《非线性最优化理论与方法》

习题解答

目　录

第1章 引论

习题

1. 证明

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x}) \Longleftrightarrow \min_{\boldsymbol{x}\in\mathbb{R}^n, t\in\mathbb{R}} \{t \mid t - f(\boldsymbol{x}) \geqslant 0\}.$$

解: $\boldsymbol{x}^*$为无约束优化问题$\min\limits_{\boldsymbol{x}\in\mathbb{R}^n} f(\boldsymbol{x})$的最优解, 当且仅当对任意的$\boldsymbol{x}\in\mathbb{R}^n$, $f(\boldsymbol{x})-f(\boldsymbol{x}^*)\geqslant 0$. 所以优化问题

$$\min\{t \mid t - f(\boldsymbol{x}) \geqslant 0\}$$

的最优值为$f(\boldsymbol{x}^*)$. 即

$$f(\boldsymbol{x}^*) = \min\{t \mid t - f(\boldsymbol{x}) \geqslant 0\}.$$

2. 设点列$\{\boldsymbol{x}_k\}$Q-超线性收敛到$\boldsymbol{x}^*$, 则

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} = 1.$$

解: 设点列$\{\boldsymbol{x}_k\}$Q-超线性收敛到$\boldsymbol{x}^*$, 则

$$\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} = 0.$$

从而由三角不等式,

$$\begin{aligned}\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} &\leqslant \lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| + \|\boldsymbol{x}_k - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} \\ &= \lim_{k\to\infty} \frac{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} = 1.\end{aligned}$$

另一方面,

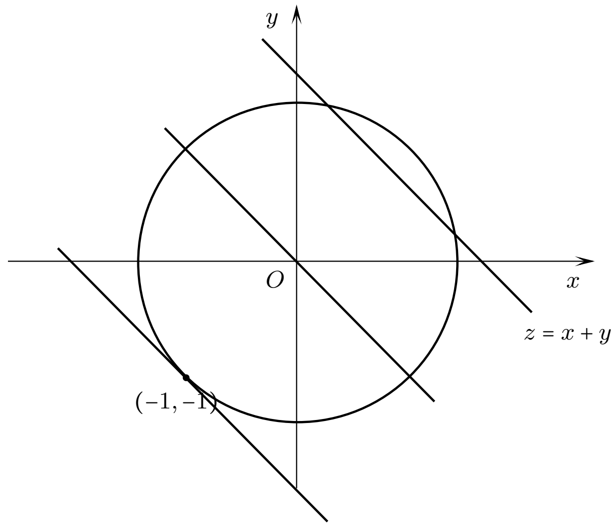
$$\begin{aligned}\lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}_k\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} &\geqslant \lim_{k\to\infty} \frac{\|\boldsymbol{x}_{k+1} - \boldsymbol{x}^*\| - \|\boldsymbol{x}_k - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} \\ &= \lim_{k\to\infty} \frac{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|}{\|\boldsymbol{x}_k - \boldsymbol{x}^*\|} = 1.\end{aligned}$$

两式结合得命题结论. 证毕

3. 用图解法求下述优化问题的最优解:

$$\min\{x_1 + x_2 \mid x_1^2 + x_2^2 \le 2\}$$

解: 根据右图, 该优化问题的最优解为 $x_1^* = x_2^* = -1$. 证毕



4. 设$\boldsymbol{a} \in \mathbb{R}^n$, $\kappa > 0, \|\boldsymbol{a}\| > 1$. 试求解优化问题

$$\begin{aligned} \min \quad & \|\boldsymbol{x}\| \\ \text{s.t} \quad & \|\boldsymbol{x}\| - \boldsymbol{a}^\top \boldsymbol{x} + \kappa \leqslant 0 \end{aligned}$$

解: 由Cauchy-Scwarzy不等式, 对同样长度的$\boldsymbol{x}$, $\boldsymbol{x}$与$\boldsymbol{a}$同向时, $-\boldsymbol{a}^\top \boldsymbol{x}$的值达到最小, $\boldsymbol{x}$的取值范围达到最大. 故该优化问题的最优解满足$\boldsymbol{x} = t\boldsymbol{a}$, 其中, $t > 0$待定. 故上述优化问题化为

$$\begin{aligned} \min \quad & t\|\boldsymbol{a}\| \\ \text{s.t} \quad & t\|\boldsymbol{a}\| - t\|\boldsymbol{a}\|^2 + \kappa \leqslant 0 \end{aligned}$$

即

$$\begin{aligned} \min \quad & t\|\boldsymbol{a}\| \\ \text{s.t} \quad & t \geqslant \frac{\kappa}{\|\boldsymbol{a}\|^2 - \|\boldsymbol{a}\|} \end{aligned}$$

显然, 其最优解为$t^* = \frac{\kappa}{\|\boldsymbol{a}\|^2 - \|\boldsymbol{a}\|}$. 这样, 原问题的最优解为

$$\boldsymbol{x}^* = t^* \boldsymbol{a} = \frac{\kappa \boldsymbol{a}}{\|\boldsymbol{a}\|^2 - \|\boldsymbol{a}\|}.$$

5. 设矩阵$\boldsymbol{\Sigma}$对称正定, $\boldsymbol{a}, \boldsymbol{x}_0 \in \mathbb{R}^n$, $\rho > 0$. 用解析法求解优化问题

$$\begin{aligned} \max \ & \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} \\ \text{s.t.} \ & (\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}} \boldsymbol{\Sigma} (\boldsymbol{x} - \boldsymbol{x}_0) \leqslant \rho \end{aligned}$$

解: 由题设, 矩阵$\boldsymbol{\Sigma}$可写成

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\Sigma}^{\frac{1}{2}}.$$

令$\boldsymbol{y}=\boldsymbol{\Sigma}^{\frac{1}{2}}(\boldsymbol{x}-\boldsymbol{x}_0)$. 则$\boldsymbol{x}=\boldsymbol{x}_0+\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{y}$. 从而原优化问题可写成

$$\begin{aligned}&\max\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_0+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{y}\\&\text{s.t. } \|\boldsymbol{y}\|^2\leqslant\rho\end{aligned}$$

由Cauchy-Schwartz不等式知,

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{y}\leqslant\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{a}\|\|\boldsymbol{y}\|,$$

且在$\boldsymbol{y}=\frac{\rho}{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{a}\|}\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{a}$时等号成立, 即目标函数取最大值. 由此, 上述优化问题的最优值为

$$\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_0+\boldsymbol{a}^{\mathrm{T}}\boldsymbol{\Sigma}^{-\frac{1}{2}}\frac{\rho\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{a}}{\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{a}\|}=\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_0+\rho\|\boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{a}\|$$

6. 设$\boldsymbol{a}\in\mathbb{R}^n,\rho>0$, 矩阵$\Sigma_0\succ\mathbf{0}$(对称正定). 试用解析法求解如下优化问题

$$\begin{aligned}&\max_{\Sigma\succ\mathbf{0}}\ \boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a}\\&\text{s.t. } \|\Sigma-\Sigma_0\|_{\mathrm{F}}\leqslant\rho\end{aligned}$$

解: 令$\Sigma=\Sigma_0+\Delta\Sigma$. 则上述约束优化问题可转化为

$$\begin{aligned}&\max\ \boldsymbol{a}^{\mathrm{T}}\Sigma_0\boldsymbol{a}+\boldsymbol{a}^{\mathrm{T}}\Delta\Sigma\boldsymbol{a}\\&\text{s.t. } \|\Delta\Sigma\|_{\mathrm{F}}\leqslant\rho\end{aligned}$$

略去常数项得

$$\begin{aligned}&\max\ \boldsymbol{a}^{\mathrm{T}}\Delta\Sigma\boldsymbol{a}\\&\text{s.t. } \|\Delta\Sigma\|_{\mathrm{F}}\leqslant\rho\end{aligned}$$

利用$\boldsymbol{a}^{\mathrm{T}}\Sigma\boldsymbol{a}=\langle\Sigma,\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\rangle$和Cauchy-Schwartz不等式,

$$\boldsymbol{a}^{\mathrm{T}}\Delta\Sigma\boldsymbol{a}=\langle\Delta\Sigma,\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\rangle\leqslant\|\Delta\Sigma\|_{\mathrm{F}}\|\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}\|_{\mathrm{F}}.$$

这里, $\langle\boldsymbol{A},\boldsymbol{B}\rangle$表示同阶矩阵$\boldsymbol{A},\boldsymbol{B}$对应元素的乘积之和.

