# 第四章 最优性理论

#### 修贤超

https://xianchaoxiu.github.io

#### 目录

- 4.1 最优化问题解的存在性
- 4.2 无约束可微问题的最优性理论
- 4.3 无约束不可微问题的最优性理论
- 4.4 对偶理论
- 4.5 一般约束优化问题的最优性理论
- 4.6 带约束凸优化问题的最优性理论

# 最优化问题解的存在性

■ 考虑优化问题

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
s.t.  $x \in \mathcal{X}$ 

- □ 首先分析最优解的存在性
- □ 然后考虑如何求出其最优解
- (Weierstrass 定理) 紧集上的连续函数一定存在最大 (最小) 值
- 而在许多实际问题中, 定义域可能不是紧的, 目标函数也不一定连续

# 推广的 Weierstrass 定理

- 若函数  $f: \mathcal{X} \to (-\infty, +\infty]$  适当且闭, 且以下条件中任意一个成立
  - $\bigcirc$  dom  $f = \{x \in \mathcal{X} : f(x) < +\infty\}$  是有界的
  - $\Box$  存在一个常数  $\bar{\gamma}$  使得下水平集

$$C_{\bar{\gamma}} = \{ x \in \mathcal{X} \mid f(x) \le \bar{\gamma} \}$$

是非空且有界的

 $oxed{o}$  f 是强制的, 即对于任一满足  $\|x^k\| \to +\infty$  的点列  $\{x^k\} \subset \mathcal{X}$ , 都有

$$\lim_{k \to \infty} f(x^k) = +\infty$$

则函数 f 的最小值点集  $\{x \in \mathcal{X} \mid f(x) \leq f(y), \forall y \in \mathcal{X}\}$  非空且紧

■ 三个条件在本质上都是保证 *f*(*x*) 的最小值不能在无穷远处取到

#### 例子

■ 当定义域不是有界闭集时, 对于强制函数

$$f(x) = x^2, x \in \mathcal{X} = \mathbb{R}$$

其全局最优解一定存在

■ 对于适当且闭的函数

$$f(x) = e^{-x}, x \in \mathcal{X} = \mathbb{R}$$

不满足三个条件中任意一个, 因此不能断言其全局极小值点存在

#### 目录

- 4.1 最优化问题解的存在性
- 4.2 无约束可微问题的最优性理论
- 4.3 无约束不可微问题的最优性理论
- 4.4 对偶理论
- 4.5 一般约束优化问题的最优性理论
- 4.6 带约束凸优化问题的最优性理论

#### 无约束可微问题的最优性理论

■ 无约束可微优化问题通常表示为

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{1}$$

■ 对于可微函数 f 和点  $x \in \mathbb{R}^n$ , 如果存在向量 d 满足

$$\nabla f(x)^{\top} d < 0$$

那么称 d 为 f 在点 x 处的一个下降方向

- 一阶最优性条件是利用梯度 (一阶) 信息来判断给定点的最优性
- 在局部最优点处不能有下降方向

# 一阶必要条件

■ 假设 f 在全空间  $\mathbb{R}^n$  可微. 如果  $x^*$  是 (1) 的一个局部极小点, 那么

$$\nabla f(x^*) = 0$$

证明 任取  $v \in \mathbb{R}^n$ , 考虑 f 在点  $x = x^*$  处的泰勒展开

$$\frac{f(x^* + tv) - f(x^*)}{t} = v^{\top} \nabla f(x^*) + o(1)$$

根据  $x^*$  的最优性, 分别对 t 取点 0 处的左、右极限可知

$$\lim_{t \to 0^+} \frac{f(x^* + tv) - f(x^*)}{t} = v^\top \nabla f(x^*) \ge 0$$
$$\lim_{t \to 0^-} \frac{f(x^* + tv) - f(x^*)}{t} = v^\top \nabla f(x^*) \le 0$$

■ 称满足  $\nabla f(x) = 0$  的点 x 为 f 的稳定点 (或驻点、临界点)

# 二阶最优性条件

- 对于  $f(x) = x^3$ , 满足 f'(x) = 0 的点为  $x^* = 0$ , 但其不是局部最优解
- 假设 f 在点 x 的一个开邻域内是二阶连续可微的,考虑

