机器学习中的优化算法

Lecture03: 最优性理论

张立平

清华大学数学科学系

办公室：理科楼#A302, Tel: 62798531

E-mail: lipingzhang@tsinghua.edu.cn

Contents and Acknowledgement

- 教材：**最优化：建模、算法与理论**

  http://bicmr.pku.edu.cn/ wenzw/bigdata2021.html

- 致谢：感谢北京大学**文再文**教授提供参考课件

Outline of Lecture03

- 最优化问题解的存在性
- 可微无约束优化的最优性理论
- 不可微无约束优化的最优性理论
- 对偶理论
- 凸优化的最优性理论
- 约束优化的最优性理论

最优化问题解的存在性

考虑优化问题

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

- Weierstrass 定理: 定义在紧集上的连续函数一定存在最大(最小) 值点.
- 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, 因此需要将此定理推广来保证最优化问题解的存在性.

最优化问题解的存在性: Weierstrass 定理

Theorem 1 (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:*

① $\mathbf{dom} f \stackrel{\text{def}}{=\!=} \{x \in \mathcal{X} \,:\, f(x) < +\infty\}$ *有界,*

② *存在一个常数$\bar{\gamma}$ 使得下水平集*

$$C_{\bar{\gamma}} \stackrel{\text{def}}{=\!=} \{x \in \mathcal{X} \,:\, f(x) \leq \bar{\gamma}\}$$

非空有界,

③ 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$\{x \in \mathcal{X} \mid f(x) \leq f(y),\ \forall y \in \mathcal{X}\}$ 非空紧.

Remarks for Weierstrass Theorem

- Weierstrass定理的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.
- 我们可以仅在一个有界的下水平集中考虑$f(x)$的最小值.
- 定理仅要求$f(x)$为适当且闭的函数, 并不需要$f(x)$的连续性.
- 当定义域不是有界闭集时, 对于强制函数, 其全局最优解一定存在; e.g., $f(x) = x^2,\ x \in \mathcal{X} = \mathbb{R}$,
- 对于适当且闭的函数$f(x) = e^{-x},\ x \in \mathcal{X} = \mathbb{R}$, 它不满足三个条件中任意一个, 因此我们不能断言其全局极小值点存在. 事实上, 其全局极小值点不存在.

Proof. 设条件②成立, 则

- 令$t \stackrel{\text{def}}{=\!=} \inf_{x\in\mathcal{X}} f(x)$, 则$t$是有限值.

  用反证法, 假设$t=-\infty$.

  ① 由下确界的定义, 存在点列$\{x^k\}\subset C_{\bar\gamma}$, 使得$\lim\limits_{k\to\infty} f(x^k)=-\infty$.

  ② 由$C_{\bar\gamma}$有界知, 点列$\{x^k\}$存在聚点x^*.

  ③ 由f是闭函数知$\mathbf{epi}f$为闭集, 故$(x^*,t)\in\mathbf{epi}f$. 由上方图的定义知$f(x^*)\le t=-\infty$, 这与f适当矛盾. 因此t有限.

- 由f为闭函数知$C_{\bar\gamma}$为闭集, 由假设知$C_{\bar\gamma}$有界, 故为紧集.

- 由$f(x^*)=t$知最小值点集非空, 且为紧集$C_{\bar\gamma}$的子集, 而紧集的闭子集也是紧集, 故最小值点集为非空紧集.

设条件①成立, 即$\mathbf{dom} f$ 有界, 我们来证明条件②成立.

- 由f 适当知存在$x_0 \in \mathcal{X}$ 使得$f(x_0) < +\infty$.
- 记$\bar{\gamma} = f(x_0)$, 由$\mathbf{dom} f$ 有界知f 的$\bar{\gamma}$ 下水平集$C_{\bar{\gamma}}$ 非空有界, 从而条件②成立.

设条件③成立, 我们来证明f 的任意下水平集都有界, 从而条件②成立.

- 用反证法. 假设存在一个无界的下水平集$C_{\bar{\gamma}}$, 那么可以取点列$\{x^k\} \subset C_{\bar{\gamma}}$ 使得$\lim\limits_{k\to\infty} \|x^k\| = +\infty$. 由③知$\lim\limits_{k\to\infty} f(x^k) = +\infty$, 这与$f(x^k) \le \bar{\gamma}$ 矛盾, 故f 的任意下水平集都有界, 从而条件②成立.

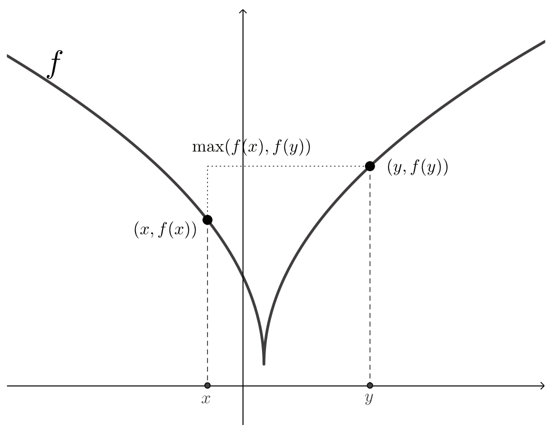
$\square$

解的存在唯一性: 强拟凸函数

Definition 1 (强拟凸函数). *给定凸集$\mathcal{X}$ 和函数$f : \mathcal{X} \to (-\infty, +\infty]$.若任取$x \neq y$ 和$\lambda \in (0, 1)$, 有*

$$f(\lambda x + (1 - \lambda)y) < \max\{f(x), f(y)\},$$

则称函数f 是强拟凸的.



注： 强拟凸函数不一定是凸函数, 但其任意一个下水平集都是凸集; 严格单调函数都是强拟凸函数.

解存在唯一性定理

Theorem 2 (唯一性定理). *设$\mathcal{X}$ 是$\mathbb{R}^n$ 的一个非空, 紧且凸的子集, 如果$f:\mathcal{X}\to(-\infty,+\infty]$ 是适当, 闭且强拟凸函数, 那么存在唯一的x^*满足*

$$f(x^*) < f(x), \quad \forall x \in \mathcal{X}\backslash\{x^*\}.$$

Proof. 由Weierstrass定理知f 至少存在一个全局极小点x^*. 若x^*, y^* 皆为全局极小点, 则

$$f(\lambda x^* + (1-\lambda)y^*) < \max\{f(x^*),\ f(y^*)\} = f(x^*), \quad \forall \lambda \in (0,1).$$

这与x^* 的全局极小性矛盾. □

- 任意强凸函数均为强拟凸的, 但凸函数并不一定是强拟凸的.
- 任何定义在有界凸集上的闭强凸函数, 其最优解都是唯一存在的.
- 对于一般的凸函数, 其最优解可能不唯一; e.g., $f(x)=\max\{x,0\}$, 任意$x\le 0$ 都是$f(x)$ 的最优解.

最优解唯一性的重要意义

- 当最优解是唯一存在时, 我们可以通过比较不同算法收敛至该解的速度来判断算法好坏.
- 如果问题有多个最优值点, 不同的算法收敛到的最优值点可能不同, 那么这些算法收敛速度的比较就失去了参考价值.
- 最优化问题解的唯一性在理论分析和算法比较中扮演着重要角色.

可微无约束优化的最优性条件

考虑无约束优化问题:

$$(\mathbf{UP}) \quad \min_{x \in \mathbb{R}^n} \quad f(x).$$

Definition 2 (下降方向)**.** *对于可微函数f 和点$x \in \mathbb{R}^n$, 如果存在向量d 满足*

$$\nabla f(x)^{\mathrm{T}} d < 0,$$

则称d 为f 在点x 处的一个下降方向.

- 由下降方向的定义, 如果f 在点x 处存在一个下降方向d, 那么对于任意的$T > 0$, 存在$t \in (0, T]$, 使得$f(x + td) < f(x)$.
- 一阶必要条件: 在局部最优解处不能有下降方向.

一阶最优性条件和二阶最优性条件

Theorem 3 (一阶必要条件). *设f 是连续可微函数, 如果x^* 是*(UP)*一个局部最优解, 则*

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

注: 若f凸, 则此条件是充要的.

Theorem 4 (二阶最优性条件). *设f 在$\mathbb{R}^n$ 上二阶可微.*

- *必要条件: 若x^* 是*(UP)*一个局部最优解, 则*

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

- *充分条件: 若$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$, 则$x^*$ 是*(UP) *一个局部最优解.*

注: 对于解的全局最优性判断, 我们还需要借助实际问题的性质, 比如目标函数是凸的、非线性最小二乘问题中目标函数值为0 等.

