

## Note: Math Tools

Nov 2024

*Lecturer:*

*Typed by: Zhuohua Shen*



# Contents

<b>1</b>	<b>Complex Analysis</b>	<b>5</b>
1.1	Complex Numbers	5
1.1.1	Transformation	5
1.2	Complex functions	6
1.2.1	Complex sequences and series	6
1.2.2	Basic complex functions	6
1.2.3	Complex differentiability	7
<b>2</b>	<b>Matrix Algebra</b>	<b>9</b>
2.1	Spaces	9
2.1.1	Spaces, rank, and related factorization	9
2.2	Inversion	9
2.2.1	Types of inversion	9
2.2.2	Special matrices	11
<b>3</b>	<b>Optimization</b>	<b>13</b>
3.1	Nonlinear Optimization	13
3.1.1	KKT Conditions	13



# Chapter 1

## Complex Analysis

References:

- CUHKSZ: MAT3253 - Complex Variables notes by Kenneth Shum (Spring 2023)

### 1.1 Complex Numbers

**Polar form of complex numbers**  $z = x + iy = r(\cos \theta + i \sin \theta)$  for  $r, \theta \geq 0$ .

- If  $z_k = r_k(\cos \theta_k + i \sin \theta_k)$  for  $k = 1, 2$ , then  $z_1 z_2 = r_1 r_2 [\cos(\theta_1 + \theta_2) + i \sin(\theta_1 + \theta_2)]$ .
- If  $z_1 z_2 z_3 = 0$ , then at least one of the three factors is zero.
- If  $\Re(z_1), \Re(z_2) > 0$ , then  $\text{Arg}(z_1 z_2) = \text{Arg}(z_1) + \text{Arg}(z_2)$ , where principal arguments in  $(-\pi, \pi]$  are used.

**Properties of complex numbers** (i)  $(z^*)^* = z$ ; (ii)  $z^* = z$  iff  $z \in \mathbb{R}$ ; (iii)  $z z^* = |z|^2 = x^2 + y^2$ ; (iv)  $z_1, z_2 \in \mathbb{C}$ ,  $(z_1 + z_2)^* = z_1^* + z_2^*$ ,  $(z_1 z_2)^* = z_1^* z_2^*$ ; (v)  $\Re(z) = (z + z^*)/2$ ,  $\Im(z) = (z - z^*)/(2i)$ ; (vi)  $|z_1 + z_2| \leq |z_1| + |z_2|$ ; (vii)  $z_1 \neq z_2$ , then  $|z_2 - z_1|^2 = r_1^2 + r_2^2 - 2r_1 r_2 \cos(\theta_1 - \theta_2)$ ; (viii)  $|z_1 + z_2|^2 \leq |z_1|^2 + 2|\Re(z_1 z_2^*)| + |z_2|^2$ , and  $|\Re(z_1 z_2^*)| \leq |z_1| |z_2|$ .

- (**DeMoivre formula**)  $\forall n \in \mathbb{Z}, \theta \in \mathbb{R}$ ,  $(\cos \theta + i \sin \theta)^n = \cos(n\theta) + i \sin(n\theta)$
- (**Binomial formula**)  $(z_1 + z_2)^m = \sum_{k=0}^m \binom{m}{k} z_1^k z_2^{m-k}$  for  $m \in \mathbb{N}^+$ ,  $z_1, z_2 \in \mathbb{C}$ .
- (Geometric series)  $\sum_{k=0}^n z^k = (1 - z^{n+1})/(1 - z)$ .

**$n$ -th root of a complex number**  $w$  is the  $n$ -th root of  $z_0$  if  $w^n = z_0$ .

- ( $n$ -th root of unity)  $\forall n \in \mathbb{N}^+$ , the solution of  $z^n = 1$  is  $z = \cos(2\pi k/n) + i \sin(2\pi k/n)$ ,  $k = 0, \dots, n-1$ . If we write  $w = \cos(2\pi/n) + i \sin(2\pi/n)$ , then the  $n$ -th root is  $w^k$ ,  $k = 0, \dots, n-1$ .