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

则以下最优性条件成立

 $\square$  (二阶必要条件) 若  $x^*$  是 f 的一个局部极小点, 则

$$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0$$

□ (二阶充分条件) 若满足

$$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$$

则  $x^*$  是 f 的一个局部极小点

#### 证明

■ 必要性 若  $\nabla^2 f(x^*)$  有负的特征值  $\lambda_- < 0$ , 设  $\nabla^2 f(x^*)d = \lambda_- d$ , 则

$$\frac{f(x^* + d) - f(x^*)}{\|d\|^2} = \frac{1}{2} \frac{d^\top}{\|d\|} \nabla^2 f(x^*) \frac{d}{\|d\|} + o(1) = \frac{1}{2} \lambda_- + o(1)$$

当 ||d|| 充分小时,  $f(x^* + d) < f(x^*)$ , 这和点  $x^*$  的最优性矛盾

■ 充分性 由  $\nabla f(x^*) = 0$  时的二阶展开

$$\frac{f(x^* + d) - f(x^*)}{\|d\|^2} = \frac{d^\top \nabla^2 f(x^*) d + o(\|d\|^2)}{\|d\|^2} \ge \frac{1}{2} \lambda_{\min} + o(1)$$

当 ||d|| 充分小时有  $f(x^* + d) \ge f(x^*)$ , 即二阶充分条件成立

# 实例: 实数情形的相位恢复

■考虑

$$\min_{x \in \mathbb{R}^n} \quad f(x) = \sum_{i=1}^m r_i^2(x)$$

其中 
$$r_i(x) = (a_i^{\top} x)^2 - b_i^2, i = 1, 2, \cdots, m$$

■ 计算梯度和的海瑟矩阵

$$\nabla f(x) = 2\sum_{i=1}^{m} r_i(x) \nabla r_i(x) = 4\sum_{i=1}^{m} ((a_i^{\top} x)^2 - b_i^2) (a_i^{\top} x) a_i$$
$$\nabla^2 f(x) = \sum_{i=1}^{m} (12(a_i^{\top} x)^2 - 4b_i^2) a_i a_i^{\top}$$

## 目录

- 4.1 最优化问题解的存在性
- 4.2 无约束可微问题的最优性理论
- 4.3 无约束不可微问题的最优性理论
- 4.4 对偶理论
- 4.5 一般约束优化问题的最优性理论
- 4.6 带约束凸优化问题的最优性理论

# 无约束不可微问题的最优性理论

■ 考虑不可微优化问题

$$\min_{x \in \mathbb{R}^n} \quad f(x) \tag{2}$$

- 假设 f 是适当且凸的函数, 则  $x^*$  为 (2) 的全局极小点当且仅当  $0 \in \partial f(x^*)$ 
  - □ 必要性 因 x\* 为全局极小点. 有

$$f(y) \ge f(x^*) = f(x^*) + 0^{\top} (y - x^*), \quad \forall y \in \mathbb{R}^n$$

$$\implies 0 \in \partial f(x^*)$$

 $\square$  充分性 如果  $0 \in \partial f(x^*)$ , 那么根据次梯度的定义

$$f(y) \ge f(x^*) + 0^{\top} (y - x^*) = f(x^*), \quad \forall y \in \mathbb{R}^n$$

⇒ x\* 为一个全局极小点

## 复合优化问题的一阶必要条件

■ 考虑一般复合优化问题

$$\min_{x \in \mathbb{R}^n} \quad \psi(x) = f(x) + h(x) \tag{3}$$

其中 ƒ 为光滑函数 (可能非凸), h 为凸函数 (可能非光滑)

■ 定理 4.5 令 x\* 为复合优化问题 (3)的一个局部极小点, 那么

$$-\nabla f(x^*) \in \partial h(x^*)$$

■ 由于目标函数可能是整体非凸的,因此一般没有一阶充分条件

# 实例: $\ell_1$ 范数优化问题

■考虑

$$\min_{x \in \mathbb{R}^n} \quad \psi(x) = f(x) + \mu ||x||_1$$

■ ||x||1 不是可微的, 但可以计算其次微分

$$\partial_i ||x||_1 = \begin{cases} \{1\}, & x_i > 0 \\ [-1, 1], & x_i = 0 \\ \{-1\}, & x_i < 0 \end{cases}$$