不可微无约束优化的最优性条件

问题: 若f不可微, (UP)的最优性条件？

Theorem 5 (凸优化问题一阶充要条件). *设f 是适当凸函数, 则x^* 是(UP)的全局最优解当且仅当$0 \in \partial f(x^*)$.*

Proof. **必要性**: 设x^* 是(UP)的全局最优解, 则

$$f(y) \geq f(x^*) = f(x^*) + 0^{\mathrm{T}}(y - x^*), \quad \forall y \in \mathbb{R}^n \Rightarrow 0 \in \partial f(x^*).$$

充分性: 设$0 \in \partial f(x^*)$, 则由次梯度的定义知,

$$f(y) \geq f(x^*) + 0^{\mathrm{T}}(y - x^*) = f(x^*), \quad \forall y \in \mathbb{R}^n.$$

故x^* 是(UP)的全局最优解. □

复合优化问题

- 在实际问题中, 目标函数不一定是凸函数, 但它可以写成一个光滑函数与一个非光滑凸函数的和, 其中目标函数的光滑项可能是凸的, 比如LASSO 问题、图像去噪问题和盲反卷积问题; 也可能是非凸的, 例如神经网络的损失函数. 因此研究此类问题的最优性条件十分必要.

- 复合优化(Compositive Optimization)问题:

$$(\mathbf{COP}) \quad \min_{x\in\mathbb{R}^n} \quad \psi(x) \stackrel{\text{def}}{=\!=} f(x) + h(x),$$

其中f 为光滑函数(可能非凸), h 为凸函数(可能非光滑).

复合优化问题的一阶必要条件

Theorem 6 (复合优化问题的一阶必要条件). *设x^* 是*(COP)*一个局部最优解, 则*

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

注: 定理给出了当目标函数含有非光滑凸函数时的一阶必要条件; 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.

【例】 设$f(x) : \mathbb{R}^n \to \mathbb{R}$ 为光滑函数, $\mu > 0$. 考虑ℓ_1 范数优化问题:

$$(\ell_1) \qquad \min_{x\in\mathbb{R}^n} \quad \psi(x) \xlongequal{\text{def}} f(x) + \mu\|x\|_1.$$

由定理6知, x^* 是(ℓ_1)局部最优解的必要条件:

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

Proof. 用反证法. 若$-\nabla f(x^*) \notin \partial h(x^*)$, 由$h$ 是凸函数知$\partial h(x^*)$ 是有界闭凸集. 根据超平面分离定理知, 存在$d \in \mathbb{R}^n, d \neq 0$ 使得

$$-\nabla f(x^*)^{\mathrm{T}} d > \sup_{\theta \in \partial h(x^*)} \theta^{\mathrm{T}} d.$$

又由$\psi(x) = f(x) + h(x)$ 及方向导数的定义知

$$\begin{aligned}
\psi'(x^*; d) &= \lim_{t \to 0+} \frac{\psi(x^* + td) - \psi(x^*)}{t} \\
&= \nabla f(x^*)^{\mathrm{T}} d + \partial h(x^*; d) \\
&= \nabla f(x^*)^{\mathrm{T}} d + \sup_{\theta \in \partial h(x^*)} \theta^{\mathrm{T}} d < 0.
\end{aligned}$$

故对充分小的非负实数t, 有

$$\psi(x^* + td) < \psi(x^*).$$

这与x^* 是(COP) 的局部最优解矛盾. 因此$-\nabla f(x^*) \in \partial h(x^*)$. □

非光滑非凸无约束优化问题的最优性条件

考虑无约束优化问题(UP), 当函数f 不可微且非凸时, 为讨论最优性理论推广次梯度和次微分的概念.

Definition 3 (次微分). *设$f: \mathbb{R}^n \to (-\infty, +\infty]$ 是适当下半连续函数.*

① *对给定的$x \in \mathrm{dom} f$, 满足如下条件的所有向量$u \in \mathbb{R}^n$ 的集合定义为f 在点x 处的Fréchet 次微分:*

$$\liminf_{y \to x, y \neq x} \frac{f(y) - f(x) - \langle u, y - x \rangle}{\|y - x\|} \geq 0,$$

记为$\hat{\partial} f(x)$.当$x \notin \mathrm{dom} f$ 时, 将$\hat{\partial} f(x)$ 定义为空集$\emptyset$.

② *f 在点$x \in \mathbb{R}^n$ 处的**极限次微分(简称次微分)**定义为*

$$\partial f(x) = \{u \in \mathbb{R}^n \ : \exists\, x^k \to x, f(x^k) \to f(x), u^k \in \hat{\partial} f(x^k) \to u\}.$$

极限次微分通过对x 附近的点处的Fréchet 次微分取极限得到.

Theorem 7 (一阶必要条件). *设f 是适当下半连续函数. 若x^* 是(UP)的局部最优解, 则$0 \in \partial f(x^*)$.*

Remarks for Subdifferential

- $\hat{\partial} f(x) \subseteq \partial f(x)$, 前者是闭凸集, 后者是闭集. 并非在所有的$x \in \mathrm{dom} f$ 处都存在*Fréchet* 次微分.
- 凸函数的次梯度要求不等式

$$f(y) \geq f(x) + \langle g, y - x \rangle, \quad g \in \partial f(x)$$

在定义域内全局成立, 而非凸函数只要求在极限意义下成立.

- 当f 是可微函数时, *Fréchet* 次微分和次微分都退化成梯度.

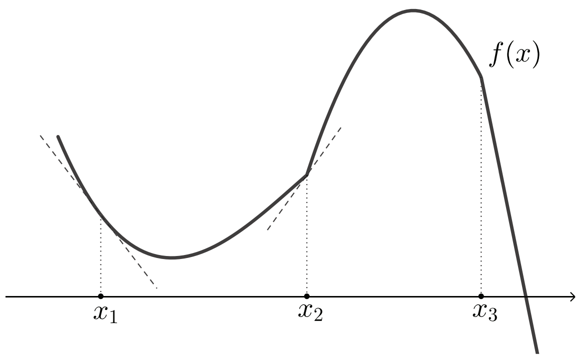


Figure 1: 非凸函数次微分; $f(x)$ 在点x_3 处不存在Fréchet 次微分, 但存在次微分.

对偶理论

考虑约束优化问题:

$$
\text{(MinCon)} \qquad \begin{aligned} \min \quad & f(x) \\ \textbf{s.t.} \quad & c_i(x) \le 0,\ i \in \mathcal{I},\ |\mathcal{I}| = m \\ & c_i(x) = 0,\ i \in \mathcal{E},\ |\mathcal{E}| = p \end{aligned}
$$

其中c_i 为定义在$\mathbb{R}^n$ 或其子集上的实值函数, $\mathcal{I}$ 和$\mathcal{E}$ 分别表示不等式约束和等式约束对应的下标集合.

- **(MinCon)**的最优值为p^*, 可行域定义为

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

- 通过将$\mathcal{X}$ 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不连续的、不可微的以及不是有限的, 这导致我们难以分析其理论性质以及设计有效的算法.

拉格朗日对偶

(MinCon)的拉格朗日函数$L:\mathbb{R}^n\times\mathbb{R}^m_+\times\mathbb{R}^p \rightarrow \mathbb{R}$($\lambda,\nu$–拉格朗日乘子)

$$L(x,\lambda,\nu)=f(x)+\sum_{i\in\mathcal{I}}\lambda_i c_i(x)+\sum_{i\in\mathcal{E}}\nu_i c_i(x). \tag{1}$$

拉格朗日对偶函数$g:\mathbb{R}^m_+\times\mathbb{R}^p \rightarrow [-\infty,+\infty)$

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

Theorem 8 (弱对偶定理). *若$\mathcal{X}\neq\emptyset$ 且$\lambda\geq 0$, 则$g(\lambda,\nu)\leq p^*$.*

Proof. 对$\forall\tilde{x}\in\mathcal{X}$, 有

$$g(\lambda,\nu)=\inf_x L(x,\lambda,\nu)\leq L(\tilde{x},\lambda,\nu)\leq f(\tilde{x}) \quad \Rightarrow g(\lambda,\nu)\leq \inf_{\tilde{x}\in\mathcal{X}} f(\tilde{x})=p^*.$$

□

常见优化问题的拉格朗日对偶

【例1】 线性方程组最小模解问题

$$\min \ x^T x, \quad \textbf{s.t.} \ Ax = b.$$

- 拉格朗日函数:

$$L(x,\nu) = x^T x - \nu^T (Ax - b).$$

- 拉格朗日对偶函数:

$$\begin{aligned} g(\nu) &= \inf_{x\in\mathbb{R}^n} L(x,\nu) = b^T\nu + \inf_{x\in\mathbb{R}^n}\{x^T x - \nu^T Ax\} \\ &= b^T\nu - \frac{1}{4}\nu^T AA^T\nu \end{aligned}$$

  是凹函数.

