

MA 5037: Optimization Methods and Applications

Mathematical Preliminaries



Suh-Yuh Yang (楊肅煜)

Department of Mathematics, National Central University
Jhongli District, Taoyuan City 320317, Taiwan

First version: April 14, 2018/Last updated: August 31, 2026

Vector space $\mathbb{R}^n$

- **Vector space $\mathbb{R}^n$:** the set of n -dimensional column vectors with real components endowed with the following component-wise addition operator and the scalar-vector product,

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} := \begin{bmatrix} x_1 + y_1 \\ x_2 + y_2 \\ \vdots \\ x_n + y_n \end{bmatrix}, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n,$$

$$\lambda \mathbf{x} = \lambda \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} := \begin{bmatrix} \lambda x_1 \\ \lambda x_2 \\ \vdots \\ \lambda x_n \end{bmatrix}, \quad \forall \mathbf{x} \in \mathbb{R}^n, \lambda \in \mathbb{R}.$$

- **Standard basis of $\mathbb{R}^n$:** $\{e_1, e_2, \dots, e_n\}$, $e_i := [\dots, 0, \underbrace{1}_{i\text{th}}, 0, \dots]^\top$.
- **Notation:** $\mathbf{e} := [1, 1, \dots, 1]^\top$ and $\mathbf{0} := [0, 0, \dots, 0]^\top$.

Important subsets of $\mathbb{R}^n$

- **Nonnegative orthant:**

$$\mathbb{R}_+^n := \{(x_1, x_2, \dots, x_n)^\top : x_i \geq 0 \forall i\}.$$

Positive orthant:

$$\mathbb{R}_{++}^n := \{(x_1, x_2, \dots, x_n)^\top : x_i > 0 \forall i\}.$$

- **Closed line segment:** let $x, y \in \mathbb{R}^n$,

$$[x, y] := \{(1 - \alpha)x + \alpha y : \alpha \in [0, 1]\}.$$

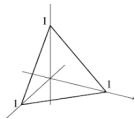
Open line segment: let $x, y \in \mathbb{R}^n$,

$$(x, y) := \{(1 - \alpha)x + \alpha y : \alpha \in (0, 1)\}.$$

$$[x, x] = \{x\} \text{ and } (x, x) = \emptyset.$$

- **Unit-simplex (單位單體):** $x + y + z = 1, x, y, z \geq 0$.

$$\Delta_n := \{x = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n : x_1, x_2, \dots, x_n \geq 0, e^\top x = 1\}.$$



Vector space $\mathbb{R}^{m \times n}$

- The set of all real-valued matrices of order $m \times n$ is denoted by $\mathbb{R}^{m \times n}$.
- The $n \times n$ identity matrix is denoted by I_n .
- The $m \times n$ zero matrix is denoted by $\mathbf{0}_{m \times n}$.
- *We will frequently omit the subscripts of these matrices when the dimensions will be clear from the context.*

Inner product on $\mathbb{R}^n$

- **Definition:** An inner product on $\mathbb{R}^n$ is a map $\langle \cdot, \cdot \rangle : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}$ with the following properties:
 - (1) *symmetry*: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle, \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
 - (2) *additivity*: $\langle \mathbf{x}, \mathbf{y} + \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{x}, \mathbf{z} \rangle, \forall \mathbf{x}, \mathbf{y}, \mathbf{z} \in \mathbb{R}^n$.
 - (3) *homogeneity*: $\langle \lambda \mathbf{x}, \mathbf{y} \rangle = \lambda \langle \mathbf{x}, \mathbf{y} \rangle, \forall \lambda \in \mathbb{R} \text{ and } \mathbf{x}, \mathbf{y} \in \mathbb{R}^n$.
 - (4) *positive definiteness*: $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0, \forall \mathbf{x} \in \mathbb{R}^n, \langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}$.
- **Example 1:** (dot product) The *standard inner product* is defined by

$$\langle \mathbf{x}, \mathbf{y} \rangle := \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^n x_i y_i, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

- **Example 2:** (weighted dot product) Let $\mathbf{w} \in \mathbb{R}_{++}^n$. Then the following weighted dot product is also an inner product:

$$\langle \mathbf{x}, \mathbf{y} \rangle_{\mathbf{w}} := \sum_{i=1}^n w_i x_i y_i, \quad \forall \mathbf{x}, \mathbf{y} \in \mathbb{R}^n.$$

