最优性理论

文再文

北京大学北京国际数学研究中心

教材《最优化：建模、算法与理论》配套电子教案

http://bicmr.pku.edu.cn/~wenzw/optbook.html

致谢：本教案由谢中林协助准备

提纲

最优化问题解的存在性

考虑优化问题

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & f(x), \\ \text{s.t.} \quad & x \in \mathcal{X}, \end{aligned}$$

其中 $\mathcal{X} \subseteq \mathbb{R}^n$ 为可行域.

- 首先要考虑的是最优解的存在性, 然后考虑如何求出其最优解.
- 在数学分析课程中, 我们学习过Weierstrass 定理, 即定义在紧集上的连续函数一定存在最大(最小) 值点.
- 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, 因此需要将此定理推广来保证最优化问题解的存在性.

定理 (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:

1. $\mathrm{dom} f \stackrel{\text{def}}{=\!=} \{x \in \mathcal{X} : f(x) < +\infty\}$ 是有界的;
2. 存在一个常数$\bar{\gamma}$ 使得下水平集
$$C_{\bar{\gamma}} \stackrel{\text{def}}{=\!=} \{x \in \mathcal{X} : f(x) \le \bar{\gamma}\}$$
是非空且有界的;
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$\{x \in \mathcal{X} \mid f(x) \le f(y), \forall y \in \mathcal{X}\}$ 非空且紧.

条件(2) 下的证明

假设条件(2) 成立，且$t \overset{\text{def}}{=\!=} \inf_{x\in\mathcal{X}} f(x) = -\infty$.

- 由下确界的定义，存在点列$\{x^k\}_{k=1}^{\infty} \subset C_{\bar{\gamma}}$，使得$\lim_{k\to\infty} f(x^k) = -\infty$.
- 由$C_{\bar{\gamma}}$ 有界知点列$\{x^k\}$ 存在聚点x^*.
- 由f 是闭函数知$\textbf{epi}f$ 为闭集，因此$(x^*, t) \in \textbf{epi}f$. 根据上方图的定义知$f(x^*) \le t = -\infty$，这与$f$ 适当矛盾，故$t \overset{\text{def}}{=\!=} \inf_{x\in\mathcal{X}} f(x)$ 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，而紧集的子集也是紧集，故最小值点集为非空紧集.

条件(1)(3)下的证明

假设条件(1)成立, 则$\mathrm{dom} f$ 是有界的.

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

假设条件(3)成立, 我们证明条件(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$\lim\limits_{k\to\infty} \|x^k\| = +\infty$, $\lim\limits_{k\to\infty} f(x^k) = +\infty$, 这与$f(x^k) \leq \bar{\gamma}$ 矛盾, 故此时f 的任意下水平集都有界, 条件(2)成立.

推广的Weierstrass定理:注记

- 推广的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.
- 因此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, 因此比数学分析中的Weierstrass定理应用范围更广.
- 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$f(x)=e^{-x},\ x\in\mathcal{X}=\mathbb{R}$, 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.

解的存在唯一性

- 推广的Weierstrass定理给出了最优解的存在性条件，但其对应的解可能不止一个.
- 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.
- 但是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.
- 因此，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.

解的存在唯一性: 拟强凸函数

定义 (拟强凸函数)

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

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

那么我们称函数f 是**强拟凸**的.

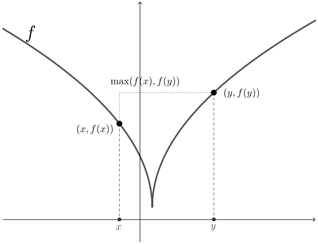


Figure: 一个强拟凸函数

拟强凸函数

- 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, 并可以包含一部分性质较好的非凸函数.
- 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.
- 任何定义在闭有界凸集上的强凸函数(如$f(x) = x^2$), 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, 比如$f(x) = \max\{x, 0\}$, 任意$x \leq 0$ 都是$f(x)$ 的最优解.

唯一性定理

定理 (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

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

证明: 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，则有:

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

这与x^* 的全局极小性矛盾.

提纲

无约束可微优化问题通常表示为如下形式：

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

其中f 是连续可微函数.

- 给定一个点$\bar{x}$, 我们想要知道这个点是否是函数f 的一个局部极小解或者全局极小解.
- 如果从定义出发, 需要对其邻域内的所有点进行判断, 这不可行.
- 因此, 需要一个更简单的方式来验证一个点是否为极小值点.我们称其为最优性条件, 它主要包含一阶最优性条件和二阶最优性条件.

一阶必要条件:下降方向

定义 (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

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

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.
- 由下降方向的定义, 容易验证: 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

$$f(x + td) < f(x).$$

因此, 在局部最优点处不能有下降方向.

一阶必要条件

定理 (一阶必要条件)

假设f 在全空间$\mathbb{R}^n$ 可微.如果x^* 是一个局部极小点, 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

$$f(x^* + tv) = f(x^*) + tv^{\mathrm{T}}\nabla f(x^*) + o(t),$$

整理得

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

根据x^* 的最优性, 在上式中分别对t 取点0 处的左, 右极限可知

$$\lim_{t\to 0^+}\frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}}\nabla f(x^*) \geq 0,$$
$$\lim_{t\to 0^-}\frac{f(x^* + tv) - f(x^*)}{t} = v^{\mathrm{T}}\nabla f(x^*) \leq 0,$$

即对任意的v 有$v^{\mathrm{T}}\nabla f(x^*) = 0$, 由$v$ 的任意性知$\nabla f(x^*) = 0$.

一阶必要性条件:注记

- 对于$f(x) = x^2, x \in \mathbb{R}$, 我们知道满足$f'(x) = 0$ 的点为$x^* = 0$, 并且其也是全局最优解.
- 对于$f(x) = x^3, x \in \mathbb{R}$, 满足$f'(x) = 0$ 的点为$x^* = 0$, 但其不是一个局部最优解.
- 称满足$\nabla f(x) = 0$ 的点x 为f 的稳定点(有时也称为驻点或临界点).
- 除了一阶必要条件, 还需要对函数加一些额外的限制条件, 才能保证最优解的充分性.

二阶最优性条件

- 在没有额外假设时，如果一阶必要条件满足，我们仍然不能确定当前点是否是一个局部极小点.
- 假设f 在点x 的一个开邻域内是二阶连续可微的.类似于一阶必要条件的推导，可以借助当前点处的二阶泰勒展开来逼近该函数在该点附近的取值情况，从而来判断最优性.
- 在点x 附近我们考虑泰勒展开

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

- 当一阶必要条件满足时，$\nabla f(x) = 0$，那么上面的展开式简化为

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

由此，我们可以导出二阶最优性条件.

定理 (二阶最优性条件)

必要条件:若x^* 是f 的一个局部极小点, 则$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0$.

充分条件:若$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$, 则$x^*$ 是f 的一个局部极小点.

