MATHEMATICAL ANALYSIS (II)

SHIBO LIU

CONTENTS

email: shibo.liu@xmu.edu.my.

Foreword

(1) Instead of $\mathbf{u}$, $\boldsymbol{u}$ or $\overrightarrow{u}$, we will simply use x, y of u to denote points (or vectors) in $\mathbb{R}^n$. This makes the symbols simple and *elegant*.

(2) For $x \in \mathbb{R}^n$, we write

$$x = (x^1, \dots, x^n).$$

That is, instead of subscript x_i, we prefer to use superscript x^i to denote the i-th coordinate or component of $x \in \mathbb{R}^n$.

This enable us to write the chain rule for partial derivative of composition $\mathbb{R}^n \to \mathbb{R}^m \to \mathbb{R}^\ell$, $x \mapsto y \mapsto u$ by

$$\frac{\partial u^k}{\partial x^i} = \sum_{j=1}^{m} \frac{\partial u^k}{\partial y^j} \frac{\partial y^j}{\partial x^i}. \tag{0.1}$$

eq0

If we use a^i_j to denote the i-th row j-th column entry of a matrix

$$A = \begin{pmatrix} a^1_1 & \cdots & a^1_n \\ \vdots & & \vdots \\ a^m_1 & \cdots & a^m_n \end{pmatrix},$$

that is, superscript for row number and subscript for column number, then (0.1) can be easily written in matrix form. This is a big advantage for using superscript to denote coordinates of $x \in \mathbb{R}^n$.

Thus, this convention has been widely adopted in mathematics and physics, in particular differential geometry.

(3) Instead of $\|x\|$, for simplicity the Euclidean norm of $x \in \mathbb{R}^n$ will be denote by $|x|$. You should not confuse this with the absolute value of a real (or complex) number. The following facts *justify* our notation:

  (a) If $n = 1$ then $x \in \mathbb{R}^n = \mathbb{R}^1$ can be regarded as a real number and the Euclidean norm of x is precisely its absolute value.
  From the context, it is easy to distinguish the meaning of $|x|$: if x is a real number, then $|x|$ is absolute value (which is exactly Euclidean norm in $\mathbb{R}^1$; if $x \in \mathbb{R}^n$, then $|x|$ is Euclidean norm.

  (b) In future study, you may encounter *norm*s for elements in abstract spaces, we leave $\|\cdot\|$ for these more advanced cases.
  For example, Let $\mathfrak{B}(\mathbb{R}^n)$ be the vector space consisted of *bounded* functions $u : \mathbb{R}^n \to \mathbb{R}$. We define the L^∞-norm of u via

$$\|u\| = \sup_{x \in \mathbb{R}^n} |u(x)|.$$

  In this notation, a sequence of functions $\{u_k\} \subset \mathfrak{B}(\mathbb{R}^n)$ converges to $u \in \mathfrak{B}(\mathbb{R}^n)$ is simply the sequence of real numbers $\{\|u_k - u\|\}$ converges to zero: $\|u_k - u\| \to 0$.
  Of course $\mathfrak{B}(\mathbb{R}^n)$ is a more *advanced* space than $\mathbb{R}^n$. Hence it is reasonable to use $\|\cdot\|$ for complicated spaces like $\mathfrak{B}(\mathbb{R}^n)$ and $|\cdot|$ for $\mathbb{R}^n$.

Note that many books use $\|\mathbf{u}\|$ to denote Euclidean norm of $\mathbf{u} \in \mathbb{R}^n$. You should adapt for such conventions when you are reading other books or papers.

1. Euclidean Space

1.1. Linear structure, scalar product.

- $\mathbb{R}^n = \{x = (x^1, \ldots, x^n) \mid x^i \in \mathbb{R}\}$, x^i is i-th coordinate of x.
- Instead of x_i, we prefer to denote coordinate of x by x^i.
- $x = y$ iff $x^i = y^i$.
- Addition $+$ and scalar multiplication $\cdot$,

$$x + y = (x^1 + y^1, \ldots, x^n + y^n),$$
$$\lambda \cdot x = \lambda x = (\lambda x^1, \ldots, \lambda x^n).$$

  A special element $0 = (0, \ldots, 0)$, called the origin or zero.
- *Proposition.* $+ : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ and $\cdot : \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n$ has the followling properties:
  (1) $x + y = y + x$, $(x + y) + z = x + (y + z)$,
  (2) $x + 0 = x$, $x + (-x) = 0$.
  (3) $1x = x$, $(\lambda\mu)\, x = \lambda\, (\mu x)$,
  (4) $(\lambda + \mu)\, x = \lambda x + \mu x$, $\lambda\, (x + y) = \lambda x + \lambda y$.
  This means $(\mathbb{R}^n, +, \cdot)$ is a vector space.
- We call $x \in \mathbb{R}^n$ points. They can be identified with vectors (arrows) from 0 to x.
- To do calculus we need to measure how close are two points. *Metric structure.*
- The scalar (or inner) product of x and y is

$$\langle x, y \rangle = x \cdot y = \sum_{i=1}^{n} x^i y^i.$$

  Equiped with $\cdot$, $\mathbb{R}^n$ is called Euclidean space.
- *Proposition.*
  (1) $\langle x, x \rangle \geq 0$, $\langle x, x \rangle = 0$ iff $x = 0$,
  (2) $\langle x, y \rangle = \langle y, x \rangle$, (in dot notation, $x \cdot y = y \cdot x$)
  (3) $\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$.
  Thus $\langle x, \lambda y + \mu z \rangle = \lambda \langle x, y \rangle + \mu \langle x, z \rangle$ as well.
- The norm of x is

$$\|x\| = \langle x, x \rangle^{1/2},$$

  For simplicity we may also write $|x| = \sqrt{x \cdot x}$. When $n = 1$, we identify $(x) \in \mathbb{R}^1$ with $x \in \mathbb{R}$, then $|x|$ is the absolute value (justify $|x|$ even $n > 1$).
- *Cauchy-Schwarz.* $|x \cdot y| \leq |x|\,|y|$.

  Proof. For $\forall t \in \mathbb{R}$,

$$0 \leq (x + ty) \cdot (x + ty)$$

$$= x \cdot x + t\,(x \cdot y) + t\,(y \cdot x) + t^2\,(y \cdot y)$$
$$= |x|^2 + 2t\,(x \cdot y) + t^2\,|y|^2 .$$

The discriminant has to be nonpositive, i.e.,

$$[2\,(x \cdot y)]^2 - 4\,|y|^2\,|x|^2 \le 0.$$

- *Proposition.*
  (1) $|x| \ge 0$, $|x| = 0$ iff $x = 0$.
  (2) $|\lambda x| = |\lambda|\,|x|$, ($\|\lambda x\| = |\lambda|\,\|x\|$)
  (3) $|x + y| \le |x| + |y|$. (triangle inequality)

  Proof. We only prove (3). By Cauchy-Schwarz,

$$\begin{aligned} |x + y|^2 &= (x + y) \cdot (x + y) \\ &= |x|^2 + 2\,(x \cdot y) + |y|^2 \\ &\le |x|^2 + 2\,|x|\,|y| + |y|^2 = (|x| + |y|)^2 . \end{aligned}$$

- We call $d(x, y) = |x - y|$ the distance between x and y. The name is justified by properties:
  (1) $d(x, y) \ge 0$, $d(x, y) = 0$ iff $x = y$,
  (2) $d(x, y) = d(y, z)$,
  (3) $d(x, y) \le d(x, z) + d(z, y)$. (triangle inequality)

  Proof. We only prove (3).

$$\begin{aligned} d(x, y) &= |x - y| = |(x - z) + (z - y)| \\ &\le |x - z| + |z - y| = d(x, z) + d(z, y). \end{aligned}$$

- If $x = \left(x^1, \ldots, x^n\right)$, $y = \left(y^1, \ldots, y^n\right)$, then

$$d(x, y) = |x - y| = \sqrt{\sum_{i=1}^{n} (x^i - y^i)^2},$$

  thus our distance is generalization of $n = 2, 3$ we learned in analytic geometry.
- For $a \in \mathbb{R}^n$, $r > 0$, the ball centered at a with radius r is

$$B_r(a) = \{x \in \mathbb{R}^n |\ |x - a| < r\}, \qquad B_r = B_r(0).$$

  We also call $B_r(a)$ an r-neighborhood of a.
- Subset $S \subset \mathbb{R}^n$ is bounded if $\exists R > 0$ such that $S \subset B_R$. This means $\forall x \in S$, $|x| < R$.
- By Cauchy-Schwarz, if $x \neq 0$, $y \neq 0$, the unique $\theta \in [0, \pi]$ s.t.

$$\cos\theta = \frac{x \cdot y}{|x|\,|y|}$$

  is called the angle between x and y.

  If $x \cdot y = 0$, then $\theta = \pi/2$. We say that x and y are orthogonal, $x \perp y$.
- *Remarks.*

(1) Scalar multiplication λx, scalar product $\langle x, y\rangle = x \cdot y$.
(2) Norm $\|x\| = |x|$. $B_r = B_r(0)$.
(3) If $\exists r > 0$ s.t. $S \subset B_r$, then for $\forall a \in \mathbb{R}^n$, $\exists \rho > 0$ s.t. $S \subset B_\rho(a)$.
In fact, $\forall x \in S$, we have $|x| < r$. Hence

$$|x - a| = |x| + |a| < r + |a|,$$

let $\rho = r + |a|$ we have $x \in B_\rho(a)$.

1.2. Convergence in $\mathbb{R}^n$.

- A sequence in $\mathbb{R}^n$ is a map $x : \mathbb{N} \to \mathbb{R}^n$. Set $x_k = x(k)$, it can be denoted by $\{x_k\}_{k=1}^{\infty}$.
- For $\{x_k\} \subset \mathbb{R}^n$, $a \in \mathbb{R}^n$. We say $x_k \to a$ if $|x_k - a| \to 0$. $\{x_k\}$ converges to a.
- That is, $\forall \varepsilon > 0$, $\exists K$, if $k \geq K$ then $|x_k - a| < \varepsilon$. We also write $\lim\limits_{k\to\infty} x_k = a$ for $x_k \to a$.
- Convergent sequences in $\mathbb{R}^n$ are bounded.

  Proof. $x_k \to a$ means $|x_k - a| \to 0$. Since convergent sequences of real numbers are bounded, $\exists R > 0$ such that $|x_k - a| \leq R$. Hence

$$\begin{aligned} |x_k| &= |(x_k - a) + a| \\ &\leq |x_k - a| + |a| \leq R + |a|. \end{aligned}$$

- *Theorem.* Suppose $x_k = \left(x_k^1, \ldots, x_k^n\right)$, $a = \left(a^1, \ldots, a^n\right)$. Then as $k \to \infty$,

$$x_k \to a \quad \Longleftrightarrow \quad x_k^i \to a^i \text{ for all } i.$$

  Proof. In (1.1) below take $A^i = x_k^i - a^i$.
  ($\Rightarrow$) Suppose $x_k \to a$. For all j, by the first inequality in (1.1)

$$\begin{aligned} \left|x_k^j - a^j\right| &\leq \sqrt{\sum_{i=1}^n \left(x_k^i - a^i\right)^2} \\ &= |x_k - a| \to 0. \end{aligned}$$

  Thus $x_k^j \to a^j$.
  ($\Leftarrow$) If $x_k^i \to a^i$ for all i, then by the second inequality in (1.1)

$$\begin{aligned} |x_k - a| &= \sqrt{\sum_{i=1}^n \left(x_k^i - a^i\right)^2} \\ &\leq \sum_{i=1}^n \left|x_k^i - a^i\right| \to 0. \end{aligned}$$

- This follows from

$$|A^j| \leq \sqrt{\sum_{i=1}^n (A^i)^2} \leq \sum_{i=1}^n |A^i|. \tag{1.1}$$

e1

Proof. The first is obvious. The second follows from

$$\begin{aligned}\left(\sum_{i=1}^{n}|A^i|\right)^2 &= \sum_{i=1}^{n}\sum_{j=1}^{n}|A^i|\,|A^j| \\ &= \sum_{i=1}^{n}\left(A^i\right)^2 + \sum_{i\neq j}|A^i|\,|A^j| \\ &\geq \sum_{i=1}^{n}\left(A^i\right)^2 .\end{aligned}$$

- *Proposition*. If $x_k \to a$, $y_k \to b$, then

$$\lambda x_k + \mu y_k \to \lambda a + \mu b, \qquad x_k \cdot y_k \to a \cdot b. \tag{1.2}$$

e2

Proof. From $x_k \to a$, $y_k \to b$ we have $x_k^i \to a^i$, $y_k^i \to b^i$ for all i. Thus $x_k^i y_k^i \to a^i b^i$ and

$$x_k \cdot y_k = \sum_{i=1}^{n} x_k^i y_k^i \to \sum_{i=1}^{n} a^i b^i = a \cdot b.$$

Proof. Since $x_k \to a$, $\exists R > 0$ s.t. $|x_k| \leq R$.

$$\begin{aligned}|x_k \cdot y_k - a \cdot b| &= |(x_k \cdot y_k - x_k \cdot b) + (x_k \cdot b - a \cdot b)| \\ &\leq |x_k \cdot y_k - x_k \cdot b| + |x_k \cdot b - a \cdot b| \\ &= |x_k \cdot (y_k - b)| + |(x_k - a) \cdot b| \\ &\leq |x_k|\,|y_k - b| + |x_k - a|\,|b| \\ &\leq R\,|y_k - b| + |x_k - a|\,|b| \to 0.\end{aligned}$$

1.3. **Open & closed sets.**

- Let $S \subset \mathbb{R}^n$, $a \in \mathbb{R}^n$. If $\exists \varepsilon > 0$ s.t. $B_\varepsilon(a) \subset S$, we call a an interior point of S.
- All interior points of S consist S°, the interior of S; also denoted by $\operatorname{int} S$.
- We always have $S^\circ \subset S$.

Proof. If $a \in S^\circ$, $\exists \varepsilon > 0$ s.t. $B_\varepsilon(a) \subset S$. But $a \in B_\varepsilon(a)$, so $a \in S$.

- For sets A_i, $i = 1, \ldots, n$, the Cartesian product

$$A_1 \times \cdots \times A_n = \{(a_1, \ldots, a_n) \,|\, a_i \in A_i\}.$$

Thus $\mathbb{R}^n = \mathbb{R} \times \cdots \times \mathbb{R}$.

- *Example*. For $Q = [a,b] \times [c,d]$, $Q^\circ = (a,b) \times (c,d)$.

Proof. Take $p = (x,y) \in Q^\circ$. Then $\exists \varepsilon > 0$ s.t. $B_\varepsilon(p) \subset Q$. Computing the distance between points in H and p we can see

$$H = \left(x - \frac{\sqrt{2}}{2}\varepsilon, x + \frac{\sqrt{2}}{2}\varepsilon\right) \times \left(y - \frac{\sqrt{2}}{2}\varepsilon, y + \frac{\sqrt{2}}{2}\varepsilon\right)$$

$$\subset B_\varepsilon(p) \subset Q = [a,b] \times [c,d].$$

Hence

$$\left(x - \frac{\sqrt{2}}{2}\varepsilon, x + \frac{\sqrt{2}}{2}\varepsilon\right) \subset [a,b],$$

$x \in (a,b)$. Similarly $y \in (c,d)$. So $p \in (a,b) \times (c,d)$.

Now if $p \in (a,b) \times (c,d)$, then

$$a < x < b, \qquad c < y < d.$$

Take

$$\varepsilon = \min\{x-a, b-a, y-c, d-y\} > 0,$$

note that ε is the minimum distance from p to the boundary of Q, e have

$$\begin{aligned} B_\varepsilon(p) &\subset (x-\varepsilon, x+\varepsilon) \times (y-\varepsilon, y+\varepsilon) \\ &\subset [a,b] \times [c,d] = Q. \end{aligned}$$

Therefore $p \in Q^\circ$.

- *Definition.* S is open if $S = S^\circ$.
- Since $S^\circ \subset S$ is automatically true, to prove that S is open we only need to show $S \subset S^\circ$.
- *Example.* $B_r(a)$ is open.

Proof. Fox $x \in B_r(a)$,

$$\varepsilon = r - |x-a| > 0.$$

We claim that

$$B_\varepsilon(x) \subset B_r(a), \tag{1.3}$$ e3

which implies that $x \in [B_r(a)]^\circ$ and $B_r(a)$ is open.

To prove (1.3), take $y \in B_\varepsilon(x)$ and estimate

$$\begin{aligned} |y-a| &\le |y-x| + |x-a| \\ &< \varepsilon + |x-a| = r. \end{aligned}$$

- *Proposition.*
  (1) If S_λ is open, so is $\bigcup_\lambda S_\lambda$.
  (2) If S_1 and S_2 are open, so is $S_1 \cap S_2$.

  Note that by definition, for a family of sets $\{S_\lambda\}_{\lambda\in\Lambda}$,

$$\bigcup_{\lambda\in\Lambda} S_\lambda = \{x \mid x \in S_\lambda \text{ for } \textit{some } \lambda \in \Lambda\},$$

$$\bigcap_{\lambda\in\Lambda} S_\lambda = \{x \mid x \in S_\lambda \text{ for } \textit{all } \lambda \in \Lambda\}.$$

Proof. (1) For $a \in \bigcup_\lambda S_\lambda$, $\exists \lambda' \in \Lambda$ s.t. $a \in S_{\lambda'}$. Since $S_{\lambda'}$ is open, $\exists \varepsilon > 0$ s.t. $B_\varepsilon(a) \subset S_{\lambda'}$. Hence $B_\varepsilon(a) \subset \bigcup_\lambda S_\lambda$ and $a \in \left(\bigcup_\lambda S_\lambda\right)^\circ$.

(2) For $a \in S_1 \cap S_2$, $\exists \varepsilon_1 > 0$ s.t. $B_{\varepsilon_1}(a) \subset S_1$ because $a \in S_1$ and S_1 is open. Similarly, $\exists \varepsilon_2 > 0$ s.t. $B_{\varepsilon_2}(a) \subset S_2$. Let $\varepsilon = \min\{\varepsilon_1, \varepsilon_2\}$. Then $\varepsilon > 0$ and

$$B_\varepsilon(a) \subset B_{\varepsilon_1}(a) \cap B_{\varepsilon_2}(a) \subset S_1 \cap S_2,$$

which implies $a \in (S_1 \cap S_2)^\circ$.

- If $\{x_k\} \subset S$, we say $\{x_k\}$ is a sequence in S.
- We say that $S \subset \mathbb{R}^n$ is closed, if

$$\left.\begin{array}{r}\{x_k\} \subset S \\ x_k \to a\end{array}\right\} \Longrightarrow a \in S;$$

i.e., the limit of any convergent sequence in S belongs to S. *Closed under limit operation.*

- *Example*. $Q = [a, b] \times [c, d]$ is closed, while $C = [0, 2) \times [0, 2]$ is neither open nor closed.

Proof. Let $\{p_k\} \subset Q$, $p_k \to p$. Write $p_k = (x_k, y_k)$, $p = (x, y)$. Then $x_k \to x$, $y_k \to y$ and

$$a \le x_k \le b, \qquad c \le y_k \le d.$$

By the properties of convergent sequences of real numbers we see that

$$a \le x \le b, \qquad c \le y \le d.$$

Thus $p \in Q$.

The point $(1, 0) \in C \backslash C^\circ$, so C is not open. Let $p = (2, 1)$,

$$p_k = \left(2 - \tfrac{1}{k}, 1\right).$$

Then $p_k \in C$, $p_k \to p$ but $p \notin C$. Thus C is not closed.

- *Theorem*. S closed iff S^c is open.

Proof. ($\Rightarrow$) Take $a \in S^c$. If $a \notin (S^c)^\circ$, then for any $k \in \mathbb{N}$, $B_{1/k}(a)$ is not contained in S^c, thus $B_{1/k}(a) \cap S \ne \emptyset$. Take x_k from the intersection we get a sequence $\{x_k\} \subset S$, with

$$|x_k - a| < \frac{1}{k}.$$

So $\{x_k\}$ is a convergent sequence in S whose limit $a \notin S$, contradicting to the closedness of S.