■ 若  $x^*$  是局部最优解, 则  $-\nabla f(x^*) \in \mu \partial \|x^*\|_1$ , 即

$$\nabla_i f(x^*) = \begin{cases} -\mu, & x_i^* > 0 \\ a \in [-\mu, \mu], & x_i^* = 0 \\ \mu, & x_i^* < 0 \end{cases}$$

# 总结

#### ■ 无约束优化问题及其最优性条件

问题	一阶条件	二阶条件
可微问题	$\nabla f(x^*) = 0$ (必要)	必要/充分
凸问题	$0 \in \partial f(x^*)$ (充要)	_
复合优化问题	$-\nabla f(x^*) \in \partial h(x^*)$ (必要)	_
非凸非光滑	$0 \in \partial f(x^*)$ (必要)	_

#### 目录

- 4.1 最优化问题解的存在性
- 4.2 无约束可微问题的最优性理论
- 4.3 无约束不可微问题的最优性理论
- 4.4 对偶理论
- 4.5 一般约束优化问题的最优性理论
- 4.6 带约束凸优化问题的最优性理论

#### 对偶理论

■ 一般的约束优化问题

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
s.t.  $c_i(x) \le 0, \ i \in \mathcal{I}$ 

$$c_i(x) = 0, \ i \in \mathcal{E}$$

■ 可行域定义为

$$\mathcal{X} = \{x \in \mathbb{R}^n \mid c_i(x) \le 0, i \in \mathcal{I} \perp c_i(x) = 0, i \in \mathcal{E}\}$$

■ 通过将 X 的示性函数加到目标函数中可以得到无约束优化问题, 但是转化后问题的目标函数是<mark>不连续的、不可微的以及不是有限的</mark>

# 拉格朗日函数

■ 拉格朗日函数  $L: \mathbb{R}^n \times \mathbb{R}^m_+ \times \mathbb{R}^p \to \mathbb{R}$ 

$$L(x, \lambda, \nu) = f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \nu_i c_i(x)$$

- ν<sub>i</sub> 为第 i 个等式约束对应的拉格朗日乘子
- 拉格朗日对偶函数  $g: \mathbb{R}^m_+ \times \mathbb{R}^p \to [-\infty, +\infty)$

$$g(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu)$$
  
=  $\inf_{x \in \mathbb{R}^n} (f(x) + \sum_{i \in \mathcal{I}} \lambda_i c_i(x) + \sum_{i \in \mathcal{E}} \nu_i c_i(x))$ 

# 拉格朗日对偶问题

■ 拉格朗日对偶问题

$$\max_{\lambda \geq 0, \nu} g(\lambda, \nu) = \max_{\lambda \geq 0, \nu} \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu)$$

- $lacksymbol{\bullet}$  设  $p^*$  是原始问题的最优解, $q^*$  是对偶问题的最优解
- 弱对偶性  $q^* \leq p^*$ 
  - 👨 对凸问题与非凸问题都成立
  - □ 可导出复杂问题的非平凡下界
- 强对偶性  $q^* = p^*$ 
  - 🗖 (通常) 对凸问题成立
  - □ 称保证凸问题强对偶性成立的条件为约束品性

# 实例: 线性规划问题的对偶

■ 考虑线性规划问题

$$\begin{aligned} \min_{x} & c^{\top} x \\ \text{s.t.} & Ax = b \\ & x \ge 0 \end{aligned}$$

■拉格朗日函数

$$L(x, s, \nu) = c^{\top} x + \nu^{\top} (Ax - b) - s^{\top} x = -b^{\top} \nu + (A^{\top} \nu - s + c)^{\top} x$$

■ 对偶函数

$$g(s,\nu) = \inf_{x} L(x,s,\nu) = \begin{cases} -b^{\top}\nu, & A^{\top}\nu - s + c = 0 \\ -\infty, &$$
其他

