $\mathbb{R}$. Let V be a vector space over $\mathbb{R}$. We denote the k-fold product $V \times \cdots \times V$ by V^k. A function $T : V^k \to \mathbb{R}$ is called *multilinear* if for each i with $1 \le i \le k$ and any $a, b \in \mathbb{R}$ we have

$$T(v_1, \ldots, av_i + bv_i', \ldots, v_k) = aT(v_1, \ldots, v_i, \ldots, v_k) + bT(v_1, \ldots, v_i', \ldots, v_k).$$

A multilinear function $T : V^k \to \mathbb{R}$ is also called a *k-tensor* on V. The set of k-tensors on V, denoted by $\mathscr{T}^k(V)$, admits a natural vector space structure defined as follows. For any $T, S \in \mathscr{T}^k(V)$ and $a, b \in \mathbb{R}$, we define

$$(aS + bT)(v_1, \ldots, v_k) = aS(v_1, \ldots, v_k) + bT(v_1, \ldots, v_k).$$

As set of functions, we may want to define multiplications on $\mathscr{T}^k(V)$ and expect the a k-tensor is generated by addition of multiplication of 1-tensors. To keep the multilinearity, there is a natural multiplicative operator on the spaces $\mathscr{T}^k(V)$ defined as follows. For any $S \in \mathscr{T}^k(V)$ and $T \in \mathscr{T}^l(V)$, we define the *tensor product* $S \otimes T$ by

$$(S \otimes T)((v_1, \ldots, v_k), (w_1, \ldots, w_l)) = S(v_1, \ldots, v_k)T(w_1, \ldots, w_l).$$

It is not hard to check that $S \otimes T$ is a multilinear function on $V^{k+l} = V^k \times V^l$. So $S \otimes T$ is a $(k+l)$-tensor. Moreover, the tensor product operation satisfies some properties as stated in the following theorem.

Theorem 4.1. *Let R, S, T be tensors on a vector space V over $\mathbb{R}$ and a be a number in $\mathbb{R}$. Then*

(a) (Distributivity)

$$(R+S)\otimes T = R\otimes T + S\otimes T$$
$$R\otimes (S+T) = R\otimes S + R\otimes T$$

(b) (Associativity)

$$(R\otimes S)\otimes T = R\otimes (S\otimes T).$$

(c) (Homogeneity)

$$(aS)\otimes T = a(S\otimes T) = S\otimes (aT).$$

Proof. Easy exercise. Q.E.D

Remark 4.2. (a) A k-tensor T on V is homogeneous, i.e. $T(av) = aT(v)$, and its image is a sub vector space of $\mathbb{R}$. .

(b) The set of 1-tensors as a vector space is exactly the dual space of V and usually denoted by V^*.

(c) Note that in general tensor product is not commutative. For example. Let $S = x + y$ and $T = x - y$ be two 1-tensors on $V = \mathbb{R}^2$. Then

$$(S\otimes T)(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix}) = -2, \quad \text{but} \quad (T\otimes S)(\begin{pmatrix}1\\1\end{pmatrix}, \begin{pmatrix}1\\2\end{pmatrix}) = 0.$$

Knowing the properties of tensors, we can prove the following theorem which tells us that k-tensors are generated by 1-tensors.

Theorem 4.3. *Let $e_1, \ldots, e_m$ be a basis of the vector space V and let $\phi_1, \ldots, \phi_m$ be the dual basis, i.e. $\phi_i(e_j) = \delta_{ij}$. Then the set of all k-tensor products*

$$\phi_{i_1}\otimes\cdots\otimes\phi_{i_k} \quad 1 \le i_1, i_2, \ldots, i_k \le m$$

is a basis for $\mathscr{T}^k(V)$. Hence $\mathscr{T}^k(V)$ has dimension m^k.

Proof. We need to show that $\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}$ span $\mathscr{T}^k(V)$ and are linearly independent. Note that

$$\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}(e_{j_1}, \ldots, e_{j_k}) = \begin{cases} 1 & \text{if } (i_1, \ldots, i_k) = (j_1, \ldots, j_k) \\ 0 & \text{otherwise.} \end{cases}$$

For any $T \in \mathscr{T}^k(V)$ and k vectors $v_1, \ldots, v_k$ with $v_j = \sum_{i=1}^m a_{ji} e_i$, we have

$$\begin{aligned} T(v_1, \ldots, v_k) &= \sum_{i_1,\ldots,i_k=1}^m a_{1i_1} \cdots a_{ki_k} T(e_{i_1}, \ldots, e_{i_k}) \\ &= \sum_{i_1,\ldots,i_k=1}^m T(e_{i_1}, \ldots, e_{i_k})(\phi_{i_1} \otimes \cdots \otimes \phi_{i_k})(v_1, \ldots, v_k). \end{aligned}$$

Therefore, $T = \sum_{i_1,\ldots,i_k=1}^m T(e_{i_1}, \ldots, e_{i_k})\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}$.

Now assume that there are numbers $a_{i_1,\ldots,i_k}$ such that

$$\sum_{i_1,\ldots,i_k=1}^m a_{i_1,\ldots,i_k} \phi_{i_1} \otimes \cdots \otimes \phi_{i_k} = 0.$$

By evaluating the left hand side tensor at each basis $(e_{i_1}, \ldots, e_{i_k})$, we see that $a_{i_1,\ldots,i_k} = 0$.

Therefore, $\phi_{i_1} \otimes \cdots \otimes \phi_{i_k}$ form a basis for $\mathscr{T}^k(V)$. Q.E.D

In the above proof, there is something interesting. If we let $I = \sum_{i=1}^m \phi_i \otimes \phi_i$. Then $I(w, v) =< w, v >$, where $< , >$ is the usual inner product. In other words, the inner product I is a special 2-tensor. Generalizing this observation, we define an *inner product* on V to be a 2-tensor T such that

(a) T is symmetric, i.e. $T(w, v) = T(v, w)$ for any $w, v \in V$,

(b) T positive-definite, i.e. $T(v, v) > 0$ if $v \neq 0$.

This point of view of inner product is widely used in differential geometry. And this generalization is not too general in the sense that up to an isomorphism T performs like the usual inner product.

Before stating the theorem, we first observe the behavior of tensors under linear transformation of base vector spaces. Let $f : V \to W$ be a linear transformation of two vector spaces. There is a natural linear transformation $f^* : \mathscr{T}^k(W) \to \mathscr{T}^k(V)$ defined by

$$f^*T(v_1, \ldots, v_k) = T(f(v_1), \ldots, f(v_k)),$$

for any $T \in \mathscr{T}^k(W)$ and $v_1, \ldots, v_k \in V$. We call f^* the *pullback* linear transformation. It is clear $f^*(S \times T) = f^*S \otimes f^*T$.

Theorem 4.4. *Let T be an inner product on V. Then there is a basis $b_1, \ldots, b_m$ for V such that $T(b_i, b_j) = \delta_{ij}$. Equivalently, there is an isomorphism $f : \mathbb{R}^m \to V$ such that $f^*T =< , >$, i.e.*

$$f^*T(w, v) = T(f(w), f(v)) =< w, v >,$$

where $< , >$ is the usual inner product on $\mathbb{R}^m$.

Proof. Given any basis $c_1, \ldots, c_m$ for V, we will construct a basis $b_1, \ldots, b_m$ with the desired properties. We first construct inductively a basis $c'_1, \ldots, c'_m$ such that $T(c'_i, c'_j) = 0$ for any $i \neq j$ and then normalized it. Define

$$c'_1 = c_1$$
$$c'_k = c_k - \sum_{i=1}^{k-1} \frac{T(c'_i, c_k)}{T(c'_i, c'_i))} c'_i.$$

It is easy to check that $T(c'_i, c'_j) = 0$ if $i \neq j$ and $T(c'_i, c'_i) \neq 0$. Define $b_i = \frac{c'_i}{T(c'_i, c'_i)}$ which is the basis are required. Since $b_1, \ldots, b_m$ is a basis for V, then there is an isomorphism $f : \mathbb{R}^m \to V$ such that $f(e_i) = b_i$, where $e_1, \ldots, e_m$ is the standard basis for $\mathbb{R}^m$. It is easy to verify that $f^*T =< , >$. Q.E.D

The basis $b_1, \ldots, b_m$ for V such that $T(b_i, b_j) = \delta_{ij}$ is called an *orthonomal* basis with respect to T.

Another interesting tensors with good geometric interpretation, which seems ubiquitous, is the determinant tensor

$$\det_\phi = \sum_{\sigma \in \Sigma_m} \text{sign}(\sigma) \phi_{\sigma(1)} \otimes \cdots \otimes \phi_{\sigma(m)} \in \mathscr{T}^m(\mathbb{R}^m),$$

where Σ_m is the symmetric groups, i.e. the set of all permutations of m distinct elements, and $\phi = \{\phi_1, \dots, \phi_m\}$ is a basis of $V^* = \mathscr{T}^1(V)$.

This tensor $\det_\phi$ is simply the generalization of the usual tensor. Let $v_1 = (v_{11}, \dots, v_{1m}), \dots, v_m = (v_{11}, \dots, v_{1m})$ be column vectors in $\mathbb{R}^m$ and take $e^* = \{e_1^*, \dots, e_i^*\}$ be the usual dual basis. Then $\det_{e^*}$ is exactly the usual determinant of matrices:

$$\begin{aligned}\det_{e^*}(v_1, \dots, v_m) &= \sum_{\sigma \in \Sigma_m} \operatorname{sign}(\sigma) e^*_{\sigma(1)} \otimes \cdots \otimes e^*_{\sigma(m)}(v_1, \dots, v_m) \\ &= \sum_{\sigma \in \Sigma_m} \operatorname{sign}(\sigma) v_{1\sigma(1)} \cdots v_{m\sigma(m)} = \det \begin{pmatrix} v_{11} & \cdots & v_{1m} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mm} \end{pmatrix}.\end{aligned}$$

We will simply denote by det for the tensor $\det_{e^*}$.

Note that if we interchange any vectors v_i and v_j, then the determinant changes the sign, i.e.

$$\det(v_1, \dots, v_i, \dots, v_j, \dots, v_m) = -\det(v_1, \dots, v_j, \dots, v_i, \dots, v_m).$$

This phenomenon is called alternating. Generalizing to k-tensors, we say that a k-tensor $\omega \in \mathscr{T}^k(V)$ is *alternating* if

$$\omega(v_1, \dots, v_i, \dots, v_j, \dots, v_k) = -\omega(v_1, \dots, v_j, \dots, v_i, \dots, v_k)$$

for any $v_1, \dots, v_k \in V$ and any $1 \le i \ne j \le k$.