($\Leftarrow$) Let $\{x_k\} \subset S$, $x_k \to a$. If $a \notin S$, then a belongs to the open set S^c. There is $\varepsilon > 0$ s.t. $B_\varepsilon(a) \subset S^c$. But $x_k \to a$, for large k, $x_k \in B_\varepsilon(a)$, thus $x_k \in S^c$, contradicting to $\{x_k\} \subset S$. Hence $a \in S$ and S is closed.

- DeMorgan's law

$$X \backslash \bigcup_\lambda A_\lambda = \bigcap_\lambda (X \backslash A_\lambda), \qquad X \backslash \bigcap_\lambda A_\lambda = \bigcup_\lambda (X \backslash A_\lambda).$$

Proof. For both equalities we need to show each side is a subset of the other side. For example we prove

$$X\backslash\bigcup_{\lambda} A_\lambda \subset \bigcap_{\lambda} (X\backslash A_\lambda).$$

Let $a \in X\backslash \bigcup_\lambda A_\lambda$, then $a \in X$, $a \notin \bigcup_\lambda A_\lambda$. Thus for $\forall\lambda$, $a \notin A_\lambda$; hence $a \in X\backslash A_\lambda$. Consequently $a \in \bigcap_\lambda (X\backslash A_\lambda)$.

- *Proposition.*
  (1) If S_λ is closed, so is $\bigcap_\lambda S_\lambda$.
  (2) If S_1 and S_2 are open, so is $S_1 \cup S_2$.

  Proof (Proof of (1)). Note that S_λ^{c} is open, hence

$$\left(\bigcap_{\lambda} S_\lambda\right)^{\mathrm{c}} = \bigcup_{\lambda} S_\lambda^{\mathrm{c}}$$

  is also open.

  Proof (Another proof of (1)). Let $\{x_k\} \subset \bigcap_\lambda S_\lambda$, $x_k \to a$. Then for $\forall\lambda$, $\{x_k\} \subset S_\lambda$; since S_λ is closed, we have $a \in S_\lambda$. Thus $a \in \bigcap_\lambda S_\lambda$ and $\bigcap_\lambda S_\lambda$ is closed.

- Let $S \subset \mathbb{R}^n$, $a \in \mathbb{R}^n$.
  (1) a is exterior point of S, if $\exists\varepsilon > 0$ s.t. $B_\varepsilon(a) \cap S = \emptyset$. All such a consist $\operatorname{ext} S$, the exterior of S.
  (2) a is boundary point of S if $\forall\varepsilon > 0$,

$$B_\varepsilon(a) \cap S \neq \emptyset, \qquad B_\varepsilon(a) \cap S^{\mathrm{c}} \neq \emptyset.$$

  All such a consist $\operatorname{bd} S$, the boundary of S. Also denote by ∂S.
  (3) Note that a is a boundary point of S iff a is a boundary point of S^{c}. Thus $\partial S = \partial S^{\mathrm{c}}$.
- *Example*. For $S = B_R$,

$$\operatorname{ext} S = \{x \in \mathbb{R}^n \mid |x| > R\},$$
$$\partial S = \{x \in \mathbb{R}^n \mid |x| = R\}.$$

- *Theorem*. S open iff $S \cap \partial S = \emptyset$.
- *Theorem*. S close iff $\partial S \subset S$.

2. Continuity, Compactness and Connectednes

2.1. Continuous functions and maps.

- *Vector valued functions*. $A \subset \mathbb{R}^n$, $f : A \to \mathbb{R}^m$.
- For $x \in A$, $f(x) \in \mathbb{R}^m$, $f^i(x)$ is the i-th coordinate of $f(x)$, new (real valued) functions $f^i : A \to \mathbb{R}$.
- $f : A \to \mathbb{R}^m$,

$$f(x) = \left(f^1(x), \dots, f^m(x)\right),$$

  we write $f = \left(f^1, \dots, f^m\right)$.

- *Defenition.* If $a \in A$, $f : A \to \mathbb{R}^m$ is continuous at a, if $f(x_k) \to f(a)$ whenever $\{x\} \subset A$ and $x_k \to a$.

  f is continuous, if it is continuous at all $x \in A$.
- *Example.* $f : (x, y) \mapsto \sin \sqrt{x^2 + y^2}$ is continuous on $\mathbb{R}^2$.
- *Proposition.* Let $f : A \to \mathbb{R}^m$, $a \in A$. Then f is continuous at a iff f^i is.
- *Continuity under operations.* If $f, g : A \to \mathbb{R}^m$ is continuous at $a \in A$, then $\lambda f + \mu g$, $f \cdot g$ are also continuous at a.

  If $m = 1$ and $g(a) \neq 0$, then f/g is also continuous at a.
- *Composition.* Suppose $A \subset \mathbb{R}^n$, $f : A \to \mathbb{R}^m$; $B \subset \mathbb{R}^m$, $g : B \to R^\ell$. If

  $$f(A) = \{ f(x) |\ x \in A\} \subset B,$$

  then for all $x \in A$ we can talk about $g(f(x))$.

  The map $g \circ f : A \to \mathbb{R}^\ell$ defined by

  $$(g \circ f)(x) = g(f(x))$$

  is called the composition of f and g.
- *Theorem.* If f is continuous at $a \in A$ and g is continuous at $b = f(a)$, then $g \circ f$ is continuous at a.
- *Example.* $f : \mathbb{R}^n \to \mathbb{R}$, $g : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}^n$,

  $$f(x) = |x|, \qquad g(x) = \frac{x}{|x|}.$$

  Both f and g are continuous, so is $f \circ g : \mathbb{R}^n \setminus \{0\} \to \mathbb{R}$. Note that $f \circ g$ is constant function.
- In Page 293, the discussion about the continuity of f is unnecessary complicated.
- *Theorem.* $f : A \to \mathbb{R}^m$ is continuous at $a \in A$ iff $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t. $|f(x) - f(a)| < \varepsilon$ for $x \in A \cap B_\delta(a)$, that is

  $$f(A \cap B_\delta(a)) \subset B_\varepsilon(f(a)).$$
- *Remark.* If A is open, then f is continuous at a iff $\forall \varepsilon > 0$, $\exists \rho > 0$. s.t.

  $$f(B_\rho(a)) \subset B_\varepsilon(f(a)).$$

  The reason is that $a \in A = A^\circ$, $\exists r > 0$ s.t. $B_r(a) \subset A$. Take $\rho = \min\{\delta, r\}$ we get $B_\rho(a) \subset A \cap B_\delta(a)$.
- *Theorem.* Let $A \subset \mathbb{R}^n$ open, $f : A \to \mathbb{R}^m$. Then f is continuous iff for any open subset $U \subset \mathbb{R}^m$, $f^{-1}(U)$ is open.
- Similarly, For closed A, $f : A \to \mathbb{R}^m$ is continuous iff for any closed subset $F \subset \mathbb{R}^m$, $f^{-1}(F)$ is closed. (De Morgan or independent proof)
- *Remark.*
  (1) By the theorem, to discuss continuity we only need to know which subsets are open. This leads to *topology*.

      Chapter 12, Metric space is a special case of topological space.

  (2) If A is neither open nor closed, we can still charaterize continuous maps by pre-image of open or closed sets in $\mathbb{R}^m$ via relative topology.

- If $f : \mathbb{R}^n \to \mathbb{R}$ is continuous, $c \in \mathbb{R}$, then $f^{-1}(-\infty, c)$ is open, $f^{-1}(-\infty, c]$ is closed.

  Since $\eta : \mathbb{R}^n \to \mathbb{R}$, $\eta(x) = |x|$ is continuous, for $r > 0$,
$$B_r = \eta^{-1}(-\infty, r)$$
  is open, $S_r = \partial B_r = \eta^{-1}(r)$ is closed.

2.2. **Sequential compactness, extreme value and uniform continuits.**

- *Notation for taking subsequences iteratively* (subsequence of subsequence). Subsequence of $\{x_k\} \subset \mathbb{R}^n$ will be denoted by $\left\{x_{k_i^{(1)}}\right\}_{i=1}^{\infty}$, the superscript (1) means the 1-st generation.

  $\left\{x_{k_i^{(1)}}\right\}$ is a sequence in its own right, whose subsequence (the 2-nd generation) will be denoted by $\left\{x_{k_i^{(2)}}\right\}$. In general, subsequence of the α-generation subsequence $\left\{x_{k_i^{(\alpha)}}\right\}$ is denoted by $\left\{x_{k_i^{(\alpha+1)}}\right\}$.

  For all α, $\left\{x_{k_i^{(\alpha)}}\right\}$ is subsequence of the original $\{x_k\}$.
- $x_{k_i^{(\alpha)}}$ is the i-th term in the α-generation subsequence.
- *Example*. If we denote the subsequence consisting by all terms with even index by $\left\{x_{k_i^{(1)}}\right\}$, then $x_{k_1^{(1)}} = x_2$, $x_{k_2^{(1)}} = x_4$, $x_{k_3^{(1)}} = x_6$.

  Denote by $\left\{x_{k_i^{(2)}}\right\}$ the subsequence of $\left\{x_{k_i^{(1)}}\right\}$ consisting by terms with odd index, then
$$x_{k_1^{(2)}} = x_{k_1^{(1)}} = x_2, \qquad x_{k_2^{(2)}} = x_{k_3^{(1)}} = x_6.$$
- We know that any bounded sequence in $\mathbb{R}$ has convergent subsequence.
- *Theorem*. If $\{x_k\} \subset \mathbb{R}^n$ is bounded, then it has a convergent subsequence.
- $A \subset \mathbb{R}^n$ is sequential compact, if whenever $\{x_k\} \subset A$, there exists $\{x_{k_i}\}$ s.t. $x_{k_i} \to a \in A$.
- *Theorem*. $A \subset \mathbb{R}^n$ is sequentially compact iff A is closed and bounded.
- Example. $\bar{B} = \{x \in \mathbb{R}^n | \, |x| \le R\}$ and $Q = [a, b] \times [c, d]$ are sequntially compact, so is n-dimensional rectangle
$$Q = \left[a^1, b^1\right] \times \cdots \times [a^n, b^n].$$
- *Theorem*. If A is sequentially compact, $f : A \to \mathbb{R}^m$ is continuous, then $f(A)$ is sequentially compact.
- *Note*. For the next result, students are expect to understand upper bounded, sup, max; as well as their differences.
- *Theorem*. If A is sequentially compact, $f : A \to \mathbb{R}$ is continuous, then there are $a, b \in A$ s.t.
$$f(a) = \inf_A f, \qquad f(b) = \sup_A f;$$
  a and b are called extreme (smallest, largest) points and $f(a)$, $f(b)$ extreme values.

- Noempty $A \subset \mathbb{R}^n$ has extreme value property (EVP), if any continuous $f : A \to \mathbb{R}$ has minimum and maximum value.
- *Uniform continuity.* $f : A \to \mathbb{R}^m$ is continuous means: $\forall a \in A$, $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t. $x \in A$,
$$|x - a| < \delta \quad \Longrightarrow \quad |f(x) - f(a)| < \varepsilon.$$
The δ depends on a, different a requires different δ. Uniform continuity means for any given $\varepsilon > 0$, there is $\delta > 0$ works for all $a \in A$.
- *Definition.* $f : A \to \mathbb{R}^m$ is uniformly continuous on A, if $\forall \varepsilon > 0$, $\exists \delta > 0$ s.t. for $x, y \in A$
$$|x - y| < \delta \quad \Longrightarrow \quad |f(x) - f(y)| < \varepsilon.$$
- *Proposition.* $f : A \to \mathbb{R}^m$ is uniformly continuous *iff* for $\{x_k\} \subset A$, $\{y_k\} \subset A$,
$$|x_k - y_k| \to 0 \quad \Longrightarrow \quad |f(x_k) - f(y_k)| \to 0.$$
- *Theorem.* For sequentially compact A, if $f : A \to \mathbb{R}^m$ is continuous, then $f : A \to \mathbb{R}^m$ is uniformly continuous.

2.3. **Pathwise connectedness, intermediate value.**

- For $a, b \in \mathbb{R}^n$, we call
$$[a, b] = \{(1 - t)\, a + tb |\; t \in [0, 1]\}$$
a segment with end points a, b.
- $S \subset \mathbb{R}^n$ is convex, if $[a, b] \subset S$ for $\forall a, b \in S$. (generalization ofinteral)
- *Example.* $B_r(a)$ is convex.
- Continuous map $\gamma : [0, 1] \to \mathbb{R}^n$ is parametrized path, the image $\gamma[0, 1]$ is called path, $\gamma(0)$ and $\gamma(1)$ are initial and end points of the (parametrized) path. If $\gamma[0, 1] \subset S$, then γ is called a parametrized path in S, we write $\gamma : [0, 1] \to S$.
- *Definition.* $S \subset \mathbb{R}^n$ is pathwise-connected if for $\forall a, b \in S$, there is path in S with end points a and b.
- Convex sets are pathwise-connected. In particular, intervals are pathwise-connected in $\mathbb{R}$.
- *Proposition.* Let $I \subset \mathbb{R}$ be pathwise-connected, $a = \inf I$, $b = \sup I$.
  (1) If $a \notin I$, $b \notin I$, then $I = (a, b)$.
  (2) If $a \in I$, $b \notin I$, then $I = [a, b)$.
  (3) If $a \notin I$, $b \in I$, then $I = (a, b]$.
  (4) If $a \in I$, $b \in I$, then $I = [a, b]$.

Proof. (1). Fot $x \in I$, $a \leq x \leq b$. Since $a \notin I$, $b \notin I$, $x \in (a, b)$.
For $x \in (a, b)$. Since
$$\inf I = a < x < b = \sup I,$$
there are $\alpha \in (a, x) \cap I$, $\beta \in (x, b) \cap I$. That is, $\alpha \in I$, $\beta \in I$ and $x \in (\alpha, \beta)$. But I is pathwise-connected, we have $\gamma : [0, 1] \to I$ continuous s.t. $\gamma(0) = \alpha$, $\gamma(1) = \beta$. Hence, $\exists \tau \in (0, 1)$ s.t. $x = \gamma(\tau) \in I$.

- If $D \subset \mathbb{R}^n$ is pathwise-connected, $f : D \to \mathbb{R}$ is continuous, then the graph
$$G_f = \{(x, f(x))|\ x \in D\}$$
is a pathwise-connected subset of $\mathbb{R}^{n+1}$.
- If S_1 and S_2 are pathwise-connected, $S_1 \cap S_2 \neq \emptyset$, then $S_1 \cup S_2$ is also pathwise-connected. Thus, the unit sphere $S^{n-1} = \partial B_1$ is pathwise-connected.
- *Theorem*. If $D \subset \mathbb{R}^n$ is pathwise-connected, $f : D \to \mathbb{R}^m$ continuous, then $f(D)$ is pathwise-connected. In particular, if $m = 1$ then $f(D)$ is an interval.
- Subset $D \subset \mathbb{R}^n$ has intermediate value property if $f(D)$ is an interval for all continuous $f : D \to \mathbb{R}$.
- Thus, pathwise-connected set has intermediate value property, but the converse is not true.

2.4. **Connectedness and the Intermediate Value Property.**

- Let $A \subset \mathbb{R}^n$, two open subset U and V separate A if both of them have nonempty intersection with A and
$$A \cap U \cap V = \emptyset, \qquad A \subset U \cup V.$$
- If there do not exist two open sets seperate A, we call A connected.
- *Theorem*. A connected iff A has intermediate value property.

Proof. ($\Rightarrow$) Else, $\exists f : A \to \mathbb{R}$ continuous, $f(A)$ is not an interval. There is $c \in \mathbb{R}$ such that $c \notin f(A)$ and
$$f^{-1}(-\infty, c) \neq \emptyset, \qquad f^{-1}(c, +\infty) \neq \emptyset.$$
We claim that there are open sets U, V such that
$$f^{-1}(-\infty, c) = A \cap U, \qquad f^{-1}(c, +\infty) = A \cap V. \tag{2.1}$$ e4

Then
$$A \cap U \cap V = f^{-1}(-\infty, c) \cap f^{-1}(c, +\infty) = \emptyset,$$
$$\begin{aligned} A &= f^{-1}(-\infty, c) \cup f^{-1}(c, +\infty) \\ &= (A \cap U) \cup (A \cap V) \subset U \cup V. \end{aligned}$$

We now show that there are open sets U and V so that (2.1) hold. For $a \in f^{-1}(-\infty, c)$, $f(a) < c$. There is $\delta_a > 0$ such that
$$f(x) < c \qquad \text{for } x \in B_{\delta_a}(a) \cap A.$$
That is
$$B_{\delta_a}(a) \cap A \subset f^{-1}(-\infty, c). \tag{2.2}$$ e5

Let
$$U = \bigcup_{a \in f^{-1}(-\infty, c)} B_{\delta_a}(a).$$

Then U is open and using (2.2) we have

$$\begin{aligned} f^{-1}(-\infty, c) &= \bigcup_{a\in f^{-1}(-\infty,c)} \{a\} \\ &\subset \bigcup_{a\in f^{-1}(-\infty,c)} (B_{\delta_a}(a)\cap A) \\ &\subset f^{-1}(-\infty, c). \end{aligned}$$

Therefore

$$f^{-1}(-\infty, c) = \bigcup_{a\in f^{-1}(-\infty,c)} (B_{\delta_a}(a)\cap A) = U\cap A.$$

($\Leftarrow$) If A is not connected, for the open sets U and V seperating A, define $f : A \to \mathbb{R}$,

$$f(x) = \begin{cases} 0 & x\in U\cap A, \\ 1 & x\in V\cap A. \end{cases}$$

Then f is locally constant therefore continuous on A, but do not satisfy IVP.

- *Corollary*. Pathwise-connected sets are connected.

2.5. Distance from a point to a set.

- Let $x \in \mathbb{R}^n$, $A\subset \mathbb{R}^n$, we call

$$d_A(x) = \inf_{y\in A} |x-y|$$

the distance from x to A. Varying x we obtain a function $d_A : \mathbb{R}^n \to [0,\infty)$.

- For $x, y\in\mathbb{R}^n$, $|d_A(x) - d_A(y)| \le |x-y|$.

Proof. Take $\{a_i\}\subset A$ such that $|y-a_i| \to d_A(y)$. Then

$$d_A(x) \le |x-a_i| \le |x-y| + |y-a_i|.$$

Let $i\to\infty$ we get $d_A(x) - d_A(y) \le |x-y|$. Now switch x and y.

- If A is closed, then $d_A(x)$ is attained: $\exists a\in A$ such that

$$d_A(x) = |x-a|.$$

- If $A \ne \emptyset$ is closed and open, then $A = \mathbb{R}^n$. This means that $\mathbb{R}^n$ is connected (having already been stated in the corollary of last section).

Proof. If $A\ne\mathbb{R}^n$, take $x\in\mathbb{R}^n\backslash A$. Because A is closed, there is $a\in A$ such that

$$|x-a| = d_A(x) = \inf_{y\in A}|x-y|. \tag{2.3}$$ e0

Since A is open and $a\in A$, $\exists\varepsilon>0$ such that $B_\varepsilon(a)\subset A$.

Take $\lambda\in(0,1)$ such that $\lambda\,|x-a| < \varepsilon$, and set

$$b = a + \lambda\,(x-a).$$

We see that $|b-a| < \varepsilon$, i.e., $b\in B_\varepsilon(a)$ hence $b\in A$; and

$$|x-b| = |x-(a+\lambda\,(x-a))|$$

$$= |(1-\lambda)(x-a)|$$
$$= (1-\lambda)|x-a| < |x-a| = \inf_{y\in A}|x-y|$$

because $\lambda \in (0,1)$. A contradiction to (2.3).

