Mathematical Analysis II

Felix Chen

Contents

§1 Introduction

Teacher: Yang Shiwu

Grading: Homework-Midterm-Endterm: 30-30-40
Contents of this course: Real analysis

§1.1 Recap

Definition 1.1.1 (Measurable space)**.** Let X be a set and $\mathcal{A}$ be a σ-algebra, we say $(X, \mathcal{A})$ is a measurable space if

- $\emptyset \in \mathcal{A}$;
- If $A \in \mathcal{A}$, then $A^c \in \mathcal{A}$;
- If $A_k \in \mathcal{A}$, then $\bigcup_{k=1}^{+\infty} \in \mathcal{A}$.

Outer measure m^*:

- $m^*(A) \geq 0$;
- $m^*(\bigcup_{k=1}^{\infty} A_k) \leq \sum_{k=1}^{\infty} m^*(A_k)$;
- $m^*(A) \leq m^*(B)$ when $A \subset B$.

Caratheodry condition:

$$m^*(T) = m^*(T \cap E) + m^*(T \cap E^c), \quad \forall T \subset X.$$

We define the measurable sets to be all the sets E satisfying above condition.

This implies the Lebesgue measure space $(\mathbb{R}^n, \mathscr{U}, m)$. It is a complete measure space, i.e. null sets are measurable.

Proposition 1.1.2

Properties of measurable sets:

- Let E be a measurable set, there exists a G_δ set G and a F_σ set F such that
$$E = G \backslash Z_1 = F \cup Z_2.$$
where Z_1, Z_2 are null sets.
- (Fatou's Lemma)
Measurable sets $E_k \nearrow E \implies \lim_{k\to\infty} m(E_k) = m(E)$ and
$$m\left(\liminf_{k\to\infty} E_k\right) \leq \liminf_{k\to\infty} m(E_k).$$

Definition 1.1.3 (Measurable function)**.** Let f be a map from measurable space $(X, \mathcal{A})$ to $(Y, \mathcal{B})$. We say f is measurable if

$$\forall B \in \mathcal{B}, \quad f^{-1}(B) \in \mathcal{A}.$$

In the case of real functions, this is equivalent to

$$\forall t \in \mathbb{R}, \quad \{x; f(x) > t\} \text{ is a measurable set.}$$

Proposition 1.1.4

Let f be a non-negative measurable function, $\exists \varphi_k \nearrow f$, where φ_k are simple functions.

For a general measurable function f, decompose it to $f = f_+ - f_-$.

Theorem 1.1.5 (Egorov)
Let E be a measurable set and $m(E) < \infty$, $f_n \to f, a.e.$, Then $\forall \varepsilon > 0$, there exists a closed set F_ε s.t. $m(E \backslash F_\varepsilon) < \varepsilon$ and $f_n \to f$ uniformly on F_ε.

Theorem 1.1.6 (Lusin)
Let E be a measurable set and $m(E) < \infty$. Then $\forall \varepsilon > 0, \exists F_\varepsilon$ such that $f\big|_{F_\varepsilon}$ is continuous.

Convergence patterns:

- Converge **almost everywhere**: $f_n \to f, a.e.$
- Converge **almost uniformly**: $f_n \to f, a.u.$
- Converge **in measure**: $f_n \xrightarrow{m} f$

§2 Lebesgue integrals

§2.1 Recap: Definition of Lebesgue integrals

- Simple functions: $f = \sum_{k=1}^N a_k \chi_{E_k}$, define
$$\int f = \sum_{k=1}^N a_k m(E_k).$$
- $f : E \to \mathbb{R}^n$, where $m(E) < \infty$, f bounded. These functions form the set $\mathcal{L}_0$. Then $\exists \varphi_k \to f$, φ_k simple, define
$$\int f = \lim_{k\to\infty} \int \varphi_k.$$
- Non-negative function:
$$\int f = \sup \left\{ \int g \;\middle|\; 0 \le g \le f, g \in \mathcal{L}_0 \right\}.$$
- General functions:
$$\int f = \int f_+ - \int f_-.$$
Integrable $\iff \int f_+, \int f_- < +\infty$.

Relations between Riemann integrals and Lebesgue integrals:

- f is Riemann integrable on $[a, b]$ iff f bounded and the discontinuous points form a null set.
- If f is Riemann integrable on $[a, b]$, then two types of integral yield the same result.

§2.2 Dominated convergence theorem

The critical question of this section: If a sequence of functions f_n converges to f (almost everywhere), when does their integrals $\int f_n$ converge to $\int f$?

We'll discuss this issue under various conditions and reach the famous Dominated Convergence Theorem.

Theorem 2.2.1

Let E be a measurable set with finite measure. Measurable functions $f_n \to f, a.e.$ on E. Furthermore, f_n is uniformly bounded almost everywhere ($|f_n| < M, a.e.$). Then we have

$$\int_E |f_n - f| \to 0 \implies \lim_{m\to\infty} \int_E f_n = \int_E f.$$

Proof. By Egorov's Theorem, $\forall \varepsilon > 0$, there exists $F_\varepsilon \subset E$ s.t. $f_n \to f$ uniformly on F_ε, and $m(E \backslash F_\varepsilon) < \varepsilon$.

Hence

$$\begin{aligned}\int_E |f_n - f| &= \int_{F_\varepsilon} |f_n - f| + \int_{E\backslash F_\varepsilon} |f_n - f| \\ &\le \varepsilon_0 m(E) + 2M\varepsilon,\end{aligned}$$

which proves the result. □

Lemma 2.2.2 (Fatou's Lemma)

If $f_n \ge 0$, then

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n.$$

Proof. For any $g \in \mathcal{L}_0$, $0 \le g \le \liminf_{n\to\infty} f_n$, we need to prove $\int g \le \liminf \int f_n$.

Let $g_k = \min\{f_k, g\}$, assmue g is uniformly bounded so that $g_k \in \mathcal{L}_0$.

We'll prove $g_k \to g$: Assmue by contradiction that $\exists \varepsilon_0 > 0, \exists x_0$ s.t.

$$g(x_0) - g_{k'}(x_0) > \varepsilon_0.$$

then $g(x_0) - f_{k'}(x_0) > \varepsilon_0$, which contradicts with $g \le \liminf_{n\to\infty} f_n$.

Thus for sufficiently large k, $g_k(x) \le g(x) \le g_k(x) + \varepsilon_0, \implies g_k \to g$.

Therefore by Theorem 2.2.1 (note $g_k \in \mathcal{L}_0$),

$$\begin{aligned}\int g &= \lim_{k\to\infty} \int g_k \\ &\le \liminf_{k\to\infty} \int f_k,\end{aligned}$$

and we're done. □

Remark 2.2.3 — This is nearly indentical to the measure version of Fatou's Lemma (Proposition 1.1.2). It shows some similarities between measure and integrals.

Theorem 2.2.4 (Beppo-Levi)
If non-negative functions $f_n \nearrow f$, we have

$$\lim_{n\to\infty} \int f_n = \int f.$$

Proof.

$$f_n \le f \implies \lim_{n\to\infty} \int f_n \le \int f.$$

By Fatou's Lemma (2.2.2),

$$\int \liminf_{n\to\infty} f_n \le \liminf_{n\to\infty} \int f_n = \lim_{n\to\infty} \int f_n,$$

$$\implies \int f \le \lim_{n\to\infty} \int f_n.$$

Combining the two inequalities we get the desired equality. □

Corollary 2.2.5
Let f_n be non-negative functions, then

$$\int \sum_{k=1}^{\infty} f_k(x) = \sum_{k=1}^{\infty} \int f_k(x).$$

Proposition 2.2.6
Let f be an integrable function, $\forall \varepsilon > 0$, we have:

- There exists a set B with finite measure s.t.

$$\int_{B^c} |f| < \varepsilon.$$

- (**Absolute continuity** of integrals) $\exists \delta > 0$ s.t. $\forall E$, if $m(E) < \delta$,

$$\int_E |f| < \varepsilon.$$

  This is equivalent to

$$\lim_{m(E)\to 0} \int_E |f| = 0.$$

Proof. Let $f_N(x) = |f(x)|$ when $|x| \le N, |f(x)| \le N$, and $f_N(x) = 0$ otherwise. Then $f_N \nearrow |f|$, so by Beppo-Levi (Theorem 2.2.4), we get

$$\lim_{N\to\infty} \int f_N = \int |f|.$$

Let $B = \{x \mid |x| \leq N, |f(x)| \leq N\}$, when N gets sufficiently large, we must have $\int_{B^c} |f| < \varepsilon$.

For the second part, when N is sufficiently large we have $\int(|f| - f_N) < \frac{\varepsilon}{2}$, so

$$\begin{aligned}\int_E |f| &= \int_E f_N + \int_E (|f| - f_N) \\ &\leq N \cdot m(E) + \frac{\varepsilon}{2}.\end{aligned}$$

Let $\delta = \frac{\varepsilon}{2N}$ to finish. □

Now we take a look at what we get so far:

- If bounded functions $f_n \in \mathcal{L}_0$, $f_n \to f$, then $\int f_n \to \int f$.
- If f_n is non-negative, then $\int \liminf f_n \leq \liminf \int f_n$. (Fatou)
  This corresponds to: $m(\liminf E_n) \leq \liminf m(E_n)$.
- If $f_n \nearrow f$, then $\int f_n \nearrow \int f$. (Beppo-Levi)
  This corresponds to: $E_n \subset E_{n+1} \implies m(\bigcup E_n) = \lim m(E_n)$.

Finally we come to the famous Lebesgue dominated convergence theorem:

Theorem 2.2.7 (Lebesgue Dominated Convergence Theorem)
Functions $f_n \to f, a.e.$, if there exists a function g s.t. $|f_n| \leq g, a.e.$, then we have:

$$\int |f - f_n| \to 0. \left(\lim_{n\to\infty} \int f_n = \int f\right)$$

Proof. By Fatou's lemma (2.2.2), $2g - |f_n - f|$ is non-negative,

$$\int \liminf(2g - |f_n - f|) \leq \liminf \int (2g - |f_n - f|)$$

$$\implies 0 \leq \liminf \left(-\int |f_n - f|\right)$$

$\implies \limsup \int |f_n - f| \leq 0$, hence it must equal to 0. □

Example 2.2.8
Non-examples of lebesgue dominated convergence theorem:

- Let $f_n = \chi_{[n,n+1]}$, $g = 1$, note that g is not integrable, so $\int f_n = 1$ while $f_n \to 0$.
- $f_n = \frac{1}{n}\chi_{[0,n]}$, $f_n \to 0$, $\int f_n = 1 \not\to 0$. Since $g(x) = \min\{\frac{1}{x}, 1\}$, which isn't integrable.
- $f_n = n\chi_{(0,\frac{1}{n})}$, $f_n \to 0$, $\int f_n = 1 \not\to 0$. Here $g(x) = \frac{1}{x}\chi_{[0,1]}$ is not integrable.

Example 2.2.9

Suppose that
$$\lim_{n\to\infty}\int_E f_n = \int_E f$$
holds for any measurable set E. Then
$$\liminf_{n\to\infty} f_n \le f \le \limsup_{n\to\infty} f_n, a.e..$$

Proof. We only need to prove the case when $f = 0$.

For $\forall\varepsilon > 0$, define
$$E_n^\varepsilon = \{x : f_n(x) < -\varepsilon\}.$$
Note that
$$\liminf E_n^\varepsilon \subset \{x : \limsup f_n \le -\varepsilon\} \subset \liminf E_n^{\frac{\varepsilon}{2}}.$$
Because when $\limsup f_n(x) \le -\varepsilon$, $\exists N$ such that $\sup_{n>N} f_n(x) < -\frac{\varepsilon}{2}$
$$\implies f_n(x) < -\frac{\varepsilon}{2}, \forall n > N$$

This implies $x \in E_n^{\frac{\varepsilon}{2}}, \forall n > N$, so $x \in \liminf E_n^{\frac{\varepsilon}{2}}$.

We proceed with the proof, by using the condition ($E = \bigcap_{k\ge N} E_k^\varepsilon$),
$$0 = \lim_{n\to\infty}\int_{\bigcap_{k\ge N} E_k^\varepsilon} f_n.$$
Since $x \in \bigcap_{k\ge N} E_k^\varepsilon \implies f_k(x) < -\varepsilon$, we deduce
$$0 = \lim_{n\to\infty}\int_{\bigcap_{k\ge N} E_k^\varepsilon} f_n \le (-\varepsilon)\cdot m\left(\bigcap_{k\ge N} E_k^\varepsilon\right)$$
Hence $E = \bigcap_{k\ge N} E_k^\varepsilon$ is a null set. □

§2.3 Integrable function space $\mathcal{L}^1(E)$

Definition 2.3.1 ($\mathcal{L}^1$ space)**.** Denoted by $\mathcal{L}^1(E)$ the space consisting of all the integrable functions on E.

If $f = g, a.e.$, then $\int|f-g| = 0$, we regard them as equivalent elements in $\mathcal{L}^1(E)$.

Observe that $\mathcal{L}^1(E)$ is a vector space, define the norm:
$$\|f\| = \int_E |f|.$$
It's easy to check that $\mathcal{L}^1(E)$ becomes a normal vector space.

Moreover, it's also a **Banach space** (complete normal vector space).

Theorem 2.3.2

$\mathcal{L}^1(E)$ is a Banach space.

Proof. Let f_n be a Cauchy sequence in $\mathcal{L}^1(E)$, suppose $\|f_{n_k} - f_{n_{k+1}}\| < 2^{-k}$.

Let $f = \sum_{k=0}^{\infty}(f_{n_{k+1}} - f_{n_k})$, where $f_{n_0} = 0$. Because

$$\int_E \sum_{k=0}^{\infty} |f_{n_{k+1}} - f_{n_k}| = \sum_{k=0}^{\infty} \int_E |f_{n_{k+1} - f_{n_k}}| \le \sum_{k=0}^{\infty} 2^{-k} < +\infty.$$

so our f is well-defined (convergent). Now we compute

$$\begin{aligned} \|f - f_m\| &= \|f_m - f_{n_l}\| + \left\| \sum_{k=l}^{\infty}(f_{n_{k+1}} - f_{n_k}) \right\| \\ &\le \|f_m - f_{n_l}\| + \sum_{k=l}^{\infty} \|f_{n_{k+1}} - f_{n_k}\| \\ &\le \|f_m - f_{n_l}\| + 2^{-l+1}. \end{aligned}$$

As m gets large, $\|f_m - f_{n_l}\|$ and 2^{-l+1} both converge to 0, so $f_n \to f$ in $\mathcal{L}^1(E)$. □

Remark 2.3.3 — Notes on multi-dimensional Riemann integrals:

For functions $f : \mathbb{R} \to \mathbb{R}$, recall that

$$\int_a^b f \, \mathrm{d}x = \lim_{\delta_i \to 0} \sum_{i=1}^{n} f(\xi_i)(x_i - x_{i-1}).$$

But in higher dimensional spaces, it's not so easy to find the suitable partition of the integral region. In fact, this requires the differential theory of multi-dimensional functions first.

As for the improper integrals, for an unbounded region D, we can simlilarly define it to be

$$\int_D f \, \mathrm{d}x = \lim_{n \to \infty} \int_{D_n} f \, \mathrm{d}x,$$

where D_n can be any shape, so the limit is actually stronger than its one-dimensional counterpart. In other words, when we partition D into small cuboids, there's an issue of the summation order.

This means the integral must be "absolutely" convergent, since by Riemann rearrangement theorem, conditional convergent sequence can be rearranged so that it becomes *divergent*.

Here we state again that if $f = g, a.e$, we regard them as the same function.

Definition 2.3.4 ($\mathcal{L}^p$ space)**.** Define the $\mathcal{L}^p$ space to be

$$\mathcal{L}^p(E) = \left\{ f \mid \left(\int_E |f|^p \right)^{\frac{1}{p}} < +\infty \right\}$$

Simliarly, it's a complete normal vector space.

In this course we mainly discuss about $\mathcal{L}^1$ instead of general $\mathcal{L}^p$.

Theorem 2.3.5

The following function spaces are dense in $\mathcal{L}^1$ space:

- Simple functions;
- Step functions;
- Continuous functions with compact support, denoted by $C_0(E)$ or $C_c(E)$.
- Smooth functions with compact support, denoted by $C_0^\infty(E)$.

Proof. • **Simple functions**:

Density is equivalent to:

$$\forall \varepsilon > 0, \exists \text{simple function } g, s.t. \|f - g\| < \varepsilon.$$

f integrable $\implies f_+, f_-$ measurable, so there exists simple functions φ_+^n and φ_-^n s.t.

$$\varphi_+^n \nearrow f_+, \varphi_-^n \nearrow f_- \overset{\text{Beppo-Levi}}{\Longrightarrow} \int \varphi_+^n \nearrow \int f_+ < \infty, \int \varphi_-^n \nearrow \int f_- < \infty$$

This implies $\int (f_\pm - \varphi_\pm^n) \to 0$.

• **Step functions**:

Let $g = \sum_{k=1}^N a_k \chi_{E_k}$, we only need to consider the case $g = \chi_{E_k}$, where E_k is a measurable set with finite measure.

Take cuboids I_j s.t. $E_k \subset \bigcup_{j=1}^\infty I_j$, and $m(E_k) + \varepsilon > \sum_{j=1}^{+\infty} |I_j|$.

Let $h = \chi_{\bigcup_{j=1}^\infty I_j}$, then

$$\begin{aligned} \int |h - g| &= \left| E_k \Delta \left(\bigcup_{j=1}^N I_j \right) \right| \\ &< \varepsilon + \sum_{j>N} |I_j| \end{aligned}$$

Let N be sufficiently large, we conclude that $\int |f - g| \to 0$.

• **1-dimensional continuous functions**:

Let $l = \begin{cases} 0, & x \in (-\infty, a] \cup [b, +\infty) \\ 1, & x \in [a + \varepsilon, b - \varepsilon] \\ linear/smooth, & otherwise \end{cases}$

Then l is a continuous/smooth function s.t. $\|\chi_{[a,b]} - l\| < 2\varepsilon$.

• **Multi-dimensional continuous functions**:

Let $I = (a_1, b_1) \times \cdots \times (a_n, b_n)$ be a cuboid. Let $l_1, \ldots, l_n$ be continuous/smooth functions on (a_i, b_i) defined earlier. We have

$$\|l_1(x_1) \cdots l_n(x_n) - \chi_I\| < C(n)\varepsilon,$$

where $C(n)$ is a constant depending on n. □

Proposition 2.3.6 (Integrals are invariance under translation and scaling)

Let $f \in \mathcal{L}^1(\mathbb{R}^n)$, for $h \in \mathbb{R}^n$, define $\tau_h(f)(x) = f(x+h)$, then $\tau_h(f) \in \mathcal{L}^1$, and $\|\tau_h(f)\| = \|f\|$.

