

优化基础

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

- 1 凸集的定义
- 2 凸函数的定义与性质
- 3 对偶理论
- 4 带约束凸优化问题的最优性理论

凸集的几何定义

在 \mathbb{R}^n 空间中, 经过不同的两点 x_1, x_2 可以确定一条直线, 其方程为

$$y = \theta x_1 + (1 - \theta)x_2, \theta \in \mathbb{R}.$$

特别, 当 $0 \leq \theta \leq 1$ 时, 直线退化为以 x_1, x_2 为端点的线段.

定义

仿射集 如果过集合 C 中的任意两点的直线都在 C 内, 则称 C 为**仿射集**, 即

$$x_1, x_2 \in C \Rightarrow \theta x_1 + (1 - \theta)x_2 \in C, \forall \theta \in \mathbb{R}.$$

例 线性方程组 $Ax = b$ 的解集 \mathcal{X} 是仿射集, 因为 $\forall x_1, x_2 \in \mathcal{X} (x_1 \neq x_2)$ 均满足 $\theta Ax_1 + (1 - \theta)Ax_2 = b$.

反之, 任何仿射集均可表示为某一线性方程组的解集.

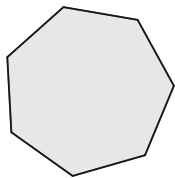
凸集的几何定义

定义

凸集 如果连接集合 C 中的任意两点的线段都在 C 内, 则称 C 为**凸集**, 即

$$x_1, x_2 \in C \Rightarrow \theta x_1 + (1 - \theta)x_2 \in C, \forall 0 \leq \theta \leq 1.$$

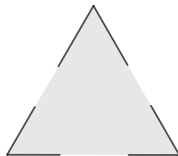
仿射集当然都是凸集.



(a)



(b)



(c)

例 在左图中我们列出了一些凸集、非凸集的例子. 其中(a)为凸集, (b)和(c)均为非凸集.

凸组合和凸包

从凸集中可以引出凸组合和凸包的概念.

定义

凸组合 形如

$$x = \theta_1 x_1 + \theta_2 x_2 + \cdots + \theta_k x_k,$$

$$\theta_1 + \cdots + \theta_k = 1, \theta_i \geq 0, i = 1, \cdots, k.$$

的点称为 x_1, \cdots, x_k 的凸组合.

定义

凸包 集合 S 的所有点的凸组合构成的点集为 S 的凸包, 记为 $\text{conv}S$.

定理

凸集和凸包的关系 若 $\text{conv}S \subseteq S$, 则 S 是凸集; 反之亦然.

上述定理并不显然, 请尝试证明. (提示: 用数学归纳法)

锥组合和凸锥

相比于凸组合和仿射组合, 锥组合不要求系数的和为1, 因此一般而言锥组合都是开放的.

定义

锥组合 形如

$$x = \theta_1 x_1 + \cdots + \theta_k x_k, \theta_i > 0 (i = 1, \cdots, k).$$

的点称为 x_1, \cdots, x_k 的锥组合.

定义

凸锥 若集合 S 中任意点的锥组合都在 S 中, 则称 S 为凸锥.

凸锥的例

例 下图显示了 \mathbb{R}^2 中两点 x_1, x_2 的凸锥. 可见 \mathbb{R}^2 中若两点不与原点 O 共线, 则其形成的凸锥为一个半径无穷的圆的扇形部分.



适当锥

一个凸锥 $K \subseteq \mathbb{R}^n$ 是适当锥, 当其还满足

- K 是闭集;
- K 是实心的, 即 $\text{int}K \neq \emptyset$;
- K 是尖的, 即内部不含有直线: 若 $x \in K, -x \in K$, 则一定有 $x = 0$.

例 非负卦限 $K = \mathbb{R}_+^n = \{x \in \mathbb{R}^n | x_i \geq 0, i = 1, \dots, n\}$ 是适当锥.

例 半正定锥 $K = \mathcal{S}_+^n$ 是适当锥.

