MATH 4211/6211 – Optimization

Review

Xiaojing Ye
Department of Mathematics & Statistics
Georgia State University

Vector spaces and matrices

A column n-vector $\boldsymbol{a}$ is denoted

$$\boldsymbol{a} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} \in \mathbb{R}^n$$

or $\boldsymbol{a}^\top = [a_1, a_2, \dots, a_n]$.

Operations on vectors:

- Sum of two vectors: $\boldsymbol{a} + \boldsymbol{b}$.
- Scalar multiplication: $\lambda \boldsymbol{a}$.

A **linear combination** of vectors $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k$ is:

$$\lambda_1 \boldsymbol{a}_1 + \lambda_2 \boldsymbol{a}_2 + \cdots + \lambda_k \boldsymbol{a}_k,$$

where $\lambda_1, \dots, \lambda_k \in \mathbb{R}$ are called combination coefficients.

The set of linear combinations of $\boldsymbol{a}_1, \dots, \boldsymbol{a}_k$ is denoted by

$$\text{span}(\boldsymbol{a}_1, \dots, \boldsymbol{a}_k) := \left\{ \sum_{i=1}^{k} \lambda_i \boldsymbol{a}_i : \lambda_i \in \mathbb{R} \right\}.$$

The span of vectors is a **vector space** $\mathcal{V}$.

Proposition. A set of vectors $\{\boldsymbol{a}_1, \dots, \boldsymbol{a}_k\}$ are linearly dependent iff* one of the vectors is a linear combination of the remaining vectors.

*"iff" stands for "if and only if".

Definition. $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ is called a **basis** of the vector space $\mathcal{V}$ if they are linearly independent and $\mathcal{V} = \mathsf{span}(\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k)$. The size k of a basis is called the dimension of $\mathcal{V}$.

Proposition. If $\{\boldsymbol{a}_1, \ldots, \boldsymbol{a}_k\}$ is a basis of $\mathcal{V}$, then any vector $\boldsymbol{a} \in \mathcal{V}$ can be represented uniquely as

$$\boldsymbol{a} = \lambda_1 \boldsymbol{a}_1 + \lambda_2 \boldsymbol{a}_2 + \cdots + \lambda_k \boldsymbol{a}_k.$$

We often denote the natural basis of $\mathcal{V} = \mathbb{R}^n$ as $\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n$ where

$$\boldsymbol{e}_i^\top = [0, \ldots, 0, \underbrace{1}_{i\text{-th}}, 0, \ldots, 0].$$

Matrices

A matrix is a rectangular array of numbers:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \vdots & a_{2n} \\ \vdots & \cdots & \ddots & \cdots \\ a_{m1} & a_{m2} & \vdots & a_{mn} \end{bmatrix} \in \mathbb{R}^{m \times n}.$$

Sum of two matrices and scalar multiplication are defined similarly.

Definition. The maximal number of linearly independent columns (or rows) of $\boldsymbol{A}$ is called the **rank** of $\boldsymbol{A}$, denoted by $\mathtt{rank}(\boldsymbol{A})$.

The following operations do not change the rank of $\boldsymbol{A}$:

- Multiplying nonzero scalars to the columns of $\boldsymbol{A}$.
- Interchanging any two columns.
- Adding a linear combination of columns to another column.

The same types of row operations do not change $\mathtt{rank}(\boldsymbol{A})$ either.

Determinant

Let $\boldsymbol{A}$ be a square matrix:

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \vdots & a_{2n} \\ \vdots & \cdots & \ddots & \cdots \\ a_{n1} & a_{2n} & \vdots & a_{nn} \end{bmatrix}.$$

The **determinant** of a square matrix $\boldsymbol{A}$ is defined recursively as

$$\det(\boldsymbol{A}) = \sum_{i=1}^{n} (-1)^{i+1} a_{i1} \det(\boldsymbol{A}_{i1}),$$

where $\boldsymbol{A}_{ij} \in \mathbb{R}^{(n-1)\times(n-1)}$ is $\boldsymbol{A}$ with its i-th row and j-th column deleted, and $\det(A_{ij})$ is called the **principal minor**.