# 实例: 线性规划问题的对偶

■ 对偶问题

$$\max_{s,\nu} \quad -b^{\top}\nu \qquad \qquad \max_{s,y} \quad b^{\top}y$$
 s.t. 
$$A^{\top}\nu - s + c = 0 \qquad \stackrel{y = -\nu}{\Leftrightarrow} \qquad \text{s.t.} \quad A^{\top}y + s = c$$
 
$$s \ge 0 \qquad \qquad s \ge 0$$

■ 若保留约束  $x \ge 0$ , 则拉格朗日函数为

$$L(x,y) = c^{\top}x - y^{\top}(Ax - b) = b^{\top}y + (c - A^{\top}y)^{\top}x$$

■ 对偶问题需要将  $x \ge 0$  添加到约束里

$$\max_{y} \left\{ \inf_{x} \ b^{\top} y + (c - A^{\top} y)^{\top} x \quad \text{s.t.} \quad x \ge 0 \right\} \quad \Rightarrow \quad \max_{y} \quad b^{\top} y$$

$$\text{s.t.} \quad A^{\top} y \le c$$

# 实例: 线性规划问题的对偶

■ 将  $\max b^{\top}y$  改写为  $\min -b^{\top}y$ , 对偶问题的拉格朗日函数为

$$L(y,x) = -b^{\mathsf{T}}y + x^{\mathsf{T}}(A^{\mathsf{T}}y - c) = -c^{\mathsf{T}}x + (Ax - b)^{\mathsf{T}}y$$

■ 得到对偶函数

$$g(x) = \inf_{y} L(y, x) = \begin{cases} -c^{\top} x, & Ax = b \\ -\infty, &$$
其他

■ 相应的对偶问题是

$$\max_{x} - c^{\top} x$$
s.t.  $Ax = b$ 

$$x > 0$$

■该问题与原始问题完全等价,表明线性规划问题与其对偶问题互为对偶

# 实例: $\ell_1$ 正则化问题的对偶

■考虑

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2} ||Ax - b||^2 + \mu ||x||_1$$

$$\min_{x,r} \quad \frac{1}{2} ||r||^2 + \mu ||x||_1$$
s.t.  $r = Ax - b$ 

■拉格朗日函数

$$L(x, r, \lambda) = \frac{1}{2} ||r||^2 + \mu ||x||_1 - \langle \lambda, Ax - b - r \rangle$$
  
=  $\frac{1}{2} ||r||^2 + \lambda^\top r + \mu ||x||_1 - (A^\top \lambda)^\top x + b^\top \lambda$ 

# 实例: $\ell_1$ 正则化问题的对偶

■ 对偶函数

$$g(\lambda) = \inf_{x,r} \ L(x,r,\lambda) = \begin{cases} b^{\top}\lambda - \frac{1}{2}\|\lambda\|^2, & \|A^{\top}\lambda\|_{\infty} \le \mu \\ -\infty, &$$
其他

■ 对偶问题

$$\max_{\lambda} \quad b^{\top} \lambda - \frac{1}{2} \|\lambda\|^{2}$$
  
s.t. 
$$\|A^{\top} \lambda\|_{\infty} \le \mu$$

# 实例: 半定规划问题的对偶问题

■ 考虑

$$\min_{X \in \mathcal{S}^n} \langle C, X \rangle 
\text{s.t.} \quad \langle A_i, X \rangle = b_i, \ i = 1, 2, \dots, m 
\quad X \succeq 0$$

■拉格朗日函数

$$L(X, y, S) = \langle C, X \rangle - \sum_{i=1}^{m} y_i (\langle A_i, X \rangle - b_i) - \langle S, X \rangle, \quad S \succeq 0$$

# 实例: 半定规划对偶问题的对偶问题

■ 对偶函数

$$g(y,S) = \inf_{X} L(X,y,S) = \begin{cases} b^{\top}y, & \sum_{i=1}^{m} y_{i}A_{i} - C + S = 0\\ -\infty, &$$
其他