例 $[0, 1]$ 上的有限非负多项式

$$K = \{x \in \mathbb{R}^n | x_1 + x_2 t + \dots + x_n t^{n-1} \geq 0, \forall t \in [0, 1]\}$$

是适当锥.

对偶锥

设 K 是一个锥.

定义

对偶锥 令锥 K 为全空间 Ω 的子集, 则 K 的对偶锥为

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

对偶锥是相对于锥 K 定义的, 因此我们知道锥的同时也可以求出对偶锥.
我们将对偶锥为自身的锥称为自对偶锥

例 $K = \mathbb{R}_+^n$ 的对偶锥是它本身, 因此是自对偶锥.

例(请自证) $K = S_+^n$ 的对偶锥是它本身, 因此是自对偶锥.

例(请自证) 锥 $K = \{(x, t) \mid \|x\|_p \leq t, t > 0, p \geq 1\}$ 的对偶锥是

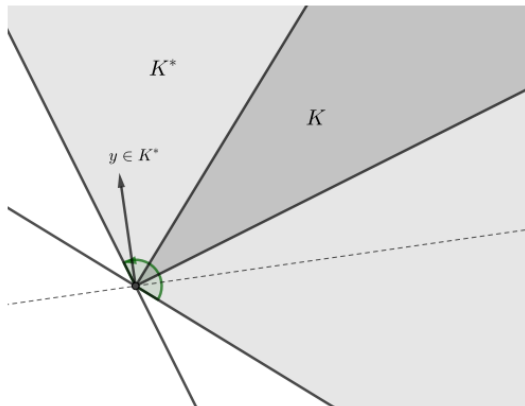
$$K^* = \{(x, t) \mid \|x\|_q \leq t, t > 0, q \geq 1, (p, q) \text{ 共轭}\}.$$

例 上例中二次锥的对偶锥是它本身, 因此是自对偶锥.

对偶锥

例 我们在下图中给出了一个 \mathbb{R}^2 平面上的一个例子. 图中深色区域表示锥 K , 根据对偶锥的定义, K^* 中的向量和 K 中所有向量夹角均为锐角或直角. 因此, 对偶锥 K^* 为图中的浅色区域.

注意, 在这个例子中, K 也为 K^* 的一部分.



提纲

1 凸集的定义

2 凸函数的定义与性质

3 对偶理论

4 带约束凸优化问题的最优性理论

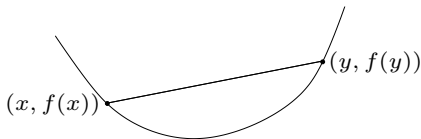
凸函数的定义

定义 (凸函数)

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 为适当函数, 如果 $\text{dom } f$ 是凸集, 且

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y)$$

对所有 $x, y \in \text{dom } f$, $0 \leq \theta \leq 1$ 都成立, 则称 f 是凸函数



- 若 f 是凸函数, 则 $-f$ 是凹函数
- 若对所有 $x, y \in \text{dom } f$, $x \neq y$, $0 < \theta < 1$, 有

$$f(\theta x + (1 - \theta)y) < \theta f(x) + (1 - \theta)f(y)$$

则称 f 是严格凸函数

凸函数判定定理

将函数限制在任意直线上，然后判断对应的一维函数是否是凸的。

定理

$f: \mathbb{R}^n \rightarrow \mathbb{R}$ 是凸函数，当且仅当对每个 $x \in \text{dom } f, v \in \mathbb{R}^n$ ，函数 $g: \mathbb{R} \rightarrow \mathbb{R}$ 是关于 t 的凸函数

$$g(t) = f(x + tv), \quad \text{dom } g = \{t | x + tv \in \text{dom } f\}$$

例： $f(X) = -\log \det X$ 是凸函数，其中 $\text{dom } f = \mathbb{S}_{++}^n$ 。任取 $X \succ 0$ 以及方向 $V \in \mathbb{S}^n$ ，将 f 限制在直线 $X + tV$ (t 满足 $X + tV \succ 0$) 上，那么

$$\begin{aligned} g(t) &= -\log \det(X + tV) = -\log \det X - \log \det(I + tX^{-1/2}VX^{-1/2}) \\ &= -\log \det X - \sum_{i=1}^n \log(1 + t\lambda_i) \end{aligned}$$