- **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

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

  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,

$$\frac{f(x^*+d)-f(x^*)}{\|d\|^2} = \frac{d^{\mathrm{T}}\nabla^2 f(x^*) d + o(\|d\|^2)}{\|d\|^2} \geq \frac{1}{2}\lambda_{\min} + o(1).$$

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

二阶最优性条件:注记

- 设点$\bar{x}$ 满足一阶最优性条件(即$\nabla f(\bar{x}) = 0$), 且该点处的海瑟矩阵$\nabla^2 f(\bar{x})$ 不是半正定的, 那么$\bar{x}$ 不是一个局部极小点.
- 进一步地, 如果海瑟矩阵$\nabla^2 f(\bar{x})$ 既有正特征值又有负特征值, 我们称稳定点$\bar{x}$ 为一个鞍点.
- 事实上, 记d_1, d_2 为其正负特征值对应的特征向量, 那么对于任意充分小的$t > 0$, 我们都有$f(\bar{x} + td_1) > f(\bar{x})$ 且$f(\bar{x} + td_2) < f(\bar{x})$.
- 注意, 二阶最优性条件给出的仍然是关于局部最优性的判断.对于给定点的全局最优性判断, 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.

- 线性最小二乘:

$$\min_{x\in\mathbb{R}^n} \quad f(x) \stackrel{\text{def}}{=\!=} \frac{1}{2}\|b-Ax\|_2^2,$$

其中$A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$.由f 可微且凸知

$$x^* \text{ 为一个全局最优解} \Leftrightarrow \nabla f(x^*) = A^{\mathrm{T}}(Ax^*-b) = 0.$$

- 非线性最小二乘(实数情形的相位恢复):

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

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

实数情形的相位恢复

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

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

- 如果x^* 为局部最优解, 那么其满足一、二阶必要条件

$$\sum_{i=1}^{m}((a_i^{\mathrm{T}}x^*)^2 - b_i)(a_i^{\mathrm{T}}x^*)a_i = 0, \quad \sum_{i=1}^{m}(12(a_i^{\mathrm{T}}x^*)^2 - 4b_i^2)a_i a_i^{\mathrm{T}} \succeq 0.$$

- 如果一个点$x^\#$ 满足二阶充分条件

$$\sum_{i=1}^{m}((a_i^{\mathrm{T}}x^\#)^2 - b_i)(a_i^{\mathrm{T}}x^\#)a_i = 0, \quad \sum_{i=1}^{m}(12(a_i^{\mathrm{T}}x^\#)^2 - 4b_i^2)a_i a_i^{\mathrm{T}} \succ 0,$$

那么$x^\#$ 为局部最优解.

提纲

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

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且各下标互不相同.

- 这个问题的可行域定义为

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

- 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.

拉格朗日函数

一般的约束优化问题:

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

变量$x \in \mathbb{R}^n$, 最优值为p^*, 定义域为

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

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

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

- λ_i 为第i 个不等式约束对应的拉格朗日乘子
- ν_i 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

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

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

定理 (弱对偶原理)

若$\lambda \geq 0$, 则$g(\lambda,\nu) \leq p^*$.

证明:若$\tilde{x}\in\mathcal{X}$, 则

$$g(\lambda,\nu) = \inf_x L(x,\lambda,\nu) \leq L(\tilde{x},\lambda,\nu) \leq f(\tilde{x}),$$

对$\tilde{x}$取下界得

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

拉格朗日对偶问题

拉格朗日对偶问题:

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

- 称 λ 和 ν 为对偶变量, 设最优值为 q^*
- q^* 为 p^* 的最优下界, 称 $p^* - q^*$ 为对偶间隙
- 拉格朗日对偶问题是一个凸优化问题
- $\mathrm{dom} g = \{(\lambda, \nu) \mid \lambda \geq 0, g(\lambda, \nu) > -\infty\}$, 称其元素为对偶可行解

例: 标准形式线性规划及其对偶

$$\begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax = b \\ & x \geq 0 \end{array} \qquad\qquad \begin{array}{ll} \max & -b^T \nu \\ \text{s.t.} & A^T \nu + c \geq 0 \end{array}$$

实例:线性方程组具有最小模的解

$$\begin{array}{ll} \min & x^T x \\ \text{s.t.} & Ax = b \end{array}$$

对偶函数

- 拉格朗日函数为$L(x,\nu) = x^T x + \nu^T (Ax - b)$
- 求L 关于x 的最小值，由一阶条件:

$$\nabla_x L(x,\nu) = 2x + A^T \nu = 0 \quad \Longrightarrow \quad x = -(1/2)A^T \nu$$

- 将上式代入L 得到对偶函数g:

$$g(\nu) = L((-1/2)A^T \nu, \nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

它是关于ν 的凹函数

弱对偶性: $p^* \geq -(1/4)\nu^T A A^T \nu - b^T \nu, \ \forall \nu$

实例:标准形式线性规划

$$
\begin{array}{ll}
\min & c^T x \\
\text{s.t.} & Ax = b, \quad x \geq 0
\end{array}
$$

对偶函数

- 拉格朗日函数为

$$
\begin{aligned}
L(x, \lambda, \nu) &= c^T x + \nu^T (Ax - b) - \lambda^T x \\
&= -b^T \nu + (c + A^T \nu - \lambda)^T x
\end{aligned}
$$

- L 关于x 线性, 因此

$$
g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = \begin{cases} -b^T \nu & A^T \nu - \lambda + c = 0 \\ -\infty & \text{otherwise} \end{cases}
$$

g 在仿射集$\{(\lambda, \nu) \mid A^T \nu - \lambda + c = 0\}$ 上线性, 因此是凹函数

弱对偶性: $p^* \geq -b^T \nu$ 若$A^T \nu + c \geq 0$

$$
\begin{aligned}
\min_{x} \quad & c^{\mathrm{T}} x \\
\text{s.t.} \quad & Ax = b \\
& x \geq 0
\end{aligned}
$$

- 拉格朗日函数:

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

- 对偶函数:

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

- 对偶问题:

$$
\begin{aligned}
\max_{s,\nu} \quad & -b^{\mathrm{T}}\nu, \\
\text{s.t.} \quad & A^{\mathrm{T}}\nu - s + c = 0, \\
& s \geq 0.
\end{aligned}
\quad \overset{y=-\nu}{\Longleftrightarrow} \quad
\begin{aligned}
\max_{s,y} \quad & b^{\mathrm{T}}y, \\
\text{s.t.} \quad & A^{\mathrm{T}}y + s = c, \\
& s \geq 0.
\end{aligned}
$$

- 若保留约束$x \geq 0$, 则拉格朗日函数为

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

- 对偶问题需要将$x \geq 0$ 添加到约束里：

$$\max_{y} \left\{ \inf_{x} \ b^{\mathrm{T}}y + (c - A^{\mathrm{T}}y)^{\mathrm{T}}x, \quad \text{s.t.} \ \ x \geq 0 \right\} \Rightarrow \begin{array}{ll} \max\limits_{y} & b^{\mathrm{T}}y, \\ \text{s.t.} & A^{\mathrm{T}}y \leq c. \end{array}$$

  此对偶问题可以通过将上页最后一个问题中的变量s 消去得到.

- 视$\max \ b^{\mathrm{T}}y$ 为$\min \ -b^{\mathrm{T}}y$, 对偶问题的拉格朗日函数为

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

线性规划问题的对偶

- 因此得到对偶函数

$$g(x) = \inf_{y} L(y,x) = \begin{cases} -c^{\mathrm{T}}x, & Ax = b, \\ -\infty, & \text{其他}. \end{cases}$$

- 相应的对偶问题是

$$\begin{aligned} \max_{x} \quad & -c^{\mathrm{T}}x, \\ \text{s.t.} \quad & Ax = b, \\ & x \geq 0. \end{aligned}$$

该问题与原始问题完全等价，这表明线性规划问题与其对偶问题互为对偶.