3. Differentiation of Functions of Several Variables

3.1. Limits.

- To define derivative of $f : \mathbb{R} \to \mathbb{R}$, we need limit
$$f'(x_0) = \lim_{x\to x_0} \frac{f(x)-f(x_0)}{x-x_0},$$
for $x \to x_0$, f need to be defined near x_0.
- Definition. $A \subset \mathbb{R}^n$, a is limit point of A if there is $\{x_k\} \subset A\backslash\{a\}$ s.t. $x_k \to a$.
- All limit points of A consists of A', the derivative set of A.
- Example. $A = B_1\backslash\{0\}$, then $A' = \{x \in \mathbb{R}^n |\ |x| \le 1\}$. Here $0 \in A'$ but $0 \notin A$.
- Let a be a limit point of $A \subset \mathbb{R}^n$, $\ell \in \mathbb{R}^m$, $f : A \to \mathbb{R}^m$. If for all $\{x_k\} \subset A\backslash\{a\}$ with $x_k \to a$ we have $f(x_k) \to \ell$, we say
$$\lim_{x\to a} f(x) = \ell.$$
Thus, if $a \in A' \cap A$, then f is continuous at a iff
$$\lim_{x\to a} f(x) = f(a).$$
If $a \in A \cap A'$, f is continuous at a automatically. If $a \notin A$, f could not be continuous at a.
- Example. $f : (x,y) \mapsto x^2y + e^{xy+1}$. For $(x_k, y_k) \to (1,2)$, by properties of convergent sequence of real numbers, and the continuity of $t \mapsto e^t$ we have
$$f(x_k,y_k) = x_k^2 y_k + e^{x_k y_k+1} \to 1^2 2 + e^{1\cdot 2+1} = 2 + e^3.$$
- Theorem. Suppose $A \subset \mathbb{R}^n$, $a \in A'$, $f, g : A \to \mathbb{R}^m$.
$$\lim_{x\to a} f(x) = \ell, \qquad \lim_{x\to a} g(x) = k,$$
then
$$\lim_{x\to a} (\lambda f(x) + \mu g(x)) = \lambda\ell + \mu k,$$
$$\lim_{x\to a} f(x)\cdot g(x) = \ell \cdot k.$$
If $m = 1$ and $k \ne 0$, then
$$\lim_{x\to a} \frac{f(x)}{g(x)} = \frac{\ell}{k}.$$

- If $(x_0, y_0) \neq (0,0)$,
$$\lim_{(x,y)\to(x_0,y_0)} \frac{xy}{x^2+y^2} = \frac{x_0 y_0}{x_0^2+y_0^2}.$$
If $(x_0, y_0) = (0,0)$, the limit does not exist because for $(x_i, y_i) = \left(\frac{\lambda}{i}, \frac{\mu}{i}\right)$,
$$\frac{xy}{x^2+y^2} \to \frac{\lambda\mu}{\lambda^2+\mu^2} \qquad \text{depends on } \lambda \text{ and } \mu.$$
$$\lim_{(x,y)\to(0,0)} \frac{x^3}{x^2+y^2} = 0, \qquad \lim_{(x,y)\to(0,0)} \frac{\sin\left(x^2+y^2\right)}{x^2+y^2} = 1.$$
- Theorem. Let $a \in A'$, $f : A \to \mathbb{R}^m$, $\ell \in \mathbb{R}^m$. Then
$$\lim_{x\to a} f(x) = \ell$$
if and only if $\forall \varepsilon > 0, \exists \delta > 0, |f(x) - \ell| < \varepsilon$ for $x \in A \cap (B_\delta(a) \setminus \{a\})$.

3.2. **Partial Derivatives.**

- Let $f : B_\varepsilon(a) \to \mathbb{R}$ and i is fixed. We have a single valiable function $\psi : (-\varepsilon, \varepsilon) \to \mathbb{R}$,
$$\begin{aligned} \psi(t) &= f(a + te_i) \qquad\qquad e_i = (0, \dots, 1, \dots, 0) \\ &= f(a^1, \dots, a^{i-1}, a^i + t, a^{i+1}, \dots, a^n). \end{aligned}$$
- If ψ is differentiable at $t = 0$, $\psi'(0)$ is called the partial derivative of f at a with respect to x^i, denoted
$$\psi'(0) = \partial_i f(a) = \left.\frac{\partial f}{\partial x^i}\right|_a,$$
sometimes we also denote this by $f_{x^i}(a)$ or even $f_i(a)$. Depending on the convention of the context.
- Geometric meaning. The intersection of G_f with the plane
$$\Pi = \left\{ \left(x^1, a^2, \dots, a^n, y\right) \middle| \ x^1 \in \mathbb{R}, y \in \mathbb{R} \right\}$$
is the curve
$$\begin{aligned} \gamma(t) &= (a + te_1, \psi(t)) \\ &= \left(a^1 + t, a^2, \dots, a^n, \psi(t)\right). \end{aligned}$$
passing $(a, f(a))$. The tangent of γ at $\gamma(0) = (a, f(a))$ is
$$\dot\gamma(0) = \left(e_1, \dot\psi(0)\right) = (e_1, \partial_1 f(a)).$$
The slope of the tangent with respect to the 1-th coordinate axis or $\partial_1 f(a)$.
- If $\partial_i f(a)$ exists for all i, we define the gradient of f at a,
$$\nabla f(a) = \left(\frac{\partial f}{\partial x^1}, \dots, \frac{\partial f}{\partial x^n}\right)_a.$$

- Example. $\nabla f(a)$ exists but f is not continuous at a.
$$f(x,y)=\begin{cases}1 & xy=0,\\ 0 & xy\neq 0.\end{cases}$$
This is a major different with single variable functions.
- Let $U\subset\mathbb{R}^n$ be open, $f:U\to\mathbb{R}$. If $\partial_i f(x)$ exists for all $x\in U$, we get a new function on U whose value at $x\in U$ is $\partial_i(x)$. This is the i-th partial derivative function of f, denoted by $\partial_i f$, $\partial f/\partial x^i$ or f_{x^i}, or even f_i.
- If for all i, $\partial_i f : U\to\mathbb{R}$ is defined and continuous, we say that f is continuously differentable in U and write $f\in C^1(U)$.
- Let $f=\left(f^1,\dots,f^m\right):B_r(a)\to\mathbb{R}^m$. Given i, if all $\partial_i f^j(a)$ exists, we have a vector
$$\partial_i f(a)=\left.\frac{\partial f}{\partial x^i}\right|_a=\begin{pmatrix}\partial_i f^1(a)\\ \vdots\\ \partial_i f^m(a)\end{pmatrix},$$
called the partial derivative of f at a with respect to x^i. If $\partial_i f(a)$ exists for $i=1,\dots,n$, the $m\times n$ matrix
$$f'(a)=\begin{pmatrix}\partial_1 f^1 & \cdots & \partial_n f^1\\ \vdots & & \vdots\\ \partial_1 f^m & \cdots & \partial_n f^m\end{pmatrix}_a \tag{3.1}$$ ef
is call the Jacobi matrix of f at a.
- If $m=n$ then $f'(a)$ is square. The determinant
$$J_f(a)=\det f'(a)=\left.\frac{\partial\left(f^1,\dots,f^n\right)}{\partial\left(x^1,\dots,x^n\right)}\right|_a$$
is called the Jacobian of f at a.
- As we will see, the linear map $f'(a):\mathbb{R}^n\to\mathbb{R}^m$ determined is a very good linear approximation of the nonlinear map
$$h\mapsto f(a+h)-f(a).$$
We will get *local properties* of f near a via the analysis of $f'(a)$, its linearization at a.
- Roughly speaking, if $f'(a)$ is invertible, the f is *locally* invertible; if $f'(a)$ is surjective (injectiv), the so is f locally.

ex-sur **Example 3.1.** *Let $f:\mathbb{R}^n\to\mathbb{R}^n$ be C^1. If $J_f(x)\neq 0$ for all $x\in\mathbb{R}^n$ and*
$$\lim_{|x|\to\infty}|f(x)|=+\infty,$$
then $f(\mathbb{R}^n)=\mathbb{R}^n$.

This means that for $\forall b\in\mathbb{R}^n$, the nonlinear equation
$$f(x)=b,$$

i.e.,

$$\begin{cases} f^1(x^1,\dots,x^n) = b^1, \\ \vdots \\ f^m(x^1,\dots,x^n) = b^m, \end{cases}$$

is solvable.

- As we know, solving equations is a central topic in mathematics. Simple equations like polynomial equations
$$p(x) = \sum_{i=0}^{n} a_{n-i}x^i = 0$$
could not be solved analyticaly if $n \geq 5$. Thus the existence of solutions is important. The Fundamental Theorem of Algebra (FTA) confirms that $p(x) = 0$ always has a solution in $\mathbb{C}$.
- In 2017, Peng Liu (undergraduate at XMU during 2014–2018, now graduate student at PKU) and I improved Example 3.1 via the following result:
- Let $f : \mathbb{R}^n \to \mathbb{R}^m$ be C^1-map so that there are at most finitely many points x_i such that rank $f'(x_i) < m$. If $f(\mathbb{R}^n)$ is closed, then $f(\mathbb{R}^n) = \mathbb{R}^m$, see [**?**].
- FTA is a direct consequence of this result.

3.3. **Differential of real valued functions.** Let $f : B_r(a) \to \mathbb{R}$, h small so that $a + h \in B_r(a)$. The difference

$$h \mapsto f(a+h) - f(a)$$

is nonlinear on h. We want to approximate this nonlinear function via linear one.

If $\exists \lambda \in \mathbb{R}^n$, such that as $h \to 0$,

$$f(a+h) - f(a) = \lambda \cdot h + o(|h|),$$

that is

$$\lim_{|h|\to 0} \frac{f(a+h) - f(a) - \lambda \cdot h}{|h|} = 0,$$

we say that f is differentiable at a.

Let $\lambda = (\lambda_1, \dots, \lambda_n)$, denote $h = \big(dx^1, \dots, dx^n\big)$, here it is reasonable to denote small change of the i-th variable by dx^i, then

$$df = \lambda \cdot h = \lambda_1 dx^1 + \cdots + \lambda_n dx^n$$

is a good approximation of $f(a+h) - f(a)$, called the differential of f at a. Thus as $|h|$ small,

$$f(a+h) - f(a) \approx df.$$

Theorem 3.2. *If f is differentiable at a, then*

(1) *f is continuous at a.*
(2) *for all i, $\partial_i f(a)$ exists and $\partial_i f(a) = \lambda_i$. Thus $\lambda = \nabla f(a)$.*

Remark 3.3. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be given by

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

Then $\nabla f(0,0) = (0,0)$ but f is not differentiable at $(0,0)$, hence f is not continuous at $(0,0)$.

Theorem 3.4. *Suppose $f : B_r(a) \to \mathbb{R}$ have all first order partial derivatives. If for all i, $\partial_i f : B_r(a) \to \mathbb{R}$ is continuous at a, then f is differentiable at a.*

Proof. To simplify the notations, we consider the case $n = 2$. Let $a = (\alpha, \beta)$ and $h = (k, l)$. For small $|h|$,

$$[\alpha, \beta] \times [\alpha + k, \beta + l] \subset B_r(a)$$

and $f(\cdot, \beta + l)$ and $f(\alpha, \cdot)$ are differentiable on $[\alpha, \alpha + k]$ and $[\beta, \beta + l]$, respectively. Applying the Lagrange mean value theorem to them, we get $\theta, \lambda \in (0,1)$ such that

$$f(\alpha + k, \beta + l) - f(\alpha, \beta + l) = \partial_1 f(\alpha + \theta k, \beta + l)k,$$
$$f(\alpha, \beta + l) - f(\alpha, \beta) = \partial_2 f(\alpha, \beta + \lambda l)l.$$

Therefore

$$\begin{aligned}
f(a+h) - f(a) &= f(\alpha + k, \beta + l) - f(\alpha, \beta) \\
&= [f(\alpha + k, \beta + l) - f(\alpha, \beta + l)] - [f(\alpha, \beta + l) - f(\alpha, \beta)] \\
&= \partial_1 f(\alpha + \theta k, \beta + l)k + \partial_2 f(\alpha, \beta + \lambda l)l \\
&= \partial_1 f(\alpha, \beta)k + \partial_2 f(\alpha, \beta)l \\
&\quad + [\partial_1 f(\alpha + \theta k, \beta + l) - \partial_1 f(\alpha, \beta)]\, k \\
&\quad + [\partial_2 f(\alpha, \beta + \lambda l) - \partial_2 f(\alpha, \beta)]\, l \\
&= \partial_1 f(\alpha, \beta)k + \partial_2 f(\alpha, \beta)l + o(|h|) \\
&= \nabla f(a) \cdot h + o(|h|),
\end{aligned}$$

because by the continuity of $\partial_1 f$ and $\partial_2 f$ at $a = (\alpha, \beta)$, as $|h| = |(k,l)| \to 0$,

$$\left| \frac{[\partial_1 f(\alpha + \theta k, \beta + l) - \partial_1 f(\alpha, \beta)]\, k}{|h|} \right| \\ \leq |\partial_1 f(\alpha + \theta k, \beta + l) - \partial_1 f(\alpha, \beta)| \to 0,$$

$$\left| \frac{[\partial_2 f(\alpha, \beta + \lambda l) - \partial_2 f(\alpha, \beta)]\, l}{|h|} \right| \\ \leq |\partial_2 f(\alpha, \beta + \lambda l) - \partial_2 f(\alpha, \beta)| \to 0.$$

3.4. **Directional Derivatives.** The partial derivative is the rate of change for $f : B_r(a) \to \mathbb{R}$ at a along the direction e_i. We can consider more general directions.

Let $\ell \in \mathbb{R}^n$, then for $\varepsilon \in (0, r/|\ell|)$, we regard $r/|\ell| = \infty$ if $|\ell| = 0$, we have a single variable function $\varphi : (-\varepsilon, \varepsilon) \to \mathbb{R}$

$$\varphi(t) = f(a + t\ell).$$

The directional derivative of f at a in the direction ℓ is denoted and defined by

$$\left.\frac{\partial f}{\partial \ell}\right|_a = \varphi'(0) = \lim_{t\to 0} \frac{f(a+t\ell) - f(a)}{t},$$

also denoted by $\partial_\ell f(a)$.

to **Theorem 3.5.** *If $f : B_r(a) \to \mathbb{R}$ is differentiable at a, then for all $\ell \in \mathbb{R}^n$,*

$$\partial_\ell f(a) = \ell \cdot \nabla f(a).$$

Proof. Since f is differentiable at a, as $h \to 0$,

$$f(a+h) - f(a) = \nabla f(a) \cdot h + o(|h|).$$

Let $h = t\ell$, as $t \to 0$ we have $h \to 0$ and the above equality becomes

$$\frac{f(a+t\ell) - f(a)}{t} = \nabla f(a) \cdot \ell + \frac{o(t)}{t}.$$

Therefore,

$$\left.\frac{\partial f}{\partial \ell}\right|_a = \lim_{t\to 0} \frac{f(a+t\ell) - f(a)}{t} = \nabla f(a) \cdot \ell.$$

Thus, $\nabla f(a)$ is the direction in which the function increase the fastest. That is, among all unit vector ℓ, $\partial_\ell f(a)$ attains its maximum $|\nabla f(a)|$ at $\ell_0 = \nabla f(a)/|\nabla f(a)|$,

$$\partial_\ell f(a) = \ell \cdot \nabla f(a) \le |\ell|\,|\nabla f(a)| \le |\nabla f(a)|.$$

Proposition 3.6. *If $f : B_r(a) \to \mathbb{R}$ is differentiable at a, $k, \ell \in \mathbb{R}^n$, $c \in \mathbb{R}$, then*

$$\partial_{k+\ell} f(a) = \partial_k f(a) + \partial_\ell f(a), \qquad \partial_{c\ell} f(a) = c\partial_\ell f(a).$$

3.5. **Mean value theorem.** The mean value theorem discloses the relation of values of f and its derivative, thus is very useful in the study of the change of f.

Theorem 3.7. *Let Ω be an open subset of $\mathbb{R}^n$, $[a, b] \subset \Omega$, $f : \Omega \to \mathbb{R}$ is differentiable at every point of $[a, b]$. Then there is $\xi \in (a, b)$ such that*

$$f(b) - f(a) = \nabla f(\xi) \cdot (b - a).$$

Or written component wise,

$$f(b) - f(a) = \sum_{i=1}^{n} \left.\frac{\partial f}{\partial x^i}\right|_\xi \left(b^i - a^i\right).$$

Proof. Let $\ell = b - a$, consider $\varphi : [0, 1] \to \mathbb{R}$,

$$\varphi(t) = f(a + t\ell) = f(a + t(b - a)).$$

By Theorem 3.5, φ is and differentiable on $[0, 1]$. Applying Lagrange to φ, $\exists \tau \in (0, 1)$ such that

$$\begin{aligned} f(b) - f(a) &= \varphi(1) - \varphi(0) = \varphi'(\tau) \\ &= \lim_{s\to 0} \frac{\varphi(\tau + s) - \varphi(\tau)}{s} \end{aligned}$$

$$\begin{aligned}
&= \lim_{s\to 0} \frac{f(a+(\tau+s)\,\ell) - f(a+\tau\ell)}{s} \\
&= \lim_{s\to 0} \frac{f((a+\tau\ell)+s\ell) - f(a+\tau\ell)}{s} \\
&= \partial_\ell f(a+\tau\ell) = \nabla f(a+\tau\ell)\cdot \ell \\
&= \nabla f(\xi)\cdot (b-a)\,, \qquad (3.2)
\end{aligned}$$

em

where $\xi = a + \tau\ell$

Example 3.8. *If Ω is a convex open subset of $\mathbb{R}^n$, $f : \mathbb{R}^2 \to \mathbb{R}$ is differentiable and $\nabla f = 0$ in Ω, then f is constant function.*

Remark 3.9. This is also true even if Ω is only a pathwise connected open subset.

Example 3.10. *Let $u : \mathbb{R}^2 \to \mathbb{R}$ be differentiable,*

$$2u_x + u_y = 0, \qquad u(x,0) = 0,$$

then $u \equiv 0$.

Proof. The idea is convert the problem into a problem in single variable functions. Given $(a,b) \in \mathbb{R}^2$, consider $\varphi : \mathbb{R} \to \mathbb{R}$,

$$\varphi(t) = u(a+2t, b+t).$$

Then for $\ell = (2,1)$,

$$\begin{aligned}
\varphi'(t) &= \partial_\ell u(a+2t, b+t) \\
&= \nabla u(a+2t, b+t)\cdot \ell \\
&= 2u_x(a+2t, b+t) + u_y(a+2t, b+t) = 0.
\end{aligned}$$

Thus φ is constnt fuction and

$$\begin{aligned}
u(a,b) &= \varphi(0) = \varphi(-b) \\
&= u(a-2b, 0) = 0.
\end{aligned}$$

4. Higher order derivatives, Taylor formula and extreme

4.1. **Higher order derivatives.** Let Ω be an open subset of $\mathbb{R}^n$, $f : \Omega \to \mathbb{R}$. If for all $x \in \Omega$, $\partial_i f(x)$, the partial derivative of f at x with respect to the i-th component exists, then we have a new function $\partial_i f : \Omega \to \mathbb{R}$.

Given $a \in \Omega$. If the partial derivative of $\partial_i f$ at a with respect to the j-th component exists, we call it a second order derivative of f at a, and denote it by $\partial_{ji} f(a)$ or $f_{ij}(a)$. Namely

$$\partial_{ji} f(a) = \left.\frac{\partial}{\partial x^j}\right|_a \left(\frac{\partial f}{\partial x^i}\right) =: \left.\frac{\partial^2 f}{\partial x^j \partial x^i}\right|_a,$$

or $f_{ij}(a) = (f_i)_j\,(a)$.

If $\partial_{ji} f(x)$ exists for every $x \in \Omega$, we have a new function $\partial_{ji} f : \Omega \to \mathbb{R}$. If the partial derivative of $\partial_{ji} f$ at $a \in \Omega$ with respect to the k-th component exists, it is caled a third order derivative of f at a and denoted by $\partial_{kji} f(a)$ or $f_{ijk}(a)$,

$$\partial_{kji} f(a) = \partial_k(\partial_{ji} f)(a), \qquad f_{ijk}(a) = (f_{ij})_k(a).$$

In this way, we can define partial derivative of any order.

An n-variable function f can have n^k k-th order derivatives $\partial_{i_1\cdots i_k} f$, here $i_1, \ldots, i_k$ run through $1, 2, \ldots, n$.

In classical notation, for example,

$$\partial_{ii} f = \frac{\partial^2 f}{\partial x^i \partial x^i} \qquad \text{and} \qquad \partial_{iij} f = \frac{\partial^3 f}{\partial x^i \partial x^i \partial x^j}$$

are written as

$$\begin{aligned} \partial_{ii} f &= \frac{\partial^2 f}{\partial \left(x^i\right)^2} = \frac{\partial^2 f}{\partial x_i^2}, \\ \partial_{iij} f &= \frac{\partial^3 f}{\partial \left(x^i\right)^2 \partial x^j} = \frac{\partial^3 f}{\partial x_i^2 \partial x_j}, \end{aligned}$$

respectively. Here in the last term, we have switched to use lower index to denote coordinates, to avoid using parenthesis.