The set of all alternating k-tensors forms a subspace $\wedge^k V$ of $\mathscr{T}^k(V)$. Is is possible the write a alternating k-tensor as a product of alternating 1-tensors? What is the dimension of $\wedge^k V$. To answer this questions, let's firs consider how to get an alternating tensor from elements in $\mathscr{T}^k V$. Inspired by the example of det, for any $T \in \mathscr{T}^k(V)$, we define a k-tensor $\operatorname{Alt}(T)$ such that

$$\operatorname{Alt}(T)(v_1, \dots, v_k) = \frac{1}{k!} \sum_{\sigma \in \Sigma_k} \operatorname{sign}(\sigma) T(v_{\sigma(1)}, \dots, v_{\sigma(k)})$$

for any k vectors $v_1, \dots, v_k \in V$. For example, $\det = k!\operatorname{Alt}(\phi_1 \otimes \cdots \otimes \phi_m)$. The connivence of dividing $k!$ on the right hand side will be clear as seen in the following theorem.

Theorem 4.5. *Let V be a vector space over $\mathbb{R}$ of dimension m.*

(a) If $T \in \mathscr{T}^k(V)$, then $\mathrm{Alt}(T) \in \wedge^k V$.

(b) If $\omega \in \wedge^k(V)$, then $\mathrm{Alt}(\omega) = \omega$.

Proof. Let $\delta = (i, j)$ be the permutation interchanging i and j and leaving all other numbers fixed.

(a) For any $\sigma \in \Sigma_m$, let $\sigma' = \sigma \cdot \delta$. Then

$$\mathrm{sign}(\sigma') = \mathrm{sign}\delta\mathrm{sign}\sigma = -\mathrm{sign}(\sigma)$$

and

$$\begin{aligned}
&\mathrm{Alt}(T)(v_1, \ldots, v_j, \ldots, v_i, \ldots, v_k)\\
&= \frac{1}{k!} \sum_{\sigma \in \Sigma_k} \mathrm{sign}(\sigma) T(v_{\sigma(1)}, \ldots, v_{\sigma(j)}, \ldots, v_{\sigma(i)}, \ldots, v_{\sigma(k)})\\
&= \frac{1}{k!} \sum_{\sigma \in \Sigma_k} \mathrm{sign}(\sigma) T(v_{\sigma'(1)}, \ldots, v_{\sigma'(i)}, \ldots, v_{\sigma'(j)}, \ldots, v_{\sigma(k)})\\
&= \frac{1}{k!} \sum_{\sigma' \in \Sigma_k} -\mathrm{sign}(\sigma') T(v_{\sigma'(1)}, \ldots, v_{\sigma'(k)})\\
&= -\mathrm{Alt}(T).
\end{aligned}$$

(b) Let $\omega \in \wedge^k(V)$. For any $\sigma \in \Sigma_k$, we know that

$$\omega(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) = \mathrm{sign}(\sigma)\omega(v_1, \ldots, v_k).$$

Consequently, we see that

$$\begin{aligned}
&\mathrm{Alt}(\omega)(v_1, \ldots, v_k)\\
&= \frac{1}{k!} \sum_{\sigma \in \Sigma_k} \mathrm{sign}(\sigma)\omega(v_{\sigma(1)}, \ldots, v_{\sigma(k)})\\
&= \frac{1}{k!} \sum_{\sigma \in \Sigma_k} \mathrm{sign}(\sigma)(\mathrm{sign}(\sigma)\omega(v_1, \ldots, v_k))\\
&= \frac{1}{k!} k!\omega(v_1, \ldots, v_k)\\
&= \mathrm{Alt}(\omega)(v_1, \ldots, v_k).
\end{aligned}$$

Q.E.D

Applying operator Alt with tensor product, we get a new product called the wedge product. For any two alternating tensors $\omega \in \wedge^k V$ and $\eta \in \wedge^l V$, we define the *wedge product* $\omega \wedge \eta \in \wedge^{k+l}$ by

$$\omega \wedge \eta = \frac{(k+l)!}{k!l!}\mathrm{Alt}(\omega \otimes \eta).$$

The wedge product has similar properties as the tensor product.

Theorem 4.6. *Let ω, η and γ be alternating tensors in $\wedge^k V$, $\wedge^l V$ and $\wedge^n V$ respectively, and $f : V \to W$ be a linear transformation. Then we have*

(a) $(\omega + \eta) \wedge \gamma = \omega \wedge \gamma + \eta \wedge \gamma$.

(b) $\omega \wedge (\eta + \gamma) = \omega \wedge \eta + \omega \wedge \gamma$.

(c) $a\omega \wedge \eta = \omega \wedge a\eta = a(\omega \wedge \eta)$ *for any* $a \in \mathbb{R}$.

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

(e) $f^*(\omega \wedge \eta) = f^*(\omega) \wedge f^*(\eta)$.

(f) $(\omega \wedge \eta) \wedge \gamma = \dfrac{k+l+n}{k!l!n!}\mathrm{Alt}(\omega \otimes \eta \otimes \gamma) = \omega \wedge (\eta \wedge \gamma)$.

Proof. The properties 4.6.(a)-4.6.(e) are easy. We prove the property 4.6.(f). By definition, we see that

$$\begin{aligned}(\omega \wedge \eta) \wedge \gamma &= \frac{k+l+n}{(k+l)!n!}\mathrm{Alt}((\omega \wedge \eta) \otimes \gamma) \\ &= \frac{k+l+n}{k!l!n!}\mathrm{Alt}(\mathrm{Alt}(\omega \otimes \eta) \otimes \gamma).\end{aligned}$$

It then suffices to show that

$$\mathrm{Alt}(\mathrm{Alt}(\omega \otimes \eta) \otimes \gamma) = \mathrm{Alt}(\omega \otimes \eta \otimes \gamma) = \mathrm{Alt}(\omega \otimes \mathrm{Alt}(\eta \otimes \gamma)). \tag{5}$$

Note that $\mathrm{Alt}((\mathrm{Alt}(\omega \otimes \eta) - (\omega \otimes \eta)) = 0$ and $\mathrm{Alt}((\mathrm{Alt}(\eta \otimes \gamma) - (\eta \otimes \gamma)) = 0$. Then the equality (5) follows from the following lemma.

Lemma 4.7. *Let* $S \in \mathscr{T}^k(V)$ *and* $T \in \mathscr{T}^l(V)$ *be tensors such that* $\mathrm{Alt}(S) = 0$*. Then*

$$\mathrm{Alt}(S \otimes T) = \mathrm{Alt}(T \otimes S) = 0.$$

Proof. By definition, we know that

$$\begin{aligned}&(k+l)!\mathrm{Alt}(S \otimes T)(v_1, \ldots, v_k, v_{k+1}, \ldots, v_{k+l}) \\ &= \sum_{\sigma \in \Sigma_{k+l}} \mathrm{sign}(\sigma) S(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) T(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}).\end{aligned}$$

Let $\sigma_0 \in \Sigma_{k+l}$ be an permutation such that $(v_{\sigma_0(k+1)}, \ldots, v_{\sigma_0(k+l)}) = (v_{k+1}, \ldots, v_{k+l})$. Fix a $\delta \in \Sigma_{k+l}$ and denote by G_δ the subset of permutations in Σ_{k+l} such that

$$(v_{\sigma(k+1)}, \ldots, v_{\sigma_0(k+l)}) = v_{\delta(k+1)}, \ldots, v_{\delta(k+l)})$$

for any $\sigma \in G_\delta$. Then $G_{\sigma_0} \cong \Sigma_k$ and $G_\delta = \delta G_{\sigma_0}$. By definition of Alt, we see that

$$\begin{aligned}&\sum_{\sigma \in G_\delta} \mathrm{sign}(\sigma) S(v_{\sigma(1)}, \ldots, v_{\sigma(k)}) T(v_{\sigma(k+1)}, \ldots, v_{\sigma(k+l)}) \\ &= (\mathrm{Alt}(S)(v_1, \ldots, v_k)) T(v_{\delta(k+1)}, \ldots, v_{\delta(k+l)}) = 0.\end{aligned}$$

By this equality, to see that $\mathrm{Alt}(S \otimes T) = 0$, it suffices to show that S_{k+l} breaks into disjoint union of such subset G_δ. We claim that if $G_{\delta_1} \cap G_{\delta_2} \neq \emptyset$ then $G_{\delta_1} = G_{\delta_2}$. Let $\sigma \in G_{\delta_1} \cap G_{\delta_2}$. Since $G_\delta = \delta G_{\sigma_0}$ Then there exists σ_1 and σ_2 in G_{σ_0} such that $\sigma = \sigma_1 \delta_1 = \sigma_2 \delta_2$. Consequently, $\delta_1 = \sigma_1^{-1} \sigma_2 \delta_2 \in G_{\delta_2}$. Therefore, $G_{\delta_1} \subset G_{\delta_2}$ and similarly $G_{\delta_2} \subset G_{\delta_1}$. Q.E.D

Q.E.D

By Theorem 4.6.(f), we can easily prove that the wedge product are generated by wedge of alternating 1-tensors.

Theorem 4.8. *Let* $\{\phi_1, \dots, \phi_m\}$ *be a basis for* $V^* = \mathscr{T}^1(C)$. *Then the set of all wedge products*

$$\phi_{i_1} \wedge \cdots \wedge \phi_{i_k} \quad 1 \le i_1 < \cdots < i_k \le m$$

is a basis for $\wedge^k V$ *which therefore has dimension* $\binom{m}{k} = \frac{n!}{k!(n-k)!}$.

Proof. Similar to the proof of Theorem 4.3 with application of Theorem 4.6.(f). Q.E.D

By the above theorem, we see that the space $\wedge^m V$ of top alternation tensors is generated by the tensor det. As an application, we have the following result.

Theorem 4.9. *Let* $b = \{b_1, \dots, b_m\}$ *be a basis for* V *and* $\omega \in \wedge^m V$. *If* $v_i = \sum_{j=1}^m v_{ij} b_j$, *then*

$$\omega(v_1, \dots, v_m) = \det(v_{ij})\omega(b_1, \dots, b_m).$$

Proof. Let $\det_{b^*}$ be the determinant tensor associate to the dual basis $b^* = \{b_1^*, \dots, b_m^*\}$ such that

$$\det_{b^*}(w_1, \dots, w_m) = \det\left(w_{ij}\right)$$

for any vectors $w_1 = \sum_{j=1}^m w_{1j} b_j, \dots, w_m = \sum_{j=1}^m w_{mj} b_j$. By Theorem 4.8, there is a number $a \in \mathbb{R}$ such that $\omega = a \det$. Since

$$\omega(b_1, \dots, b_m) = a \det_{b^*}(b_1, \dots, b_m) = a,$$

then $\omega(v_1, \dots, v_m) = \det\left(v_{ij}\right) \omega(b_1, \dots, b_m)$. Q.E.D

The det tensor is geometrically very significant. By Theorem 4.9, the a non-zero element $\omega \in \wedge^m V$ splits the set $\mathscr{B}_V = \{\{b_1, \dots, b_m\}\}$ of bases of V into two subsets

$$\omega_+ = \{\{b_1, \dots, b_m\} \mid \omega(b_1, \dots, b_m) > 0, \{b_1, \dots, b_m\} \in \mathscr{B}_V\}$$

$$\omega_- = \{\{b_1, \dots, b_m\} \mid \omega(b_1, \dots, b_m) < 0, \{b_1, \dots, b_m\} \in \mathscr{B}_V\}$$

Moreover, the splitting is independent of choice of ω, i.e. if $\omega' = \lambda\omega$, then $\omega_+ = \omega'_+$ if and only if $\lambda > 0$.