#### Example 1.1.1.

- (*Summation of  $\cos k\theta$* )

$$1 + \cos \theta + \cos 2\theta + \dots + \cos n\theta = \frac{1}{2} + \frac{\sin((2n+1)\theta/2)}{2 \sin(\theta/2)}, \quad 0 < \theta < 2\pi.$$

- (*Chebyshev polynomials*) Let  $m = n/2$  if  $n$  is even and  $(n-1)/2$  if  $n$  is odd, then

$$\cos n\theta = \sum_{k=0}^m \binom{n}{2k} (-1)^k \cos^{n-2k}(\theta) \sin^{2k}(\theta), \quad n \in \mathbb{N}.$$

Write  $x = \cos \theta$ , the above becomes a polynomial  $T_n(x)$  of degree  $n$  in the variable  $x$ .

#### 1.1.1 Transformation

**Linear fractional/Möbius/bilinear transformation**

$$f(z) = \frac{az + b}{cz + d} = \frac{a}{c} + \frac{bc - ad}{c} \frac{1}{cz + d}, \quad a, b, c, d \in \mathbb{C}, \quad ad - bc \neq 0.$$

- $b = 0, c = 0, d = 1$ , rotation  $f(z) = az = re^{i\theta}z$ ;
- $a = 1, c = 0, d = 1$ , translation  $f(z) = z + b$ ;
- $a = 0, b = 1, c = 1, d = 0$ , inversion function  $f(z) = 1/z$ , that maps circles and straight lines to circles and straight lines;
- $f(z) = rz$ ,  $0 < r \in \mathbb{R}$ , scaling.

All four types of transformation maps circle/line to circle/line. If  $ad - bc = 0$ , then  $f(z)$  is a constant.

When  $z = -d/c$ ,  $f(z) = \infty$ , we extend the domain. The **Riemann sphere** is three-dimensional sphere with the south pole touching the origin of the complex plane. The **stereographic projection** is a function that maps a complex number  $z = x + iy$  in the complex plane to the point  $P(x, y)$  on the Riemann sphere such that  $(x, y), P(x, y)$  and the north pole of the sphere are colinear. The north pole of the sphere does not correspond to any point on the complex

plane and is called the **point at infinity**, and is denoted by the symbol  $\infty$ . The Riemann sphere is often called the **one-point compactification** of the complex plane.

**Extended complex number system/extended complex plane**  $\bar{\mathbb{C}}, \hat{\mathbb{C}} = \mathbb{C} \cup \{\infty\}$ .

- Given complex numbers  $a, b, c, d \in \mathbb{C}$  with  $ad - bc \neq 0$ , define a linear fractional transformation on  $\bar{\mathbb{C}}$  by

$$f(z) = \begin{cases} \frac{az+b}{cz+d} & \text{if } z \neq -d/c, z \neq \infty \\ \infty & \text{if } z = -d/c \\ a/c & \text{if } z = \infty \end{cases}$$

which is a bijection on the Riemann sphere.

