

# 第四章 最优性理论

修贤超

<https://xianchaoxiu.github.io>

- 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
- 4.7 约束优化最优性理论应用实例

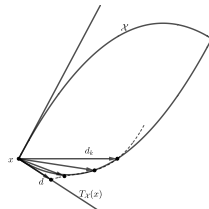
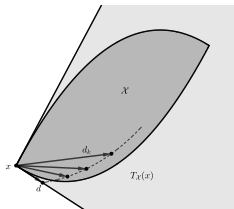
# 切锥

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

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

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

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \leq 0, \quad i \in \mathcal{I} \\ & c_i(x) = 0, \quad i \in \mathcal{E} \end{aligned}$$

- **定理 4.6** 假设可行点  $x^*$  是上述问题的一个局部极小点. 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

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

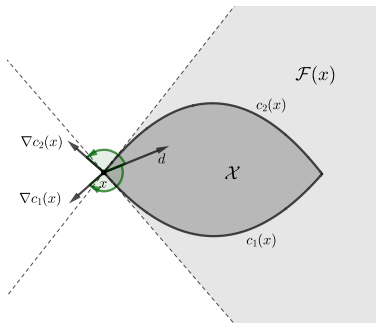
等价于

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

# 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

$$\mathcal{F}(x) = \left\{ d \mid \begin{array}{l} d^\top \nabla c_i(x) = 0, \forall i \in \mathcal{E} \\ d^\top \nabla c_i(x) \leq 0, \forall i \in \mathcal{A}(x) \cap \mathcal{I} \end{array} \right\}$$



# 线性化可行锥包含切锥

- **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

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

- 反之, 切锥未必包含线性化可行锥

$$\begin{aligned} \min_{x \in \mathbb{R}} \quad & f(x) = x \\ \text{s.t.} \quad & c(x) = -x + 3 \leq 0 \end{aligned}$$

- 则  $T_{\mathcal{X}}(3) = \{d \mid d \geq 0\}, \mathcal{F}(3) = \{d \mid d \geq 0\}$ , 于是  $T_{\mathcal{X}}(3) = \mathcal{F}(3)$

- 将问题的约束变形为

$$c(x) = (-x + 3)^3 \leq 0$$

因为可行域不变, 故点  $x^* = 3$  处, 切锥  $T_{\mathcal{X}}(x^*) = \{d \mid d \geq 0\}$  不变. 由  $c'(x^*) = -3(x^* - 3)^2 = 0$  知线性化可行锥  $\mathcal{F}(x^*) = \{d \mid d \in \mathbb{R}\}$

- 因此  $\mathcal{F}(x^*) \supset T_{\mathcal{X}}(x^*)$  (严格包含)

# 约束品性的引入

- 线性化可行方向锥  $\mathcal{F}(x)$  受可行域  $\mathcal{X}$  代数表示方式的影响
- 切锥  $T_{\mathcal{X}}(x)$  仅由可行域  $\mathcal{X}$  决定
- 线性可行化方向锥容易计算, 但不能反映可行域  $\mathcal{X}$  的本质特征
- 切锥能反映可行域  $\mathcal{X}$  的本质特征, 但不容易计算
- 引入约束品性来沟通两者, 确保最优点  $x^*$  处  $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ , 从而可以用  $\mathcal{F}(x)$  取代  $T_{\mathcal{X}}(x)$

- **定义 4.7** 给定可行点  $x$  及相应的积极集  $\mathcal{A}(x)$ . 如果积极集对应的约束函数的梯度, 即  $\nabla c_i(x)$ ,  $i \in \mathcal{A}(x)$  是线性无关的, 则称线性无关约束品性 (LICQ) 在点  $x$  处成立

- **定义 4.8** 给定可行点  $x$  及积极集  $\mathcal{A}(x)$ . 如果存在一个向量  $w \in \mathbb{R}^n$ , 使得

$$\begin{aligned}\nabla c_i(x)^\top w &< 0, \quad \forall i \in \mathcal{A}(x) \cap \mathcal{I} \\ \nabla c_i(x)^\top w &= 0, \quad \forall i \in \mathcal{E}\end{aligned}$$

并且等式约束对应的梯度集  $\{\nabla c_i(x), i \in \mathcal{E}\}$  是线性无关的, 则称点  $x$  处 Mangasarian-Fromovitz 约束品性 (MFCQ) 成立

- **定义 4.9** 若所有的约束函数  $c_i(x)$ ,  $i \in \mathcal{I} \cup \mathcal{E}$  都是线性的, 则称线性约束品性成立