强对偶性: 若一个线性规划问题有最优解, 则其对偶问题有最优解, 且最优值相等.

dual \ primal	finite	unbounded	infeasible
finite	$\surd$	$\times$	$\times$
unbounded	$\times$	$\times$	$\surd$
infeasible	$\times$	$\surd$	$\surd$

- 若$p^* = -\infty$, 则$d^* \le p^* = -\infty$, 因此对偶问题不适定
- 若$d^* = +\infty$, 则$+\infty = d^* \le p^*$, 因此原问题不适定

$$\begin{array}{ll} \min & x_1 + 2x_2 \\ \text{s.t.} & x_1 + x_2 = 1 \\ & 2x_1 + 2x_2 = 3 \end{array} \qquad \begin{array}{ll} \max & p_1 + 3p_2 \\ \text{s.t.} & p_1 + 2p_2 = 1 \\ & p_1 + 2p_2 = 2 \end{array}$$

$$\begin{array}{ll} \min & \|x\| \\ \text{s.t.} & Ax = b \end{array}$$

对偶函数

$$g(\nu) = \inf_x (\|x\| - \nu^T Ax + b^T \nu) = \begin{cases} b^T \nu & \|A^T \nu\|_* \le 1 \\ -\infty & \text{其他} \end{cases}$$

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

证明: 利用 $\inf_x (\|x\| - y^T x)$ 在 $\|y\|_* \le 1$ 时等于0 否则等于 $-\infty$

- 若 $\|y\|_* \le 1$, 则 $x - y^T x \ge 0$ 对任意 x 都成立, 当 $x = 0$ 时取等
- 若 $\|y\|_* > 1$, 取 $x = tu$, 其中 $\|u\| \le 1$, $u^T y = \|y\|_* > 1$:

$$\|x\| - y^T x = t(\|u\| - \|y\|_*) \to -\infty \quad \text{当} t \to \infty$$

弱对偶性: $p^* \ge b^T \nu$ 若 $\|A^T \nu\|_* \le 1$

实例: 二路划分

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

- 非凸问题; 可行域由2^n 个离散点组成
- 含义: 将$\{1, \ldots, n\}$ 划分为两个集合; W_{ij} 是将i, j 划入同一集合的代价; $-W_{ij}$ 则是将i, j 划入不同集合的代价

对偶函数

$$g(\nu) = \inf_x (x^T W x + \sum_i \nu_i (x_i^2 - 1)) = \inf_x x^T (W + \mathrm{diag}(\nu)) x - \mathbf{1}^T \nu$$

$$= \begin{cases} -\mathbf{1}^T \nu & W + \mathrm{diag}(\nu) \succeq 0 \\ -\infty & \text{其他} \end{cases}$$

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

例: 取$\nu = -\lambda_{\min}(W)\mathbf{1}$ 即得下界的估计, $p^* \geq n\lambda_{\min}(W)$

$$\begin{array}{ll} \min & f_0(x) \\ \text{s.t.} & Ax \leq b, \quad Cx = d \end{array}$$

对偶函数

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

- 回顾共轭函数的定义$f^*(y) = \sup_{x \in \mathbf{dom}\, f}(y^T x - f(x))$
- 在f_0 的共轭函数已知时可以简化对偶函数的推导

例：最大化熵

$$f_0(x) = \sum_{i=1}^n x_i \log x_i, \qquad f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

弱对偶性与强对偶性

弱对偶性: $d^* \le p^*$

- 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, 例如, SDP问题

$$\begin{array}{ll} \max & -\mathbf{1}^T \nu \\ \text{s.t.} & W + \operatorname{diag}(\nu) \succeq 0 \end{array}$$

  给出了二路划分问题的一个下界:

$$\min\ x^T W x, \qquad \text{s.t. } x_i^2 = 1, \quad i = 1, \ldots, n$$

强对偶性: $d^* = p^*$

- 对一般问题而言通常不成立
- (通常) 对凸问题成立
- 称保证凸问题强对偶性成立的条件为**约束品性**

问题形式与对偶性

- 一个问题不同等价形式的对偶可能差异巨大
- 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
- 将显式约束隐式化或将隐式约束显式化
- 改变目标函数或者约束函数的形式
  例如，用 $\phi(f_0(x))$ 取代$f_0(x)$，其中 ϕ 是凸的增函数

引入新变量与等式约束

$$\min \quad f_0(Ax+b)$$

- 对偶函数为常数
- 强对偶性成立，但对偶问题无意义

改写原问题及其对偶

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

对偶函数为

$$\begin{aligned} g(\nu) &= \inf_{x,y}(f_0(y) - \nu^T y + \nu^T Ax + b^T \nu) \\ &= \begin{cases} -f_0^*(\nu) + b^T\nu & A^T\nu = 0 \\ -\infty & \text{其他} \end{cases} \end{aligned}$$

范数逼近问题: $\min \|Ax - b\|$

$$\begin{aligned} \min \quad & \|y\| \\ \text{s.t.} \quad & y = Ax - b \end{aligned}$$

由$\|\cdot\|$的共轭函数知其对偶函数为:

$$\begin{aligned} g(\nu) &= \inf_{x,y}(\|y\| + \nu^T y - \nu^T Ax + b^T \nu) \\ &= \begin{cases} b^T \nu + \inf_y(\|y\| + \nu^T y) & A^T \nu = 0 \\ -\infty & \text{其他} \end{cases} \\ &= \begin{cases} b^T \nu & A^T \nu = 0, \quad \|\nu\|_* \le 1 \\ -\infty & \text{其他} \end{cases} \end{aligned}$$

范数逼近问题的对偶

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

带边界约束的线性规划: 原问题与对偶问题

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

通过隐式化边界约束改写原问题

$$
\begin{array}{ll}
\min & f_0(x) = \begin{cases} c^T x & -\mathbf{1} \le x \le \mathbf{1} \\ -\infty & \text{其他} \end{cases} \\
\text{s.t.} & Ax = b
\end{array}
$$

对偶函数为

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

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

带广义不等式约束优化问题

- 问题

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

  中的不等式约束$c_i(x)$, $i \in \mathcal{I}$ 都是实值函数的形式.

- 在许多实际应用中，我们还会遇到大量带广义不等式约束的优化问题，例如自变量x 可能取值于半正定矩阵空间中.

- 对于这类约束我们不易将其化为$c_i(x) \leq 0$ 的形式，此时又该如何构造拉格朗日对偶函数呢?

定义 (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), 即$\mathbf{int}\,K \neq \emptyset$ ；
4. K 是尖的(pointed), 即对任意非零向量x, 若$x \in K$, 则$-x \notin K$, 也即K 中无法容纳直线.