Simlilarly, define $D_\delta f(x) = f(\delta x)$, then $D_\delta f \in \mathcal{L}^1$, $\|D_\delta f\| = \delta^{-n}\|f\|$.

Theorem 2.3.7 (Translation and scaling are continuous)
For $h \in \mathbb{R}^n$ and $\delta \in \mathbb{R}$,

$$\lim_{h\to 0}\|\tau_h f - f\| = 0, \quad \lim_{\delta\to 1}\|D_\delta f - f\| = 0.$$

Proof. $\forall \varepsilon > 0$, $\exists$ step function g such that $\|g - f\| < \frac{\varepsilon}{3}$.

$$\begin{aligned}\|\tau_h f - f\| &= \|\tau_h(f-g) - (f-g) + (\tau_h g - g)\| \\ &= \|\tau_h(f-g)\| + \|f-g\| + \|\tau_h g - g\| \\ &= \|\tau_h g - g\| + \frac{2}{3}\varepsilon.\end{aligned}$$

Suppose $g = \sum_{k=1}^N a_k \chi_{I_k}$, it's sufficient to prove the case $g = \chi_I$:

$$\lim_{h\to 0}\|\tau_h g - g\| = \lim_{h\to 0}\|I\Delta(I+h)\| = 0.$$

Similarly D_δ is continuous. □

§3 Fubini's theorem

This theorem provides a way to compute nulti-dimensional integrals.

Let $f(x,y) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$. We wonder if the following equation holds:

$$\int f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_{\mathbb{R}^{d_1}}\left(\int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y\right)\mathrm{d}x = \int_{\mathbb{R}^{d_2}}\left(\int_{\mathbb{R}^{d_1}} f(x,y)\,\mathrm{d}x\right)\mathrm{d}y?$$

In fact, this formula somehow says the same thing as the area of a rectangle is equal to its width and length, and this multiplication is commutative.

Theorem 3.0.1 (Fubini's Theorem)
Let $f(x,y) : \mathbb{R}^{d_1} \times \mathbb{R}^{d_2} \to \mathbb{R}$, and f is integrable on $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$.

1. $f(x,y)$ as a function of y is integrable on $\mathbb{R}^{d_2}$ for $x \in \mathbb{R}^{d_1}\backslash Z$ with $m(Z) = 0$.
2. Let $g(x) = \int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y$, for $x \in \mathbb{R}^{d_1}\backslash Z$, where Z is a null set. We have g is integrable on $\mathbb{R}^{d_1}$.
3. $$\int_{\mathbb{R}^{d_1+d_2}} f(x,y) = \int_{\mathbb{R}^{d_1}}\left(\int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y\right)\mathrm{d}x.$$

Proof. Let $\mathscr{F}$ be the space consisting of all the integrable functions that satisfy Fubini's theorem.

Lemma 3.0.2
$\mathscr{F}$ is a vector space. Furthermore, for non-negative monotone sequence $f_n \in \mathscr{F}$, if $\lim f_n$ is integrable, then $\lim f_n \in \mathscr{F}$ as well.

Proof of the lemma. First notice that $f \in \mathscr{F} \implies cf \in \mathscr{F}$.

If $f, g \in \mathscr{F}$, consider $f + g$:

By our conditions, there exists $X_f, X_g \subset \mathbb{R}^{d_1}$, s.t. $f(x, y)$ integrable on $\mathbb{R}^{d_2}$, $\forall x \notin X_f$, and $g(x, y)$ integrable on $\mathbb{R}^{d_2}$, $\forall x \notin X_g$.

This implies $f(x, y) + g(x, y)$ integrable on $\mathbb{R}^{d_2}$ for $x \notin X_f \cup X_g$, which proves (1).

$$\int_{\mathbb{R}^{d_2}} f(x,y) + g(x,y)\,\mathrm{d}y = \int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y + \int_{\mathbb{R}^{d_2}} g(x,y)\,\mathrm{d}y.$$

So the LHS is integrable on $\mathbb{R}^{d_1}$ (this is (2)), taking the integral we get

$$\int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x + \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} g(x,y)\,\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f(x,y) + g(x,y)\,\mathrm{d}y \right) \mathrm{d}x.$$

Therefore $\mathscr{F}$ is a vector space.

For a monotone non-negative sequence f_n, $\exists X_n \subset \mathbb{R}^{d_1}$ s.t. f_n is integrable with respect to y for $x \notin X_n$.

Similarly, when $x \notin \bigcup_{n=1}^{\infty} X_n$, as a function of y, by Beppo-Levi (or Dominated convergence),

$$\int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y = \lim_{n\to\infty} \int_{\mathbb{R}^{d_2}} f_n(x,y)\,\mathrm{d}y.$$

This equation holds when $\int f(x,y)\,\mathrm{d}y$ is finite, so we need to prove it is finite almost everywhere.

For $x \notin \bigcup X_n$, we have:

$$\int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f_n(x,y)\,\mathrm{d}y \right) \mathrm{d}x \to \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x$$

$$\int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f_n(x,y)\,\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^{d_1+d_2}} f_n \to \int_{\mathbb{R}^{d_1+d_2}} f$$

Compare these relations we deduce

$$\int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y \right) \mathrm{d}x = \int_{\mathbb{R}^{d_1+d_2}} f < +\infty.$$

so $\int_{\mathbb{R}^{d_2}} f(x,y)\,\mathrm{d}y$ is finite almost everywhere. This gives (1), and (2), (3) follows immediatedly. $\square$

Back to the proof of the original theorem, we want to prove $\mathscr{F} = \mathcal{L}^1$.

We prove the indicator function of following sets are in $\mathscr{F}$:

- Cuboids;
- Finite open sets;
- G_δ sets;
- Null sets;
- General measurable sets.

Let I be a cuboid, $I = I_x \times I_y$, so $\chi_I = \chi_{I_x}\chi_{I_y}$.

$$\int \chi_I = |I| = |I_x||I_y| = \int \chi_{I_x}|I_y|\,\mathrm{d}x = \int \int (\chi_{I_x}\chi_{I_y}\,\mathrm{d}y)\,\mathrm{d}x.$$

Let O be a finite open set, $O = \bigcup_{n=1}^{\infty} I_n$, where I_n are pairwise disjoint cuboids.

$$\chi_O = \lim_{n\to\infty} \chi_{\bigcup_{k=1}^{n} I_k} \in \mathscr{F},$$

as it's an inceasing sequence.

For $G_\delta = \bigcap_{n=1}^{\infty} O_n$, $\chi_{O_n} \searrow \chi_{G_\delta}$. $\implies \chi_{G_\delta} \in \mathscr{F}$.

For null set E, if $\chi_E \in \mathscr{F}$, $\forall A \subset E$,

$$0 = \int \chi_E = \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} \chi_E \, \mathrm{d}y \right) \mathrm{d}x.$$

hence $\int_{\mathbb{R}^{d_2}} \chi_E \, \mathrm{d}y = 0$, for $x, a.e.$ $\implies \int_{\mathbb{R}^{d_2}} \chi_A \, \mathrm{d}y = 0$ for $x, a.e.$.

Taking the integral with respect to x, we have $\chi_A \in \mathscr{F}$.

Therefore if E is a null set, by taking its equi-measure hull we deduce $\chi_E \in \mathscr{F}$.

Finally, for a general measurable set E, let O be its equi-measure hull, and $E = O \backslash A$. since $\mathscr{F}$ is a vector space, $\chi_E \in \mathscr{F}$.

The rest is trival now: Because all the simple functions are in $\mathscr{F}$, and any measurable functions can be expressed as limits of increasing simple functions, so $\mathscr{F} = \mathcal{L}^1(\mathbb{R}^{d_1+d_2})$. □

Theorem 3.0.3 (Tonelli's theorem)

Let f be a non-negative measurable function on $\mathbb{R}^d$.

- $f(x, y)$ is measurable on $\mathbb{R}^{d_2}$ for x almost everywhere;
- $\int_{\mathbb{R}^{d_2}} f(x, y) \, \mathrm{d}y$ as a function of x is measurable;
- The integral satisfies:
$$\int_{\mathbb{R}^d} f = \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} f(x, y) \, \mathrm{d}y \right) \mathrm{d}x.$$

Proof. Consider the truncation function $f(x, y)\chi_{|x|+|y|<k}\chi_{f<k}$. □

Proposition 3.0.4

Let E be a measurable set on $\mathbb{R}^d$. For x almost everywhere, $E^x = \{y \mid (x, y) \in E\}$ is measurable on $\mathbb{R}^{d_2}$.

As a function of x, $m(E^x)$ satisfies

$$m(E) = \int_{\mathbb{R}^{d_1}} m(E^x).$$

Proof. Consider $f = \chi_E$ and use Tonelli's theorem. □

§3.1 Applications of Fubini's theorem

Definition 3.1.1 (Product measure)**.** Let $(X, \mathscr{F}, m)$ and $(Y, \mathscr{G}, m)$ be measure spaces, define a measure on $X \times Y$: The measure m induces an outer measure on $X \times Y$, and complete it to a normal measure by using Caratheodory conditions. This measure is called the **product measure** on $X \times Y$.

Theorem 3.1.2

Let $\mathbb{R}^d = \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$, E_1, E_2 are subsets of $\mathbb{R}^{d_1}, \mathbb{R}^{d_2}$, respectively.

- If E_1, E_2 are measurable, then E is measurable as well, and $m(E) = m(E_1)m(E_2)$.
- If E is measurable, then E_1, E_2 are measurable, and $m(E) = m(E_1)m(E_2)$, unless one of E_1, E_2 is null set, which means E is null as well.

Proof. First it's easy to note that

$$m^*(E) \leq m^*(E_1)m^*(E_2).$$

So we directly conclude that if one of E_1, E_2 is null set, E must be null.

Thus we may assume below that E_1, E_2 have finite nonzero measure. By taking the equimeasure hull of E_1, E_2 (denoted by F_1, F_2), let $Z_1 = F_1 \backslash E_1, Z_2 = F_2 \backslash E_2$, we have

$$(F_1 \times F_2) \backslash (Z_1 \times F_2 \cup F_1 \backslash Z_2) \subset E \subset F_1 \times F_2,$$

so E is measurable.

Conversely, if E is measurable, consider the measurable function χ_E, by definition $\chi_E = \chi_{E_1}\chi_{E_2}$, hence by Tonelli's theorem, for x almost everywhere, $\chi_{E_1}(x)\chi_{E_2}$ is measurable on $\mathbb{R}^{d_2}$ $\implies$ E_2 is measurable.

Therefore we have the equation

$$m(E) = \int_{\mathbb{R}^d} \chi_E = \int_{\mathbb{R}^{d_1}} \left(\int_{\mathbb{R}^{d_2}} \chi_{E_1}\chi_{E_2} \right) = m(E_1)m(E_2).$$

This proves the theorem. $\square$

Corollary 3.1.3

Let $f(x)$ be a measurable function on $\mathbb{R}^{d_1}$, we have $g(x, y) = f(x)$ is measurable on $\mathbb{R}^{d_2}$.

Proof. It's sufficient to prove that $\{(x, y) | f(x) > t\}$ is measurable in $\mathbb{R}^d$. This follows from the fact that

$$\{(x, y) | f(x) > t\} = \{x | f(x) > t\} \times \mathbb{R}^{d_2},$$

and the previous theorem. $\square$

Proposition 3.1.4

Let L be a linear map $\mathbb{R}^d \to \mathbb{R}^d$, $E \subset \mathbb{R}^d$ a measurable set, then $L(E)$ is measurable, and

$$m(L(E)) = |\det L| m(E).$$

Proof. In fact we only need to prove it for cuboids E and elementary linear transformation L.

Now we only need to look at the case where $L = \begin{pmatrix} 1 & c & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$ since the other cases are trivial or similar to this case.

Thus by Fubini's theorem, WLOG E is the unit cube,

$$m(L(E)) = \int \chi_{L(E)} = \int_{\mathbb{R}^{d-1}} \left(\int_{\mathbb{R}} \chi_{L(E)\,\mathrm{d}x_1} \right) = \int_{\mathbb{R}^{d-1}} \chi_{E'} \cdot 1 = 1 = |\det L| m(E),$$

where $E' = \{(x_2, \ldots, x_n) | 0 \leq x_i \leq 1\}$. □

From this transformation formula we deduce the integral version:
Let f be an integrable function on $\mathbb{R}^d$, then $f(L(x))$ is also integrable, and

$$\int f(L(x)) = \frac{1}{|\det L|} \int f(x).$$

Here we require $L \in \mathrm{GL}(n)$, since if $\det L = 0$, the function $f(L(x))$ need not be measurable.
At last we take a look at Fubini's theorem with the convolution product.

Definition 3.1.5 (Convolution)**.** Let f, g be smooth functions with compact support, define their **convolution** to be

$$f * g = \int f(x-y) g(y) \,\mathrm{d}y.$$

Then $f * g$ is also a smooth function with compact support.

In fact we can generalize this definition for $f, g \in L^1$.
First note that $f(x-y), g(y)$ are measurable functions on $\mathbb{R}^{2d}$, by Tonelli's theorem,

$$\int_{\mathbb{R}^{2d}} |f(x-y)||g(y)| \,\mathrm{d}x \,\mathrm{d}y = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} |f(x-y)||g(y)| \,\mathrm{d}x \right) \mathrm{d}y = \|f\|_{L^1} \|g\|_{L^1} < +\infty.$$

This shows that $f(x-y)g(y)$ is integrable on $\mathbb{R}^{2d}$. Hence by Fubini's theorem $f(x-y)g(y)$ is integrable as a function of y, and $f * g$ is integrable on $\mathbb{R}^d$.
Moreover we have

$$\|f * g\|_{L^1} \leq \|f\|_{L^1} \|g\|_{L^1}.$$

The equality holds when both f and g are non-negative.
Fubini's theorem is also useful when computing integrals.

Example 3.1.6 (Gauss integral)
Recall the Gauss integral:

$$\int_0^{\infty} e^{-x^2} \,\mathrm{d}x = \frac{\sqrt{\pi}}{2}.$$

Here we give a different proof:

$$\begin{aligned} \int e^{-x^2} \,\mathrm{d}x \int e^{-y^2} \,\mathrm{d}y &= \iint_{\mathbb{R}^2} e^{-(x^2+y^2)} \,\mathrm{d}x \,\mathrm{d}y \\ &= \int_0^{+\infty} e^{-r^2} \,\mathrm{d}r^2 \cdot \pi = \pi. \end{aligned}$$

§4 Lebesgue differentiation

The most important theorem in calculus is no doubt the Fundamental theorem of Calculus (which is also called Newton-Lebniz formula). Since we generalized the integrals, there must be a generalized version of this theorem:

Theorem 4.0.1 (Lebesgue differentiation theorem, part 1)
If f is integrable on $\mathbb{R}^d$, for any ball $B \subset \mathbb{R}^d$, we have

$$\lim_{x\in B, |B|\to 0} \frac{1}{m(B)} \int_B f(y)\,\mathrm{d}y = f(x), a.e.$$

This theorem clearly holds for continuous points of f.

Our basic idea is to take a continuous g, such that $\|g - f\|_{\mathcal{L}^1} < \varepsilon$.
and to prove

$$\left\{ x : \limsup_{x\in B, |B|\to 0} \frac{1}{m(B)} \int_B |f(y) - f(x)|\,\mathrm{d}y \geq \varepsilon_0 \right\}$$

is a null set.

Now we estimate

$$\begin{aligned}\frac{1}{m(B)} \int_B |f(y) - f(x)|\,\mathrm{d}y &\leq \frac{1}{m(B)} \int_B \left(|f(y) - g(y)| + |g(y) - g(x)| + |g(x) - f(x)|\right) \mathrm{d}y \\ &= |f(x) - g(x)| + \varepsilon + \frac{1}{m(B)} \int_B |f(y) - g(y)|\,\mathrm{d}y\end{aligned}$$

We find that the last term is pretty hard to deal with, so we'll introduce some new tools:

Definition 4.0.2 (Hardy-Littlewood maximal function)**.** Let f be an integrable function on $\mathbb{R}^d$. Define

$$Mf(x) = \sup_{x\in B} \frac{1}{m(B)} \int_B |f(y)|\,\mathrm{d}y.$$

to be the **maximal function** of f.

Theorem 4.0.3 (Hardy-Littlewood)
The maximal function Mf satisfies:

- Mf is measurable;
- For x almost everywhere, $|f(x)| \leq Mf(x) < +\infty$.
- There exists a constant C s.t.

$$|\{x : Mf > \alpha\}| \leq \frac{C}{\alpha} \|f\|_{\mathcal{L}^1}.$$

Proof. First we prove $\{Mf > \alpha\}$ is measurable. If $Mf(x_0) > \alpha$, then exists an open ball $B \ni x_0$,

$$\int_B |f(y)|\,\mathrm{d}y > \alpha m(B).$$

This implies that $B \subset \{Mf > \alpha\} \implies \{Mf > \alpha\}$ is an open set.

For the second part, we'll prove for $\forall \varepsilon_0 > 0, N > 0$,

$$m(\{x : Mf(x) + \varepsilon_0 < |f(x)| \leq N\}) = 0.$$

Otherwise denote the above set as E, for $\forall 0 < \lambda < 1$, $\exists B$ s.t. $|E \cap B| > \lambda |B|$.

Thus for $x \in E$,

$$\begin{aligned} Mf(x) &\geq \frac{1}{m(B)} \int_B |f(y)| \,\mathrm{d}y \\ &\geq \frac{1}{m(B)} \int_{E\cap B} |f(y)| \,\mathrm{d}y \\ &\geq \frac{1}{m(B)} \int_{E\cap B} \varepsilon_0 + Mf(y) \,\mathrm{d}y \\ &= \frac{m(E\cap B)}{m(B)}\varepsilon_0 + \frac{1}{|B|} \int_{E\cap B} Mf(y) \,\mathrm{d}y. \end{aligned}$$

Taking the integral with respect to x:

$$\left(1 - \frac{|E\cap B|}{|B|}\right) \int_{E\cap B} Mf \geq \frac{|E\cap B|^2}{|B|}\varepsilon_0.$$

This implies $(1-\lambda)N \geq \lambda\varepsilon_0$, which is impossible as $\lambda \to 1$.

Now for the last part, since $\{Mf > \alpha\}$ is open, $\forall x \in \{Mf > \alpha\}$, $\exists B$ s.t.

$$\int_B |f(y)| \,\mathrm{d}y > \alpha m(B).$$

Hence for disjoint balls B_{i_k},

$$\|f\|_{\mathcal{L}^1} \geq \sum_{l=1}^{k} \int_{B_{i_l}} |f(y)| \,\mathrm{d}y > \alpha \sum_{l=1}^{k} |B_{i_l}|.$$

If we could select B_{i_l}'s such that their measure achieves say 1% of E, then we're done.