# Karush-Kuhn-Tucker(KKT) 条件: 引入

## ■ 回顾几何最优性条件

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

## ■ $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ 时 (约束品性成立), 上述条件变为

$$\left\{ d \left| \begin{array}{l} d^\top \nabla f(x^*) < 0, \\ d^\top \nabla c_i(x^*) = 0, \ i \in \mathcal{E}, \\ d^\top \nabla c_i(x^*) \leq 0, \ i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{array} \right. \right\} = \emptyset$$

## ■ 上式依然难以验证, 但可使用 Farkas 引理进行化简

- **引理 4.3** 设  $p$  和  $q$  为两个非负整数, 给定  $\mathbb{R}^n$  中的向量  $\{a_i\}_{i=1}^p, \{b_i\}_{i=1}^q$  和  $c$ , 则满足

$$d^\top a_i = 0, \quad i = 1, 2, \dots, p$$

$$d^\top b_i \geq 0, \quad i = 1, 2, \dots, q$$

$$d^\top c < 0$$

的  $d$  不存在当且仅当存在  $\{\lambda_i\}_{i=1}^p$  和  $\mu_i \geq 0, i = 1, 2, \dots, q$ , 使得

$$c = \sum_{i=1}^p \lambda_i a_i + \sum_{i=1}^q \mu_i b_i$$

# 从 Farkas 引理到 KKT 条件

- 由 Farkas 引理, 取  $a_i = \nabla c_i(x^*), i \in \mathcal{E}, b_i = \nabla c_i(x^*), i \in \mathcal{A}(x^*) \cap \mathcal{I}$  以及  $c = -\nabla f(x^*)$ , 则  $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$  时几何最优性条件等价于

$$-\nabla f(x^*) = \sum_{i \in \mathcal{E}} \lambda_i^* \nabla c_i(x^*) + \sum_{i \in \mathcal{A}(x^*) \cap \mathcal{I}} \lambda_i^* \nabla c_i(x^*)$$

- 如果补充定义  $\lambda_i^* = 0, i \in \mathcal{I} \setminus \mathcal{A}(x^*)$ , 那么

$$-\nabla f(x^*) = \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i^* \nabla c_i(x^*) \quad \Rightarrow \quad \text{一阶最优性条件}$$

- 对于任意的  $i \in \mathcal{I}$ , 有

$$\lambda_i^* c_i(x^*) = 0 \quad \Rightarrow \quad \text{互补松弛条件}$$

- $x^*$  称为 KKT 点,  $(x^*, \lambda^*)$  称为 KKT 对

■ **定理 4.7** 假设  $x^*$  是一般优化问题

$$\begin{aligned} \min_{x \in \mathbb{R}^n} \quad & f(x) \\ \text{s.t.} \quad & c_i(x) \leq 0, \quad i \in \mathcal{I} \\ & c_i(x) = 0, \quad i \in \mathcal{E} \end{aligned}$$

的一个局部最优点. 如果  $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$  成立, 那么存在拉格朗日乘子  $\lambda_i^*$  使得如下条件成立

稳定性条件  $\nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i \in \mathcal{I} \cup \mathcal{E}} \lambda_i^* \nabla c_i(x^*) = 0$

原始可行性条件  $c_i(x^*) = 0, \quad \forall i \in \mathcal{E}$

原始可行性条件  $c_i(x^*) \leq 0, \quad \forall i \in \mathcal{I}$

对偶可行性条件  $\lambda_i^* \geq 0, \quad \forall i \in \mathcal{I}$

互补松弛条件  $\lambda_i^* c_i(x^*) = 0, \quad \forall i \in \mathcal{I}$

## 二阶最优性条件: 引入

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

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

- 一阶条件无法判断  $x^*$  是否是最优值点
- 若  $d^\top \nabla f(x^*) = 0$ , 则需用二阶信息来进一步判断可行域内的目标函数值

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

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

其中  $\mathcal{F}(x^*)$  为点  $x^*$  处的线性化可行方向锥

- 临界锥是线性化可行方向锥  $\mathcal{F}(x^*)$  的子集
- 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** 假设  $x^*$  是问题的一个局部最优解, 并且  $T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$  成立. 令  $\lambda^*$  为相应的拉格朗日乘子, 即  $(x^*, \lambda^*)$  满足 KKT 条件, 那么

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

- **定理 4.9** 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

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

那么  $x^*$  为问题的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

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

## ■ 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 $x = (x_1, x_2)^\top$ 处的线性化可行方向锥为

$$\mathcal{F}(x) = \{(d_1, d_2) \mid \frac{x_1}{4}d_1 + x_2d_2 = 0\}$$

因为只有一个等式约束且其对应函数的梯度非零, 故有 LICQ 成立, 于是  $\mathcal{F}(x) = T_{\mathcal{X}}(x)$ . 若  $(x, \lambda)$  为 KKT 对, 由于无不等式约束, 故  $\mathcal{C}(x, \lambda) = (x)$



