数据拟合

- 对于问题(**??**)，令$t = \|Ax - b\|_\infty$，则得到等价问题

$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n,\, t \in \mathbb{R}} & t, \\ \text{s.t.} & \|Ax - b\|_\infty \leq t. \end{array}$$

- 利用ℓ_∞ 范数的定义，可以进一步写为

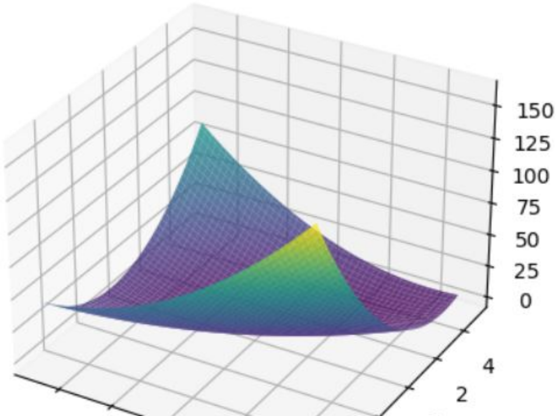
$$\begin{array}{ll} \min\limits_{x \in \mathbb{R}^n,\, t \in \mathbb{R}} & t, \\ \text{s.t.} & -t\mathbf{1} \leq Ax - b \leq t\mathbf{1}, \end{array}$$

这是一个线性规划问题．

二次规划问题（Quadratic Programming, QP）

$$\min \quad f(x_1, x_2) = x_1^2 + 2x_2^2 - 2x_1x_2 + 2x_1 - 6x_2$$

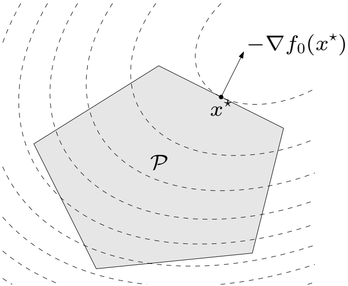
$$\text{s.t.} \quad \begin{aligned} & x_1 + x_2 \geq 2 \\ & x_1 - 2x_2 \leq 4 \\ & 2x_1 + x_2 = 6 \end{aligned}$$



二次规划问题（Quadratic Programming, QP）

$$
\begin{aligned}
\min \quad & \frac{1}{2}x^T P x + q^T x + r \\
\text{s.t.} \quad & Gx \leq h \\
& Ax = b
\end{aligned} \tag{1}
$$

- $P \in \mathcal{S}_+^n$,故目标函数是二次的
- 在一个多面体内最小化一个二次凸问题



二次规划的应用

- 最小二乘问题：
$$\min \|Ax-b\|_2^2 \tag{2}$$
  - 该问题的解析解为$x^* = A^\dagger b$ (其中$A^\dagger$为广义逆)
- 随机线性规划
$$\begin{aligned} \min \quad & \bar{c}^T x + \gamma x^T \Sigma x = \mathbf{E} c^T x + \gamma \mathrm{var}(c^T x) \\ \text{s.t.} \quad & Gx \le h, \quad Ax = b \end{aligned}$$
- c 是随机向量并且均值为$\bar{c}$，方差为Σ
- $c^T x$ 均值为$\bar{c}^T x$,方差为$x^T \Sigma x$
- $\gamma > 0$ 为风险参数,控制预期成本与风险.

带有二次约束的二次规划问题(QCQP)

考虑带有二次约束的二次规划问题：

$$
\begin{aligned}
\min \quad & (1/2)x^T P_0 x + q_0^T x + r_0 \\
\text{s.t.} \quad & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, \quad i = 1, \dots, m \\
& Ax = b
\end{aligned}
$$

- $P_i \in \mathbb{S}_+^n$;即目标函数与约束均为二次凸的
- 如果$P_1, \dots, P_m \in \mathbb{S}_{++}^n$, 则可行域为$m$个椭球与一个仿射集的交集.

- 凸问题的广义不等式约束

$$
\begin{aligned}
\min \quad & f_0(x) \\
\text{s.t.} \quad & f_i(x) \preceq_{K_i} 0, \quad i = 1, \ldots, m \\
& Ax = b
\end{aligned}
$$

  - $f_0 : \mathbb{R}^n \to \mathbb{R}$ 为凸函数; $f_i : \mathbb{R}^n \to \mathbb{R}^{k_i}$ 关于适当锥K_i为K_i-凸的.
  - 与标准凸问题有相同的性质

- 锥形式问题(具有仿射目标函数与约束的特殊情况)

$$
\begin{aligned}
\min \quad & c^T x \\
\text{s.t.} \quad & Fx + g \preceq_K 0 \\
& Ax = b
\end{aligned}
$$

将线性规划问题延伸到了非多面体锥上($K = \mathbb{R}^m_+$)

QP问题的例子

- Least-squares and regression
- Distance between polyhedra

Least-squares and regression

- 问题描述

$$\text{minimize} \quad \|Ax - b\|_2^2$$
$$\|Ax - b\|_2^2 = x^T A^T Ax - 2b^T Ax + b^T b$$

具有解析解$x = A^\dagger b$，$A^\dagger$是A的伪逆

多面体间距离

- 问题描述：R^n上多面体
  $P_1 = \{x \mid A_1 x \prec= b_1\}$和$P_2 = \{x \mid A_2 x \prec= b_2\}$的（Euclid）距离定义为

$$\mathrm{dist}(P_1, P_2) = \inf\{\|x_1 - x_2\| \mid x_1 \in P_1, x_2 \in P_2\}$$
$$P_1 = \{x \mid A_1 x \prec= b_1\} \quad P_2 = \{x \mid A_2 x \prec= b_2\}$$

- QP形式

$$\begin{array}{ll} \text{minimize} & \|x_1 - x_2\|_2^2 \\ \text{subject to} & A_1 x \prec= b_1, A_2 x \prec= b_2 \end{array}$$

这一问题无可行解的充要条件是，其中一个多面体是空的。其最优解为零的充要条件是，多面体相交，这种情况下，最优的x_1和x_2是相等的（并且是交集$\mathcal{P}_1 \cap \mathcal{P}_2$中的点）。否则最优的$x_1$和$x_2$分别在$\mathcal{P}_1$和$\mathcal{P}_2$中，并且是最接近的。（我们将在第8章中更加详细地讨论涉及距离的几何问题。）

- SOCP问题的基本描述：

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subject to} & \|A_i x + b\|_2 \leq c_i^T x + d_i, i = 1, \cdots, m \\ & Fx = g, \end{array}$$