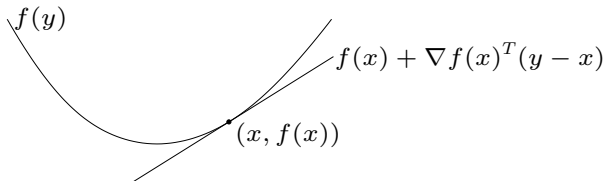
其中 λ_i 是 $X^{-1/2}VX^{-1/2}$ 第 i 个特征值。对每个 $X \succ 0$ 以及方向 V ， g 关于 t 是凸的，因此 f 是凸的。

一阶条件

定理

一阶条件：对于定义在凸集上的可微函数 f ， f 是凸函数当且仅当

$$f(y) \geq f(x) + \nabla f(x)^T(y - x) \quad \forall x, y \in \text{dom } f$$



几何直观： f 的一阶逼近始终在 f 的图像下方

提纲

1 凸集的定义

2 凸函数的定义与性质

3 对偶理论

4 带约束凸优化问题的最优性理论

对偶理论：一般的约束优化问题

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

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

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

- 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:

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

变量 $x \in \mathbb{R}^n$, 最优值为 p^* , 定义域为

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

拉格朗日函数 $L: \mathbb{R}^n \times \mathbb{R}_+^m \times \mathbb{R}^p \rightarrow \mathbb{R}$

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

- 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 $g: \mathbb{R}_+^m \times \mathbb{R}^p \rightarrow [-\infty, +\infty)$

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

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
 - $\text{dom}g = \{(\lambda, \nu) \mid \lambda \geq 0, g(\lambda, \nu) > -\infty\}$, 称其元素为对偶可行解
- 例: 标准形式线性规划及其对偶(参考后面具体推导)

$$\min \quad c^T x$$

$$\text{s.t.} \quad Ax = b$$

$$x \geq 0$$

$$\max \quad -b^T \nu$$

$$\text{s.t.} \quad A^T \nu + c \geq 0$$

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:

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

- 将上式代入 L 得到对偶函数 g , 它是关于 ν 的凹函数

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

- 对偶问题: $\min_{\nu} \frac{1}{4}\nu^T A A^T \nu + b^T \nu$

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

线性规划的对偶

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

- 拉格朗日函数:

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

- 对偶函数:

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

- 对偶问题:

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

线性规划的对偶：保留简单约束推导

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

- 若保留约束 $x \geq 0$, 则拉格朗日函数为

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

- 对偶问题需要将 $x \geq 0$ 添加到约束里：

$$\max_y \left\{ \inf_x b^T y + (c - A^T y)^T x, \quad \text{s.t. } x \geq 0 \right\} \Rightarrow \begin{aligned} \max_y \quad & b^T y, \\ \text{s.t.} \quad & A^T y \leq c. \end{aligned}$$

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.

线性规划对偶问题的对偶

考虑对偶问题: $\max_y b^T y$, s.t. $A^T y \leq c$

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

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

- 因此得到对偶函数

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

- 相应的对偶问题是

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

与原始问题完全等价, 表明线性规划与其对偶问题互为对偶.

线性规划问题的对偶性

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	✓	×	×
unbounded	×	×	✓
infeasible	×	✓	✓

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

$$\min \quad x_1 + 2x_2$$

$$\text{s.t.} \quad x_1 + x_2 = 1$$

$$2x_1 + 2x_2 = 3$$

$$\max \quad p_1 + 3p_2$$

$$\text{s.t.} \quad p_1 + 2p_2 = 1$$

$$p_1 + 2p_2 = 2$$

实例:等式约束下的范数最小化

$$\min_x \|x\| \quad \text{s.t.} \quad Ax = b$$

对偶函数

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

其中 $\|v\|_* = \sup_{\|u\| \leq 1} u^T v$ 是 $\|\cdot\|$ 的对偶范数证明: 利用 $\inf_x (\|x\| - y^T x)$ 在 $\|y\|_* \leq 1$ 时等于0 否则等于 $-\infty$