由于$\|\Delta\Sigma\|_{\mathrm{F}}\leqslant\rho$. 显然, 在$\Delta\Sigma$与矩阵$\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}$同向时, 不等式取等号, 且在$\|\Delta\Sigma\|_{\mathrm{F}}=\rho$时, 不等式左侧取最大值. 考虑到约束条件, 上述优化问题的最优解为

$$\Sigma^*=\Sigma_0+\Delta\Sigma=\Sigma_0+\rho\frac{\boldsymbol{a}\boldsymbol{a}^{\mathrm{T}}}{\|\boldsymbol{a}\|^2}$$

进而上述优化问题的最优值为

$$\boldsymbol{a}^{\mathrm{T}}\Sigma^*\boldsymbol{a}=\boldsymbol{a}^{\mathrm{T}}\Sigma_0\boldsymbol{a}+\rho\boldsymbol{a}^{\mathrm{T}}\boldsymbol{a}.$$

第2章 凸分析基础

习题

1. (1) 设$x,y\geqslant 0$, $\alpha\in(0,1)$. 证明$x^\alpha y^{1-\alpha}\leqslant\alpha x+(1-\alpha)y$.

(2) 设$x_i>0, i=1,2,\cdots,n$. 证明

$$\sum_{i=1}^n x_i\sum_{i=1}^n\frac{1}{x_i}\geqslant n^2,$$

且等号成立的充分必要条件是所有的x_i都相等.

解: (1) 设$f(x)=-\ln x, x>0$. 容易验证, $f(x)$关于$x\in(0,\infty)$为凸函数. 所以, 对任意的$x,y\in(0,\infty)$和$\alpha\in(0,1)$,

$$f(\alpha x+(1-\alpha)y)\leqslant\ \alpha f(x)+(1-\alpha)f(y).$$

而这依次等价于

$$\begin{aligned}&-\ln(\alpha x+(1-\alpha)y)\leqslant-\alpha\ln x-(1-\alpha)\ln y\\ \Leftrightarrow\ &\alpha\ln x+(1-\alpha)\ln y\leqslant\ln(\alpha x+(1-\alpha)y)\\ \Leftrightarrow\ &\ln x^\alpha+\ln y^{1-\alpha}\leqslant\ln(\alpha x+(1-\alpha)y).\end{aligned}$$

利用对数函数的单调性得,

$$x^\alpha y^{1-\alpha}\leqslant\alpha x+(1-\alpha)y.$$

命题结论得证.

(2) 取$f(x)=x+\frac{1}{x}, x>0$. 则$f(x)$关于$x\in(0,\infty)$为凸函数. 从而对任意的$x_1,x_2,\cdots,x_n\in(0,\infty)$,

$$f\left(\frac{x_1+x_2+\cdots+x_n}{n}\right)\leqslant\ \frac{f(x_1)+f(x_2)+\cdots+f(x_n)}{n}.$$

而这依次等价于

$$
\begin{aligned}
& \sum_{j=1}^{n} x_j \Big/ n + \frac{n}{\sum\limits_{j=1}^{n} x_j} \leqslant \sum_{j=1}^{n}\left(x_j + \frac{1}{x_j}\right)\Big/ n \\
\Leftrightarrow\ & \frac{1}{n}\sum_{j=1}^{n} x_j + \frac{n}{\sum\limits_{j=1}^{n} x_j} \leqslant \frac{1}{n}\sum_{j=1}^{n} x_j + \frac{1}{n}\sum_{j=1}^{n}\frac{1}{x_j} \\
\Leftrightarrow\ & \frac{n}{\sum\limits_{j=1}^{n} x_j} \leqslant \frac{1}{n}\sum_{j=1}^{n}\frac{1}{x_j} \\
\Leftrightarrow\ & \sum_{i=1}^{n} x_i \sum_{i=1}^{n}\frac{1}{x_i} \geq n^2.
\end{aligned}
$$

其中，“=”成立的充分必要是$x_1 = x_2 = \cdots = x_n$. 证毕

2. 用Cauchy-Schwarz不等式证明

$$\|x\|_1 \leqslant \sqrt{n}\|x\|_2, \quad \forall\ x \in \mathbb{R}^n.$$

解: 对 $x = (x_1; x_2; \cdots; x_n)$, $\mathbf{1} = (1; 1; \cdots; 1)$, $|x| = (|x_1|; |x_2|; \cdots; |x_n|)$, 由Cauchy-Schwarz不等式,

$$\|x\|_1 = \mathbf{1}^{\mathrm{T}}|x| \leqslant \|\mathbf{1}\|_2\|x\|_2 = \sqrt{n}\|x\|_2.$$ 证毕

3. 证明函数$f: \mathbb{R}^n \to \mathbb{R}$为凸函数的充分必要条件是其上图

$$\mathrm{epi}f = \{(x, \alpha) \mid x \in \mathbb{R}^n, \alpha \geqslant f(x)\}$$

为凸集.

解: 必要性. 对任意的 $(x_1, \alpha_1), (x_2, \alpha_2) \in \mathrm{epi}f$ 和$\lambda \in [0, 1]$,

$$f(x_1) \leqslant \alpha_1, \qquad f(x_2) \leqslant \alpha_2.$$

由 $f(x)$ 为凸函数知

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda)f(x_2) \leqslant \lambda\alpha_1 + (1-\lambda)\alpha_2.$$

从而,

$$\lambda(x_1, \alpha_1) + (1-\lambda)(x_2, \alpha_2) = (\lambda x_1 + (1-\lambda)x_2, \lambda\alpha_1 + (1-\lambda)\alpha_2) \in \mathrm{epi}f.$$

充分性：对任意的 $x_1, x_2 \in \mathbb{R}^n$ 和 $\lambda \in [0,1]$, 显然有

$$(x_1, f(x_1)), (x_2, f(x_2)) \in \mathrm{epi} f.$$

由上图是凸集, 则

$$(\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda) f(x_2)) = \lambda(x_1, f(x_1)) + (1-\lambda)(x_2, f(x_2)) \in \mathrm{epi} f.$$

也就是

$$f(\lambda x_1 + (1-\lambda)x_2) \leqslant \lambda f(x_1) + (1-\lambda) f(x_2).$$

证毕

4. 讨论由$\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m \in \mathbb{R}^n$所生成的凸包、仿射包、子空间之间的关系.