There are so many k-th order derivatives. Foutunately, in many cases most of them are equal. For example, in general we have

$$\frac{\partial^2 f}{\partial x^1 \partial x^2} = \frac{\partial^2 f}{\partial x^2 \partial x^1}.$$

We prove this kind of result for 2-variable functions. For convenience, instead of $\left(x^1, x^2\right)$, we use (x, y) to denote the variables.

t1 **Theorem 4.1.** *If $f : B_r^2(a) \to \mathbb{R}$ has second order derivatives f_{xy} and f_{yx}. If $f_{xy}, f_{yx} : B_r(a) \to \mathbb{R}$ is continuous at $a = (x_0, y_0)$, then $f_{xy}(a) = f_{yx}(a)$, i.e.,*

$$f_{xy}(x_0, y_0) = f_{yx}(x_0, y_0).$$

Proof. Consider functions φ and ψ defined in $(x_0 - r, x_0 + r)$ and $(y_0 - r, y_0 + r)$ given by

$$\begin{aligned} \varphi(x) &= f(x, y_0 + k) - f(x, y_0), \\ \psi(y) &= f(x_0 + h, y) - f(x_0, y). \end{aligned}$$

Let

$$\begin{aligned} \Delta = {} & f(x_0 + h, y_0 + k) - f(x_0 + h, y_0) \\ & - f(x_0, y_0 + k) + f(x_0, y_0). \end{aligned}$$

Then for some $\theta_1, \theta_2 \in (0, 1)$,

$$\begin{aligned} \Delta &= \varphi(x_0 + h) - \varphi(x_0) = \varphi'(x_0 + \theta_1 h)h \\ &= \left[f_x(x_0 + \theta_1 h, y_0 + k) - f_x(x_0 + \theta_1 h, y_0)\right] h \\ &= f_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 k)kh. \end{aligned}$$

Similarly, for some $\theta_3, \theta_4 \in (0,1)$ we also have

$$\Delta = f_{yx}(x_0 + \theta_3 h, y_0 + \theta_4 k)hk.$$

Thus

$$f_{xy}(x_0 + \theta_1 h, y_0 + \theta_2 k) = f_{yx}(x_0 + \theta_3 h, y_0 + \theta_4 k).$$

By the continuity of f_{xy} and f_{yx} at $a = (x_0, y_0)$, letting $(h,k) \to (0,0)$, we get the desired result.

For functions of more variables, we have the following.

Corollary 4.2. *If the second derivatives f_{ij} and f_{ji} of $f : B_r^n(a) \to \mathbb{R}$ are continuous at a, then $f_{ij}(a) = f_{ji}(a)$.*

Proof. Consider the two-variable function φ given by

$$\varphi(x,y) = f(a^1, \dots . a^{i-1}, x, a^{i+1}, \dots, a^{j-1}, y, a^{j+1}, \dots, a^n).$$

Then

$$f_{ij}(a) = \left.\frac{\partial^2 \varphi}{\partial y \partial x}\right|_{(a^i, a^j)} = \left.\frac{\partial^2 \varphi}{\partial x \partial y}\right|_{(a^i, a^j)} = f_{ji}(a).$$

We introduce a convenient notation for higher order derivatives. Recall that $\mathbb{N}$ is the set of nonnegative integers. Let Ω be an open subset of $\mathbb{R}^n$, and $f : \Omega \to \mathbb{R}$. For $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}^n$, $h = (h_1, \dots, h_n) \in \mathbb{R}^n$ we denote

$$|\alpha| = \alpha_1 + \cdots + \alpha_n, \qquad \alpha! = \alpha_1! \cdots \alpha_n!.$$

$$\partial^\alpha f = \partial^{(\alpha_1, \dots, \alpha_n)} f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_n^{\alpha_n}}, \qquad h^\alpha = h_1^{\alpha_1} \cdots h_n^{\alpha_n}.$$

For example, if $\alpha = (1,2,0,2)$, then $|\alpha| = 5$,

$$\partial^\alpha f = \frac{\partial^5 f}{\partial x_1 \partial x_2^2 \partial x_4^2}.$$

Let Ω be an open subset of $\mathbb{R}^n$, and $f : \Omega \to \mathbb{R}$, $k \in \mathbb{N}$, if for all $\alpha \in \mathbb{N}^n$ with $|\alpha| \le k$, $\partial^\alpha f : \Omega \to \mathbb{R}$ exists and continuous, we say that f is k-order continuously differentiable, and write $f \in C^k(\Omega)$.

Obviously, if $f \in C^k(\Omega)$, then for all i, $\partial_i f \in C^{k-1}(\Omega)$.

For vector valued functions, or maps $f : \Omega \to \mathbb{R}^m$. If $f^i \in C^k(\Omega)$ for all i, we say that f is C^k (k-order continuously differentiable), and write $f \in C^k(\Omega, \mathbb{R}^m)$.

If in addition there is an open set $D \subset \mathbb{R}^m$ such that $f(\Omega) \subset D$, we can consider f as a map $f : \Omega \to D$ and write $f \in C^k(\Omega, D)$. If $f : \Omega \to D$ is bijective, and the inverse $f^{-1} : D \to \Omega$ is C^k, then we say that f is a C^k-diffeomorphism between Ω and D.

4.2. **Taylor formula.** Let Ω be an open subset of $\mathbb{R}^n$, $f \in C^k(\Omega)$, if you feel difficult, in what follows you may restrict to $k = 2$.

Suppose $[a, a+h] \subset \Omega$, we consider the derivatives of $\varphi : [0,1] \to \mathbb{R}$,

$$\varphi(t) = f(a + th).$$

In the proof of mean value theorem, we know that (see (3.2)) φ is differentiable and

$$\varphi'(t) = \nabla f(a+th) \cdot h = \sum_{i=1}^{n} \partial_i f(a+th) h^i.$$

Now, $\partial_i f \in C^{k-1}(\Omega)$, as for φ, the differentiability of $t \mapsto \partial_i f(a+th)$ enables us to differentiate again and get

$$\begin{aligned}
\varphi''(t) &= \sum_{i=1}^{n} h^i \frac{d}{dt} \left(\partial_i f(a+th)\right) \\
&= \sum_{i=1}^{n} h^i \left(\sum_{j=1}^{n} \partial_j \left(\partial_i f\right)(a+th) \cdot h^j \right) \\
&= \sum_{i=1}^{n} h^i \sum_{j=1}^{n} h^j \partial_{ji} f(a+th) = \partial_{ji} f(a+th) h^i h^j.
\end{aligned}$$

Here, we adopt the Einstein summation convention (repeated indices are implicitly summed over their range).

Apply the same argument again and again, we conclude for $\ell \le k$,

$$\varphi^{(\ell)}(t) = \partial_{j_1 \cdots j_\ell} f(a+th) h^{j_1} \cdots h^{j_\ell} = f_{j_1 \cdots j_\ell}(a+th) h^{j_1} \cdots h^{j_\ell}.$$

Theorem 4.3 (Taylor)**.** *Let Ω be an open subset of $\mathbb{R}^n$, $f \in C^k(\Omega)$. If $[a, a+h] \subset \Omega$, then $\exists \xi \in [a, a+h]$ such that*

$$\begin{aligned}
f(a+h) = f(a) + f_i(a) h^i + \frac{1}{2!} f_{ij}(a) h^i h^j + \cdots \\
+ \frac{1}{(k-1)!} f_{j_1 \cdots j_{k-1}}(a) h^{j_1} \cdots h^{j_{k-1}} + \frac{1}{k!} f_{j_1 \cdots j_k}(\xi) h^{j_1} \cdots h^{j_k}.
\end{aligned}$$

Proof. Apply the Taylor theorem for single variable functions to φ, for some $\tau \in [0,1]$ we get

$$\begin{aligned}
f(a+h) = \varphi(1) &= \sum_{\ell=0}^{k-1} \frac{1}{\ell!} \varphi^{(\ell)}(0) + \frac{1}{k!} \varphi^{(k)}(\tau) \\
&= \sum_{\ell=0}^{k-1} \frac{1}{\ell!} f_{j_1 \cdots j_\ell}(a) h^{j_1} \cdots h^{j_\ell} + \frac{1}{k!} f_{j_1 \cdots j_k}(a+\tau h) h^{j_1} \cdots h^{j_k}.
\end{aligned}$$

lc **Corollary 4.4.** *Let $f \in C^k(B_r(a))$, then as $h \to 0$,*

$$f(a+h) = f(a) + f_i(a) h^i + \frac{1}{2!} f_{ij}(a) h^i h^j + \cdots$$

$$+ \frac{1}{k!} f_{j_1\cdots j_k}(a) h^{j_1} \cdots h^{j_k} + o(|h|^k).$$

Proof. By the continuity of $f_{j_1\cdots j_k}$ at a, as $h \to 0$,

$$\begin{aligned}
&\left| \frac{1}{|h|^k} \left(f(a+h) - \sum_{\ell=0}^{k} \frac{1}{\ell!} f_{j_1\cdots j_\ell}(a) h^{j_1} \cdots h^{j_\ell} \right) \right| \\
&= \left| \frac{1}{|h|^k} \left[f_{j_1\cdots j_k}(a+\tau h) - f_{j_1\cdots j_k}(a) \right] h^{j_1} \cdots h^{j_k} \right| \\
&\le \left| f_{j_1\cdots j_k}(a+\tau h) - f_{j_1\cdots j_k}(a) \right| \to 0.
\end{aligned}$$

Example 4.5. *We consider the case $k = 2$. For $f \in C^2(B_r(a))$, as $h \to 0$,*

$$\begin{aligned}
f(a+h) &= f(a) + f_i(a) h^i + \frac{1}{2} f_{ij}(a) h^i h^j + o(|h|^2) \\
&= f(a) + \nabla f(a) \cdot h + \frac{1}{2} h^{\mathrm{T}} \nabla^2 f(a) h + o(|h|^2). \qquad (4.1)
\end{aligned}$$ et

Here,

$$\nabla^2 f(a) = \begin{pmatrix} f_{11} & \cdots & f_{1n} \\ \vdots & & \vdots \\ f_{n1} & \cdots & f_{nn} \end{pmatrix}_a = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}_a .$$

is the Hessian of f at a, also denoted by $H_f(a)$.

4.3. **Extreme values.** From a real symmetric matrix $A = \left(a_{ij}\right)_{n\times n}$, we construct a function $Q : \mathbb{R}^n \to \mathbb{R}$,

$$Q(x) = x^{\mathrm{T}} A x = (Ax) \cdot x = a_{ij} x^i x^j .$$

Q is the quadratic form associated to A. From Linear Algebra we know that A has n real eigenvalues

$$\lambda_1 \le \cdots \le \lambda_n,$$

the corresponding unit eigenvectors form an orthonormal base $\{\varepsilon_i\}_{i=1}^n$ of $\mathbb{R}^n$, for $x \in \mathbb{R}^n$ we have

$$\lambda_1 |x|^2 \le Q(x) \le \lambda_n |x|^2 .$$

If v is an eigenvector of λ_i, then $Q(v) = \lambda_i |v|^2$.

If $\lambda_1 > 0$, then A is positive definite, if $\lambda_n < 0$, A is negative definite. The case $\lambda_1 \lambda_n = 0$ is semi definite. If $\lambda_1 \lambda_n < 0$, A is indefinite.

Now, let $\Omega \subset \mathbb{R}^n$, $a \in \Omega^\circ$, $f : \Omega \to \mathbb{R}$. If $\exists \delta > 0$ such that

$$f(x) \ge f(a) \qquad \text{for } x \in B_\delta(a),$$

we call a a local minimizer of f and $f(a)$ local mimmum. Replacing $\ge$ by $>$ we define strict local minimizer. Similarly we define maximizer.

lf **Lemma 4.6** (Fermat)**.** *If $a \in \Omega^\circ$ is a local minimizer of $f : \Omega \to \mathbb{R}$ and $\partial_i f(a)$ exists, then $\partial_i f(a) = 0$.*

As a consequence, if $\nabla f(a)$ exists, then $\nabla f(a) = 0$. For a function f, if $\nabla f(a) = 0$ we say that a is a critical point of f.

Proof. We choose $\delta > 0$ such that $B_\delta(a) \subset \Omega$. Then $t = 0$ is local minimizer of the differentiable function $\varphi : (-\delta, \delta) \to \mathbb{R}$,

$$\varphi(t) = f(a + te_i).$$

Hence $\partial_i f(a) = \varphi'(0) = 0$.

Example 4.7. *Let A be a positive definite $n \times n$ matrix, $F \in C^1(\mathbb{R}^n)$ satisfies*

$$|F(x)| \le C\left(1 + |x|^\alpha\right), \qquad x \in \mathbb{R}^n$$

for some $C > 0$ and $\alpha \in [1, 2)$. Let $f = \nabla F$, then the nonlinear equation

$$Ax = f(x)$$

is solvable.

Proof. We define $\Phi : \mathbb{R}^n \to \mathbb{R}$,

$$\Phi(x) = \frac{1}{2} Ax \cdot x - F(x).$$

Then as $|x| \to \infty$, since $\alpha < 2$,

$$\begin{aligned} \Phi(x) &\ge \frac{1}{2}\lambda_1 |x|^2 - |F(x)| \\ &\ge \frac{1}{2}\lambda_1 |x|^2 - C|x|^\alpha - C \to +\infty. \end{aligned}$$

Thus, $\exists \xi \in \mathbb{R}^n$ such that

$$\Phi(\xi) = \min_{\mathbb{R}^n} \Phi.$$

By Lemma 4.6,

$$0 = \nabla\Phi(\xi) = A\xi - \nabla F(\xi) = A\xi - f(\xi).$$

So ξ is a solution of $Ax = f(x)$.

Remark 4.8. Seeking solutions of certain equations via looking for critical points is a powerful method in modern mathematics, called variational method.

According to Lemma 4.6, local extreme points of $f : \Omega \to \mathbb{R}$ must be critical points of f, or points where ∇f does not exist. Given a critical point a of f, how to determine whether it is an extreme point?

Theorem 4.9. *Let $f \in C^2(B_r(a))$, $\nabla f(a) = 0$. Then*

(1) *If $\nabla^2 f(a)$ is positive definite, then a is a local minimizer of f.*
(2) *If $\nabla^2 f(a)$ is indefinite, then a is not a local extreme point of f.*
(3) *If $\nabla^2 f(a)$ is negative definite, then a is a local maximizer of f.*

Proof. Using Taylor, there is a function $\eta : B_r(a) \to \mathbb{R}$ such that

$$\lim_{h\to 0} \frac{\eta(h)}{|h|^2} = 0$$

and

$$f(a+h) = f(a) + \nabla f(a)\cdot h + \frac{1}{2}h^{\mathrm{T}}\nabla^2 f(a)h + \eta(h).$$

(1) If $\nabla^2 f(a)$ is positive definite, then $\lambda_1 > 0$, for $|h| < r$,

$$\begin{aligned} f(a+h) &= f(a) + \nabla f(a)\cdot h + \frac{1}{2}h^{\mathrm{T}}\nabla^2 f(a)h + \eta(h) \\ &\geq f(a) + \frac{\lambda_1}{2}|h|^2 + \eta(h) \\ &= f(a) + |h|^2\left(\frac{\lambda_1}{2} + \frac{\eta(h)}{|h|^2}\right) \\ &= f(a) + |h|^2\,\Psi(h). \end{aligned}$$

Here as $h \to 0$,

$$\Psi(h) = \frac{\lambda_1}{2} + \frac{\eta(h)}{|h|^2} \to \frac{\lambda_1}{2} > 0,$$

hence there is $\varepsilon > 0$ such that for $h \in B_\varepsilon(0)$, $\Psi(h) > 0$ and

$$f(a+h) > f(a).$$

So a is a strict local minimizer.

(2) If $\nabla f^2(a)$ is indefinite, it has eigenvelues $\lambda^- < 0$ and $\lambda^+ > 0$. Let h^- be a unit eigenvector of λ^-, then

$$\lim_{t\to 0}\frac{\eta(th^-)}{t^2} = \lim_{t\to 0}\frac{\eta(th^-)}{|th^-|^2} = 0.$$

$$\begin{aligned} f(a+th^-) &= f(a) + \nabla f(a)\cdot(th^-) + \frac{1}{2}(th^-)^{\mathrm{T}}\ \nabla^2 f(a)\,(th^-) + \eta(th^-) \\ &= f(a) + t^2\left(\frac{\lambda^-}{2} + \frac{\eta(th^-)}{t^2}\right) \\ &= f(a) + t^2\psi(t), \end{aligned}$$

where as $t \to 0$,

$$\psi(t) = \frac{\lambda^-}{2} + \frac{\eta(th^-)}{t^2} \to \frac{\lambda^-}{2} < 0.$$

Hence, there is $\varepsilon^- > 0$ such that for $t \in (0, \varepsilon^-)$,

$$f(a+th^-) < f(a).$$

Similarly we find $\varepsilon^+ > 0$, for $t \in \left(0, \varepsilon^+\right)$ we hace

$$f(a+th^+) > f(a),$$

where h^+ is a unit eigenvector of λ^+.

The above discussion means that, for $\forall \varepsilon > 0$, there always exist

$$a^{\pm} = a + th^{\pm} \in B_{\varepsilon}(a),$$

such that

$$f(a^{+}) > f(a) > f(a^{-}).$$

So a is not a local extreme point.

es **Example 4.10.** *Let $f : \mathbb{R}^n \to \mathbb{R}^n$ be a C^1-map, for all $x \in \mathbb{R}^n$, $J_f(x) \neq 0$. If*

$$\lim_{|x|\to\infty} |f(x)| = +\infty, \tag{4.2}$$

eli

then $f(\mathbb{R}^n) = \mathbb{R}^n$.

Proof. For $b \in \mathbb{R}^n$, consider $\varphi : \mathbb{R}^n \to \mathbb{R}$,

$$\varphi(x) = \frac{1}{2}|f(x) - b|^2 = \frac{1}{2}\sum_{i=1}^{n}\left(f^i(x) - b^i\right)^2.$$

Since f is C^1, we see that $\varphi \in C^1(\mathbb{R}^n)$,

$$\partial_j\varphi(x) = \sum_{i=1}^{n}\left(f^i(x) - b^i\right)\partial_j f^i(x).$$

Thus for

$$\begin{aligned}\nabla\varphi(x) &= \left(f^1(x) - b^1, \dots, f^n(x) - b^n\right)\begin{pmatrix}\partial_1 f^1 & \cdots & \partial_n f^1\\ \vdots & & \vdots\\ \partial_1 f^1 & \cdots & \partial_n f^1\end{pmatrix}_x\\ &= (f(x) - b)\, f'(x).\end{aligned}$$

On the other hand, by (4.2), as $|x| \to \infty$ we have

$$\begin{aligned}\varphi(x) &= \frac{1}{2}|f(x) - b|^2\\ &\geq \frac{1}{2}(|f(x)| - |b|)^2 \to +\infty.\end{aligned}$$

There is $\xi \in \mathbb{R}^n$ such that

$$\varphi(\xi) = \min_{\mathbb{R}^n}\varphi.$$

By Fermat,

$$0 = \nabla\varphi(\xi) = (f(\xi) - b)\, f'(\xi).$$

Since $J_f(\xi) = \det f'(\xi) \neq 0$, we must have $f(\xi) = b$.

4.4. **k-th order approximations.** Let $f, g : B_r(a) \to \mathbb{R}$, we say that g is an k-th order approximation of f at a, if

$$\lim_{h\to 0} \frac{f(a+h) - g(a+h)}{|h|^k} = 0.$$

Of course, we want g to be an simpler function so that we can study the local behavior of f via g. The simplest functions are polynormial.

Example 4.11. *If $f \in C^k(B_r(a))$, let*

$$g(a+h) = \sum_{\ell=0}^{k} \frac{1}{\ell!} f_{j_1\cdots j_\ell}(a) h^{j_1} \cdots h^{j_\ell}.$$

By Lemma 4.4, g is an k-th order approximation of f at a.

5. Differential calculus for vector valued functions

5.1. **Derivative of nonlinear maps.** Reall that an $m \times n$ matrix A can be consider as a linear map $A : \mathbb{R}^n \to \mathbb{R}^m$, $h \mapsto Ah$. The norm of A, is defined by

$$\|A\| = \max_{|h|=1} |Ah|. \tag{5.1}$$ eu

We know that for all $h \in \mathbb{R}^n$ we have $|Ah| \le \|A\| \, |h|$. $\|A\|$ can be viewed as the magnifying factor of A.