- 适当锥K 可以诱导出广义不等式, 它定义了全空间上的偏序关系：
$$x \preceq_K y \iff y - x \in K.$$
- 类似地, 可以定义严格广义不等式：
$$x \prec_K y \iff y - x \in \mathbf{int}\,K.$$
- 当$K = \mathbb{R}^n_+$ 时, $x \preceq_K y$ 是我们之前经常使用的记号$x \leq y$
- 当$K = \mathcal{S}^n_+$ 时, $X \preceq_K Y$ 表示$Y - X \succeq 0$, 即$Y - X$ 是半正定矩阵

对偶锥与拉格朗日乘子

- 对广义不等式，该如何构造广义不等式约束所对应的乘子？

定义 (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.

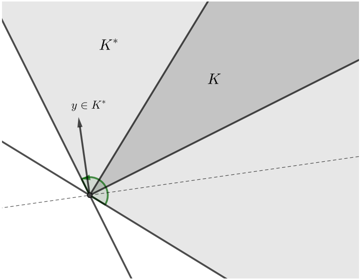


Figure: $\mathbb{R}^2$ 平面上的锥K 及其对偶锥K^*

对偶锥与拉格朗日乘子:注记

- 如果$K=\mathbb{R}^n_+$, $\Omega=\mathbb{R}^n$ 并且定义$\langle x,y\rangle=x^\mathrm{T}y$, 那么易知$K^*=\mathbb{R}^n_+$.
- 假设$K=\mathcal{S}^n_+$, $\Omega=\mathcal{S}^n$ 并且定义

$$\langle X,Y\rangle=\mathrm{Tr}(XY^\mathrm{T}),$$

可以证明

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

即半正定锥的对偶锥仍为半正定锥.

- 称满足$K=K^*$ 的锥K 为**自对偶锥**, 因此非负锥和半正定锥都是自对偶锥.
- 直观来说, 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:

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

  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:

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

- 容易验证$L(x, \lambda, \nu) \le f(x), \ \forall\, x \in \mathcal{X}, \ \lambda_i \in K_i^*, \nu_i \in \mathbb{R}$.

- 对偶函数$g(\lambda, \nu) = \inf_{x\in\mathbb{R}^n} L(x, \lambda, \nu)$, 对偶问题为

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

半定规划问题的对偶问题

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

其中$A_i \in \mathcal{S}^n, \ i = 1, 2, \cdots, m, \ C \in \mathcal{S}^n, b \in \mathbb{R}^m$

- 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:

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

- 对偶问题:

$$
\min_{y \in \mathbb{R}^m} \quad -b^{\mathrm{T}} y, \quad \text{s.t.} \quad \sum_{i=1}^{m} y_i A_i - C + S = 0, \ S \succeq 0
$$

半定规划对偶问题的对偶问题

$$\min_{y\in\mathbb{R}^m} \quad -b^{\mathrm{T}}y, \quad \text{s.t.} \quad \sum_{i=1}^m y_iA_i \preceq C$$

- 拉格朗日函数:

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

- 对偶函数:

$$g(X) = \inf_y \ L(y,X) = \begin{cases} -\langle C,X\rangle, & \langle A_i, X\rangle = b_i, \ i=1,2,\cdots,m \\ -\infty, & \text{其他} \end{cases}$$

- 对偶问题:

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

ℓ_1 正则化问题的对偶

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

令$r=Ax-b$, 问题等价于$\min_{x\in\mathbb{R}^n} \frac{1}{2}\|r\|^2+\mu\|x\|_1$， s.t. $r=Ax-b$

- 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:

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

- 对偶问题:

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

$$\begin{aligned} \max_{x \in \mathbb{R}^n} \quad & x^{\mathrm{T}} C x \\ \text{s.t.} \quad & x_i^2 = 1,\ i = 1, 2, \cdots, n \end{aligned}$$

- 拉格朗日函数:

$$L(x, y) = -x^{\mathrm{T}} C x + \sum_{i=1}^{n} y_i (x_i^2 - 1) = x^{\mathrm{T}} (\mathrm{Diag}(y) - C) x - \mathbf{1}^{\mathrm{T}} y$$

- 对偶函数:

$$g(y) = \inf_x L(x, y) = \begin{cases} -\mathbf{1}^{\mathrm{T}} y, & \mathrm{Diag}(y) - C \succeq 0 \\ -\infty, & \text{其他} \end{cases}$$

- 对偶问题:

$$\begin{aligned} \min_{y \in \mathbb{R}^n} \quad & \mathbf{1}^{\mathrm{T}} y \\ \text{s.t.} \quad & \mathrm{Diag}(y) - C \succeq 0 \end{aligned}$$

最大割对偶问题的对偶问题

$$\begin{aligned} \min_{y\in\mathbb{R}^n} \quad & \mathbf{1}^{\mathrm{T}} y \\ \text{s.t.} \quad & \mathrm{Diag}(y) - C \succeq 0 \end{aligned}$$

- 拉格朗日函数:

$$L(y,X) = \mathbf{1}^{\mathrm{T}} y - \langle \mathrm{Diag}(y) - C, X\rangle = \sum_{i=1}^{n} (1 - X_{ii}) y_i + \langle C, X\rangle$$

- 对偶函数:

$$g(X) = \inf_y L(y,X) = \begin{cases} \langle C, X\rangle, & X_{ii} = 1,\ i = 1,2,\cdots,n \\ -\infty, & \text{其他} \end{cases}$$

- 对偶问题:

$$\begin{aligned} \max \quad & \langle C, X\rangle \\ \text{s.t.} \quad & X_{ii} = 1,\ i = 1,2,\cdots,n \\ & X \succeq 0 \end{aligned}$$

$$\textbf{(P)} \quad \min \quad c^\top x \quad \text{s.t.} \quad Ax = b, x_Q \succeq 0$$

$$\textbf{(D)} \quad \max \quad b^\top y \quad \text{s.t.} \quad A^\top y + s = c, s_Q \succeq 0$$

$$\begin{aligned} \textbf{(P)} \quad \min \quad & \langle C, X \rangle \\ \text{s.t.} \quad & \langle A_1, X \rangle = b_1 \\ & \ldots \\ & \langle A_m, X \rangle = b_m \\ & X \succeq 0 \end{aligned}$$

$$\begin{aligned} \textbf{(D)} \quad \max \quad & b^\top y \\ \text{s.t.} \quad & \sum_i y_i A_i + S = C \\ & S \succeq 0 \end{aligned}$$

Strong duality

- If $p^* > -\infty$, (P) is **strictly** feasible, then (D) is feasible and $p^* = d^*$
- If $d^* < +\infty$, (D) is **strictly** feasible, then (P) is feasible and $p^* = d^*$
- If (P) and (D) has **strictly** feasible solutions, then both have optimal solutions.

Failure of SOCP Duality

$$\begin{array}{llcll} \inf & (1,-1,0)x & & \sup & y \\ \text{s.t.} & (0,0,1)x = 1 & & \text{s.t.} & (0,0,1)^\top y + z = (1,-1,0)^\top \\ & x_{\mathcal{Q}} \succeq 0 & & & z_{\mathcal{Q}} \succeq 0 \end{array}$$

- primal: $\min\ x_0 - x_1,$ s.t. $x_0 \geq \sqrt{x_1^2 + 1}$; It holds $x_0 - x_1 > 0$ and $x_0 - x_1 \to 0$ if $x_0 = \sqrt{x_1^2 + 1} \to \infty$. Hence, $p^* = 0$, no finite solution
- dual: $\sup\ y$ s.t. $1 \geq \sqrt{1 + y^2}$. Hence, $y = 0$

$p^* = d^*$ but primal is not attainable.