Either ω_+ or ω_- is called an *orientation* for V. The orientation to which a basis $\{b_1, \dots, b_m\}$ belongs is denoted by $[b_1, \dots, b_m]$ and the other orientation is denoted by $-[b_1, \dots, b_m]$. The *usual orientation* for $\mathbb{R}^n$ is defined as $[e_1, \dots, e_m]$.

For an orientation $\mu = [b_1, \ldots, b_m]$ associated to a orthonormal basis $b = \{b_1, \ldots, b_m\}$, there is an unique alternating m-tensor ω such that $\omega(v_1, \ldots, v_m) = 1$ for any orthonormal basis $v = \{v_1, \ldots, v_m\}$ such that $[v_1, \ldots, v_m] = \mu$. This unique ω is called the *volume element* of V.

4.2 Differential forms

Having discussed the algebra preliminaries of wedge product, now we return to discuss the more geometric object, differential forms. We understand the differential forms are viewed as functions on the "tangent bundles". So we start define tangent bundles.

Definition 4.10. For any point $p \in \mathbb{R}^m$ and a vector $v \in \mathbb{R}^m$, the pair $\{(p; v) \mid v \in \mathbb{R}\}$ is called a *tangent vector* of $\mathbb{R}^m$ at p. The set of all tangent vectors at $p \in \mathbb{R}^m$ forms a vector space defined by

$$(p; v_1) + (p; v_2) = (p; v_1 + v_2) \quad a(p; v) = (p; av), a \in \mathbb{R}.$$

We call this vector space the *tangent space* to $\mathbb{R}^m$ at p and denoted by $T_p\mathbb{R}^m$. The tangent bundle of $\mathbb{R}^m$, denoted by $T\mathbb{R}^m$ is the union $\cup_{p\in\mathbb{R}^m} T_p\mathbb{R}^m$ of the tangent spaces

Although the vector space $T_p\mathbb{R}^m$ has the exactly same structure at $\mathbb{R}^m$, geometrically, we show think of $T_p\mathbb{R}^m$ as the shift of $T_o\mathbb{R}^m = \mathbb{R}^m$, i.e. we view $(p; v)$ as a vector initiated at p and simply write $(p; v)$ as v_p.

Playing with the tangent space $T\mathbb{R}^m$, we can define different kind of functions from it or to it.

Definition 4.11. A *tangent vector field* on $\mathbb{R}^m$ is a function $F : \mathbb{R}^m \to T\mathbb{R}^m$ such that $F(p)$ is a tangent vector at p for any $p \in \mathbb{R}^m$.

By choosing a "compatible" basis, we may define k-tensors on $T\mathbb{R}^m$. In particular, let $e_1, \ldots, e_m$ be the usual basis for $\mathbb{R}^m$. Then the basis $(e_1)_p, \ldots, (e_m)_p$, called the usual basis of $T_p\mathbb{R}^m$, changes continuous as a function of p. More than that, it changes rigidly as p changes. This is the "compatibility" I am talking about. We will discuss this idea later when we talk about manifolds.

Definition 4.12. For each $i = 1, \ldots, m$, we define a differential 1-form on $\mathbb{R}^m$, $\phi_i : T\mathbb{R}^m \to \mathbb{R}$ such that $\phi_i(p; e_j) = \delta_{ij}$. The forms $\phi_1, \ldots, \phi_m$ are called the elementary 1-forms on $\mathbb{R}^m$.

Notice that an elementary differential 1-form on $\mathbb{R}^m$ is equivalent to a function $\mathbb{R}^m \to \mathscr{T}^1(T\mathbb{R}^m)$.

Definition 4.13. A *differential k-form*, or simply *differential form*, ω on $\mathbb{R}^m$ is a function such that

$$\omega(p) = \sum_{1 \le i_1 < \cdots < i_k \le m} w_{i_1,\ldots,i_k}(p)(\phi_{i_1}(p) \wedge \cdots \phi_{i_k}(p)),$$

where $w_{i_1,\ldots,i_k}(p)$ are continuous functions, $\phi_1(p), \ldots, \cdots \phi_m(p))$ is the dual basis to $(e_1)_p$, $\ldots$, $(e_m)_p$. A 0-form on $\mathbb{R}^m$ is a function. The k-form ω is called a C^r form if $w_{i_1,\ldots,i_k}(p)$ are C^r functions. The sum $\omega + \eta$, product $f\omega$ and wedge product $\omega \wedge \eta$ are defined in the same way.

To simplify discussion of differentiation, from now on, we will assume that forms and vector fields are all smooth, i.e. the functions are C^∞.

Definition 4.14. Let $f : \mathbb{R}^m \to \mathbb{R}$ be a smooth function. We define a 1-form $\mathrm{d}f$, called the differential of f, by

$$\mathrm{d}f(p)(v_p) = Df(p) \cdot v_p, \tag{6}$$

for any $v_p \in T_p\mathbb{R}^m$.

In particular, considering the i-th component projection $x_i : \mathbb{R}^m \to \mathbb{R}$, we see that $\mathrm{d}x_1(p), \ldots, \mathrm{d}x_m(p)$ is just the dual basis of $(e_1)_p, \ldots, (e_m)_p$. Therefore, a k-form ω can be written as

$$\omega = \sum_{1 \le i_1 < \cdots < i_k \le m} w_{i_1,\ldots,i_k}(\mathrm{d}x_{i_1} \wedge \cdots \mathrm{d}x_{i_k}).$$

In other words, $\mathrm{d}x_1(p), \ldots, \mathrm{d}x_m(p)$ are the elementary 1-forms.

From the definition of differential of a 0-form, we can easily prove the following theorem.

Theorem 4.15. *Let $f : \mathbb{R}^m \to \mathbb{R}$ be a smooth function. Then*

$$\mathrm{d}f = D_1 f \mathrm{d}x^1 + \cdots + D_m f \mathrm{d}x^m.$$

In Leibniz notation,

$$\mathrm{d}f = \frac{\partial f}{\partial x_1}\mathrm{d}x^1 + \cdots + \frac{\partial f}{\partial x_1}\mathrm{d}x^m$$

The way of writing the same index as sup-index and sub-index is commonly used in differential geometry. Differential geometers write $f_I \mathrm{d}x^I$ for $\sum_I f_I \mathrm{d}x^I$, where $I = (i_1, \ldots, i_k)$ are vectors of numbers $1 \leq i_j \leq m$, $\mathrm{d}x^I = \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}$ and $f_I = f_{i_1 \cdots i_k}$. The notation $f_I \mathrm{d}x^I = \sum_I f_I \mathrm{d}x^I$ is called the Einstein summation convention.

How should we define differential of k-forms? Recall that a differential k-form ω can be considered as a function $\mathbb{R}^m \to \wedge^k(\mathbb{R}^m) \cong \mathbb{R}^{\binom{m}{k}}$ given by

$$\omega(p) = (w_{1,\ldots,k}(p), \ldots, w_{m-k,\ldots,m}(p)).$$

In this sense, we can take differentials of each $w_{i_1,\ldots,i_k}$ which suggests the following definition of differential of k-forms.

Definition 4.16. Let

$$\omega = \sum_{1 \leq i_1 < \cdots < i_k \leq m} w_{i_1,\ldots,i_k}(\mathrm{d}x^{i_1} \wedge \cdots \mathrm{d}x^{i_k})$$

be a differential k-form on $\mathbb{R}^m$, we define a $(k+1)$-form $\mathrm{d}\omega$, called the *differential of the k-form* ω, by

$$\mathrm{d}\omega = \sum_{1 \leq i_1 < \cdots < i_k \leq m} \sum_{\alpha=1}^{m} \frac{\partial w_{i_1,\ldots,i_k}}{\partial x_\alpha} \mathrm{d}x^\alpha \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_k}.$$

We denote by $\Omega^k(\mathbb{R}^m)$ the set of all differential k-forms on $\mathbb{R}^m$.

Similar to the set of alternating k-tensors, the set $\Omega^k(\mathbb{R}^m)$ has a natural vector bundle structure. We will call $\Omega^k(\mathbb{R}^m)$ the space of k-forms on $\mathbb{R}^m$.

Theorem 4.17. *Let* $\mathrm{d} : \Omega^k(\mathbb{R}^m) \to \Omega^{k+1}(\mathbb{R}^m)$ *be a map defined by* $\mathrm{d}(\omega) = \mathrm{d}\omega$.

(a) (Linearity) For any two k-forms ω and η, and real number a and b, $\mathrm{d}(a\omega + b\eta) = a\mathrm{d}\omega + b\mathrm{d}\eta$.

(b) (Leibniz rule) If ω is a k-form and η is an l-form, then

$$\mathrm{d}(\omega \wedge \eta) = \mathrm{d}\omega \wedge \eta + (-1)^k \omega \wedge \mathrm{d}\eta.$$

(c) ($d^2 = 0$) For any k-form ω, $\mathrm{d}(\mathrm{d}\omega) = 0$.

Proof. The conclusion (4.17.(a)) follows from definition.

To prove (4.17.(b)), again by linearity, it suffices to show that (4.17.(b)) holds for $\omega = f_I \mathrm{d}x^I$ and $\eta = g_J \mathrm{d}x^J$. Note that (4.17.(b)) holds for 0-forms, since $\mathrm{d}(fg) = (\mathrm{d}f)g + f(\mathrm{d}g)$ by Leibniz rule for functions.