Definition. A square matrix $\boldsymbol{A}$ is called **invertible** (or **nonsingular**) if there exists a matrix $\boldsymbol{B}$ such that $\boldsymbol{AB} = \boldsymbol{BA} = \boldsymbol{I}$. We denote $\boldsymbol{A}^{-1} = \boldsymbol{B}$.

Proposition. $\mathsf{det}(\boldsymbol{A}) \neq 0$ iff $\mathsf{rank}(A) = n$ iff A is invertible.

Definition. Let $\boldsymbol{A}$ be a square matrix. Then

- $\boldsymbol{A}$ is **symmetric** if $\boldsymbol{A} = \boldsymbol{A}^\top$.
- $\boldsymbol{A}$ is **orthogonal** if $\boldsymbol{AA}^\top = \boldsymbol{A}^\top\boldsymbol{A} = \boldsymbol{I}$. Clearly an orthogonal matrix is invertible.

Inner Products and Norms

For $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, the inner product of $\boldsymbol{x}$ and $\boldsymbol{y}$ is

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

Properties:

- Positivity: $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$; and $= 0$ iff $\boldsymbol{x} = \boldsymbol{0}$.
- Symmetry: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle$.
- Additivity: $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$.
- Homogeneity: $\langle \lambda \boldsymbol{x}, \boldsymbol{y} \rangle = \lambda \langle \boldsymbol{x}, \boldsymbol{y} \rangle$ for any $\lambda \in \mathbb{R}$.

Due to symmetry, additivity and homogeneity also hold for the second argument.

Norms

The (Euclidean) **norm** of $\boldsymbol{x}$ is defined by $\|\boldsymbol{x}\| = \sqrt{\langle \boldsymbol{x}, \boldsymbol{x} \rangle}$.

Properties:

- Positivity: $\|\boldsymbol{x}\| \geq 0$; and $= 0$ iff $\boldsymbol{x} = \mathbf{0}$.
- Homogeneity: $\|\lambda \boldsymbol{x}\| = |\lambda| \|\boldsymbol{x}\|$ for any $\lambda \in \mathbb{R}$.
- Triangle inequality: $\|\boldsymbol{x} + \boldsymbol{y}\| \leq \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

Cauchy-Schwarz inequality. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, there is

$$|\langle \boldsymbol{x}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{x}\| \|\boldsymbol{y}\|.$$

The equality holds iff $\boldsymbol{x} = \lambda \boldsymbol{y}$ for some $\lambda \in \mathbb{R}$ or $\boldsymbol{y} = \boldsymbol{0}$.

Proposition. For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, there is

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 = \|\boldsymbol{x}\|^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + \|\boldsymbol{x}\|^2.$$

General vector norms. We define p-norm of x as

$$\|x\|_p = \begin{cases} \Big(|x_1|^p + \cdots + |x_n|^p\Big)^{1/p}, & \text{if } 1 \leq p < \infty, \\ \max\left(|x_1|, \ldots, |x_n|\right), & \text{if } p = \infty. \end{cases}$$

Eigenvalues and eigenvectors

Definition. Let $\boldsymbol{A}$ be a square matrix. If $\lambda \in \mathbb{C}$ and nonzero $\boldsymbol{x} \in \mathbb{C}^n$ are such that

$$\boldsymbol{A}\boldsymbol{x} = \lambda \boldsymbol{x}.$$

Then λ and $\boldsymbol{x}$ are respectively called **eigenvalue** and **eigenvector** of $\boldsymbol{A}$.

Note that $\mathtt{det}(\lambda \boldsymbol{I} - \boldsymbol{A})$ is a polynomial of λ of degree n. It is called the **characteristic polynomial** of $\boldsymbol{A}$.