- 弱对偶性: $p^* \geq b^T\nu - \frac{1}{4}\nu^T AA^T\nu, \ \forall\nu.$

【例2】 二路划分问题：将n 个点划分为两个集合; W_{ij} 是将i 点和j 点划入同一集合的代价; $-W_{ij}$ 则是将i 点和j 点划入不同集合的代价.

$$\begin{array}{ll} \min & x^T W x \\ \textbf{s.t.} & x_i^2 = 1, \quad i = 1, \ldots, n. \end{array}$$

这是一个非凸问题.

- **对偶函数:**

$$\begin{aligned} g(\nu) &= \inf_x \{x^T W x - \sum_i^n \nu_i (x_i^2 - 1)\} = \mathbf{1}^T \nu + \inf_x x^T (W - \mathrm{diag}(\nu)) x \\ &= \begin{cases} \mathbf{1}^T \nu, & W - \mathrm{diag}(\nu) \succeq 0, \\ -\infty, & \text{otherwise.} \end{cases} \end{aligned}$$

- **弱对偶性**: $p^* \geq \mathbf{1}^T \nu, \quad \forall \nu : \ W - \mathrm{diag}(\nu) \succeq 0.$

注: 取$\nu = \lambda_{\min}(W)\mathbf{1}$ 可得具体的下界估计, $p^* \geq n\lambda_{\min}(W)$.

线性约束优化问题的拉格朗日对偶与共轭函数

考虑线性约束优化问题:

$$(\mathbf{LCOP}) \qquad \min \quad f_0(x), \quad \textbf{s.t.} \;\; Ax \geq b, \;\; Cx = d.$$

拉格朗日函数: $L(x, \lambda, \nu) = f_0(x) - \lambda^T(Ax - b) - \nu^T(Cx - d)$.

对偶函数:

$$\begin{aligned} g(\lambda, \nu) &= \inf_{x \in \mathrm{dom}\ f_0} L(x, \lambda, \nu) \\ &= b^T\lambda + d^T\nu + \inf_{x \in \mathrm{dom}\ f_0} \left\{ f_0(x) - (A^T\lambda + C^T\nu)^T x \right\} \\ &= b^T\lambda + d^T\nu - \sup_{x \in \mathrm{dom}\ f_0} \left\{ (A^T\lambda + C^T\nu)^T x - f_0(x) \right\} \\ &= b^T\lambda + d^T\nu - f_0^*(A^T\lambda + C^T\nu). \end{aligned}$$

注: 当f_0 的共轭函数已知时可以简化对偶函数的推导;

函数$f(x)$的共轭函数: $f^*(y) = \sup_{x \in \mathrm{dom}\ f} \{y^T x - f(x)\}$.

【例3】 等式约束范数最小化问题

$$\min \ \|x\|, \quad \textbf{s.t. } Ax = b.$$

- 拉格朗日对偶函数:

$$g(\nu) = \inf_{x\in\mathbb{R}^n} \{\|x\| - \nu^T(Ax-b)\} = \begin{cases} b^T\nu & \|A^T\nu\|_* \le 1 \\ -\infty & \textbf{otherwise} \end{cases}$$

其中$\|v\|_* = \sup_{\|u\|\le 1} u^T v$ 是$\|\cdot\|$ 的对偶范数.

- 弱对偶性: $p^* \ge b^T\nu, \quad \forall \nu : \|A^T\nu\|_* \le 1.$

Proof. 由范数$f(x) = \|x\|$的共轭函数:

$$f^*(y) = \sup_x \{y^T x - \|x\|\} = \begin{cases} 0 & \|y\|_* \le 1 \\ +\infty & \|y\|_* > 1 \end{cases}$$

得$g(\nu) = \inf_x\{\|x\| - \nu^T Ax + b^T\nu\} = b^T\nu - \sup_x\{(A^T\nu)^T x - \|x\|\}.$ □

拉格朗日对偶问题

原问题**(MinCon)**的Lagrange对偶问题:

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

- 称λ 和ν 为对偶变量, 设**(D-MinCon)**的最优值为d^*.
- d^* 为p^* 的最优下界, 称$p^* - d^*$ 为对偶间隙.
- **(D-MinCon)**是一个凸优化问题.
- $\mathbf{dom} g = \{(\lambda,\nu) \mid \lambda \geq 0, g(\lambda,\nu) > -\infty\}$; $(\lambda,\nu) \in \mathbf{dom} g$ 称为对偶可行解.
- 弱对偶性: $d^* \leq p^*$ 对凸问题与非凸问题都成立.
- 强对偶性: $d^* = p^*$ 对一般问题通常不成立; (通常)对凸问题成立, 称保证凸问题强对偶性成立的条件为**约束规范**.

Lagrange对偶问题形式与对偶性

- 同一个问题不同等价形式的Lagrange对偶问题可能差异巨大.
- 当对偶问题难以推导或没有价值时, 可以尝试改写原问题的形式.

常用的改写技巧

- 引入新变量与等式约束: $\min f_0(Ax+b)$

  等价原问题及其Lagrange对偶问题:

$$\begin{array}{ll} \min & f_0(y) \\ \textbf{s.t.} & Ax+b-y=0 \end{array} \qquad\qquad \begin{array}{ll} \max & b^T\nu - f_0^*(\nu) \\ \textbf{s.t.} & A^T\nu = 0 \end{array}$$

对偶函数:
$$\begin{aligned} g(\nu) &= \inf_{x,y}\{f_0(y) - \nu^T y + \nu^T Ax + b^T\nu\} \\ &= b^T\nu + \inf_x (A^T\nu)^T x - \sup_y\{\nu^T y - f_0(y)\} \\ &= \begin{cases} b^T\nu - f_0^*(\nu) & A^T\nu = 0 \\ -\infty & \textbf{otherwise}. \end{cases} \end{aligned}$$

【例】 $\min \|Ax - b\|$的Lagrange对偶问题:

$$\begin{aligned} \max \quad & b^T\nu \\ \textbf{s.t.} \quad & A^T\nu = 0, \quad \|\nu\|_* \le 1. \end{aligned}$$

【例】 LASSO的对偶问题

$$\min_{x\in\mathbb{R}^n} \quad \tfrac{1}{2}\|Ax-b\|_2^2+\mu\|x\|_1 \quad \Leftrightarrow \quad \begin{cases} \displaystyle\min_{x\in\mathbb{R}^n} \frac{1}{2}\|r\|_2^2+\mu\|x\|_1 \\ \text{s.t.} \quad r = Ax-b \end{cases}$$

- 拉格朗日函数:

$$\begin{aligned} L(x,r,\lambda) &= \frac{1}{2}\|r\|^2+\mu\|x\|_1-\langle\lambda, Ax-b-r\rangle \\ &= \frac{1}{2}\|r\|^2+\lambda^{\mathrm{T}}r+\mu\|x\|_1-(A^{\mathrm{T}}\lambda)^{\mathrm{T}}x+b^{\mathrm{T}}\lambda \end{aligned}$$

- 对偶函数:

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

- 对偶问题:

$$\max \quad b^{\mathrm{T}}\lambda-\frac{1}{2}\|\lambda\|^2, \quad \text{s.t.} \quad \|A^{\mathrm{T}}\lambda\|_\infty \le \mu$$

- 将显式约束隐式化或将隐式约束显式化

【例】 带边界约束的线性规划: 原问题与对偶问题

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b, \\ & -\mathbf{1} \le x \le \mathbf{1}. \end{array} \quad \xrightarrow{\text{dual}} \quad \begin{array}{ll} \max & b^T \nu - \mathbf{1}^T \lambda_1 + \mathbf{1}^T \lambda_2 \\ \text{s.t.} & A^T \nu + \lambda_1 + \lambda_2 = c, \\ & \lambda_1 \ge 0, \quad \lambda_2 \le 0. \end{array}$$

通过隐式化边界约束的对偶函数为:

$$\begin{aligned} g(\nu) &= \inf_{-\mathbf{1} \le x \le \mathbf{1}} \{c^T x - \nu^T (Ax - b)\} \\ &= b^T \nu - \sup_{-\mathbf{1} \le x \le \mathbf{1}} (A^T \nu - c)^T x \\ &= b^T \nu - \|A^T \nu - c\|_1 \end{aligned}$$

对偶问题: $\max \quad b^T \nu - \|A^T \nu - c\|_1$

- 约束不能化为$c_i(x) \le 0$ 的形式, 该如何构造拉格朗日对偶函数呢?