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

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

对偶问题: $\max_{\nu} b^T \nu$, s.t. $\|A^T \nu\|_* \leq 1$

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

实例：最大割问题

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

- 拉格朗日函数:

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

- 对偶函数:

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

- 对偶问题:

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

拉格朗日对偶与共轭函数

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

对偶函数

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

- 回顾共轭函数的定义 $f^*(y) = \sup_{x \in \text{dom } f} (y^T x - f(x))$
- 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

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

弱对偶性与强对偶性

弱对偶性: $d^* \leq p^*$

- 对凸问题与非凸问题都成立
- 可导出复杂问题的非平凡下界, 例如, SDP问题

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

给出了二路划分问题的一个下界:

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

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

$$\min \quad f_0(y)$$

$$\text{s.t.} \quad Ax + b - y = 0$$

$$\max \quad b^T \nu - f_0^*(\nu)$$

$$\text{s.t.} \quad A^T \nu = 0$$

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: $\min \|Ax - b\|$

等价形式:

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

由 $\|\cdot\|$ 的共轭函数知其对偶函数为:

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

范数逼近问题的对偶

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

ℓ_1 正则化问题的对偶

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

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:

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

● 对偶函数:

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

● 对偶问题:

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

隐式约束

带边界约束的线性规划：原问题与对偶问题

$$\min \quad c^T x$$

$$\text{s.t.} \quad Ax = b$$

$$-\mathbf{1} \leq x \leq \mathbf{1}$$

$$\max \quad -b^T \nu - \mathbf{1}^T \lambda_1 - \mathbf{1}^T \lambda_2$$

$$\text{s.t.} \quad c + A^T \nu + \lambda_1 - \lambda_2 = 0$$

$$\lambda_1 \geq 0, \quad \lambda_2 \geq 0$$

通过隐式化边界约束改写原问题

$$\min \quad f_0(x) = \begin{cases} c^T x & -\mathbf{1} \leq x \leq \mathbf{1} \\ -\infty & \text{其他} \end{cases}$$

$$\text{s.t.} \quad Ax = b$$

对偶函数为

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

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

适当锥与广义不等式

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 $\text{int } K \neq \emptyset$;
- 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:

$$x \prec_K y \iff y - x \in \text{int } K.$$

- 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.

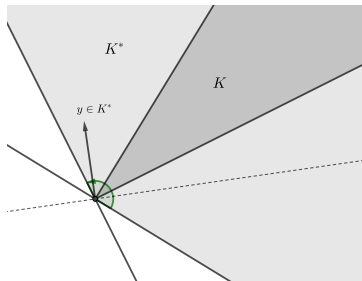


Figure: \mathbb{R}^2 平面上的锥 K 及其对偶锥 K^*

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

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

可以证明

$$\langle X, Y \rangle \geq 0, \forall X \in \mathcal{S}_+^n \iff Y \in \mathcal{S}_+^n,$$

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拉格朗日函数的构造

- 广义不等式约束优化问题:

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

其中 $c_i: \mathbb{R}^n \rightarrow \mathbb{R}^{k_i}$, $k_i \in \mathbb{N}_+$, $i \in \mathcal{I}$ 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 $L(x, \lambda, \nu) \leq f(x)$, $\forall x \in \mathcal{X}$, $\lambda_i \in K_i^*, \nu_i \in \mathbb{R}$.
- 对偶函数 $g(\lambda, \nu) = \inf_{x \in \mathbb{R}^n} L(x, \lambda, \nu)$, 对偶问题为

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

半定规划问题的对偶问题

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

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

- 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:

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

- 对偶问题:

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

半定规划对偶问题的对偶问题

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

- 拉格朗日函数:

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

- 对偶函数:

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

- 对偶问题:

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

最大割对偶问题的对偶问题

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

- 拉格朗日函数:

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

- 对偶函数:

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

- 对偶问题:

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

SOCP/SDP Duality

$$\begin{array}{ll} \text{(P)} & \min \quad c^\top x \\ & \text{s.t.} \quad Ax = b, x_{\mathcal{Q}} \succeq 0 \end{array}$$

$$\begin{array}{ll} \text{(D)} & \max \quad b^\top y \\ & \text{s.t.} \quad A^\top y + s = c, s_{\mathcal{Q}} \succeq 0 \end{array}$$

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

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

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

$$\begin{array}{ll} \inf & (1, -1, 0)x \\ \text{s.t.} & (0, 0, 1)x = 1 \\ & x_Q \succeq 0 \end{array} \qquad \begin{array}{ll} \sup & y \\ \text{s.t.} & (0, 0, 1)^\top y + z = (1, -1, 0)^\top \\ & z_Q \succeq 0 \end{array}$$

- primal: $\min x_0 - x_1$, s.t. $x_0 \geq \sqrt{x_1^2 + 1}$; It holds $x_0 - x_1 > 0$ and $x_0 - x_1 \rightarrow 0$ if $x_0 = \sqrt{x_1^2 + 1} \rightarrow \infty$. 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

$$\begin{aligned} \min \quad & \left\langle \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}, X \right\rangle \\ \text{s.t.} \quad & \left\langle \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, X \right\rangle = 0 \\ & \left\langle \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix}, X \right\rangle = 2 \\ & X \succeq 0 \end{aligned} \quad \begin{aligned} \max \quad & 2y_2 \\ \text{s.t.} \quad & \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} y_1 + \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 2 \end{pmatrix} y_2 \preceq \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

- 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

1 凸集的定义

2 凸函数的定义与性质

3 对偶理论

4 带约束凸优化问题的最优性理论

带约束凸优化问题

- 前述问题都可以写为

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

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

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

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

- 自变量 x 还受约束的限制, 定义可行域

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

- 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:相对内点

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

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

定义(相对内点)

集合 \mathcal{D} 的相对内点集定义为

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

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,但在它的仿射包 $\mathbf{affine}\mathcal{D}$ 中考虑,则 \mathcal{D} 可能有相对内点.

Slater约束品性

定义 (Slater约束品性)

若对凸优化问题

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

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.

- 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 $x \in \text{relint}\mathcal{D}$, 满足

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

即对线性不等式约束无要求其存在严格可行点.

Slater约束品性与强对偶原理

定理

若凸优化问题满足Slater条件, 则强对偶原理成立.

- 这里假设集合 \mathcal{D} 内部非空(即 $\text{relint}\mathcal{D} = \text{int}\mathcal{D}$), 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.
- 需证明的结论: 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 $g(\lambda^*, \nu^*) = d^* = p^*$.
- 定义集合

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

- 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 $f(x) \leq t < p^*$, 这与 p^* 是原始问题最优值矛盾.

Slater约束品性与强对偶原理

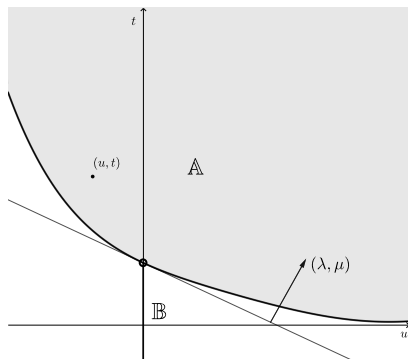


Figure: 集合 \mathbb{A} 和 \mathbb{B} 在 $u-t$ 方向投影的示意图
(\mathbb{A} 一般为有内点的凸集, \mathbb{B} 是一条射线且不含点 $(0, 0, p^*)$)

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

$$\begin{aligned}\lambda^T u + \nu^T v + \mu t &\geq \alpha, \quad \forall (u, v, t) \in \mathbb{A}, \\ \lambda^T u + \nu^T v + \mu t &\leq \alpha, \quad \forall (u, v, t) \in \mathbb{B}.\end{aligned}$$

Slater约束品性与强对偶原理

- 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 $\lambda^T u + \mu t$ 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