Proposition. λ is an eigenvalue of $\boldsymbol{A}$ iff $\mathtt{det}(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$ (i.e., λ is a root of the characteristic polynomial of $\boldsymbol{A}$).

Theorem. If $\mathtt{det}(\lambda \boldsymbol{I} - \boldsymbol{A}) = 0$ has n distinct roots $\lambda_1, \dots, \lambda_n$, then there exist n linearly independent eigenvectors $\boldsymbol{v}_1, \dots, \boldsymbol{v}_n$ such that

$$\boldsymbol{A}\boldsymbol{v}_i = \lambda_i \boldsymbol{v}_i, \quad i = 1, \dots, n.$$

Theorem. All eigenvalues of a symmetric matrix $\boldsymbol{A}$ are real. If in addition $\boldsymbol{A}$ is real, then all the corresponding eigenvectors are mutually orthogonal, i.e., $\langle \boldsymbol{v}_i, \boldsymbol{v}_j \rangle = 0$ for all $i \neq j$.

Orthogonal projections

Let $\mathcal{V}$ be a linear subspace of $\mathbb{R}^n$. Then the **orthogonal complement** of $\mathcal{V}$ is defined by

$$\mathcal{V}^\perp := \{\boldsymbol{u} \in \mathbb{R}^n : \boldsymbol{v}^\top \boldsymbol{u} = 0, \ \forall\, \boldsymbol{v} \in \mathcal{V}\}.$$

Then any $\boldsymbol{x} \in \mathbb{R}$ can be uniquely decomposed as

$$\boldsymbol{x} = \boldsymbol{u} + \boldsymbol{v}, \quad \boldsymbol{v} \in \mathcal{V},\ \boldsymbol{u} \in \mathcal{V}^\perp.$$

We also write $\mathbb{R}^n = \mathcal{V} \oplus \mathcal{V}^\perp$, called the direct sum of $\mathcal{V}$ and $\mathcal{V}^\perp$.

We say $\boldsymbol{P} \in \mathbb{R}^{n\times n}$ the **orthogonal projector** onto $\mathcal{V}$ if for all $\boldsymbol{x} \in \mathbb{R}^n$ we have $\boldsymbol{P}\boldsymbol{x} \in \mathcal{V}$ and $\boldsymbol{x} - \boldsymbol{P}\boldsymbol{x} \in \mathcal{V}^\perp$.

Kernel and range of matrices

Let $\boldsymbol{A} \in \mathbb{R}^{m \times n}$. Then the **range** of $\boldsymbol{A}$ is

$$\mathcal{R}(\boldsymbol{A}) := \{\boldsymbol{A}\boldsymbol{x} : \boldsymbol{x} \in \mathbb{R}^n\} \subset \mathbb{R}^m$$

which is the span of the columns of $\boldsymbol{A}$. So $\mathcal{R}(\boldsymbol{A})$ is also called the **column space** of $\boldsymbol{A}$.

The **kernel** of $\boldsymbol{A}$ is

$$\mathcal{N}(\boldsymbol{A}) := \{\boldsymbol{x} : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\} \subset \mathbb{R}^n$$

which is the orthogonal complement of the span of the rows of $\boldsymbol{A}$. So $\mathcal{N}(\boldsymbol{A}) = \mathcal{R}(\boldsymbol{A}^\top)^\perp$.

Theorem. $\boldsymbol{P}$ is an **orthogonal projector** (onto the subspace $\mathcal{V} = \mathcal{R}(\boldsymbol{P})$) iff $\boldsymbol{P}^2 = \boldsymbol{P} = \boldsymbol{P}^\top$.

Quadratic forms

We call $f : \mathbb{R}^n \to \mathbb{R}$ a **quadratic form** if

$$f(\boldsymbol{x}) = \boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x}$$

for some real square matrix $\boldsymbol{Q}$.