解: 对向量组$\boldsymbol{a}_1, \boldsymbol{a}_2, \cdots, \boldsymbol{a}_m \in \mathbb{R}^n$, 其多面胞为

$$A_1 = \mathrm{co}\ \{a_1, a_2, \cdots, a_m\} = \{\sum_{j=1}^m \lambda_j a_j \mid \sum_{j=1}^m \lambda_j = 1, \lambda_j \geqslant 0, j = 1, 2, \cdots, m\}.$$

仿射集为

$$A_2 = \mathrm{aff}\ \{a_1, a_2, \cdots, a_m\} = \{\sum_{j=1}^m \lambda_j a_j \mid \sum_{j=1}^m \lambda_j = 1, \lambda_j \in \mathbb{R}, j = 1, 2, \cdots, m\}.$$

子空间为

$$A_3 = \mathrm{span}\ \{a_1, a_2, \cdots, a_m\} = \{\sum_{j=1}^m \lambda_j a_j \mid \lambda_j \in \mathbb{R}, j = 1, 2, \cdots, m\}.$$

它们三者之间的关系如下: 多面胞A_1为有界闭凸集, 仿射集A_2为与子空间结构一样的一个集合, 它包含多面胞A_1. 如果该向量组线性无关, 则该向量组生成的仿射子空间的维数为$m-1$, 他们生成的子空间的维数为m.

5. 设$\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{b} \in \mathbb{R}^n$. 证明下述两线性系统恰好一个有解.

(1) $\boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{0},\ \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} > 0,\ \boldsymbol{x} \geqslant \boldsymbol{0}$;

(2) $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y} \geqslant \boldsymbol{b},\ \boldsymbol{y} \leqslant \boldsymbol{0}$.

解: 显然, 系统(1)式等价于

$$\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{I} \end{pmatrix} \boldsymbol{x} \geqslant \boldsymbol{0}, \quad \boldsymbol{b}^{\mathrm{T}}\boldsymbol{x} > 0.$$

由Farkas引理,

Farkas**引理** 设$\boldsymbol{A}\in\mathbb{R}^{n\times m}$, $\boldsymbol{b}\in\mathbb{R}^n$, 则下述两系统恰有一个有解.

(1) $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}\geqslant\boldsymbol{0}$, $\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}<0$;

(2) $\boldsymbol{A}\boldsymbol{y}=\boldsymbol{b}$, $\boldsymbol{y}\geqslant\boldsymbol{0}$.

上述系统有解等价于下述系统无解

$$(\boldsymbol{A}^{\mathrm{T}}\quad \boldsymbol{I})\boldsymbol{y}=-\boldsymbol{b},\ \boldsymbol{y}\geqslant\boldsymbol{0}.$$

也就是系统$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}\leqslant-\boldsymbol{b}$, $\boldsymbol{y}\geqslant\boldsymbol{0}$无解. 也即系统$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}\geqslant\boldsymbol{b}$, $\boldsymbol{y}\leqslant\boldsymbol{0}$无解. 证毕

6. 设$\mathcal{S}$为$\mathbb{R}^n$的闭凸集. 若存在$\boldsymbol{x}_0\in\mathcal{S}$和$\boldsymbol{d}\in\mathbb{R}^n$, 使对任意的$\alpha\geqslant0$, 均有$\boldsymbol{x}_0+\alpha\boldsymbol{d}\in\mathcal{S}$, 则对任意的$\boldsymbol{x}\in\mathcal{S}$和$\beta\geqslant0$, 均有$\boldsymbol{x}+\beta\boldsymbol{d}\in\mathcal{S}$.

解: 由题设, 对任意的$\boldsymbol{x}\in\mathcal{S}$, $\lambda\in(0,1)$ 和$t,\alpha>0$,

$$\lambda\boldsymbol{x}_0+(1-\lambda)\boldsymbol{x}+t\lambda\alpha\boldsymbol{d}\in\mathcal{S}.$$

令$\lambda=\frac{1}{\alpha}\to0$, 并利用集合的闭性得, $\boldsymbol{x}+t\boldsymbol{d}\in\mathcal{S}$. 证毕

7. 设函数函数$f,g:\mathbb{R}^n\to\mathbb{R}$为二次齐次函数. 则集合

$$\mathcal{M}=\{(f(\boldsymbol{x}),g(\boldsymbol{x}))\mid\boldsymbol{x}\in\mathbb{R}^n\}$$

为凸集.

解: 取$\boldsymbol{u}=(\boldsymbol{u}_f,\boldsymbol{u}_g),\boldsymbol{v}=(\boldsymbol{v}_f,\boldsymbol{v}_g)$. 若这两点的连线过原点, 则由函数的齐次性得, $[\boldsymbol{u},\boldsymbol{v}]\subset\mathcal{M}$. 命题结论得证. 否则, 存在$\boldsymbol{x}_{\boldsymbol{u}},\boldsymbol{x}_{\boldsymbol{v}}\in\mathbb{R}^n$使得

$$\boldsymbol{u}_f=f(\boldsymbol{x}_u),\quad \boldsymbol{u}_g=g(\boldsymbol{x}_u);\qquad \boldsymbol{v}_f=f(\boldsymbol{x}_v),\quad \boldsymbol{v}_g=g(\boldsymbol{x}_v).$$

为证命题结论, 只需证存在$\lambda\in[0,1]$使得$\boldsymbol{x}_\lambda$满足

$$(f(\boldsymbol{x}_\lambda),g(\boldsymbol{x}_\lambda))=(1-\lambda)u+\lambda v.$$

为此令

$$\boldsymbol{x}_\lambda=\rho(\boldsymbol{x}_{\boldsymbol{u}}\cos\theta+\boldsymbol{x}_{\boldsymbol{v}}\sin\theta),$$

其中$\theta \in [-\frac{\pi}{2}, \frac{\pi}{2}], \rho > 0$待定.

利用函数的二次齐次性, 将其代入上式得

$$\rho^2 f(\boldsymbol{x_u} \cos\theta + \boldsymbol{x_v} \sin\theta) = (1-\lambda)\boldsymbol{u}_f + \lambda \boldsymbol{v}_f,$$

$$\rho^2 g(\boldsymbol{x_u} \cos\theta + \boldsymbol{x_v} \sin\theta) = (1-\lambda)\boldsymbol{u}_g + \lambda \boldsymbol{v}_g.$$

两式相除, 解得

$$\lambda(\theta) = \frac{\boldsymbol{u}_g f(\boldsymbol{x_u} \cos\theta + \boldsymbol{x_v} \sin\theta) - \boldsymbol{u}_f g(\boldsymbol{x_u} \cos\theta + \boldsymbol{x_v} \sin\theta)}{(\boldsymbol{u}_g - \boldsymbol{v}_g) f(\boldsymbol{x_u} \cos\theta + \boldsymbol{x_v} \sin\theta) - (\boldsymbol{u}_f - \boldsymbol{v}_f) g(\boldsymbol{x_u} \cos\theta + \boldsymbol{x_v} \sin\theta)}$$

由函数f, g的的二次齐次性, 上式中的分母关于$\cos\theta$和$\sin\theta$为二次函数, 因此可记为

$$T(\theta) = a\cos^2\theta + b\sin^2\theta + 2c\cos\theta\sin\theta.$$

由$T(0) = T(\pm\frac{\pi}{2}) = \boldsymbol{v}_f \boldsymbol{u}_g - \boldsymbol{v}_g \boldsymbol{u}_f \triangleq d$得, $a = b$. 不妨设$d = \boldsymbol{v}_f \boldsymbol{u}_g - \boldsymbol{v}_g \boldsymbol{u}_f > 0$. 则

$$T(\theta) = d + 2c\cos\theta\sin\theta.$$

显然, 若$c \geqslant 0$, 则

$$T(\theta) = d + 2c\cos\theta\sin\theta > 0, \quad \forall \theta \in [0, \frac{\pi}{2}];$$

否则若$c < 0$, 则

$$T(\theta) = d + 2c\cos\theta\sin\theta > 0, \quad \forall \theta \in [-\frac{\pi}{2}, 0].$$

不妨设前一种情况成立. 则$\lambda(\theta)$在$[0, \frac{\pi}{2}]$上有定义连续. 从而由$\lambda(0) = 0$和$\lambda(\frac{\pi}{2}) = 1$得存在$\theta \in [0, \frac{\pi}{2}]$使得$\lambda(\theta) \in [0, 1]$满足题设条件, 结论得证.