Therefore, it suffices to show that (4.17.(b)) holds for $\omega = \mathrm{d}x^I$ and $\eta = \mathrm{d}x^J$. This is obvious, since by definition

$$0 = \mathrm{d}(\mathrm{d}x^I \wedge \mathrm{d}x^J) = 0 + 0 = \mathrm{d}\omega \wedge \eta + (-1)^k \omega \wedge \mathrm{d}\eta.$$

To prove (4.17.(c)), by linearity, it suffices to show that $\mathrm{d}(\mathrm{d}\omega) = 0$ for $\omega = f_I \mathrm{d}x^I$. By definition,

$$\begin{aligned}
\mathrm{d}(\mathrm{d}(f_I \mathrm{d}x^I)) &= \mathrm{d}(\sum_{\alpha=1}^{m} \frac{\partial f_I}{\partial x_\alpha} \mathrm{d}x^\alpha \wedge \mathrm{d}x^I) \\
&= \sum_{\beta=1}^{m} \sum_{\alpha=1}^{m} \frac{\partial^2 f_I}{\partial x_\alpha \partial x_\beta} \mathrm{d}x^\beta \wedge \mathrm{d}x^\alpha \wedge \mathrm{d}x^I) \\
&= \sum_{\alpha=1}^{m} \frac{\partial^2 f_I}{\partial x_\alpha \partial x_\alpha} (\mathrm{d}x^\alpha \wedge \mathrm{d}x^\alpha) \wedge \mathrm{d}x^I) \\
&\quad + \sum_{1 \leq \alpha < \beta \leq m}^{m} (\frac{\partial^2 f_I}{\partial x_\alpha \partial x_\beta} - \frac{\partial^2 f_I}{\partial x_\beta \partial x_\alpha}) \mathrm{d}x^\beta \wedge \mathrm{d}x^\alpha) \wedge \mathrm{d}x^I \\
&= 0
\end{aligned}$$

In the second last equality, we use the facts that $\mathrm{d}x^\alpha \wedge \mathrm{d}x^\alpha = 0$ and $\frac{\partial^2 f_I}{\partial x_\alpha \partial x_\beta} = \frac{\partial^2 f_I}{\partial x_\beta \partial x_\alpha}$.

Q.E.D

By Theorem 4.17.(a), we see that $d : \Omega^k(\mathbb{R}^m) \to \Omega^{k+1}(\mathbb{R}^m)$ is a linear transformation and called a differential operator.

Remark 4.18. The differential operator d defined in Definition 4.16 is the only linear transformation $\Omega^k(\mathbb{R}^m) \to \Omega^{k+1}(\mathbb{R}^m)$ that satisfies 4.17.(b), 4.17.(c) and the equality (6).

From 4.17.(c), we see that $\mathrm{d}\omega = 0$ if $\omega = \mathrm{d}\eta$. Motivating by this fact, we define closed forms and exact forms.

Definition 4.19. A k-form ω is called a *closed form* if $\mathrm{d}\omega = 0$. It is called a *exact form* if $\omega = \mathrm{d}\eta$ for some $(k-1)$-form.

Using this definition, by 4.17.(c), we can see that any exact form on $\mathbb{R}^m$ is closed. The converse is not true in general. The following example shows that not every closed form is exact.

Example 4.20. Let

$$\omega = \frac{-y}{x^2+y^2}\mathrm{d}x + \frac{x}{x^2+y^2}\mathrm{d}y$$

be a 1-form defined over $\mathbb{R}^2 \setminus \{0\}$. It is easy to check that $\mathrm{d}\omega = 0$. However, if $\omega = \mathrm{d}f$ for some smooth function f, then

$$\frac{-y}{x^2+y^2} = \frac{\partial f}{\partial x} \qquad \frac{x}{x^2+y^2} = \frac{\partial f}{\partial y}.$$

Since f is continuous, then $\frac{\partial^2 f}{\partial y \partial x} = \frac{\partial^2 f}{\partial x \partial y}$. However,

$$\frac{\partial}{\partial y}\left(\frac{-y}{x^2+y^2}\right) = \frac{x^2-y^2}{(x^2+y^2)^2} \neq \frac{-x^2+y^2}{(x^2+y^2)^2} = \frac{\partial}{\partial x}\left(\frac{-y}{x^2+y^2}\right).$$

The failure of closed forms being exact depends on the "shape" of the domain where the differential form is defined. This observation leads to a very important topic in modern mathematics, called de Rham cohomology. Using d, we may define a sequence

$$\Omega^0 \xrightarrow{\mathrm{d}} \Omega^1 \xrightarrow{\mathrm{d}} \cdots \xrightarrow{\mathrm{d}} \Omega^n$$

called de Rham complex

On $\mathbb{R}^3$, the de Rham complex is closely related to the div-grad-curl sequence, namely, we have the following commutative diagram

$$\begin{array}{ccccccc}
\Omega^0 & \xrightarrow{\text{grad}} & \mathfrak{X} & \xrightarrow{\text{curl}} & \mathfrak{X} & \xrightarrow{\text{div}} & \Omega^0 \\
{\scriptstyle i}\downarrow\,\| & & \downarrow{\scriptstyle m} & & \downarrow{\scriptstyle c} & & \downarrow{\scriptstyle v} \\
\Omega^0 & \xrightarrow{\mathrm{d}} & \Omega^1 & \xrightarrow{\mathrm{d}} & \Omega^2 & \xrightarrow{\mathrm{d}} & \Omega^3
\end{array}$$

where $\mathfrak{X}$ is the set of tangent fields, m, c and v are isomorphisms given by

$$m(\sum_{i=1}^{3} g_i \frac{\partial}{\partial x_i}) = \sum_{i=1}^{3} g_i \mathrm{d}x^i,$$

$$c(\sum_{i=1}^{3} g_i \frac{\partial}{\partial x_i}) = g_2 \mathrm{d}x^2 \wedge \mathrm{d}x^3 - g_1 \mathrm{d}x^1 \wedge \mathrm{d}x^3 + g_3 \mathrm{d}x^1 \wedge \mathrm{d}x^2,$$

$$v(f) = f \mathrm{d}x^1 \wedge \mathrm{d}x^2 \wedge \mathrm{d}x^3.$$

Recall that

$$\mathrm{grad}(g) = \sum_{i=1}^{3} \frac{\partial g}{\partial x_i} \frac{\partial}{\partial x_i},$$

$$\mathrm{curl}(\sum_{i=1}^{3} g_i \frac{\partial}{\partial x_i}) = (\frac{\partial g_3}{\partial x_2} - \frac{\partial g_2}{\partial x_3})\frac{\partial}{\partial x_1} - (\frac{\partial g_3}{\partial x_1} - \frac{\partial g_1}{\partial x_3})\frac{\partial}{\partial x_2} + (\frac{\partial g_2}{\partial x_1} - \frac{\partial g_1}{\partial x_2})\frac{\partial}{\partial x_3},$$

$$\mathrm{div}(\sum_{i=1}^{3} g_i \frac{\partial}{\partial x_i}) = \sum_{i=1}^{3} \frac{\partial g_i}{\partial x_i}.$$

4.3 Pullback of a differential form

Let $f = (f_1, \ldots, f_n) : \mathbb{R}^m \to \mathbb{R}^n$ be a differentiable function. Motivated by the derivative Df and that $\mathrm{d}f := (\mathrm{d}f_1, \ldots, \mathrm{d}f_n) = Df \cdot (\mathrm{d}x^1, \ldots, \mathrm{d}x^m)$, we can define a map $f_* : T\mathbb{R}^m \to T\mathbb{R}^n$ between the tangent spaces by

$$f_*(p; v) = (f(p), Df(p)(v)),$$

for any vector filed v on $\mathbb{R}^m$. It can be check that f_* is a linear transformation, called the pushforward map by f.

By taking dual, we define another map $f^* : \Omega^k(\mathbb{R}^n) \to \Omega^k(\mathbb{R}^m)$, called the pullback map by f, such that

$$f^*\omega((v_1)_p, \ldots, (v_k)_p) = \omega(f_*(v_1)_p, \ldots, f_*(v_1)_p)$$

for any $\omega \in \Omega^k(\mathbb{R}^n)$ any k tangent vectors $(v_1)_p, \ldots, (v_k)_p$ in $T\mathbb{R}^m$.

Again, the pullback map f^* is a linear transformation. Moreover, it commutes with wedge products and the differential operator d.

Theorem 4.21. *Let $f = (f_1, \dots, f_n) : \mathbb{R}^m \to \mathbb{R}^n$ be a smooth map. Then*

(a) $f^*\mathrm{d}y^i = \mathrm{d}f_i = \sum_{j=1}^m \frac{\partial f_i}{\partial x_j}\mathrm{d}x^j$.

(b) $f^*(\omega + \eta) = f^*\omega + f^*\eta$.

(c) $f^*(g \cdot \omega) = (g \circ f) \cdot f^*\omega$.

(d) $f^*(\omega \wedge \eta) = f^*\omega \wedge f^*\eta$. *In particular, if $m = n$, then*

$$f^*(g\mathrm{d}y^1 \wedge \cdots \wedge \mathrm{d}y^n) = (g \circ f)(\det Df)\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n.$$

(e) $f^*(\mathrm{d}\omega) = \mathrm{d}f^*\omega$.

Proof. For vector field v on $\mathbb{R}^m$ by definition of f^*,

$$f^*\mathrm{d}y^i(v) = \mathrm{d}y^i(\mathrm{d}f(v)) = \mathrm{d}y^i(\mathrm{d}f_1(v), \dots, \mathrm{d}f_m(v)) = \mathrm{d}f_i(v).$$

The statement 4.21.(b) and the first part of 4.21.(c) follow easily from the linearity of differential forms. The second part of 4.21.(c) follows from Theorem 4.9 and the definition of f_*.

The fourth statement 4.21.(d) follows from the definitions of pullback and wedge product.

Now we prove 4.21.(e). By 4.17.(a) and 4.21.(c), it suffices to prove 4.21.(e) for $\omega = \mathrm{d}y^I$. Since $\mathrm{d}^2 = 0$, then $f^*\mathrm{d}(\mathrm{d}y^I) = 0$. We need to show that $\mathrm{d}(f^*\mathrm{d}y^I) = 0$. We may assume that $I = (1, \dots, k)$. By 4.21.(d) and 4.21.(a), we have

$$\mathrm{d}f^*\mathrm{d}y^I = \mathrm{d}(\mathrm{d}f_1 \wedge \cdots \wedge \mathrm{d}f_k).$$

Since $\mathrm{d}^2 = 0$, then $\mathrm{d}(\mathrm{d}f_i) = 0$. Then by applying the Leibniz rule 4.17.(b) repeatedly, we know that

$$\mathrm{d}(f^*\mathrm{d}y^I) = \mathrm{d}(\mathrm{d}f_1 \wedge \cdots \wedge \mathrm{d}f_k) = 0.$$

Therefore, $f^*\mathrm{d} = \mathrm{d}f^*$. Q.E.D

4.4 Integration on chains

This section is a bridge to differential manifolds. We will discuss Stokes' Theorem and Poincare's Lemma. We discussed differential form in order to generalize the Fundamental Theorem of Calculus to multivariable functions. For this purpose, we need to generalized integrals and its boundary which consists points to higher dimension.

Definition 4.22. A *singular n-cell* in a set $D \subset \mathbb{R}^n$ is a continuous function $c : [0,1]^n \to D$. A finite formal sum $\sum_{i=1}^{r} a_i c_i$ of singular n-cells $c_1, \ldots, c_r$ with integer coefficients is called an *n-chain*. We let $\mathbb{R}^0$ and $[0,1]^0$ both denote the 0-cell $\{0\}$.