Without loss of generality, we assume $\boldsymbol{Q} = \boldsymbol{Q}^\top$: if $\boldsymbol{Q}$ is not symmetric, then replace it with $\frac{1}{2}(\boldsymbol{Q} + \boldsymbol{Q}^\top)$ because

$$\boldsymbol{x}^\top \boldsymbol{Q} \boldsymbol{x} = \boldsymbol{x}^\top \boldsymbol{Q}^\top \boldsymbol{x} = \frac{1}{2} \boldsymbol{x}^\top (\boldsymbol{Q} + \boldsymbol{Q}^\top) \boldsymbol{x}$$

for any $\boldsymbol{x}$.

Positive definite matrices

Definition. We say Q is **positive semidefinite** (denoted $Q \succeq 0$) if $x^\top Q x \geq 0$ for all $x \in \mathbb{R}$. If in addition $=$ holds only at $x = 0$, then we say Q is **positive definite**, denoted $Q \succ 0$. We say Q is **negative (semi)definite** if $-Q$ is positive (semi)definite.

Sylvester's criterion. A symmetric Q is positive definite iff all its leading principal minors are positive.

Theorem. A symmetric Q is positive definite (or positive semidefinite) iff all eigenvalues of Q are positive (or nonnegative).

Hyperplanes and half-spaces

Let $\boldsymbol{a} \in \mathbb{R}^n$ and $b \in \mathbb{R}$, then

$$H := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} = b\}$$

is called a **hyperplane** in $\mathbb{R}^n$.

A hyperplane divides $\mathbb{R}^n$ into two **half-spaces**:

$$H_+ := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} \geq b\}$$
$$H_- := \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}^\top \boldsymbol{x} \leq b\}$$

Linear varieties

Let $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ and $\boldsymbol{b} \in \mathbb{R}^m$ be such that $\boldsymbol{b} \in \mathcal{R}(\boldsymbol{A})$, then the **linear variety** is defined by

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}.$$

If $\text{dim}(\mathcal{N}(\boldsymbol{A})) = r$, the linear variety has dimension r.

It is obvious that

$$\{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\} = \bigcap_{i=1}^{m} \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{a}_i^\top \boldsymbol{x} = b_i\}$$

where $\boldsymbol{a}_i^\top$ is the i-th row of $\boldsymbol{A}$.

A linear variety is a subspace iff $\boldsymbol{b} = \boldsymbol{0}$.

Convex sets

For any $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, the **line segment** between $\boldsymbol{x}$ and $\boldsymbol{y}$ is

$$\{\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} : \lambda \in [0, 1]\}$$

A set $C \subset \mathbb{R}^n$ is called **convex** if

$$\lambda \boldsymbol{x} + (1 - \lambda)\boldsymbol{y} \in C$$

for any $\boldsymbol{x}, \boldsymbol{y} \in C$ and $\lambda \in [0, 1]$.

In other words, C is convex iff the line segment between any two points in C lies in C.

Examples of convex sets include:

- the empty set
- a set consisting of a single point
- a line or a line segment
- a subspace
- hyperplane
- balls and ellipses

Theorem. Let C_1 and C_2 be two convex sets, then $C_1 \cap C_2$ is convex, and

$$C_1 + C_2 := \{\boldsymbol{x}_1 + \boldsymbol{x}_2 : \boldsymbol{x}_1 \in C_1,\ \boldsymbol{x}_2 \in C_2\}$$

is also convex.

Neighborhoods

A **neighborhood** of a point $\boldsymbol{x} \in \mathbb{R}^n$ is defined by

$$B_\varepsilon(\boldsymbol{x}) := \{\boldsymbol{y} \in \mathbb{R}^n : \|\boldsymbol{y} - \boldsymbol{x}\| < \varepsilon\}$$

for some $\varepsilon > 0$. Note that $B_\varepsilon(\boldsymbol{x})$ is open.

Let $S \subset \mathbb{R}^n$, then $\boldsymbol{x}$ is called an **interior point** of S if there exists $\varepsilon > 0$ such that $B_\varepsilon(\boldsymbol{x}) \subset S$. The set of interior points of S is called the **interior** of S, denoted by $\mathtt{int}(S)$.