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

- 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 $x \in \mathcal{D}$,

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

- 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

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

- 根据 $(\lambda, \nu, \mu) \neq 0$ 可知 $\nu \neq 0$, 结合 A 行满秩有 $A^T \nu \neq 0$. 因为 $x_S \in \text{int } \mathcal{D}$, 则存在 e 使得 $\tilde{x} = x_S + e \in \mathcal{D}$, 且 $\nu^T A e < 0$. 另一方面有 $\nu^T A e = \nu^T A(\tilde{x} - x_S) = \nu^T(A\tilde{x} - b) \geq 0$, 矛盾, $\mu = 0$ 不成立.
- 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：必要性简化证明

假设Slater条件满足， x^*, λ^* 分别是原始，对偶最优解。则有：

$$\begin{aligned} f(x^*) = g(\lambda^*) &= \inf_x f(x) + \sum_{i \in \mathcal{I}} \lambda_i^* c_i(x) + \sum_{i \in \mathcal{E}} \lambda_i^* (a_i^T x - b_i) \\ &\leq f(x^*) + \sum_{i \in \mathcal{I}} \lambda_i^* c_i(x^*) + \sum_{i \in \mathcal{E}} \lambda_i^* (a_i^T x^* - b_i) \\ &\leq f(x^*) \end{aligned}$$

因此，上述两个不等式中的等式成立

- x^* 极小化 $L(x, \lambda^*)$
- $\lambda_i^* c_i(x^*) = 0, i \in \mathcal{I}$ (互补条件, complementary slackness):

$$\lambda_i^* > 0 \implies c_i(x^*) = 0, \quad c_i(x^*) < 0 \implies \lambda_i^* = 0$$

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, $\nabla f, \nabla c_i$ 表示梯度, 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

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

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

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

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

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

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

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

- 当固定 $\lambda = \bar{\lambda}$ 时, 注意到 $\bar{\lambda}_i \geq 0, i \in \mathcal{I}$ 以及 $\bar{\lambda}_i (a_i^T x), i \in \mathcal{E}$ 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.
- 根据KKT条件计算最优化问题显式解:
 - 写出KKT条件
 - 由“稳定性条件”将原问题的解 x 写成关于乘子 λ 的函数
 - 将 x 代入原问题的约束得到关于 λ 的方程, 看是否能求解出 λ
 - 检查上述 x 和 λ 是否完全满足KKT条件
- 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 $\alpha_i > 0$)

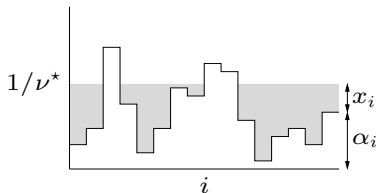
$$\min \quad - \sum_{i=1}^n \log(x_i + \alpha_i)$$

$$\text{s.t.} \quad x \geq 0, \quad \mathbf{1}^T x = 1$$

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

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

- 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 $1/\nu^*$

光滑凸优化实例:仿射空间的投影问题

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

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

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

- 由上述KKT 条件第一式, 等号左右两边同时左乘 A 可得

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

- 将 λ^* 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 $y - A^T (AA^T)^{-1} (Ay - b)$.

非光滑凸优化实例:基追踪问题

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

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

$$\left\{ \begin{array}{l} \mathbf{1} + [A, -A]^T \nu^* - s^* = 0, \\ [A, -A]y^* = b, \\ y^* \geq 0, \\ s^* \geq 0, \\ s^* \odot y^* = 0, \end{array} \right.$$

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: $L(x, \nu) = \|x\|_1 + \nu^T(Ax - b)$.
- x^* 为全局最优解当且仅当存在 $\nu^* \in \mathbb{R}^m$ 使得

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

- 令 $x_i^* = y_i^* - y_{n+i}^*$, $i = 1, 2, \dots, n$, 则

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

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

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

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 $(A^T \nu^*)_i \in [-1, 1]$.
- 若 $x_i^* < 0$, 则有 $y_i^* = 0, y_{n+i}^* > 0$, 此时有 $(A^T \nu^*)_i = 1$.
- 对于 $x_i^* > 0$, 我们有 $(A^T \nu^*)_i = -1$.
- 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.