8.(S-引理) 对二次函数$f, g : \mathbb{R}^n \to \mathbb{R}$, 设存在$\bar{\boldsymbol{x}} \in \mathbb{R}^n$使得$g(\bar{\boldsymbol{x}}) < 0$. 试借助上题结论和凸集分离定理证明下述两结论等价

(1) 不存在$\boldsymbol{x} \in \mathbb{R}^n$满足$f(\boldsymbol{x}) < 0,\ g(\boldsymbol{x}) \leqslant 0$;

(2) 存在$\mu \geqslant 0$使对任意的$\boldsymbol{x} \in \mathbb{R}^n$, 成立$f(\boldsymbol{x}) + \mu g(\boldsymbol{x}) \geqslant 0$.

解: 显然, 若(2)成立, 则(1)成立. 故只需证明(1)⇒(2).

首先设f,g为二次齐次函数. 则由上题结论, 集合

$$\mathcal{M}=\{(f(\boldsymbol{x}),g(\boldsymbol{x}))\mid \boldsymbol{x}\in\mathbb{R}^n\}$$

为凸集. 由题设, 集合$\mathcal{M}$与凸集$\mathcal{C}=\{(u_1,u_2)\mid u_1<0,u_2\leqslant 0\}$不相交. 由凸集分离定理, 存在非零数$\lambda_1\lambda_2$使得

$$\lambda_1u_1+\lambda_2u_2\leqslant 0,\quad \forall(u_1,u_2)\in\mathcal{C};\qquad \lambda_1 f(\boldsymbol{x})+\lambda_2 g(\boldsymbol{x})\geqslant 0,\quad \forall\boldsymbol{x}\in\mathbb{R}^n.$$

由于$(-1,0),(-\varepsilon,-1)\in\mathcal{C}$, 故$\lambda_1,\lambda_2\geqslant 0$. 进一步, 由题设, 存在$\bar{\boldsymbol{x}}\in\mathbb{R}^n$使得$g(\bar{\boldsymbol{x}})<0$, 推知$\lambda_1>0$. 由此, 令$\lambda=\frac{\lambda_2}{\lambda_1}$, 即由上式得命题结论.

下讨论f,g为二次非齐次函数的情形. 在此不妨设$\bar{\boldsymbol{x}}=0$. 否则可通过线性平移变换实现. 不妨设函数形式为

$$f(\boldsymbol{x})=\boldsymbol{x}^\top\boldsymbol{A}_f\boldsymbol{x}+b_f^\top\boldsymbol{x}+c_f,\quad g(\boldsymbol{x})=\boldsymbol{x}^\top\boldsymbol{A}_g\boldsymbol{x}+b_g^\top\boldsymbol{x}+c_g.$$

由此定义二次齐次函数

$$\tilde{f}(\boldsymbol{x},\tau)=\boldsymbol{x}^\top\boldsymbol{A}_f\boldsymbol{x}+\tau b_f^\top\boldsymbol{x}+c_f\tau^2,\quad \tilde{g}(\boldsymbol{x},\tau)=\boldsymbol{x}^\top\boldsymbol{A}_g\boldsymbol{x}+\tau b_g^\top\boldsymbol{x}+c_g\tau^2.$$

容易验证, $\tilde{f},\tilde{g}$满足(1), 且$\tilde{g}(0,1)<0$. 这样由刚证明的结论, 得(2)成立.

9. 计算点$\boldsymbol{x}\in\mathbb{R}^n$到闭凸集$S\subset\mathbb{R}^n$的距离函数

$$d_S(\boldsymbol{x})=\min_{\boldsymbol{y}\in S}\|\boldsymbol{y}-\boldsymbol{x}\|$$

的次梯度.

解: 容易验证, $d_S(\boldsymbol{x})$关于$\boldsymbol{x}\in S$为凸函数. 下分情况讨论.

(1) 若$\boldsymbol{x}\notin S$, 则$d_S(\boldsymbol{x})>0$. 显然, $d_S(\boldsymbol{x})$为点$\boldsymbol{x}$到S上的投影之间的距离. 由$d_S(\boldsymbol{x})$的连续性, $d_S(\boldsymbol{x})$在该点连续可微. 从而由链式法则,

$$\nabla d_S(\boldsymbol{x})=\nabla\sqrt{d_S^2(\boldsymbol{x})}=\frac{\nabla d_S^2(\boldsymbol{x})}{2d_S(\boldsymbol{x})}.\tag{1}$$

由此, 为计算$\nabla d_S(\boldsymbol{x})$ 下求$\nabla d_S^2(\boldsymbol{x})$.

记$P_S(\boldsymbol{x})$为$\boldsymbol{x}$在S上的投影, 并对任意模充分小的$\boldsymbol{h}\in\mathbb{R}^n$, 记

$$\triangle = d_S^2(\boldsymbol{x}+\boldsymbol{h}) - d_S^2(\boldsymbol{x}).$$

根据定义,

$$d_S^2(\boldsymbol{x}) \leqslant \|\boldsymbol{x} - P_S(\boldsymbol{x}+\boldsymbol{h})\|^2.$$

所以,

$$\begin{aligned}\triangle &= d_S^2(\boldsymbol{x}+\boldsymbol{h}) - d_S^2(\boldsymbol{x})\\ &\geqslant \|\boldsymbol{x}+\boldsymbol{h} - P_S(\boldsymbol{x}+\boldsymbol{h})\|^2 - \|\boldsymbol{x} - P_S(\boldsymbol{x}+\boldsymbol{h})\|^2\\ &= \|\boldsymbol{h}\|^2 + 2\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{x} - P_S(\boldsymbol{x}+\boldsymbol{h})).\end{aligned}$$

利用投影算子的非扩张性

$$\|P_S(\boldsymbol{x}+\boldsymbol{h}) - P_S(\boldsymbol{x})\| \leqslant \|\boldsymbol{h}\|$$

得

$$\begin{aligned}\triangle \geqslant \|\boldsymbol{h}\|^2 + 2\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{x} - P_S(\boldsymbol{x}+\boldsymbol{h})) &\geqslant \|\boldsymbol{h}\|^2 + 2\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{x} - P_S(\boldsymbol{x})) - o(\|\boldsymbol{h}\|)\\ &= 2\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{x} - P_S(\boldsymbol{x})) + o(\|\boldsymbol{h}\|).\end{aligned} \tag{2}$$