For example, the identity function $I^n : [0,1]^n \to \mathbb{R}^n$ given by $I^n(x) = x$ is a singular n-cell, called the standard n-cell.

Singular cells are fundamental concept in algebraic topology. They are used to define homology groups which are topological invariants of topological spaces.

The way a n-chain is defined suggests that we can add two n-chains and multiply an n-chain by a number. For example,

$$2(c_1 + c_2 - 3c_3) - 3(c_2 - c_1 - c_4) = 5c_1 - c_2 - 6c_3 + 3c_4.$$

Why should we care about the n-chains? Does the negative sign has a geometric meaning?

Consider the interval $[0,1]$, as a topological space, its boundary is a set consists two points 0 and 1. For a continuous function $f(x)$ on $\mathbb{R}$, we know that

$$\int_0^1 f \, \mathrm{d}x = -\int_1^0 f \, \mathrm{d}x.$$

In terms of 1-cells, we can write $\int_I f \, \mathrm{d}x = \int_0^1 f \, \mathrm{d}x$ and $\int_1^0 f \, \mathrm{d}x = \int_{-I} f \, \mathrm{d}x$. Then the equality $\int_0^1 f \, \mathrm{d}x = -\int_1^0 f \, \mathrm{d}x$ is equivalent to $\int_{-I} f \, \mathrm{d}x = -\int_I f \, \mathrm{d}x$. In other words, the negative sign can be viewed as the opposite of an orientation.

In fact, this convention of negative sign is important in the generalization of the fundamental theorem of calculus in single variable.

For each n-cell c in a set D, we can define $(n-1)$-chain ∂c called the boundary in the following way. We first consider the standard n-cell of I^n. For each $1 \leq i \leq n$, we define two singular $(n-1)$-cells $I^n_{(i,0)}$ and $I^n_{(i,1)}$ associated to I^n. For each $x \in [0,1]^{n-1}$, we define

$$I^n_{(i;0)}(x) = I^n(x_1, \ldots, x_{i-1}, 0, x_i, \ldots, x_{n-1})$$

and

$$I^n_{(i;1)}(x) = I^n(x_1, \ldots, x_{i-1}, 1, x_i, \ldots, x_{n-1}).$$

We call $I^n_{(i;0)}$ the $(i;0)$-face and $I^n_{(i;1)}$ the $(i;1)$-face of I^n. The boundary of the standard n-cell I^n is defined as

$$\partial I^n = \sum_{i=1}^{n} \sum_{a=0,1} (-1)^{i+1} I^n_{(i;a)}.$$

For a general singular n-cell $c : [0,1]^n \to D$, we define the $(i;a)$-face $c_{(i;a)}$ of c as $c_{(i;a)} = c \circ (I^n_{(i;a)})$ and the boundary ∂c of c as

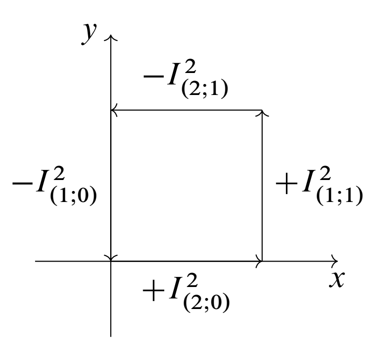
$$\partial c = \sum_{i=1}^{n} \sum_{a=0,1} (-1)^{i+1} c_{(i;a)}.$$

Generalizing to n-chains, we define the boundary $\partial(\sigma)$ of a n-chain $\sigma = \sum a_i c_i$ as

$$\partial \sigma = \sum a_i \partial(c_i).$$

For example, the boundary of $I^2 = [0,1]^2$ is

$$\partial I^2 = I^2_{(2;0)} + I^2_{(1;1)} - I^2_{(2;1)} - I^2_{(1;0)}$$



If we view those faces as unit vectors, we can conclude that the sum is 0. Because an initial point is also an end point. In terms of singular cell, this is equivalent to say that

$$\begin{aligned}\partial(\partial I^2) &= \partial(I^2_{(2;0)} + I^2_{(1;1)} - I^2_{(2;1)} - I^2_{(1;0)}) \\ &= [\{0\} \times \{0\} - \{1\} \times \{0\}] + [\{1\} \times \{0\} - \{1\} \times \{1\}] \\ &\quad - [\{1\} \times \{0\} - \{1\} \times \{1\}] - [\{0\} \times \{0\} - \{0\} \times \{1\}] \\ &= 0\end{aligned}$$

In general, by definition, for any $x \in [0,1]^{n-2}$ and $1 \leq i \leq j \leq n-1$, the $(n-2)$-cell $(I^n_{(i;a)})_{(j;b)}$ is defined by

$$\begin{aligned}(I^n_{(i;a)})_{(j;b)}(x) &= I^n_{(i;a)}(I^{(n-1)}_{(j;b)}(x)) \\ &= I^n_{(i;a)}(x_1, \ldots, \ldots, x_{j-1}, b, x_j, \ldots, x_{n-2}) \\ &= (x_1, \ldots, x_{i-1}, a, x_i, \ldots, x_{j-1}, b, x_{j+1}, \ldots, x_{n-2}).\end{aligned}$$

Equivalently, the function $(I^n_{(i;a)})_{(j;b)}$ first inserts a between the $(i-1)$-th and i-th components of x to get a point, say x_a, in $[0,1]^{n-1}$, and then inserts b between the (j)-th and $j+1$-th components of x_a to get a point, say x_{ab}, in $[0,1]^n$. In other words,

$$(I^n_{(i;a)})_{(j;b)} = (I^n_{(j+1;b)})_{(i;a)}.$$

In particular, if $i = j$, then we have

$$(I^n_{(i;a)})_{(i;b)}(x) = (x_1, \ldots, x_{i-1}, a, b, x_i, \ldots, x_{n-1}).$$

If $i > j$, then $(I^n_{(i;a)})_{(j;b)} = (I^n_{(j;b)})_{(i-1;a)}$.

More generally, we have the following theorem.

Theorem 4.23 ($\partial^2 = 0$)**.** *For any n-chain c in $D \subset \mathbb{R}^m$, we have $\partial(\partial(c)) = 0$.*

Proof. For any $1 \leq i \leq j \leq n-1$, we can check that

$$(I^n_{(i;a)})_{(j;b)} = (I^n_{(j+1;b)})_{(i;a)}.$$

Moreover,

By definition, for any singular n-cell c, we have $(c_{(i;a)})_{(j;b)} = (c_{(j+1;b)})_{(i;a)}$. It then follows that

$$
\begin{aligned}
\partial(\partial c) &= \partial \sum_{i=1}^{n} \sum_{a=0,1} (-1)^{i+a} c_{(i;a)} \\
&= \sum_{i=1}^{n} \sum_{a=0,1} (-a)^{i+a} \partial c_{(i;a)} \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n-1} \sum_{a,b=0,1} (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} \\
&= \sum_{a,b=0,1} \left(\sum_{1 \le j < i \le n} (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} + \sum_{1 \le i \le j \le n-1} (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} \right) \\
&= \sum_{a,b=0,1} \left(\sum_{1 \le j < i \le n} (-1)^{i+j+a+b} (c_{(j;b)})_{(i-1;a)} + \sum_{1 \le i \le j \le n-1} (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} \right) \\
&= \sum_{a,b=0,1} \left(\sum_{1 \le j \le k \le n-1} (-1)^{k+1+j+a+b} (c_{(j;b)})_{(k;a)} + \sum_{1 \le i \le j \le n-1} (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} \right) \\
&= \sum_{1 \le i \le j \le n-1} \left(\sum_{a,b=0,1} \left(-1)^{i+1+j+a+b} (c_{(i;b)})_{(j;a)} + (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} \right) \right) \\
&= \sum_{1 \le i \le j \le n-1} \left(- \sum_{a,b=0,1} (-1)^{i+j+a+b} (c_{(i;b)})_{(j;a)} + \sum_{a,b=0,1} (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} \right) \\
&= \sum_{1 \le i \le j \le n-1} \left(- \sum_{a,b=0,1} (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} + \sum_{a,b=0,1} (-1)^{i+j+a+b} (c_{(i;a)})_{(j;b)} \right) \\
&= 0.
\end{aligned}
$$

Q.E.D

Recall that the differential operator d also has the property that $\mathrm{d}^2 = 0$. In fact, d and ∂ are both certain kind of boundary maps and related to each other. More precisely, there are connected in the sense of duality given by integration.

Let ω be a k-form on the k-cell $[0,1]^k$. Then $\omega = f\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k$ for a unique function f. We define the integral of ω over $[0,1]^k$ to be

$$\int_{[0,1]^k} \omega = \int_{[0,1]^k} f.$$

Note that if we change the order of the coordinates $x^1, \ldots, x^k$, the integral $\int_{[0,1]^k} \omega$ may change by a negative sign. This is because that the function f would change to $-f$.

With the usual order, we could also write

$$\int_{[0,1]^k} f\,\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^k = \int_{[0,1]^n} f\,\mathrm{d}x^1 \cdots \mathrm{d}x^k$$

as we usually do in multivariable calculus.

Using pullback of differential forms, we define the *integral of a k-form* $\omega = \sum_I f_I \,\mathrm{d}x^I$ on a singular k-cell c in D to be

$$\int_c \omega = \int_{[0,1]^k} c^*\omega.$$

For consistency, we define the integral of a 0-form ω on a 0-cell c as $\int_c \omega = \omega(c(o))$.

Generalize to k-chains, we define the integral of a k-form ω and a k-chain $\sigma = \sum a_i c_i$ in D to be

$$\int_\sigma \omega = \sum a_i \int_{c_i} \omega.$$

The integral of a 1-form (2-form) on an 1-chain (a 2-chain) is often called a *line integral* (*surface integral*).

Remark 4.24. Let $P\,\mathrm{d}x + Q\,\mathrm{d}y$ be a 1-form on a singular 1-cell in $\mathbb{R}^2$, one can prove that

$$\int_c P\,\mathrm{d}x + Q\,\mathrm{d}y = \lim \sum_{i=0}^{n} (c^1(t_i) - c^1(t_{i-1})P(c(\xi_i))) + (c^2(t_i) - c^2(t_{i-1})Q(c(\xi_i)))$$

where $t_0, \ldots, t_n$ is a partial of $[0,1]$, $\xi_i \in [t_{i-1}, t_i]$ is an arbitrary number, and the limit is take over all partitions. The right hand sides is often taken as the definition of line integral.

Remark 4.25. In our definition of integral of a k-form over a k-chain, we implicitly used the change of variable theorem.