# 例子

- 可以计算出其 4 个 KKT 对

$$(x^\top, \lambda) = (2, 0, -4), \quad (-2, 0, -4), \quad (0, 1, -1) \quad \text{和} \quad (0, -1, -1)$$

- 第一个 KKT 对  $y = (2, 0, -4)$ , 计算可得

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

取  $d = (0, 1)$ , 则

$$d^\top \nabla_{xx}^2 L(y) d = -6 < 0$$

因此  $y$  不是局部最优点

- 类似地对第三个 KKT 对  $z = (0, 1, -1)$ , 计算可得

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

对于任意的  $d = (d_1, 0)$  且  $d_1 \neq 0$ , 有

$$d^\top \nabla_{xx}^2 L(z) d = \frac{3}{2} d_1^2 > 0$$

因此  $z$  为一个严格局部最优点

- 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
- 4.7 约束优化最优性理论应用实例

# 带约束凸优化问题

- 前述问题都可以写为

$$\begin{array}{ll}\min_{x \in \mathcal{D}} & f(x) \\ \text{s.t.} & c_i(x) \leq 0, \quad i = 1, 2, \dots, m \\ & Ax = b\end{array}$$

- $A \in \mathbb{R}^{p \times n}, b \in \mathbb{R}^p$  是已知的
- $f(x)$  为适当的凸函数,  $c_i(x)$  是凸函数且  $\text{dom } c_i = \mathbb{R}^n$
- 集合  $\mathcal{D}$  表示自变量  $x$  的自然定义域, 即

$$\mathcal{D} = \{x \mid f(x) < +\infty\}.$$

- 自变量  $x$  还受约束的限制, 定义可行域

$$\mathcal{X} = \{x \in \mathcal{D} \mid c_i(x) \leq 0, i = 1, 2, \dots, m, Ax = b\}$$

- 给定集合  $\mathcal{D}$ , 记其仿射包为

$$\text{affine}\mathcal{D} = \{x \mid x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k, x_1, x_2, \cdots, x_k \in \mathcal{D}, \sum_{i=1}^k \theta_i = 1\}$$

- **定义 4.11** 集合  $\mathcal{D}$  的相对内点集定义为

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

- 相对内点是内点的推广

■ **定义 4.12** 若对凸优化问题

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

存在  $x \in \text{relint}\mathcal{D}$  满足

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

则称对此问题 Slater 约束品性满足. 该约束品性也称为 Slater 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, 用  $a_i$  表示矩阵  $A^\top$  的第  $i$  列,  $\partial f, \partial c_i$  表示次梯度, 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

$$\text{稳定性条件} \quad 0 \in \partial f(x^*) + \sum_{i \in \mathcal{I}} \lambda_i^* \partial c_i(x^*) + \sum_{i \in \mathcal{E}} \lambda_i^* a_i$$

$$\text{原始可行性条件} \quad Ax^* = b, \forall i \in \mathcal{E}$$

$$\text{原始可行性条件} \quad c_i(x^*) \leq 0, \forall i \in \mathcal{I}$$

$$\text{对偶可行性条件} \quad \lambda_i^* \geq 0, \forall i \in \mathcal{I}$$

$$\text{互补松弛条件} \quad \lambda_i^* c_i(x^*) = 0, \forall i \in \mathcal{I}$$

## 关于充分性的评述

- 对于一般的约束优化问题, 当问题满足特定约束品性时, 我们知道 KKT 条件是局部最优解处的必要条件
- 对于凸优化问题, 当 Slater 条件满足时, KKT 条件则变为局部最优解的充要条件 (根据凸性, 局部最优解也是全局最优解) 定理的充分性说明, 若能直接求解出凸优化问题的 KKT 对, 则其就是对应问题的最优解.
- Slater 条件的意义在于当问题最优解存在时, 其相应 KKT 条件也会得到满足



- 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
- 4.7 约束优化最优性理论应用实例

## 实例：仿射空间的投影问题

- 考虑

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

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

- Slater 条件成立,  $x^*$  为一个全局最优解当且仅当存在  $\lambda^* \in \mathbb{R}^m$  使得

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

## 实例：仿射空间的投影问题

- 由上述 KKT 条件第一式, 等号左右两边同时左乘  $A$  可得

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

- 将  $\lambda^*$  代回 KKT 条件第一式可知

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

因此点  $y$  到集合  $\{x \mid Ax = b\}$  的投影为  $y - A^T(AA^T)^{-1}(Ay - b)$

## ■ 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