Failure of SDP Duality

Consider

$$\begin{aligned} \min \quad & \left\langle \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, X \right\rangle \\ \text{s.t.} \quad & \left\langle \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, X \right\rangle = 0 \\ & \left\langle \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix}, X \right\rangle = 2 \\ & X \succeq 0 \end{aligned}$$

$$\begin{aligned} \max \quad & 2y_2 \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

- primal: $X^* = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$, $p^* = 1$
- dual: $y^* = (0, 0)$. Hence, $d^* = 0$

Both problems have finite optimal values, but $p^* \neq d^*$

提纲

带约束凸优化问题

- 稀疏优化问题：

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

- 低秩矩阵恢复问题：

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

- 矩阵分离问题：

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

- 回归分析中的问题：

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

- 前述问题都可以写为

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

其中$f(x)$ 为适当的凸函数, $\forall i, c_i(x)$ 是凸函数且$\mathrm{dom} c_i=\mathbb{R}^n$.

- $A\in\mathbb{R}^{p\times n}, b\in\mathbb{R}^p$ 是已知的.
- 集合$\mathcal{D}$ 表示自变量x 的自然定义域, 即

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

- 自变量x 还受约束的限制, 定义可行域

$$\mathcal{X}=\{x\in\mathcal{D} : c_i(x)\le 0,\ i=1,2,\cdots,m;\ Ax=b\}.$$

- 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.

Slater约束品性与强对偶原理

- 凸优化问题有很多好的性质.一个自然的问题是:我们能否像研究无约束问题那样找到该问题最优解的一阶充要条件?如果这样的条件存在, 它在什么样的约束品性下成立?
- 在通常情况下, 优化问题的对偶间隙大于0, 即强对偶原理不满足.
- 但对很多凸优化问题, 在特定约束品性下可以证明强对偶原理.
- 直观的一种约束品性是存在满足所有约束条件的严格可行解.

Slater约束品性与强对偶原理:相对内点

首先给出集合$\mathcal{D}$ 的相对内点集relint$\mathcal{D}$ 的定义.给定集合$\mathcal{D}$, 记其仿射包为

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

定义 (相对内点)

集合$\mathcal{D}$ 的相对内点集定义为

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

相对内点是内点的推广, 若$\mathcal{D}$ 本身的“维数”较低, 则$\mathcal{D}$ 不可能有内点, 但如果在它的仿射包**affine** $\mathcal{D}$ 中考虑, 则$\mathcal{D}$ 可能有相对内点.

定义 (Slater约束品性)

若对凸优化问题

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

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

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

则称对此问题Slater 约束品性满足.该约束品性也称为 Slater 条件.

- Slater约束品性实际上是要求自然定义域$\mathcal{D}$ 的相对内点中存在使得不等式约束严格成立的点, **affine** $\mathcal{D}=\mathbb{R}^n$ 时相对内点就是内点.
- 不等式约束是仿射函数时, Slater条件可以放宽.设前k 个不等式约束是仿射的, 此时Slater约束品性变为:存在$x\in \mathrm{relint}\mathcal{D}$, 满足

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

即对线性不等式约束无需要求其存在严格可行点.

定理

若凸优化问题满足*Slater*条件, 则强对偶原理成立.

- 当$d^* > -\infty$ 时, 对偶问题的最优解可以取到, 即存在对偶可行解(λ^*, ν^*), 满足$g(\lambda^*, \nu^*) = d^* = p^*$.
- 假设集合$\mathcal{D}$ 内部非空(即$\mathrm{relint}\mathcal{D} = \mathbf{int}\,\mathcal{D}$), A 行满秩(否则可以去掉多余的线性等式约束)以及原始问题最优函数值p^* 有限.
- 定义集合

$$
\begin{aligned}
\mathbb{A} = \{(u, v, t) \mid \exists x \in \mathcal{D},\ c_i(x) \le u_i,\ i = 1, 2, \cdots, m, \\
Ax - b = v, f(x) \le t\}. \\
\mathbb{B} = \{(0, 0, s) \in \mathbb{R}^m \times \mathbb{R}^p \times \mathbb{R} \mid s < p^*\}.
\end{aligned}
$$

- 可以证明集合$\mathbb{A}$ 和$\mathbb{B}$ 是不相交的.
- 假设$(u, v, t) \in \mathbb{A} \cap \mathbb{B}$.根据$(u, v, t) \in \mathbb{B}$, 有$u = 0, v = 0$ 和$t < p^*$.
- 由$(u, v, t) \in \mathbb{A}$, 可知$f(x) \le t < p^*$, 这与p^* 是原始问题最优值矛盾.

Slater约束品性与强对偶原理

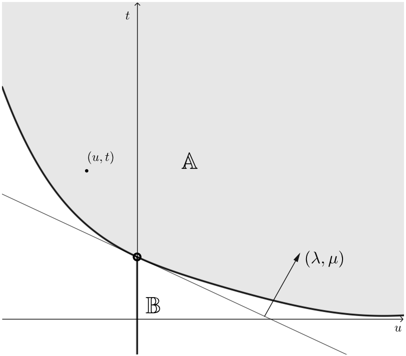


Figure: 集合$\mathbb{A}$ 和$\mathbb{B}$ 在u - t 方向投影的示意图
($\mathbb{A}$ 一般为有内点的凸集, $\mathbb{B}$ 是一条射线且不含点$(0,0,p^*)$)

因为$\mathbb{A}$ 和$\mathbb{B}$ 均为凸集, 由超平面分离定理, 存在$(\lambda,\nu,\mu)\neq 0$ 和α, 使得

$$\lambda^{\mathrm{T}}u+\nu^{\mathrm{T}}v+\mu t\geq\alpha,\quad\forall\,(u,v,t)\in\mathbb{A},$$

$$\lambda^{\mathrm{T}}u+\nu^{\mathrm{T}}v+\mu t\leq\alpha,\quad\forall\,(u,v,t)\in\mathbb{B}.$$

- 我们断言$\lambda \geq 0$ 和$\mu \geq 0$ (否则可以取u_i 和t 为任意大的正实数以及$\nu = 0$, 这会导致$\lambda^{\mathrm{T}} u + \mu t$ 在集合$\mathbb{A}$ 上无下界).
- 同时, 由于$\mu t \leq \alpha$ 对于所有$t < p^*$ 成立, 可得$\mu p^* \leq \alpha$.
- 对任意$x \in \mathcal{D}$, 取$(u, v, t) = (c_i(x), Ax - b, f(x)) \in \mathbb{A}$, 可知

$$\sum_{i=1}^{m} \lambda_i c_i(x) + \nu^{\mathrm{T}}(Ax - b) + \mu f(x) \geq \alpha \geq \mu p^*.$$

- 假设$\mu > 0$, 则

$$L(x, \frac{\lambda}{\mu}, \frac{\nu}{\mu}) \geq p^*.$$

  进一步地, 我们有$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) \geq p^*$, 根据弱对偶性$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) \leq p^*$ 自然成立.因此, 必有$g(\frac{\lambda}{\mu}, \frac{\nu}{\mu}) = p^*$ 成立.说明在此情况下强对偶性满足, 且对偶最优解可以达到.