Example 4.26. Let's find the integral of the 1-form $\omega = x\mathrm{d}y$ on the 1-chain ∂I^2 in $\mathbb{R}^2$. Recall that $I^2_{(1;a)}(t) = (a, t)$ and $I^2_{(2;a)}(t) = (t, a)$, where $t \in [0, 1]$. Then

$$I^2_{(2;a)}{}^* x\mathrm{d}y = t\mathrm{d}a = 0 \quad \text{and} \quad I^2_{(1;a)}{}^* x\mathrm{d}y = a\mathrm{d}t.$$

Therefore,

$$\begin{aligned}
\int_{\partial I^2} x\mathrm{d}y &= \int_{I^2_{(2;0)}} x\mathrm{d}y + \int_{I^2_{(1;1)}} x\mathrm{d}y - \int_{I^2_{(2;1)}} x\mathrm{d}y - \int_{I^2_{(1;0)}} x\mathrm{d}y \\
&= \int_{[0,1]} 0 + \int_{[0,1]} \mathrm{d}t - \int_{[0,1]} 0 - \int_{[0,1]} 0 \\
&= 1
\end{aligned}$$

Note that if we integrate $\mathrm{d}(x\mathrm{d}y)$ over I^2, we get

$$\int_{I^2} \mathrm{d}(x\mathrm{d}y) = \int_{[0,1]^2} \mathrm{d}(x\mathrm{d}y) = \int_{[0,1]^2} \mathrm{d}x \wedge \mathrm{d}y = 1.$$

This is not an coincidence. You can also try $y\mathrm{d}x$, then you will have

$$\int_{\partial I^2} y\mathrm{d}x = -1 = \int_{I^2} \mathrm{d}(y\mathrm{d}x).$$

In fact, those are very special cases of Stoke's theorem, a.k.a. fundamental theorem of calculus in higher dimension.

Theorem 4.27 (Stokes' Theorem)**.** *Let ω be a $(k-1)$-form on an open set $D \in \mathbb{R}^n$ and c be a k-chain in D. Then*

$$\int_{\partial c} \omega = \int_c \mathrm{d}\omega.$$

Proof. By definition of integrals of k-forms on k-chains and linearity of integration and differential forms, we may assume that c is the standard singular k-cell I^k and ω is a $(k-1)$-form of the type

$$f\mathrm{d}x^1 \wedge \cdots \wedge \widehat{\mathrm{d}x^i} \wedge \cdots \wedge \mathrm{d}x^k,$$

where $\widehat{\mathrm{d}x^i}$ means removing $\mathrm{d}x^i$ from the wedge product and f is a C^∞ function on $[0, 1]^k$.

As we have seen in Example 4.26, we know that

$$I^{k}_{(j;a)}{}^{*}(\omega) = \begin{cases} f(x_1,\dots,a,\dots x_k)\mathrm{d}x^1\wedge\cdots\wedge\widehat{\mathrm{d}x^i}\wedge\cdots\wedge\mathrm{d}x^k & i=j \\ 0 & i\neq j. \end{cases}$$

Then

$$\begin{aligned}
&\int_{\partial I^k} f\mathrm{d}x^1\wedge\cdots\wedge\widehat{\mathrm{d}x^i}\wedge\cdots\wedge\mathrm{d}x^k) \\
=&(-1)^i\int_{[0,1]^{k-1}} f(x_1,\dots,0,\dots x_k)\mathrm{d}x^1\wedge\cdots\wedge\widehat{\mathrm{d}x^i}\wedge\cdots\wedge\mathrm{d}x^k) \\
&+(-1)^{i+1}\int_{[0,1]^{k-1}} f(x_1,\dots,1,\dots x_k)\mathrm{d}x^1\wedge\cdots\wedge\widehat{\mathrm{d}x^i}\wedge\cdots\wedge\mathrm{d}x^k).
\end{aligned}$$

By Fubini's theorem and fundamental theorem of calculus in dimension 1, we have

$$\begin{aligned}
&\int_{I^k}\mathrm{d}(f\mathrm{d}x^1\wedge\cdots\wedge\widehat{\mathrm{d}x^i}\wedge\cdots\wedge\mathrm{d}x^k) \\
=&\int_{I^k}\frac{\partial f}{\partial x_i}\mathrm{d}x^i\wedge\mathrm{d}x^1\wedge\cdots\wedge\widehat{\mathrm{d}x^i}\wedge\cdots\wedge\mathrm{d}x^k) = (-1)^{i-1}\int_{I^k}\frac{\partial f}{\partial x_i}\mathrm{d}x^1\wedge\cdots\wedge\mathrm{d}x^k \\
=&(-1)^{i-1}\int_{I^k}\frac{\partial f}{\partial x_i}\mathrm{d}x^1\cdots\mathrm{d}x^k = (-1)^{i-1}\int_{I^{k-1}}\left(\int_0^1\frac{\partial f}{\partial x_i}\mathrm{d}x^i\right)\mathrm{d}x^1\cdots\widehat{\mathrm{d}x^i}\cdots\mathrm{d}x^k) \\
=&(-1)^{i-1}\int_{I^{k-1}}(f(x_1,\dots,1,\dots x_k)-f(x_1,\dots,0,\dots x_k))\mathrm{d}x^1\cdots\widehat{\mathrm{d}x^i}\cdots\mathrm{d}x^k) \\
=&(-1)^{i-1}\int_{I^{k-1}}(f(x_1,\dots,1,\dots x_k)-f(x_1,\dots,0,\dots x_k))\mathrm{d}x^1\cdots\widehat{\mathrm{d}x^i}\cdots\mathrm{d}x^k)
\end{aligned}$$

Therefore,

$$\int_{\partial I^k}\omega = \int_{I^k}\mathrm{d}(\omega).$$

Q.E.D

There are many application of Stokes' theorem. For example, we can consider Green's theorem as a special case of Stoke's theorem, i.e.

$$\int_{\partial S} P\mathrm{d}x + Q\mathrm{d}y = \int_S(\frac{\partial Q}{\partial x}-\frac{\partial P}{\partial y})\mathrm{d}x\mathrm{d}y,$$

where S is a singular 2-cell and ∂S is the boundary.

In the following, we give a proof of Brouwer Fixed Point Theorem. We first prove a lemma.

Lemma 4.28. *Let $S^{n-1} = \{x \in \mathbb{B}^n \mid \|x\| = 1\}$ be the surface of $\mathbb{B}^n$. There is no C^∞ mapping $g : \mathbb{B}^n \to S^{n-1}$ such that $g(p) = p$ for any $p \in S^{n-1}$.*

Proof. Assume that $g = (g^1, \ldots, g^n)$ is such a mapping. Let

$$\omega = \sum_{i=1}^{n} (-1)^{i-1} x_i \mathrm{d}x^1 \wedge \cdots \wedge \widehat{\mathrm{d}x^i} \wedge \cdots \wedge \mathrm{d}x^n.$$

Then $g^*\omega = \omega$ on S^{n-1}, since g is identity on S^{n-1}. Since $\|g\| = 1$, then $\sum_{i=1}^n g^i \mathrm{d}g^i = 0$ which implies that

$$\mathrm{d}g^*\omega = g^*\mathrm{d}\omega = g^*(n\mathrm{d}x^1 \wedge \cdots \wedge \mathrm{d}x^n) = n\mathrm{d}g^1 \wedge \cdots \wedge \mathrm{d}g^n = 0$$

on $\mathbb{B}^n$. Apply Stokes' theorem to ω and $g^*\omega$ respectively, we see that

$$\int_{S^{n-1}} \omega = \int_{\mathbb{B}^n} n\mathrm{d}\omega = n\mathrm{vol}\mathbb{B}^n$$

and

$$\int_{S^{n-1}} g^*\omega = \int_{\mathbb{B}^n} \mathrm{d}g^*\omega = 0.$$

However, $g^*\omega = \omega$ on S^{n-1}. It's a contradication. Q.E.D

Theorem 4.29 (Brouwer Fixed Point Theorem)**.** *Let $f : \mathbb{B}^n \to \mathbb{B}^\ltimes$ be a C^∞ mapping between the unit n-ball $\mathbb{B}^n$. Then there is a point $p \in \mathbb{B}^n$ such that $f(p) = p$.*

Proof. Assume in the contrary that $f(p) \neq p$ for any $p \in \mathbb{B}^n$. We can construct a C^∞ mapping $g : \mathbb{B}^n \to S^{n-1}$ such that $g(p) = p$ for any $p \in S^{n-1}$.

For each $p \in \mathbb{B}^n$, consider the function $g : \mathbb{B}^n \to \mathbb{R}^n$ given by $g(p) = p + t(p)(p - f(p))$ where $t : \mathbb{B}^n \to \mathbb{R}$ is a nonnegative C^∞ function. We will prove that t can be chosen such that $\|g(p)\| = 1$. Since

$$\begin{aligned}\|g(p)\| =& \sum_{i=1}^{n} (p_i + t(p)(p_i - f_i(p)))^2 \\ =& t^2(p)\|p - f(p)\|^2 + 2t < p, p - f(p) > + \|p\|^2,\end{aligned}$$

and $\|p\| \leq 1$, then $\|g(p)\| = 1$ has a unique positive solution

$$t(p) = \frac{-<p, p - f(p)> + \sqrt{<p, p - f(p)>^2 + \|p - f(p)\|^2(1 - \|p\|^2)}}{\|p - f(p)\|^2}$$

and t is C^∞. Moreover, $g(p) = p$ if and only if $t(p) = 0$ if and only if $\|p\| = 1$. Q.E.D

5 Manifolds

In this section, we will generalize calculus to smooth manifolds. Roughly speaking, a smooth manifolds is set which locally looks like a linear spaces. For example, consider the unit 2-sphere, $S^2 = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$. If we remove the north pole or south pole, then we can expand the sphere to a plane. However, we can not defined a global coordinate system on S^2. For derivatives, we can focus locally around a point. For integrals, we have the additivity, i.e

$$\int_{S_1 \cup S_2} f = \int_{S_1} f + \int_{S_2} f - \int_{S_1 \cap S_2} f$$

and partition of unity. So essentially we hope to work locally but we have to take care of the compatibility of local results.

From now on, we will call a C^∞ function $f : A \subset \mathbb{R}^m \to B \subset \mathbb{R}^n$ a smooth map.

5.1 Manifold-with-boundary

Definition 5.1. Let U and V be two open sets in $\mathbb{R}^m$. A smooth map $f : U \to V$ is called a *diffeomorphism* if there is a smooth map $g : V \to U$ such that $g \circ f = \mathrm{id}$ is the identity map.

Our first definition of manifold is in terms of the ambient space.

Definition 5.2. A subset M of $\mathbb{R}^m$ is called a k-dimensional manifold if for any point $p \in M$ there exist an open set $U \subset \mathbb{R}^m$ containing p, an open set $V \subset \mathbb{R}^m$, and a diffeomorphism $h : U \to V$ such that

$$h(U \cap M) = V \cap (\mathbb{R}^k \times \{0\}) = \{y \in V \mid y_{k+1} = \cdots = y_m = 0\}.$$

A trivial example is an open set $O \subset \mathbb{R}^m$.