适当锥与广义不等式

Definition 4 (适当锥). 称满足如下条件的锥K 为适当锥(proper cone):

① K 是凸锥;

② K 是闭集;

③ K 是实心的(solid), **i.e.**, $\mathbf{int}K \neq \emptyset$;

④ K 是尖的(pointed), **i.e.**, **对任意非零向量**x**, 若**$x \in K$**, 则**$-x \notin K$.

- 适当锥K 可以诱导出广义不等式并定义偏序关系:

$$x \preceq_K y \iff y - x \in K; \quad x \prec_K y \iff y - x \in \mathbf{int}K.$$

- 当$K = \mathbb{R}^n_+$, $x \preceq_K y \Leftrightarrow x \leq y$.

- 当$K = \mathcal{S}^n_+$, $X \preceq_K Y$ 表示$Y - X$ 是半正定矩阵.

对偶锥与拉格朗日乘子

Definition 5 (对偶锥). *令K 为全空间Ω 的子集, 称集合*

$$K^* = \{y \in \Omega \mid \langle x, y\rangle \geq 0,\ \forall x \in K\}$$

为其对偶锥.

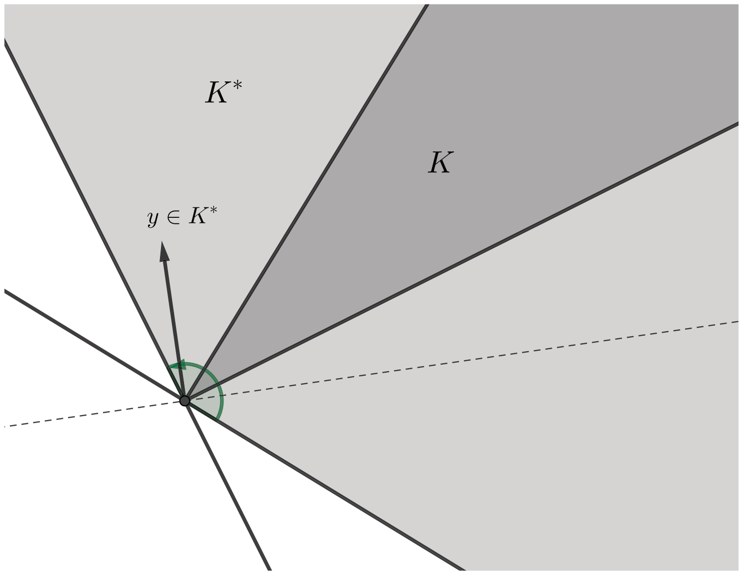


Figure 2: $\mathbb{R}^2$ 平面上的锥K 及其对偶锥K^*

注:

- 设$K = \mathbb{R}^n_+, \Omega = \mathbb{R}^n$ 且定义$\langle x, y\rangle = x^{\mathrm{T}} y$, 则$K^* = \mathbb{R}^n_+$.
- 设$K = \mathcal{S}^n_+, \Omega = \mathcal{S}^n$ 且定义$\langle X, Y\rangle = \mathrm{Tr}(XY^{\mathrm{T}})$, 则$K^* = \mathcal{S}^n_+$

$$\langle X, Y\rangle \geq 0,\ \forall X \in \mathcal{S}^n_+ \quad \Longleftrightarrow \quad Y \in \mathcal{S}^n_+.$$

半正定锥的对偶锥仍为半正定锥.

- 自对偶锥: $K = K^*$. 非负锥和半正定锥都是自对偶锥.
- 对偶锥K^* 中向量和原锥K 中向量的内积恒非负, 这一性质可以被用来构造拉格朗日对偶函数.

广义不等式约束优化问题拉格朗日函数的构造

考虑广义不等式约束优化问题:

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

其中$c_i:\mathbb{R}^n\to\mathbb{R}^{k_i},\ k_i\in\mathbb{N}_+, i\in\mathcal{I}$ 为向量值函数, f 与$c_i, i\in\mathcal{E}$ 为实值函数, $K_i, i\in\mathcal{I}$ 为适当锥.

- 拉格朗日函数:

$$L(x,\lambda,\nu)=f(x)+\sum_{i\in\mathcal{I}}\langle c_i(x),\lambda_i\rangle+\sum_{i\in\mathcal{E}}\nu_i c_i(x),\quad \lambda_i\in K_i^*,\nu_i\in\mathbb{R}.$$

- $L(x,\lambda,\nu)\le f(x),\ \forall\, x\in\mathcal{X},\ \lambda_i\in K_i^*,\nu_i\in\mathbb{R}.$
- 对偶函数$g(\lambda,\nu)=\inf_{x\in\mathbb{R}^n}\ L(x,\lambda,\nu)$, 对偶问题为

$$\max_{\lambda_i\in K_i^*,\ \nu_i\in\mathbb{R}} \quad g(\lambda,\nu).$$

半定规划(SDP)的对偶问题

设$A_i \in \mathcal{S}^n,\ i=1,\ldots,m;\ C \in \mathcal{S}^n, b \in \mathbb{R}^m$, 考虑半定规划(SDP):

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

- 拉格朗日函数:

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

- 对偶函数:

$$g(y, S) = \inf_X\ L(X, y, S) = \begin{cases} b^{\mathrm{T}} y, & \sum_{i=1}^m y_i A_i - C + S = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

- 对偶问题(D-SDP): $\max_{y \in \mathbb{R}^m} \quad b^{\mathrm{T}} y, \quad \text{s.t.} \quad \sum_{i=1}^m y_i A_i - C + S = 0,\ S \succeq 0.$

半定规划对偶问题的对偶问题

(D-SDP): $\min_{y\in\mathbb{R}^m} \quad -b^{\mathrm{T}}y,$ **s.t.** $\sum_{i=1}^m y_i A_i \preceq C$

- 拉格朗日函数:

$$
\begin{aligned}
L(y,X) &= -b^{\mathrm{T}}y + \langle X, \sum_{i=1}^m y_i A_i - C\rangle \\
&= \sum_{i=1}^m y_i(-b_i + \langle A_i, X\rangle) - \langle C, X\rangle
\end{aligned}
$$

- 对偶函数:

$$
g(X) = \inf_y \; L(y,X) = \begin{cases} -\langle C, X\rangle, & \langle A_i, X\rangle = b_i,\; i=1,\ldots,m, \\ -\infty, & \text{otherwise.} \end{cases}
$$

- 对偶问题: $\min_{X\in\mathcal{S}^n} \quad \langle C, X\rangle,$ **s.t.** $\langle A_i, X\rangle = b_i,\; i=1,\ldots,m;\; X\succeq 0.$

约束凸优化问题

设$A \in \mathbb{R}^{p\times n}, b \in \mathbb{R}^p$, $f(x)$ 为适当的凸函数; $\forall i, c_i(x)$ 是凸函数且$\mathrm{dom} c_i = \mathbb{R}^n$. 则约束凸优化问题:

$$(\mathbf{CovP}) \quad \begin{aligned} \min_{x\in\mathcal{D}} \quad & f(x), \\ \text{s.t.} \quad & c_i(x) \leqslant 0, \quad i = 1,\ldots,m, \\ & Ax = b. \end{aligned}$$

- 集合$\mathcal{D}$ 表示目标函数的定义域:

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

- 可行域: $\mathcal{X} = \{x \in \mathcal{D} \,:\, c_i(x) \leq 0,\, i = 1,\ldots,m;\ Ax = b\}$.

- 由于凸优化问题的可行域是凸集, 因此等式约束只能是线性约束.

- 凸优化及其对偶问题的强对偶定理在约束规范下可以成立.