## 1.2 Complex functions

### 1.2.1 Complex sequences and series

- Distance  $d(z_1, z_2) = |z_1 - z_2|$ .
- Open disc** of radius  $r$  centered at  $z_0$ :  $D(z_0, r) = \{z \in \mathbb{C} : |z - z_0| < r\}$ . Neighborhood of  $\infty$  is  $\{z \in \mathbb{C} : |z| > R\}$  for some large  $R$ .
- Convergence** of complex sequence  $(z_n)_{n=1}^\infty$ : converges to  $L \in \mathbb{C}$  if  $\forall \epsilon > 0, \exists N \in \mathbb{N}$  s.t.  $|z_n - L| < \epsilon, \forall n > N$ . We write  $\lim_{n \rightarrow \infty} z_n = L$ .  $(z_n)_{n=1}^\infty$  converges to  $\infty$  iff  $1/|z_n| \rightarrow 0$  as  $n \rightarrow \infty$  ( $|z_n| \rightarrow \infty$ ). It **diverges** if  $z_n$  does not converge to any  $L \in \mathbb{C}$  ( $\rightarrow \infty$  is also divergent for  $\mathbb{C}$ ).
- Cauchy sequence**  $\forall \epsilon > 0, \exists N \in \mathbb{N}$  s.t.  $|z_m - z_n| \leq \epsilon, \forall m, n \geq N$ .
  - If  $z_n = x_n + iy_n$ , then  $z_n$  is Cauchy iff  $x_n, y_n$  Cauchy.
  - $z_n$  converges iff  $z_n$  is Cauchy.
- Complex series**  $\sum_{k=1}^\infty z_k := \lim_n \sum_{k=1}^n z_k$  if the limit exists. We call it **converges absolutely** if  $\sum_{k=1}^\infty |z_k|$  converges. We call it **converges conditionally** if  $\sum_{k=1}^\infty z_k$  converges but  $\sum_{k=1}^\infty |z_k|$  diverges.
  - (n-th term test)** If  $\sum_{k=1}^\infty z_k$  converges, then  $\lim_n |z_n| = 0$ . If  $|z_n| \not\rightarrow 0$ , then  $\sum_k z_k$  diverges.
  - (Absolute convergence test)**  $\sum_{k=1}^\infty |z_k|$  converges, then  $\sum_{k=1}^\infty z_k$  converges.
  - (Limit ratio test)** Assume  $\lim_n |a_{n+1}/a_n|$  exists and is equal to  $L$ . (a)  $L > 1 \Rightarrow \sum_{k=1}^\infty a_n$  diverges, (b)  $L < 1 \Rightarrow \sum_{k=1}^\infty a_n$  converges absolutely, (c)  $L = 1$ , no conclusion.
  - If  $\sum_{k=0}^\infty a_k$  and  $\sum_{k=0}^\infty b_k$  converges absolutely, then  $(\sum_{k=0}^\infty a_k)(\sum_{k=0}^\infty b_k) = (\sum_{k=0}^\infty c_k)$ , where  $c_k = \sum_{j=0}^k a_j b_{k-j}$ .
  - If a series converges absolutely, then a series obtained by rearranging the terms converges to the same limit.

### 1.2.2 Basic complex functions

**Power series** A complex power series centered at the origin is a series in the form  $\sum_{k=0}^\infty a_k z_k^k$ ,  $a_k \in \mathbb{C}$ .

**Definition 1.2.1.** For  $z \in \mathbb{C}$ , define

- (complex exponential function)**

$$e^z := \exp(z) := \sum_{n=0}^\infty \frac{z^n}{n!},$$

we have  $e^{z_1+z_2} = e^{z_1}e^{z_2}$ ,  $e^{-z} = (e^z)^{-1}$ ,  $e^z \neq 0, \forall z \in \mathbb{C}$ , and  $e^{a+ib} = e^a e^{ib}$ ,  $a, b \in \mathbb{R}$ .

- (complex trigonometric, hyperbolic trigonometric)**

$$\begin{aligned} \sin(z) &:= \sum_{n=0}^\infty (-1)^n \frac{z^{2n+1}}{(2n+1)!}, & \cos(z) &:= \sum_{n=0}^\infty (-1)^n \frac{z^{2n}}{(2n)!}, & \tan(z) &:= \frac{\sin(z)}{\cos(z)}, \\ \sinh(z) &:= \sum_{n=0}^\infty \frac{z^{2n+1}}{(2n+1)!}, & \cosh(z) &:= \sum_{n=0}^\infty \frac{z^{2n}}{(2n)!}, & \tanh(z) &:= \frac{\sinh(z)}{\cosh(z)}. \end{aligned}$$

They are all converges absolutely.

**Theorem 1.2.2.**  $\forall z \in \mathbb{C}$ ,

- (Euler's formula)**  $e^{iz} = \cos z + i \sin z$

•

$$\begin{aligned} \cos z &= \frac{e^{iz} + e^{-iz}}{2}, & \sin z &= \frac{e^{iz} - e^{-iz}}{2i}, \\ \cosh(z) &= \frac{e^z + e^{-z}}{2}, & \sinh(z) &= \frac{e^z - e^{-z}}{2}, \end{aligned}$$

thus we have  $\cosh(iz) = \cos z$ ,  $\sinh(iz) = i \sin z$ .