另一方面, 由

$$d_S^2(\boldsymbol{x}+\boldsymbol{h}) \leqslant \|\boldsymbol{x}+\boldsymbol{h} - P_S(\boldsymbol{x})\|^2$$

得

$$\begin{aligned}\triangle &= d_S^2(\boldsymbol{x}+\boldsymbol{h}) - d_S^2(\boldsymbol{x})\\ &\leqslant \|\boldsymbol{x}+\boldsymbol{h} - P_S(\boldsymbol{x})\|^2 - \|\boldsymbol{x} - P_S(\boldsymbol{x})\|^2\\ &= \|\boldsymbol{h}\|^2 + 2\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{x} - P_S(\boldsymbol{x})).\end{aligned}$$

结合(2)得

$$\triangle = 2\boldsymbol{h}^{\mathrm{T}}(\boldsymbol{x} - P_S(\boldsymbol{x})) + o(\|\boldsymbol{h}\|).$$

从而

$$\nabla d_S^2(\boldsymbol{x}) = 2(\boldsymbol{x} - P_S(\boldsymbol{x})).$$

代入(1)得

$$\nabla d_S(\boldsymbol{x}) = \frac{(\boldsymbol{x} - P_S(\boldsymbol{x}))}{\|\boldsymbol{x} - P_S(\boldsymbol{x})\|}.$$

(2) 若$\boldsymbol{x} \in \operatorname{int}(S)$, 则$d_S(\boldsymbol{x}) = 0$, 且对任意模充分小的$\boldsymbol{h} \in \mathbb{R}^n$, $d_S(\boldsymbol{x}+\boldsymbol{h}) = 0$. 因此$\nabla d_S(\boldsymbol{x}) = \mathbf{0}$.

(3) 若$\boldsymbol{x} \in \operatorname{bd}(S)$. 则对任意的$\xi \in \partial d_S(\boldsymbol{x})$, 由次梯度的定义,

$$d_S(\boldsymbol{y}) - d_S(\boldsymbol{x}) = d_S(\boldsymbol{y}) \geqslant \boldsymbol{\xi}^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^n.$$

取$\boldsymbol{y} \in S$, 则

$$0 = d_S(\boldsymbol{y}) \geqslant \xi^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}), \quad \forall \boldsymbol{y} \in S.$$

这说明, $\boldsymbol{\xi} \in N_S(\boldsymbol{x})$, 即$\partial d_S(\boldsymbol{x}) \subset N_S(\boldsymbol{x})$.

取$\boldsymbol{y} = \boldsymbol{x} + \xi$. 则

$$\|\xi\|^2 = \xi^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) \leqslant d_S(\boldsymbol{x}+\xi) \leqslant \|\boldsymbol{x}+\xi-\boldsymbol{x}\| = \|\xi\|.$$

这说明$\|\xi\| \leqslant 1$.

结合前面的讨论得,

$$\partial d_S(\boldsymbol{x}) \subset N_S(\boldsymbol{x}) \cap B(0,1). \tag{3}$$

下证反向包含式成立.

设$\xi \in N_S(\boldsymbol{x}) \cap B(0,1)$. 对任意的$\boldsymbol{y} \in \mathbb{R}^n$,

$$\xi^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) = \xi^{\mathrm{T}}(\boldsymbol{y}-P_S(\boldsymbol{y})) + \xi^{\mathrm{T}}(P_S(\boldsymbol{y})-\boldsymbol{x}).$$

另一方面, 由极锥$N_S(\boldsymbol{x})$的定义,

$$\xi^{\mathrm{T}}(P_S(\boldsymbol{y})-\boldsymbol{x}) \leqslant 0.$$

即

$$\xi^{\mathrm{T}} P_S(\boldsymbol{y}) \leqslant \xi^{\mathrm{T}} \boldsymbol{x}.$$

所以, 由$\|\xi\| \leqslant 1$得

$$\begin{aligned}\xi^{\mathrm{T}}(\boldsymbol{y}-\boldsymbol{x}) \leqslant & \xi^{\mathrm{T}}(\boldsymbol{y}-P_S(\boldsymbol{y})) \leqslant \|\xi\|\|\boldsymbol{y}-P_S(\boldsymbol{y})\| \\ \leqslant & \|\boldsymbol{y}-P_S(\boldsymbol{y})\| = d_S(\boldsymbol{y}) = d_S(\boldsymbol{y}) - d_S(\boldsymbol{x}).\end{aligned}$$

这说明$\xi \in \partial d_S(\boldsymbol{x})$. 由$\xi \in N_S(\boldsymbol{x}) \cap B(0,1)$的任意性,

$$N_S(\boldsymbol{x}) \cap B(0,1) \subset \partial d_S(\boldsymbol{x}).$$

结合(3)式得

$$\partial d_S(\boldsymbol{x}) = N_S(\boldsymbol{x}) \cap B(0,1).$$

综上所述得,

$$\partial d_S(\boldsymbol{x}) = \begin{cases} \mathbf{0}, & \boldsymbol{x} \in \mathrm{int}(S); \\ N_S(\boldsymbol{x}) \cap B(0,1), & \boldsymbol{x} \in \mathrm{bd}(S); \\ \dfrac{\boldsymbol{x} - P(\boldsymbol{x})}{\|\boldsymbol{x} - P_S(\boldsymbol{x})\|}, & \boldsymbol{x} \notin S. \end{cases}$$

10. 计算闭凸集$S \subset \mathbb{R}^n$的示性函数

$$\delta_S(\boldsymbol{x}) = \begin{cases} 0, & \text{若 } \boldsymbol{x} \in S \\ \infty, & \text{若 } \boldsymbol{x} \notin S \end{cases}$$

的次梯度.

解: 容易验证, 示性函数是凸函数. 下面讨论其次梯度.

(1) 若$\boldsymbol{x} \in \mathrm{int}(S)$, 则对于任意的$\boldsymbol{h} \in \mathbb{R}^n$, 只要$\|\boldsymbol{h}\| > 0$充分小, 总有$\boldsymbol{x} + \boldsymbol{h} \in S$. 因此

$$\triangle = \delta_S(\boldsymbol{x} + \boldsymbol{h}) - \delta_S(\boldsymbol{x}) \equiv 0.$$

所以, 示性函数$\delta_S(\boldsymbol{x})$在该点邻域内恒为常数, 即$\nabla \delta_S(\boldsymbol{x}) = \mathbf{0}$.

(2) 若$\boldsymbol{x} \in \mathrm{bd}S$. 对任意的$\xi \in \partial \delta_S(\boldsymbol{x})$, 根据次梯度的定义,

$$\delta_S(\boldsymbol{y}) - \delta_S(\boldsymbol{x}) = \delta_S(\boldsymbol{y}) \geqslant \xi^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^n.$$

令$\boldsymbol{y} \in S$得

$$0 = \delta_S(\boldsymbol{y}) \geqslant \xi^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}).$$

这说明, $\xi \in N_S(\boldsymbol{x})$, 即示性函数在该点的次梯度属于集合在该点的极锥, $\partial \delta_S(\boldsymbol{x}) \subset N_S(\boldsymbol{x})$.