Lemma 4.0.4

Let $B_1, \dots, B_n$ be open balls in $\mathbb{R}^d$. There exists $i_1, \dots, i_k$ such that B_{i_j}'s are pairwise disjoint, and

$$\bigcup_{i=1}^{n} B_i \subset \bigcup_{j=1}^{k} 3B_{i_j}.$$

Here $3B$ means to multiply the radius of the ball by 3.

Proof of lemma. Trivial, just take the largest ball first and using greedy algorithm. □

Remark 4.0.5 — For countable many balls, the conculsion holds with 3 replaced by 5.

In particular, for all compact sets $K \subset \{Mf > \alpha\}$, there exists a finite open cover $B_1, B_2, \dots, B_n$ of K. By lemma we can select B_{i_j}'s satisfying

$$\sum_{i=1}^{k} m(B_{i_j}) \geq \frac{1}{3^d} m\left(\bigcup_{i=1}^{n} B_i\right) \geq \frac{1}{3^d} m(K).$$

Combining with our previous discussion we get $\|f\|_{\mathcal{L}^1} \geq \frac{\alpha}{3^d} m(K)$.

□

Returning to the proof of Theorem 4.0.1, we can assmue g is continuous with compact support,

$$\frac{1}{m(B)}\int_B |f(y)-g(y)|\,\mathrm{d}y \le M(f-g)(x)$$

Hence by taking B sufficiently small s.t. $|g(y)-g(x)| \le \varepsilon_0$ for all $x, y \in B$,

$$\frac{1}{m(B)}\int_B f(y)\,\mathrm{d}y \ge 3\varepsilon_0$$
$$\Longleftarrow \ |f(x)-g(x)| + M(f-g)(x) \ge 2\varepsilon_0.$$

But

$$m\{|f(x)-g(x)| \ge \varepsilon_0\} + m\{M(f-g) > \varepsilon_0\} \le \frac{\|f-g\|_{\mathcal{L}^1}}{\varepsilon_0} + \frac{3^d}{\varepsilon_0}\|f-g\|_{\mathcal{L}^1} \le \frac{3^d+1}{\varepsilon_0}\varepsilon.$$

This completes the proof.

Definition 4.0.6 (Lebesgue points)**.** Let $|f(x)| < \infty$, f is *locally integrable.* If x satisfies

$$\lim_{|B|\to 0, B\ni x} \frac{1}{|B|}\int_B |f(y)-f(x)|\,\mathrm{d}y = 0,$$

we say x is a **Lebesgue point** of f.

Remark 4.0.7 — Here "locally integrable" means for all bounded measurable sets E, $f\chi_E \in \mathcal{L}^1$. This is denoted by $f \in \mathcal{L}^1_{loc}$.

Let E be a measurable set, χ_E locally integrable, If point x is called a **density point** of E if it's a Lebesgue point of χ_E.

Theorem 4.0.8

Let E be a measurable set, then almost all the points in E are density points of E, almost all the points outside of E are not density points of E.

Proof. This is a direct corollary of Theorem 4.0.1. □

The differentiation theorem has some applications in convolution:

$$\begin{aligned}\frac{1}{|B|}\int_B f(y)\,\mathrm{d}y &= c_d^{-1}\varepsilon^{-d}\int_{B(x,\varepsilon)} f(y)\,\mathrm{d}y \\ &= \int f(x-y)\cdot c_d^{-1}\varepsilon^{-d}\chi_{B(0,\varepsilon)}(y)\,\mathrm{d}y \\ &= f * K_\varepsilon.\end{aligned}$$

where $K_\varepsilon = c_d^{-1}\varepsilon^{-d}\chi_{B(0,\varepsilon)}$, c_d is the measure of a unit sphere in $\mathbb{R}^d$.

By differentiation theorem, $\lim_{\varepsilon\to 0} f * K_\varepsilon = f(x), a.e.$. In the homework we proved that there doesn't exist a function I s.t. $f * I = f$ for all $f \in \mathcal{L}^1$, but the functions K_ε is approximating this "convolution identity".

Definition 4.0.9. In general, if $\int K_\varepsilon = 1$, $|K_\varepsilon| \le A\min\{\varepsilon^{-d}, \varepsilon|x|^{-d-1}\}$ for some constant A, we say K_ε is an **approximation to the identity**.

"convolution kernel"

Let φ be a smooth function whose support is in $\{|x| \leq 1\}$, and $\int \varphi = 1$. The function $K_\varepsilon := \varepsilon^{-d}\varphi(\varepsilon^{-1}x)$ is called the Friedrichs smoothing kernel.

Theorem 4.0.10

If K_ε is an approximation to the identity, f integrable,

$$\lim_{\varepsilon\to 0}\|f * K_\varepsilon - f\|_{\mathcal{L}^1} = 0.$$

Proof.

$$\begin{aligned}
|(f * K_\varepsilon)(x) - f(x)| &= \left|\int f(x-y)K_\varepsilon(y)\,\mathrm{d}y - f(x)\right| \\
&\leq \int |f(x-y) - f(x)||K_\varepsilon(y)|\,\mathrm{d}y \\
&\leq \int_{|y|\leq R} |f(x-y) - f(x)|A\varepsilon^{-d}\,\mathrm{d}y + \int_{|y|>R} |f(x-y) - f(x)|A\varepsilon|y|^{-d-1}\,\mathrm{d}y.
\end{aligned}$$

Taking the integral over $\mathbb{R}^d$:

$$\begin{aligned}
&\|K_\varepsilon * f - f\|_{\mathcal{L}^1} \\
&\leq A\varepsilon^{-d}\int\int_{|y|\leq R} |f(x-y) - f(x)|\,\mathrm{d}y\,\mathrm{d}x + A\varepsilon\int\int_{|y|>R} |f(x-y) - f(x)||y|^{-d-1}\,\mathrm{d}y\,\mathrm{d}x \\
&\leq A\varepsilon^{-d}\int\int_{|y|\leq R} |\tau_{-y}f(x) - f(x)|\,\mathrm{d}y\,\mathrm{d}x + A\varepsilon\int_{|y|>R} |y|^{-d-1}\int |\tau_{-y}f(x)| + |f(x)|\,\mathrm{d}x\,\mathrm{d}y \\
&\leq A\varepsilon^{-d}\int_{|y|\leq R} \|\tau_{-y}f - f\|_{\mathcal{L}^1}\,\mathrm{d}y + A\varepsilon\int_{|y|>R} |y|^{-d-1}2\|f\|_{\mathcal{L}^1}\,\mathrm{d}y.
\end{aligned}$$

By the continuity of translation, $\forall\varepsilon_0$, let R be sufficiently small we have

$$\|\tau_{-y}f - f\|_{\mathcal{L}^1} < \varepsilon_0, \forall|y| < R.$$

$$\|K_\varepsilon * f - f\|_{\mathcal{L}^1} \leq A\varepsilon^{-d}R^d c_d\varepsilon_0 + \varepsilon R^{-1}C$$

where C is a constant.

Take suitable $\varepsilon, \varepsilon_0$ s.t. $R = \varepsilon\varepsilon_0^{-\frac{1}{2d}}$, then $LHS \leq C_1\varepsilon_0^{\frac{1}{2}} + C_2\varepsilon_0^{\frac{1}{2d}} \to 0$. □

Theorem 4.0.11

Let K_ε be an approximation to the identity, f integrable,

$$\lim_{\varepsilon\to 0} f * K_\varepsilon = f(x)$$

holds for Lebesgue points x of f.

Proof. WLOG $x = 0$, let

$$\omega(r) = \frac{1}{r^d}\int_{B(0,r)} |f(y) - f(0)|\,\mathrm{d}y,$$

we have $\lim_{r\to 0}\omega(r) = 0$, and ω is continuous.

$$\omega(r) \le \frac{\|f\|_{\mathcal{L}^1}}{r^d} + |f(0)|,$$

Thus ω is bounded.

Therefore we can compute

$$\begin{aligned}
|K_\varepsilon * f(x) - f(x)| &\le \int |K_\varepsilon(y)||f(x-y) - f(x)|\,\mathrm{d}y \\
&\le \int_{B(0,r)} |K_\varepsilon(y)||f(x-y) - f(x)|\,\mathrm{d}y + \sum_{k\ge 0}\int_{2^k r\le |y| < 2^{k+1}r} |K_\varepsilon(y)||f(x-y) - f(x)|\,\mathrm{d}y \\
&\le A\varepsilon^{-d}\cdot r^d\omega(r) + \sum_{k\ge 0}\int_{2^k r\le |y|<2^{k+1}r} A\varepsilon|y|^{-d-1}|f(x-y) - f(x)|\,\mathrm{d}y \\
&\le A\varepsilon^{-d}r^d\omega(r) + \sum_{k\ge 0} A\varepsilon 2^{-k(d+1)}r^{-d-1}\cdot 2^{(k+1)d}r^d\omega(2^{k+1}r) \\
&= A\varepsilon^{-d}r^d\omega(r) + A2^d\varepsilon r^{-1}\sum_{k\ge 0} 2^{-k}\omega(2^{k+1}r).
\end{aligned}$$

Let $r = \varepsilon$, since $\omega(r)$ is continuous and bounded, we're done. □

§4.1 Lebesgue Differentiation theorem for monotone functions

Recall that Fundamental theorem of Calculus has a second part, Given any differentiable function $F(x)$, if $F'(x)$ Riemann integrable, then

$$F(b) - F(a) = \int_a^b F'(x)\,\mathrm{d}x.$$

When F has nice properties (smooth), then it has a generalization to higher dimensions (Stokes' theorem).

But in Lebesgue integration theory, the derivative of a function is hard to define. Hence we'll find an alternative for $F'(x)$.

Example 4.1.1

Consider Heaviside function $H(t) = \begin{cases} 1, & t > 0 \\ 0, & t \le 0 \end{cases}$

Then H is differentiable almost everywhere, but $\int_{-1}^1 H'(t)\,\mathrm{d}t = 0 \ne H(1) - H(-1)$.

Example 4.1.2

Consider Cantor-Lebesgue function F, similarly we have $F'(x) = 0, a.e.$, but $\int_0^1 F'(x)\,\mathrm{d}x = 0 \ne F(1) - F(0)$.

Definition 4.1.3 (Dini derivatives)**.** Let $f(x)$ be a measurable function, define

$$D^+(f)(x) = \limsup_{h>0,h\to 0} \frac{f(x+h)-f(x)}{h}, \quad D^-(f)(x) = \limsup_{h<0,h\to 0} \frac{f(x+h)-f(x)}{h}.$$

$$D_+(f)(x) = \liminf_{h>0,h\to 0} \frac{f(x+h)-f(x)}{h}, \quad D_-(f)(x) = \liminf_{h<0,h\to 0} \frac{f(x+h)-f(x)}{h}.$$

Theorem 4.1.4 (Lebesgue Differentiation theorem for increasing functions)
Let f be an increasing function on $[a,b]$, then $F'(x)$ exists almost everywhere, and

$$\int_a^b F'(x)\,\mathrm{d}x \le F(b) - F(a).$$

Sketch of the proof. The outline of the proof is as follows:
Step 1: Decompose $F = F_c + J$, where F_c is continuous, J is a jump function.
Step 2: Prove F_c increasing and $J' = 0, a.e.$.
Step 3: Prove $D^+(F) < +\infty, a.e.$, $D^+(F) \le D_-(F), a.e.$, and $D^-(F) \le D_+(F), a.e.$. □

We proceed step by step.

Step 1 Denote $F(x+0) = \lim_{h\to 0^+} F(x+h), F(x-0) = \lim_{h\to 0^-} F(x+h)$.
Since F increasing, let $\{x_n\}$ be all the discontinuous points of F. Define:

$$j_n(x) = \begin{cases} 0, & x < x_n \\ \beta_n, & x = x_n \\ \alpha_n, & x > x_n \end{cases}$$

where $\alpha_n = F(x_n+0) - F(x_n-0), \beta_n = F(x_n) - F(x_n-0)$.
Hence the jump function

$$J_F(x) = \sum_{n=1}^{+\infty} j_n(x) \le \sum_{n=1}^{+\infty} \alpha_n = \sum_{n=1}^{+\infty} (F(x_n+0) - F(x_n-0)) \le F(b) - F(a)$$

is well-defined and increasing.

Lemma 4.1.5
$F - J_F$ is continuous and increasing.

Proof. First note that

$$\lim_{h\to 0^+} (F(x+h) - J_F(x+h)) = F(x+0) - \lim_{h\to 0^+} J_F(x+h) = F(x-0) - \lim_{h\to 0^+} J_F(x-h)$$

This can be derived from the definition of J_F: If F is continuous at x, the equality is obvious;
If $x = x_n$ for some n,

$$\lim_{h\to 0^+} J_F(x+h) = \sum_{x_k \le x_n} \alpha_k + \lim_{h\to 0^+} \sum_{x_n < x_k \le x_n + h} j_k(x+h) = \sum_{x_k \le x_n} \alpha_k$$

$$\lim_{h\to 0^+} J_F(x-h) = \lim_{h\to 0^+} \sum_{x_k<x_n-h} \alpha_k + \lim_{j\to 0^+} \sum_{x_k=x_n-h} \beta_k = \sum_{x_k<x_n} \alpha_k$$

Note that $\alpha_n = F(x_n+0) - F(x_n-0)$, thus $F - J_F$ is continuous.

Secondly,

$$F(x) - J_F(x) \le F(y) - J_F(y), \quad \forall a \le x \le y \le b.$$

Because

$$J_F(y) - J_F(x) = \sum_{x<x_j<y} \alpha_j + \sum_{x_k=y} \beta_k - \sum_{x_k=x} \beta_k \le \sum_{x<x_j<y} \alpha_j + F(y) - F(y-0) \le F(y) - F(x).$$

which means $F - J_F$ is increasing. □

Step 2

Proposition 4.1.6

The jump function $J(x)$ is differentiable almost everywhere, and $J'(x) = 0, a.e.$.

Proof. The Dini derivatives of $J(x)$ exist and are non-negative (since J is increasing).

$$\overline{D}(J)(x) = \max\{D^+(J)(x), D^-(J)(x)\}.$$

Let $E_\varepsilon = \{\overline{D}(J)(x) > \varepsilon > 0\}$. We'll prove E_ε is null for all ε. If $x \in E_\varepsilon$, $\exists h$ s.t.

$$\frac{J(x+h) - J(x)}{h} > \varepsilon \implies J(x+h) - J(x-h) > \varepsilon h.$$

Let $N \in \mathbb{N}$ s.t. $\sum_{n>N} \alpha_n < \frac{\varepsilon\delta}{10}$. Define $J_N(x) = \sum_{n>N} j_n(x)$.

$$E_{\varepsilon,N} = \{\overline{D}(J_N)(x) > \varepsilon\}, \quad E_\varepsilon \subset E_{\varepsilon,N} \cup \{x_1, \dots, x_N\},$$

Since for $x \ne x_i$,

$$\begin{aligned}\overline{D}(J)(x) &= \limsup_{h\to 0} \frac{J(x+h) - J(x)}{h} \\ &= \limsup_{h\to 0} \left(\frac{J_N(x+h) - J_N(x)}{h} + \frac{1}{h} \sum_{n=1}^{N} (j_n(x+h) - j_n(x)) \right) = \overline{D}(J_N)(x).\end{aligned}$$

Next we need to control the measure of $E_{\varepsilon,N}$.

For all $y \in E_{\varepsilon,N}$, there exists sufficiently small h s.t. $J_N(y+h) - J_N(y) > h\varepsilon$. So the intervals $(y-h, y+h)$ is a covering of $E_{\varepsilon,N}$, and it can be controled using the value of J_N. Therefore we hope to find some *disjoint* intervals which cover certain ratio of $E_{\varepsilon,N}$.

Lemma 4.1.7

Let $\mathcal{B}$ be a collection of balls with bounded radius in $\mathbb{R}^d$. There exists countably many disjoint balls B_i s.t.

$$\bigcup_{B\in\mathcal{B}} B \subset \bigcup_{i=1}^{\infty} 5B_i.$$

Proof. Let $r(B)$ denote the radius of B. Take B_1 s.t. $r(B_1) > \frac{1}{2}\sup_{b\in\mathcal{B}} r(B)$.
The rest is the same as before. $\square$

By lemma, there exists countably many disjoint intervals $(x_i + h_i, x_i - h_i)$ s.t.

$$\begin{aligned} m^*(E_{\varepsilon,N}) &\le 5\sum_{i=1}^{\infty} 2h_i \\ &\le 10\sum_{i=1}^{\infty} \varepsilon^{-1}(J_N(x_i+h_i) - J_N(x_i-h_i)) \\ &\le 10\varepsilon^{-1}(J_N(b) - J_N(a)) < \delta. \end{aligned}$$

Hence $m^*(E_\varepsilon) \le m^*(E_{\varepsilon,N}) < \delta \implies m^*(E_\varepsilon) = 0$, which gives $\overline{D}(J) = 0, a.e.$. $\square$

Step 3 First we prove $D^+(F) < \infty, a.e.$.
Let $E_\gamma = \{x : D^+(F)(x) > \gamma\}$.
When $h \in [\frac{1}{n+1}, \frac{1}{n}]$:

$$\begin{aligned} \frac{F(x+h) - F(x)}{h} &\le \frac{n+1}{n}\frac{F(x+\frac{1}{n}) - F(x)}{\frac{1}{n}}, \\ &\ge \frac{n}{n+1}\frac{F(x+\frac{1}{n+1}) - F(x)}{\frac{1}{n+1}}. \end{aligned}$$

Thus

$$D^+(F)(x) = \limsup_{h\to 0}\frac{F(x+h)-F(x)}{h} = \limsup_{n\to\infty}\frac{F(x+\frac{1}{n}) - F(x)}{\frac{1}{n}}$$

is measurable.

Lemma 4.1.8 (Riesz sunrise lemma)
Let $G(x)$ be a continuous function on $\mathbb{R}$. Define

$$E = \{x : \exists h > 0, s.t.\ G(x+h) > G(x)\}.$$

Then E is open and $E = \bigcup_{i=1}^{\infty}(a_i, b_i)$, where (a_i, b_i) are disjoint finite intervals s.t. $G(a_i) = G(b_i)$.
When G is defined on finite interval $[a, b]$, we also have $G(a) \le G(b_1)$.