Since  $e^z = e^{z+2\pi ki}$ ,  $k \in \mathbb{Z}$ , the inverse function of  $e^z$  is multi-valued. For  $0 \neq w \in \mathbb{C}$ , define the **complex log function** as

$$\log(w) := \log|w| + i(\arg(w) + 2\pi k), \quad k \in \mathbb{Z}.$$

Define the **principal complex log function** as

$$\text{Log}(w) := \log|w| + i \arg(w), \quad \arg(w) \in (-\pi, \pi] \text{ or } [0, 2\pi).$$

Given  $0 \neq z \in \mathbb{C}$ , define the **complex power** by

$$z^w := \exp(w \log(z)).$$

### The angle function, parametric curve and winding number

Suppose  $\theta(z)$  is continuous,  $\theta(z_0) = 0$ , then as  $z \rightarrow z_0$  from the right,  $\theta(z) \rightarrow 2\pi \neq 0$ . To prevent closed cycle around the origin, let the domain of **angle function** be the half plane.

- If  $H = \{x + iy : y > 0\}$ , the range is  $(0, \pi)$ , then for  $z \in H$ , define  $F(x, y) := \cos^{-1}(x/\sqrt{x^2 + y^2})$ .
- If  $H_\alpha = \{x + iy : y > \tan(\alpha)x\}$  for  $\alpha > 0$ , define  $F_\alpha := F(e^{-i\alpha}z) + \alpha$ .

The **parametric curve** is a function  $\gamma : [a, b] \rightarrow \mathbb{C}$  continuous,  $\gamma(t)$  is the location at time  $t$ . If  $\gamma(a) = \gamma(b)$ , then we call it **closed curve**.

Given  $\gamma : [a, b] \rightarrow \mathbb{C} \setminus \{0\}$ , divide  $[a, b]$  into  $a = t_0 < t_1 < \dots < t_{n-1} < t_n = b$  s.t.  $\gamma(t)$ ,  $t_k \leq t \leq t_{k+1}$ , is inside  $H_{\alpha(k)}$  for  $k = 0, \dots, n-1$ . Define

- **change in angle** in the  $k$ th part:  $F_{\alpha(k)}(\gamma(t_{k+1})) - F_{\alpha(k)}(\gamma(t_k))$ ,
- **overall change of angle**:  $\sum_{k=0}^{n-1} [F_{\alpha(k)}(\gamma(t_{k+1})) - F_{\alpha(k)}(\gamma(t_k))]$ ,
- **branch**: A continuous angle function as a function of  $t$ .

Note that it doesn't depend on the sub-division of the curve and how we parameterize the curve. The **winding number/index** of a closed parametric curve not passing through the origin is  $(2\pi)^{-1}(\text{change in angle})$ .

### 1.2.3 Complex differentiability

#### Limit and continuity





# Chapter 2

## Matrix Algebra

References:

- Peng Ding - *Linear Model and Extensions*: appendix A.
- Robb J. Muirhead - *Aspects of multivariate statistical theory* [1]: appendix A.

Notations:

- **Column space/range/image** of  $n \times m$  matrix  $A = (\mathbf{a}_1, \dots, \mathbf{a}_m)$  is  $\text{Col}(A) = \{\alpha_1 \mathbf{a}_1 + \dots + \alpha_m \mathbf{a}_m : \alpha_1, \dots, \alpha_m \in \mathbb{R}\}$ .
- **Row space** of  $A$  is  $\text{Col}(A^\top)$ .
- **Kernel/null space/nullspace** of  $A$  is  $\ker(A) = \{\mathbf{v} \in \mathbb{R}^m : A\mathbf{v} = \mathbf{0}_n\}$ .
- **Rank** of  $A$  is  $r(A)$ .
- **Trace** of  $A$  is  $\text{tr}(A)$ .
- **Eigenvalues** of  $A$  are  $\lambda(A)$ .