■ 对偶问题

$$\min_{y \in \mathbb{R}^m} -b^{\top} y$$
s.t. 
$$\sum_{i=1}^m y_i A_i - C + S = 0$$

$$S \succeq 0$$

#### 目录

- 4.1 最优化问题解的存在性
- 4.2 无约束可微问题的最优性理论
- 4.3 无约束不可微问题的最优性理论
- 4.4 对偶理论
- 4.5 一般约束优化问题的最优性理论
- 4.6 带约束凸优化问题的最优性理论

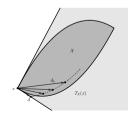
# 切锥

■ 给定可行域  $\mathcal{X}$  及  $x \in \mathcal{X}$ , 若存在序列  $\{z_k\}_{k=1}^{\infty} \subset \mathcal{X}$ ,  $\lim_{k\to\infty} z_k = x$  以及正标量序列  $\{t_k\}_{k=1}^{\infty}$ ,  $t_k \to 0$  满足

$$\lim_{k \to \infty} \frac{z_k - x}{t_k} = d$$

则称向量 d 为  $\mathcal{X}$  在点 x 处的一个切向量

■ 所有点 x 处的切向量构成的集合称为切锥, 用  $T_{\mathcal{X}}(x)$  表示





## 几何最优性条件

■ 一般优化问题

$$\min_{x \in \mathbb{R}^n} f(x)$$
s.t.  $c_i(x) \le 0, i \in \mathcal{I}$ 

$$c_i(x) = 0, i \in \mathcal{E}$$

$$(4)$$

■ 定理 4.6 假设可行点  $x^*$  是问题 (4) 的一个局部极小点. 如果 f(x) 和  $c_i(x)$ ,  $i \in \mathcal{I} \cup \mathcal{E}$  在点  $x^*$  处是可微的, 那么

$$d^{\top} \nabla f(x^*) \ge 0, \quad \forall d \in T_{\mathcal{X}}(x^*)$$

等价于

$$T_{\mathcal{X}}(x^*) \cap \{d \mid \nabla f(x^*)^{\top} d < 0\} = \emptyset$$

#### 线性化可行锥

■ 定义 4.6 对于可行点 $x \in \mathcal{X}$ , 定义积极集  $\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \mid c_i(x) = 0\}$ , 点 x 处的线性化可行方向锥定义为

$$\mathcal{F}(x) = \left\{ d \middle| d^{\top} \nabla c_i(x) = 0, \ \forall \ i \in \mathcal{E} \\ d^{\top} \nabla c_i(x) \le 0, \ \forall \ i \in \mathcal{A}(x) \cap \mathcal{I} \right\}$$

■ 命题 4.1 设  $c_i(x), i \in \mathcal{E} \cup \mathcal{I}$  一阶连续可微, 则对任意可行点 x 有

$$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$$

- 线性可行化方向锥容易计算, 但不能反映可行域 *X* 的本质特征
- 切锥能反映可行域 X 的本质特征, 但不容易计算
- 引入约束品性, 确保  $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ , 从而用  $\mathcal{F}(x)$  取代  $T_{\mathcal{X}}(x)$

#### KKT 条件

■ 定理 4.7 假设  $x^*$  是一般优化问题 (4) 的一个局部最优点

$$\min_{x \in \mathbb{R}^n} \quad f(x)$$
s.t.  $c_i(x) \leq 0, \ i \in \mathcal{I}$  
$$c_i(x) = 0, \ i \in \mathcal{E}$$
如果  $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$  成立, 那么存在拉格朗日乘子  $\lambda_i^*$  使得稳定性条件  $\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \sum_i \lambda_i^* \nabla c_i(x^*) = 0$ 

 $i \in \mathcal{I} \cup \mathcal{E}$ 

原始可行性条件 
$$c_i(x^*)=0, \ \forall i\in\mathcal{E}$$
 原始可行性条件  $c_i(x^*)\leq 0, \ \forall i\in\mathcal{I}$  对偶可行性条件  $\lambda_i^*\geq 0, \ \forall i\in\mathcal{I}$  互补松弛条件  $\lambda_i^*c_i(x^*)=0, \ \forall i\in\mathcal{I}$ 