常见的约束凸优化问题

- 稀疏优化问题:

$$\min_{x \in \mathbb{R}^n} \quad \|x\|_1 \qquad \text{s.t.} \quad Ax = b.$$

- 低秩矩阵恢复问题:

$$\min_{X \in \mathbb{R}^{m \times n}} \|X\|_* \qquad \text{s.t.} \quad X_{ij} = M_{ij}, \quad (i,j) \in \Omega.$$

- 矩阵分离问题:

$$\min_{X,S \in \mathbb{R}^{m \times n}} \|X\|_* + \mu \|S\|_1 \qquad \text{s.t.} \quad X + S = M.$$

- 回归分析中的问题:

$$\min_{x \in \mathbb{R}^n} \frac{1}{2} \|Ax - b\|_2^2 \qquad \text{s.t.} \quad \|x\|_1 \leqslant \sigma.$$

强对偶定理与Slater约束规范

- 集合$\mathcal{D}$的仿射包:

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

- 集合$\mathcal{D}$的相对内部:

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

- $\textbf{affine}\mathcal{D} = \mathbb{R}^n$ 时相对内点就是内点.

Definition 6 (Slater约束规范). 若对凸优化问题**(CovP)**存在$x \in \textbf{relint}\mathcal{D}$满足

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

则称凸优化**(CovP)**满足**Slater约束规范**Slater 条件.

注: Slater 条件实际上是要求可行域的相对内点中存在使得不等式约束严格成立的点. 特别地, 线性不等式约束无需要求其存在严格可行点.

Theorem 9. *若凸优化问题***(CovP)***满足Slater条件, 则强对偶定理成立:* $d^* = p^*$.

Theorem 10 (凸优化问题的一阶充要条件). 对凸优化问题**(CovP)**, 用a_i 表示矩阵A^{T} 的第i 列, $\partial f, \partial c_i$ 表示次梯度. 若Slater条件成立, 则x^*, λ^* 分别是原始, 对偶全局最优解当且仅当

$$\text{KKT条件}\begin{cases} \textbf{稳定性条件} & 0 \in \partial f(x^*) + \sum_{i\in\mathcal{I}} \lambda_i^* \partial c_i(x^*) + \sum_{i\in\mathcal{E}} \lambda_i^* a_i, \\ \textbf{原始可行性条件} & Ax^* = b,\ \forall i \in \mathcal{E}, \\ \textbf{原始可行性条件} & c_i(x^*) \le 0,\ \forall i \in \mathcal{I}, \\ \textbf{对偶可行性条件} & \lambda_i^* \ge 0,\ \forall i \in \mathcal{I}, \\ \textbf{互补松弛条件} & \lambda_i^* c_i(x^*) = 0,\ \forall i \in \mathcal{I}. \end{cases}$$

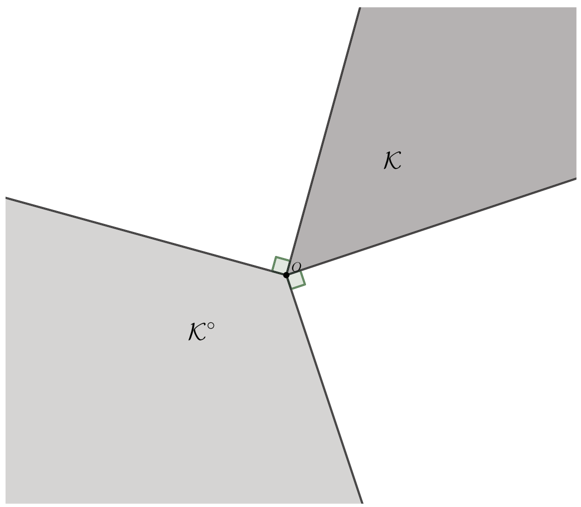
- 凸优化问题的KKT对就是最优解, 不需要Slater条件. 反之不成立.
- Slater条件的意义在于当问题最优解存在时, 其KKT条件也会得到满足.
- 线性约束的凸优化, KKT对等价最优解.

一般形式的凸优化问题最优性条件

设$\mathcal{X}$ 是闭凸集并且$f(x)$ 是凸函数, 考虑一般形式的凸优化问题:

$$\min_{x\in\mathcal{X}} \quad f(x).$$

- 集合$K \subset \mathbb{R}^n$ 的极锥: $K^\circ = \{y \mid \langle x, y\rangle \leq 0, \quad \forall x \in K\}.$
- 集合K的对偶锥: $K^* = \{y \in \Omega \mid \langle x, y\rangle \geq 0, \quad \forall x \in K\}.$
- 极锥的性质: ① 设K 为$\mathbb{R}^n$ 上的凸锥, 则K° 是闭凸锥且$K^\circ = (\bar{K})^\circ$.
  ② 若K 为闭凸锥, 则$K^{\circ\circ} = K$.



法锥

Definition 7 (法锥). 设$\mathcal{X}$ 为凸集, $x \in \mathcal{X}$. 称集合

$$N_{\mathcal{X}}(x) \overset{\text{def}}{=\!=} \{z \mid \langle z, y-x\rangle \leq 0,\ \forall y \in \mathcal{X}\}$$

为$\mathcal{X}$ 在点x 处的法锥.

注: 法锥实际上是将x 平移到原点后的极锥: $N_{\mathcal{X}}(x) = (\mathcal{X} - \{x\})^{\circ}$.

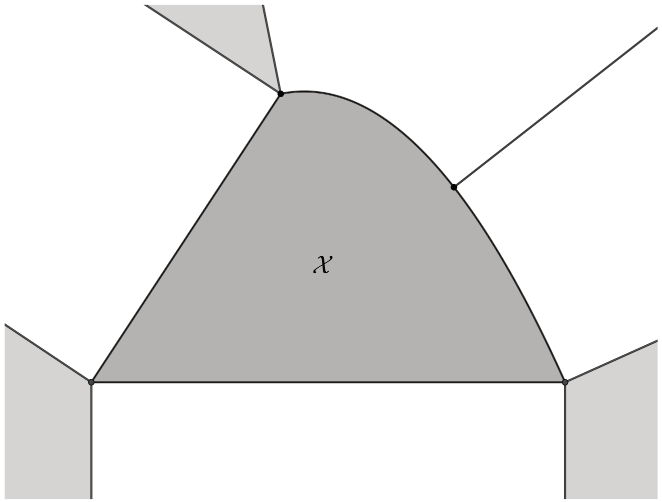


Figure 3: 集合$\mathcal{X}$ 在不同点的法锥; 在$\mathcal{X}$ 光滑边界上, $N_{\mathcal{X}}(x)$ 退化成外法线.

Theorem 11 (法锥与示性函数次微分的等价性). *设$\mathcal{X}$ 是闭凸集, $x \in \mathcal{X}$, 则$N_{\mathcal{X}}(x) = \partial I_{\mathcal{X}}(x)$, 其中$I_{\mathcal{X}}(x)$ 是集合$\mathcal{X}$ 的示性函数.*

Proof. 若$z \in \partial I_{\mathcal{X}}(x)$, 则对任意$y \in \mathrm{dom} I_{\mathcal{X}}$ 有

$$0 = I_{\mathcal{X}}(y) - I_{\mathcal{X}}(x) \geq z^{\mathrm{T}}(y - x),$$

因此$\partial I_{\mathcal{X}}(x) \subseteq N_{\mathcal{X}}(x)$. □

Theorem 12 (Moreau-Rockafellar定理). *设$f_1, f_2 : \mathbb{R}^n \to (-\infty, +\infty]$ 是两个凸函数, 则对任意的$x_0 \in \mathbb{R}^n$,*

$$\partial f_1(x_0) + \partial f_2(x_0) \subseteq \partial(f_1 + f_2)(x_0).$$

若$\mathbf{int\ dom} f_1 \cap \mathbf{dom} f_2 \neq \emptyset$, 则对任意的$x_0 \in \mathbb{R}^n$,

$$\partial(f_1 + f_2)(x_0) = \partial f_1(x_0) + \partial f_2(x_0).$$

Theorem 13. *设Slater条件满足. $x^* \in \mathcal{X}$是凸优化$\min\limits_{x \in \mathcal{X}} f(x)$ 的全局最优解当且仅当存在$s \in \partial f(x^*)$ 使得$-s \in N_{\mathcal{X}}(x^*)$.*

Proof. 充分性: 假设存在$s \in \partial f(x^*)$ 满足$-s \in N_{\mathcal{X}}(x^*)$.由次梯度的定义,

$$f(y) \geq f(x^*) + s^{\mathrm{T}}(y - x^*), \quad \forall y \in \mathcal{X}.$$

由$-s \in N_{\mathcal{X}}(x^*)$知, $\quad s^{\mathrm{T}}(y - x^*) \geq 0, \quad \forall y \in \mathcal{X}.$
于是有$f(y) \geq f(x^*), \quad \forall y \in \mathcal{X}$. 因此, x^* 为$\min\limits_{x \in \mathcal{X}} f(x)$ 的全局最优解.