(3) 若$\boldsymbol{x} \notin S$. 对任意的$\xi \in \partial \delta_S(\boldsymbol{x})$, 根据次梯度的定义,

$$\delta_S(\boldsymbol{y}) - \delta_S(\boldsymbol{x}) \geqslant \xi^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}), \quad \forall \boldsymbol{y} \in \mathbb{R}^n.$$

令$\boldsymbol{y} \in S$得

$$\delta_S(\boldsymbol{y}) - \delta_S(\boldsymbol{x}) = -\infty \geqslant \xi^{\mathrm{T}}(\boldsymbol{y} - \boldsymbol{x}).$$

这说明, 示性函数在该点不存在次梯度, 即$\partial d_S(\boldsymbol{x}) = \Phi$.

综上所述得,

$$\partial \delta_S(\boldsymbol{x}) = \begin{cases} 0, & 若\ \boldsymbol{x} \in \mathrm{int}(S); \\ \subset N_S(\boldsymbol{x}), & 若\ \boldsymbol{x} \in \mathrm{bd}(S), \\ \Phi, & 若\ \boldsymbol{x} \notin S. \end{cases}$$

证毕

第3章 无约束优化最优性条件

1. 设$\boldsymbol{b}\in\mathbb{R}^n$, 矩阵$\boldsymbol{Q}\in\mathbb{R}^{n\times n}$对称正定. 试求下述优化问题的最优解和最优值

$$\max\{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}\mid \boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}\le 1\}$$

并利用该结果证明对任意的$\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$有,

$$(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y})^2\leqslant(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x})(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{y}).$$

解: 由于$\boldsymbol{Q}$对称正定, 故可以分解成非奇异矩阵$\boldsymbol{B}$和$\boldsymbol{B}^{\mathrm{T}}$的乘积, 即$\boldsymbol{Q}=\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B}$. 令$\boldsymbol{y}=\boldsymbol{B}\boldsymbol{x}$, 则$\boldsymbol{x}=\boldsymbol{B}^{-1}\boldsymbol{y}$, 原问题化为

$$\max\{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{B}^{-1}\boldsymbol{y}\mid \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\le 1\}$$

也就是

$$\max\{(\boldsymbol{B}^{-\mathrm{T}}\boldsymbol{b})^{\mathrm{T}}\boldsymbol{y}\mid \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\le 1\}$$

由Cauchy-Schwarz不等式知, 上述问题的最优解为$\boldsymbol{y}^*=\frac{\boldsymbol{B}^{-\mathrm{T}}\boldsymbol{b}}{\|\boldsymbol{B}^{-\mathrm{T}}\boldsymbol{b}\|}$. 从而原问题的最优解为

$$\boldsymbol{x}^*=\boldsymbol{B}^{-1}\boldsymbol{y}^*=\frac{\boldsymbol{B}^{-1}\boldsymbol{B}^{-\mathrm{T}}\boldsymbol{b}}{\|\boldsymbol{B}^{-\mathrm{T}}\boldsymbol{b}\|}=\frac{\boldsymbol{Q}^{-1}\boldsymbol{b}}{\sqrt{\boldsymbol{b}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{b}}}.$$

另外,

$$\begin{aligned}(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y})^2&=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{y}=\left\langle\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}},\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\right\rangle,\\&=\left\langle\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}},\boldsymbol{Q}\boldsymbol{Q}^{-1}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\right\rangle=\left\langle\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q},\boldsymbol{Q}^{-1}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\right\rangle,\\&\le\|\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\|_F\|\boldsymbol{Q}^{-1}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\|_F=\mathrm{tr}\left(\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\right)\mathrm{tr}\left(\boldsymbol{Q}^{-1}\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}\right),\\&=\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{Q}\boldsymbol{x}\right)\left(\boldsymbol{y}^{\mathrm{T}}\boldsymbol{Q}^{-1}\boldsymbol{y}\right).\end{aligned}$$

第二部分得证. 证毕

2. 试在$\mathbb{R}^3$中确定一通过点$(3;4;5)$的平面, 使其与非负象限中的三个坐标面构成的四面体的体积最小.

解: 三维欧式空间中过点(3,4,5)的平面方程可表示为

$$a(x-3)+b(y-4)+c(z-5)=0,$$

其中待求量(a,b,c)为该平面的任一法向量. 容易计算该平面在坐标轴上的截距分别为

$$x_0=\frac{3a+4b+5c}{a},\ y_0=\frac{3a+4b+5c}{b},\ z_0=\frac{3a+4b+5c}{c}.$$

根据四面体的体积公式, 该平面和坐标轴围成的四面体的体积为

$$V=\frac{1}{6}x_0y_0z_0=\frac{(3a+4b+5c)^3}{6abc},$$

其中$abc\neq 0$. 从而, 原问题可化为如下优化问题

$$\min_{a,b,c}\frac{(3a+4b+5c)^3}{6abc}$$

根据无约束优化问题的最优性条件, 将上述函数关于a,b,c分别求导并令之为零得

$$\begin{cases}9a-(3a+4b+5c)=0,\\12b-(3a+4b+5c)=0,\\15c-(3a+4b+5c)=0.\end{cases}$$

化简得

$$\begin{cases}6a-4b-5c=0,\\8b-3a-5c=0,\\10c-3a-4b=0.\end{cases}$$

整理得

$$3a=4b=5c.$$

由此得使得体积最大的过点(3,4,5)的平面方程的法向量为$(1/3,1/4,1/5)$. 代入体积公式得体积=270. 证毕

3. 设$\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{b}\in\mathbb{R}^m$. 试给出无约束优化问题

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2$$

的一阶最优性条件, 并验证该条件是否是充分的. 它的最优解唯一吗?

解: 该优化问题的一阶最优性条件为

$$\nabla_{\boldsymbol{x}}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2=\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})=0.$$

由于$\nabla^2\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$半正定, 故$\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2$关于$\boldsymbol{x}\in\mathbb{R}^n$为凸函数, 所以满足$\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b})=0$的$\boldsymbol{x}$也是优化问题的最优解, 即该最优性条件是充分的. 特别地, 当$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A}$正定时, 最优解唯一. 证毕

4. 设$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\in\mathbb{R}^n$. 求下述问题的最优解

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\sum_{i=1}^m\|\boldsymbol{a}_i-\boldsymbol{x}\|^2$$

解: 对该优化问题, 由一阶最优性条件, 最优解满足

$$\sum_{i=1}^m(\boldsymbol{a}_i-\boldsymbol{x}^*)=0.$$

此即最优解为

$$\boldsymbol{x}^*=\frac{1}{m}\sum_{i=1}^m\boldsymbol{a}_i.$$

证毕

5. 设$\boldsymbol{a}\in\mathbb{R}^n,\lambda>0$. 试利用目标函数的可分离性质给出如下优化问题的最优解

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\frac{1}{2}\|\boldsymbol{x}-\boldsymbol{a}\|^2+\lambda\|\boldsymbol{x}\|_1\qquad\min_{\boldsymbol{x}\in\mathbb{R}^n}\lambda\|\boldsymbol{x}\|_1-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}$$

解: 前一优化问题等价于

$$\sum_{i=1}^n\left(\min_{\boldsymbol{x}_i\in\mathbb{R}}\frac{1}{2}(x_i-a_i)^2+\lambda|x_i|\right)$$

这样, 要求解原问题, 只需求解如下子问题

$$\min_{\boldsymbol{x}_i\in\mathbb{R}}\frac{1}{2}(x_i-a_i)^2+\lambda|x_i| \tag{1.6.2}$$

对此, 我们分情况讨论.