A typical example is the sphere $S^{m-1} \subset \mathbb{R}^m$. As we have mentioned, for any $p \in S^{m-1}$, denote by q the antipode. Up to a linear transformation, we may assume that

$p = (0, \ldots, 0, 0)$ and $q = (0, \ldots, 2)$. Let $U = \mathbb{R}^m \setminus \{x_n = 2\} = V$. Consider the map $h : U \to V$ defined by

$$h(x_1, \ldots, x_n) = \frac{2t - 4}{x_n - 2}(x_1, \ldots, x_{n-1}, x_n - 2) + (0, \ldots, 0, 2),$$

where t is the number such that x is on the sphere centered at $(0, \ldots, 0, t)$ with radius $|2 - t|$. Note that t is a differential function of x which can be solve from a quadratic equation. Then h is a diffeomorphism such that

$$h(S^{m-1} \cap U) = V \cap (\mathbb{R}^{m-1} \times \{0\}).$$

As you can see, it is no easy to figure out the diffeomorphism explicitly in general. However, the implicit function theorem can help us prove the existence of local diffeomorphisms.

Theorem 5.3. *Let $D \subset \mathbb{R}^m$ be an open subset and $f : D \to \mathbb{R}^n$ be a differentiable map with $n \leq m$. Denote by $M = \{x \in D \mid f(x) = 0\}$. Assume that $Df(x)$ has rank n for any $x \in M$. Then M is a $m - n$-dimensional manifold in $\mathbb{R}^m$.*

Proof. This follows from the rank theorem (implicit function theorem). Q.E.D

An alternative definition inspired by the above theorem is the following.

Theorem 5.4 (Theorem-Definition)**.** *A subset $M \subset \mathbb{R}^m$ is a k-dimensional manifold if and only if for each point $x \in M$ there exists an open set $U \subset \mathbb{R}^m$ containing x, an open set $W \subset \mathbb{R}^k$ and a* 1*-to-*1 *differentiable function $f : W \to \mathbb{R}^m$ such that*

(a) $f(W) = M \cap U$,

(b) $Df(y)$ has rank k for each $y \in W$,

(c) $f^{-1} : f(W) \to W$ is continuous.

Proof. If M is a k-dimensional manifold in $\mathbb{R}^m$, choose open sets U, V and a diffeomorphism $h : U \to V$ as in the definition. Let $W = \{a \in \mathbb{R}^k \mid (a, 0) \in h(M)\}$ and define $f(a) = h^{-1}(a, 0)$. Then $f(W) = M \cap U$ by our definition and f^{-1} is continuous. Let $i : \mathbb{R}^k \to \mathbb{R}^m$ be the map given by $i(y) = (y, 0)$. Then $Df(y) = Dh^{-1}Di(y)$ since h is a diffeomorphism and $Di(y) = I_{k\times k}$. Then, for any point $y \in W$, $Df(y)$ has rank k.

Conversely, we need to to find a diffeomrophism $h : U \to V$ such that $h(U \cap M) = V \cap \mathbb{R}^k \times \{0\}$. We will use the inverse function theorem.

Let $f : W \to \mathbb{R}^m$ be a 1-to-1 differentiable function satisfying 5.4.(a)-5.4.(c). Up to a linear transformation of $\mathbb{R}^m$, we may assume that the matrix $\left(D_i f^j(y)\right)_{1\leq,i,j\leq k}$ has rank k. Define $g : W \otimes \mathbb{R}^{m-k} \to \mathbb{R}^m$ by $g(a, b) = f(a) + (0, b)$. Then

$$Dg(a,b) = \begin{pmatrix} \left(D_i f^j(a)\right)_{1\leq,i,j\leq k} & 0 \\ * & I_{m-k\times m-k} \end{pmatrix}.$$

By inverse function theorem, there is an open neighborhood V_1' of $(y, 0)$ and an open neighborhood V_2' of $g(y, 0) = x$ such that $g : V_1' \to V_2'$ is a diffeomorphism. Since f^{-1} is continuous, there is an open set $U \subset \mathbb{R}^m$ such that $U \cap f(W) = \{f(a) \mid (a, 0) \in V_1'\}$. Let $V_2 = V_2' \cap U$ and $V_1 = g^{-1}(V_2)$. Then

$$V_2 \cap M = \{f(a) \mid (a, 0) \in V_1\} = \{g(a, 0) \mid (a, 0) \in V_1\}.$$

Define $h = g^{-1} : V_2 \to V_1$. We see that

$$h(V_2 \cap M) = g^{-1}(V_2 \cap M) = V_1 \cap \mathbb{R}^k \times \{0\}.$$

Q.E.D

The differential map f in Theorem 5.4 is called a *coordinate system* around x. From Theorem 5.4, we note that any two coordinate systems $f_1 : W_1 \to \mathbb{R}^n$ and $f_2 : W_2 \to \mathbb{R}^n$ are compatible, i.e. the map

$$f_2^{-1} \circ f_1 : f_1^{-1} f_2(W_2) \to \mathbb{R}^k$$

is differentiable and has non-singular Jacobian matrix. More precisely,

$$f_2^{-1} \circ f_1 : f_1^{-1}(f_2(W_2) \cap f_1(W_1)) \to f_2^{-1}(f_1(W_1) \cap f_2(W_2))$$

is a diffeomorphism.

As an application of the above theorem, one can check easily using spherical coordinate system that a open ball $B \in\subset \mathbb{R}^3$ is a manifold.

Now we define manifold-with-boundary.

We call the set $\mathbb{H}^k = \{x \in \mathbb{R}^k \mid x_k \geq 0\} \subset \mathbb{R}^k$ the *half-plane* in $\mathbb{R}^k$.

Definition 5.5. A subset M of $\mathbb{R}^m$ is called a k-dimensional *manifold-with-boundary* if for any point $p \in M$ there exist an open set $U \subset \mathbb{R}^m$ containing p, an open set $V \subset \mathbb{R}^m$, and a diffeomorphism $h : U \to V$ such that one of the following two conditions holds

(a) $h(U \cap M) = V \cap (\mathbb{R}^k \times \{0\})$, or

(b) $h(U \cap M) = V \cap (\mathbb{H}^k \times \{0\})$ and the k-th component of $h(p)$ is 0.

The set of all points $p \in M$ such that the condition 5.5.(b) holds is called the boundary of M and denoted by ∂M.

Note that a point $p \in \partial M$ can not satisfies 5.5.(a) and a point $p \in M \setminus \partial M$ can not satisfies 5.5.(b). Roughly speaking, that is because $V \cap \mathbb{R}^k \times \{0\}$ is open, but $V \cap \mathbb{H}^k \times \{0\}$ is not.

It is also clear that the boundary itself is a $(k-1)$-manifold. In fact, for any $p \in \partial M$, we see that $h(U \cap \partial M) = V \cap \mathbb{R}^{k-1} \times \{0\}$. This fact suggests that the boundary of a manifold $M \subset \mathbb{R}^m$ is different from the topological boundary.

For example, consider the interval $[0, 1)$. As a manifold in $\mathbb{R}$, its boundary is $\{0\}$, but as a topological subset its boundary is $\{0, 1\}$.

Let $B = \{x \in \mathbb{R}^2 \mid \|x\| < 1\}$. Then B is a manifold in $\mathbb{R}^2$ and $\partial B = 0$, since B is has a global coordinate system: $\mathrm{id} : B \to B$. But as a topological subspace in $\mathbb{R}^2$, the boundary of B is the circle $S = \{x \in \mathbb{R}^2 \mid \|x\| = 1\}$.

5.2 Differential forms on manifolds

Let M be a k-dimensional manifold in $\mathbb{R}^m$ and $f : W \to \mathbb{R}^m$ a coordinate system around $x = f(a)$. Using pullback and pushforward we can define tangent bundles and differential forms on M. The only thing we need to check is the pullback or pullforward is independent of choice of local coordinate systems.

By the definition of local coordinate system around x, we know that the matrix $Df(a)$ has rank k. Therefore, the linear transformation $f_* : \mathbb{R}^k_a \to \mathbb{R}^m_x$ is 1-to-1 which realizes $f_*\mathbb{R}^k_a$ as a sub-vector space in $\mathbb{R}^m_x$. If $g : V \to \mathbb{R}^m$ is another coordinate system around $x = g(b)$, then we have $f^{-1} \circ g : g^{-1}(f(W_1) \cap g(W_2)) \to \mathbb{R}^k$ is locally a diffeomorphism around a. we will have

$$g_*\mathbb{R}^k_b = f_*((f^{-1} \circ g)_*\mathbb{R}^k_b) = f_*\mathbb{R}^k_a.$$

We call the k-dimensional vector space $f_*\mathbb{R}^k_a$ the *tangent space* of M at x and denote it by T_xM.

Locally at each point $x \in M$, we can define forms as alternating tensor on T_xM. Globally, we view a form as a function from M to the space of alternating tensors. This suggests the following definition of p-forms on M.

We call a function ω which assigns to each $x \in M$ a p-form $\omega(x) \in \wedge^k T_xM$ a *p-form on the manifold* M. A p-form ω on M is *differentiable* if $f^*\omega$ is differentiable, where $f : W \to \mathbb{R}^n$ is a local coordinate system and $\mathrm{d}x^{i_1}, \ldots, \mathrm{d}x^{i_p}$ may subject to some relations.

Note that the tangent space $T_xM \subset \mathbb{R}^m$ is a sub vector space. Then under the inclusion $T_xM \subset \mathbb{R}^m$, a p-form on M can be written as

$$\omega = \sum_{i_1 < \cdots < i_p} \omega_{i_1,\ldots,i_p} \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_p}.$$

where $\omega_{i_1,\ldots,i_p}$ are functions defined only on M.

For example, the form $\omega = y\mathrm{d}x + x\mathrm{d}y$ as a differentiable 1-form on $S^1 \subset \mathbb{R}^2$ is equivalent to $\frac{y^2-x^2}{y}\mathrm{d}x$. In fact, consider polar coordinate system $f : \mathbb{R} \to \mathbb{R}^2$ given by $f(t) = (\cos t, \sin t)$. For any $p = (\cos a, \sin a)$, take $W = (-\pi + a, a + \pi)$, then $f : W \to \mathbb{R}^2$ is a coordinate system around p. Then $f^*\omega = f^*(\frac{y^2-x^2}{y}\mathrm{d}x)$.

So it doesn't make sense to defined $\mathrm{d}\omega$ as $\sum_{i_1<\cdots<i_p} \sum_{i=1}^m D_i\omega_{i_1,\ldots,i_p} \mathrm{d}x^i \wedge \mathrm{d}x^{i_1} \wedge \cdots \wedge \mathrm{d}x^{i_p}$. Instead, we really should consider M locally as an open subset in $\mathbb{R}^k$. And there is a unique way to define d on $\wedge^p T_xM$.