必要性: 基于凸集$\mathcal{X}$ 的示性函数$I_{\mathcal{X}}(x)$, 凸优化$\min\limits_{x \in \mathcal{X}} f(x)$可转化为无约束优化问题:

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

由凸优化问题一阶充要条件可知, x^* 为其全局最优解 $\Leftrightarrow \ 0 \in \partial(f + I_{\mathcal{X}})(x^*)$.
由于Slater条件成立, 根据Moreau-Rockafellar定理, 有

$$\partial(f + I_{\mathcal{X}})(x^*) = \partial f(x^*) + \partial I_{\mathcal{X}}(x^*).$$

因此, 一定存在一个次梯度$s \in \partial f(x^*)$, 使得

$$-s \in \partial I_{\mathcal{X}}(x^*).$$

又$\partial I_{\mathcal{X}}(x^*) = N_{\mathcal{X}}(x^*)$, 故有$-s \in N_{\mathcal{X}}(x^*)$. $\square$

交集的法锥为各集合对应法锥的和

对凸优化问题**(CovP)**, 其可行域$\mathcal{X}$可写成若干集合的交:

$$\mathcal{X} = \left(\bigcap_{i \in \mathcal{I}} \{x \mid c_i(x) \leq 0\} \right) \cap \{x \mid Ax = b\}.$$

Q: 可行域$\mathcal{X}$的法锥?

Theorem 14 (交集的法锥为各集合对应法锥的和). *给定$\mathbb{R}^n$ 中的一列闭凸集$\mathcal{X}_1, \mathcal{X}_2, \cdots, \mathcal{X}_m$, 记其交集为$\mathcal{X} = \bigcap_{i=1}^m \mathcal{X}_i$. 设$x \in \mathcal{X}$, 若*

$$\mathbf{int}\mathcal{X}_1 \cap \mathbf{int}\mathcal{X}_2 \cap \cdots \cap \mathbf{int}\mathcal{X}_m \neq \emptyset,$$

则

$$N_{\mathcal{X}}(x) = N_{\mathcal{X}_1}(x) + N_{\mathcal{X}_2}(x) + \cdots + N_{\mathcal{X}_m}(x).$$

若某$\mathcal{X}_i$ 是多面体, 则在上式中$\mathbf{int}\mathcal{X}_i$ 可减弱为$\mathcal{X}_i$.

$\{x \mid c_i(x) \le 0\}$ 法锥的具体形式

Theorem 15. *设$c(x)$ 是适当的闭凸函数, $c(x_0) = 0$, $\partial c(x_0)$ 存在非空,*

$$\mathcal{X} \stackrel{\text{def}}{=\!=} \{x \mid c(x) \le 0\},$$

*且Slater条件满足: 存在x_s 使得$c(x_s) < 0$, 则有$N_{\mathcal{X}}(x_0) = \textbf{cone}\,\partial c(x_0)$, 其中**cone**$S$ 是集合S 的锥包,* **cone**$S \stackrel{\text{def}}{=\!=} \{tx \mid x \in S,\ t \ge 0\}$.

Lemma 1. *设f 为适当凸函数, 且在x_0 处次微分不为空集, 则对任意$d \in \mathbb{R}^n$ 有*

$$\partial f(x_0; d) = \sup_{g \in \partial f(x_0)} g^{\mathrm{T}} d,$$

且当$\partial f(x_0; d)$ 不为无穷时, 上确界可以取到.

- 定义$K_0 = \textbf{cone}(\mathcal{X} - \{x_0\})$, $\mathcal{G}(x_0) = \textbf{cone}\,\partial c(x_0)$.

- 根据极锥和法锥的定义, $N_{\mathcal{X}}(x_0) = (K_0)^{\circ}$.

- 先证 $K_0 \subseteq \big(\mathcal{G}(x_0)\big)^\circ$: 设$d \in K_0$, 由$\mathcal{X}$ 的凸性知, 当τ 充分小时有

$$x_0 + \tau d \in \mathcal{X}.$$

根据引理1及$c(x_0) = 0$知, x_0 处的方向导数$\partial c(x_0; d) \le 0$且

$$0 \ge \partial c(x_0; d) = \sup_{s \in \partial c(x_0)} d^{\mathrm{T}} s \Rightarrow d^{\mathrm{T}} s \le 0,\ \forall s \in \mathcal{G}(x_0) \Rightarrow K_0 \subseteq \big(\mathcal{G}(x_0)\big)^\circ.$$

- 再证 $\big(\mathcal{G}(x_0)\big)^\circ \subseteq \overline{K_0}$: 设$d \in \big(\mathcal{G}(x_0)\big)^\circ$, 则根据引理1, $\partial c(x_0; d) \le 0$.

① $\partial c(x_0; d) < 0$. 由方向导数的定义知, 当τ 充分小时有$c(x_0 + \tau d) < 0$,
$\Rightarrow \quad x_0 + \tau d \in \mathcal{X} \Rightarrow d \in K_0$.

② $\partial c(x_0; d) = 0$. 构造$\{d^k\}$: $\quad d^k = (1 - \alpha_k)d + \alpha_k(x_s - x_0)$, 其中$\{\alpha_k\}$ 为单调下降趋于0 的正序列, 则由$c(x_s) < 0$ 及$c(x_0) = 0$ 可知d^k 处的方向导数:

$$\begin{aligned}\partial c(x_0; d^k) = \sup_{s \in \partial c(x_0)} s^{\mathrm{T}} d^k &\le (1 - \alpha_k) \sup_{s \in \partial c(x_0)} s^{\mathrm{T}} d + \alpha_k \sup_{s \in \partial c(x_0)} s^{\mathrm{T}} (x_s - x_0) \\ &\le (1 - \alpha_k)\partial c(x_0; d) + \alpha_k c(x_s) < 0.\end{aligned}$$

由①可知$d^k \in K_0,\ \forall k$. 令$k \to \infty$ 得$d \in \overline{K_0}$.

因此, 由①②可知$(\mathcal{G}(x_0))^\circ \subseteq \overline{K_0}$.

- 于是有$K_0 \subseteq (\mathcal{G}(x_0))^\circ \subseteq \overline{K_0} \;\Rightarrow\; (\mathcal{G}(x_0))^\circ = \overline{K_0}$. 从而,

$$N_{\mathcal{X}}(x_0) = (\overline{K_0})^\circ = (\mathcal{G}(x_0))^{\circ\circ} = \mathcal{G}(x_0).$$

Theorem 16 (法锥$N_{\mathcal{X}}(x)$ 的分解形式). *考虑凸优化*(CovP)*. 设其满足Slater条件, $x \in \mathcal{X}$ 且$\mathcal{A}(x)$ 为x 的积极集,则*

$$N_{\mathcal{X}}(x) = \sum_{i \in \mathcal{A}(x) \cap \mathcal{I}} \mathcal{G}_i(x) + \mathcal{R}(A^{\mathrm{T}}),$$

其中$\mathcal{G}_i(x) = \mathbf{cone}\partial c_i(x), \mathcal{R}(A^{\mathrm{T}})$ 是A^{T} 的像空间. $\Rightarrow$ ***KKT条件***

Proof. 令$\mathcal{L} = \{x|\; Ax = b\}$, 只需证$\mathcal{R}(A^{\mathrm{T}}) = N_{\mathcal{L}}(x)$.

$$A(y - x) = 0, \forall y \in \mathcal{L} \quad \Rightarrow \quad \mathcal{R}(A^{\mathrm{T}}) \subseteq N_{\mathcal{L}}(x).$$

对任意$w \in \mathbb{R}^n$ 有

$$w = z + A^{\mathrm{T}} s, \quad z \in \mathcal{N}(A),\; s \in \mathbb{R}^p,$$

可知若$w \in N_{\mathcal{L}}(x)$ 则必有$z = 0$. 这说明$N_{\mathcal{L}}(x) \subseteq \mathcal{R}(A^{\mathrm{T}})$. □

基追踪问题的最优性条件

$$\begin{array}{ll} \min\limits_{x\in\mathbb{R}^n} & \|x\|_1, \\ \text{s.t.} & Ax=b. \end{array} \iff \begin{array}{ll} \min & \sum\limits_i x_i^+ + x_i^-, \\ \text{s.t.} & Ax^+ - Ax^- = b, \\ & x^+, x^- \geq 0. \end{array} \iff \begin{array}{ll} \min\limits_{y\in\mathbb{R}^{2n}} & \mathbf{1}^{\mathrm{T}} y, \\ \text{s.t.} & [A, -A]y = b, \\ & y \geq 0. \end{array}$$

根据线性规划问题的最优性条件, 求解基追踪问题等价于求解

$$\begin{cases} \mathbf{1} + [A, -A]^{\mathrm{T}} \nu^* - s^* = 0, \\ [A, -A]y^* = b, \\ y^* \geq 0, \\ s^* \geq 0, \\ s^* \odot y^* = 0, \end{cases}$$

其中$s^* \in \mathbb{R}^{2n}, \nu^* \in \mathbb{R}^m$.