(1) $a_i\geqslant 0$. 显然, 该子问题的最优解非负, 即该子问题等价于

$$\min_{x_i\geqslant 0}\frac{1}{2}(x_i-a_i)^2+\lambda x_i$$

利用优化问题的最优性条件, 上述问题的最优解为

$$x_i^*=\begin{cases}a_i-\lambda, & 若\ a_i>\lambda;\\ 0, & 若\ a_i\leqslant\lambda.\end{cases}$$

(2) 若$a_i<0$, 则子问题的最优解非正. 该子问题可化为

$$\min_{x_i\leqslant 0}\frac{1}{2}(x_i-a_i)^2-\lambda x_i.$$

利用优化问题的最优性条件, 该子问题的最优解为

$$x_i^* = \begin{cases} 0, & 若\ a_i \geqslant -\lambda; \\ a_i + \lambda, & 若\ a_i < -\lambda. \end{cases}$$

引入符号函数

$$\operatorname{sgn}(a_i) = \begin{cases} 1, & 若\ a_i > 0; \\ 0, & 若\ a_i = 0; \\ -1, & 若\ a_i < 0. \end{cases}$$

则问题的最优解为

$$x_i^* = \begin{cases} 0, & 若\ |a_i| < \lambda \\ a_i - \operatorname{sgn}(a_i)\lambda, & 若\ |a_i| > \lambda \end{cases} = \operatorname{sgn}(a_i)(|a_i| - \lambda)_+.$$

由此得原问题的最优解.

对后一优化问题, 先考虑如下一元优化问题

$$\min_{x \in R} \lambda |x| - ax$$

若$a \geq 0$, 则最优解$x^* \geq 0$, 从而原问题可写成

$$\min_{x \geq 0} (\lambda - a)x$$

显然, 该问题的最优解为

$$x^* = \begin{cases} +\infty, & 如果\lambda < a, \\ 0, & 如果\lambda \geq a. \end{cases}$$

从而, 最优值

$$\min_{x \geq 0} (\lambda - a)x = \begin{cases} -\infty, & 如果\lambda < a, \\ 0, & 如果\lambda \geq a. \end{cases}$$

若$a \leq 0$, 则最优解$x^* \leq 0$, 从而原问题可写成

$$\min_{x \leq 0} -(\lambda + a)x$$

显然, 该问题的最优解为

$$x^* = \begin{cases} -\infty, & 如果\lambda < -a, \\ 0, & 如果\lambda \geq -a. \end{cases}$$

从而, 最优值

$$\min_{x\le 0} -(\lambda + a)x = \begin{cases} -\infty, & \text{如果}\lambda < -a, \\ 0, & \text{如果}\lambda \ge -a. \end{cases}$$

综上所述, 优化问题

$$\min_{\boldsymbol{x}\in R^n} \lambda|\boldsymbol{x}|_1 - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = \begin{cases} -\infty, & \text{如果}\lambda < \|\boldsymbol{a}\|_\infty, \\ 0, & \text{如果}\lambda \ge \|\boldsymbol{a}\|_\infty. \end{cases}$$

第4章 线搜索方法

习题

1. 设函数$f:\mathbb{R}^n\to\mathbb{R}$连续可微, $\boldsymbol{d}\in\mathbb{R}^n$为该函数在点$\boldsymbol{x}\in\mathbb{R}^n$的下降方向. 试建立函数$\phi(\alpha)=f(\boldsymbol{x}+\alpha\boldsymbol{d})$在$\alpha\geqslant 0$上的最小值点的必要条件, 并讨论在什么条件下该条件是充分的.

解: 由于$\boldsymbol{d}$为点$\boldsymbol{x}\in\mathbb{R}^n$的下降方向, 故

$$f'_\alpha(\boldsymbol{x}+\alpha\boldsymbol{d})|_{\alpha=0}=\boldsymbol{d}^{\mathrm{T}}\nabla f(\boldsymbol{x})<0.$$

这说明, $f(\boldsymbol{x}+\alpha\boldsymbol{d})$关于$\alpha>0$的最小值点满足$\alpha^*\in(0,\infty)$.

由无约束优化问题的最优性条件, 函数$\phi(\alpha)=f(\boldsymbol{x}+\alpha\boldsymbol{d})$关于$\alpha>0$的最小值点满足

$$f'_\alpha(\boldsymbol{x}+\alpha\boldsymbol{d})=\boldsymbol{d}^{\mathrm{T}}\nabla f(\boldsymbol{x}+\alpha\boldsymbol{d})=0.$$

如果函数$f(\boldsymbol{x}+\alpha\boldsymbol{d})$关于$\alpha$为凸函数, 则上述条件是充分的. 证毕

2. 求严格凸二次函数$f(\boldsymbol{x})=\dfrac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}$在$\boldsymbol{x}_k$点沿下降方向$\boldsymbol{d}_k$的最优步长. 若取$\boldsymbol{d}_k=-\boldsymbol{g}_k$, 试计算目标函数在每一迭代步的下降量.

解: 由最优步长的定义,

$$\alpha_k=\arg\min_{\alpha\geq 0}f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k).$$

则由

$$\boldsymbol{d}_k^{\mathrm{T}}\nabla f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)=\boldsymbol{d}_k^{\mathrm{T}}[\boldsymbol{A}(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)+\boldsymbol{b}]=0$$

知最优步长

$$\alpha_k=\frac{-\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\boldsymbol{d}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_k},$$

其中, $\boldsymbol{g}_k=\boldsymbol{A}\boldsymbol{x}_k+\boldsymbol{b}$.

若 $\boldsymbol{d}_k=-\boldsymbol{g}_k$, 则 $\alpha_k=\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k}$. 从而,

$$\begin{aligned}
f(\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k)=&f(\boldsymbol{x}_k-\alpha_k\boldsymbol{g}_k)\\
=&\frac{1}{2}(\boldsymbol{x}_k-\alpha_k\boldsymbol{g}_k)^{\mathrm{T}}\boldsymbol{A}(\boldsymbol{x}_k-\alpha_k\boldsymbol{g}_k)+\boldsymbol{b}^{\mathrm{T}}(\boldsymbol{x}_k-\alpha_k\boldsymbol{g}_k)\\
=&\frac{1}{2}\boldsymbol{x}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_k+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}_k-\alpha_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_k+\frac{1}{2}\alpha_k^2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k-\alpha_k\boldsymbol{b}^{\mathrm{T}}\boldsymbol{g}_k\\
=&f(\boldsymbol{x}_k)-\alpha_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_k+\frac{1}{2}\alpha_k^2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k-\alpha_k\boldsymbol{b}^{\mathrm{T}}\boldsymbol{g}_k.
\end{aligned}$$

由此,目标函数在该迭代步的下降量为

$$\begin{aligned}
f(\boldsymbol{x}_k+\alpha_k\boldsymbol{d}_k)-f(\boldsymbol{x}_k) &= -\alpha_k\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_k+\frac{1}{2}\alpha_k^2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k-\alpha_k\boldsymbol{b}^{\mathrm{T}}\boldsymbol{g}_k\\
&= -\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_k+\frac{1}{2}\left(\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k}\right)^2\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k-\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k}\boldsymbol{b}^{\mathrm{T}}\boldsymbol{g}_k\\
&= -\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k}\left[\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}_k-\frac{1}{2}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{g}_k\right]\\
&= -\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k}\frac{1}{2}\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{g}_k=-\frac{1}{2}\frac{\|\boldsymbol{g}_k\|^4}{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{A}\boldsymbol{g}_k}.
\end{aligned}$$

证毕

第5章 最速下降法与牛顿方法

习题

1. 证明牛顿算法对于严格凸二次函数一次迭代即得到该函数的全局最优值点.