### 2.1 Spaces

#### 2.1.1 Spaces, rank, and related factorization

**Theorem 2.1.1** (Rank). Let  $A \in \mathbb{R}^{n \times m}$ .

1.  $r(A) = r(A^\top)$ .
2.  $r(A) \leq \min(n, m)$ .
3.  $r(AB) \leq \min\{r(A), r(B)\}$ .
4. If  $A, B$  have same size, then  $r(A + B) \leq r(A) + r(B)$ .
5. If  $A \in \mathbb{R}^{n \times n}$ ,  $B \in \mathbb{R}^{n \times m}$ ,  $C \in \mathbb{R}^{m \times m}$ , and  $A$  and  $C$  are nonsingular, then  $r(ABC) = r(B)$ .
6. If  $B \in \mathbb{R}^{m \times p}$  such that  $AB = \mathbf{0}$ , then  $r(B) \leq m - r(A)$ .
7.  $A$  is full column/row rank iff  $A^\top A / AA^\top$  is nonsingular.

Point 7 is proved by SVD.

**Theorem 2.1.2** (Rank-related factorization).

1. (**Full-rank factorization**) If  $A \in \mathbb{R}^{n \times m}$  with  $r(A) = k$ , then  $A = BC$  for some full column rank  $B \in \mathbb{R}^{n \times k}$  and full rank  $C \in \mathbb{R}^{k \times m}$ .
2. (**Non-negative definite**) If  $n \times n$   $A \succeq \mathbf{0}$ ,  $r(A) = r$ , then
  - $\exists B \in \mathbb{R}^{n \times r}$  of rank  $r$  such that  $A = BB^\top$ ;
  - $\exists C \in \mathbb{R}^{n \times n}$  nonsingular such that  $A = C \begin{bmatrix} I_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} C^\top$ .

### 2.2 Inversion

#### 2.2.1 Types of inversion

Left and right inverse

**Definition 2.2.1** (Left/right inverse). Let  $A \in \mathbb{R}^{n \times m}$ . The **left inverse**  $A_{left}^{-1} \in \mathbb{R}^{m \times n}$  satisfies  $A_{left}^{-1}A = I_m$ , and the **right inverse**  $A_{right}^{-1} \in \mathbb{R}^{m \times n}$  satisfies  $AA_{right}^{-1} = I_n$ .

We can find these inverses by Theorem 2.1.1: if

- 10 • A full column rank  $r(A) = m$ , then  $A^\top A$  nonsingular, and  $A_{left}^{-1} = (A^\top A)^{-1}A^\top$ ; CHAPTER 2. MATRIX ALGEBRA
- A full row rank  $r(A) = n$ , then  $AA^\top$  nonsingular, and  $A_{right}^{-1} = A^\top(AA^\top)^{-1}$ .

### Moore–Penrose inverse

It is used when left/right inverse cannot be obtained.

**Definition 2.2.2** (Moore–Penrose inverse/pseudoinverse). Let  $A \in \mathbb{R}^{n \times m}$ .  $A^+ \in \mathbb{R}^{m \times n}$  is defined as a **Moore–Penrose (M-P) inverse** of  $A$  if

1.  $AA^+$  and  $A^+A$  are symmetric;
2.  $AA^+A = A$ ;
3.  $A^+AA^+ = A^+$ .

**Example 2.2.3** (M-P inverse of a diagonal matrix). Let  $D \in \mathbb{R}^{n \times m}$ , WLOG, assume  $n \geq m$ .  $D_{ii} = d_i$  for  $i = 1, \dots, m$ , while others are zero. Then  $D^+ \in \mathbb{R}^{m \times n}$  satisfies  $D_{ii}^+ = 1/d_i$  if  $d_i \neq 0$  and zero otherwise,  $i = 1, \dots, m$ .