We extend the differential of real valued functions to $f : B_r(a) \to \mathbb{R}^m$. The idea is again approximating f by linear functions.

If there is an $m \times n$ matrix A, such that as $h \to 0$,

$$f(a+h) - f(a) = Ah + o(|h|), \tag{5.2}$$ ed

this means

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Ah}{|h|} = 0,$$

or equivalently, there is $\eta : B_r(a) \to \mathbb{R}^m$ satisfying

$$\lim_{h\to 0} \frac{\eta(h)}{|h|} = 0,$$

such that

$$f(a+h) - f(a) - Ah = \eta(h),$$

then we say that f is differentiable at a.

We will show that such A is unique, thus denote it by $f'(a)$, called the derivative of f at a.

Suppose $f = \left(f^1, \dots, f^m\right)$, and the i-th row of A is A^i. Then the i-th component of (5.2) is

$$f^i(a+h) - f^i(a) = A^i h + o(|h|).$$

This proves that the real valued function f^i is differentiable at a and $\nabla f^i(a) = A^i$. Thus

$$A = \begin{pmatrix} A^1 \\ \vdots \\ A^m \end{pmatrix} = \begin{pmatrix} \partial_1 f^1 & \cdots & \partial_n f^1 \\ \vdots & & \vdots \\ \partial_1 f^m & \cdots & \partial_n f^m \end{pmatrix}_a .$$

Thus we have determined the matrix A. Denoting A by $f'(a)$ is consistent with (3.1).

We will see that the matrix $f'(a)$ encodes local properties of f near a.

Proposition 5.1. *If $f : B_r(a) \to \mathbb{R}^m$ is differentiable at a, then f is continuous at a.*

Proposition 5.2. *The map $f : B_r(a) \to \mathbb{R}^m$ is differentiable at a, if and only if all $f^i : B_r(a) \to \mathbb{R}$ are differentiable at a.*

This follows from

$$\begin{aligned} \left| f^i(a+h) - f^i(a) - A^i \cdot h \right| &\le |f(a+h) - f(a) - Ah| \\ &\le \sum_{j=1}^m \left| f^j(a+h) - f^j(a) - A^j \cdot h \right| . \end{aligned}$$

An $m \times n$ matrix $A = (a_{ij})_{m\times n}$ can be considered as a point in $\mathbb{R}^{m\times n}$, so we can define the Euclidean norm of A by

$$|A| = \sqrt{\sum_{i=1}^m \sum_{j=1}^n a_{ij}^2}.$$

For the norm defined in (5.1), in general $\|A\| \ne \|A^{\mathrm{T}}\|$. But we always have $|A| = |A^{\mathrm{T}}|$.

From (5.1) we easily obtain a useful property of matrix norm: if A is $m \times n$ and B is $n \times \ell$, then

$$\|AB\| \le \|A\| \, \|B\| .$$

This is also true for $|\cdot|$ but slightly difficult to prove.

If $A = (a_{ij})$ is an $m \times n$ real matrix, using Cauchy inequality it is easy to prove

$$\frac{1}{\sqrt{n}} |A| \le \|A\| \le |A| .$$

Thus, let $x \mapsto F(x)$ be a matrix valued function taking values in $\mathbb{R}^{m\times n}$, $A \in \mathbb{R}^{m\times n}$. Then as $x \to x_0$

$$|F(x) - A| \to 0 \qquad \Longleftrightarrow \qquad \|F(x) - A\| \to 0.$$

We say that as $x \to a$, the limit of F is A, denote by

$$\lim_{x\to x_0} F(x) = A.$$

If $A = F(x_0)$, we say that F is continuous at x_0.

Let $F(x) = (a^i_j(x))$, then F is continuous at x_0 if and only if all a^i_j are continuous at x_0. Then, both $x \mapsto |F(x)|$ and $x \mapsto \|F(x)\|$ are all continuous at x_0.

p1 **Proposition 5.3.** *If $f, g : B_r(a) \to \mathbb{R}^m$ are differentiable at a, then*

(1) *for $\lambda, \mu \in \mathbb{R}$, $\lambda f + \mu g$ is differentiable at a and*

$$(\lambda f + \mu g)'(a) = \lambda f'(a) + \mu g'(a).$$

(2) *$f \cdot g$ is differentiable at a and*

$$(f \cdot g)'(a) = f^{\mathrm{T}}(a) g'(a) + g^{\mathrm{T}}(a) f'(a). \tag{5.3}$$ ev

Proof. Since f, g are differentiable at a, we have $\eta, \delta : B_r(a) \to \mathbb{R}^m$ such that

$$\lim_{h \to 0} \frac{\eta(h)}{|h|} = 0, \qquad \lim_{h \to 0} \frac{\delta(h)}{|h|} = 0 \tag{5.4}$$ ehd

and

$$f(a+h) = f(a) + f'(a)h + \eta(h), \qquad g(a+h) = g(a) + g'(a)h + \delta(h).$$

We have

$$\begin{aligned} (f \cdot g)(a+h) &= \big(f(a) + f'(a)h + \eta(h)\big) \cdot \big(g(a) + g'(a)h + \delta(h)\big) \\ &= f(a) \cdot g(a) + f(a) \cdot g'(a)h + f'(a)h \cdot g(a) + \Theta(h) \\ &= (f \cdot g)(a) + f^{\mathrm{T}}(a) g'(a) h + g^{\mathrm{T}}(a) f'(a) h + \Theta(h) \\ &= (f \cdot g)(a) + \big(f^{\mathrm{T}}(a) g'(a) + g^{\mathrm{T}}(a) f'(a)\big) h + \Theta(h), \end{aligned} \tag{5.5}$$ ew

where $\Theta : B_r(a) \to \mathbb{R}^m$ is given by

$$\begin{aligned} \Theta(h) = f(a) \cdot \delta(h) + g(a) \cdot \eta(h) + f'(a)h \cdot g'(a)h \\ + f'(a)h \cdot \delta(h) + \eta(h) \cdot g'(a)h + \eta(h) \cdot \delta(h). \end{aligned}$$

Using

$$\begin{aligned} |\Theta(h)| &\le |f(a)|\,|\delta(h)| + |g(a)|\,|\eta(h)| + \left|f'(a)h\right| \left|g'(a)h\right| \\ &\quad + \left|f'(a)h\right| |\delta(h)| + |\eta(h)| \left|g'(a)h\right| + |\eta(h)|\,|\delta(h)| \\ &\le |f(a)|\,|\delta(h)| + |g(a)|\,|\eta(h)| + \left\|f'(a)\right\| |h| \cdot \left\|g'(a)\right\| |h| \\ &\quad + \left\|f'(a)\right\| |h|\,|\delta(h)| + |\eta(h)| \cdot \left\|g'(a)\right\| |h| + |\eta(h)|\,|\delta(h)| \end{aligned}$$

and (5.4), it is easy to see that

$$\lim_{h \to 0} \frac{\Theta(h)}{|h|} = 0.$$

This and (5.5) imply the differentiability of $f \cdot g$ at a and (5.3).

5.2. **Chain Rule.** Sometimes the following equivalent description of differentiability is useful.

ydjms **Lemma 5.4.** *The map* $f : B_r(a) \to \mathbb{R}^m$ *is differentiable at* a *if and only if there are* $m \times n$ *matrix* A *and map* $\theta : B_r(0)\backslash 0 \to \mathbb{R}^m$ *such that*

$$f(a+h) - f(a) = Ah + |h|\,\theta(h), \qquad \lim_{|h|\to 0} \theta(h) = 0. \tag{5.6}$$ ykw

In this case $f'(a) = A$.

Proof. ($\Leftarrow$) If the conditions are satisfied

$$\lim_{h\to 0} \frac{f(a+h) - f(a) - Ah}{|h|} = \lim_{|h|\to 0} \theta(h) = 0,$$

so f is differentiable at a and $f'(a) = A$.

($\Rightarrow$) If f is differentiable at a, take $A = f'(a)$ and define $\theta : B_r(0)\backslash 0 \to \mathbb{R}^m$,

$$\theta(h) = \frac{f(a+h) - f(a) - Ah}{|h|}.$$

By the differentiability of f at a we get 5.6.

The chain rule is a basic tool for computing derivative of composition of maps.

ychn **Theorem 5.5** (Chain rule)**.** *Let* $g : B_r(a) \to \mathbb{R}^m$ *be differentiable at* a, U *is an open subset in* $\mathbb{R}^n$ *containning* $g(B_r(a))$, $f : U \to \mathbb{R}^\ell$ *is differentiable at* $b = g(a)$, *then* $f \circ g : B_r(a) \to \mathbb{R}^\ell$ *is differentiable at* a *and*

$$(f \circ g)'(a) = f'(b)g'(a). \tag{5.7}$$ emx

The equality (5.7) means that the Jacobi matrix of $f \circ g$ at a equals the multiplication of the Jacobi matrix of f at $b = g(a)$ and the Jacobi matrix of g at a, namely, if $g : x \mapsto u$ is differentiable at a, $f : u \mapsto y$is differentiable at $b = g(a)$, then

$$\begin{pmatrix} \frac{\partial y^1}{\partial x^1} & \cdots & \frac{\partial y^1}{\partial x^n} \\ \vdots & & \vdots \\ \frac{\partial y^\ell}{\partial x^1} & \cdots & \frac{\partial y^\ell}{\partial x^n} \end{pmatrix} = \begin{pmatrix} \frac{\partial y^1}{\partial u^1} & \cdots & \frac{\partial y^1}{\partial u^m} \\ \vdots & & \vdots \\ \frac{\partial y^\ell}{\partial u^1} & \cdots & \frac{\partial y^\ell}{\partial u^m} \end{pmatrix} \begin{pmatrix} \frac{\partial u^1}{\partial x^1} & \cdots & \frac{\partial u^1}{\partial x^n} \\ \vdots & & \vdots \\ \frac{\partial u^m}{\partial x^1} & \cdots & \frac{\partial u^m}{\partial x^n} \end{pmatrix},$$

that is , for $i = 1, \ldots, m$ and $k = 1, \ldots, \ell$ we have

$$\left.\frac{\partial y^k}{\partial x^i}\right|_a = \sum_{j=1}^n \left.\frac{\partial y^k}{\partial u^j}\right|_b \cdot \left.\frac{\partial u^j}{\partial x^i}\right|_a.$$

Proof (Thm 5.5). Denote $A = f'(b)$, $B = g'(a)$. Since f and g are differentiable at b and a, there are $\lambda : B^m_\rho \to \mathbb{R}^\ell$ and $\eta : B^n_r \to \mathbb{R}^m$ such that

$$\lim_{k\to 0} \frac{\lambda(k)}{|k|} = 0, \qquad \lim_{h\to 0} \frac{\eta(h)}{|h|} = 0 \tag{5.8}$$ eq

and

$$f(b+k) - f(b) = Ak + \lambda(k), \tag{5.9}$$ er

$$g(a+h)-g(a)=Bh+\eta(h). \tag{5.10}$$ eg

Note that $\lambda(0)=0$. From (5.8), we may assume that for $h \in B_r^n$ we have $|\eta(h)| \leq |h|$.

Substituting

$$k=g(a+h)-g(a)=Bh+\eta(h)$$

in (5.9), we get

$$\begin{aligned}(f\circ g)(a+h)-(f\circ g)(a)&=f(b+k)-f(b)\\&=A\,(Bh+\eta(h))+\lambda(Bh+\eta(h))\\&=(AB)h+\left[A\eta(h)+\lambda(Bh+\eta(h))\right].\end{aligned}$$

Since as $h\to 0$ we have

$$\frac{|A\eta(h)|}{|h|}\leq\frac{\|A\|\,|\eta(h)|}{|h|}\to 0,$$

it suffices to show that

$$\lim_{h\to 0}\frac{\lambda(Bh+\eta(h))}{|h|}=0. \tag{5.11}$$ l

From (5.8) and $\lambda(0)=0$, for $\forall\varepsilon>0$, $\exists\delta>0$ such that for $k\in B_\delta^m$ we have

$$|\lambda(k)|\leq\frac{\varepsilon}{\|B\|+1}\,|k|\,.$$

If $|h|<(\|B\|+1)^{-1}\,\delta$, then

$$\begin{aligned}|Bh+\eta(h)|&\leq|Bh|+|\eta(h)|\\&\leq\|B\|\,|h|+|h|<\delta,\end{aligned}$$

and

$$\begin{aligned}|\lambda(Bh+\eta(h))|&\leq\frac{\varepsilon}{\|B\|+1}\,|Bh+\eta(h)|\\&\leq\frac{\varepsilon}{\|B\|+1}\,(\|B\|\,|h|+|h|)=\varepsilon\,|h|\,.\end{aligned}$$

This proves (5.11).

ywt0 **Corollary 5.6.** *If for $g:B_r(a)\to\mathbb{R}^m$, the partial derivative $\partial_i g(a)$ exists, f is differentiable at $b=g(a)$, then for $f\circ g:B_r(a)\to\mathbb{R}^\ell$ the partial derivative $\partial_i(f\circ g)$ exists and we have*

$$\partial_i(f\circ g)(a)=f'(b)\partial_i g(a). \tag{5.12}$$ ywch

Proof. Since $\partial_i g(a)$ exists, the map $\varphi:(-r,r)\to\mathbb{R}^m$, $\varphi(t)=g(a+te_i)$ is differentiable at $t=0$.[1] The result follows from applying the chain rule to the composition

$$(-r,r)\xrightarrow{\varphi}U\xrightarrow{f}\mathbb{R}^\ell$$

[1] For single variable functions $\varphi:(-r,r)\to\mathbb{R}^m$, differentiability is equivalent to the existence of derivative.

Remark 5.7. If $f : B_r(a) \to \mathbb{R}^m$ is differentiable at $a, h \in \mathbb{R}^n$. Let $g : t \mapsto a + th$, applying the chain rule to the composition

$$(-\varepsilon, \varepsilon) \xrightarrow{g} B_r(a) \xrightarrow{f} \mathbb{R}^n$$

we get

$$f'(a)h = \frac{d}{dt}\bigg|_{t=0} f(a + th).$$

Note that if $m = 1$, $f'(a)h = \nabla f(a) \cdot h$, we recover the result in (3.2).

Let Ω be an open subset of $\mathbb{R}^n$, from Remark 5.7, the mean value theorem for differentiable $f : \Omega \to \mathbb{R}$ and $[a, b] \subset \Omega$ can be written as

$$f(b) - f(a) = f'(\xi)(b - a), \qquad \xi \in [a, b].$$

This could not be generized to a map $f : \Omega \to \mathbb{R}^m$, simply because the mean point ξ for f^i maybe different.

Theorem 5.8. *Let Ω be an open subset of $\mathbb{R}^n$, $f \in C^1(\Omega, \mathbb{R}^m)$, $[a, b] \subset \Omega$. Then $\exists \xi \in [a, b]$ such that*

$$|f(b) - f(a)| \le \left\| f'(\xi) \right\| |b - a|.$$

Proof. We convert vector valued functions into real valued functions via inner product. Consider $\varphi : \Omega \to \mathbb{R}$,

$$\varphi(x) = (f(b) - f(a)) \cdot f(x).$$

Then by Proposition 5.3, $\varphi \in C^1(\Omega)$,

$$\varphi'(x) = (f(b) - f(a))^{\mathrm{T}} f'(x).$$

By mean value theorem, $\exists \xi \in [a.b]$ such that

$$\begin{aligned} |f(b) - f(a)|^2 &= \varphi(b) - \varphi(a) = \varphi'(\xi)\,(b - a) \\ &= (f(b) - f(a))^{\mathrm{T}} f'(\xi)\,(b - a) \\ &= (f(b) - f(a)) \cdot f'(\xi)\,(b - a) \\ &\le |f(b) - f(a)| \left| f'(\xi)\,(b - a) \right| \\ &\le |f(b) - f(a)| \left\| f'(\xi) \right\| |b - a|. \end{aligned}$$

5.3. **Revision.** Points in Euclidean space $\mathbb{R}^n$ is denoted $x = \left(x^1, \dots, x^n\right)$.

We define scalar product and norm for $x, y \in \mathbb{R}^n$

$$x \cdot y = \sum_{i=1}^{n} x^i y^i, \qquad |x| = \sqrt{x \cdot x}.$$

Distance between x, y is $|x - y|$. Ball $B_r(a)$ is r-neighbourhood of a.

For $\{x_k\} \subset \mathbb{R}^n$, $x_k \to a$ iff $|x_k - a| \to 0$. We know

$$x_k \to a \qquad \Longleftrightarrow \qquad x_k^i \to a^i.$$

If $A \subset \mathbb{R}^n$, $a \in A^\circ$ if $\exists r > 0$, $B_r(a) \subset A$.

$a \in A'$ if $\exists \{x_k\} \subset A \backslash a$, $x_k \to a$. This iff $\forall \varepsilon > 0$, $B_\varepsilon(a) \cap (A \backslash a) \neq \emptyset$.

An n-variable (real valued) function on A is a map $f : A \to \mathbb{R}$, that is, any $x = \left(x^1, \ldots, x^n\right) \in A$ is assigned a real number depending on x, thus denoted by $f(x) = f(x^1, \ldots, x^n)$.

For $x \neq \tilde{x}$, it is possible that $f(x) = f(\tilde{x})$. Example constant functions.

Vector valued function is a map $f : A \to \mathbb{R}^m$. Thus, $x \in A$ is assigned to a point y in $\mathbb{R}^m$:

$$x = \left(x^1, \ldots, x^n\right) \longmapsto y = \left(y^1, \ldots, y^m\right).$$

Let $f^i : x \mapsto y^i$, we get (real valued) function $f^i : A \to \mathbb{R}$, namely $y^i = f^i(x)$.

Thus the map $f : A \to \mathbb{R}^m$ is complitely determined by $f^1, \ldots, f^m$,

$$\begin{aligned} x = \begin{pmatrix} x^1 \\ \vdots \\ x^n \end{pmatrix} \longmapsto y &= \begin{pmatrix} y^1 \\ \vdots \\ y^m \end{pmatrix} \\ &= \begin{pmatrix} f^1(x) \\ \vdots \\ f^m(x) \end{pmatrix} = \begin{pmatrix} f^1(x^1, \ldots, x^n) \\ \vdots \\ f^m(x^1, \ldots, x^n) \end{pmatrix}. \end{aligned}$$

We express this by $f = \left(f^1, \ldots, f^m\right)$.

Continuity means that when the change of x is small, the change of $y = f(x)$ is also small. In other words, if x is close to a sufficiently, $f(x)$ should be arbitrarily close to $f(a)$. Thus, f is continuous at $a \in A$ if $f(x_k) \to f(a)$ for all $\{x_k\} \subset A$ verifying $x_k \to a$.

An equivalent description is: $\forall \varepsilon > 0$, $\exists \delta > 0$,

$$|f(x) - f(a)| < \varepsilon \qquad \text{for all } x \in B_\delta(a) \cap A.$$

For $f : A \backslash a \to \mathbb{R}^m$ with $a \in A'$, and $b \in \mathbb{R}^m$ we say

$$\lim_{x \to a} f(x) = b \tag{5.13}$$ elm

if $f(x_k) \to b$ whenever $x_k \in A \backslash a$, $x_k \to a$. Condition $a \notin A'$ is to ensure $\exists x_k \in A \backslash a$ so that $x_k \to a$.

We know that (5.13) is equivalent to $\forall \varepsilon > 0$, $\exists \delta > 0$,

$$|f(x) - A| < \varepsilon \qquad \text{for all } x \in B_\delta(a) \cap (A \backslash a). \tag{5.14}$$ eda

I remind you that to have (5.13), f need not be defined at a.

The order of ε and δ is crucial. Logically, for ANY $\varepsilon > 0$, we look for the δ, δ depends on ε. So you need to fix an ε, which is arbitrary but once choosen, it is fixed. Then we use this fixed ε to find the δ so that (5.14) is valid.

We are studying higher dimensional functions $f : A \to \mathbb{R}$ or $f : A \to \mathbb{R}^m$, where the domain of definition $A \subset \mathbb{R}^n$. For $x \in A$, $x = \left(x^1, \ldots, x^n\right)$. For such object, unless $n = 1$, NO ordering is defined! Thus for real numbers k you could not say $x > k$ or $x < -k$!