$\boldsymbol{x}$ is called a **boundary point** of S if any neighborhood of $\boldsymbol{x}$ contains a point in S and a point in S^c. A boundary point may or may not be in S. The set of boundary points of S is called the **boundary** of S.

Open sets, closed sets, compact sets

A set $S \subset \mathbb{R}^n$ is called **open** if all its point are an interior points. S is called **closed** if S^c is open. S is called **bounded** if $S \subset B_R(\mathbf{0})$ for some $R > 0$. S is called **compact** if S is closed and bounded.

Weierstrass theorem. Let $S \subset \mathbb{R}^n$ be compact and $f : S \to \mathbb{R}$ be continuous, then f attains maximum and minimum in S.

Polytopes and polyhedra

The intersection of finitely many half-spaces is called a **polytope**. Note that a polytope is convex, since all half-spaces are convex.

A nonempty bounded polytope is called a **polyhedron**.

Sequences and limits

Let $\boldsymbol{x}^{(1)}, \ldots, \boldsymbol{x}^{(k)}, \ldots$ be a sequence in $\mathbb{R}^n$, then we say $\boldsymbol{x}^{(k)}$ **converges** to $\boldsymbol{x}^*$ if for any $\varepsilon > 0$, there exists $K \in \mathbb{N}$ (depending on ε) such that

$$\|\boldsymbol{x}_k - \boldsymbol{x}^*\| < \varepsilon$$

for all $k \geq K$. This is denoted by $\lim_{k\to\infty} \boldsymbol{x}^{(k)} = \boldsymbol{x}^*$ or $\boldsymbol{x}^{(k)} \to \boldsymbol{x}^*$. $\boldsymbol{x}^*$ is called the **limit** of the sequence $(\boldsymbol{x}^{(k)})_{k=1}^{\infty}$. If a sequence is convergent, then the limit is unique. Note that $\boldsymbol{x}^{(k)} \to \boldsymbol{x}^*$ iff $x_i^{(k)} \to x_i^*$ for all $i = 1, \ldots, n$.

Theorem. A convergent sequence is bounded. A bounded sequence has at least one convergent subsequence.

Theorem. A sequence $(\boldsymbol{x}^{(k)})_{k=1}^{\infty}$ converges to $\boldsymbol{x}^*$ iff every subsequence of $(\boldsymbol{x}^{(k)})_{k=1}^{\infty}$ converges to $\boldsymbol{x}^*$.

Continuous functions

We say $f : \mathbb{R}^n \to \mathbb{R}^m$ is **continuous** at $\boldsymbol{x} \in \mathbb{R}^n$ if

$$f(\boldsymbol{x}^{(k)}) \to f(\boldsymbol{x})$$

for any sequence $\boldsymbol{x}^{(k)} \to \boldsymbol{x}$.

We say f is continuous on $S \subset \mathbb{R}^n$ if f is continuous at every point of S.

Gradient and Jacobian

Let $f : \mathbb{R}^n \to \mathbb{R}$, then the **gradient** of f at $\boldsymbol{x}$ is

$$\nabla f(\boldsymbol{x}) := \left[\frac{\partial f}{\partial x_1}(\boldsymbol{x}), \dots, \frac{\partial f}{\partial x_n}(\boldsymbol{x})\right] \in \mathbb{R}^{1\times n}$$

where $\frac{\partial f}{\partial x_i}(\boldsymbol{x})$ is the i-th partial derivative of f at $\boldsymbol{x}$:

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}) := \lim_{h\to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{e}_i) - f(\boldsymbol{x})}{h}.$$

Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$, then the **Jacobian** of $\boldsymbol{f} = [f_1, \dots, f_m]^\top$ at $\boldsymbol{x}$ is

$$D\boldsymbol{f}(\boldsymbol{x}) = \begin{bmatrix} \nabla f_1(\boldsymbol{x}) \\ \vdots \\ \nabla f_m(\boldsymbol{x}) \end{bmatrix} \in \mathbb{R}^{m\times n}$$