- 二次锥约束条件：

$$\|Ax + b\|_2 \leq c^T x + d$$

当$c_i = 0, i = 1, \cdots, m$, SOCP等同于QCQP;
$A_i = 0, i = 1, \cdots, m$, SOCP退化为线性规划。

SOCP问题的例 – Robust linear programming

- 问题描述：

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & a_i^T x \le b_i \end{array}$$

其中参数c,a_i和b_i含有一些不确定性或变化。

- 例c,b_i为确定的常数，a_i为变量，其范围满足：

$$a_i \in \varepsilon_i = \{\bar{a}_i + P_i u \mid \|u\|_2 \le 1\}$$

- SOCP形式

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & \bar{a}_i^T x + \left\| P_i^T x \right\|_2 \le b_i \end{array}$$

- 几何规划的基本描述：

$$
\begin{aligned}
\text{minimize} \quad & f_0(x) \\
\text{subject to} \quad & f_i(x) \le 1, i = 1, \ldots, m \\
& h_i(x) = 1, i = 1, \ldots, p
\end{aligned}
$$

  其中f_i为正项式，h_i为单项式，$D = R^n_{++}$。

- 单项式与正项式($c > 0, c_k > 0,$)：

$$f(x) = cx_1^{a_1} \cdots x_n^{a_n}$$

$$f(x) = \sum_{k=1}^{K} c_k x_1^{a_{1k}} \cdots x_n^{a_{nk}}$$

- 正项式对于加法，数乘和非负的伸缩变化是封闭的。
- 单项式对于数乘和除是封闭的。
- 一个正项式和一个单项式相乘，其结果为一个正项式。
- 一个正项式除以一个单项式，其结果仍为正项式。

几何规划（一般）不是凸优化问题，但是通过变量代换以及目标、约束函数的转化，GP问题可以转化为凸优化问题。

- 令：

$$y_i = \log x_i$$

- 几何规划的凸形式：

$$\begin{aligned} \text{minimize} \quad & \tilde{f}_0(y) = \log\left(\sum_{k=1}^{K_0} e^{a_{0k}^T y + b_{0k}}\right) \\ \text{subject to} \quad & \tilde{f}_i(y) = \log\left(\sum_{k=1}^{K_i} e^{a_{ik}^T y + b_{ik}}\right) \leq 0, i = 1, \ldots, m \\ & \tilde{h}_i(y) = g_i^T y + h_i = 0, i = 1, \ldots, p \end{aligned}$$

广义不等式约束

- 广义不等式约束的优化问题：

$$\begin{aligned} \text{minimize} \quad & f_0(x) \\ \text{subject to} \quad & f_i(x) \preceq_{K_i} 0, i = 1, \ldots, m \\ & Ax = b \end{aligned}$$

- SOCP的描述：

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & -\left(A_i x + b_i, c_i^T x + d_i\right) \preceq_{K_i} 0, i = 1, \ldots, m \\ & Fx = g \\ & K_i = \left\{(y, t) \in R^{k_i+1} \mid \|y\|_2 \leq t\right\} \end{aligned}$$

半定规划问题的一般形式如下：

$$\begin{aligned} \min \quad & c^{\mathrm{T}}x, \\ \text{s.t.} \quad & x_1A_1 + x_2A_2 + \cdots + x_nA_n + B \preceq 0, \\ & Gx = h, \end{aligned} \tag{3}$$

其中$c \in \mathbb{R}^n, A_i \in \mathcal{S}^m, i = 1, 2, \cdots, m, B \in \mathcal{S}^m, G \in \mathbb{R}^{p\times n}, h \in \mathbb{R}^p$ 为已知的向量和矩阵，$x = (x_1, x_2, \cdots, x_n) \in \mathbb{R}^n$是自变量．

- 半定规划（SDP）是线性规划在矩阵空间中的一种推广．它的目标函数和等式约束均为关于矩阵的线性函数，而它与线性规划不同的地方是其自变量取值于半正定矩阵空间．
- 作为一种特殊的矩阵优化问题，半定规划在某些结构上和线性规划非常相似，很多研究线性规划的方法都可以作为研究半定规划的基础．

类似于线性规划问题，我们考虑半定规划的标准形式

$$\begin{aligned} \min \quad & \langle C, X\rangle, \\ \text{s.t.} \quad & \langle A_1, X\rangle = b_1, \\ & \cdots \\ & \langle A_m, X\rangle = b_m, \\ & X \succeq 0, \end{aligned} \tag{4}$$

和对偶形式：

$$\begin{aligned} \min \quad & -b^{\mathrm{T}} y, \\ \text{s.t.} \quad & y_1 A_1 + y_2 A_2 + \cdots + y_n A_n \preceq C. \end{aligned} \tag{5}$$

形如(3) 式的优化问题都可以转化成(4) 式或者(5) 式的形式.

LP与SDP

$$\text{LP:}\quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & Ax \le b \end{array} \qquad \text{SDP:}\quad \begin{array}{ll} \min & c^T x \\ \text{s.t.} & \operatorname{diag}(Ax - b) \preceq 0 \end{array}$$

SOCP 与SDP

$$\text{SOCP:}\quad \begin{array}{ll} \min & f^T x \\ \text{s.t.} & \|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i = 1, \dots, m \end{array}$$

$$\text{SDP:}\quad \begin{array}{ll} \min & f^T x \\ \text{s.t.} & \begin{bmatrix} (c_i^T x + d_i)I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \dots, m \end{array}$$

二次约束二次规划问题的半定规划松弛

- 考虑二次约束二次规划问题

$$
\begin{aligned}
\min_{x\in\mathbb{R}^n} \quad & x^{\mathrm{T}}A_0x+2b_0^{\mathrm{T}}x+c_0, \\
\text{s.t.} \quad & x^{\mathrm{T}}A_ix+2b_i^{\mathrm{T}}x+c_i\le 0, \quad i=1,2,\cdots,m,
\end{aligned} \tag{6}
$$

其中A_i 为$n\times n$ 对称矩阵．当部分A_i为对称不定矩阵时，问题(6)是NP 难的非凸优化问题．

- 写出问题(6) 的半定规划松弛问题．对任意$x\in\mathbb{R}^n$以及$A\in\mathcal{S}^n$，有恒等式

$$x^{\mathrm{T}}Ax=\mathrm{Tr}(x^{\mathrm{T}}Ax)=\mathrm{Tr}(Axx^{\mathrm{T}})=\left\langle A,xx^{\mathrm{T}}\right\rangle,$$