Proof. Note that E is open since G is continuous.
Take a maximum open interval $(a, b) \subset E$, i.e. $a, b \notin E$, so $G(a) \ge G(b)$.
Since $b \notin E, G(x) \le G(b), \forall x > b$. If $G(a) > G(b)$, Let $G(a+\varepsilon) > G(b)$, as $a+\varepsilon \in E$, exists $h > 0$ s.t. $G(a+\varepsilon+h) > G(a+\varepsilon)$.
But G has a maximum on $[a+\varepsilon, b]$, say $G(c)$, we must have $c \ne a+\varepsilon, b$. This leads to a contradiction. $\square$

Remark 4.1.9 — This lemma provides a better estimation than previous covering lemmas, since it directly claims that E can be broken into disjoint intervals.

For $x \in E_\gamma$, $\exists h > 0$ s.t. $F(x+h) - F(x) > \gamma h$, by Lemma 4.1.8 on $F(x) - \gamma x$,

$$m(E_\gamma) \le \sum_{k=1}^{\infty}(b_k - a_k) \le \gamma^{-1}\sum_{k=1}^{\infty}(F(b_k) - F(a_k)) \le \gamma^{-1}(F(b) - F(a)).$$

Therefore when $\gamma \to \infty$, $m(E_\gamma) \to 0$.

The last part is $D^+(F) \le D_-(F), a.e.$.

Similarly it's sufficient to prove the following set is null for all rational numbers $r < R$:

$$E_{r,R} = \{D^+(F)(x) > R, D_-(F)(x) < r\}.$$

Since $D^+(F)$ is measurable, $E_{r,R}$ is measurable. If $m(E_{r,R}) > 0$, we can restrict it to a smaller interval $[c,d] \subset [a,b]$ such that $d - c < \frac{R}{r}m(E_{r,R})$.

Let $G(x) = F(-x) + rx$, by Lemma 4.1.8 on $[-d,-c]$,

$$\{s : \exists h > 0, G(x+h) > G(x)\} = \bigcup_k (-b_k, -a_k).$$

Note that $-E_{r,R}$ is contained in the above set, and $G(-b_k) \le G(-a_k) \iff F(b_k) - F(a_k) \le r(b_k - a_k)$,

We use Lemma 4.1.8 again on each (a_k, b_k) and $F(x) - Rx$,

$$E_{r,R} \cap (a_k, b_k) \subset \{x : \exists h > 0, F(x+h) - F(x) \ge Rh\} = \bigcup_{l=1}^{\infty}(a_{k,l}, b_{k,l}).$$

Hence

$$\begin{aligned} m(E_{r,R}) &\le \sum_{k,l=1}^{\infty}(b_{k,l} - a_{k,l}) \\ &\le R^{-1}\sum_{k,l=1}^{\infty}(F(b_{k,l}) - F(a_{k,l})) \le R^{-1}\sum_{k=1}^{\infty}(F(b_k) - F(a_k)) \\ &\le R^{-1}r\sum_{k=1}^{\infty}(b_k - a_k) \le R^{-1}r(d-c), \end{aligned}$$

which gives a contradiction! So $m(E_{r,R}) = 0$ for all rationals $r < R$. Therefore we're done by

$$m(\{D^+(F) > D_-(F)\}) \le \sum_{r,R} m(E_{r,R}) = 0$$

Now we can complete the proof of Theorem 4.1.4. Here we state the theorem again:

Let F be an increasing function on $[a,b]$, then F is differentiable almost everywhere, and

$$\int_a^b F'(x)\,\mathrm{d}x \le F(b) - F(a).$$

Let $F_n(x) = n(F(x + \frac{1}{n}) - F(x))$, where $F(x) = F(b)$ for $x > b$. Since $F_n \ge 0$, by Fatou's

Lemma, (we've already proved F is differentiable almost everywhere and $F' \geq 0$)

$$
\begin{aligned}
\int_a^b \liminf_{n\to\infty} F_n \,\mathrm{d}x &\leq \liminf_{n\to\infty} \int_a^b F_n \,\mathrm{d}x \\
\implies \int_a^b F'(x)\,\mathrm{d}x &\leq \liminf_{n\to\infty} \int_a^b n\left(F\left(x+\frac{1}{n}\right) - F(x)\right)\mathrm{d}x \\
&= \liminf_{n\to\infty} n\left(\int_{a+\frac{1}{n}}^{b+\frac{1}{n}} F(x) - \int_a^b F(x)\right)\mathrm{d}x \\
&= \liminf_{n\to\infty}\left(F(b) - n\int_a^{a+\frac{1}{n}} F(x)\,\mathrm{d}x\right) \\
&\leq F(b) - F(a)
\end{aligned}
$$

§4.2 Absolute continuous functions

Definition 4.2.1 (Absolute continuity)**.** We say a function $F(x)$ is **absolutely continuous** on interval $[a,b]$, if $\forall \varepsilon > 0, \exists \delta > 0$, such that for all disjoint intervals $(a_k, b_k), k = 1, \dots, N$ with

$$\sum_{k=1}^{N}(b_k - a_k) < \delta,$$

we must have

$$\sum_{k=1}^{N} |F(b_k) - F(a_k)| < \varepsilon.$$

The space consisting of all the absolutely continuous functions on $[a,b]$ is denoted by $Ac([a,b])$.

Example 4.2.2

A C^1 function with bounded derivative or a Lipschtiz function is absolutely continuous.

Some obvious properties of absolutely continuous function F:

- F is continuous;
- F has bounded variation, i.e. $F \in BV$.
- F is differentiable almost everywhere, since $F = F_1 - F_2$, where F_1, F_2 are increasing. In fact we have
$$T_F([a,b]) = \int_a^b |F'(x)|\,\mathrm{d}x.$$
- If N is a null set, then $F(N)$ is also null. In particular F maps measurable sets to measurable sets.

Proof of the last property. Take intervals (a_k, b_k) s.t. $N \subset \bigcup_{k=1}^{\infty}(a_k, b_k)$. Since $F(N) \subset F(\bigcup(a_k, b_k))$,

$$|F(N)| \leq \sum_{k=1}^{\infty} |F([a_k, b_k])| \leq \sum_{k=1}^{\infty} |F(\tilde{b}_k) - F(\tilde{a}_k)| < \varepsilon.$$

$\square$

Proposition 4.2.3

The space $Ac([a,b]) \subset BV([a,b])$, moreover it's an algebra, and it's a separable Banach space under the norm induced from BV.

Finally we come to the full generalization of Newton-Lebniz formula:

Theorem 4.2.4 (Fundamental theorem of Calculus)

A function $F \in Ac([a,b]) \implies F$ is differentiable almost everywhere, F' is integrable, and

$$F(x) - F(a) = \int_a^x F'(\tilde{x})\,\mathrm{d}\tilde{x}, \quad \forall x \in [a,b].$$

Proof. Let $\tilde{F}(x) = F(a) + \int_a^b F'(y)\,\mathrm{d}y \in Ac([a,b])$ (by the absolute continuity of integrals).

We have $F - \tilde{F} \in Ac([a,b])$ and $(F - \tilde{F})' = 0, a.e.$.

Thus it suffices to prove the following theorem:

Theorem 4.2.5

Let $F \in Ac([a,b])$, and $F' = 0, a.e.$, then $F(a) = F(b)$, i.e. F is constant on $[a,b]$.

To prove this, we'll need Vitali covering theorem:

Definition 4.2.6 (Vitali covering)**.** Let $\mathcal{B} = \{B_\alpha\}$, where B_α are closed balls in $\mathbb{R}^d$. We say $\mathcal{B}$ is a **Vitali covering** of a set E, if $\forall x \in E, \forall \eta > 0$, exists $B_\alpha \in \mathcal{B}$ s.t. $m(B_\alpha) < \eta$, $x \in B_\alpha$.

Theorem 4.2.7 (Vitali)

Let $E \subset \mathbb{R}^d$ with $m^*(E) < \infty$, for any Vitali covering $\mathcal{B}$ of E and $\delta > 0$, exists disjoint balls $B_1, \ldots, B_n \in \mathcal{B}$, such that

$$m^*\left(E \backslash \bigcup_{i=1}^n B_i\right) < \delta.$$

Proof. For all $\varepsilon > 0$, exists an open set A s.t. $E \subset A$ and $m(A) < m^*(E) + \varepsilon < +\infty$.

Remove all the balls in $\mathcal{B}$ with radius greater than 1. Each time we take a ball B_i with radius greater than $\frac{1}{2}\sup_{B \in \mathcal{B}'} r(B)$, where $\mathcal{B}'$ are the remaining balls, and remove all the balls which intersect with B_i.

If we end up with finitely many balls $B_1, \ldots, B_n$, we must have $E \subset \bigcup_{i=1}^n B_i$, otherwise $x \notin B_i \implies \exists B \ni x$, $B \cap B_i = \emptyset$, contradiction!

If we take out countably many balls $B_1, B_2, \cdots \subset A$, since $\sum_{i=1}^\infty m(B_i) \le m(A) < \infty$, there exists N s.t. $\sum_{i=N+1}^\infty m(B_i) < 5^{-d}\delta$.

Now we only need to prove

$$E \backslash \bigcup_{i=1}^N B_i \subset \bigcup_{i>N} 5B_i.$$

□

Let $E = \{x : F'(x) = 0\}$, $\forall x \in E$, $\exists \delta(x) > 0$, s.t.

$$|F(y) - F(x)| < \varepsilon|y - x|, \forall |y - x| < \delta(x).$$

Hence $[x - h, x + h], 0 < h < \delta(x)$ is a Vitali covering of E. By Theorem 4.2.7, there exists finitely many disjoint intervals $[x_k - h_k, x_k + h_k] = I_k$ s.t.

$$m^*\left(E \backslash \bigcup_{k=1}^{N} I_k\right) < \varepsilon.$$

Assmue $a \leq a_1 < b_1 < \cdots < a_N < b_N \leq b$, by absolute continuity and $|F(b_k) - F(a_k)| < \varepsilon(b_k - a_k)$,

$$F(b) - F(a) \leq \sum_{k=1}^{N} |F(b_k) - F(a_k)| + \sum_{k=0}^{N} |F(a_{k+1}) - F(b_k)| \leq \varepsilon(b - a) + \delta.$$

Here we complete the proof of the generalized Fundamental theorem of Calculus. □

There's another version of this thoerem which looks like Newton-Lebniz formula more:

Theorem 4.2.8

Let F be a differentiable function on $[a, b]$, if F' is Lebesgue integrable, then

$$F(b) - F(a) = \int_a^b F'(x)\,\mathrm{d}x.$$

We need to prove a lemma first.

Theorem 4.2.9

Let F be real function on $[a, b]$, if F is differentiable on E, and $|F'| \leq M$ in E, then

$$m^*(F(E)) \leq M m^*(E).$$

Proof. For all $\varepsilon > 0$, $x \in E$, $\exists \delta > 0$,

$$\left|\frac{F(x + h) - F(x)}{h} - M\right| < \varepsilon, \quad \forall |h| < \delta.$$

So $[x - h, x + h]$ is a Vitali covering of E. By Vitali's theorem (4.2.7), exists disjoint intervals $I_i = [x_i - h_i, x_i + h_i]$ s.t.

$$m^*\left(E \backslash \bigcup_{i=1}^{\infty} I_i\right) = 0, \quad \sum_{i=1}^{\infty} 2h_i \leq m^*(E) + \varepsilon.$$

But for $y \in I_i$, $|F(y) - F(x_i)| \leq (M + \varepsilon)h_i$, thus $m^*(F(I_i)) \leq 2(M + \varepsilon)h_i = (M + \varepsilon)|I_i|$.

$$\begin{aligned}
m^*(F(E)) &\leq m^*(F(E \cap \bigcup_{i=1}^{\infty} I_i)) + m^*(F(E \backslash \bigcup_{i=1}^{\infty} I_i)) \\
&\leq \sum_{i=1}^{\infty} m^*(F(I_i)) + m^*(F(E \backslash \bigcup_{i=1}^{\infty} I_i)) \\
&\leq (M + \varepsilon)(m^*(E) + \varepsilon) + m^*(F(E \backslash \bigcup_{i=1}^{\infty} I_i))
\end{aligned}$$

So it suffices to prove the case when E is null. Define

$$E_n = \left\{ x \in E : |F(y) - F(x)| \le (M+\varepsilon)|y-x|, \forall |y-x| < \frac{1}{n} \right\}.$$

Observe that $E_n \nearrow E$ and $F(E_n) \nearrow F(E)$. There exists disjoint intervals $J_{n,k}$ s.t.

$$E_n \subset \bigcup_{k=1}^{\infty} J_{n,k}, \quad \sum_{k=1}^{\infty} |J_{n,k}| \le \min\left\{\frac{1}{n}, \varepsilon\right\}.$$

Thus

$$m^*(F(E_n)) \le \sum_{k=1}^{\infty} m^*(F(E_n \cap J_{n,k})) \le \sum_{k=1}^{\infty} (M+\varepsilon)|J_{n,k}| \le \varepsilon(M+\varepsilon).$$

Taking $\varepsilon \to 0$ we get $F(E_n)$ is null. So $F(E) = \lim_{n\to\infty} F(E_n)$ is null, which completes the proof. □

Returning to the proof of the theorem, in fact we only need to prove

$$|F(b) - F(a)| \le \int_a^b |F'(x)| \, \mathrm{d}x,$$

since this implies F is absolutely continuous. For all $\varepsilon > 0$, let

$$E_n = \{x \in [a,b] : n\varepsilon \le |F'(x)| < (n+1)\varepsilon\}.$$

By our lemma, $m^*(F(E_n)) \le (n+1)\varepsilon m(E_n) \le \varepsilon m(E_n) + \int_{E_n} |F'(x)| \, \mathrm{d}x$.

Hence

$$\begin{aligned} |F(b) - F(a)| \le m(F([a,b])) &\le \sum_{n=0}^{\infty} m^*(F(E_n)) \\ &\le \varepsilon(b-a) + \int_a^b |F'(x)| \, \mathrm{d}x. \end{aligned}$$

Theorem 4.2.10

A rectifiable curve $\gamma(t) = (x(t), y(t))$ with x, y absolutely continuous has length

$$L(\gamma) = \int_a^b |\gamma'(t)| \, \mathrm{d}t.$$

Proof. Since $|\gamma(t_i) - \gamma(t_{i-1})| = |\int_{t_{i-1}}^{t_i} \gamma'(t) \, \mathrm{d}t| \le \int_{t_{i-1}}^{t_i} |\gamma'(t)| \, \mathrm{d}t$, thus $L(\gamma) \le \int_a^b |\gamma'(t)| \, \mathrm{d}t$.

$\forall \varepsilon > 0$, we can take a step function (with vector values) g s.t. $\gamma' = g + h$, and $\int_a^b |h| \, \mathrm{d}x < \varepsilon$.

Define

$$G(x) = G(a) + \int_a^x g(t) \, \mathrm{d}t, \quad H(x) = H(a) + \int_a^x h(t) \, \mathrm{d}t.$$

We have $\gamma(t) = G(t) + h(t)$, and $T_\gamma([a,b]) \ge T_G([a,b]) - T_H([a,b])$.

$$\begin{aligned} L(\gamma) = T_\gamma([a,b]) &\ge \int_a^b |g| \, \mathrm{d}t - \int_a^b |h| \, \mathrm{d}t \\ &\ge \int_a^b |\gamma'(t)| \, \mathrm{d}t - 2\int_a^b |h| \, \mathrm{d}t \\ &\ge \int_a^b |\gamma'(t)| \, \mathrm{d}t - 2\varepsilon. \end{aligned}$$

which gives the opposite inequality. □

Proposition 4.2.11 (substitution formula)
Let $\phi : [a, b] \to [c, d]$ be strictly increasing AC function. For a function f on $[c, d]$, we have

$$\int_c^d f(y)\,\mathrm{d}y = \int_a^b f(\phi(x))\phi'(x)\,\mathrm{d}x.$$

Proof. It's equivalent to $m(\phi(E)) = \int_E \phi'\,\mathrm{d}x$. □

§5 Multi-dimensional Calculus

In this section we'll generalize the differentiation and integration theory to higher dimensions, and finally reach the generalized Fundamental Theorem of Calculus (Stokes' formula). Recall that Lebesgue integral is already defined on higher dimensions, so here we mainly study the differentiation of multi-dimensional functions.

§5.1 Directional derivatives

Let Ω be a simply connected open set in $\mathbb{R}^d$. f is a multi-variable function on Ω. Let $(x_1, \ldots, x_n)$ be a coordinate system on Ω, we can write $f = f(x_1, \ldots, x_n)$.

Definition 5.1.1 (Directional derivatives)**.** Let $v \in \mathbb{R}^d$ be a nonzero vector. If

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

exists, then we say the **directional derivative** of f in direction v exists at x_0, denoted by

$$\frac{\partial f}{\partial v}(x_0) = (\nabla_v f)(x_0) = \lim_{h\to 0} \frac{f(x_0 + h) - f(x_0)}{h}.$$

Definition 5.1.2 (Partial derivatives)**.** Let $(x_1, \ldots, x_n)$ be a coordinate system, let $e_i = (0, \ldots, 1, \ldots, 0)$ be the i-th vector of the standard basis. The directional derivative in e_i

$$(\nabla_{e_i} f)(x_0) = \frac{\partial f}{\partial x_i}(x_0) = \lim_{h\to 0} \frac{f(x_1, \ldots, x_i + h, \ldots, x_n) - f(x_1, \ldots, x_n)}{h}.$$

is called the i-th **partial derivative** of f. Here $\frac{\partial}{\partial x_i}$ is also called a "vector field".

Remark 5.1.3 — The partial derivatives rely on the coordinate, but the directional derivatives is independent of the coordinate (i.e. geometry quantities).

Example 5.1.4
Let $f : \mathbb{R}^2 \to \mathbb{R}$, and $f(x, y) = g(x)$ for some g.

$$\frac{\partial f}{\partial y} = \lim_{h\to 0} \frac{f(x, y + h) - f(x, y)}{h} = 0.$$

Example 5.1.5
Consider $f : \mathbb{R}^2 \to \mathbb{R}$,

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

The partial derivative

$$\frac{\partial f}{\partial x}(0, 0) = \lim_{h \to 0} \frac{f(h, 0) - f(0, 0)}{h} = 0 = \frac{\partial f}{\partial y}(0, 0).$$

But the directional derivative in $v = (v_1, v_2)$ is

$$(\nabla_v f)(0, 0) = \lim_{h \to 0} \frac{f(hv_1, hv_2) - f(0, 0)}{h} = \lim_{h \to 0} \frac{v_1 v_2}{h(v_1^2 + v_2^2)},$$

which doesn't exist for $v_1 v_2 \neq 0$.