Slater约束品性与强对偶原理

- 考虑$\mu = 0$ 的情况, 可以从上面得到对于所有的$x \in \mathcal{D}$,

$$\sum_{i=1}^{m} \lambda_i c_i(x) + \nu^{\mathrm{T}}(Ax - b) \geq 0.$$

- 取满足Slater条件的点x_S, 有$\sum_{i=1}^{m} \lambda_i c_i(x_S) \geq 0$.

- 又$c_i(x_S) < 0$ 和$\lambda_i \geq 0$, 我们得到$\lambda = 0$, 上式化为

$$\nu^{\mathrm{T}}(Ax - b) \geq 0, \quad \forall\, x \in \mathcal{D}.$$

  根据$(\lambda, \nu, \mu) \neq 0$ 可知$\nu \neq 0$, 结合A 行满秩可以得到$A^{\mathrm{T}}\nu \neq 0$.由于x_S 是可行解, 我们有$\nu^{\mathrm{T}}(Ax_S - b) = 0$.

- 因为$x_S \in \textbf{int}\,\mathcal{D}$, 则存在点$\tilde{x} = x_S + e \in \mathcal{D}$, 满足$\nu^{\mathrm{T}}(A\tilde{x} - b) < 0$.这与$\nu^{\mathrm{T}}(Ax - b) \geq 0$, $\forall\, x \in \mathcal{D}$ 矛盾.

- 综上所述, Slater条件能保证强对偶性.

- 在定理的证明中, Slater条件保证了$\mu \neq 0$.

一阶充要条件

- 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.
- 而对于凸优化问题，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 (凸优化问题的一阶充要条件)

对于凸优化问题，用a_i 表示矩阵A^{T} 的第i 列，$\partial f, \partial c_i$ 表示次梯度，如果*Slater*条件成立，那么x^*, λ^* 分别是原始，对偶全局最优解当且仅当

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

一阶充要条件:充分性

- 设存在$(\bar{x},\bar{\lambda})$ 满足KKT条件, 我们考虑凸优化问题的拉格朗日函数

$$L(x,\lambda)=f(x)+\sum_{i\in\mathcal{I}}\lambda_i c_i(x)+\sum_{i\in\mathcal{E}}\lambda_i(a_i^{\mathrm{T}}x-b_i).$$

- 当固定$\lambda=\bar{\lambda}$ 时, 注意到$\bar{\lambda}_i\geq 0, i\in\mathcal{I}$ 以及$\bar{\lambda}_i(a_i^{\mathrm{T}}x), i\in\mathcal{E}$ 是线性函数可知$L(x,\bar{\lambda})$ 是关于x 的凸函数.

- 由凸函数全局最优点的一阶充要性可知, 此时$\bar{x}$ 就是$L(x,\bar{\lambda})$ 的全局极小点.根据拉格朗日对偶函数的定义,

$$L(\bar{x},\bar{\lambda})=\inf_{x\in\mathcal{D}}L(x,\bar{\lambda})=g(\bar{\lambda}).$$

- 根据原始可行性条件$A\bar{x}=b$ 以及互补松弛条件$\bar{\lambda}_i c_i(\bar{x})=0, i\in\mathcal{I}$,

$$L(\bar{x},\bar{\lambda})=f(\bar{x})+0+0=f(\bar{x}).$$

- 根据弱对偶原理,

$$L(\bar{x},\bar{\lambda})=f(\bar{x})\geq p^*\geq d^*\geq g(\bar{\lambda})=L(\bar{x},\bar{\lambda})\Rightarrow p^*=d^*,$$

故$\bar{x},\bar{\lambda}$ 分别是原始问题和对偶问题的最优解.

关于充分性的评述

- 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.
- 在充分性部分的证明中，我们没有使用Slater条件，这是因为在证明的一开始假设了KKT点是存在的.
- 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.
- 当Slater条件不满足时，即使原始问题存在全局极小值点，也可能不存在(x^*, λ^*) 满足KKT条件.

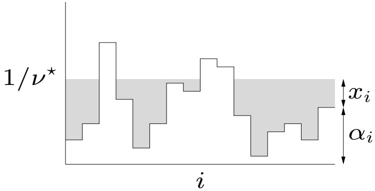
example:water-filling (assume $\alpha_i > 0$)

$$\begin{array}{ll} \min & -\sum_{i=1}^{n} \log(x_i + \alpha_i) \\ \text{s.t.} & x \geq 0, \quad \mathbf{1}^T x = 1 \end{array}$$

x is optimal iff $x \geq 0$, $\mathbf{1}^T x = 1$, and there exist $\lambda \in \mathbb{R}^n$, $\nu \in \mathbb{R}$ such that

$$\lambda \geq 0, \qquad \lambda_i x_i = 0, \qquad \frac{1}{x_i + \alpha_i} + \lambda_i = \nu$$

- if $\nu < 1/\alpha_i$: $\lambda_i = 0$ and $x_i = 1/\nu - \alpha_i$
- if $\nu \geq 1/\alpha_i$: $\lambda_i = \nu - 1/\alpha_i$ and $x_i = 0$
- determine ν from $\mathbf{1}^T x = \sum_{i=1}^{n} \max\{0, 1/\nu - \alpha_i\} = 1$



interpretation

- n patches; level of patch i is at height α_i
- flood area with unit amount of water
- resulting level is $1/\nu^*$

非光滑凸优化实例:基追踪问题

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

根据一般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$s^* \in \mathbb{R}^{2n}, \nu^* \in \mathbb{R}^m$.

基追踪问题最优性条件:直接推导

- 拉格朗日函数: $L(x,\nu) = \|x\|_1 + \nu^{\mathrm{T}}(Ax - b)$.
- x^* 为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$x_i^* = y_i^* - y_{n+i}^*,\ i = 1,2,\cdots,n$, 则

$$[A,-A]y^* = b \implies Ax^* = b.$$

- 对于等式$\mathbf{1} + [A,-A]^{\mathrm{T}}\nu^* - s^* = 0$, 我们有

$$(A^{\mathrm{T}}\nu^*)_i = -1 + s_i^* = 1 - s_{n+i}^*, \quad i = 1,2,\cdots,n.$$

因此, $s_i^* + s_{n+i}^* = 2$.根据互补条件, $y_i^*,\ y_{n+i}^*$ 至少有一个为0.

基追踪问题最优性条件:直接推导

- 若$x_i^* = 0$, 则有$y_i^* = y_{n+i}^* = 0$. 由s_i^*, $s_{n+i}^* \geq 0$ 知$(A^\mathrm{T}\nu^*)_i \in [-1, 1]$.
- 若$x_i^* < 0$, 则有$y_i^* = 0$, $y_{n+i}^* > 0$, 此时有$(A^\mathrm{T}\nu^*)_i = 1$.
- 对于$x_i^* > 0$, 我们有$(A^\mathrm{T}\nu^*)_i = -1$.
- 以上过程均可逆推, 这就证明了利用非光滑凸优化理论推导的最优性条件和利用线性规划推导的最优性条件是等价的.

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

其中$A\in\mathbb{R}^{m\times n}$, $b\in\mathbb{R}^m$ 以及$y\in\mathbb{R}^n$ 为给定的矩阵和向量且A 满秩.