因此问题(6) 中所有的二次项均可用下面的方式进行等价刻画：

$$x^{\mathrm{T}}A_ix+2b_i^{\mathrm{T}}x+c_i=\left\langle A_i,xx^{\mathrm{T}}\right\rangle+2b_i^{\mathrm{T}}x+c_i.$$

二次约束二次规划问题的半定规划松弛

由上述分析，原始问题等价于

$$\begin{aligned} \min_{x\in\mathbb{R}^n} \quad & \langle A_0, X\rangle + 2b_0^{\mathrm{T}}x + c_0 \\ \text{s.t.} \quad & \langle A_i, X\rangle + 2b_i^{\mathrm{T}}x + c_i \leq 0, \quad i = 1, 2, \cdots, m, \\ & X = xx^{\mathrm{T}}. \end{aligned} \tag{7}$$

进一步地，

$$\begin{aligned} x^{\mathrm{T}}A_i x + 2b_i^{\mathrm{T}}x + c_i &= \left\langle \begin{pmatrix} A_i & b_i \\ b_i^{\mathrm{T}} & c_i \end{pmatrix}, \begin{pmatrix} X & x \\ x^{\mathrm{T}} & 1 \end{pmatrix} \right\rangle, \\ &\overset{\text{def}}{=\!=} \langle \overline{A}_i, \overline{X}\rangle, \quad i = 0, 1, \cdots, m. \end{aligned}$$

接下来将等价问题(7) 松弛为半定规划问题.

- 在问题(7) 中，唯一的非线性部分是约束$X = xx^{\mathrm{T}}$，我们将其松弛成半正定约束$X \succeq xx^{\mathrm{T}}$. 可以证明，$\overline{X} \succeq 0$ 与$X \succeq xx^{\mathrm{T}}$ 是等价的.
- 因此这个问题的半定规划松弛可以写成

$$
\begin{aligned}
\min \quad & \langle \overline{A_0}, \overline{X} \rangle \\
\text{s.t.} \quad & \langle \overline{A_i}, \overline{X} \rangle \leq 0, \quad i = 1, 2, \cdots, m, \\
& \overline{X} \succeq 0, \\
& \overline{X}_{n+1,n+1} = 1.
\end{aligned}
$$

其中“松弛”来源于我们将$X = xx^{\mathrm{T}}$ 替换成了$X \succeq xx^{\mathrm{T}}$.

- 令G为一个无向图，其节点集合为$V=\{1,2,\cdots,n\}$和边的集合为E. 令$w_{ij}=w_{ji}$为边$(i,j)\in E$上的权重，并假设$w_{ij}\geq 0,\ (i,j)\in E$. 最大割问题是找到节点集合$V$的一个子集$S$使得$S$与它的补集$\overline{S}$之间相连边的权重之和最大化.
- 可以将最大割问题写成如下整数规划的形式：令$x_j=1,\ j\in S$和$x_j=-1,\ j\in\overline{S}$，则

$$\begin{aligned} \max \quad & \frac{1}{2}\sum_{i<j}(1-x_ix_j)w_{ij} \\ \text{s.t.} \quad & x_j\in\{-1,1\},\ j=1,2,\cdots,n. \end{aligned} \tag{8}$$

- 在问题(8)中，只有当x_i与x_j不同时，目标函数中w_{ij}的系数非零. 最大割问题是一个离散优化问题，很难在多项式时间内找到它的最优解.

二次约束二次规划问题的半定规划松弛

接下来介绍如何将问题(8) 松弛成一个半定规划问题.

- 令$W=(w_{ij})\in\mathcal{S}^n$，并定义$C=-\frac{1}{4}(\mathrm{Diag}(W\mathbf{1})-W)$为图$G$的拉普拉斯矩阵的$-\frac{1}{4}$倍，则问题(8) 可以等价地写为

$$\begin{aligned}\min\quad & x^\mathrm{T}Cx,\\ \text{s.t.}\quad & x_i^2=1,\ i=1,2,\cdots,n.\end{aligned}$$

  由于目标函数是关于x的二次函数，可将其等价替换为$\langle C,xx^\mathrm{T}\rangle$.
- 接下来令$X=xx^\mathrm{T}$，注意到约束$x_i^2=1$，这意味着矩阵X对角线元素$X_{ii}=1$. 因此利用矩阵形式我们将最大割问题转化为

$$\begin{aligned}\min\quad & \langle C,X\rangle,\\ \text{s.t.}\quad & X_{ii}=1,\ i=1,2,\cdots,n,\\ & X\succeq 0,\ \mathrm{rank}(X)=1.\end{aligned}\tag{9}$$

- 问题(9) 和(8) 是等价的，这是因为$X=xx^\mathrm{T}$ 可以用约束$X\succeq 0$和$\mathrm{rank}(X)=1$等价刻画.

- SDP问题描述：

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & x_1 F_1 + x_2 F_2 + \cdots + x_n F_n + G \preceq 0 \\ & Ax = b \end{aligned}$$

其中$F_i, G \in \mathbf{S}^k$。

- 不等式约束包含多个线性矩阵不等式（LMI）：

$$x_1 \hat{F}_1 + \cdots + x_n \hat{F}_n + \hat{G} \preceq 0, \quad x_1 \tilde{F}_1 + \cdots + x_n \tilde{F}_n + \tilde{G} \preceq 0$$

等价于单个LMI：

$$x_1 \begin{bmatrix} \hat{F}_1 & 0 \\ 0 & \tilde{F}_1 \end{bmatrix} + x_2 \begin{bmatrix} \hat{F}_2 & 0 \\ 0 & \tilde{F}_2 \end{bmatrix} + \cdots + x_n \begin{bmatrix} \hat{F}_n & 0 \\ 0 & \tilde{F}_n \end{bmatrix} + \begin{bmatrix} \hat{G} & 0 \\ 0 & \tilde{G} \end{bmatrix} \preceq 0$$

- LP与等价的SDP：

$$\text{LP}: \text{minimize} \quad c^T x$$
$$\text{subject to} \quad Ax \preceq b$$

$$\text{SDP}: \text{minimize} \quad c^T x$$
$$\text{subject to} \quad diag(Ax - b) \preceq 0$$