The main idea of differentiation in 1 dimensional is to estimate a function locally using a straight line. Likely, in higher dimensions, the differentiation is also estimating a function locally using a *linear map*.

Definition 5.1.6 (Differentiation)**.** Let $f : \Omega \to \mathbb{R}$, $x_0 \in \Omega$. If there exists a linear map $A : \mathbb{R}^d \to \mathbb{R}$ s.t.

$$f(x_0 + v) = f(x_0) + A(v) + o(|v|) \iff \lim_{|v| \to 0} \frac{|f(x_0 + v) - f(x_0) - A(v)|}{|v|} = 0,$$

then we say f is **differentiable** at x_0, and the linear map A is called the **differentiation** of f at x_0, denoted by

$$\mathrm{d}f\big|_{x_0} = \mathrm{d}f(x_0) = A : \mathbb{R}^d \to \mathbb{R}.$$

If f is differentiable everywhere, we say f is a differentiable function.

Remark 5.1.7 — In fact this definition can be generalized to any Banach space. Keep in mind that $\mathrm{d}f(x_0)$ is a *linear map* instead of a number, the reason why the one dimensional differentiation is a number is that a linear map in one dimension is identical to a scalar.

Theorem 5.1.8
Let f be a function differentiable at x_0, then its directional derivatives exist at x_0, $\forall v \in \mathbb{R}^d$,

$$(\nabla_v f)(x_0) = (\mathrm{d}f(x_0))(v) = \sum_{i=1}^{d} \frac{\partial f}{\partial x_i} \cdot v_i = \nabla f \cdot v.$$

where $\nabla f = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})$ is the **gradient vector** of f.

Proof. Note that

$$\frac{f(x_0 + hv) - f(x_0)}{h} = \frac{\mathrm{d}f(x_0)(hv) + o(h|v|)}{h} \to \mathrm{d}f(x_0)(v).$$

$$\mathrm{d}f(x_0)(v) = \mathrm{d}f(x_0)\left(\sum_{i=1}^{d} v_i e_i\right) = \sum_{i=1}^{d} v_i \, \mathrm{d}f(x_0)(e_i). = \sum_{i=1}^{d} v_i \frac{\partial f}{\partial x_i}.$$

□

Example 5.1.9

Let $f : \mathbb{R}^2 \to \mathbb{R}$,

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

Note that the directional derivatives of f exists at $(0,0)$, but f is not continuous at x_0, so not differentiable.

Theorem 5.1.10

Let $\Omega \subset \mathbb{R}^d$. If the partial derivatives of f exists and are continuous at x_0, then f is differentiable at x_0.

Proof. Let $u_j = (v_1, \ldots, v_j, 0, \ldots, 0)$.

$$\begin{aligned} f(x_0+v) - f(x_0) - (\nabla f)(x_0) \cdot v &= \sum_{j=1}^{d} f(x_0+u_j) - f(x_0+u_{j-1}) - \frac{\partial f}{\partial x_j}(x_0) v_j \\ &= \sum_{j=1}^{d} \frac{\partial f}{\partial x_j}(x_0 + u_{j-1} + \xi_j e_j) v_j - \frac{\partial f}{\partial x_j}(x_0) v_j \end{aligned}$$

where the last step used Lagrange's theorem. Since $v_j < |v|$ and the partial derivatives are continuous at x_0, so when $|v| \to 0$, the above also approach to 0. □

Corollary 5.1.11

If f is differentiable on Ω, and $\mathrm{d}f = 0$, then f is constant on Ω.

Proposition 5.1.12

Let $f : \Omega \to \mathbb{R}$ be a function differentiable at x_0, and f achieves its local extremum at x_0, then $\mathrm{d}f(x_0) = 0$.

Proof. Trivial. □

If we want to study the second derivative of multi-variable functions, since the derivative is a function $\mathbb{R}^d \to \mathbb{R}^d$ (there are d partial derivatives), we need to study the differentiation for vector-valued functions.

§5.2 Jacobi matrices

Definition 5.2.1. Let $\Omega \subset \mathbb{R}^d$, $\Omega' \subset \mathbb{R}^{d'}$, $f : \Omega \to \Omega'$. If there exists a linear map

$$\mathrm{d}f\big|_{x_0} : \mathbb{R}^d \to \mathbb{R}^{d'},$$

s.t.

$$f(x_0 + v) = f(x_0) + \mathrm{d}f(x_0)(v) + o(|v|),$$

then we say f is differentiable at x_0, the linear map $\mathrm{d}f(x_0)$ is called the differentiation of f at x_0.

Proposition 5.2.2

Let $f = (f_1, \dots, f_{d'})$. f is differentiable at x_0 is equivalent to f_i is differentiable at x_0, and $\mathrm{d}f(x_0) : \mathbb{R}^d \to \mathbb{R}^{d'}$ can be represent as the matrix

$$\left(\frac{\partial f_i}{\partial x_j}(x_0)\right)_{i,j}$$

this is called the **Jacobi matrix** of f at x_0, denoted by $J(f)(x_0)$.

For a function $f : \mathbb{R}^d \to \mathbb{R}$, $\mathrm{d}f(x_0) = (\nabla f)(x_0)$ is a function $\mathbb{R}^d \to \mathbb{R}^d$, hence $\mathrm{d}(\mathrm{d}f)(x_0) = J(\nabla f)$ is a matrix. If we look at the higher derivatives, it will become an n dimensional array, which is hard to represent.

When we have multiple functions to deal with, the differentiation is almost the same as 1 dimensional case:

Proposition 5.2.3 (Chain rule)

Let $\Omega_i \subset \mathbb{R}^{n_i}, 1 \le i \le 3$ be open sets, and $f : \Omega_1 \to \Omega_2, g : \Omega_2 \to \Omega_3$ be differentiable functions. Then $g \circ f : \Omega_1 \to \Omega_3$ is differentiable, and

$$\mathrm{d}(g \circ f)(x) = \mathrm{d}g\big|_{f(x)} \cdot \mathrm{d}f(x).$$

where $\mathrm{d}g$ is a $n_3 \times n_2$ matrix, $\mathrm{d}f$ is a $n_2 \times n_1$ matrix, so $\mathrm{d}(g \circ f)$ is a $n_3 \times n_1$ matrix, as defined above.

Proof. Let $f(x_0) = y_0$,

$$f(x_0 + v) = y_0 + \mathrm{d}f(x_0)v + o(|v|),$$

and

$$g(y_0 + w) = g(y_0) + \mathrm{d}g(y_0)w + o(|w|).$$

Now we compute

$$\begin{aligned} g(f(x_0 + v)) &= g(y_0 + \mathrm{d}f(x_0)v + o(|v|)) \\ &= g(y_0) + \mathrm{d}g(y_0)(\mathrm{d}f(x_0)v + o(|v|)) + o(|\,\mathrm{d}f(x_0)v + o(|v|)|) \\ &= g(y_0) + \mathrm{d}g(y_0)\,\mathrm{d}f(x_0)v + \mathrm{d}g(y_0)o(|v|) + o(|\,\mathrm{d}f(x_0)v + o(|v|)|), \end{aligned}$$

so we only need to verify that

$$\lim_{|v| \to 0} \frac{|\,\mathrm{d}g(y_0)o(|v|) + o(\mathrm{d}f(x_0)v + o(v))|}{|v|} = 0.$$

Note that $|A \cdot v| \leq \|A\||v|$, where the norm of a matrix is defined as $(\sum A_{ij}^2)^{\frac{1}{2}}$, so it's clear the above limit holds. $\square$

Corollary 5.2.4

Let $\Omega_1 \subset \mathbb{R}^{n_1}, \Omega \subset \mathbb{R}^{n_2}$, let f be a differentiable map $\Omega_1 \to \Omega_2$.

If f is a bijection and f^{-1} is differentiable, then:

- $n_1 = n_2$;
- $\mathrm{d}f^{-1}(y) = (\mathrm{d}f)^{-1}(x)$, where $x = f^{-1}(y)$.

Proof. Consider the composite function $\mathrm{id} = f \circ f^{-1} : \Omega_2 \to \Omega_2$, by chain rule,

$$I_{n_2} = \mathrm{d}(f \circ f^{-1}) = \mathrm{d}f \cdot \mathrm{d}f^{-1}.$$

since I_{n_2} has rank n_2, we know that $n_1 \geq n_2$. Similarly $n_2 \geq n_1$, so $n_1 = n_2$.

Hence the inverse of $\mathrm{d}f$ exists and is equal to $\mathrm{d}f^{-1}$. $\square$

Example 5.2.5

Consider the exponential map:

$$\exp : M_n(\mathbb{R}) \to M_n(\mathbb{R}), A \mapsto \sum_{k=0}^{\infty} \frac{A^k}{k!} =: e^A.$$

then $\mathrm{d}\exp(A)$ is a linear map $M_n(\mathbb{R}) \to M_n(\mathbb{R})$.

By definition,

$$e^{A+V} - e^A = \mathrm{d}\exp(A) \cdot V + o(|V|).$$

The left hand side is equal to

$$\sum_{k=0}^{\infty} \frac{(A+V)^k - A^k}{k!} = \sum_{k=0}^{\infty} \frac{\sum_{l=0}^{k-1} A^l V A^{k-1-l} + O(|V|^2)}{k!}.$$

since $\|AB\| \leq \|A\|\|B\|$, the $O(|V|^2)$ part has norm at most $2^k\|V\|^2\|A\|^{k-2}$.

$$\implies e^{A+V} - e^A = \sum_{k=0}^{\infty} \frac{\sum_{l=0}^{k-1} A^l V A^{k-1-l}}{k!} + o(\|V\|).$$

In particular,

- $\mathrm{d}\exp(I)(V) = \sum_{k=0}^{\infty} \frac{kV}{k!} = eV$;
- $\mathrm{d}\exp(0)(V) = V$;
- If A and V is commutative, $\mathrm{d}\exp(A)(V) = \exp(A)V$.

Theorem 5.2.6 (Substitution formula)
Let $\phi : U \to V$ be a bijection, ϕ, ϕ^{-1} are C^1 functions, and Jacobi determinant

$$J_\phi(x) := \det(J(\phi)(x)) \neq 0, \quad \forall x \in U.$$

If f is Lebesgue integrable on V, then

$$\int_V f(y)\,\mathrm{d}y = \int_U f(\phi(x))|J_\phi(x)|\,\mathrm{d}x.$$

Remark 5.2.7 — In fact we only need to check for cuboid I,

$$m(\phi(I)) = \int_I |J_\phi(x)|\,\mathrm{d}x.$$

and ϕ maps null sets to null sets.

Proof. Since $\phi \in C^1$, exists constant M s.t.

$$M^{-1} \leq \|\mathrm{d}\phi\|, \|\mathrm{d}\phi^{-1}\|, |J_\phi| \leq M.$$

$\forall \varepsilon > 0$, divide I into sufficiently small cuboids I_j, such that

$$\phi(x) - \phi(x_j) - \mathrm{d}\phi(x_j)(x - x_j) \leq M\varepsilon|x - x_j|, \quad \forall x \in I_j,$$

where x_j is the center of I_j, because

$$\begin{aligned}
\phi(x) - \phi(x_j) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\phi(tx + (1-t)x_j)\,\mathrm{d}t \\
&= \int_0^1 \mathrm{d}\phi(tx + (1-t)x_j)(x - x_j)\,\mathrm{d}t \\
&= \mathrm{d}\phi(x_j)(x - x_j) + \int_0^1 (\mathrm{d}\phi(tx + (1-t)x_j) - \mathrm{d}\phi(x_j))\,\mathrm{d}t \cdot (x - x_j)
\end{aligned}$$

Hence there exists K independent of ε,

$$m(\phi(I_j)) \leq (|J_\phi(x_j)| + MK\varepsilon)m(I_j).$$

since the image $\phi(I_j)$ is a subset of $\mathrm{d}\phi(x_j)(I_j)$ (which is a parallogram) extending $M\varepsilon|x - x_j|$ on each side.

By taking sufficiently small ε,

$$m(\phi(I)) \leq \sum_j (|J_\phi(x_j)| + MK\varepsilon)m(I_j) = 2MK\varepsilon m(I) + \int_I |J_\phi(x)|\,\mathrm{d}x.$$

Therefore

$$\int_V f(y)\,\mathrm{d}y \leq \int_U f(\phi(x))|J_\phi|\,\mathrm{d}x.$$

apply this to ϕ^{-1} we'll get the equality:

$$m(E) \leq \int_{\phi^{-1}(E)} |J_\phi(x)|\,\mathrm{d}x \leq \int_E |J_\phi(\phi^{-1}(x))||J_{\phi^{-1}}(x)|\,\mathrm{d}x = m(E).$$

$\square$

Example 5.2.8

Consider the spherical coordinates $x = r\sin\theta\sin\varphi$, $y = r\sin\theta\cos\varphi$, $z = r\cos\theta$.

Let $F : (r, \theta, \varphi) \mapsto (x, y, z)$.

$$J_F = \begin{pmatrix} \sin\theta\sin\varphi & r\cos\theta\sin\varphi & r\sin\theta\cos\varphi \\ \sin\theta\cos\varphi & r\cos\theta\cos\varphi & -r\sin\theta\sin\varphi \\ \cos\theta & -\sin\theta & 0 \end{pmatrix}$$

So $\det(J_F) = r^2\sin\theta$. Thus

$$\int_{\mathbb{R}^3} f(x,y,z)\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z = \int_{(0,2\pi)^2}\int_0^{+\infty} f(r,\theta,\phi)r^2\sin\theta\,\mathrm{d}r\,\mathrm{d}\theta\,\mathrm{d}\varphi.$$

Theorem 5.2.9 (Clairaut-Schwarz)

Given an open set $\Omega \subset \mathbb{R}^n$ and $f : \Omega \to \mathbb{R}$. Assume $\frac{\partial f}{\partial x_i}(x)$, $\frac{\partial f}{\partial x_j}(x)$, $\frac{\partial}{\partial x_i}(\frac{\partial f}{\partial x_j})(x)$ exists and are continuous, then $\frac{\partial}{\partial x_j}(\frac{\partial f}{\partial x_i})(x)$ exists and

$$\frac{\partial}{\partial x_i}\left(\frac{\partial f}{\partial x_j}\right)(x) = \frac{\partial}{\partial x_j}\left(\frac{\partial f}{\partial x_j}\right)(x).$$

Proof. WLOG $n = 2$. We'll just expand and compute:

$$\begin{aligned}\frac{\partial}{\partial y}\frac{\partial f}{\partial x}(x_0,y_0) &= \lim_{s\to 0}\frac{1}{s}\left(\frac{\partial f}{\partial x}(x_0,y_0+s) - \frac{\partial f}{\partial x}(x_0,y_0)\right) \\ &= \lim_{s\to 0}\frac{1}{s}\lim_{t\to 0}\frac{1}{t}(f(x_0+t,y_0+s) - f(x_0,y_0+s) - f(x_0+t,y_0) + f(x_0,y_0)).\end{aligned}$$

Since

$$(f(x_0+t,y_0+s) - f(x_0,y_0+s) - f(x_0+t,y_0) + f(x_0,y_0)) = \int_0^s\int_0^t \frac{\partial}{\partial x}\frac{\partial f}{\partial y}(x_0+\tilde{t},y_0+\tilde{s})\,\mathrm{d}\tilde{t}\,\mathrm{d}\tilde{s}.$$

So by Fubini's theorem, □

Notation: Let $\alpha = (\alpha_1, \dots, \alpha_n)$ be a multiple index, where $\alpha_i \geq 0$ are integers. define

$$\partial^\alpha f = \left(\frac{\partial f}{\partial x_1}\right)^{\alpha_1}\cdots\left(\frac{\partial f}{\partial x_n}\right)^{\alpha_n} f.$$

or we can write

$$\partial^\alpha = \frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\cdots\partial x_n^{\alpha_n}}, \quad |\alpha| = \sum_{i=1}^n \alpha_i.$$

Theorem 5.2.10 (Multi-dimensional Taylor expansion)
Let $\Omega \subset \mathbb{R}^n$ be a convex open set. Let $f \in C^{k+1}(\Omega)$, for all $x, y \in \Omega$, then $\exists \theta \in (0, 1]$ s.t.

$$f(y) = \sum_{|\alpha| \le k} \frac{\partial^\alpha f(x)}{\alpha!}(y-x)^\alpha + \sum_{|\alpha|=k+1} \frac{\partial^\alpha f(x+\theta(y-x))}{\alpha!}(y-x)^\alpha.$$

where $(y-x)^\alpha = \prod_{i=1}^n (x_i - y_i)^{\alpha_i}$, $\alpha! = \prod_{i=1}^n \alpha_i!$.

Proof. Let $g(t) = f(ty + (1-t)x)$, $g \in C^{k+1}((-1, 1))$. By Taylor expansion, there exists $\theta \in [0, 1]$,

$$g(1) = \sum_{l=0}^k \frac{g^{(l)}(0)}{l!} + \frac{g^{(k+1)}(\theta)}{(k+1)!}.$$

so it's just a differential formula of composite function, which can be easily proved by induction, and I don't bother to write it down. □

§5.3 Implicit function theorem

As usual let $C^k(\Omega)$ denote the k times continuously differentiable functions on Ω.

Definition 5.3.1 (Differential homeomorphisms)**.** Let $U, V \subset \mathbb{R}^n$, if there exists a bijection $f : U \to V$, such that f, f^{-1} are smooth, then we say U and V are **smoothly homeomorphic**.

Denoted by $C^\infty(U, V)$ the smooth homeomorphisms from U to V.

Example 5.3.2
Let $f : \mathbb{R} \to \mathbb{R}$ by $x \mapsto x^3$, then f is a smooth bijection, but f^{-1} is not differentiable at 0.

Recall that in $\mathbb{R}$ we have the following results:

- If f is strictly increasing and continuous, then f^{-1} is continuous.
- If f is strictly increasing and C^1, $f' \ne 0$, then $f^{-1} \in C^1$.

Theorem 5.3.3
Let Φ be an differential homeomorphism $U \to V$, $f \in C^k(V)$. Then $f \circ \Phi =: \Phi^* f \in C^k(\Omega)$, this is called the **pullback** of f by Φ.

Proof. We proceed by induction on k. When $k = 0$, this is just the continuity of composite functions.

Assume $k = n$ holds, then for $k = n + 1$,

$$\frac{\partial \Phi^* f}{\partial x_j} = \frac{\partial f(\Phi(x))}{\partial x_j} = \sum_{i=1}^n \frac{\partial f}{\partial y_i}(\Phi(x)) \cdot \frac{\partial \Phi_i(x)}{\partial x_j}.$$

Since $f \in C^{n+1} \implies \frac{\partial f}{\partial y_i} \in C^n$, and $\frac{\partial \Phi_i}{\partial x_j}$ is smooth, so $\frac{\partial \Phi^* f}{\partial x_j} \in C^n$. □

Note that the condition $f' \ne 0$ grants that f is indeed a bijection locally. In higher dimensional spaces, the derivatives are more complex, so let's look at some simple cases first.