▶ 根据凸约束优化问题的一阶充要条件, x^* 为基追踪问题的全局最优解当且仅当存在$\nu^* \in \mathbb{R}^m$ 使得

$$\begin{cases} 0 \in \partial \|x^*\|_1 + A^{\mathrm{T}}\nu^*, \\ b = Ax^*. \end{cases}$$

- 令$x_i^* = y_i^* - y_{n+i}^*,\ i = 1, \ldots, n$, 则
  $x_i^* < 0 \Rightarrow y_i^* = 0,\ y_{n+i}^* = -x_i^* > 0, (A^{\mathrm{T}}\nu^*)_i = 1.$
  $x_i^* > 0 \Rightarrow y_i^* = x_i^* > 0,\ y_{n+i}^* = 0, (A^{\mathrm{T}}\nu^*)_i = -1.$
  $x_i^* = 0 \Rightarrow y_i^* = y_{n+i}^* = 0, (A^{\mathrm{T}}\nu^*)_i \in [-1, 1].$
  $(A^{\mathrm{T}}\nu^*)_i = -1 + s_i^* = 1 - s_{n+i}^*.$
- $[A, -A]y^* = b \ \Leftrightarrow \ Ax^* = b.$
- $\mathbf{1} + [A, -A]^{\mathrm{T}}\nu^* - s^* = 0.$
- $y^*\mathbb{R}_+^n,\ s^*\mathbb{R}_+^n,\ s^* \odot y^* = 0.$

约束优化的最优性理论

Definition 8 (切锥). *给定可行域$\mathcal{X}$ 及$x \in \mathcal{X}$, 若存在序列$\{z_k\} \subset \mathcal{X}$, $\lim_{k\to\infty} z_k = x$ 以及正数列$\{t_k\}$, $t_k \to 0$ 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.

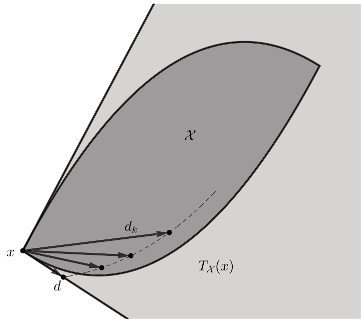


Figure 4: 不等式约束

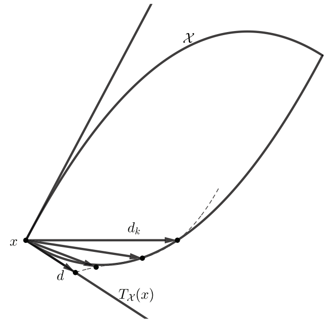


Figure 5: 等式约束

几何最优性条件

考虑约束优化:

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

Theorem 17 (几何最优性条件). *假设可行点x^* 是(P)的一个局部最优解, 若$f(x)$ 和$c_i(x)$, $i \in \mathcal{I} \cup \mathcal{E}$ 在点x^* 处是可微的, 则*

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

i.e.,

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

Proof. 用反证法.

- 若$T_{\mathcal{X}}(x^*) \cap \{d \mid \nabla f(x^*)^{\mathrm{T}} d < 0\} \neq \emptyset$, 取$d \in T_{\mathcal{X}}(x^*)$ 且$\nabla f(x^*)^{\mathrm{T}} d < 0$.

- 存在$\{t_k\}_{k=1}^{\infty}$ 和$\{d_k\}_{k=1}^{\infty}$ 使得$x^* + t_k d_k \in \mathcal{X}$, 其中$t_k \to 0$ 且$d_k \to d$.

- 由于$\nabla f(x^*)^{\mathrm{T}} d < 0$, 对于充分大的$k$, 我们有

$$
\begin{aligned}
f(x^* + t_k d_k) &= f(x^*) + t_k \nabla f(x^*)^{\mathrm{T}} d_k + o(t_k) \\
&= f(x^*) + t_k \nabla f(x^*)^{\mathrm{T}} d + t_k \nabla f(x^*)^{\mathrm{T}} (d_k - d) + o(t_k) \\
&= f(x^*) + t_k \nabla f(x^*)^{\mathrm{T}} d + o(t_k) \\
&< f(x^*)
\end{aligned}
$$

  这与x^* 的局部极小性矛盾.

$\square$

线性化可行方向锥

Definition 9 (线性化可行方向锥). *对于可行点*$x \in \mathcal{X}$, *定义该点的积极集*$\mathcal{A}(x) = \mathcal{E} \cup \{i \in \mathcal{I} \,:\, c_i(x) = 0\}$, *点*$x$ *处的*线性化可行方向锥*定义为*

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

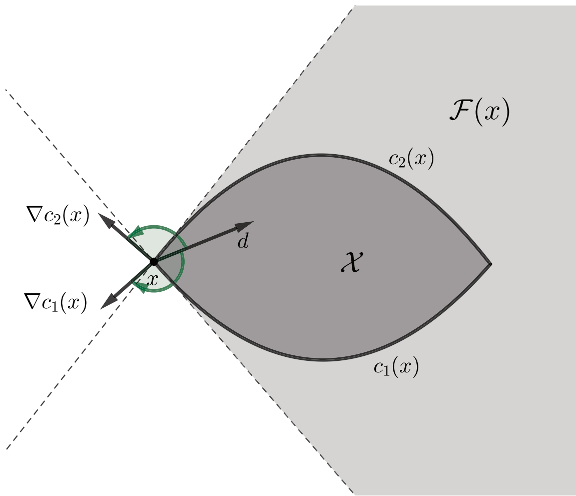


Figure 6: $\mathbb{R}^2$ 上的约束集合和线性化可行方向锥

线性化可行方向锥包含切锥

Theorem 18. *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$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$.*

Proof. 设$d \in T_{\mathcal{X}}(x)$, 由定义

$$\lim_{k\to\infty} t_k = 0, \quad \lim_{k\to\infty} \frac{z_k - x}{t_k} = d \Leftrightarrow z_k = x + t_k d + e_k$$

其中残量e_k 满足$\|e_k\| = o(t_k)$.对$i \in \mathcal{E}$, 根据泰勒展开,

$$\begin{aligned} 0 &= \frac{1}{t_k} c_i(z_k) \\ &= \frac{1}{t_k}(c_i(x) + \nabla c_i(x)^{\mathrm{T}}(t_k d + e_k) + o(t_k)) \\ &= \nabla c_i(x)^{\mathrm{T}} d + \frac{\nabla c_i(x)^{\mathrm{T}} e_k}{t_k} + o(1). \end{aligned}$$

注意到$\frac{\|e_k\|}{t_k} \to 0$, 令$k \to \infty$ 即可得到

$$\nabla c_i(x)^{\mathrm{T}} d = 0, \quad i \in \mathcal{E}.$$

对$i \in \mathcal{A}(x) \cap \mathcal{E}$, 根据泰勒展开,

$$\begin{aligned} 0 &\geq \frac{1}{t_k} c_i(z_k) \\ &= \frac{1}{t_k}\left(c_i(x) + \nabla c_i(x)^{\mathrm{T}}(t_k d + e_k) + o(t_k)\right) \\ &= \nabla c_i(x)^{\mathrm{T}} d + \frac{\nabla c_i(x)^{\mathrm{T}} e_k}{t_k} + o(1). \end{aligned}$$