- SOCP与等价的SDP：

$$\text{SOCP}: \text{minimize} \quad f^T x$$
$$\text{subject to} \quad \|A_i x + b_i\|_2 \le c_i^T x + d_i, \quad i = 1, \ldots, m$$

$$\text{SDP}: \text{minimize} \quad f^T x$$
$$\text{subject to} \quad \begin{bmatrix} (c_i^T x + d_i) I & A_i x + b_i \\ (A_i x + b_i)^T & c_i^T x + d_i \end{bmatrix} \succeq 0, \quad i = 1, \ldots, m$$

- 特征值最小化：

$$\text{minimize} \quad \lambda_{\max}(A(x))$$

给定$A_i \in \mathbf{S}^k$，$A(x) = A_0 + x_1 A_1 + \cdots + x_n A_n$。

- 等价SDP：

$$\begin{array}{ll} \text{minimize} & t \\ \text{subject to} & A(x) \preceq tI \end{array}$$

其中$x \in \mathbf{R}^n, t \in \mathbf{R}$。

$$\lambda_{\max}(A) \le t \quad \Longleftrightarrow \quad A \preceq tI$$

第5章
对偶问题

优化问题的拉格朗日函数

- 设优化问题：

$$\begin{array}{lll} \text{minimize} & f_0(x), & x \in \mathcal{R}^n \\ \text{subject to} & f_i(x) \leq 0, & i = 1, \ldots, m \\ & h_i(x) = 0, & j = 1, \ldots, p \end{array}$$

- 拉格朗日函数$L : D \times R^n \times R^p \to R$：

$$L(x, \lambda, v) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x)$$

- 对固定的x,拉格朗日函数$L(x, \lambda, v)$为关于λ和v的仿射函数。

拉格朗日对偶函数

- 拉格朗日对偶函数(lagrange dual function)：
$g(\lambda, v) = \inf_{x \in D} L(x, \lambda, v) = \inf_{x \in D} \left(f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} v_i h_i(x)\right)$

  若拉格朗日函数没有下界，则令
$$g(\lambda, v) = -\infty$$

- 拉格朗日对偶函数为凹函数。

- 对$\forall \lambda \succ= 0$ 和$\forall v$，若原最优化问题有最优值p^*，则
$$g(\lambda, v) \leq p^*$$

对偶函数的例子

- 线性方程的最小二乘解
- 标准形式的线性规划问题
- 双向划分问题

线性方程的最小二乘解

- 原问题：

$$\begin{array}{rl} \text{minimize} & x^T x, x \in \mathcal{R}^n \\ \text{subject to} & Ax = b \end{array}$$

- 拉格朗日函数：
$$L(x, v) = x^T x + v^T (Ax - b)$$

- 拉格朗日对偶函数：
$$g(v) = -\frac{1}{4} v^T A A^T v - b^T v$$

$g(v)$是一个二次凹函数，定义域为R^p。利用对偶函数给出原问题下界的性质，从而对于任意的$v \in R^p$,有$-\frac{1}{4} v^T A A^T v - b^T v \leq inf\{x^T x | Ax = b\}$

- 原问题：

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \succeq 0 \end{aligned}$$

- 拉格朗日函数：

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

- 拉格朗日对偶函数：

$$g(\lambda, v) = \begin{cases} -b^T v & A^T v - \lambda + c = 0 \\ -\infty & \text{其他情况} \end{cases}$$

注：对偶函数g只有在$R^m \times R^p$上的一个正常方式仿射子集上取值有限。

双向划分问题

- 原问题：

$$\begin{array}{ll} \text{minimize} & x^T W x, W \in S^n \\ \text{subject to} & x_i^2 = 1, i = 1, \ldots, n \end{array}$$

- 拉格朗日函数：

$$\begin{aligned} L(x, v) &= x^T W x + \textstyle\sum_i^n v_i \left(x_i^2 - 1\right) \\ &= x^T (W + \operatorname{diag}(v)) x - 1^T v \end{aligned}$$

- 拉格朗日对偶函数：

$$g(v) = \begin{cases} -1^T v & W + \operatorname{diag}(v) \succeq 0 \\ -\infty & \text{otherwise} \end{cases}$$

- 弱对偶性: $p^* \geq -\mathbf{1}^T \nu$ 若 $W + \operatorname{diag}(\nu) \succeq 0$
- 取 $\nu = -\lambda_{\min}(W)\mathbf{1}$ 得最优解 p^* 的一个下界的估计, $p^* \geq n\lambda_{\min}(W)$

- 共轭函数
$$f^*(y) = \sup_{x \in \mathrm{dom} f} \left(y^T x - f(x) \right)$$
- 共轭函数与对偶函数存在密切联系
- 具有线性不等式约束和线性等式约束的优化问题：

$$\begin{array}{ll} \text{minimize} & f_0(x) \\ \text{subject to} & Ax \prec = b, Cx = d \end{array}$$

对偶函数：
$$g(\lambda, v) = -b^T \lambda - d^T v - f_0^* \left(-A^T \lambda - C^T v \right)$$
$$\mathrm{dom} g = \left\{ (\lambda, v) \mid -A^T \lambda - C^T v \in \mathrm{dom} f_0^* \right\}$$

实例:等式约束下的范数最小化

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

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 $\|u\| \le 1$, $u^T y = \|y\|_* > 1$:

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

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

熵的最大化

- 原始问题：

$$\begin{aligned} \text{minimize} \quad & \sum_{i=1}^{\text{n}} x_i \log x_i, \quad D = R_+^n \\ \text{subject to} \quad & Ax \prec= b, 1^T x = 1 \end{aligned}$$

- 共轭函数：

$$f_0^*(y) = \sum_{i=1}^n e^{y_i - 1}$$

- 对偶函数：

$$\begin{aligned} g(\lambda, v) \quad &= \quad -b^T\lambda - v - f_0^*\left(-A^T\lambda - 1v\right) \\ &= -b^T\lambda - v - \sum_{i=1}^m e^{-a_i^T\lambda - v - 1} \\ &= -b^T\lambda - v - e^{-v-1}\sum_{i=1}^m e^{-a_i^T\lambda} \end{aligned}$$

- 原始问题：

$$\begin{array}{ll} \text{minimize} & \log\det X^{-1}, \quad D = S^n_{++} \\ \text{subject to} & a_i^T X a_i \le 1, i = 1, \ldots, m \end{array}$$

- 共轭函数：

$$f_0^*(Y) = \log\det(-Y)^{-1} - n$$

- 对偶函数：

$$g(\lambda) = \begin{cases} \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - 1^T\lambda + n & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty & \text{otherwise} \end{cases}$$

拉格朗日对偶问题

- 拉格朗日对偶问题的描述：

$$\begin{aligned} \text{maximize} \quad & g(\lambda, v) \\ \text{subject to} \quad & \lambda \succ= 0 \end{aligned}$$