Differentiation rules

Chain rule. Let $\boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^m$ and $\boldsymbol{g} : \mathbb{R}^m \to \mathbb{R}^k$, then their composition is $\boldsymbol{g} \circ \boldsymbol{f} : \mathbb{R}^n \to \mathbb{R}^k$, and the Jacobian of $\boldsymbol{g} \circ \boldsymbol{f}$ at $\boldsymbol{x}$ is

$$D(\boldsymbol{g} \circ \boldsymbol{f})(\boldsymbol{x}) = \underbrace{D\boldsymbol{g}(\boldsymbol{f}(\boldsymbol{x}))}_{k \times m} \underbrace{D\boldsymbol{f}(\boldsymbol{x})}_{m \times n} \in \mathbb{R}^{k \times n}.$$

Product rule. Let $\boldsymbol{f}, \boldsymbol{g} : \mathbb{R}^n \to \mathbb{R}^m$, then $\boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{g}(\boldsymbol{x}) \in \mathbb{R}$ for any $\boldsymbol{x} \in \mathbb{R}^n$ and

$$\nabla(\boldsymbol{f}(\boldsymbol{x})^\top \boldsymbol{g}(\boldsymbol{x})) = \underbrace{\boldsymbol{f}(\boldsymbol{x})^\top}_{1 \times m} \underbrace{D\boldsymbol{g}(\boldsymbol{x})}_{m \times n} + \underbrace{\boldsymbol{g}(\boldsymbol{x})^\top}_{1 \times m} \underbrace{D\boldsymbol{f}(\boldsymbol{x})}_{m \times n} \in \mathbb{R}^{1 \times n}$$

Level sets

The **level set** of a function $f : \mathbb{R}^n \to \mathbb{R}$ at level $c \in \mathbb{R}$ is

$$S_c := \{\boldsymbol{x} \in \mathbb{R}^n : f(\boldsymbol{x}) = c\}$$

If $n = 2$ then S_c is a curve. If $n = 3$ then S_c is a surface.

Theorem. For any c, $\nabla f(\boldsymbol{x})$ is orthogonal to the tangent of S_c at $\boldsymbol{x} \in S_c$.

In fact, $\frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|}$ is the direction of fastest increase (steepest ascent direction) of f at $\boldsymbol{x}$ (if $\nabla f(\boldsymbol{x}) \neq \boldsymbol{0}$).

Taylor theorem

Let $f:\mathbb{R}\to\mathbb{R}$ and $f\in\mathcal{C}^m$, and denote $h=b-a$, then

$$f(b)=f(a)+\frac{h}{1!}f^{(1)}(a)+\frac{h^2}{2!}f^{(2)}(a)+\cdots+\frac{h^{m-1}}{(m-1)!}f^{(m-1)}(a)+R_m$$

where $f^{(i)}$ is the i-th derivative of f and

$$R_m=\frac{h^m(1-\theta)^{m-1}}{(m-1)!}f^{(m)}(a+\theta h)=\frac{h^m}{m!}f^{(m)}\left(a+\theta' h\right)$$

with $\theta,\theta'\in(0,1)$.

Let $f:\mathbb{R}^n\to\mathbb{R}$ and $f\in\mathcal{C}^2$, and denote $\boldsymbol{h}=\boldsymbol{b}-\boldsymbol{a}$, then

$$f(\boldsymbol{b})=f(\boldsymbol{a})+Df(\boldsymbol{a})\boldsymbol{h}+\frac{1}{2}\boldsymbol{h}^\top D^2f(\boldsymbol{a})\boldsymbol{h}+o(\|\boldsymbol{h}\|^2),$$

where $\lim_{\|\boldsymbol{h}\|\to 0} o(\|\boldsymbol{h}\|^2)/\|\boldsymbol{h}\|^2=0$.