Vector norms

- **Definition:** A norm $\|\cdot\|$ on $\mathbb{R}^n$ is a function $\|\cdot\| : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfying the following properties:
 - (1) *nonnegativity:* $\|x\| \geq 0$, $\forall x \in \mathbb{R}^n$, and $\|x\| = 0 \Leftrightarrow x = 0$.
 - (2) *positive homogeneity:* $\|\lambda x\| = |\lambda| \|x\|$, $\forall \lambda \in \mathbb{R}$ and $x \in \mathbb{R}^n$.
 - (3) *triangle inequality:* $\|x + y\| \leq \|x\| + \|y\|$, $\forall x, y \in \mathbb{R}^n$.
- **The associated norm with an inner product:** One natural way to generate a norm on $\mathbb{R}^n$ is to take any inner product $\langle \cdot, \cdot \rangle$ on $\mathbb{R}^n$ and define the associated norm

$$\|x\| := \sqrt{\langle x, x \rangle}, \quad \forall x \in \mathbb{R}^n.$$

If the inner product is the dot product (i.e., the standard inner product), then the associated norm is the so-called *Euclidean norm* or *ℓ_2 -norm*:

$$\|x\|_2 = \sqrt{\sum_{i=1}^n x_i^2}, \quad \forall x \in \mathbb{R}^n.$$

By default, the underlying norm on $\mathbb{R}^n$ is $\|\cdot\|_2$ and the subscript 2 will be frequently omitted.

- The ℓ_p -norm, $p \geq 1$, is defined by

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

- The ℓ_∞ -norm is defined by

$$\|\mathbf{x}\|_\infty = \max_{i=1,2,\dots,n} |x_i|, \quad \forall \mathbf{x} \in \mathbb{R}^n,$$

and unsurprisingly, it can be shown that

$$\|\mathbf{x}\|_\infty = \lim_{p \rightarrow \infty} \|\mathbf{x}\|_p.$$

- **The Cauchy-Schwarz inequality:** *For any $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$, we have*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| (= |\mathbf{x}^\top \mathbf{y}|) \leq \|\mathbf{x}\|_2 \|\mathbf{y}\|_2.$$

Equality is satisfied if and only if $\mathbf{x}$ and $\mathbf{y}$ are linearly dependent.

Supplement

- For $0 < p < 1$,

$$\|\mathbf{x}\|_p = \left(\sum_{i=1}^n |x_i|^p \right)^{1/p}, \quad \forall \mathbf{x} \in \mathbb{R}^n.$$

is not a norm on $\mathbb{R}^n$.

Give an example to explain why $\|\cdot\|_{\frac{1}{2}}$ is not a norm!

- Let $\|\mathbf{x}\|_0 := \#\{i : x_i \neq 0\}$. Since

$$\lim_{p \rightarrow 0^+} |x_i|^p = \begin{cases} 1 & \text{if } x_i \neq 0 \\ 0 & \text{if } x_i = 0 \end{cases} \quad \forall \mathbf{x} = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n,$$

we have

$$\lim_{p \rightarrow 0^+} \|\mathbf{x}\|_p^p = \lim_{p \rightarrow 0^+} \sum_{i=1}^n |x_i|^p = \sum_{i=1}^n \lim_{p \rightarrow 0^+} |x_i|^p = \#\{i : x_i \neq 0\} = \|\mathbf{x}\|_0.$$

Therefore, in general, $\lim_{p \rightarrow 0^+} \|\mathbf{x}\|_p \neq \|\mathbf{x}\|_0$.

Matrix norms

- **Definition:** A norm $\|\cdot\|$ on $\mathbb{R}^{m \times n}$ is a function $\|\cdot\| : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}$ satisfying the following properties:
 - (1) *nonnegativity:* $\|A\| \geq 0$, $\forall A \in \mathbb{R}^{m \times n}$, and $\|A\| = 0 \Leftrightarrow A = \mathbf{0}$.
 - (2) *positive homogeneity:* $\|\lambda A\| = |\lambda| \|A\|$, $\forall \lambda \in \mathbb{R}$, $A \in \mathbb{R}^{m \times n}$.
 - (3) *triangle inequality:* $\|A + B\| \leq \|A\| + \|B\|$, $\forall A, B \in \mathbb{R}^{m \times n}$.
- **Induced norms:** Given a matrix $A \in \mathbb{R}^{m \times n}$ and two norms $\|\cdot\|_a$ and $\|\cdot\|_b$ on $\mathbb{R}^n$ and $\mathbb{R}^m$, respectively, the induced matrix norm $\|A\|_{a,b}$ is defined by

$$\|A\|_{a,b} := \max\{\|Ax\|_b : x \in \mathbb{R}^n \text{ and } \|x\|_a \leq 1\}.$$

Note: An induced norm is a norm.