- 对偶可行域

$$\lambda \succ= 0$$
$$g(\lambda, v) > -\infty$$

- 最优值

$$d^*$$

- 最优解

$$(\lambda^*, v^*)$$

LP问题的对偶问题

- 标准LP问题

$$
\begin{aligned}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & Ax = b, x \succ= 0
\end{aligned}
$$

- 对偶函数

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

- 对偶问题

$$
\begin{aligned}
\text{maximize} \quad & g(\lambda, v) \\
\text{subject to} \quad & \lambda \succ= 0
\end{aligned}
$$

- 等价描述

$$
\begin{aligned}
\text{maximize} \quad & g(\lambda, v) \\
\text{subject to} \quad & A^T \nu - \lambda + c = 0, \lambda \succ= 0
\end{aligned}
$$

LP问题的对偶问题

- 标准LP问题

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \succ= 0 \end{aligned}$$

- 对偶函数

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

- 对偶问题

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & \lambda \succ= 0 \end{aligned}$$

- 等价描述

$$\begin{aligned} \text{maximize} \quad & g(\lambda, \nu) \\ \text{subject to} \quad & A^T \nu - \lambda + c = 0 \\ & \lambda \succ= 0 \end{aligned}$$

弱对偶性

定理

弱对偶性：设原始问题的最优值为p^*,对偶问题的最优值为d^*,则$d^* \leq p^*$成立。

- 最优对偶间隙：optimal duality gap
$$p^* - d^*$$
- 可以利用对偶问题找到原始问题最优解的下界。

强对偶性

定义

强对偶性　设原始问题的最优值为p^*,对偶问题的最优值为d^*。若$d^*=p^*$成立，则称原始问题和对偶问题之间具有强对偶性。

- 强对偶性并不是总是成立的。
- 凸优化问题通常（但并不总是）具有强对偶性。

定理

*Slater*定理：若凸优化问题存在严格可行解，即存在$x\in \mathrm{relint}\ D$，满足

$$f_i(x)<0, i=1,\cdots,m,$$
$$Ax=b$$

则优化问题具有强对偶性。该条件称为*Slater*条件。

Slater约束品性与强对偶原理

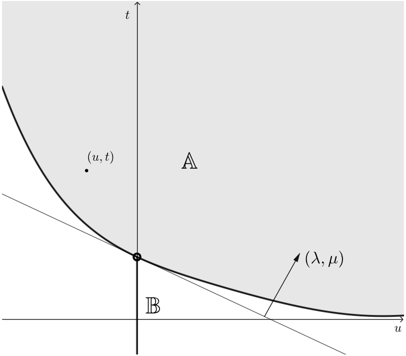
定理

若凸优化问题满足*Slater*条件，则强对偶原理成立.

- 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$\mathrm{relint}\mathcal{D} = \mathcal{D}$)，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.



justification=centering

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

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

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$\lambda^{\mathrm{T}} u + \mu t$ 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$x_S \in \mathcal{D}$, 则存在点e使得$\tilde{x} = x_S + e \in \mathcal{D}$,
且$\nu^{\mathrm{T}}Ae = \nu^{\mathrm{T}}(A\tilde{x} - b) < 0$.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

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

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.

强对偶性

- 弱化的Slater条件：若不等式约束条件的前k个约束函数是仿射，则Slater条件可以弱化为：
  存在$x \in \mathrm{relint}\ D$，满足

$$
\begin{aligned}
& f_i(x) \le 0, i = 1, \cdots, k, \\
& f_i(x) < 0, i = k+1, \cdots, m, \\
& Ax = b
\end{aligned}
$$

线性方程组的最小二乘解

- 原问题：

$$\begin{aligned} \text{minimize} \quad & x^T x, x \in \mathcal{R}^n \\ \text{subject to} \quad & Ax = b \end{aligned}$$

- 对偶问题：

$$\text{maximize} \quad g(\nu) = -\frac{1}{4}\nu^T A A^T \nu - b^T \nu$$

- 具有强对偶性

Lagrange dual of QCQP

$$\begin{array}{ll} \text{minimize} & (1/2)x^T P_0 x + q_0^T x + r_0 \\ \text{subject to} & (1/2)x^T P_i x + q_i^T x + r_i \leq 0, i = 1, \cdots, m \end{array}$$

- 拉格朗日函数：

$$L(x, \lambda) = \frac{1}{2} x^T P(\lambda) x + q(\lambda)^T x + r(\lambda)$$

$$P(\lambda) = P_0 + \sum_{i=1}^m \lambda_i P_i \quad q(\lambda) = q_0 + \sum_{i=1}^m \lambda_i q_i \quad r(\lambda) = r_0 + \sum_{i=1}^m \lambda_i r_i$$

- 对偶函数：

$$g(\lambda) = \inf_x L(x, \lambda) = -\frac{1}{2} q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda)$$

Lagrange dual of QCQP

- 对偶问题：

$$\begin{array}{ll} \text{maximize} & -\frac{1}{2}q(\lambda)^T P(\lambda)^{-1} q(\lambda) + r(\lambda) \\ \text{subject to} & \lambda \succ= 0 \end{array}$$

- Slater条件：存在x，满足

$$(1/2)x^T P_i x + q_i^T x + r_i < 0, i = 1, \cdots, m$$

Entropy maximization

- 原始问题：

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{n} x_i \log x_i, \quad D = R_+^n \\ \text{subject to} & Ax \preceq= b \\ & 1^T x = 1 \end{array}$$

- 对偶函数：

$$g(\lambda, \nu) = -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-a_i^T\lambda}$$

- 对偶问题：

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - e^{-\nu-1} \sum_{i=1}^{n} e^{-\alpha_i^T\lambda} \\ \text{subject to} & \lambda \succ= 0 \end{array}$$

Entropy maximization 简化对偶问题

关于对偶变量ν解析求最大可以简化对偶问题(5.30)。对于任意固定λ，当目标函数对ν的导数为零时，即

$$\nu = \log \sum_{i=1}^{n} e^{-a_i^T \lambda} - 1.$$

目标函数取最大值。将ν的最优值代入对偶问题可以得到

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \log\left(\sum_{i=1}^{n} e^{-a_i^T\lambda}\right) \\ \text{subject to} & \lambda \succ= 0 \end{array}$$