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

$$\begin{cases} x^*-y+A^{\mathrm{T}}\lambda^*=0,\\ Ax^*=b.\end{cases}$$

- 由上述KKT 条件第一式, 等号左右两边同时左乘A 可得

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

- 将λ^* 代回KKT 条件第一式可知

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

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

提纲

切锥

定义 (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.

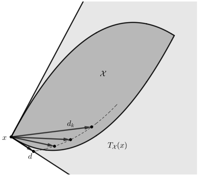


Figure: 不等式约束

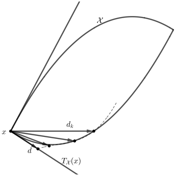


Figure: 等式约束

几何最优性条件

一般优化问题:

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

定理 (几何最优性条件)

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

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

等价于

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

证明

- 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$f(x^*)^\mathrm{T} d < 0$.
- 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.

线性化可行锥

定义 (线性化可行锥)

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

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

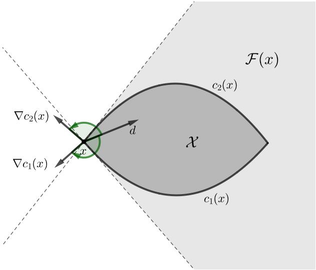


Figure: $\mathbb{R}^2$ 上的约束集合和线性化可行方向锥

线性化可行锥包含切锥

定理 (线性化可行锥包含切锥)

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

证明:不妨设积极集$\mathcal{A}(x) = \mathcal{E} \cup \mathcal{I}$, 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,

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

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

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

同理, 对$i \in \mathcal{I}$, 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

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

结合以上两点, 最终可得到$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$

反例:切锥未必包含线性化可行锥

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

- 则$T_{\mathcal{X}}(3) = \{d \mid d \ge 0\}, \mathcal{F}(3) = \{d : d \ge 0\}$, 于是$T_{\mathcal{X}}(3) = \mathcal{F}(3)$
- 将问题的约束变形为

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

因为可行域不变, 故点$x^* = 3$处, 切锥$T_{\mathcal{X}}(x^*) = \{d : d \ge 0\}$ 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)

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

线性无关约束品性

定义 (线性无关约束品性)

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.

定理

给定任意可行点$x \in \mathcal{X}$, 若在该点LICQ 成立, 则有$T_{\mathcal{X}}(x) = \mathcal{F}(x)$.

- **证明**:不失一般性, 我们假设积极集$\mathcal{A}(x) = \mathcal{E} \cup \mathcal{I}$.记矩阵

$$A(x) = [\nabla c_i(x)]^{\mathrm{T}}_{i \in \mathcal{I} \cup \mathcal{E}}.$$

- 假设集合$\mathcal{A}(x)$ 的元素个数为m, 那么矩阵$A(x) \in \mathbb{R}^{m \times n}$ 并且$\mathrm{rank}(A) = m$.
- 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

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

线性无关约束品性

- 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，$c_i(z_k) \geq 0$，$i \in \mathcal{I}$，$c_i(z_k) = 0$，$i \in \mathcal{E}$，即z_k 为可行点.

- 由泰勒展开得

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

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$T_{\mathcal{X}}(x) \subseteq \mathcal{F}(x)$, 则两集合相同.

定义 (Mangasarian–Fromovitz约束品性)

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

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

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*MFCQ*成立.

- Mangasarian–Fromovitz约束品性是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$T_{\mathcal{X}}(x) = \mathcal{F}(x)$.

线性约束品性

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

定义 (线性约束品性)

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.

- 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, 线性约束品性和LICQ之间没有互相包含的关系.

Karush-Kuhn-Tucker(KKT)条件:引入

- 回顾几何最优性条件:

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

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

$$\left\{ d \;\middle|\; \begin{array}{l} d^{\mathrm{T}} \nabla f(x^*) < 0, \\ d^{\mathrm{T}} \nabla c_i(x^*) = 0, \ i \in \mathcal{E}, \\ d^{\mathrm{T}} \nabla c_i(x^*) \le 0, \ i \in \mathcal{A}(x^*) \cap \mathcal{I} \end{array} \right\} = \varnothing.$$

- 上式依然难以验证, 但可使用Farkas引理进行化简.

Farkas引理

定理 (Farkas引理)

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:

$$\begin{aligned} d^{\mathrm{T}} a_i &= 0,\ i = 1, 2, \cdots, p, \\ d^{\mathrm{T}} b_i &\geq 0,\ i = 1, 2, \cdots, q, \\ d^{\mathrm{T}} c &< 0 \end{aligned}$$

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

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

证明:仅证必要性，若这样的λ_i和μ_i不存在，定义集合

$$S = \{z \mid z = \sum_{i=1}^p \lambda_i a_i + \sum_{i=1}^q \mu_i b_i,\ \lambda_i \in \mathbb{R}, \mu_i \geq 0\}.$$

Farkas引理

- S 是一个闭凸锥.因为$c \notin S$, 由凸集的严格分离超平面定理可知:存在$d \in \mathbb{R}^n, \alpha \in \mathbb{R}$ 使得

$$d^\mathrm{T} c < \alpha < d^\mathrm{T} z, \quad \forall z \in S.$$

- 因为$0 \in S$, 所以

$$d^\mathrm{T} c < \alpha < d^\mathrm{T} 0 = 0.$$

- 任取b_i 与$t \geq 0$, 有$tb_i \in S$, 故$\alpha < td^\mathrm{T} b_i$. 由$t$ 的任意性知$d^\mathrm{T} b_i \geq 0$.
- 同理, 任取$t \in \mathbb{R}$ 与a_i, 有$ta_i \in S$, 故$td^\mathrm{T} a_i < \alpha$.由$t$ 的任意性知$d^\mathrm{T} a_i = 0$.
- 综上, 此时的d 为不等式系统的解.

- 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：

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

  其中$\lambda_i^* \in \mathbb{R}, i \in \mathcal{E}$，$\lambda_i^* \geq 0, i \in \mathcal{A}(x^*) \cap \mathcal{I}$.

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

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

  这恰好对应于拉格朗日函数关于x 的一阶最优性条件.

- **互补松弛条件**:对于任意的$i \in \mathcal{I}$，我们注意到

$$\lambda_i^* c_i(x^*) = 0.$$

  这说明$i \in \mathcal{A}(x^*) \cap \mathcal{I}$ 时乘子$\lambda_i^* = 0$ 或$c_i(x^*) = 0$ 至少出现一种，当两种情况恰好只有一种满足时，我们也称严格互补松弛条件.

假设x^* 是一般优化问题

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

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λ_i^* 使得如下条件成立:

$$\begin{aligned} \text{稳定性条件} \quad & \nabla_x L(x^*, \lambda^*) = \nabla f(x^*) + \sum_{i\in\mathcal{I}\cup\mathcal{E}} \lambda_i^* \nabla c_i(x^*) = 0, \\ \text{原始可行性条件} \quad & c_i(x^*) = 0,\ \forall i \in \mathcal{E}, \\ \text{原始可行性条件} \quad & c_i(x^*) \le 0,\ \forall i \in \mathcal{I}, \\ \text{对偶可行性条件} \quad & \lambda_i^* \ge 0,\ \forall i \in \mathcal{I}, \\ \text{互补松弛条件} \quad & \lambda_i^* c_i(x^*) = 0,\ \forall i \in \mathcal{I}. \end{aligned}$$

KKT条件:注记

- 称满足KKT条件的变量对(x^*, λ^*) 为KKT对.
- 称x^* 为KKT点.
- 如果局部最优点x^* 处$T_{\mathcal{X}}(x^*) \neq \mathcal{F}(x^*)$, 那么$x^*$ 不一定是KKT点.
- KKT条件只是必要的, KKT点不一定是局部最优点.

