Mathematical Analysis II

Felix Chen

Contents

Now we can complete the proof of **??**. 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 $F' \ge 0$)

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

§0.1 Absolute continuous functions

Definition 0.1 (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 0.2

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

Proposition 0.3

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 0.4 (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 0.5

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 0.6 (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 0.7 (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 0.7, 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 $a \le a_1 < b_1 < \cdots < a_N < b_N \le b$, by absolute continuity and $|F(b_k) - F(a_k)| < \varepsilon(b_k - a_k)$,

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

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 0.8

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 0.9

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

$$m^*(F(E)) \le Mm^*(E).$$

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 (0.7), exists disjoint intervals $I_i = [x_i - h_i, x_i + h_i]$ s.t.

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

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

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

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

Hence

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

Theorem 0.10

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

$\forall \varepsilon > 0$, 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 $T_\gamma([a,b]) \geq T_G([a,b]) - T_H([a,b])$.

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

which gives the opposite inequality. □

Proposition 0.11 (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 $m(\phi(E)) = \int_E \phi' \, \mathrm{d}x$. □

§1 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.

§1.1 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 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 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 1.3 — The partial derivatives rely on the coordinate, but the directional derivatives is independent of the coordinate (i.e. geometry quantities).

Example 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 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 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 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 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 $\nabla f = (\frac{\partial f}{\partial x_1}, \dots, \frac{\partial f}{\partial x_n})$ 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 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 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 $u_j = (v_1, \dots, v_j, 0, \dots, 0)$.

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

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 1.11

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

Proposition 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.

Definition 1.13. 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 1.14

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

Corollary 1.16
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 $n_1 \ge n_2$. Similarly $n_2 \ge n_1$, so $n_1 = n_2$.

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

Example 1.17

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 1.18 (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 $f(U)$, then

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

Remark 1.19 — 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.

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,

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

$\square$