Points in $\mathbb{R}^n$ could not be denominator of fractions. Thus, we say that $f : B_r^n(a) \to \mathbb{R}^m$ is differentiable at a, if there is an $m \times n$ matrix A such that for

$\eta : B_r^n \to \mathbb{R}^m$ given by

$$\eta(h) = f(a+h) - f(a) - Ah,$$

we have

$$\lim_{h\to 0} \frac{\eta(h)}{|h|} = 0.$$

You could not write the above limit as

$$\lim_{h\to 0} \frac{\eta(h)}{h} = 0$$

because it makes no sense to put a point (or vector) $h \in \mathbb{R}^n$ in denominator.

The method to master this course, is to understand the concepts you encounter. Think of its geometric meaning whenever possible. Once you get the points of the concepts, every thing will be easy.

For example, the problem 3 in the quiz is: for continuous $f : \mathbb{R}^n \to \mathbb{R}^m$ with

$$\lim_{|x|\to\infty} |f(x)| = +\infty, \tag{5.15}$$ eml

proving $f(\mathbb{R}^n)$ is bounded.

By definition, you need to show that whenever $y_k \in f(\mathbb{R}^n)$, $y_k \to y$, then $y \in f(\mathbb{R}^n)$.

Of course such y_k can be written as $f(x_k)$ for some $x_k \in \mathbb{R}^n$. But instead of assuming $y_k \to y$, many students wrote $x_k \to x^*$. This is wrong.

The correct proof is expressing $y_k = f(x_k)$, then using (5.15) to get the bouned-ness of $\{x_k\}$, so that we can extract a convergent subsequence.

Also, to get the boundness of $\{x_k\}$, many students wrote: if not, then $|x_k| \to \infty$. In fact, for unbounded $\{x_k\}$, we don't need to have $|x_k| \to \infty$. An example is

$$x_k = \left(k\left(1 + (-1)^k\right), 0, \dots 0\right).$$

So you need to enforce your logical trainning.

Note that you could not write $\lim_{|x|\to\infty} f(x) = +\infty$, why?

5.4. **Computation with chain rule.** Chain rule $x \mapsto u \mapsto y$,

$$\frac{\partial y}{\partial x^i} = \sum_{j=1}^n \frac{\partial y}{\partial u^j}\frac{\partial u^j}{\partial x^i}.$$

Example 5.9. *$u = f(xy, \frac{y}{x}, yz)$ is composition of*

$$(x, y, z) \mapsto \left(xy, \frac{y}{x}, yz\right)$$

and f.

$$\begin{aligned}\frac{\partial u}{\partial x} &= f_1\,(xy)_x + f_2\left(\frac{y}{x}\right)_x + f_3\,(yz)_x \\ &= yf_1 - \frac{y}{x^2} f_2, \end{aligned} \tag{5.16}$$ eb

here f_i is partial derivative of f with respect to the i-component. Note that in (5.16) the f_i is evaluated at $(xy, \frac{y}{x}, yz)$. Thus, it you want to compute u_{xy}, the f_i should be considered as composition of

$$(x,y,z) \mapsto (xy, \frac{y}{x}, yz)$$

and f_i.

$$\begin{aligned}
\frac{\partial^2}{\partial y \partial x} &= \frac{\partial}{\partial y}\frac{\partial u}{\partial x} = \frac{\partial}{\partial y}\left(y f_1 - \frac{y}{x^2} f_2\right) \\
&= f_1 + y\,(f_1)_y - \frac{1}{x^2} f_2 - \frac{y}{x^2}\,(f_2)_y \\
&= f_1 + y\left[f_{11}\,(xy)_y + f_{12}\left(\frac{y}{x}\right)_y + f_{13}\,(yz)_y\right] \\
&\quad - \frac{1}{x^2} f_2 - \frac{y}{x^2}\left[f_{21}\,(xy)_y + f_{22}\left(\frac{y}{x}\right)_y + f_{23}\,(yz)_y\right] \\
&= f_1 - \frac{1}{x^2} f_2 + y\left(x f_{11} + \frac{1}{x} f_{12} + z f_{13}\right) - \frac{y}{x^2}\left(x f_{21} + \frac{1}{x} f_{22} + z f_{23}\right).
\end{aligned}$$

Let Ω be open subset of $\mathbb{R}^n$, $u \in C^2(\Omega)$ is harmonic if

$$\Delta u = \sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2} = 0.$$

Example 5.10. *Suppose $u \in C^2(\mathbb{R}^2)$ is harmonic, then $v : \mathbb{R}^2 \to \mathbb{R}$ given by*

$$v\,(x,y) = u(x^2 - y^2, 2xy)$$

is C^2 and harmonic.

Proof. Since v is composition of $u \in C^2(\mathbb{R}^2)$ and

$$(x,y) \mapsto (x^2 - y^2, 2xy),$$

it follows that $v \in C^2(\mathbb{R}^2)$.

$$\begin{aligned}
v_x &= u_1\left(x^2 - y^2\right)_x + u_2\,(2xy)_x \\
&= 2xu_1 + 2yu_2, \\
v_{xx} &= (2xu_1 + 2yu_2)_x \\
&= 2u_1 + 2x\,(u_1)_x + 2y\,(u_2)_x \\
&= 2u_1 + 2x\left(u_{11}\left(x^2 - y^2\right)_x + u_{12}\,(2xy)_x\right) \\
&\quad + 2y\left(u_{21}\left(x^2 - y^2\right)_x + u_{22}\,(2xy)_x\right) \\
&= 2u_1 + 2x\,(2xu_{11} + 2yu_{12}) + 2y\,(2xu_{21} + 2yu_{22}) \\
&= 2u_1 + 4x^2 u_{11} + 8xyu_{12} + 4y^2 u_{22}, \\
v_y &= u_1\left(x^2 - y^2\right)_y + u_2\,(2xy)_y \\
&= -2yu_1 + 2xu_2,
\end{aligned}$$

$$\begin{aligned} v_{yy} &= (-2yu_1 + 2xu_2)_y \\ &= -2u_1 - 2y\,(u_1)_y + 2x\,(u_2)_y \\ &= -2u_1 - 2y\,(-2yu_{11} + 2xu_{12}) + 2x\,(-2yu_{21} + 2xu_{22}) \\ &= -2u_1 + 4y^2u_{11} - 8xyu_{12} + 4x^2u_{22}. \end{aligned}$$

Thus

$$v_{xx} + v_{yy} = u_{11} + u_{22} = 0.$$

5.5. **Geometric applications.** Let $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^n$ be a C^1-curve passing p, that is $p = \gamma(0)$. To find the tangent vector of γ at p, we consider the vector $\overrightarrow{p\gamma(t)}$from p to a nearby point $\gamma(t)$. As $t \to 0$, $\overrightarrow{p\gamma(t)} \to 0$, we get nothing. But

$$\frac{\gamma(t) - p}{t} = \frac{\gamma(t) - \gamma(0)}{t - 0} \to \dot{\gamma}(0).$$

If $\dot{\gamma}(0) \neq 0$, it determines a line passing p, we call it the tangent line of γ at p,

$$\dot{\gamma}(0) = \left(\dot{\gamma}^1(0), \dots, \dot{\gamma}^n(0)\right)$$

is a tangent vector.

Let Ω be an open subset of $\mathbb{R}^n$ and $f : \Omega \to \mathbb{R}$ be a C^1 function, $p \in \Omega$ such that $f(p) = 0$. Under suitable conditions, near the point p, from

$$f(x^1, \dots, x^n) = 0$$

we can express x^i via $x^1, \dots, x^{i-1}, x^{i+1}, \dots, x^n$. Thus the portion S of the set $\{f = 0\}$ near p is graph of a C^1-function

$$\varphi : \left(x^1, \dots, \widehat{x^i}, \dots, x^n\right) \mapsto x^i.$$

Namely, locally S is a smooth hypersurface near p. We want to defermine the normal vector of S at p.

For any curve $\gamma : (-\varepsilon, \varepsilon) \to S$ being $\gamma(0) = p$, we have

$$f(\gamma(t)) = 0.$$

By chain rule,

$$\nabla f(p) \cdot \dot{\gamma}(0) = 0.$$

This means that $\nabla f(p)$ is orthorgonal to $\dot{\gamma}(0)$. This means the tangent lines of all curves γ lying on S and passing p belongs to a hyperplane containning p with $\nabla f(p)$ as normal vector, if $\nabla f(p) \neq 0$.

This plane is called tangent plane of S at p, and the line through p with direction $\nabla f(p)$ is the normal line of S at p.

A surface S in $\mathbb{R}^3$ can be parametrized by

$$x = x(u, v), \qquad y = y(u, v), \qquad z = z(u, v).$$

That is a C^1-map $\varphi : U \to \mathbb{R}^3$,

$$\varphi(u, v) = (x(u, v), y(u, v), z(u, z))$$

such that $S = \varphi(U)$, where U is an open set in $\mathbb{R}^2$. Let $p = \varphi(a, b)$, $(a, b) \in U$. Then $p \in S$, we want to find the normal vector of S at p.

Fixing $v = b$, then $\gamma : u \mapsto \varphi(u, b)$ is a curve on S passing p. The tangent vector of this curve at p is

$$\dot{\gamma}(a) = \partial_u \varphi(a, b).$$

Similary, the tangent vector of $v \mapsto \varphi(a, v)$ is also a curve on S passing p, with tangent vector $\partial_v \varphi(a, b)$.

If

$$\begin{aligned} N &= \partial_u \varphi \times \partial_v \varphi|_{(a,b)} \\ &= \left(\frac{\partial(y, z)}{\partial(u, v}, \frac{\partial(z, x)}{\partial(u, v)}, \frac{\partial(x, y)}{\partial(u, v)} \right)_{(a,b)} \neq 0, \end{aligned} \qquad (5.17) \quad \boxed{\texttt{eno}}$$

then N is a normal vector of S at $p = \varphi(a, b)$.

Let U be an open subset of $\mathbb{R}^{n-1}$, $a \in U$. $\varphi : U \to \mathbb{R}^n$ is C^1 and satisfies

$$\operatorname{rank} \varphi'(a) = n - 1.$$

Then near $p = \varphi(a)$, the image $S = \varphi(U)$ is a hypersurface in $\mathbb{R}^n$ containning p. Except $n = 3$ we do not have cross product, how to find a normal vector N of S at p?

For any $h \in \mathbb{R}^{n-1}$, $t \mapsto a + th$ is a segment in U passing a, $\gamma : t \mapsto \varphi(a + th)$ is a curve on S passing p, whose velocity is

$$\dot{\gamma}(0) = \left.\frac{d}{dt}\right|_{t=0} \varphi(a + th) = \varphi'(a)h.$$

Our normal vector N must be orthogonal to $\dot{\gamma}(0)$, thus

$$\begin{aligned} 0 = \dot{\gamma}(0) \cdot N &= \left(\varphi'(a)h\right)^{\mathrm{T}} N = h^{\mathrm{T}} \left(\varphi'(a)\right)^{\mathrm{T}} N \\ &= h \cdot \left(\left(\varphi'(a)\right)^{\mathrm{T}} N \right). \end{aligned}$$

Since h is arbitrary, we must have

$$\left(\varphi'(a)\right)^{\mathrm{T}} N = 0. \qquad (5.18) \quad \boxed{\texttt{eN}}$$

Since

$$\operatorname{rank} \left(\varphi'(a)\right)^{\mathrm{T}} = \operatorname{rank} \varphi'(a) = n - 1,$$

the solutions of (5.18) form a 1-dimensional subspace of $\mathbb{R}^N$, the normal space (normal line) of S at p. A base vector is

$$N = \left(\frac{\partial(x^2, \ldots, x^n)}{\partial(u^1, \ldots, u^{n-1})}, -\frac{\partial(x^1, x^3, \ldots, x^n)}{\partial(u^1, \ldots, u^{n-1})}, \ldots, (-1)^{n+1} \frac{\partial(x^1, \ldots, x^{n-1})}{\partial(u^1, \ldots, u^{n-1})} \right). \qquad (5.19) \quad \boxed{\texttt{eR}}$$

When $n = 3$, we recover (5.17).

To prove (5.19), since $\operatorname{rank} \varphi'(a) = n - 1$, we asume

$$N^n = (-1)^{n+1} \frac{\partial(x^1, \ldots, x^{n-1})}{\partial(u^1, \ldots, u^{n-1})} \neq 0$$

and write the (5.18) component-wise as follows

$$\partial_1 x^1 N^1 + \cdots + \partial_1 x^{n-1} N^{n-1} = -\partial_1 x^n N^n,$$

$$\vdots$$

$$\partial_{n-1}x^1N^1 + \cdots + \partial_{n-1}x^{n-1}N^{n-1} = -\partial_{n-1}x^nN^n.$$

The coefficient determinant of this system of linear equations for $\left(N^1, \ldots, N^{n-1}\right)$ is $(-1)^{n+1}N^n \neq 0$, by Cramer rule,

$$\begin{aligned} N^i &= \frac{1}{(-1)^{n+1}N^n}\det\begin{pmatrix} x_1^1 & \cdots & x_1^{i-1} & -x_1^nN^n & x_1^{i+1} & \cdots & x_1^{n-1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ x_{n-1}^1 & \cdots & x_{n-1}^{i-1} & -x_{n-1}^nN^n & x_{n-1}^{i+1} & \cdots & x_{n-1}^{n-1} \end{pmatrix} \\ &= (-1)^n\det\begin{pmatrix} x_1^1 & \cdots & x_1^{i-1} & x_1^n & x_1^{i+1} & \cdots & x_1^{n-1} \\ \vdots & & \vdots & \vdots & \vdots & & \vdots \\ x_{n-1}^1 & \cdots & x_{n-1}^{i-1} & x_{n-1}^n & x_{n-1}^{i+1} & \cdots & x_{n-1}^{n-1} \end{pmatrix} \\ &= (-1)^{i+1}\frac{\partial(x^1, \ldots, \widehat{x^i}, \ldots, x^n)}{\partial(u^1, \ldots, u^{n-1})}, \end{aligned}$$

where $x_j^i = \partial_jx^i(a)$ and $\widehat{x^i}$ means that x^i is missing.

6. The Inverse and Implicit Function Theorems

6.1. The inverse function theorem.

Proposition 6.1. *Let A be $n \times n$ matrix, then A is invertible if and only if $\exists c > 0$ such that for all $h \in \mathbb{R}^n$,*

$$|Ah| \geq c\,|h|.$$

Thus, if A is invertible then for $x, y \in \mathbb{R}^n$,

$$|Ax - Ay| \geq c\,|x - y|.$$

Let U be open subset of $\mathbb{R}^n$, $f : U \to \mathbb{R}^n$ is stable, if $\exists c > 0$ such that for $x, y \in U$,

$$|f(x) - f(y)| \geq c\,|x - y|.$$

ekr **Lemma 6.2.** *Let $A : \Omega \to \mathbb{R}^{m\times n}$ be a matrix valued function with entries $\alpha_j^i : \Omega \to \mathbb{R}$ continuous at $a \in \Omega$, then $\psi : \Omega \to \mathbb{R}$, $\psi(x) = \|A(x)\|$ is continuous at a.*

Proof. Firstly, for $m \times n$ matrix $A = \left(\alpha_j^i\right)_{m\times n}$, we have

$$\frac{1}{\sqrt{n}}|A| \leq \|A\| \leq |A|, \tag{6.1}$$ enr

where $|A|$ is the Euclidean norm of A,

$$|A| = \sqrt{\sum_{i=1}^m\sum_{j=1}^n\left(\alpha_j^i\right)^2} = \sqrt{\sum_{i=1}^m|A^i|^2} = \sqrt{\sum_{j=1}^n|A_j|^2}.$$

where A^i be the i-th row and A_j be the j-column of A.

For $h \in \mathbb{R}^n$, $|h| = 1$ we have

$$\begin{aligned} |Ah|^2 &= \sum_{i=1}^{m} \left(A^i \cdot h\right)^2 \le \sum_{i=1}^{m} \left|A^i\right|^2 |h|^2 \\ &\le \sum_{i=1}^{m} \left|A^i\right|^2 = |A|^2 . \end{aligned}$$

Thus

$$\|A\| = \max_{|h|=1} |Ah| \le |A| .$$

On the other hand, since $\left|e_j\right| = 1$,

$$\left|A_j\right|^2 = \left|Ae_j\right|^2 \le \|A\|^2 ,$$

so

$$|A| = \sqrt{\sum_{j=1}^{n} \left|A_j\right|^2} \le \sqrt{n}\, \|A\| ,$$

(6.1) is proved.

Now, for $x_k \in \Omega$, $x_k \to a$, we have $\alpha^i_j(x_k) \to \alpha^i_j(a)$. Thus apply the second inequality of (6.1) to the matrix $A(x_k) - A(a)$, we get

$$\begin{aligned} \left|\psi(x_j) - \psi(a)\right| &= \left|\|A(x_k)\| - \|A(a)\|\right| \\ &\le \|A(x_k) - A(a)\| \\ &\le |A(x_k) - A(a)| \\ &= \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} \left(\alpha^i_j(x_k) - \alpha^i_j(a)\right)^2} \to 0. \end{aligned}$$

Remark 6.3. The proof of Lemma 6.2 only needs the second inequality of (6.1). But the whold (6.1) implies that $x \mapsto \|A(x)\|$ is continuous iff $x \mapsto |A(x)|$ is.

lin **Lemma 6.4.** *Let Ω be an open subset of $\mathbb{R}^n$, $f \in C^1(\Omega, \mathbb{R}^n)$, $a \in \Omega$. If $\det f'(a) \neq 0$, then there are $\varepsilon > 0$ and $\lambda > 0$ such that for all $x, y \in B_\varepsilon(a)$,*

$$|f(x) - f(y)| \ge \lambda |x - y| .$$

Proof. Let $A = f'(a)$, consider the C^1-map $\varphi : \Omega \to \mathbb{R}^n$,

$$\varphi(x) = x - A^{-1} f(x).$$

Then

$$\varphi'(a) = I_n - A^{-1} f'(a) = 0_n,$$

so $\|\varphi'(a)\| = 0$. By Lemma 6.2, $x \mapsto \|\varphi'(x)\|$ is continuous. Thus there is $\varepsilon > 0$ such that

$$\left\|\varphi'(x)\right\| \le \frac{1}{2}, \qquad \forall x \in B_\varepsilon(a).$$

For $x, y \in B_\varepsilon(a)$, $\exists \xi \in [x, y]$ such that

$$\begin{aligned}
\frac{1}{2}|x-y| \geq \left\|\varphi'(\xi)\right\| |x-y| &\geq |\varphi(x)-\varphi(y)| \\
&= \left|\left(x - A^{-1}f(x)\right) - \left(y - A^{-1}f(y)\right)\right| \\
&= \left|(x-y) - A^{-1}(f(x)-f(y))\right| \\
&\geq |x-y| - \left|A^{-1}(f(x)-f(y))\right| \\
&\geq |x-y| - \left\|A^{-1}\right\| |f(x)-f(y)| .
\end{aligned}$$

This implies

$$|f(x)-f(y)| \geq \frac{1}{2\left\|A^{-1}\right\|} |x-y| .$$

emi

Lemma 6.5. *Let G be an open subset of $\mathbb{R}^n$, $f \in C^1(G, \mathbb{R}^n)$. If for all $x \in G$, $\det f'(x) \neq 0$, then $f(G)$ is an open subset of $\mathbb{R}^n$.*

Remark 6.6. Comparing this lemma with Lemma 16.11 in the textbook, the stability of f is not assumed here.