这是一个非负约束的几何规划问题（凸优化问题）。弱化的Slater条件：存在$x \succ 0$，满足

$$Ax \prec= b$$

$$1^T x = 1$$

Minimum volume covering ellipsoid

- 原始问题：

$$\begin{array}{ll} \text{minimize} & \log\det X^{-1}, D = S^n_{++} \\ \text{subject to} & a_i^T X a_i \leq 1, i = 1, \ldots, m \end{array}$$

- 对偶函数：

$$g(\lambda) = \begin{cases} \log\det\left(\sum_{i=1}^m \lambda_i a_i a_i^T\right) - 1^T\lambda + n & \sum_{i=1}^m \lambda_i a_i a_i^T \succ 0 \\ -\infty & otherwise \end{cases}$$

- 对偶问题：

$$\begin{array}{ll} \text{minimize} & \log\det\left(\sum\limits_{i=1}^m \lambda_i a_i a_i^T\right) - 1^T\lambda + n \\ \text{subject to} & \lambda \succ= 0 \end{array}$$

- 弱化的**Slater**条件：存在$X \in S^n_{++}$，满足

$$a_i^T X a_i \leq 1, i = 1, \ldots, m$$

- 弱化的**Slater**条件总成立，因此该优化问题具有强对偶性。

次优解决认证和终止条件

- 对于一优化问题，若x为可行解，(λ, v)为对偶问题可行解，则有如下不等式：

$$f_0(x) - p^* \leq f_0(x) - g(\lambda, v)$$

x为ε次优解，其中

$$\varepsilon = f_0(x) - g(\lambda, v)$$

- 不等式可以用于对次优解的精度估计。

互补松弛条件

设x^*为原始优化问题的最优解，(λ^*, ν^*) 为对偶问题的最优解，若两者具有强对偶性，则

$$\begin{aligned} f_0(x^*) &= g(\lambda^*, \nu^*) \\ &= \inf_x \left(f_0(x) + \sum_i \lambda_i^* f_i(x) + \sum_i \nu_i^* h_i(x) \right) \\ &\leq f_0(x^*) + \sum_i \lambda_i^* f_i(x^*) + \sum_i \nu_i^* h_i(x^*) \\ &\leq f_0(x^*) \end{aligned}$$

所以 $\sum_i \lambda_i^* f_i(x^*) = 0$.

即 $\lambda_i^* f_i(x^*) = 0, i = 1, \ldots, m$.

KKT优化条件

设优化问题中，函数$f_0(x),\ldots,f_m(x), h_0(x),\ldots,h_p(x)$可微。设$x^*$为原始优化问题的最优解，$(\lambda^*,\nu^*)$ 为对偶问题的最优解，且两者具有强对偶性，则(x^*,λ^*,ν^*)满足如下Karush-Kuhn-Tucker(KKT)条件：

- $f_i\left(x^*\right) \leq 0, i=1,\ldots,m.$
- $h_i\left(x^*\right) = 0, i=1,\ldots,p.$
- $\lambda_i^* \geq 0, i=1,\ldots,m.$
- $\lambda_i^* f_i\left(x^*\right) = 0, i=1,\ldots,m.$
- $\nabla f_0\left(x^*\right)+\sum_{i=1}^m \lambda_i^* \nabla f_i\left(x^*\right)+\sum_{i=1}^p \nu_i^* \nabla h_i\left(x^*\right)=0.$

凸优化问题的KKT条件

设原始问题为凸优化问题中，函数$f_0(x), \dots, f_m(x), h_0(x), \dots, h_p(x)$可微，设$(\tilde{x}, \tilde{\lambda}, \tilde{v})$满足**KKT**条件，则$\tilde{x}$为原始问题的最优解，而$(\tilde{\lambda}, \tilde{v})$ 为对偶问题的最优解，且两者具有强对偶性。

- $f_i(\tilde{x}) \leq 0, i = 1, \dots, m.$
- $h_i(\tilde{x}) = 0, i = 1, \dots, p.$
- $\tilde{\lambda}_i \geq 0, i = 1, \dots, m$
- $\tilde{\lambda}_i f_i(\tilde{x}) = 0, i = 1, \dots, m$
- $\nabla f_0(\tilde{x}) + \sum_{i=1}^m \tilde{\lambda}_i \nabla f_i(\tilde{x}) + \sum_{i=1}^p \tilde{\nu}_i \nabla h_i(\tilde{x}) = 0.$

等式约束二次凸问题求极小

例5.1 等式约束凸二次最小化。我们考虑这个问题：

$$\begin{array}{ll} \text{minimize} & (1/2)x^TPx + q^Tx + r \\ \text{subject to} & Ax = b, \end{array}$$

其中，$P \in \mathbf{S}_+^n$。这个问题的KKT条件是

$$Ax^\star = b, \quad Px^\star + q + A^T\nu^\star = 0,$$

我们可以写成

$$\begin{bmatrix} P & A^T \\ A & 0 \end{bmatrix} \begin{bmatrix} x^\star \\ \nu^* \end{bmatrix} = \begin{bmatrix} -q \\ b \end{bmatrix} \tag{10}$$

解决了$m+n$个变量的$m+n$个方程。优化问题等价为方程组求解问题。

原始凸优化问题：

$$\begin{aligned} \text{minimize} \quad & -\sum_i \log(\alpha_i + x_i) \\ \text{subject to} \quad & x \succ= 0 \\ & 1^T x = 1 \end{aligned}$$

KKT条件：

- $x^* \succ= 0, 1^T x^* = 1, \lambda^* \succ= 0,$
- $\lambda_i^* x_i^* = 0, i = 1, \ldots, n,$
- $-1/(\alpha_i + x_i^*) - \lambda_i^* + v^* = 0, i = 1, \ldots, n$

解得：

$$x_i^* = \begin{cases} 1/v^* - \alpha_i & v^* < 1/\alpha_i \\ 0 & v^* \geq 1/\alpha_i \end{cases}$$

其中：

n

凸优化问题的对偶求解

设原始优化问题与对偶问题具有强对偶性，且(λ^*, ν^*) 为对偶问题的最优解，假设$L(x, \lambda^*, \nu^*)$存在唯一的最小解，即

$$\text{minimize} \quad f_0(x) + \sum_{i=1}^{m} \lambda_i^* f_i(x) + \sum_{i=1}^{p} \nu_i^* h_i(x)$$

存在唯一解x^*。若x^*为原问题的可行解，则x^*即为原始问题的最优解；若x^*不是原始问题的可行解，则原始问题不存在最优点，即原问题最优解无法达到。当对偶问题比原问题更容易求解时候，例如对偶问题可以解析求解或者有某些结构更加容易分析，上述方法有意义。