It can be shown that for any $x \in \mathbb{R}^n$, we have

$$\|Ax\|_b \leq \|A\|_{a,b} \|x\|_a.$$

- We refer to the matrix-norm $\|\cdot\|_{a,b}$ as the (a,b) -norm. When $a = b$, we will simply refer to it as an a -norm.

Matrix norms (cont'd)

- **spectral norm or ℓ_2 -norm:** If $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_2$, the induced $(2,2)$ -norm of $A = (A_{ij}) \in \mathbb{R}^{m \times n}$ is the maximum singular value of A ,

$$\|A\|_2 = \|A\|_{2,2} := \sqrt{\lambda_{\max}(A^\top A)} =: \sigma_{\max}(A).$$

This norm is called the spectral norm or ℓ_2 -norm. *Note that the eigenvalues λ_i ($i = 1, 2, \dots, n$) of $A^\top A$ are real and nonnegative.*

- **ℓ_1 -norm:** If $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_1$, the induced $(1,1)$ -norm of $A = (A_{ij}) \in \mathbb{R}^{m \times n}$ is given by

$$\|A\|_1 = \|A\|_{1,1} := \max_{j=1,2,\dots,n} \sum_{i=1}^m |A_{ij}|.$$

- **ℓ_∞ -norm:** If $\|\cdot\|_a = \|\cdot\|_b = \|\cdot\|_\infty$, the induced (∞, ∞) -norm of $A = (A_{ij}) \in \mathbb{R}^{m \times n}$ is given by

$$\|A\|_\infty = \|A\|_{\infty,\infty} := \max_{i=1,2,\dots,m} \sum_{j=1}^n |A_{ij}|.$$

- **Frobenius norm:** A non-induced norm is defined by

$$\|A\|_F := \left(\sum_{i=1}^m \sum_{j=1}^n A_{ij}^2 \right)^{1/2}, \quad \forall A = (A_{ij}) \in \mathbb{R}^{m \times n}.$$

Eigenvalues and eigenvectors

- **Definition:** Let $A \in \mathbb{R}^{n \times n}$. Then a nonzero vector $v \in \mathbb{C}^n$ is called an *eigenvector* of A if there exists a $\lambda \in \mathbb{C}$ for which $Av = \lambda v$. The scalar λ is called the *eigenvalue* corresponding to the eigenvector v .

Note: $\exists \mathbf{0} \neq v \in \mathbb{C}^n$ s.t. $Av = \lambda v \Rightarrow Av - \lambda Iv = (A - \lambda I)v = \mathbf{0}$.
 $\Rightarrow \det(A - \lambda I) = 0$.

- $f_A(\lambda) := \det(A - \lambda I)$ is called the *characteristic polynomial* of A .

$$f_A(\lambda) = (-1)^n \lambda^n + (-1)^{n-1} \underbrace{(a_{11} + \cdots + a_{nn})}_{:=\text{trace}(A)} \lambda^{n-1} + \cdots + \det(A).$$

- *In general, real-valued matrices can have complex eigenvalues, but it is well known that all the eigenvalues of symmetric matrices are real.*

The eigenvalues of a symmetric matrix $A \in \mathbb{R}^{n \times n}$ are denoted by

$$\underbrace{\lambda_1(A)}_{:=\lambda_{\max}(A)} \geq \lambda_2(A) \geq \cdots \geq \lambda_{n-1}(A) \geq \underbrace{\lambda_n(A)}_{:=\lambda_{\min}(A)}$$

The spectral decomposition (factorization) theorem

- **The spectral decomposition theorem:** *Let $A \in \mathbb{R}^{n \times n}$ be a symmetric matrix. Then there exists an orthogonal matrix $U \in \mathbb{R}^{n \times n}$, $U^\top U = UU^\top = I$, and a diagonal matrix $D = \text{diag}(d_1, d_2, \dots, d_n)$ such that $A = UDU^\top$.*

Note that the columns of the matrix U in the factorization constitute an orthonormal basis comprised of eigenvectors of A and the diagonal elements of D are the corresponding eigenvalues.