#### 临界锥

■ 若  $x^*$  是满足 KKT 条件的点, 假设  $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ , 则  $\forall d \in \mathcal{F}(x^*)$  有

$$d^{\top} \nabla f(x^*) = -\sum_{i \in \mathcal{E}} \underbrace{\lambda_i^* d^{\top} \nabla c_i(x^*)}_{= 0} - \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{I}} \underbrace{\lambda_i^* d^{\top} \nabla c_i(x^*)}_{\leq 0} \ge 0$$

■ 定义 4.10 设 $(x^*, \lambda^*)$  是满足 KKT 条件的 KKT 对, 定义临界锥为

$$\mathcal{C}(x^*, \lambda^*) = \{ d \in \mathcal{F}(x^*) \mid d^\top \nabla c_i(x^*) = 0, \ \forall i \in \mathcal{A}(x^*) \cap \mathcal{I} \ \stackrel{\text{def}}{=} \lambda_i^* > 0 \}$$

 $\blacksquare$  当  $d \in \mathcal{C}(x^*, \lambda^*)$  时,  $\forall i \in \mathcal{E} \cup \mathcal{I}$  有  $\lambda_i^* \nabla c_i(x^*)^\top d = 0$ , 故

$$d^{\top} \nabla f(x^*) = \sum_{i \in \mathcal{E} \cup \mathcal{I}} \lambda_i^* d^{\top} \nabla c_i(x^*) = 0$$

#### 二阶最优性条件

■ 定理 4.8 (二阶必要条件) 假设  $x^*$  是一个局部最优解, 且  $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ . 令  $\lambda^*$  为相应的拉格朗日乘子, 即  $(x^*, \lambda^*)$  满足 KKT 条件, 那么

$$d^{\top} \nabla^2_{xx} L(x^*, \lambda^*) d \ge 0, \quad \forall d \in \mathcal{C}(x^*, \lambda^*)$$

■ 定理 4.9 (二阶充分条件) 假设在可行点  $x^*$  处, 存在一个拉格朗日乘子  $\lambda^*$ , 使得  $(x^*, \lambda^*)$  满足 KKT 条件. 如果

$$d^{\top} \nabla^2_{xx} L(x^*, \lambda^*) d > 0, \quad \forall d \in \mathcal{C}(x^*, \lambda^*), \ d \neq 0$$

那么 x\* 为一个严格局部极小解

■ 回顾无约束优化问题的二阶最优性条件

# 例子

■考虑

min 
$$x_1^2 + x_2^2$$
 s.t.  $\frac{x_1^2}{4} + x_2^2 - 1 = 0$ 

■ 拉格朗日函数为

$$L(x,\lambda) = x_1^2 + x_2^2 + \lambda(\frac{x_1^2}{4} + x_2^2 - 1)$$

■ 该问题可行域在任意一点  $x = (x_1, x_2)^{\mathsf{T}}$  处的线性化可行方向锥为

$$\mathcal{F}(x) = \{ (d_1, d_2) \mid \frac{x_1}{4} d_1 + x_2 d_2 = 0 \}$$

■ 根据  $C(x,\lambda) = F(x)$ , 计算出 4 个 KKT 对

$$(x^{\mathsf{T}}, \lambda) = (2, 0, -4), (-2, 0, -4), (0, 1, -1), (0, -1, -1)$$

# 例子

■ 第一个 KKT 对 y = (2, 0, -4), 计算可得

$$\nabla^2_{xx}L(y) = \begin{bmatrix} 0 & 0 \\ 0 & -6 \end{bmatrix}, \quad \mathcal{C}(y) = \{(d_1, d_2) \mid d_1 = 0\}$$

取 d = (0,1), 则  $d^{\top} \nabla^2_{xx} L(y) d = -6 < 0$ , 因此 y 不是局部最优点

■ 第三个 KKT 对 z = (0, 1, -1), 计算可得

$$\nabla^2_{xx}L(z) = \begin{bmatrix} \frac{3}{2} & 0\\ 0 & 0 \end{bmatrix}, \quad \mathcal{C}(z) = \{(d_1, d_2) \mid d_2 = 0\}$$

对于任意的  $d=(d_1,0)$  且  $d_1\neq 0$ , 有  $d^{\top}\nabla^2_{xx}L(z)d=\frac{3}{2}d_1^2>0$ , 因此 z 是一个严格局部最优点