注意到$c_i(x) = 0,\ i \in \mathcal{I}$, 因此我们有

$$\nabla c_i(x)^{\mathrm{T}} d \leq 0, \quad i \in \mathcal{A}(x) \cap \mathcal{E}.$$

因此有$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$. $\square$

切锥未必包含线性化可行方向锥

反例:

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

- 则$T_{\mathcal{X}}(3) = \{d \mid d \geq 0\}, \mathcal{F}(3) = \{d : d \geq 0\}$, 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$T_{\mathcal{X}}(x^*) = \{d : d \geq 0\}$ 不变

- 由$c'(x^*) = -3(x^* - 3)^2 = 0$ 知线性化可行锥$\mathcal{F}(x^*) = \{d \mid d \in \mathbb{R}\}$

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

注：切锥与线性化可行方向锥

- 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
- 引入约束规范来沟通两者, 确保最优点x^* 处$T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$, 从而可以用$\mathcal{F}(x)$ 取代$T_{\mathcal{X}}(x)$

线性无关约束规范

Definition 10 (线性无关约束规范). *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, 则称线性无关约束规范*(LICQ) *在点x 处成立.*

Theorem 19. 设LICQ*在可行点$x \in \mathcal{X}$处成立, 则有$T_{\mathcal{X}}(x) = \mathcal{F}(x)$.*

Proof. 不妨设积极集$\mathcal{A}(x) = \mathcal{E} \cup \mathcal{I}$, 其元素个数为$m$, 则矩阵

$$A(x) = [\nabla c_i(x)]_{i\in\mathcal{A}(x)}^{\mathrm{T}} \in \mathbb{R}^{m\times n}, \quad \text{rank}(A) = m.$$

- 令矩阵$Z \in \mathbb{R}^{n\times(n-m)}$ 为$A(x)$ 的零空间的基矩阵, 则Z 满足

$$\mathrm{rank}(Z) = n - m, \quad A(x)Z = 0.$$

- 令$d \in \mathcal{F}(x)$ 为任意线性化可行方向, 给定$\lim_{k\to\infty} t_k = 0$ 的正标

量$\{t_k\}_{k=1}^{\infty}$, 定义映射$R : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}^n$:

$$R(z,t) = \begin{bmatrix} c(z) - tA(x)d \\ Z^{\mathrm{T}}(z - x - td) \end{bmatrix},$$

其中$c(z)$ 为向量值函数, 其第i 个分量为$c_i(z)$.

- 由$\mathcal{A}(x) = \mathcal{E} \cup \mathcal{I}$ 知

$$R(x,0) = \begin{bmatrix} c(x) \\ Z^{\mathrm{T}}(x - x) \end{bmatrix} = 0, \quad \frac{\partial R(x,0)}{\partial z} = \begin{bmatrix} A(x) \\ Z^{\mathrm{T}} \end{bmatrix}.$$

- 根据Z 的构造, 雅可比矩阵$\frac{\partial R(x,0)}{\partial z}$ 是非奇异的.因此, 由隐函数定理, 对任意充分小的t_k, 都存在唯一的z_k, 使得$R(z_k, t_k) = 0$.

- 由于$R(z_k, t_k) = 0$, 故$c_i(z_k) = t_k \nabla c_i(x)^{\mathrm{T}} d,\ i \in \mathcal{I} \cup \mathcal{E}$.根据线性化可行方向$d$ 的定义, $c_i(z_k) \leq 0,\ i \in \mathcal{I},\ c_i(z_k) = 0,\ i \in \mathcal{E}$, 即$z_k$ 为可行点.

- 由泰勒展开得

$$0 = R(z_k, t_k) = \begin{bmatrix} c(z_k) - t_k A(x) d \\ Z^{\mathrm{T}}(z_k - x - t_k d) \end{bmatrix} = \begin{bmatrix} A(x)(z_k - x) + e_k - t_k A(x) d \\ Z^{\mathrm{T}}(z_k - x - t_k d) \end{bmatrix}$$

$$= \begin{bmatrix} A(x) \\ Z^{\mathrm{T}} \end{bmatrix} (z_k - x - t_k d) + \begin{bmatrix} e_k \\ 0 \end{bmatrix}.$$

其中残量e_k 满足$\|e_k\| = o(t_k)$.

- 两边同时作用$\begin{bmatrix} A(x)^{\mathrm{T}} & Z \end{bmatrix}^{-\mathrm{T}}$ 并除以t_k, 则有

$$\frac{z_k - x}{t_k} = d + \begin{bmatrix} A(x) \\ Z^{\mathrm{T}} \end{bmatrix}^{-1} \begin{bmatrix} \frac{e_k}{t_k} \\ 0 \end{bmatrix} \Rightarrow \lim_{k \to \infty} \frac{z_k - x}{t_k} = d,$$

即$d \in T_{\mathcal{X}}(x)$.故$\mathcal{F}(x) \subseteq T_{\mathcal{X}}(x)$.又$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$,从而$T_{\mathcal{X}}(x) = \mathcal{F}(x)$.

□

Mangasarian – Fromovitz约束规范(MFCQ)

Definition 11 (Mangasarian – Fromovitz约束规范). *给定可行点x 及相应的积极集$\mathcal{A}(x)$.如果存在一个向量$w \in \mathbb{R}^n$, 使得*

$$\nabla c_i(x)^{\mathrm{T}} w < 0, \quad \forall i \in \mathcal{A}(x) \cap \mathcal{I},$$

$$\nabla c_i(x)^{\mathrm{T}} w = 0, \quad \forall i \in \mathcal{E},$$

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MFCQ成立.

- Mangasarian – Fromovitz约束规范是LICQ的的一个常用推广, 简称为MFCQ.
- LICQ可以推出MFCQ, 但是反过来不成立.在MFCQ 成立的情况下, 我们也可以证明$T_{\mathcal{X}}(x) = \mathcal{F}(x)$.

线性约束规范

- 另外一个用来保证$T_{\mathcal{X}}(x)=\mathcal{F}(x)$ 的约束规范是线性约束规范.

Definition 12 (线性约束规范). *若所有的约束函数$c_i(x)$, $i\in\mathcal{I}\cup\mathcal{E}$都是线性的, 则称线性约束规范成立.*

- 当线性约束品性成立时, 也有$T_{\mathcal{X}}(x)=\mathcal{F}(x)$.

- 因此对只含线性约束的优化问题, 例如线性规划、二次规划, 很自然地有$T_{\mathcal{X}}(x)=\mathcal{F}(x),\forall\, x$.我们无需再关注约束函数的梯度是否线性无关.

- 一般来说, 线性约束规范和LICQ之间没有互相包含的关系.

► 在LICQ或MFCQ或线性约束下, $T_{\mathcal{X}}(x)=\mathcal{F}(x)$成立, 根据Farkas' Lemma, KKT条件成立.

二阶最优性条件

考虑一般优化问题:

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

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

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

此时一阶条件无法判断x^* 是否是最优值点.

- 若$d^{\mathrm{T}}\nabla f(x^*)=0$, 则需要利用二阶信息来进一步判断在其可行邻域内的目标函数值.
- 拉格朗日函数在这些方向上的曲率即可用来判断x^* 的最优性.
- 首先引入临界锥来精确刻画这些方向.

临界锥

Definition 13 (临界锥). *设(x^*, λ^*) 是满足KKT条件的KKT对, 定义临界锥为*

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

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.
- 沿着临界锥中的方向进行优化, 所有等式约束和$\lambda_i^* > 0$ 对应的不等式约束(此时这些不等式均取等)都会尽量保持不变.
- 当$d \in \mathcal{C}(x^*, \lambda^*)$ 时, $\forall i \in \mathcal{E} \cup \mathcal{I}$ 有$\lambda_i^* \nabla c_i(x^*)^{\mathrm{T}} d = 0$, 故

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

- 临界锥定义了依据一阶导数不能判断是否为下降或上升方向的线性化可行方向, 必须使用高阶导数信息加以判断.

二阶最优性条件

考虑一般优化问题:

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

Theorem 20 (二阶最优性条件). **必要性**:*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λ^* 为相应的拉格朗日乘子, 即(x^*,λ^*) 满足KKT条件, 那么*

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

充分性:*假设在可行点x^* 处, 存在一个拉格朗日乘子λ^*, 使得(x^*,λ^*) 满足KKT条件.如果*

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

那么x^* 为问题的一个严格局部极小解.