- Let $A \in \mathbb{R}^{n \times n}$. The trace and the determinant of A can be expressed via its eigenvalues:

$$\text{trace } A = \sum_{i=1}^n \lambda_i(A) \quad \text{and} \quad \det A = \prod_{i=1}^n \lambda_i(A).$$

Note: If A is symmetric, then $f_D(\lambda) = \det(D - \lambda I) = \det(U^\top (A - \lambda I) U)$
 $= \det(U^\top) \det(A - \lambda I) \det(U) = \det(A - \lambda I) = f_A(\lambda).$

Rayleigh quotient

- **Definition:** For a symmetric matrix $A \in \mathbb{R}^{n \times n}$, the Rayleigh quotient is defined by

$$R_A(\mathbf{x}) := \frac{\mathbf{x}^\top A \mathbf{x}}{\|\mathbf{x}\|^2}, \quad \forall \mathbf{x} \neq \mathbf{0}.$$

- **Lower and upper bounds on the Rayleigh quotient:**

Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then

$$\lambda_{\min}(A) \leq R_A(\mathbf{x}) \leq \lambda_{\max}(A), \quad \forall \mathbf{x} \neq \mathbf{0}.$$

Proof.

- (i) By the spectral decomposition theorem, $\exists$ an orthogonal $U \in \mathbb{R}^{n \times n}$ such that $U^\top A U = D$, $D = \text{diag}(d_1, d_2, \dots, d_n)$, and $\lambda_{\max}(A) = d_1 \geq d_2 \geq \dots \geq d_n = \lambda_{\min}(A)$.
- (ii) $\forall \mathbf{x} \neq \mathbf{0}$, making the change of variables $\mathbf{x} = U\mathbf{y}$, then $\mathbf{y} \neq \mathbf{0}$ and we have

$$\frac{\mathbf{x}^\top A \mathbf{x}}{\|\mathbf{x}\|^2} = \frac{\mathbf{y}^\top U^\top A U \mathbf{y}}{\|U\mathbf{y}\|^2} = \frac{\mathbf{y}^\top D \mathbf{y}}{\mathbf{y}^\top \underbrace{U^\top U}_{I} \mathbf{y}} = \frac{\sum_i d_i y_i^2}{\sum_i y_i^2}.$$

$$\Rightarrow \lambda_{\min}(A) = d_n = \frac{d_n(\sum_i y_i^2)}{\sum_i y_i^2} \leq R_A(\mathbf{x}) \leq \frac{d_1(\sum_i y_i^2)}{\sum_i y_i^2} = \lambda_{\max}(A) = d_1. \quad \square$$

The minimal and maximal eigenvalues

Let $A \in \mathbb{R}^{n \times n}$ be symmetric. Then

- $\min_{x \neq 0} R_A(x) = \lambda_{\min}(A)$, and the eigenvectors of A corresponding to the minimal eigenvalue are minimizers.

Proof: Let v be an eigenvector corresponding to the minimal eigenvalue of A . Then

$$R_A(v) = \frac{v^\top A v}{\|v\|^2} = \frac{\lambda_{\min}(A) \|v\|^2}{\|v\|^2} = \lambda_{\min}(A),$$

which combined with the lower bound on the Rayleigh quotient lead to the desired result. $\square$

- $\max_{x \neq 0} R_A(x) = \lambda_{\max}(A)$, and the eigenvectors of A corresponding to the maximal eigenvalue are maximizers.

Proof: Let w be an eigenvector corresponding to the maximal eigenvalue of A . Then

$$R_A(w) = \frac{w^\top A w}{\|w\|^2} = \frac{\lambda_{\max}(A) \|w\|^2}{\|w\|^2} = \lambda_{\max}(A),$$

which combined with the upper bound on the Rayleigh quotient lead to the desired result. $\square$

Basic topological concepts

- **Open ball:** The open ball with center $\mathbf{c} \in \mathbb{R}^n$ and radius $r > 0$ is defined by

$$B(\mathbf{c}, r) := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{c}\| < r\}.$$

The open ball $B(\mathbf{c}, r)$ is also referred to as a neighborhood of $\mathbf{c}$.

