Mathematical Analysis II

Felix Chen

Contents

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

Theorem 0.0.1

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 0.0.2

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

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

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

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

But for $y \in I_i$, $|F(y) - F(x_i)| \leq (M+\varepsilon)h_i$, 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

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

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

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

Thus

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

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. $\square$

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

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

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

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

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 0.0.3

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 0.0.4 (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 1.1.1 (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 1.1.2 (Partial derivatives)**.** Let $(x_1, \dots, x_n)$ be a coordinate system, let $e_i = (0, \dots, 1, \dots, 0)$ be the i-th vector of the standard basis. The directional derivative in e_i

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

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

Remark 1.1.3 — The partial derivatives rely on the coordinate, but the directional derivatives is independent of the coordinate (i.e. geometry quantities).

Example 1.1.4

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 1.1.5

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 1.1.6 (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 1.1.7 — 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 1.1.8

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 1.1.9

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 1.1.10

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 1.1.11
If f is differentiable on Ω, and $\mathrm{d}f = 0$, then f is constant on Ω.

Proposition 1.1.12
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.

§1.2 Jacobi matrices

Definition 1.2.1. 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 1.2.2
Let $f = (f_1, \ldots, f_{d'})$. 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 1.2.3 (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 1.2.4

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 1.2.5

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 $\|AB\| \le \|A\|\|B\|$, 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 1.2.6 (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 1.2.7 — 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.

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

$\forall \varepsilon > 0$, 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:

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

□

Example 1.2.8

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

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

§1.3 Implicit function theorem

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

Definition 1.3.1 (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 1.3.2

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, $f' \neq 0$, then $f^{-1} \in C^1$.

Theorem 1.3.3

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 $\frac{\partial \Phi^* f}{\partial x_j} \in C^n$. □

Theorem 1.3.4 (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 1.3.5 (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.

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

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. □