Theorem 5.6. *Let ω be a differential p-form on $M \subset \mathbb{R}^m$. There is a unique differential $(p+1)$-form* $\mathrm{d}\omega$ *on M such that for any coordinate system $f : W \to M$ we have*

$$f^*\mathrm{d}\omega = \mathrm{d}f^*\omega.$$

Proof. Let $f : W \to \mathbb{R}^m$ be a coordinate system such that $f(a) = x$ and $v_1 \ldots, v_{p+1} \in T_xM$. There are unique vectors $w_1, \ldots, w_{p+1}$ in R^k_a such that $f_*w_i = v_i$. Define

$$\mathrm{d}\omega(x)(v_1, \ldots, v_{p+1}) = \mathrm{d}f^*\omega(a)(w_1, \ldots, w_{p+1}).$$

We need to check that $\mathrm{d}\omega$ is independent of the choice of local coordinates. Let $g : V \to \mathbb{R}^m$ be another local system around x such that $g(b) = x$. Then $(g^{-1} \circ f)_* = g_*^{-1}(f_*)$ is a linear transformation between $\mathbb{R}^k_a \to \mathbb{R}^k_b$. Therefore, $g_*^{-1}(f_*w_i)$ are the unique vectors such that $g_*(g_*^{-1}(f_*w_i)) = f_*w_i = v_i$ which proves that $\mathrm{d}\omega$ is well-defined. Moreover, the way that we define it makes it unique. Q.E.D

For a manifold $M \in \mathbb{R}^m$, at each point $x \in M$, we can choose an orientation μ_x for T_xM which can be considered as an equivalent class of an ordered basis for T_xM in the sense that the linear transformation between two bases in the class has positive determinant. However, moving on M, you may choose different orientations for tangent spaces which might cause some unnecessary trouble. We want to a consistent choice in the following sense.

Let μ be an orientation on $\mathbb{R}^m$. On $T\mathbb{R}^m$, for each point x, we identify $T_x\mathbb{R}^m$ with $\mathbb{R}^m$ and choose an orientation $\mu_x = \mu$ for $T_x\mathbb{R}^m = \mathbb{R}^m$. Viewing the choice of orientation as a "function" on $\mathbb{R}^m$, the above choice of orientation for $\mathbb{R}^m$ is simply "constant".

Definition 5.7. Let $f : W \to \mathbb{R}^m$ be an local coordinate system and $a, b \in W$ two points. We say that a choice of orientations μ_x for TM is *consistent* provided that the relation

$$[f_*((e_1)_a), \ldots, f_*((e_k)_a)] = \mu_{f(a)}$$

holds if and only if the relation

$$[f_*((e_1)_b), \ldots, f_*((e_k)_b)] = \mu_{f(b)}$$

holds.

In other words, for any $a, b \in W$, the natural isomorphism between $f_*\mathbb{R}^k_a$ and $f_*\mathbb{R}^k_b$ via the inclusion in $\mathbb{R}^m$ is orientation preserving.

Let (W, μ) and (V, η) be two m-dimensional vector spaces with orientations. An isomorphism $T : W \to V$ is called *orientation preserving* if $[T(v_1), \ldots, T(v_m)] = \eta$ whenever $[v_1, \ldots, v_m] = \mu$.

More generally, let $f : U \to V$ be a diffeomorphism between two open subsets in $\mathbb{R}^k$. We say f is *orientation preserving* if $\det(Df) > 0$. We say f is *orientation reversing* if $\det(Df) < 0$.

Definition 5.8. A manifold $M \in \mathbb{R}^m$ for which an orientation μ can be chosen consistently is called *orientable manifold* and μ is called the orientation of M.

Example 5.9. The share $S^2 \subset \mathbb{R}^3$ is orientable, it has an orientation μ_x such that $[n_x, \mu_x]$ is the standard orientation (right-hand orientation) of $\mathbb{R}^3$, where n_x the outward unit normal vector at x, i.e. the vector n_x from the origin to x.

Example 5.10. The Möbius band is a non-orientable manifold because of the twist.

We can generalize the idea in Example 5.9 to manifolds with boundaries.

Let M be a k-dimensional orientable manifold with boundary and $x \in \partial M$ be a point. Then the tangent space $T_x(\partial M)$ is a subspace of $T_x M$. There are exactly two vectors in $T_x M$ that are perpendicular to $T_x(\partial M)$. We distinguish them as follows. If $f : W \subset \mathbb{H}^k \to \mathbb{R}^n$ is a local coordinate system at x with $f(x) = 0$, then only one of the two unit vectors in $T_x M$ perpendicular to $T_x(\partial M)$, say v_0, has the property that the k-th component v^k of $f_*(v_0) = (v^1, \ldots, v^m)$ is positive. We call v_0 the *outward unit normal vector* and denote it by n_x. Note that the definition does not depend on f, since $g^{-1} \circ f$ has non-singular Jacobian.

Let M be a k-dimensional orientable manifold with boundary and μ is an orientation on M. Then there is an induced orientation for ∂M given in the following way. For any $x \in \partial M$, choose $v_1, \ldots, v_{k-1}$ in $T_x(\partial M)$ such that $[n_x, v_1, \ldots, v_{k-1}] = \mu$. Then $[v_1, \ldots, v_{k-1}]$ determines an orientation for ∂M. We denote this orientation by $\partial\mu$ and call it the *induced orientation*.

Example 5.11. The upper half plane $\mathbb{H}^k = \{(x_1, \ldots, x_k) \mid x_k \geq 0\}$ has a standard orientation $\mu = [e_1, \ldots, e_k]$. The boundary of H^k is $\mathbb{R}^k \times \{0\} = \{(x_1, \ldots, x_k) \mid x_k =$

$0\}$. The outward unit normal vector is simple e_k. The induced orientation is $\partial\mu = (-1)^k[e_1, \ldots, e_{k-1}]$.

5.3 Stokes' theorem on manifolds

Let ω be a differentiable k-form on a k-dimension orientable manifold M with boundary ∂M. Let $f : W \to \mathbb{R}^m$ be a local coordinate system. Up to a orientation preserving map between $\mathbb{R}^k$, we may assume that $[0,1]^k \subset W$. Therefore, we will assume that a k-cell c is of the type $c(x) = f(x)$. A k-cell c is orientation preserving if f is orientation preserving.

On manifolds, it could happen that two local coordinate system has overlaps. So in order to define integral on manifolds using integration on k-chains, we need the following theorem.

Theorem 5.12. *Let $c_1, c_2 : [0,1]^k \to M$ be two orientation preserving singular k-cell in a k-dimension orientable manifold M and ω be a k-form on M such that $\omega = 0$ outside of $c_1([0,1]^k) \cap c_2([0,1]^k)$. Then*

$$\int_{c_1} \omega = \int_{c_2} \omega.$$

Proof. Let $D = c_1([0,1]^k) \cap c_2([0,1]^k) \subset [0,1]^k$. Then we have a diffeomorphism $c_2^{-1} \circ c_1 : c_1^{-1}(D) \to c_2^{-1}(D)$. Composing with $c_2 : c_2^{-1}(D) \to D$, we see that $c_1^*\omega = (c_2^{-1} \circ c_1)^* c_2^*\omega$. Since c_1 and c_2 are both orientation preserving then $\det D((c_2^{-1} \circ c_1)) > 0$, apply change of variable theorem, we see that

$$\int_{c_2} \omega = \int_{c_2^{-1}(D)} c_2^*\omega = \int_{c_1^{-1}(D)} (c_2^{-1} \circ c_1)^* c_2^*\omega.$$

Therefore,

$$\int_{c_1} \omega = \int_{c_1^{-1}(D)} c_1^*\omega = \int_{c_1^{-1}(D)} (c_2^{-1} \circ c_1)^* c_2^*\omega = \int_{c_2} \omega.$$

Q.E.D

Let ω be a k-form on an oriented k-dimension manifold M such that $\omega = 0$ outside an orientation preserving singular k-cell c in M. We define

$$\int_M \omega = \int_c \omega.$$

For arbitrary k-forms on M, we need partition of unity.

From now on we assume that M is compact, i.e. M is a compact subset in $\mathbb{R}^m$. Let $\emptyset = \{U\}$ be an open cover of M such that for each $U \in \emptyset$ there is an orientation preserving singular k-cell c containing U. Take a partial of unit Φ subject to $\emptyset$, i.e. for each U, there is a $\phi_U > 0$ such that $\sum_{\phi_U} = 0$. Since we assume that M is compact, this can always be realized. Similar to before, we define

$$\int_M \omega = \sum_{\phi \in \Phi} \int_M \phi\omega.$$

Let c be a singular k-cell in M such that $c_{k,0}$ lies in ∂M. By Example 5.11, orientation of $c_{k,0}$ may changes to the opposite if k is odd. Therefore, for a $(k-1)$-form on M which is 0 outside $c_{k,0}$, we define

$$\int_{c_{k,0}} \omega = (-1)^k \int_{\partial M} \omega.$$

In particular, if an orientation preserving k-cell only has the face $c_{k,0}$ in ∂M, then

$$\int_{\partial c} \omega = \int_{(-1)^k c_{k,0}} \omega = (-1)^k \int_{c_{k,0}} \omega = \int_{\partial M} \omega.$$

This observation is one of the two special cases in the proof of the following generalized Stokes' theorem.

Theorem 5.13 (Stokes's theorem on manifolds)**.** *Let M be a compact oriented manifold with boundary and ω be a $k-1$-form on ∂M which has the induced orientation. Then*

$$\int_{\partial M} \omega = \int_M \mathrm{d}\omega.$$

Proof. First, if $\omega = 0$ outside a singular k-cell c in $M \setminus \partial M$, the by Stokes' theorem for k-cells in $\mathbb{R}^k$, we are done.

Assume that c is an orientation preserving singular k-cell with only one face $c_{k,0}$ in ∂M and $\omega = 0$ outside of c. Again by Stokes' theorem for k-cells in $\mathbb{R}^k$, we are done.

In general, we take an open cover and partition Φ of unit subordinate to this cover such that for each $\phi \in \Phi$ the form $\phi\omega$ is one of two types we considered. Since $\sum_{\phi\in\Phi} \phi = 1$, then $\sum_{\phi\in\Phi} \mathrm{d}\phi = 0$ which implies $\sum_{\phi\in\Phi} \mathrm{d}\phi \wedge \omega = 0$. Then $\mathrm{d}\sum_{\phi\in\Phi} \phi\omega = \sum_{\phi\in\Phi} \phi \mathrm{d}\omega$. Since M is compact, we may assume the open cover is finite so that the sum $\sum_{\phi\in\Phi}$ interchanges with $\int_M$. Therefore, we have

$$\int_M \omega = \int_{\partial M} \mathrm{d}\omega.$$

Q.E.D