- **Close ball:** The close ball with center $\mathbf{c} \in \mathbb{R}^n$ and radius $r > 0$ is defined by

$$B[\mathbf{c}, r] := \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{c}\| \leq r\}.$$

- **Interior point:** Given a set $U \subseteq \mathbb{R}^n$, a point $\mathbf{c} \in U$ is an interior point of U if there exists $r > 0$ for which $B(\mathbf{c}, r) \subseteq U$.
- **Interior set:** The set of all interior points of a given set U is called the interior of the set and is denoted by $\text{int}(U)$, i.e.,

$$\text{int}(U) := \{\mathbf{x} \in U : B(\mathbf{x}, r) \subseteq U \text{ for some } r > 0\}.$$

Example: (1) $\text{int}(\mathbb{R}_+^n) = \mathbb{R}_{++}^n$. (2) $\text{int}(B[\mathbf{c}, r]) = B(\mathbf{c}, r)$.

Open set, closed set, and boundary point

- **Open set:** $U \subseteq \mathbb{R}^n$ is an open set if and only if for every $\mathbf{x} \in U$ there exists $r > 0$ such that $B(\mathbf{x}, r) \subseteq U$.

Example: $\mathbb{R}^n$, open balls, positive orthant $\mathbb{R}_{++}^n$ are open sets.

Note: (1) A union of any number of open sets is an open set.

(2) The intersection of a finite number of open sets is open.

- **Closed set:** A set $U \subseteq \mathbb{R}^n$ is said to be closed if for every sequence of points $\{\mathbf{x}_k\} \subseteq U$ satisfying $\mathbf{x}_k \rightarrow \mathbf{x}^*$ as $k \rightarrow \infty$, it holds that $\mathbf{x}^* \in U$.

Example: closed ball $B[\mathbf{c}, r]$, closed line segments, nonnegative orthant $\mathbb{R}_+^n$, unit simplex Δ_n , $\mathbb{R}^n$ are closed sets.

- **Proposition:** Let f be a continuous function defined over a closed set $S \subseteq \mathbb{R}^n$. Then for any $\alpha \in \mathbb{R}$ the following sets are closed:

$$\text{Lev}(f, \alpha) := \{\mathbf{x} \in S : f(\mathbf{x}) \leq \alpha\}, \quad (\text{level set})$$

$$\text{Con}(f, \alpha) := \{\mathbf{x} \in S : f(\mathbf{x}) = \alpha\}. \quad (\text{contour set})$$

Boundedness and compactness

- **Boundary point:** Given a set $U \subseteq \mathbb{R}^n$, a boundary point of U is a point $\mathbf{x} \in \mathbb{R}^n$ satisfying the following: any neighborhood of $\mathbf{x}$ contains at least one point in U and at least one point in U^c , i.e.,

$$\forall r > 0, B(\mathbf{x}, r) \cap U \neq \emptyset \text{ and } B(\mathbf{x}, r) \cap U^c \neq \emptyset.$$

- **Boundary of U :** The set of all boundary points of a set U is called the boundary of U and is denoted by $bd(U)$.

Example: $bd(B(\mathbf{c}, r)) = bd(B[\mathbf{c}, r]) = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x} - \mathbf{c}\| = r\}$.

- **Closure of U :** The closure of a set $U \subseteq \mathbb{R}^n$ is defined to be the smallest closed set containing U and denoted by $cl(U)$, i.e.,

$$cl(U) := \cap \{T : U \subseteq T, T \text{ is closed}\}.$$

Note: (1) The closure set is indeed a closed set as an intersection of closed sets containing U . (2) $cl(U) = U \cup bd(U)$.

- **Boundedness:** A set $U \subseteq \mathbb{R}^n$ is called bounded if $\exists M > 0$ s.t. $U \subseteq B(\mathbf{0}, M)$.
- **Compactness:** A set $U \subseteq \mathbb{R}^n$ is called compact if it is closed and bounded.

Differentiability

- **Directional derivative:** Let f be a real-valued function defined on a set $S \subseteq \mathbb{R}^n$. Let $\mathbf{x} \in \text{int}(S)$ and let $\mathbf{0} \neq \mathbf{d} \in \mathbb{R}^n$. If

$$\lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{d}) - f(\mathbf{x})}{t}$$

exists, then it is called the directional derivative of f at $\mathbf{x}$ along the direction $\mathbf{d}$ and is denoted by $f'(\mathbf{x}; \mathbf{d})$.