Lemma 5.3.4
Let $U, V \subset \mathbb{R}^d$ be open regions. Let $f : U \to V$ be a C^1 bijection, and $J(f)$ is non-degenerate (i.e. $\det J(f) \neq 0$). Then $f^{-1} : V \to U$ is continuously differentiable.

Proof. Let $x_0 \in U$, $y_0 = f(x_0) = V$,

$$f(x_0 + v) = y_0 + A \cdot v + o(|v|).$$

Let $E(\delta)$ be a function which satisfies

$$f^{-1}(y_0 + \delta) = x_0 + A^{-1}\delta + E(\delta).$$

this can be derived from taking f^{-1} on both sides of the above equation.

$$y_0 + \delta = f(x_0 + A^{-1}\delta + E(\delta)) = y_0 + \delta + AE(\delta) + o(|A^{-1}\delta + E(\delta)|)$$
$$\implies AE(\delta) + o(A^{-1}\delta + E(\delta)) = 0.$$

From this we can calculate

$$\frac{|E(\delta)|}{|\delta|} = \frac{|o(A^{-2}\delta + A^{-1}E(\delta))|}{|\delta|} \leq \frac{|o(A^{-2}\delta + A^{-1}E(\delta))|}{|A^{-2}\delta + A^{-1}E(\delta)|} \cdot \frac{|A^{-2}\delta + A^{-1}E(\delta)|}{|\delta|}$$
$$\leq o(1)\left(C + C\frac{|E(\delta)|}{|\delta|}\right).$$

Hence $\lim_{|\delta| \to 0} \frac{|E(\delta)|}{|\delta|} = 0$. □

In this case we are given f^{-1} exists, but generally we need to prove this existence.

Theorem 5.3.5 (Inverse function theorem)
Let $f : \Omega \to \mathbb{R}^d$ be a C^1 map, and $\mathrm{d}f(x_0)$ is non-degenerate, then f is a C^1 differential homeomorphism in some neighborhood of x_0.
This is to say, $\exists U \ni x_0, V \ni f(x_0)$ s.t. f is a bijection from U to V and $f^{-1} : V \to U$ is a C^1 map.

Proof. WLOG $x_0 = 0$, $f(x_0) = 0$, also we can apply a linear transformation such that $\mathrm{d}f(x_0) = I$.
There exists $\delta > 0$, s.t.

$$|f(v) - v| < \varepsilon_0|v|, \quad \|J(f)(v) - I\| < \varepsilon_0, \quad \forall |v| < \delta.$$

note that

$$f(v) - f(u) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t} f(tv + (1-t)u)\,\mathrm{d}t = \int_0^1 \mathrm{d}f(tv + (1-t)u) \cdot (v - u)\,\mathrm{d}t$$
$$= \int_0^1 (Jf(tv + (1-t)u) - I) \cdot (v - u)\,\mathrm{d}t + (u - v).$$

but when $|u|, |v| < \delta$, $|f(v) - f(u) - (v - u)| < \varepsilon_0|v - u|$.
Hence $f(u) = f(v) \implies u = v$, f is injective in $B_\delta(0)$.
As for surjectivity, it's sufficient to prove $f(\overline{B_\delta(0)})$ contains a neighborhood of $f(0) = 0$. i.e. $\forall |v| < \delta_1, \exists |u| < \delta$ s.t. $f(u) = u + o(u) = v$.

Since we don't know the non-linear term $o(u)$, we'll iterate to get a solution u: let $u_0 = v$. Define $u_{k+1} = v - (f(u_k) - u_k)$. When δ_1 is sufficiently small,

$$|u_{k+1}| \le |v| + |f(u_k) - u_k| \le |v| + \varepsilon_0 |u_k| \le \delta_1 + \varepsilon_0 \delta \le \delta.$$

Now we prove the convergency:

$$\begin{aligned} |u_{k+2} - u_{k+1}| &= |f(u_{k+1}) - u_{k+1} - f(u_k) + u_k| \\ &= |\int_0^1 (J(f)(tu_{k+1} + (1-t)u_k) - I)\,\mathrm{d}t(u_{k+1} - u_k)| \\ &\le \varepsilon_0 |u_{k+1} - u_k|. \end{aligned}$$

by contraction mapping principle we're done. □

Remark 5.3.6 — This theorem holds for any Banach space.

Corollary 5.3.7

Let $k \ge 2$ be an integer, when $f \in C^k$ in the above theorem, we can imply that $f^{-1} \in C^k(V)$.

Proof. Since

$$\mathrm{d}f^{-1}(u) = (\mathrm{d}f)^{-1}(f^{-1}(u)),$$

so $\mathrm{d}f \in C^{k-1} \implies (\mathrm{d}f)^{-1} \in C^{k-1}$. □

Theorem 5.3.8 (Implicit function theorem)

Let $f : \Omega \subset \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^p$ be a continuously differentiable function. If $\exists (x^*, y^*) \in \Omega$ s.t. $f(x^*, y^*) = 0$, and $\mathrm{d}_y f(x^*, y^*)$ is inversible, then there exists an open neighborhood $U \subset \mathbb{R}^n$ of x^*, $V \subset \mathbb{R}^p$ of y^*, and a C^1 map $\phi : U \to V$ such that:

$$f(x, \phi(x)) = 0, \quad \mathrm{d}\phi(x) = -(\mathrm{d}_y f(x, \phi(x)))^{-1} \cdot \mathrm{d}_x f(x, \phi(x)).$$

Also if $x \in U$ and $f(x, y) = 0$, we must have $y = \phi(x)$.

Remark 5.3.9 — This is to say, if $f(x, y) = 0$, $x \in U, y \in V$, then $y = \phi(x)$.
Also remember that $\mathrm{d}_y f$ is a $p \times p$ matrix, $\mathrm{d}_x f$ is a $p \times n$ matrix.

Proof. By the inverse function theorem, let $F(x, y) := \mathbb{R}^{n+p} \to \mathbb{R}^{n+p}$ with

$$(x, y) \mapsto (x, f(x, y))$$

So $F(x^*, y^*) = (x^*, 0)$, and

$$\mathrm{d}F(x^*, y^*) = \begin{pmatrix} I_n & 0 \\ \mathrm{d}_x f(x^*, y^*) & \mathrm{d}_y f(x^*, y^*) \end{pmatrix}.$$

Since $\mathrm{d}_y f(x^*, y^*)$ is inversible, $\mathrm{d}F(x^*, y^*)$ is inversible as well. Hence there exists neighborhoods of (x^*, y^*) and $(x^*, 0)$, say $\widetilde{\Omega}$ and $\widetilde{\Omega}_1$, such that F is a C^1 homeomorphism $\widetilde{\Omega} \to \widetilde{\Omega}_1$.

We can find $U \ni x^*, V \ni y^*$ s.t. $U \times V \subset \widetilde{\Omega}$. Let T be the C^1 map s.t.

$$F^{-1}(x,z) = (x, T(x,z)).$$

Let $\phi(x) = T(x,0)$, we have

$$F(x,\phi(x)) = F(x,T(x,0)) = (x,0) \implies f(x,\phi(x)) = 0.$$

Since F is a bijection, clearly $f(x,y) = 0 \implies y = \phi(x)$. By taking the differentiation of $f(x,\phi(x)) = 0$,

$$(\mathrm{d}_x f, \mathrm{d}_y f) \cdot \begin{pmatrix} I_n \\ \mathrm{d}\phi(x) \end{pmatrix} = 0 \implies \mathrm{d}_x f(x,\phi(x)) + \mathrm{d}_y f(x,\phi(x)) \cdot \mathrm{d}\phi(x) = 0.$$

□

§5.4 Submanifolds in Euclid space

The implicit function theorem actually gives an example of manifolds: the preimage of $f(x,y) = 0$ is an n-dimensional manifold in $\mathbb{R}^{n+p}$.

Definition 5.4.1 (Manifolds)**.** Let $M \subset \mathbb{R}^n$ be a nonempty set. If $\exists d \geq 0$, $\forall x \in M$ exists open sets $U \subset \mathbb{R}^n$, $V \subset \mathbb{R}^d$, and a differential homeomorphism $\Phi : U \to V$, such that

$$\Phi(U \cap M) = V,$$

we say M is a **d-dimensional differential manifold**. Denote $\dim M = d$, and $n - d$ is called the **codimension** of M.

Remark 5.4.2 — There might be different maps $\phi_1 : U_1 \to V_1, \phi_2 : U_2 \to V_2$, when $U_1 \cap U_2 \cap M \neq \emptyset$, we must have $\phi_2 \circ \phi_1^{-1}$ is a differential map from V_1 to V_2. In fact when M isn't a subset of $\mathbb{R}^n$, this is the original definition of differential manifolds.

Corollary 5.4.3 (Regular value theorem)

Let $f : \Omega \to \mathbb{R}^p$ be a smooth map, where $\Omega \subset \mathbb{R}^n$ is an open set, $n \geq p$. For all $c \in \mathbb{R}^p$, we call the **fibre** of c to be its preiamge:

$$f^{-1}(c) = \{x \in \mathbb{R}^n \mid f(x) = c\}.$$

If $\forall x \in f^{-1}(c)$, $\operatorname{rank} \mathrm{d}f(x) = p$, then $f^{-1}(c)$ is a manifold with **codimension** p.

Example 5.4.4

Let S^{n-1} be the unit sphere in $\mathbb{R}^n$. Let $f : \mathbb{R}^n \to \mathbb{R}$ with $x \mapsto |x|^2 - 1$, then $S^{n-1} = f^{-1}(0)$.

Since $\mathrm{d}f = (2x_1, 2x_2, \ldots, 2x_n)$, clearly $\operatorname{rank} \mathrm{d}f = 1$ for all $x \in S^{n-1}$, so S^{n-1} is a manifold with codimension 1.

Example 5.4.5

Consider a surface in $\mathbb{R}^4 = \mathbb{C}^2$:

$$T^2 = \{(z_1, z_2) \mid |z_1| = 1, |z_2| = 1\}.$$

Let $f(x, y, z, w) = x^2 + y^2 - 1, g(x, y, z, w) = z^2 + w^2 - 1$, then $T^2 = \begin{pmatrix} f \\ g \end{pmatrix}^{-1} (0)$.

The differentiation is

$$\mathrm{d}\begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} 2x & 2y & 0 & 0 \\ 0 & 0 & 2z & 2w \end{pmatrix},$$

so T^2 is a manifold with codimension 2.

Definition 5.4.6. Let $M \subset \mathbb{R}^n$ be a manifold. If $\dim M = 1$, we say M is a curve; if $\dim M = 2$, M is a surface; and if $\dim M = n - 1$, we say M is a hyperplane.

Lemma 5.4.7

Let $f : \mathbb{R}^n \to \mathbb{R}$ is a smooth function, if $\forall x_0 \in f^{-1}(0)$, $\mathrm{d}f(x_0) \neq 0$, then $f^{-1}(0)$ is a smooth hyperplane, it is called the hyperplane globally determined by an equation.

Example 5.4.8

In $\mathbb{R}^3$, f, g are smooth functions. If for all $x \in \mathbb{R}^3$ with $f(x) = g(x) = 0$ we have $\nabla f, \nabla g$ are linearly independent, then $\{f = g = 0\}$ is a smooth curve.

Theorem 5.4.9 (Parametrization of manifolds)

Let Ω be an open set in $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}^{n+p}$ is a smooth map. If $\forall x^* \in \Omega$, $\operatorname{rank} \mathrm{d}f(x^*) = n$, then there exists an open set U, $x^* \in U$ s.t. $f(U) \subset \mathbb{R}^{n+p}$ is an n-dimensional manifold.

Proof. Let x_i be a coordinate in $\mathbb{R}^{n+p}$.

WLOG $(\frac{\partial f_i}{\partial x_j})_{1 \le i,j \le n}$ is non-degenerate, let $F = (f_1, \dots, f_n)$, $G = (f_{n+1}, \dots, f_{n+p})$ and apply inverse function theorem on F, there exists open neighborhoods $U \ni x, V \ni F(x) =: y$, s.t. $F : U \to V$ is a smooth homeomorphism.

$$\begin{tikzcd} U \subset \Omega \arrow[r, "F"] \arrow[d, "f"] & V \subset \mathbb{R}^n \arrow[dl, "\phi"] \\ \mathbb{R}^{n+p} & \end{tikzcd}$$

So $f(x) = (F(x), G(x)) = (y, GF^{-1}(y))$. Let

$$\phi : V \to \mathbb{R}^n, \quad y \mapsto (y, GF^{-1}(y)).$$

We can see that ϕ is a homeomorphism $V \to f(U)$. (Indeed it's a bijection) So by definition we know $f(U)$ is a manifold. $\square$

Example 5.4.10

Let

$$\phi(\theta, r) = \begin{cases} x = \left(1 + r\cos\dfrac{\theta}{2}\right)\cos\theta \\ y = \left(1 + r\cos\dfrac{\theta}{2}\right)\sin\theta \\ z = r\sin\dfrac{\theta}{2} \end{cases}, \quad I = [0, 2\pi] \times (-1, 1).$$

Then $M = \phi(I)$ is a Mobius strip, which is a two dimensional smooth manifold in $\mathbb{R}^3$, as $\mathrm{d}\phi$ has rank 2 everywhere.

Besides, there doesn't exist a function $f : \mathbb{R}^3 \to \mathbb{R}$ s.t. $M = f^{-1}(0)$. Basically this is because M is not orientable, but ∇f and $-\nabla f$ are "normal" directions of M, which makes it orientable. Below we give a sketch:

Proof. Let $v(\theta) = \cos\frac{\theta}{2} e_2(\theta) - \sin\frac{\theta}{2} e_1(\theta)$, where $e_2(\theta) = (0, 0, 1), e_1(\theta) = (\cos\theta, \sin\theta, 0)$.

Note that $e_1 \perp e_2$, consider the curve $\beta : [0, 2\pi] \to \mathbb{R}^3$

$$\theta \mapsto (\cos\theta, \sin\theta, 0) + \varepsilon v(\theta).$$

Let ε be sufficiently small, when $\varepsilon \neq 0$ we can check β and M do not intersect. We can take ε s.t. $f(\beta(0)) > 0$ as $\mathrm{d}f \neq 0$. (ε can be negative)

Since $\beta(0) = (1, 0, \varepsilon), \beta(2\pi) = (1, 0, -\varepsilon)$, when $f(\beta(0)) > 0$, we must have $f(\beta(2\pi)) < 0$. By continuity, $\exists\theta_0$ s.t. $f(\beta(\theta_0)) = 0$, which means $\beta(\theta_0) \in M$, contradiction! $\square$

Midterm exam....qaq

Proposition 5.4.11

Let $\Omega \subset \mathbb{R}^n$, and $f : \Omega \to \mathbb{R}^m$ is a smooth map. Let $S \subset \mathbb{R}^m$ be a differential manifold, if for all $x \in f^{-1}(S)$, we have $\operatorname{rank} \mathrm{d}f(x) = m$, then $f^{-1}(S)$ is a differential manifold with codimension same as S.

Proof. For any $x \in S$, let Φ be the homeomorphism from an open neighborhood of x to $\mathbb{R}^m$.

Suppose $\dim S = d$, let

$$F(x) = ((\Phi \circ f)_{d+1}, \ldots, (\Phi \circ f)_m).$$

Note that $\mathrm{d}(\Phi \circ f)$ is an $m \times n$ matrix, and its rank is m. Since

$$\mathrm{d}(\Phi \circ f) = \begin{pmatrix} \mathrm{d}(\Phi \circ f)_1 \\ \vdots \\ \mathrm{d}(\Phi \circ f)_d \\ \mathrm{d}F \end{pmatrix}$$

Thus $\mathrm{d}F$ is a $(m-d) \times n$ matrix with rank $m - d$. So $F^{-1}(0) = f^{-1}(S)$ is a manifold with dimension $n - (m - d)$. $\square$

§5.5 Tangent space

Since the differentiation relies on the choice of coordinates, if we want to study the manifold more geometrically, we should look at the tangent lines or tangent planes of the manifold.

Definition 5.5.1 (Tangent vectors)**.** Let M be a differential manifold. Let $p \in M$, for all parametrized curve $\gamma : (-\varepsilon, \varepsilon) \to M$ with $\gamma(0) = p$, we say the vector $\gamma'(0) \in \mathbb{R}^n$ is the **tangent vector** of γ at point p.

Let T_pM denote the **tangent space** at p, which is defined as

$$T_pM = \{\gamma'(0) \in \mathbb{R}^n \mid \gamma(0) = p\}.$$

It's clear that T_pM should be a vector space of dimension $\dim M$, next we'll prove this fact.

Proposition 5.5.2 (Push forward of tangent spaces under differential homeomorphism)
Let $\Phi : U \to V$ be a differential homeomorphism, $M \subset U$ be a manifold, then

$$T_{\Phi(p)}\Phi(M) = (\mathrm{d}\Phi)\big|_p \cdot T_pM.$$

Proof. Let γ be a parametrized curve on M with $\gamma(0) = p$. Note that $\Phi \circ \gamma$ is a curve on $\Phi(M)$ passing through $\Phi(p)$. Since

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi \circ \gamma(t)\Big|_{t=0} = \mathrm{d}\Phi(p) \cdot \gamma'(0).$$

Thus $\mathrm{d}\Phi(p) \cdot T_pM \subset T_{\Phi(p)}\Phi(M)$.

Now we do the same thing for Φ^{-1}, we can get the desired equality. □

Now we can easily calculate the tangent space: since M is locally homeomorphic to $\mathbb{R}^d$, and obviously $T_{\Phi(p)}(\mathbb{R}^d \times \{0\}) = \mathbb{R}^d \times \{0\}$, by above proposition, $T_pM = (\mathrm{d}\Phi) \cdot T_{\Phi(p)}(\mathbb{R}^d \times \{0\})$ is a vector space of dimension d.

Theorem 5.5.3
Let M be a manifold, T_pM is a vector space of dimension $\dim M$.

Proposition 5.5.4
Let $f : \mathbb{R}^{n+d} \to \mathbb{R}^n$ be a smooth map, $\operatorname{rank} \mathrm{d}f = n$. Let $M = f^{-1}(f(p))$, then $T_pM = \ker \mathrm{d}f(p)$.

Proof. Let

$$F(x, y) = (x, f(x, y)), x \in \mathbb{R}^d, y \in \mathbb{R}^n.$$

F is a homeomorphism, so $T_pM = (\mathrm{d}F^{-1})T_{F(p)}F(M)$.