#### 目录

- 4.1 最优化问题解的存在性
- 4.2 无约束可微问题的最优性理论
- 4.3 无约束不可微问题的最优性理论
- 4.4 对偶理论
- 4.5 一般约束优化问题的最优性理论
- 4.6 带约束凸优化问题的最优性理论

#### 带约束凸优化问题

■ 考虑带约束的凸优化问题

$$\min_{x \in \mathcal{D}} f(x)$$
s.t.  $c_i(x) \le 0$ ,  $i = 1, 2, \dots, m$ 

$$Ax = b$$
(5)

- □ f(x) 为适当的凸函数
- $\square$  集合  $\mathcal D$  表示自变量 x 的定义域, 即  $\mathcal D=\{x\mid f(x)<+\infty\}$

#### Slater 约束品性与强对偶原理

■ 定义 4.11 集合 D 的相对内点集定义为

relint 
$$\mathcal{D} = \{x \in \mathcal{D} \mid \exists r > 0, \ \mathbf{使} \in \mathcal{B}(x,r) \cap \text{affine } \mathcal{D} \subseteq \mathcal{D}\}$$

■ 定义 4.12 若对凸优化问题 (5) 存在  $x \in \text{relint } \mathcal{D}$  满足

$$c_i(x) < 0, \quad i = 1, 2, \dots, m, \quad Ax = b$$

则称满足 Slater 约束条件

■ 定理 4.10 若凸优化问题满足 Slater 条件, 则强对偶原理成立

#### 一阶充要条件

■ 定理 4.11 对于凸优化问题 (5), 如果 Slater 条件成立, 那么  $x^*$ ,  $\lambda^*$  分别是原始、对偶全局最优解当且仅当

稳定性条件 
$$0 \in \partial f(x^*) + \sum_{i \in \mathcal{I}} \lambda_i^* \partial c_i(x^*) + \sum_{i \in \mathcal{E}} \lambda_i^* a_i$$
 原始可行性条件  $Ax^* = b, \ \forall i \in \mathcal{E}$  原始可行性条件  $c_i(x^*) \leq 0, \ \forall i \in \mathcal{I}$  对偶可行性条件  $\lambda_i^* \geq 0, \ \forall i \in \mathcal{I}$  互补松弛条件  $\lambda_i^* c_i(x^*) = 0, \ \forall i \in \mathcal{I}$ 

# 实例: 仿射空间的投影问题

■考虑

$$\min_{x \in \mathbb{R}^n} \quad \frac{1}{2} ||x - y||_2^2 \qquad \text{s.t.} \quad Ax = b$$

- 拉格朗日函数  $L(x,\lambda) = \frac{1}{2} ||x-y||^2 + \lambda^\top (Ax-b)$
- KKT 条件

$$\begin{cases} x^* - y + A^{\top} \lambda^* = 0 \\ Ax^* = b \end{cases}$$

 $\blacksquare$  第一式左右两边同时左乘 A 可得

$$Ax^* - Ay + AA^{\mathsf{T}}\lambda = 0 \quad \Rightarrow \quad \lambda^* = (AA^{\mathsf{T}})^{-1}(Ay - b)$$

■ 将 λ\* 代回第一式可知

$$x^* = y - A^{\top} (AA^{\top})^{-1} (Ay - b)$$

## 总结

#### ■ 无约束优化问题及其最优性条件

问题	一阶条件	二阶条件
可微问题	$\nabla f(x^*) = 0$ (必要)	必要/充分
凸问题	$0 \in \partial f(x^*)$ (充要)	_
复合优化问题	$-\nabla f(x^*) \in \partial h(x^*)$ (必要)	_
非凸非光滑	$0 \in \partial f(x^*)$ (必要)	_

#### ■ 约束优化问题的最优性条件和相应约束品性

问题	一阶条件	二阶条件	约束品性
一般问题	KKT 条件 (必要)	必要/充分	LICQ
凸问题	KKT 条件 (充要)	_	Slater

# Q&A

# Thank you!

感谢您的聆听和反馈