Proof. For $b \in f(G)$, we need to show $b \in [f(G)]^\circ$. Take $a \in G$ such that $f(a) = b$, by Lemma 6.4, $\exists \varepsilon > 0$ such that $f : \overline{B_\varepsilon(a)} \to \mathbb{R}^n$ is injective. Thus for $x \in \partial B_\varepsilon(a)$ we have $f(x) \neq f(a)$, and

$$\mu = \inf_{x \in \partial B_\varepsilon(a)} |f(x) - f(a)| > 0.$$

For $y \in B_{\mu/2}(b)$, consider the $C^1$0function $\psi : \bar{B}_\varepsilon(a) \to \mathbb{R}$,

$$\psi(x) = |f(x) - y|^2 .$$

For $x \in \partial B_\varepsilon(a)$,

$$\begin{aligned}
\psi(x) &= |f(x)-y|^2 \\
&\geq \{|f(x)-f(a)| - |f(a)-y|\}^2 \\
&> \left\{\mu - \frac{\mu}{2}\right\}^2 = \frac{\mu^2}{4} \\
&> |f(a)-y|^2 = \psi(a).
\end{aligned}$$

Hence ψ attains ins minimum at some $\xi \in B_\varepsilon(a)$. Thus

$$0 = \psi'(\xi) = (f(\xi) - y)^{\mathrm{T}} f'(\xi).$$

But $\det f'(\xi) \neq 0$, we must have $y = f(\xi) \in f(G)$. This proves $B_{\mu/2}(b) \subset f(G)$ and $b \in [f(G)]^\circ$.

Proof (Another proof of Example 4.10). We know that $f(\mathbb{R}^n)$ is a closed subset of $\mathbb{R}^n$. By Lemma 6.5, $f(\mathbb{R}^n)$ is an open subset of $\mathbb{R}^n$. Thus $f(\mathbb{R}^n) = \mathbb{R}^n$ because $\mathbb{R}^n$ is the only subset of $\mathbb{R}^n$ that is closed and open.

Let Ω be an open subset of $\mathbb{R}^n$, $f \in C^1(\Omega, \mathbb{R}^n)$, $a \in \Omega$. If $f'(a)$, the linearization of f at a, is invertible (as a linear map on $\mathbb{R}^n$), we expect that f is also invertible. But this can only be true locally, because the differentiability of f at a and the derivative $f'(a)$, depends only on the properties of f near a.

For $a \in \mathbb{R}^n$, by $\mathcal{N}_a = \mathcal{N}_a^n$ we denote the family of open subsets containning a.

Let $\Omega \subset \mathbb{R}^n$, $a \in \Omega^\circ$. A C^1-map $f : \Omega \to \mathbb{R}^m$ is call a local diffeomorphism at a, if there are $U \in \mathcal{N}_a^n$ and $V \in \mathcal{N}_{f(a)}^m$ such that $f : U \to V$ is surjective and $f^{-1} \in C^1(V, U)$. In this case, by the chain rule we have $m = n$.

tiv **Theorem 6.7.** *Let Ω be an open subset of $\mathbb{R}^n$, $f \in C^1(\Omega, \mathbb{R}^n)$, $a \in \Omega$. If* $\det f'(a) \neq 0$, *then f is a local diffeomorphism at a with*

$$(f^{-1})'(y) = \left[f'(x)\right]^{-1},$$

where $x \in U$, $y = f(x)$.

Proof. Because $\det f'(a) \neq 0$, by Lemma 6.4 and the continuity of $x \mapsto \det f'(x)$, there exists $\varepsilon > 0$ such that

(1) f is injective in $U = B_\varepsilon(a)$, for some $\lambda > 0$ we have
$$|f(x_1) - f(x_2)| \geq \lambda \, |x_1 - x_2| , \qquad \forall x_1, x_2 \in U. \tag{6.2}$$ ein

(2) $\forall x \in U$ we have $\det f'(x) \neq 0$. By Lemma 6.5, $V = f(U)$ is an open subset containning $b = f(a)$.

Now, $f : U \to V$ is surjective. Let $\varphi : V \to U$ be its inverse, then (6.2) is precisely
$$|\varphi(y_1) - \varphi(y_2)| \leq \frac{1}{\lambda} \, |y_1 - y_2| .$$
This means that φ is continuous.

For $y \in V$, we should prove that φ is differentiable at y. For $k \in \mathbb{R}^n \backslash 0$ small, Let $x = \varphi(y)$,
$$h = \varphi(y + k) - \varphi(y).$$
Then
$$y + k = f(\varphi(y + k)) = f(x + h),$$
$$|h| = |\varphi(y + k) - \varphi(y)| \leq \frac{1}{\lambda} \, |k| .$$
Since $k \neq 0$ and φ is injective, we have $h \neq 0$. Moreover, as $k \to 0$ we have $h \to 0$. From
$$\frac{\left|\varphi(y + k) - \varphi(y) - [f'(x)]^{-1} k\right|}{|k|} = \frac{\left|h - [f'(x)]^{-1} k\right|}{|k|}$$
$$= \frac{\left|[f'(x)]^{-1} \left(f'(x)h - (f(x + h) - f(x))\right)\right|}{|k|}$$
$$\leq \frac{\left\|[f'(x)]^{-1}\right\| \, |f'(x)h - (f(x + h) - f(x))|}{|h|} \frac{|h|}{|k|}$$

$$\leq \frac{\left\| [f'(x)]^{-1} \right\|}{\lambda} \frac{|f(x+h) - f(x) - f'(x)h|}{|h|}$$

and the differentiability of f at x, we get

$$\lim_{k \to 0} \frac{\left| \varphi(y+k) - \varphi(y) - [f'(x)]^{-1} k \right|}{|k|} = 0.$$

Thus, φ is differentiable at y and $\varphi'(y) = [f'(x)]^{-1}$, that is

$$(f^{-1})'(y) = \left[f'(x)\right]^{-1} = \left[f'(f^{-1}(y))\right]^{-1}.$$

By the formula for inverse matrix and continuity of f' and f^{-1}, it follows that f^{-1} is C^1, see Exercise 10.

The inverse function theorem says that for $f : \Omega \to \mathbb{R}^n$, if the linerization $f'(a) : \mathbb{R}^n \to \mathbb{R}^n$ is invertible, the locally f is invertible near a. In the same spirit, for $f : \Omega \to \mathbb{R}^m$, if $f'(a) : \mathbb{R}^n \to \mathbb{R}^m$ is suejective, we expect f to be locally surjective.

top

Theorem 6.8. *Let Ω be open subset in $\mathbb{R}^n$, $a \in \Omega$. Consider $f : \Omega \to \mathbb{R}^m$, $f(a) = b$. If* rank $f'(a) = m$*, then $b \in [f(\Omega)]^\circ$.*

Remark 6.9. That $b \in [f(\Omega)]^\circ$ means that all points near b are contained in the image of f. For this reason we say that f is locally surjective.

If U is open subset of $\mathbb{R}^m$, $f : U \to \mathbb{R}^n$ is a C^1-map, $a \in U$. If

$$\operatorname{rank} f'(a) < n,$$

we say that a is f a critical point of f. Thus, Theorem 6.8 says that if a is not a critical point of f, then f is locally surjective at a.

In particular, If for $\forall x \in \Omega$ we have rank $f'(x) = m$, then $f(\Omega)$ is open subset of $\mathbb{R}^m$. Thus Lemma 6.5 is a special case of Theorem 6.8.

Proof. Let $f = \left(f^1, \dots, f^m\right)$. We may assume

$$\det \left(f_i^j(a) \right)_{i,j=1,\dots,m} \neq 0.$$

Define $\Phi : \mathbb{R}^n \to \mathbb{R}^n$,

$$\Phi(x) = \left(f(x), x^{m+1} - a^{m+1}, \dots, x^n - a^n\right).$$

Then $\Phi(a) = (b, 0)$,

$$\Phi'(a) = \begin{pmatrix} f_1^1 & \cdots & f_m^1 & f_{m+1}^1 & \cdots & f_n^1 \\ \vdots & & \vdots & \vdots & & \vdots \\ f_1^m & \cdots & f_m^m & f_{m+1}^m & \cdots & f_n^m \\ 0 & \cdots & 0 & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 1 \end{pmatrix}_a$$

is invertible. By Theorem 6.7, there are $U \in \mathcal{N}_a^n$ and $V \in \mathcal{N}_{(b,0)}^n$ such that $\Phi : U \to V$ is diffeomorphism.

Hence, for some $\varepsilon > 0$ we have

$$B_\varepsilon^m(b, 0) \subset V = \Phi(U) \subset \Phi(\Omega).$$

By the definition of Φ we see $B_\varepsilon^n(b) \subset f(\Omega)$. Indeed, if $y \in B_\varepsilon^n(b)$ then $(y, 0) \in B_\varepsilon^m(b, 0)$, so there is $x \in \Omega$ such that

$$(y, 0) = \Phi(x) = \left(f(x), x^{n+1} - p^{n+1}, \ldots, x^m - p^m\right),$$

That is $y = f(x) \in f(\Omega)$.

As generalization of Lemma 6.5, the proof of Theorem 6.8 depends on the inverse function theorem (Theorem 6.7), which in turn depends on Lemma 6.5. However, we can also prove Theorem 6.7 using the contraction mapping principle, then deduce Theorem 6.8. Lemma 6.5 is now indeed a consequence of Theorem 6.8.

6.2. **Dini's Theorem.** Given an equation $F(x, y) = 0$, we want to solve for y from this equation.

Example 6.10. *Given $F(x, y) = x^2 + y^2 - 1 = 0$. If $F(a, b) = 0$ and $b \neq 0$, we can find $\varepsilon > 0$ and a unique continuous function $y : (a - \varepsilon, a + \varepsilon) \to \mathbb{R}$ such that*

(1) *$y(a) = b$,*
(2) *$F(x, y(x)) = 0$ for all $x \in (a - \varepsilon, a + \varepsilon)$.*

However, if $b = 0$, this is no longer true.

Note that in the above example, $b \neq 0$ is quivalent to $\partial_y F(a, b) \neq 0$. It turns out that this is a crucial condition.

Theorem 6.11 (Dini)**.** *Let $F : (\alpha, \beta) \times (\gamma, \delta) \to \mathbb{R}$ be a C^1-function, $(a, b) \in (\alpha, \beta) \times (\gamma, \delta)$, $F(a, b) = 0$. If $\partial_y F(a, b) \neq 0$, then there are $r > 0$ and a C^1-function $\varphi : (a - r, a + r) \to \mathbb{R}$ such that*

(1) *$\varphi(a) = b$,*
(2) *$F(x, \varphi(x)) = 0$ for all $x \in (a - r, a + r)$,*
(3) *if $|x - a| < r$, $|y - b| < r$, $F(x, y) = 0$, then $y = \varphi(x)$.*

Proof. Without lost of generality, assume $\partial_y F(a, b) > 0$. By continuity, there is $\rho > 0$ such that

$$\partial_y F(x, y) > 0 \qquad \text{for } (x, y) \in [a - \rho, a + \rho] \times [b - \rho, b + \rho].$$

Note that $\partial_x F/\partial_y F$ is continuous on the above closed rectangle,

$$M = \sup_{[a-\rho,a+\rho]\times[b-\rho,b+\rho]} \left|\frac{\partial_x F}{\partial_y F}\right| < \infty.$$

Since $F(a, b) = 0$ and $\partial_y F(a, y) > 0$ means that $y \mapsto F(a, y)$ is strictly increasing on $[b - \rho, b + \rho]$, we deduce

$$F(a, b - \rho) < 0 < F(a, b + \rho).$$

By continuity of $x \mapsto F(x, b \pm \rho)$, there is $r > 0$ such that

$$F(x, b-\rho) < 0 < F(x, b+\rho) \qquad \text{for } x \in (a-r, a+r)\,. \tag{6.3}$$

$\boxed{\texttt{ej}}$

Given $x \in (a-r, a+r)$, consider $\psi : y \mapsto F(x, y)$. We have

$$\psi'(y) = \partial_y F(x, y) > 0, \qquad y \in [b-\rho, b+\rho]\,.$$

So ψ is strictly increasing on $[b-\rho, b+\rho]$, while (6.3) means that

$$\psi(b-\rho) < 0 < \psi(b+\rho).$$

So we get a unique $y \in (b-\rho, b+\rho)$ such that

$$F(x, y) = \psi(y) = 0.$$

This y, depending on x, is denoted by $\varphi(x)$, we thus get a function $\varphi : (a-r, a+r) \to \mathbb{R}$ such that

$$F(x, \varphi(x)) = \psi(\varphi(x)) = 0.$$

Note that $\varphi(a) = b$.

We prove that φ differentiable at a. The same argument can be apply to any $x \in (a-r, a+r)$, thus proving our theorem.

For $x \in (a-r, a+r)$, by the mean value theorem, there is a point

$$\xi_x \in [(a, b)\,, (x, \varphi(x))]$$

such that

$$\begin{aligned} 0 &= F(x, \varphi(x)) - F(a, b) \\ &= \partial_x F(\xi_x)\,(x-a) + \partial_y F(\xi_x)\,(\varphi(x) - b)\,. \end{aligned}$$

So

$$\varphi(x) - \varphi(a) = \varphi(x) - b = -\frac{\partial_x F(\xi_x)}{\partial_y F(\xi_x)}\,(x-a)\,. \tag{6.4}$$

$\boxed{\texttt{ez}}$

As $x \to a$,

$$|\varphi(x) - \varphi(a)| \le \left| \frac{\partial_x F(\xi_x)}{\partial_y F(\xi_x)}\,(x-a) \right| \le M\,|x-a|\,.$$

Thus φ is continuous at a, which implies $\xi_x \to (a, b)$ as $x \to a$. Then by (6.4) and the continuity of $\partial_x F$ and $\partial_y F$ we see that φ is differentiable at a with

$$\varphi'(a) = \lim_{x \to a} \frac{\varphi(x) - \varphi(a)}{x-a} = -\lim_{x \to a} \frac{\partial_x F(\xi_x)}{\partial_y F(\xi_x)} = -\frac{\partial_x F(a, b)}{\partial_y F(a, b)}.$$

Example 6.12. *If z is function of (x, y) determined by*

$$F(xz, yz) = 0$$

for a given C^2-function F. Compute z_{xy}.

Proof. Let $G(x,y,z) = F(xz,yz)$, if
$$G_z = F_1\,(xz)_z + F_2\,(yz)_z = xF_1 + yF_2 \neq 0,$$
z would be a function of (x,y) thanks to Dini Theorem.

To compute z_{xy}, we compute
$$\begin{aligned}
0 &= \partial_x\,(F(xz,yz)) \\
&= F_1\,(xz)_x + F_2\,(yz)_x \\
&= F_1\,(z + xz_x) + F_2\,(yz_x)\,, \qquad (6.5) \quad \boxed{\texttt{ec}}\\
0 &= \partial_y\,(F_1\,(z+xz_x) + F_2\,(yz_x)) \\
&= F_1\,(z+xz_x)_y + (z+xz_x)\,\partial_y F_1 \\
&+ F_2\,(yz_x)_y + (yz_x)\,\partial_y F_2 \\
&= F_1\left(z_y + xz_{xy}\right) + (z+xz_x)\left[F_{11}\,(xz)_y + F_{12}\,(yz)_y\right] \\
&+ F_2\left(z_x + yz_{xy}\right) + yz_x\left[F_{21}\,(xz)_y + F_{22}\,(yz)_y\right] \\
&= F_1\left(z_y + xz_{xy}\right) + (z+xz_x)\left[F_{11}\left(xz_y\right) + F_{12}\left(z + yz_y\right)\right] \\
&+ F_2\left(z_x + yz_{xy}\right) + yz_x\left[F_{21}\left(xz_y\right) + F_{22}\left(z + yz_y\right)\right].
\end{aligned}$$
Solving for z_{xy}, we get an expression for z_{xy} involving z_x and z_y, which could be solved from (6.5) and its analogue. We eventually get the expression for z_{xy} involving (x,y,z), F_i and F_{ij}.

6.3. **The Implicit Function Theorem.** Let U and V be open subset of $\mathbb{R}^m$ and $\mathbb{R}^n$, $F : U \times V \to \mathbb{R}^p$, $(a,b) \in U \times V$. Then we have a map $F_2 : V \to \mathbb{R}^p$, $y \mapsto F(a,y)$. We define
$$\partial_y F(a,b) = F_2'(b).$$
Similarly we define $\partial_x F(a,b)$. Then $\partial_x F$ and $\partial_y F$ are linear maps from $\mathbb{R}^m$ and $\mathbb{R}^n$ to $\mathbb{R}^p$ respectively, with the matrices
$$\partial_x F(a,b) = \begin{pmatrix} \partial_{x^1} F^1 & \cdots & \partial_{x^m} F^1 \\ \vdots & & \vdots \\ \partial_{x^1} F^p & \cdots & \partial_{x^m} F^p \end{pmatrix},$$
$$\partial_y F(a,b) = \begin{pmatrix} \partial_{y^1} F^1 & \cdots & \partial_{y^n} F^1 \\ \vdots & & \vdots \\ \partial_{y^1} F^p & \cdots & \partial_{y^n} F^p \end{pmatrix}.$$

Proposition 6.13. *Suppose $F : U \times V \to \mathbb{R}^p$, $(a,b) \in U \times V$.*

(1) *If F is differentiable at (a,b), then $F_1 : x \mapsto F(x,b)$ is differentiable at a, $F_2 : y \mapsto F(a,y)$ is differentiable at b, and we have*
$$F'(a,b)(h,k) = \partial_x F(a,b)h + \partial_y F(a,b)k, \qquad (h,k) \in \mathbb{R}^m \times \mathbb{R}^n. \qquad (6.6) \quad \boxed{\texttt{par}}$$
(2) *If $\partial_x F$ and $\partial_y F$ are continuous at (a,b), then F is differentiable at (a,b) and we have 6.6.*

By considering the components of F, the proof is easy..

Theorem 6.14 (Implicit function theorem). *Let U and V be open sets in $\mathbb{R}^m$ and $\mathbb{R}^n$, $F \in C^1(U \times V, \mathbb{R}^n)$, $(a,b) \in U \times V$. If*

$$F(a,b) = 0, \qquad \det\left[\partial_y F(a,b)\right] \neq 0,$$

then there are $r > 0$ and a C^1-map $\varphi : B_r^m(a) \to V$ such that $B_r^m(a) \subset U$ and

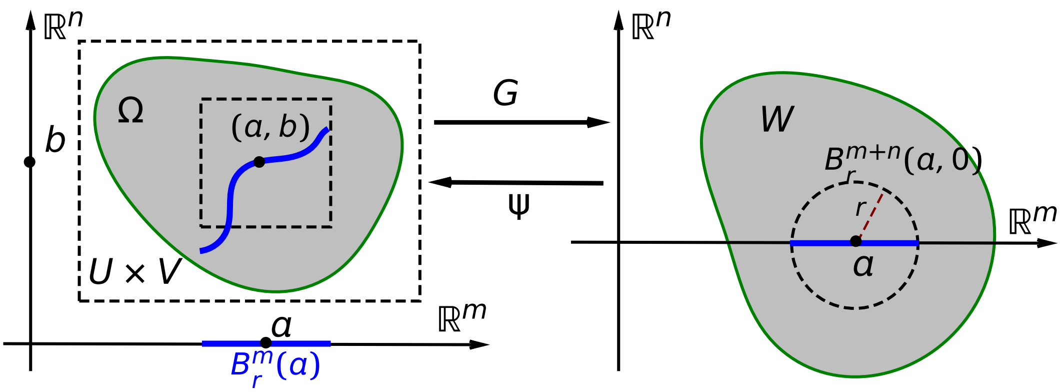
1. *$\varphi(a) = b$,*
2. *for $\forall x \in B_r^m(a)$ we have $F(x, \varphi(x)) = 0$.*
3. *if $(x,y) \in B_r^m(a) \times B_r^n(b)$ such that $F(x,y) = 0$, then $y = \varphi(x)$.*

Proof. Define $G : U \times V \to \mathbb{R}^m \times \mathbb{R}^n$, $G(x,y) = (x, F(x,y))$. Then $G \in C^1$,

$$G'(a,b) = \begin{pmatrix} I_m & 0 \\ \partial_x F(a,b) & \partial_y F(a,b) \end{pmatrix}.$$

Obviously $\det G'(a,b) \neq 0$, $G(a,b) = (a,0)$. By inverse function theorem, there are $\Omega \in \mathcal{N}^{m+n}_{(a,b)}$ and $W \in \mathcal{N}^{m+n}_{(a,0)}$ such that $G : \Omega \to W$ is diffeomorphism. Let $\Psi : W \to \Omega$ be the local inverse of G. From the definition of G, for $(x,z) \in W$ we have $\Psi^1(x,z) = x$, thus

$$\Psi(x,z) = (x, \Psi^2(x,z)).$$



Take $r > 0$ such that $B_r^{m+n}(a,0) \subset W$, then for $x \in B_r^m(a)$ we have $(x,0) \in W$. We may also assume that

$$B_r^m(a) \times B_r^n(b) \subset \Omega. \tag{6.7}$$ einj

Define $\varphi : B_r^m(a) \to \mathbb{R}^n$ by

$$\varphi(x) = \Psi^2(x,0),$$

then φ is a C^1-map. For $x \in B_r^m(a)$ we have

$$\begin{aligned}(x, F(x,\varphi(x))) &= G(x,\varphi(x)) = G(x, \Psi^2(x,0)) \\ &= G(\Psi(x,0)) = (x,0),\end{aligned}$$

Thai is $F(x,\varphi(x)) = 0$.