**Theorem 2.2.4.** Each matrix  $A$  has an  $A^+$ .

If  $A = \mathbf{0}$ , then  $A^+ = \mathbf{0}$ . If  $A \neq \mathbf{0}$ , two ways to construct  $A^+$ :

1. Full-rank factorization  $A = B_{n \times r}C_{r \times m}$  of rank  $r$ . We have

$$A^+ = C^\top(CC^\top)^{-1}(B^\top B)^{-1}B^\top.$$

2. SVD  $A = UDV^\top$ , then  $A^+ = VD^+U^\top$ .

**Theorem 2.2.5.** Let  $A \in \mathbb{R}^{n \times m}$ .

1. The M-P inverse is unique.
2.  $(A^\top)^+ = (A^+)^\top$ .
3.  $r(A^+) = r(A)$ .
4. If  $A$  is symmetric, then  $A^+ = (A^+)^\top$ .
5. If  $A$  is nonsingular, then  $A^{-1} = A^+$ .
6. If  $A$  is symmetric idempotent, then  $A^+ = A$ .
7. If  $r(A) = m$ , then  $A^+ = A_{left}^{-1} = (A^\top A)^{-1}A^\top$ .
8. If  $r(A) = n$ , then  $A^+ = A_{right}^{-1} = A^\top(AA^\top)^{-1}$ .
9. The matrices  $AA^+, A^+A, I_n - AA^+$ , and  $I_m - A^+A$  are all symmetric idempotent.

### Generalized Inverse

**Definition 2.2.6** (Generalized inverse). Let  $A \in \mathbb{R}^{n \times m}$ . The **generalized inverse (G-inverse)**  $A^- \in \mathbb{R}^{m \times n}$  satisfies

$$AA^-A = A.$$

It is obvious that the M-P inverse is also a G-inverse.

**Theorem 2.2.7.**

1. G-inverse may not be unique.
2. Let  $A \in \mathbb{R}^{n \times m}$  with rank  $k > 0$ .
  - (a)  $r(A^-) \geq k$ .
  - (b)  $A^-A$  and  $AA^-$  are idempotent.
  - (c)  $r(A^-A) = r(AA^-) = k$ .
  - (d)  $A^-A = I_m$  iff  $r(A) = m$ .
  - (e)  $AA^- = I_n$  iff  $r(A) = n$ .
  - (f)  $\text{tr}(A^-A) = \text{tr}(AA^-) = k = r(A)$ .
  - (g) If  $A^-$  is any G-inverse of  $A$ , then  $(A^-)^\top$  is a G-inverse of  $A^\top$ .
3. Let  $K = A(A^\top A)^-A^\top \in \mathbb{R}^{n \times n}$ ,
  - (a)  $K$  is invariant for any G-inverse of  $A^\top A$ .
  - (b)  $K = AA^+$ .
  - (c)  $K$  is symmetric and idempotent.
  - (d)  $r(K) = \text{tr}(K) = r(A)$ .

(e)  $KA = A; A^\top K = A^\top$ .

(f)  $(A^\top A)^- A^\top$  is a  $G$ -inverse of  $A$  for any  $G$ -inverse of  $A^\top A$ .

(g)  $A(A^\top A)^-$  is a  $G$ -inverse of  $A^\top$  for any  $G$ -inverse of  $A^\top A$ .

## 2.3 Special matrices

### 2.3.1 Idempotent matrices

**Definition 2.3.1** (Idempotent matrices).  $A \in \mathbb{R}^{n \times n}$  if  $A^2 = A$ .

**Theorem 2.3.2.**  $A$  is idempotent.

1. All idempotent matrices(except  $I$ ) are singular.
2.  $r(A) = \text{tr}(A)$ .
3.  $\lambda(A)$  is either 0 or 1.
4. If  $A$  symmetric, all  $\lambda(A)$ 's are 0 or 1, then  $A$  is idempotent.