Note that $F(M) = \{(x, p) \mid \exists y, f(x, y) = f(p)\}$, it must be a vector space of dimension d, so $T_{F(p)}F(M) = \mathbb{R}^d \times \{0\}$,

$$T_pM = (\mathrm{d}F^{-1})T_{F(p)}F(M) = \ker \mathrm{d}f(p).$$

□

Example 5.5.5

Let M be a manifold determined by $f : \mathbb{R}^n \to \mathbb{R}$,

$$T_pM = \ker \mathrm{d}f = \{v \in \mathbb{R}^n \mid \mathrm{d}f(p)v = 0\}.$$

Here $\mathrm{d}f = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n}) = \nabla f$. So $v \in T_pM \iff \nabla f \cdot v = 0$, the dot means the inner product. In this case the vector ∇f is called **normal direction vector**.

§5.6 Smooth maps between manifolds

Next we'll briefly introduce the differential maps between manifolds. Since the differentiation on manifolds is hard to define, so what we do is actually regarding manifolds as $\mathbb{R}^d$ locally and define the differentiablity using the maps between Eucild spaces.

Definition 5.6.1. Let M, N be manifolds in $\mathbb{R}^m, \mathbb{R}^n$, respectively. $f : M \to N$ is a map, if $\forall p \in M$, there exists $p \in U \subset \mathbb{R}^m, V \subset \mathbb{R}^d$, $\Phi : U \to V$ s.t.

$$f_\Phi = f \circ \Phi^{-1}$$

is a smooth map from V to N. We say f is a smooth map from M to N.

We need to check this definition is well-defined: if there's another homeomorphism Φ', $f \circ \Phi' = (f \circ \Phi^{-1}) \circ (\Phi \circ \Phi'^{-1})$ is indeed a smooth map.

Lemma 5.6.2 (Smooth maps are locally restrictions of smooth maps in Eucild spaces)

Let $f : M \to N$ be a map, then f is smooth $\iff \forall p \in M, \exists p \in U \subset \mathbb{R}^m$ and a smooth map $F : U \to \mathbb{R}^n$ s.t.

$$f\big|_{U\cap M} = F\big|_{U \cap M}.$$

Proof. Let τ denote the embedding from $M \cap U$ to U. Since $f \circ \Phi^{-1} = F \circ \tau \circ \Phi^{-1}$, so F smooth $\implies f_\Phi$ smooth $\implies f$ smooth.

$$\begin{tikzcd}
V \subset \mathbb{R}^d \arrow[dr, "f\circ\Phi^{-1}"'] & M \cap U \arrow[l, "\Phi"'] \arrow[r, "\tau"] \arrow[d, "f"] & U \arrow[dl, "F"] \\
& N \subset \mathbb{R}^n &
\end{tikzcd}$$

TODO: fix this

On the other hand, let $\widetilde{\tau}$ be the projection from U to V, then $F = f \circ \Phi^{-1} \circ \widetilde{\tau} \circ \Phi$ satisfies the desired condition. $\square$

Example 5.6.3

Let A be an orthogonal map in $\mathbb{R}^3$, then A can be restricted to $S^2 \to S^2$.

Definition 5.6.4 (Tangent map)**.** Let $f : M \to N$ be a map between manifolds, $v \in T_pM$. Let $\gamma : (-\varepsilon, \varepsilon) \to M$ be a parametrized curve with $\gamma(0) = p, \gamma'(0) = v$, then $f(\gamma(t))$ is a curve on N.

$$\mathrm{d}f(p)(v) = \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\Big|_{t=0} \in T_{f(p)}N.$$

Thus $\mathrm{d}f(p) : T_pM \to T_{f(p)}N$ is a map between tangent spaces.

In fact, if $f = F\big|_M$, then $\mathrm{d}f(p)(v) = \mathrm{d}F(p) \cdot v$.

Definition 5.6.5 (Tangent bundle)**.** Let M be a manifold, $\forall p \in M$, there's a tangent space T_pM. Define the **tangent bundle** of M to be

$$TM = \bigsqcup_{p \in M} T_pM.$$

If X is a map $M \to TM$: $p \mapsto X(p)$, with $X(p) \in T_pM$, then it's called a **tangent vector field**.

In other words, a tangent vector field is just to assign a tangent vector to every point in M.

Proposition 5.6.6

Let $M \subset \mathbb{R}^n$ be a manifold, all its tangent vector field form a C^∞ module $T(M, TM)$, i.e. $\forall f \in C^\infty(M)$, X, Y are smooth vector fields, then $fX, X + Y$ are both smooth vector fields.

Proposition 5.6.7

Let $M \subset \mathbb{R}^n$ be a smooth manifold, we have

$$TM = \{(x, v) \mid x \in M, v \in T_xM\}$$

is a smooth manifold in $\mathbb{R}^{2n}$, and $\dim TM = 2 \dim M$.

Proof. There exists a local homeomorphism $\phi : V \to U \subset \mathbb{R}^n$ s.t. $V \subset \mathbb{R}^d$, $\phi(V) = M \cap U$.

Define map $T\phi : V \times \mathbb{R}^n \to U \times \mathbb{R}^n$, $(x, v) \mapsto (\phi(x), \mathrm{d}\phi(x) \cdot v)$. Since $T\phi$ is injective(ϕ is homeomorphism), and

$$\mathrm{d}T\phi = \begin{pmatrix} \mathrm{d}\phi & 0 \\ \mathrm{d}(\mathrm{d}\phi)(v) & \mathrm{d}\phi \end{pmatrix}$$

is non-degenerate, so $T\phi$ is a bijection and hence differential homeomorphism.

Since the tangent space of V is just $\mathbb{R}^d$, so $T(U \cap M)$ is the image of $T\phi$ restricted on $V \times \mathbb{R}^d$. (Note that $\mathrm{d}\phi(x) \cdot v \in T_{\phi(x)}M$) Thus TM is a manifold in $\mathbb{R}^{2n}$ with dimension $2d$. $\square$

Definition 5.6.8 (Tangent maps)**.** Earlier we know that $\mathrm{d}f(p)$ is a map $T_pM \to T_{f(p)}N$, combined with tangent bundle we can write $\mathrm{d}f : TM \to TN$, this map is called the **tangent map** or the **differentiation** of f.

If we have a vector field X and a smooth function $f : M \to \mathbb{R}^n$, consider

$$X(f)(p) = \mathrm{d}f(X)(p) := \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\Big|_{t=0}, \quad \gamma(0) = p, \gamma'(0) = X(p).$$

So X induces a smooth map $C^\infty(M) \to C^\infty(M)$.

Now we can generalize a well known result to manifolds:

Proposition 5.6.9

Let $M \subset \mathbb{R}^n$ be a smooth manifold, $f \in C^\infty(M)$. If f achieves a local extremum at $p \in M$, we must have $\mathrm{d}f(p) = 0$.

Proof. It suffices to prove $\mathrm{d}f(p)(v) = 0$, $\forall v \in T_pM$. Take γ s.t. $\gamma(0) = p, \gamma'(0) = v$, then $f(\gamma(t))$ achieves its extremum at $t = 0$, so $\frac{\mathrm{d}}{\mathrm{d}t} f(\gamma(t))\big|_{t=0} = 0 = \mathrm{d}f(p)(v)$. $\square$

§5.7 Conditional extremum problem

Consider a function $f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$ and some constraint conditions

$$\begin{cases} g_1(x_1, \ldots, x_n) = 0 \\ \quad\vdots \\ g_m(x_1, \ldots, x_n) = 0 \end{cases}$$

We want to compute the extremum of f under these conditions.

Well, you probably heard of *Lagrange multipliers*, i.e. let

$$L(x_1, \ldots, x_n, \lambda_1, \ldots, \lambda_m) = f(x) - \sum_{j=1}^{m} \lambda_j g_j(x).$$

But here we'll provide a different point of view. Let M be the manifold in $\mathbb{R}^n$ under those conditions, Suppose $p \in M$ is a local extremum of f, then $T_pM \subset \ker \mathrm{d}f(p)$.

Also recall that $T_pM = \ker \mathrm{d}g(p) = \bigcap_{j=1}^{m} \ker \mathrm{d}g_j(p)$. This means that, $\exists \lambda_1, \ldots, \lambda_m$ s.t.

$$\mathrm{d}f(p) = \sum_{j=1}^{m} \lambda_j \,\mathrm{d}g_j(p).$$

Surprisingly, we get the same result of Lagrange multipliers! Hence what we've done is to give a geometrical comprehension of Lagrange multipliers.

Example 5.7.1

Let $g : \mathbb{R}^n \to \mathbb{R}$ be the constraint function, then f can achieve its extremum only if $\mathrm{d}f = \lambda\, \mathrm{d}g$.

For example, let $f(x) = d(x, z)^2$, $\mathrm{d}f(x) = 2(x_1 - z_1, \ldots, x_n - z_n)$, so $\mathrm{d}f = \lambda\, \mathrm{d}g$ means the vector $\mathrm{d}f(p)$ is orthogonal to the tangent plane of $M = \{g = 0\}$.

Proposition 5.7.2 (Hadamard's inequality)

Let $v_1, \ldots, v_n \in \mathbb{R}^n$, then

$$|\det(v_1, \ldots, v_n)| \le |v_1| \cdots |v_n|.$$

Proof. Let $f : \mathbb{R}^{n^2} \to \mathbb{R}$, $(v_1, \ldots, v_n) \mapsto \det(v_1, \ldots, v_n)$ with constraint $|v_i| = 1$. Let $v_{ij} \in \mathbb{R}$,

$$g_i(V) = -1 + \sum_{j=1}^{n} v_{ij}^2.$$

The manifold determined by g_i is $M = (S^{n-1})^n$. The extremum point of f in M must satisfy:

$$\frac{\partial f}{\partial v_{i_0 j}} - \lambda_{i_0} \frac{\partial g_{i_0}}{\partial v_{i_0 j}} = 0.$$

This implies $v_{i_0 j}^* = 2\lambda_{i_0} v_{i_0 j}$, where $v_{i_0 j}^*$ is the *cofactors* of $v_{i_0 j}$.

This means that $\sum_{j=1}^{n} v_{i_0 j} v_{kj} = 0$, so V must be an orthogonal matrix, so $|f| \le 1$. $\square$

§5.8 Convex functions

Definition 5.8.1 (Hesse matrix)**.** Let $f : \Omega \subset \mathbb{R}^n \to \mathbb{R}$ be a C^2 function, we call the Jacobi matrix of ∇f to be the **Hesse matrix** of f. (Also called Hessian matrix)

$$H_f(p) = \nabla^2 f(p) = \left(\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} f(p)\right)_{i,j}.$$

Since the partial derivatives commute, so H_f is a symmetrical matrix, hence diagonalizable.

Proposition 5.8.2
Let $f \in C^2(\Omega)$, let x_0 be a minimum of f, then $\nabla f(x_0) = 0$, and $H_f(x_0)$ is semi positive definite.

Proof. By Taylor's expansion,

$$f(x) = f(x_0) + \mathrm{d}f(x_0)(x - x_0) + \frac{1}{2}(x - x_0)^T H_f(x_0)(x - x_0) + o(|x - x_0|^2).$$

If $H_f(x_0)$ has a negative eigenvalue $-\lambda$, with eigenvector v, then $f(x_0 + tv) = f(x_0) - \frac{1}{2}\lambda t^2 |v|^2 + o(|tv|^2)$, which contradicts with the minimality of x_0. ☐

Proposition 5.8.3
If $\nabla f(x_0) = 0$, $H_f(x_0)$ is positive definite, then x_0 is a local minimum of f.

Proof. Same as previous one. ☐

Definition 5.8.4 (Convex functions)**.** If f and Ω satisfies:

$$\forall x, y \in \Omega, tx + (1-t)y \in \Omega, \quad f(tx + (1-t)y) \le tf(x) + (1-t)f(y)$$

we say Ω is a **convex set** and f a **convex function**.

Theorem 5.8.5 (Jensen's inequality)
Let f be a convex function on Ω. Real numbers $t_i \ge 0, \sum_{i=1}^N t_i = 1$, for $x_i \in \Omega$,

$$f\left(\sum_{i=1}^N t_i x_i\right) \le \sum_{i=1}^N t_i f(x_i).$$

Example 5.8.6 (Convex functions)
Linear functions $f(x) = Ax + b$ are convex.
The norm function $\|\cdot\| : \mathbb{R}^n \to \mathbb{R}$ is convex. Also let A be an $n \times n$ positive definite matrix, then $f(x) = x^T Ax$ is convex.

Just like the one dimensional case, convex functions have nice properties.

Theorem 5.8.7

Let f be a convex function on an open convex set Ω, then f is continuous, and Lipschitz continuous in any compact set, i.e.

$$|f(x) - f(y)| \leq M|x - y|, \quad x, y \in U$$

where U is a compact set.

Proof. WLOG $0 \in \Omega$, take an orthogonal basis $e_1, \dots, e_n$. Let

$$x = \sum_{i=1}^{n} \lambda_i \bar{e}_i, \quad \bar{e}_i = e_i \text{ or } -e_i, \lambda_i \geq 0.$$

When $|x|$ sufficiently small, $\sum_{i=1}^{n} \lambda_i < 1$, so by Jensen's inequality,

$$f(x) \leq \sum_{i=1}^{n} \lambda_i f(\bar{e}_i) + \lambda f(0),$$

$$f(x) - f(0) \leq \sum_{i=1}^{n} \lambda_i (f(\bar{e}_i) - f(0)) \leq \left(\sum_{i=1}^{n} \lambda_i^2 \right)^{\frac{1}{2}} \left(\sum_{i=1}^{n} (f(\bar{e}_i) - f(0))^2 \right)^{\frac{1}{2}} \leq |x| C,$$

since we can change the length of e_i, and f is continuous on a straight line.

This means f is continuous. For the second part, let $\lambda_0 = \frac{1}{1 + \sum_{i=1}^{n} \lambda_i}$, since $0 = \lambda_0 x + \sum_{i=1}^{n} \lambda_0 \lambda_i (-\bar{e}_i)$, by Jensen's inequality, we'll get the desired property. □

Proposition 5.8.8

Let f be a differentiable function on a covex set Ω,
f is convex $\iff f(x) \geq f(x_0) + \mathrm{d}f(x_0)(x - x_0)$.

Proof. If f is convex, just use the definition and let $t \to 0$:

$$f(x_0) + f'(x_0) t(x - x_0) + o(t(x - x_0)) \leq t f(x) + (1 - t) f(x_0).$$

Conversely, let $z = tx + (1 - t)x_0$,

$$f(x) \geq f(z) + f'(z)(1 - t)(x - x_0), f(x_0) \geq f(z) + f'(z) t(x_0 - x).$$

Thus adding these together we get

$$t f(x) + (1 - t) f(x_0) \geq f(z).$$

□

Theorem 5.8.9

Let $\Omega \subset \mathbb{R}^n$ be an open convex set, $f \in C^2(\Omega)$, f convex $\iff H_f(x)$ semi positive definite.

Proof. One direction can be proved using Taylor's expansion.

On the other hand, let $H(t) = f(x_0 + t(x - x_0)) - f(x_0) - t\,\mathrm{d}f(x_0)(x - x_0)$, then $H'(t) = \mathrm{d}f(x_0 + t(x - x_0))(x - x_0) - \mathrm{d}f(x_0)(x - x_0)$,

$$H''(t) = (x - x_0)^T H_f(p)(x_0 + t(x - x_0))(x - x_0) \geq 0.$$

So $H(t)$ is a convex function, $H(0) = 0, H'(0) = 0$. $\square$

§6 Integrals on surfaces

§6.1 Measures on manifolds

To define integrals, we need to define a measure on it first.

For example, let $v_1, \dots, v_d \in \mathbb{R}^n$ be linearly independent vectors, and unit vectors $v_{d+1}, \dots, v_n$ complete them to a basis, satisfying $v_j \perp v_i, j > d, j > i$.

Let A be a linear map s.t. $Ae_i = v_i$, then the volume of $A(E)$ is $|\det A| = \sqrt{\det(G \cdot G^T)}$, where $G = \begin{pmatrix} v_1 \\ \vdots \\ v_d \end{pmatrix}$ is a $d \times n$ matrix.

Since $AA^T = \begin{pmatrix} GG^T & 0 \\ 0 & I_{n-d} \end{pmatrix}$, $|\det A| = \sqrt{\det GG^T}$, we say GG^T is the **Gram matrix** of G.

Another example is the length of a curve. Recall that we have the formula

$$L(\gamma) = \int_0^1 |\gamma'(t)| \,\mathrm{d}t.$$

The length of a curve is essentially the "volume" of a 1-dimensional manifold, so the idea of higher dimensional manifold is the same.

Definition 6.1.1. Let M be a manifold in $\mathbb{R}^n$. Let $\Phi : V \subset \mathbb{R}^d \to U \subset M$ be a smooth homeomorphism, $\operatorname{rank} \Phi = d$. We can split U to many small regions and use the paraloids to approximate the volume of each regoin.

Thus we define:

$$m(U) = \int_V \sqrt{\det(\mathrm{d}\Phi(x)^T \,\mathrm{d}\Phi(x))} \,\mathrm{d}x_1 \,\mathrm{d}x_2 \cdots \mathrm{d}x_d.$$

The above differential form inside the integral is called the **volume form**:

$$\mathrm{d}\sigma(y) = \sqrt{\det(\mathrm{d}\Phi^T \,\mathrm{d}\Phi)} \,\mathrm{d}y.$$

Moreover, with this measure we can define the integral of general measurable function f (measurable means locally measurable on $\mathbb{R}^d$):

$$\int_U f \,\mathrm{d}\sigma = \int_V f(\Phi(x)) \sqrt{\det(\mathrm{d}\Phi^T \,\mathrm{d}\Phi)} \,\mathrm{d}x.$$

Just like the length of a curve, the volume defined above is also a geometric quantity, i.e. independent of coordinates.

Example 6.1.2

Let $d = 1$, $\gamma : (-1, 1) \to \mathbb{R}^n$, $\gamma'(0) \neq 0$. For fixed $-1 < a < b < 1$ and a function f on γ, let C_a^b denote the curve between $\gamma(a), \gamma(b)$,

$$\int_{C_a^b} f \,\mathrm{d}\sigma = \int_a^b f(\gamma(t))|\gamma'(t)| \,\mathrm{d}t$$

is called the **curve integral of the first type**.