If $(x,y) \in B_r^m(a) \times B_r^n(b)$ is such that $F(x,y) = 0$, then

$$G(x,y) = G(x,0) = G(x,\varphi(x)).$$

By (6.7), G is injective in $B_r^m(a) \times B_r^n(b)$, we deduce that $y = \varphi(x)$. In particular, $b = \varphi(a)$.

How to compute the derivative of $y = \varphi(x)$? Let $\Phi : x \mapsto F(x, \varphi(x))$, it is the composition of $g : x \mapsto (x, \varphi(x))$ and F. Since for $\forall x \in O$ we have $\Phi(x) = 0$, we deduce

$$\begin{aligned} 0 = \Phi'(x) &= F'(x, \varphi(x)) g'(x) \\ &= \big(\partial_x F(x, \varphi(x)), \partial_y F(x, \varphi(x))\big) \begin{pmatrix} I_m \\ \varphi'(x) \end{pmatrix} \\ &= \partial_x F(x, \varphi(x)) + \partial_y F(x, \varphi(x)) \varphi'(x), \end{aligned}$$

Note that

$$\partial_y F(a, b) = \partial_y F(a, \varphi(a))$$

is invertible, by continuity, for smaller O we may assume that $\partial_y F(x, \varphi(x))$ is invertible for $x \in O$. For such x, multiplying $\left[\partial_y F(x, \varphi(x))\right]^{-1}$ to both sides of the above equality we get

$$\begin{aligned} \varphi'(x) &= -\left[\partial_y F(x, \varphi(x))\right]^{-1} \partial_x F(x, \varphi(x)) \\ &= -\left[\partial_y F(x, y)\right]^{-1} \partial_x F(x, y). \end{aligned}$$

In practical computation, we take derivative with respect to x^k on both sides of

$$F^i(x^1, \dots, x^m, y^1, \dots, y^n) = 0, \qquad i = 1, \dots, n$$

to get

$$\frac{\partial F^i}{\partial x^k} + \sum_{j=1}^n \frac{\partial F^i}{\partial y^j} \frac{\partial y^j}{\partial x^k} = 0, \qquad i = 1, \dots, n,$$

then solve for $\partial y^j / \partial x^k$ using Cramer rule.

Now we look back to surfaces in $\mathbb{R}^n$. For surface, we mean subset of R^n which is locally a graph $G_f = \{(z, \varphi(z))\}$ of smooth function $\varphi : \mathbb{R}^{n-1} \to \mathbb{R}$.

Example 6.15. *Let $F : \mathbb{R}^n \to \mathbb{R}$ be a C^1-function, $M = F^{-1}(0)$ is not empty. For $x \in M$ we have $\nabla F(x) \neq 0$. Consider $a \in M$, we may assume $\partial_n F(a) \neq 0$, then by implicit function theorem, from*

$$F(x^1, \dots, x^n) = 0$$

we may locally express x^n via $\left(x^1, \dots, x^{n-1}\right)$,

$$x^n = \varphi(x^1, \dots, x^{n-1}),$$

where φ is a C^1-function. Near the point a, $x \in M$ iff x lies on the graph of φ. Thus M is a surface. We know that $\nabla F(a)$ is a normal vector of M at a.

Example 6.16. *Let U be open subset of $\mathbb{R}^{n-1}$, $x : U \to \mathbb{R}^n$ is a C^1-map. If for all $u \in U$,*

$$\operatorname{rank} x'(u) = n - 1,$$

then $S = x(U)$ is a surface in $\mathbb{R}^n$.

For $a = x(\alpha) \in S$, where $\alpha \in U$, since

$$\operatorname{rank} x'(\alpha) = n - 1,$$

we may assume that

$$\left.\frac{\partial(x^1,\dots,x^{n-1})}{\partial(u^1,\dots,u^{n-1})}\right|_\alpha \neq 0.$$

By inverse function theorem, near $\left(a^1,\dots,a^{n-1}\right)$ *and* α*, the map*

$$\left(u^1,\dots,u^{n-1}\right) \mapsto \left(x^1,\dots,,x^{n-1}\right)$$

is invertible, that is, we can express u^i *by* $z = \left(x^1,\dots,,x^{n-1}\right)$*,*

$$u^i = u^i(z) = u^i(x^1,\dots,,x^{n-1}).$$

Consequently, near a*,* S *is graph of the* C^1*-function*

$$\begin{aligned} x^n &= x^n(u^1,\dots,u^{n-1}) \\ &= x^n(u^1(z),\dots,u^{n-1}(z)) \\ &= \varphi(z) = \varphi(x^1,\dots,,x^{n-1}). \end{aligned}$$

So S *is a smooth surface. We also know that the normal vector of* S *at* $a = x(\alpha)$ *is*

$$N = \left(\frac{\partial(x^2,\dots,x^n)}{\partial(u^1,\dots,u^{n-1})},\dots,(-1)^{n+1}\frac{\partial(x^1,\dots,x^{n-1})}{\partial(u^1,\dots,u^{n-1})}\right)_\alpha.$$

Example 6.17. *More generaly, let* U *be open subset of* $\mathbb{R}^k$*,* $x : U \to \mathbb{R}^n$ *is a* C^1*-map. If for all* $u \in U$*,*

$$\operatorname{rank} x'(u) = k,$$

then $S = x(U)$ *is a* k*-dimensional surface in* $\mathbb{R}^n$*. If* $k = 1$*, this is a* 1*-dmensional surface, that is a curve.*

Assume that $\alpha \in U$ *and*

$$\left.\frac{\partial(x^1,\dots,x^k)}{\partial(u^1,\dots,u^k)}\right|_\alpha \neq 0,$$

then near $a = x(\alpha)$*,* S *is graph of the vector valued function* $\varphi : \mathbb{R}^k \to \mathbb{R}^{n-k}$*,*

$$\varphi : \left(x^1,\dots,x^k\right) \mapsto \left(x^{k+1},\dots,x^n\right).$$

That is

$$x^i = x^i(x^1,\dots,x^k), \qquad i = k+1,\dots,n.$$

In geometric application we have proved that $\nabla F(p_0)$ is orthogonal to tangent vector of curves γ in $S = F^{-1}(0)$ passing p_0. Now we prove the inverse: if h is orthogonal to $\nabla F(p_0)$, then there is a curve γ in S passing p_0 such that $\dot\gamma(0) = h$.

wcur

Proposition 6.18. *Let* $g : \mathbb{R}^m \to \mathbb{R}^n$ *be* C^1*-map,* $p \in M = g^{-1}(0)$*,* $\operatorname{rank} g'(p) = n$*. If* $g'(p)h = 0$*, there is a* C^1*-curve* $\gamma : (-\varepsilon,\varepsilon) \to M$ *such that* $\gamma(0) = p$*,* $\dot\gamma(0) = h$*.*

Proof (Proof via implicite function theorem). Consider

$$g'(p) = \begin{pmatrix} g_1^1 & \cdots & g_{m-n}^1 & g_{m-n+1}^1 & \cdots & g_m^1 \\ \vdots & & \vdots & \vdots & & \vdots \\ g_1^n & \cdots & g_{m-n}^n & g_{m-n+1}^n & \cdots & g_m^n \end{pmatrix},$$

we may assume

$$\det\left(g_i^j(p)\right)_{i=m-n+1,\ldots,m;\, j=1,\ldots,n} \neq 0. \tag{6.8}$$ wdet

Denote $x = (z, y) \in \mathbb{R}^{m-n} \times \mathbb{R}^n$, $p = (a, b)$, $h = (k, l)$.

Note that 6.8 means $\partial_y g(p)$ is invertible. We have

$$0 = g'(p)h = \left(\partial_z g, \partial_y g\right)\begin{pmatrix} k \\ l \end{pmatrix} = \left(\partial_z g\right)k + \left(\partial_y g\right)l. \tag{6.9}$$ wkl

By implicity function theorem, there is $\varepsilon > 0$ and C^1-map $\varphi : B_\varepsilon^{m-n}(a) \to \mathbb{R}^n$, such that $\varphi(a) = b$,

$$\varphi'(a) = -\left[\partial_y g\right]^{-1}\left(\partial_z g\right); \qquad g(z, \varphi(z)) = 0,\ \forall z \in B_\varepsilon^{m-n}(a).$$

Define $\gamma : (-\varepsilon, \varepsilon) \to \mathbb{R}^m$,

$$\gamma(t) = (a + tk, \varphi(a + tk)),$$

then $\gamma(0) = p$. For $t \in (-\varepsilon, \varepsilon)$ we have

$$g(\gamma(t)) = g(a + tk, \varphi(a + tk)) = 0,$$

Thus γ is a curve on M passing p. Moreover, using 6.9 we obtain

$$\dot{\gamma}(0) = \left(k, \varphi'(a)k\right) = \left(k, -\left[\partial_y g\right]^{-1}\left(\partial_z g\right)k\right) = (k, l) = h.$$

Proof (Proof via ODE). Since $\left\{\nabla g^i(p)\right\}_{i=1}^n$ are linearly independent, $\exists \varepsilon > 0$, such that if $q \in B_\varepsilon(p)$, $\left\{\nabla g^i(q)\right\}_{i=1}^n$ is also independent. Apply the Gram-Schmidt orthogonalization to$\left\{\nabla g^i(q)\right\}_{i=1}^n$, we obtain vector fields $v^i : B_\varepsilon(p) \to \mathbb{R}^m$, such that

$$v^i(q) \cdot v^j(q) = \delta_{ij}, \qquad q \in B_\varepsilon(p); \qquad h \cdot v^i(p) = 0.$$

Here we have used $g'(p)h = 0$, that is $h \cdot \nabla g^i(p) = 0$.

Define continuous vector field $Y : B_\varepsilon(p) \to \mathbb{R}^m$,

$$Y(q) = h - \sum_{k=1}^n \left(h \cdot v^k(q)\right)v^k(q).$$

Geometricly, for $q \in M$, $Y(q)$ is the projection of the constant vector h on the tangent space $T_q M$ of M at q. Since $\nabla g^i(q)$ is linear combination of $\left\{v^j(q)\right\}_{j=1}^n$,

$$\nabla g^i(q) = \sum_{j=1}^n \lambda_j^i(q) v^j(q),$$

we deduce

$$\begin{aligned}
\nabla g^i(q)\cdot Y(q) &= \sum_{j=1}^{n}\lambda^i_j(q)v^j(q)\cdot Y(q)\\
&= \sum_{j=1}^{n}\lambda^i_j(q)v^j(q)\cdot\left(h-\sum_{k=1}^{n}\left(h\cdot v^k(q)\right)v^k(q)\right)\\
&= \sum_{j=1}^{n}\lambda^i_j(q)\left(v^j(q)\cdot h - v^j(q)\cdot\sum_{i=1}^{n}\left(h\cdot v^k(q)\right)v^k(q)\right)\\
&= \sum_{j=1}^{n}\lambda^i_j(q)\left(v^j(q)\cdot h - \sum_{i=1}^{n}\left(h\cdot v^k(q)\right)\delta^{jk}\right) = 0. \qquad (6.10)
\end{aligned}$$

we007

The integral curve $\gamma : (-\varepsilon,\varepsilon)\to\mathbb{R}^m$ passing p of the continuous vector field Y verifies

$$\dot\gamma(t) = Y(\gamma(t)), \qquad \gamma(0) = p.$$

Since $h\cdot v^k(p) = 0$, we have

$$\dot\gamma(0) = Y(p) = h.$$

From 6.10 we get

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}g^i(\gamma(t)) &= \nabla g^i(\gamma(t))\cdot\dot\gamma(t)\\
&= \nabla g^i(\gamma(t))\cdot Y(\gamma(t)) = 0,
\end{aligned}$$

Thus

$$g^i(\gamma(t)) = g^i(\gamma(0)) = g^i(p) = 0.$$

That is $g(\gamma(t))\equiv 0$, $\gamma(t)\in M$.

6.4. **Constraint Extreme Problems and Lagrange Multipliers.** Let $D\subset\mathbb{R}^4$ be open set, $f, g_1, g_2\in C^1(D)$. We want to fine extrume value of $z = f(x,y,u,v)$ under the constrain

$$\begin{cases} g_1(x,y,u,v) = 0,\\ g_2(x,y,u,v) = 0\end{cases} \qquad (6.11)$$

wxe1

Geometrically, 6.11 represents a surface S in D, we need to find extreme of f restrict on S. In in D we have

$$\operatorname{rank}\begin{pmatrix} g_{1x} & g_{1y} & g_{1u} & g_{1v}\\ g_{2x} & g_{2y} & g_{2u} & g_{2v}\end{pmatrix} = 2, \qquad (6.12)$$

we2

for example if $\dfrac{\partial(g_1,g_2)}{\partial(u,v)}\neq 0$, then from 6.11 we can apply the inverse function theorem and get

$$\begin{cases} u = \varphi_1(x,y),\\ v = \varphi_2(x,y).\end{cases}$$

The problem is now to find free extreme of

$$z = f(x, y, \varphi_1(x, y), \varphi_2(x, y))$$

In practical situation, it is hard to solve for u, v from 6.11. To go arround this difficulty we have the Lagrange multiplier method..

Consider the most simplest case. If

$$f : (x, y) \mapsto z = f(x, y)$$

attains extreme value c under the constrain $g(x, y) = 0$ at at (x_0, y_0), then the curve $f(x, y) = c$ must be tangent to $g(x, y) = 0$ at (x_0, y_0). If $\nabla g(x_0, y_0) \neq 0$, we have $\lambda \in \mathbb{R}$ such that

$$\nabla f(x_0, y_0) + \lambda \nabla g(x_0, y_0) = 0.$$

xlag **Theorem 6.19** (Lagrange)**.** *Let Ω be open subset in $\mathbb{R}^m$, $f \in C^1(\Omega)$, $g : \Omega \xrightarrow{C^1} \mathbb{R}^n$, $M = g^{-1}(0)$. If $x_0 \in M$ is a minimal point of $f : M \to \mathbb{R}$,* rank $g'(x_0) = n$, *then $\exists \lambda \in \mathbb{R}^n$ such that*

$$\nabla f(x_0) + \lambda^{\mathrm{T}} g'(x_0) = 0. \tag{6.13}$$ wee6

Proof (A). We denote $x = (z, y) \in \mathbb{R}^{m-n} \times \mathbb{R}^n$, $x_0 = (a, b)$. Assume $\det \partial_y g(x_0) \neq 0$, by the implicit function theorem, there is a function y defined near a such that

$$g(z, y(z)) = 0, \qquad y(a) = b.$$

Then a is minimizer of $\psi = f(\cdot, y(\cdot))$, hence

$$\partial_z f(a, b) + \partial_y f(a, b) y'(a) = \nabla \psi(a) = 0. \tag{6.14}$$ wee3

From $g(z, y(z)) = 0$ we have

$$\partial_z g(a, b) + \partial_y g(a, b) y'(a) = 0. \tag{6.15}$$ wee4

Since $\det \partial_y g(x_0) \neq 0$, by Cramer rule, there is $\lambda \in \mathbb{R}^n$ such that

$$\partial_y f(a, b) + \lambda^{\mathrm{T}} \partial_y g(a, b) = 0. \tag{6.16}$$ wee5

Multiplying λ^{T} from left to both sides of 6.15, adding the result to 6.14 and using 6.16 we get

$$\partial_z f(a, b) + \lambda^{\mathrm{T}} \partial_z g(a, b) = 0.$$

Combining with 6.16 we get 6.13,

$$\begin{aligned} \nabla f(x_0) + \lambda^{\mathrm{T}} g'(x_0) &= \left(\partial_z f(x_0), \partial_y f(x_0)\right) + \lambda^{\mathrm{T}} \left(\partial_z g(x_0), \partial_y g(x_0)\right) \\ &= \left(\partial_z f(x_0) + \partial_z g(x_0), \partial_y f(x_0) + \lambda^{\mathrm{T}} \partial_y g(x_0)\right) = (0, 0). \end{aligned}$$

Proof (B). Let $x_0 \in M$ be mimimizer of $f : M \to \mathbb{R}$, for some $\delta > 0$ we have

$$x \in B_\delta(x_0) \cap M \quad \Longrightarrow \quad f(x) \geq f(x_0). \tag{6.17}$$ we

Define $F : B_\delta(x_0) \to \mathbb{R}^{n+1}$, $F(x) = (f(x), g(x))$. Write

$$F(x_0) = y_0 = (f(x_0), 0).$$

If

$$\operatorname{rank} F'(x_0) = \operatorname{rank} \begin{pmatrix} \nabla f(x_0) \\ \nabla g^1(x_0) \\ \vdots \\ \nabla g^n(x_0) \end{pmatrix} = n+1$$

By Theorem 6.8, $\exists \varepsilon > 0$ such that

$$B_\varepsilon^{n+1}(y_0) \subset F(B_\delta(x_0)).$$

Noting that $\left(f(x_0) - \frac{\varepsilon}{2}, 0\right) \in B_\varepsilon^{n+1}(y_0)$, there is $x^* \in U$ such that

$$\left(f(x_0) - \frac{\varepsilon}{2}, 0\right) = F(x^*) = (f(x^*), g(x^*)).$$

That is $f(x^*) = f(x_0) - \frac{\varepsilon}{2}$, $g(x^*) = 0$. This means $x^* \in B_\delta(x_0) \cap M$, but $f(x^*) < f(x_0)$, contradicting 6.17.

Thus $\operatorname{rank} F'(x_0) = n$. Since the last n rows of $F'(x_0)$, $\nabla g^1(x_0), \ldots, \nabla g^n(x_0)$ are linearly independent, we have $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$ such that

$$\nabla f(x_0) + \sum_{i=1}^n \lambda_i \nabla g^i(x_0) = 0,$$

this is 6.13.

We conclude with another geometric proof.

wlin **Lemma 6.20.** *Suppose $v_1, \ldots, v_n \in \mathbb{R}^m$ are linearly independent, $v \in \mathbb{R}^m$. If*

$$h \perp \operatorname{span}\{v_1, \ldots, v_n\} \implies h \perp v,$$

then $v \in \operatorname{span}\{v_1, \ldots, v_n\}$.

Proof. Extend $v_1, \ldots, v_n$ to a base of $\mathbb{R}^m$, $v_1, \ldots, v_m$, such that for $i, j \in \{n+1, \ldots, m\}$ we have

$$v_i \perp \operatorname{span}\{v_1, \ldots, v_n\}, \qquad v_i \cdot v_j = \delta_{ij}.$$

Let $v = \sum_{i=1}^m \kappa^i v_i$. If $j > n$ we have $v_j \perp \operatorname{span}\{v_1, \ldots, v_n\}$, thus $v_j \perp v$ and

$$0 = v \cdot v_j = \sum_{i=1}^m \kappa^i v_i \cdot v_j = \kappa^j.$$

So

$$v = \sum_{i=1}^n \kappa^i v_i \in \operatorname{span}\{v_1, \ldots, v_n\}.$$

Proof (C). Let $x_0 \in M$ be minimizer of $f : M \to \mathbb{R}$. For $\forall h \in \ker g'(x_0)$, there is a curve $\gamma : (-\varepsilon, \varepsilon) \to M$ on M passing x_0, such that $\dot{\gamma}(0) = h$. We have

$$\nabla f(x_0) \cdot h = \nabla f(x_0) \cdot \dot{\gamma}(0) = \left.\frac{\mathrm{d}}{\mathrm{d}t}\right|_{t=0} f(\gamma(t)) = 0.$$

Noting that $h \in \ker g'(x_0)$ means h is perpendicular to the row vectors of $g'(x_0)$, we deduce from Lemma 6.20

$$\nabla f(x_0) \in \operatorname{span}\left\{\nabla g^1(x_0), \ldots, \nabla g^n(x_0)\right\},$$

thus 6.13 is proved.