二阶最优性条件:引入

依旧考虑一般优化问题:

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

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

定义 (临界锥)

设(x^*,λ^*) 是满足KKT条件的KKT对，定义临界锥为

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

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:

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

定理 (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

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

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

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

那么x^* 为问题的一个严格局部极小解.

二阶最优性条件:无约束VS有约束

回顾无约束优化问题的二阶最优性条件:

- 问题: $\min_{x\in\mathbb{R}^n} f(x)$.
- 必要条件:若x^* 是f 的一个局部极小点, 则$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succeq 0$.
- 充分条件:若$\nabla f(x^*) = 0, \nabla^2 f(x^*) \succ 0$, 则$x^*$ 是f 的一个局部极小点.

约束优化问题的二阶最优性条件也要求某种"正定性", 但只需要考虑临界锥$\mathcal{C}(x^*, \lambda^*)$ 中的向量而无需考虑全空间的向量.

有些教材中将其称为"投影半正定性".

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

其拉格朗日函数为

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

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

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

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$\mathcal{C}(x,\lambda) = \mathcal{F}(x)$.
可以计算出其4 个KKT对

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

约束优化的最优性理论:例子

考虑第一个KKT对$y=(2,0,-4)^{\mathrm{T}}$, 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

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

因此y 不是局部最优点.类似地, 对第三个KKT对$z=(0,1,-1)$,

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

对于任意的$d=(d_1,0)$ 且$d_1\neq 0$,

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

因此, z 为一个严格局部最优点.

$$\begin{aligned} \max \quad & \langle C, X\rangle \\ \text{s.t.} \quad & X_{ii} = 1,\ i = 1, 2, \cdots, n \\ & X \succeq 0. \end{aligned}$$

- 拉格朗日函数:

$$L(X, \mu, \Lambda) = \langle C, X\rangle + \sum_{i=1}^{n} \mu_i (X_{ii} - 1) - \mathrm{Tr}(X\Lambda), \quad \Lambda \in \mathcal{S}_+^n,\ \mu \in \mathbb{R}^n.$$

- Slater条件成立, 最优性条件:

$$\begin{cases} C + \mathrm{Diag}(\mu^*) - \Lambda^* = 0, \\ X_{ii}^* = 1,\ i = 1, \cdots, n, \\ X^* \succeq 0, \\ \Lambda^* \succeq 0, \\ \mathrm{Tr}(X^*\Lambda^*) = 0. \end{cases}$$

由于X^* 与Λ^* 的半正定性, $\mathrm{Tr}(X^*\Lambda^*) = 0$ 等价于$X^*\Lambda^* = 0$.

$$\begin{aligned} \max_{Y\in\mathbb{R}^{n\times p}} \quad & \mathrm{Tr}(CYY^{\mathrm{T}}), \\ \text{s.t.} \quad & \mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}. \end{aligned}$$

- 利用$X = YY^{\mathrm{T}}$, $Y \in \mathbb{R}^{p\times n}$ 来减小规模, 此时$YY^{\mathrm{T}} \succeq 0$ 自然满足.
- 可行点$Y = [y_1, y_2, \cdots, y_n]^{\mathrm{T}}$ 处约束为$c_i(Y) = \|y_i\|^2 - 1 = 0, \forall i$.

$$\nabla c_i(Y) = 2[0, \cdots, 0, y_i, 0, \cdots 0]^{\mathrm{T}}.$$

  因为$y_i \neq 0$, 故$\{c_i(Y)\}_{i=1}^n$ 是线性无关的, 即LICQ 成立.
- 拉格朗日函数: $L(Y, \lambda) = \mathrm{Tr}(CYY^{\mathrm{T}}) + \sum_{i=1}^n \lambda_i c_i(Y)$.
- KKT条件:

$$\begin{cases} 2CY - 2[\lambda_1 y_1, \lambda_2 y_2, \cdots, \lambda_n y_n]^{\mathrm{T}} = 0, \\ \mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}. \end{cases}$$

- 令$\Lambda = \mathrm{Diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$, KKT条件第一式可以转换为

$$(C - \Lambda)Y = 0.$$

- 因为该问题只有等式约束, 故临界锥就是切锥, 即

$$\mathcal{C}(Y, \lambda) = \{D \in \mathbb{R}^{n \times p} \mid \mathrm{diag}(Y^{\mathrm{T}} D) = 0\}.$$

- 拉格朗日函数的海瑟矩阵算子形式为

$$\nabla^2_{YY} L(X, \lambda)[D] = 2(C - \Lambda)D.$$

- **必要性**:假设Y 为一个局部最优解, 则存在λ_i, 使得

$$\begin{cases} \langle (C - \Lambda)D, D \rangle \geq 0, & \forall \mathrm{diag}(Y^{\mathrm{T}} D) = 0, \\ \mathrm{diag}(YY^{\mathrm{T}}) = \mathbf{1}, \\ (C - \Lambda)Y = 0. \end{cases}$$

- **充分性:**假设在一点Y处, 存在λ_i, $i = 1, 2, \cdots, n$, 使得

$$\begin{cases} \langle (C - \Lambda)D, D\rangle > 0, \quad \forall \mathrm{diag}(Y^\mathrm{T} D) = 0, \\ \mathrm{diag}(YY^\mathrm{T}) = \mathbf{1}, \\ (C - \Lambda)Y = 0, \end{cases}$$

那么Y为问题的一个严格局部最优解.

- 利用关系式$(C - \Lambda)Y = 0$和约束$\mathrm{diag}(YY^\mathrm{T}) = \mathbf{1}$可以显式求得$\lambda = \mathrm{diag}(CYY^\mathrm{T})$.
- 在这个例子中根据约束的特殊结构, 我们能显式给出乘子λ的表达式.这个性质在一般约束优化问题中是没有的.

提纲

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)	$\nabla^2 f(x^*) \succeq 0$ (必要) $\nabla^2 f(x^*) \succ 0$ (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)	—

总结: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)	$d^{\mathrm{T}}\nabla_{xx}^2 L(x^*,\lambda^*)d \geq 0, \quad \forall d \in \mathcal{C}(x^*,\lambda^*)$ (必要) $d^{\mathrm{T}}\nabla_{xx}^2 L(x^*,\lambda^*)d > 0, \quad \forall d \in \mathcal{C}(x^*,\lambda^*),\ d \neq 0$, (充分)[1]	LICQ[2]
凸问题	KKT 条件(充要)	—	Slater

1. 一般约束优化问题的二阶充分条件不需要LICQ 作为前提.
2. 或其他可推出$T_{\mathcal{X}}(x^*) = \mathcal{F}(x^*)$ 的约束品性.