Example 6.1.3

Let $d = n - 1$, $f : \mathbb{R}^{n-1} \to \mathbb{R}$, the graph of f is a hyper-surface $\Gamma_f = \{(x, f(x)) \mid x \in \mathbb{R}^{n-1}\}$. It has a parametrization $\Phi(x) = (x, f(x))$, so

$$\mathrm{d}\Phi = \begin{pmatrix} I_{n-1} \\ \nabla f \end{pmatrix} \implies \mathrm{d}\Phi^T \,\mathrm{d}\Phi = I_{n-1} + \nabla f^T \cdot \nabla f.$$

Hence $\det(\mathrm{d}\Phi^T \,\mathrm{d}\Phi) = 1 + |\nabla f|^2$. (This can be obtained by looking at the eigenvectors)

Therefore for φ on $\mathbb{R}^n$, we have

$$\int_{\Gamma_f} \varphi \,\mathrm{d}\sigma = \int_{\mathbb{R}^{n-1}} \varphi(x, f(x))\sqrt{1 + |\nabla f|^2} \,\mathrm{d}x.$$

Next we'll compute the surface area of unit sphere S^{n-1}.

Let c_n denote the volume of unit sphere in $\mathbb{R}^n$,

$$c_n = \frac{\pi^{\frac{n}{2}}}{\Gamma(\frac{n}{2} + 1)}.$$

We claim in advance that the surface area of unit sphere $\omega_{n-1} = nc_n$. Here we use the sphere coordinates:

$$x_1 = r \prod_{k=1}^{n-1} \sin\theta_k, \quad x_i = r \sin\theta_{n-1} \cdots \sin\theta_i \cos\theta_{i-1}, 2 \leq i \leq n.$$

Let $F_n(r, \theta_1, \ldots, \theta_{n-1}) = (x_1, \ldots, x_n)$.

$$\mathrm{d}F_n = \begin{pmatrix} \sin\theta_{n-1} \,\mathrm{d}F_{n-1} & \cos\theta_{n-1} F_{n-1}^T \\ (\cos\theta_{n-1}, 0, \ldots, 0) & -r\sin\theta_{n-1} \end{pmatrix}.$$

So we can compute its determinant (expand using the last row), note that the first column of $\mathrm{d}F_{n-1}$ is $r^{-1}F_{n-1}^T$,

$$\begin{aligned} \det \mathrm{d}F_n =& - r\sin\theta_{n-1}(\sin\theta_{n-1})^{n-1} \det(\mathrm{d}F_{n-1}) \\ &+ (-1)^{n-1}(\cos\theta_{n-1})^2(-1)^{n-2} r(\sin\theta_{n-1})^{n-2} \det(\mathrm{d}F_{n-1}) \\ =& - r(\sin\theta_{n-1})^{n-2} \det(\mathrm{d}F_{n-1}). \end{aligned}$$

Hence $\mathrm{d}x = r^{n-1} \sin^{n-2}\theta_{n-1} \cdots \sin\theta_2 \,\mathrm{d}r \,\mathrm{d}\theta_1 \cdots \mathrm{d}\theta_{n-1} = r^{n-1} \,\mathrm{d}r \,\mathrm{d}\omega$.

Denote F_n^S to be the function F_n restricted to S^{n-1}. Then $\mathrm{d}F_n = (r^{-1}F_n^T, \mathrm{d}F_n^S)$. We can compute that the Gram determinant of $\mathrm{d}F_n^S$ is just $\det \mathrm{d}F_n$ with $r = 1$.

The rest is some integrals with gamma function and beta function, which is left out.

§6.2 Stolkes' formula

Intuitively, Stolkes' formula states that: Let D be a region, $\mathrm{d}\omega$ be a differential form, then

$$\int_D \mathrm{d}\omega = \int_{\partial D} \omega.$$

Here ∂D means the "boundary" of D.

Of course we need some "regularity" requirements of D and ω, and it's the generalization of Newton-Lebniz formula into higher dimensions.

Definition 6.2.1 (Bounded regions with boundary)**.** Let $\Omega \subset \mathbb{R}^n$ be a compact set, we say it's a **bounded region with boundary** if $\forall x \in \partial\Omega$, there exists open sets $U, V \subset \mathbb{R}^n$, $x \in U$ and a continuous homeomorphism $\Phi : U \to V$, such that

$$\Phi(U \cap \Omega) = V \cap \{x_n \geq 0\}, \quad \Phi(U \cap \partial\Omega) = V \cap \{x_n = 0\}.$$

If Φ is also C^1, we say $x \in \partial\Omega$ is a **regular point**, otherwise a **singular point**.

Lemma 6.2.2

Let Ω be a bounded region with boundary, for all regular $p \in \partial\Omega$, there exists a unique unit vector $\nu(p) \in \mathbb{R}^n$, and $\varepsilon > 0$, s.t.

$$\nu(p) \perp T_p\partial\Omega, |\nu(p)| = 1, p - t\nu(p) \in \Omega, \forall 0 < t < \varepsilon.$$

We call $\nu(p)$ the **outward unit normal vector** of p.

Proof. By the definition of regular points, we may assume that:

$$\Omega \cap V = \{x \in V \mid f(x) \geq 0\}, \quad \partial\Omega \cap V = \{x \in V \mid f(x) = 0\}.$$

Where f is a C^1 function.

Since ∇f is nonzero, the tangent space $T_p\partial\Omega = \{v \mid v \cdot \nabla f = 0\}$.

Let $\nu(p) = -\frac{\nabla f}{|\nabla f|}$, then it's obvious $\nu(p)$ points outside of Ω. □

Now for a cuboid I and a C^1 function ϕ,

$$\begin{aligned}\int_I \frac{\mathrm{d}\phi}{\mathrm{d}x_n}\,\mathrm{d}x &= \int_{I_{n-1}} \phi(x_1, \ldots, x_{n-1}, b_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} - \int_{I_{n-1}} \phi(x_1, \ldots, x_{n-1}, a_n)\,\mathrm{d}x_1 \cdots \mathrm{d}x_{n-1} \\ &= \int_{\partial I} \phi \cdot \nu_n \,\mathrm{d}\sigma.\end{aligned}$$

Where σ is the measure on the boundary, ν is the outward unit normal vector.

Lemma 6.2.3

Let K be a compact set in $\mathbb{R}^n$, $U \supset K$ is open, there exists a smooth function f such that $\operatorname{supp} f \subset U$, and $f\big|_K > 0$.

Proof. Let $\rho(x)$ be a smooth function s.t. $\rho(x) = 1$ for $|x| \leq 1$ and $\rho(x) = 0$ for $|x| \geq 2$. Let

$$g(x) = \int_{|y| \leq 2} f(x - \delta y)\rho(y)\,\mathrm{d}y.$$

Then g is a smooth non-negative function. □

Theorem 6.2.4 (Unit decomposition on compact sets)
Let K be a compact set, $\{U_1, \dots, U_k\}$ is an open covering of K. There exists smooth functions $f_1, \dots, f_k$ s.t.
$$1 = f_1(x) + f_2(x) + \cdots + f_k(x), \quad \operatorname{supp} f_i(x) \subset U_i.$$

Proof. For $1 \le i \le k$, $\delta > 0$, define
$$K_i^\delta = \{x \in U_i \mid d(x, U_i^c) > \delta\}.$$
Note that $\{\bigcup_{i=1}^k K_i^{\frac{1}{m}}\}_{m=1}^\infty$ is also an open covering of K, thus there exists N s.t.
$$K \subset \bigcup_{i=1}^k K_i^{\frac{1}{N}}.$$
Hence by lemma we have g_i s.t. $\operatorname{supp} g_i \subset U_i$ and $g_i > 0$ on the closure of $K_i^{\frac{1}{N}}$. Similarly we have a smooth function g s.t. $g(x) = 0$ on K, and $g > 0$ outside of the closure of $\bigcup_{i=1}^k K_i^{\frac{1}{N}}$.

Let $G(x) = g_1(x) + \cdots + g_k(x) + g(x) > 0$ on $\bigcup_{i=1}^k U_i$, then we can define $f_i(x) = \frac{g_i(x)}{G(x)}$ which satisfy the condition. ☐

Theorem 6.2.5
Let Φ be a C^1 homeomorphism from a cuboid I to Ω, then Ω satisfies Stolkes' formula:
$\forall \phi \in C^1(\mathring{D}) \cap C(\overline{D})$, we have
$$\int_D \nabla\phi \,\mathrm{d}x = \int_{\partial D} \phi\nu \,\mathrm{d}\sigma.$$

Proof. Since $\Omega = \Phi(I)$, let y be the coordinates on I, $x = \Phi(y)$,
$$\int_{\Phi(I)} \nabla\varphi \,\mathrm{d}x = \int_I \nabla\varphi(\Phi(y))(\mathrm{d}\Phi)^{-1} J_\Phi \,\mathrm{d}y.$$
Let $A = \mathrm{d}\Phi$, WLOG $J_\Phi > 0$. Using the index notation and Einstein summation,
$$A_{kj}A^{ji} = A^{kj}A_{ji} = \delta_{ki}.$$
Thus
$$\partial_{y_j}\varphi A^{ji}|A| = \partial_{y_j}(\varphi A^{ji}|A|) - \varphi\partial_{y_j}(A^{ji}|A|)$$
Since $|A| = A_{kl}A^{kl}|A|$, $A_{kl} = \frac{\partial\Phi_k}{\partial y_l}$.
$$\begin{aligned}\partial_{y_j}(A^{ji}|A|) &= |A|\partial_{y_j}A^{ji} + A^{ji}\partial_{y_j}|A| \\ &= |A|\partial_{y_j}A^{ji} + A^{ji}|A|\partial_{y_j}A_{kl}A^{kl} \\ &= |A|(\partial_{y_j}A^{ji} + \partial_{y_l}A_{kj}A^{kl}) \\ &= |A|(\partial_{y_j}A^{ji} - \partial_{y_j}A^{ji}) = 0.\end{aligned}$$
Hence by our previous work,
$$\int_I \partial_{y_j}(\varphi A^{ji}|A|) \,\mathrm{d}y = \int_{\partial I} \varphi A^{ji}|A|\nu_j \,\mathrm{d}\sigma.$$

Putting this together for all i's, note that $\widetilde{\nu} = \frac{\nabla \Phi_n^{-1}}{|\nabla \Phi_n^{-1}|}$,

TODO $\square$

Let $(\phi_1, \ldots, \phi_n)$ be an element in the tangent boundle TM, it can represent a vector field

$$X = (\phi_1, \ldots, \phi_n) = \sum_{i=1}^{n} \phi_i \frac{\partial}{\partial x_i}.$$

Here $X \in TM, X(p) \in T_pM$.

We define the **divergence** of X to be

$$\operatorname{div}(X) = \sum_{i=1}^{n} \frac{\partial \phi_i}{\partial x_i}.$$

The Stolke's formula can be presented as divergence theorem:

Theorem 6.2.6 (Divergence theorem)
Let X be a vector field,

$$\int_D \operatorname{div}(X) \, \mathrm{d}x = \int_{\partial D} X \cdot \nu \, \mathrm{d}\sigma.$$

Another commonly-used operator is the **Laplace operator**:

$$\Delta = \operatorname{div} \cdot \nabla, \quad \Delta\phi = \operatorname{div}(\nabla\phi) = \operatorname{tr}(H_\phi) = \sum_{i=1}^{n} \frac{\partial^2 \phi}{\partial x_i^2}.$$

When $n = 2$, we have $X = P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y}$, $\operatorname{div}(X) = \frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y}$,

$$\int_D \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} \right) \mathrm{d}x \, \mathrm{d}y = \int_{\partial D} X \cdot \nu \, \mathrm{d}\sigma.$$

Since ∂D is a curve $\gamma(t)$, so $\mathrm{d}\sigma = |\gamma'(t)| \, \mathrm{d}t$. Let $\gamma(t) = (x(t), y(t))$, then $\nu(t) = \frac{(y'(t), -x'(t))}{|\gamma'(t)|}$. Here we must take $\gamma(t)$ to be *counterclockwise* to ensure ν points outside of D.

Thus we get

$$\int_{\partial D} X \cdot \nu \, \mathrm{d}\sigma = \int_\gamma \frac{Py'(t) - Qx'(t)}{|\gamma'(t)|} |\gamma'(t)| \, \mathrm{d}t = \int_{\partial D} (P \, \mathrm{d}y - Q \, \mathrm{d}x).$$

This result is known as *Green's formula*.

This leads to the curve integrals of the second type: let $\gamma(t) \in \mathbb{R}^d$, X a vector field, we call the integral

$$\int_\gamma \sum_{i=1}^{d} X^i \, \mathrm{d}x_i = \int_\gamma X \cdot \mathrm{d}\gamma(t).$$

the **curve integral of the second type**.

When $n = 3$, the result is called *Gauss's formula*, we have $X = P\frac{\partial}{\partial x} + Q\frac{\partial}{\partial y} + R\frac{\partial}{\partial z}$,

$$\int_D \operatorname{div}(X) \, \mathrm{d}x \, \mathrm{d}y \, \mathrm{d}z = \int_{\partial D} X \cdot \nu \, \mathrm{d}\sigma.$$

Let $\gamma(u,v) = (x,y,z)$ be a parametrization of ∂D. We have two tangent vector γ_u, γ_v, so the normal vector is defined as $\nu = \frac{\gamma_u \times \gamma_v}{|\gamma_u \times \gamma_v|}$. Also $\mathrm{d}\sigma = |\gamma_u \times \gamma_v| \,\mathrm{d}u \,\mathrm{d}v$. After some computation we can get

$$\nu \,\mathrm{d}\sigma = (\mathrm{d}y \,\mathrm{d}z, \mathrm{d}z \,\mathrm{d}x, \mathrm{d}x \,\mathrm{d}y).$$

$$\int_D \mathrm{div}(X) \,\mathrm{d}x \,\mathrm{d}y \,\mathrm{d}z = \int_{\partial D} (P \,\mathrm{d}y \,\mathrm{d}z + Q \,\mathrm{d}z \,\mathrm{d}x + R \,\mathrm{d}x \,\mathrm{d}y).$$

§6.3 Differential forms

Let $T_p^* M$ denote the *dual space* of $T_p M$, and $\mathrm{d}x_i$ is the dual basis of $\frac{\partial}{\partial x_i}$. The linear combination of $\mathrm{d}x_i$ are called **differential forms**, and a differential form on a manifold can be written as $\sum_{i=1}^n a_i \,\mathrm{d}x_i$, where a_i are functions on M.

We can construct differential forms of higher order, the order is $1 \le k \le n$, called **k-forms**, which is a linear combination of

$$\mathrm{d}x_{i_1} \,\mathrm{d}x_{i_2} \cdots \mathrm{d}x_{i_k}, \quad i_1 < i_2 < \cdots < i_k.$$

Here the product is *wedge product*, i.e. $\mathrm{d}x_i \,\mathrm{d}x_j = -\,\mathrm{d}x_j \,\mathrm{d}x_i$. We denote the space of all k-forms by $\Lambda^k(\Omega)$.

We can define the multiplication of forms: let $\omega_1 \in \Lambda^{k_1}, \omega_2 \in \Lambda^{k_2}$, then $\omega_1 \wedge \omega_2 \in \Lambda^{k_1+k_2}$ by multiplying the coefficients and $\mathrm{d}x_i$'s respectively.

There's also an operator called **exterior differentiation** $\mathrm{d} : \Lambda^k \to \Lambda^{k+1}$, where

$$\mathrm{d}(a \,\mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_k}) = \sum_{j=1}^n \frac{\partial a}{\partial x_j} \,\mathrm{d}x_j \,\mathrm{d}x_{i_1} \cdots \mathrm{d}x_{i_k}.$$

This operator behaves like the derivatives very much:

$$\mathrm{d}(\omega_1 + \omega_2) = \mathrm{d}\omega_1 + \mathrm{d}\omega_2, \quad \mathrm{d}(\omega_1 \omega_2) = \mathrm{d}\omega_1 \wedge \omega_2 + (-1)^{k_1 k_2} \omega_1 \wedge \mathrm{d}\omega_2.$$

Note that $\omega_1 \wedge \omega_2 = (-1)^{k_1 k_2} \omega_2 \wedge \omega_1$, so when $k_1 k_2$ is even, the wedge product may not be anti-symmetrical.

If we have a coordinate transformation $\Phi : (x_1, \ldots, x_n) \to (y_1, \ldots, y_n)$, we have

$$\mathrm{d}y_i = \sum_{j=1}^n \frac{\partial y_i}{\partial x_i} \,\mathrm{d}x_i,$$

thus $\mathrm{d}y_1 \cdots \mathrm{d}y_n = J_\Phi \,\mathrm{d}x_1 \cdots \mathrm{d}x_n$. Here J_Φ can be negative, so the differential forms already contains the information of orientation.

Theorem 6.3.1

Let ω be a differential form, $\mathrm{d}(\mathrm{d}\omega) = 0$.

Proof. Partial derivatives commute. □

Definition 6.3.2. Let ω be a differential form, if $\mathrm{d}\omega = 0$, we say ω is a **closed form**, if there exists ω_1 s.t. $\mathrm{d}\omega_1 = \omega$, then ω is a **exact form**.

The theorem above tells us that exact forms must be closed, but in general closed forms may not be exact, it depends on the topology structure of Ω.

Theorem 6.3.3 (Poincare)
The closed forms on $\mathbb{R}^n$ must be exact.

Proof. Use induction, when ω is an n-form this can be proved by computation.

For a generic form $\omega = \omega_1 + \mathrm{d}x_1 \wedge \omega_2$, where ω_1, ω_2 do not contain $\mathrm{d}x_1$. We want to find ω_3 s.t. $\mathrm{d}\omega_3 = \mathrm{d}x_1 \wedge \omega_2 + \omega_4$, where ω_3, ω_4 don't contain $\mathrm{d}x_1$ as well. (The construction is direct)

Since $\omega - \mathrm{d}\omega_3 = \omega_1 - \omega_4$, and

$$\mathrm{d}(\omega - \mathrm{d}\omega_3) = \mathrm{d}\omega = 0 \implies \mathrm{d}(\omega_1 - \omega_4) = 0.$$

Since $\mathrm{d}(\omega_1 - \omega_4) = 0$ and it doesn't contain $\mathrm{d}x_1$, hence all its coefficients can't contain $\mathrm{d}x_1$. Thus we can view it as a differential form in $\mathbb{R}^{n-1}$. □

Remark 6.3.4 — When Ω is simply connected, then all the closed 1-forms are exact. Also this is equivalent to the integral on any closed curves are 0.

We can rewrite Stolkes' formula using differential forms:

Theorem 6.3.5 (Stolkes' formula)
Let D be a $k+1$ dimensional orientable manifold, $\omega \in \Lambda^k(\mathbb{R}^n)$, we have

$$\int_D \mathrm{d}\omega = \int_{\partial D} \omega.$$