例 5.3　熵的最大化。 考虑熵的最大化问题

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n x_i \log x_i \\ \text{subject to} & Ax \preceq b \\ & \mathbf{1}^T x = 1 \end{array}$$

其中定义域为 $\mathbf{R}^n_{++}$，其对偶问题为

$$\begin{array}{ll} \text{maximize} & -b^T\lambda - \nu - e^{-\nu-1}\sum_{i=1}^n e^{-a_i^T\lambda} \\ \text{subject to} & \lambda \succeq 0 \end{array}$$

(参见第 213 页以及第 220 页。) 假设 Slater 条件的弱化形式成立，即存在 $x \succ 0$ 使得 $Ax \preceq b$ 以及 $\mathbf{1}^T x = 1$，因此强对偶性成立，存在一个对偶最优解 $(\lambda^\star, \nu^\star)$。

设对偶问题已经解出。$(\lambda^\star, \nu^\star)$ 处的 Lagrange 函数为

$$L(x, \lambda^\star, \nu^\star) = \sum_{i=1}^n x_i \log x_i + \lambda^{\star T}(Ax - b) + \nu^\star(\mathbf{1}^T x - 1)$$

它在 $\mathcal{D}$ 上严格凸且有下界，因此有一个唯一解 $x^\star$，

$$x_i^\star = 1/\exp(a_i^T\lambda^\star + \nu^\star + 1), \quad i = 1, \cdots, n,$$

其中 a_i 是矩阵 A 的列向量。如果 $x^\star$ 是原问题可行解，则其必是原问题 (5.13) 的最优解。如果 $x^\star$ 不是原问题可行解，那么我们可以说原问题的最优解不能达到。

Example 5.4 *Minimizing a separable function subject to an equality constraint.* We consider the problem

$$\begin{array}{ll} \text{minimize} & f_0(x) = \sum_{i=1}^n f_i(x_i) \\ \text{subject to} & a^T x = b, \end{array}$$

where $a \in \mathbf{R}^n$, $b \in \mathbf{R}$, and $f_i : \mathbf{R} \to \mathbf{R}$ are differentiable and strictly convex. The objective function is called *separable* since it is a sum of functions of the individual variables $x_1, \ldots, x_n$. We assume that the domain of f_0 intersects the constraint set, *i.e.*, there exists a point $x_0 \in \mathbf{dom}\, f_0$ with $a^T x_0 = b$. This implies the problem has a unique optimal point $x^\star$.

The Lagrangian is

$$L(x, \nu) = \sum_{i=1}^n f_i(x_i) + \nu(a^T x - b) = -b\nu + \sum_{i=1}^n (f_i(x_i) + \nu a_i x_i),$$

which is also separable, so the dual function is

$$\begin{aligned} g(\nu) &= -b\nu + \inf_x \left(\sum_{i=1}^n (f_i(x_i) + \nu a_i x_i) \right) \\ &= -b\nu + \sum_{i=1}^n \inf_{x_i} (f_i(x_i) + \nu a_i x_i) \\ &= -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i). \end{aligned}$$

The dual problem is thus

$$\text{maximize} \quad -b\nu - \sum_{i=1}^n f_i^*(-\nu a_i),$$

with (scalar) variable $\nu \in \mathbf{R}$.

Now suppose we have found an optimal dual variable $\nu^\star$. (There are several simple methods for solving a convex problem with one scalar variable, such as the bisection method.) Since each f_i is strictly convex, the function $L(x, \nu^\star)$ is strictly convex in x, and so has a unique minimizer $\tilde{x}$. But we also know that $x^\star$ minimizes $L(x, \nu^\star)$, so we must have $\tilde{x} = x^\star$. We can recover $x^\star$ from $\nabla_x L(x, \nu^\star) = 0$, *i.e.*, by solving the equations $f_i'(x_i^\star) = -\nu^\star a_i$.

扰动及灵敏度分析

- 扰动问题:

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subject to} & f_i(x) \le u_i, & i = 1, \ldots, m \\ & h_j(x) = v_j, & j = 1, \ldots, p \end{array}$$

- 当$u = 0, v = 0$ 时即为原始问题。
- 若u_i 为正, 则第i 个不等式约束被放宽; 若u_i 为负, 则第i 个不等式约束被收紧。
- 记$p^*(u, v)$ 为扰动问题的最优解。若扰动问题无最优解, 则记

$$p^*(u, v) = \infty$$

全局不等式和局部灵敏度分析

- 设对偶问题存在最优解, 且与原始问题具有强对偶性, 若非扰动问题的最优对偶解为(λ^*, ν^*), 则有:

$$p^*(u,v) \geq p^*(0,0) - {\lambda^*}^T u - {\nu^*}^T v$$

- 当若$p^*(u,v)$ 在$u=0, v=0$ 处可微, 则

$$\lambda_i^* = -\frac{\partial p^*(0,0)}{\partial u_i} \quad \nu_i^* = -\frac{\partial p^*(0,0)}{\partial v_i}$$

全局不等式证明

定理

对于所有的u 和v，我们有

$$p^\star(u,v) \geq p^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v.$$

证明：假设x 扰动问题的任一可行解, 即, $f_i(x) \leq u_i$ 对$i = 1, \ldots, m$成立, 同时$h_i(x) = v_i$ 对$i = 1, \ldots, p$成立. 利用强对偶性有,

$$\begin{aligned} p^\star(0,0) = g\left(\lambda^\star, \nu^\star\right) &\leq f_0(x) + \sum_{i=1}^{m} \lambda_i^\star f_i(x) + \sum_{i=1}^{p} \nu_i^\star h_i(x) \\ &\leq f_0(x) + \lambda^{\star T} u + \nu^{\star T} v. \end{aligned}$$

(第一个不等式利用$g\left(\lambda^\star, \nu^\star\right)$的定义得到; 第二个不等式由$\lambda^\star \succeq 0$得到。)因此，我们有对于扰动问题的任意可行解$x$，有下式成立

$$f_0(x) \geq p^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v,$$

从而全局不等式$p^\star(u,v) \geq p^\star(0,0) - \lambda^{\star T} u - \nu^{\star T} v$成立.