# Chapter 3

## Optimization

References:

- CUHKSZ: MAT3007 - Optimization I
- CUHKSZ: MAT3220 - Optimization II. Textbook:
- 1. *Introduction to Nonlinear Optimization: Theory, Algorithms, and Applications with MATLAB*, Amir Beck.
- 2. *Convex Optimization*, S. Boyd and L. Vandenberghe.
- 3. *Nonlinear Programming*, D. Bertsekas.
- 4. *First-Order Methods in Optimization*, Amir Beck

### 3.1 Nonlinear Optimization

#### 3.1.1 KKT Conditions

**Theorem 3.1.1** (The Fritz-John necessary conditions). *Let  $\mathbf{x}^*$  be a local minimum of the problem*

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m \end{aligned}$$

*where  $f, g_1, \dots, g_m \in C^1(\mathbb{R}^n)$ . Then  $\exists$  multipliers  $\lambda_0, \dots, \lambda_m \geq 0$ , which are not all zeros, such that*

$$\begin{aligned} \lambda_0 \nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}^*) &= \mathbf{0} \\ \lambda_i g_i(\mathbf{x}^*) &= 0, \quad i = 1, 2, \dots, m. \end{aligned}$$

A major drawback of the Fritz-John conditions is, they allow  $\lambda_0 = 0$ . Under an additional [regularity condition](#), we can assume  $\lambda_0 = 1$ . Let  $I(\mathbf{x}^*)$  be the set of active constraints at  $\mathbf{x}^*$ :

$$I(\mathbf{x}^*) = \{i : g_i(\mathbf{x}^*) = 0\}.$$

**Theorem 3.1.2** (The KKT conditions for inequality constrained problems). *Let  $\mathbf{x}^*$  be a local minimum of*

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m \end{aligned}$$

*where  $f, g_1, \dots, g_m \in C^1(\mathbb{R}^n)$ . If  $\{\nabla g_i(\mathbf{x}^*)\}_{i \in I(\mathbf{x}^*)}$  are linearly independent. Then  $\exists \lambda_1, \dots, \lambda_m \geq 0$  such that*

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}^*) &= \mathbf{0} \\ \lambda_i g_i(\mathbf{x}^*) &= 0, \quad i = 1, 2, \dots, m. \end{aligned}$$

**Theorem 3.1.3** (The KKT conditions for inequality/equality constrained problems). *Let  $\mathbf{x}^*$  be a local minimum of*

$$\begin{aligned} \min \quad & f(\mathbf{x}) \\ \text{s.t.} \quad & g_i(\mathbf{x}) \leq 0, \quad i = 1, 2, \dots, m \\ & h_j(\mathbf{x}) = 0, \quad j = 1, \dots, p \end{aligned} \tag{3.1}$$

*where  $f, g_1, \dots, g_m, h_1, \dots, h_p \in C^1(\mathbb{R}^n)$ . If  $\{\nabla g_i(\mathbf{x}^*), \nabla h_j(\mathbf{x}^*), i \in I(\mathbf{x}^*), j = 1, \dots, p\}$  are linearly independent.*

Then  $\exists \lambda_1, \dots, \lambda_m \geq 0, \mu_1, \dots, \mu_p \in \mathbb{R}$ , such that

$$\begin{aligned} \nabla f(\mathbf{x}^*) + \sum_{i=1}^m \lambda_i \nabla g_i(\mathbf{x}^*) + \sum_{j=1}^p \mu_j \nabla h_j(\mathbf{x}^*) &= \mathbf{0}, \\ \lambda_i g_i(\mathbf{x}^*) &= 0, \quad i = 1, 2, \dots, m. \end{aligned} \tag{3.2}$$

Consider problem (1), a feasible point  $\mathbf{x}^*$  is called a **KKT point** if  $\exists \lambda_1, \dots, \lambda_m \geq 0, \mu_1, \dots, \mu_p \in \mathbb{R}$ , such that (3.2) holds.  $\mathbf{x}^*$  is called **regular** if  $\{\nabla g_i(\mathbf{x}^*), \nabla h_j(\mathbf{x}^*), i \in I(\mathbf{x}^*), j = 1, \dots, p\}$  are linearly independent.

- The additional requirement of regularity is not required in linearly constrained problems in which no such assumption is needed.

# Bibliography

- [1] R. J. Muirhead. *Aspects of multivariate statistical theory*. John Wiley & Sons, 1982. [2](#)