Note that here we do not assume that $\mathbf{d}$ is a unit vector $\|\mathbf{d}\| = 1$.

- **Partial derivatives:** For $i = 1, 2, \dots, n$, the directional derivative of f at $\mathbf{x}$ along the direction $\mathbf{e}_i$ is called the i th partial derivative, i.e.,

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) := \lim_{t \rightarrow 0} \frac{f(\mathbf{x} + t\mathbf{e}_i) - f(\mathbf{x})}{t}.$$

- The gradient of f at $\mathbf{x}$ is defined as

$$\nabla f(\mathbf{x}) = \left[\frac{\partial f}{\partial x_1}(\mathbf{x}), \frac{\partial f}{\partial x_2}(\mathbf{x}), \dots, \frac{\partial f}{\partial x_n}(\mathbf{x}) \right]^\top.$$

Continuous differentiability

- **Definition:** A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called continuously differentiable over U if all the partial derivatives exist and are continuous on U . (i.e., $f \in C^1(U)$)
- **Definition:** A function f is said to be continuously differentiable over a set C if there exists an open set U containing C on which the function is also defined and continuously differentiable.
- Let f be continuously differentiable over open set U . Then

$$f'(x; d) = \nabla f(x)^\top d, \quad \forall x \in U, d \in \mathbb{R}^n.$$

- **Proposition:** Let $f : U \rightarrow \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. Assume that f is continuously differentiable over U . Then

$$\lim_{d \rightarrow 0} \frac{f(x + d) - f(x) - \nabla f(x)^\top d}{\|d\|} = 0, \quad \forall x \in U,$$

or equivalently,

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + o(\|y - x\|),$$

where $o(\cdot) : \mathbb{R}_+ \rightarrow \mathbb{R}$ satisfies $\frac{o(t)}{t} \rightarrow 0$ as $t \rightarrow 0^+$.

Twice continuous differentiability

- **Definition:** A function f defined on an open set $U \subseteq \mathbb{R}^n$ is called twice continuously differentiable over U if all the second order partial derivatives exist and are continuous over U . (i.e., $f \in C^2(U)$)
- **Proposition:** Let $f : U \rightarrow \mathbb{R}$ be defined on an open set $U \subseteq \mathbb{R}^n$. If f is twice continuously differentiable, then for any $i \neq j$ and any $\mathbf{x} \in U$,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) = \frac{\partial^2 f}{\partial x_j \partial x_i}(\mathbf{x}).$$

- The Hessian of f at a point $\mathbf{x} \in U$ is the $n \times n$ matrix

$$\nabla^2 f(\mathbf{x}) := \nabla(\nabla^\top f(\mathbf{x})) := \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n}(\mathbf{x}) \\ \frac{\partial^2 f}{\partial x_2 \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_2^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_2 \partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \cdots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1}(\mathbf{x}) & \frac{\partial^2 f}{\partial x_n \partial x_2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_n^2}(\mathbf{x}) \end{bmatrix}.$$

If f is twice continuously differentiable over U , then the Hessian matrix is symmetric (by Clairaut-Schwarz theorem).

Linear and quadratic approximation theorems

There are two main approximation results which are consequences of Taylor's approximation theorem:

- ① **Linear approximation theorem:** *Let $f : U \rightarrow \mathbb{R}$ be a twice continuously differentiable function over an open set $U \subseteq \mathbb{R}^n$, and let $x \in U$, $r > 0$ satisfy $B(x, r) \subseteq U$. Then for any $y \in B(x, r)$, there exists $\xi \in (x, y)$ such that*

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2}(y - x)^\top \nabla^2 f(\xi)(y - x).$$

- ② **Quadratic approximation theorem:** *Let $f : U \rightarrow \mathbb{R}$ be a twice continuously differentiable function over an open set $U \subseteq \mathbb{R}^n$, and let $x \in U$, $r > 0$ satisfy $B(x, r) \subseteq U$. Then for any $y \in B(x, r)$,*

$$f(y) = f(x) + \nabla f(x)^\top (y - x) + \frac{1}{2}(y - x)^\top \nabla^2 f(x)(y - x) + o(\|y - x\|^2).$$