灵敏度分析

灵敏度解释：$p^*(u,v) \geq p^*(0,0) - \lambda^{*T}u - v^{*T}v$

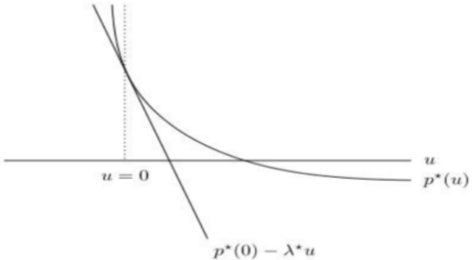


图5.10 具有一个约束$f_1(x) \leqslant u$ 的凸问题的最优值$p^\star(u)$ 的图像，它是u 的函数。当$u=0$时，对应原始未被扰动的问题；当$u<0$ 时，约束加强了，当$u>0$ 时，约束放松了。仿射函数$p^\star(0)-\lambda^\star u$ 是$p^\star$ 的一个下界。

- 如果$\lambda_i^\star$ 比较大，我们加强第i 个约束(即选择$u_i < 0$)，则最优值$p^\star(u,v)$ 必会大幅增加。
- 如果$\nu_i^\star$ 较大且大于零，我们选择$v_i < 0$，或者如果$\nu_i^\star$ 较大且小于零，我们选择$v_i > 0$，在这两种情况下最优值$p^\star(u,v)$ 必会大幅增加。
- 如果$\lambda_i^\star$ 较小，我们放松第i 个约束($u_i > 0$)，那么最优值$p^\star(u,v)$ 不会减小太多。
- 如果$\nu_i^\star$ 较小且大于零，$v_i > 0$ 或者如果$\nu_i^\star$ 较小且小于零，$v_i < 0$，那么最优值$p^\star(u,v)$ 不会减小太多。

局部灵敏度分析

局部灵敏度: 假设$p^\star(u,v)$ 在$(0,0)$ 处是可微的, 然后

$$\lambda_i^\star = -\frac{\partial p^\star(0,0)}{\partial u_i}, \quad \nu_i^\star = -\frac{\partial p^\star(0,0)}{\partial v_i}$$

利用全局不等式结果证明(对于$\lambda_i^\star$)

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t \searrow 0} \frac{p^\star(te_i, 0) - p^\star(0,0)}{t} \geq -\lambda_i^\star$$

$$\frac{\partial p^\star(0,0)}{\partial u_i} = \lim_{t > 0} \frac{p^\star(te_i, 0) - p^\star(0,0)}{t} \leq -\lambda_i^\star$$

因此$p^\star(u)$对于一个不等式的问题也有同样约束约束

广义不等式

本节将Lagrange 对偶理论扩展到具有广义不等式约束的问题，即广义不等式

$$\begin{array}{lll} \text{minimize} & f_0(x) & \\ \text{subjectto} & f_i(x) \preceq_{K_i} 0, & i=1,\cdots,m \\ & h_i(x)=0, & i=1,\cdots,p, \end{array}$$

$\preceq_{K_i}$ 是$\mathbf{R}^{k_i}$上的广义不等式对于问题(5.91) 中的每个广义不等式$f_i(x) \preceq_{K_i} 0$, 引入Lagrange 乘子向量$\lambda_i \in \mathbf{R}^{k_i}$ 并定义相关的Lagrange 函数如下

$$L(x,\lambda,\nu) = f_0(x) + \lambda_1^T f_1(x) + \cdots + \lambda_m^T f_m(x) + \nu_1 h_1(x) + \cdots + \nu_p h_p(x),$$

其中$\lambda = (\lambda_1, \cdots, \lambda_m)$, $\nu = (\nu_1, \cdots, \nu_p)$ 。对偶函数的定义和原问题只有数值不等式的情形一样,对偶函数$g : \mathbf{R}^{k_1} \times \cdots \times \mathbf{R}^{k_m} \times \mathbf{R}^p \to \mathbf{R}$, 被定义为

$$g(\lambda,\nu) = \inf_{x\in\mathcal{D}} L(x,\lambda,\nu) = \inf_{x\in\mathcal{D}} \left(f_0(x) + \sum_{i=1}^m \lambda_i^T f_i(x) + \sum_{i=1}^p \nu_i h_i(x) \right)$$

广义不等式

下界性质：如果$\lambda_i \succeq_{K_i^*} 0$，同时$g(\lambda_1, \ldots, \lambda_m, \nu) \le p^\star$ 证明：如果$\tilde{x}$ 是可行点，同时$\lambda \succeq_{K_i^*} 0$，有

$$\begin{aligned} f_0(\tilde{x}) &\ge f_0(\tilde{x}) + \sum_{i=1}^{m} \lambda_i^T f_i(\tilde{x}) + \sum_{i=1}^{p} \nu_i h_i(\tilde{x}) \\ &\ge \inf_{x \in \mathcal{D}} L(x, \lambda_1, \ldots, \lambda_m, \nu) \\ &= g(\lambda_1, \ldots, \lambda_m, \nu) \end{aligned}$$

最小化所有可行点$\tilde{x}$ 可以得到$p^\star \ge g(\lambda_1, \ldots, \lambda_m, \nu)$
Lagrange对偶优化问题为：

$$\begin{array}{ll} \text{maximize} & g(\lambda_1, \ldots, \lambda_m, \nu) \\ \text{subjectto} & \lambda_i \succeq_{K_i^*} 0, \quad i = 1, \ldots, m \end{array}$$

弱对偶性总是成立，即：$p^\star \ge d^\star$ 强对偶性在广义不等式的情况下同样成立：$p^\star = d^\star$，前提是原问题是凸的且满足合适的约束准则

半定规划

primal SDP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subjectto} & x_1 F_1 + \cdots + x_n F_n + G \preceq 0 \end{array}$$

- 拉格朗日乘数是矩阵$Z \in \mathbf{S}^k$
- 拉格朗日$L(x, Z) = c^T x + \operatorname{tr}\left(Z\left(x_1 F_1 + \cdots + x_n F_n - G\right)\right)$
- 对偶函数

$$g(Z) = \inf_x L(x, Z) = \begin{cases} -\operatorname{tr}(GZ) & \operatorname{tr}\left(F_i Z\right) + c_i = 0, \quad i = 1, \ldots, n \\ -\infty & \text{otherwise} \end{cases}$$

对偶SDP(半定锥自对偶)

$$\begin{array}{ll} \text{maximize} & -\operatorname{tr}(GZ) \\ \text{subjectto} & Z \succeq 0, \quad \operatorname{tr}\left(F_i Z\right) + c_i = 0, \quad i = 1, \ldots, n \end{array}$$

如果primal SDP 是严格可行的($\exists x$ with $x_1 F_1 + \cdots + x_n F_n \prec G$)，$p^\star = d^\star$