解: 对严格凸二次函数为 $f(\boldsymbol{x})=\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b}^{\mathrm{T}}\boldsymbol{x}+c$, 其中 $\boldsymbol{A}$对称正定,

$$\boldsymbol{g}_k=\nabla f(\boldsymbol{x}_k)=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b},\ \boldsymbol{d}_k^N=-\boldsymbol{A}^{-1}\boldsymbol{g}_k.$$

对初始点 $\boldsymbol{x}_0$, 迭代过程为

$$\begin{aligned}&\boldsymbol{d}_1=-\boldsymbol{A}^{-1}\boldsymbol{g}_0=-\boldsymbol{A}^{-1}(\boldsymbol{A}\boldsymbol{x}_0+\boldsymbol{b})=-\boldsymbol{x}_0-\boldsymbol{A}^{-1}\boldsymbol{b}\ ;\\&\boldsymbol{x}_1=\boldsymbol{x}_0+\boldsymbol{d}_1=-\boldsymbol{A}^{-1}\boldsymbol{b}.\end{aligned}$$

从而, $\boldsymbol{g}_1=-\boldsymbol{A}\boldsymbol{A}^{-1}\boldsymbol{b}+\boldsymbol{b}=0$. $\boldsymbol{x}_1$为凸函数的稳定点, 算法终止. 证毕

2. 设$f:\mathbb{R}^n\to\mathbb{R}$为连续可微函数, $\boldsymbol{B}_k$为n阶对称正定矩阵. 考虑由下述公式产生搜索方向

$$\boldsymbol{B}_k\boldsymbol{d}_k=-\boldsymbol{g}_k$$

的下降算法. 分析$\boldsymbol{B}_k$的不同取法对算法收敛速度的影响.

解: 考虑搜索方向的下降性

$$\begin{aligned}f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)&=f(\boldsymbol{x}_k)+\alpha\nabla f(\boldsymbol{x}_k)^{\mathrm{T}}\boldsymbol{d}_k+o(\alpha\|\boldsymbol{d}_k\|)\\&=f(\boldsymbol{x}_k)+\alpha\boldsymbol{g}_k^{\mathrm{T}}B_k^{-1}\boldsymbol{g}_k+o(\alpha\|\boldsymbol{d}_k\|)\le f(\boldsymbol{x}_k).\end{aligned}$$

若 $\boldsymbol{B}_k=\boldsymbol{I}$, 则该搜索方向为最速下降方向, 收敛速度至多为线性的; 若$\boldsymbol{B}_k=\boldsymbol{G}_k$, 目标函数$f(\boldsymbol{x})$ 在极小点处 Lipschitz 连续, 则算法二阶收敛. 证毕

第6章 共轭梯度法

习题

1. 设矩阵$\boldsymbol{A}\in\mathbb{R}^{n\times n}$对称正定, 向量组$\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_n$关于矩阵$\boldsymbol{A}$共轭. 试证明:

(1) 对任意的$\boldsymbol{x}\in\mathbb{R}^n$, $\boldsymbol{x}=\sum\limits_{i=1}^{n}\dfrac{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}\boldsymbol{d}_i$;

(2) $\boldsymbol{A}^{-1}=\sum\limits_{i=1}^{n}\dfrac{\boldsymbol{d}_i\boldsymbol{d}_i^{\mathrm{T}}}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}$.

解: (1) 由向量组$\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_n$关于矩阵$\boldsymbol{A}$共轭知, $\boldsymbol{d}_1,\boldsymbol{d}_2,\cdots,\boldsymbol{d}_n\in\mathbb{R}^n$ 线性无关. 故对任意的$\boldsymbol{x}\in\mathbb{R}^n$, 存在数组 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 使得

$$\boldsymbol{x}=\sum_{i=1}^{n}\alpha_i\boldsymbol{d}_i.$$

从而

$$\langle\boldsymbol{x},\boldsymbol{A}\boldsymbol{d}_j\rangle=\left\langle\sum_{i=1}^{n}\alpha_i\boldsymbol{d}_i,\boldsymbol{A}\boldsymbol{d}_j\right\rangle=\sum_{i=1}^{n}\alpha_i\langle\boldsymbol{d}_i,\boldsymbol{A}\boldsymbol{d}_j\rangle=\alpha_j\boldsymbol{d}_j^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_j,\ j=1,2,\cdots,n.$$

由此,

$$\alpha_j=\frac{\langle\boldsymbol{x},\boldsymbol{A}\boldsymbol{d}_j\rangle}{\boldsymbol{d}_j^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_j}=\frac{\boldsymbol{d}_j^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{d}_j^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_j},\ j=1,2,\cdots,n.$$

也即

$$\boldsymbol{x}=\sum_{i=1}^{n}\frac{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}\boldsymbol{d}_i.$$

(2) 设$\boldsymbol{A}^{-1}=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)$, $\boldsymbol{a}_j\in\mathbb{R}^n$, $j=1,2,\cdots,n$. 则由$\boldsymbol{I}=\boldsymbol{A}\boldsymbol{A}^{-1}=(\boldsymbol{A}\boldsymbol{a}_1,\boldsymbol{A}\boldsymbol{a}_2,\cdots,\boldsymbol{A}\boldsymbol{a}_n)$知$\boldsymbol{A}\boldsymbol{a}_j=\boldsymbol{e}_j$, $j=1,2,\cdots,n$, 其中 $\boldsymbol{e}_j\in\mathbb{R}^n$ 表示第 j 个单位向量.

由(1)知,

$$\boldsymbol{a}_j=\sum_{i=1}^{n}\frac{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{a}_j}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}\boldsymbol{d}_i=\sum_{i=1}^{n}\frac{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{e}_j}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}\boldsymbol{d}_i=\sum_{i=1}^{n}\frac{(\boldsymbol{d}_i)_j}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}\boldsymbol{d}_i.$$

从而

$$\begin{aligned}\boldsymbol{A}^{-1}&=(\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_n)\\&=\left(\sum_{i=1}^{n}\frac{\boldsymbol{d}_{i1}}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}\boldsymbol{d}_i,\sum_{i=1}^{n}\frac{\boldsymbol{d}_{i2}}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}\boldsymbol{d}_i,\cdots,\sum_{i=1}^{n}\frac{\boldsymbol{d}_{in}}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}\boldsymbol{d}_i\right)\\&=\sum_{i=1}^{n}\frac{\boldsymbol{d}_i}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}(\boldsymbol{d}_{i1},\boldsymbol{d}_{i2},\cdots,\boldsymbol{d}_{in})=\sum_{i=1}^{n}\frac{\boldsymbol{d}_i\boldsymbol{d}_i^{\mathrm{T}}}{\boldsymbol{d}_i^{\mathrm{T}}\boldsymbol{A}\boldsymbol{d}_i}.\end{aligned}$$

证毕

2. 在共轭梯度法的迭代格式中, 若参数β由D-Y公式确定, 则无论采用何种步长规则产生下一迭代点都有

$$\beta_k = \frac{\boldsymbol{d}_{k+1}^{\mathrm{T}} \boldsymbol{g}_{k+1}}{\boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{g}_k}.$$

解: 先给出DY共轭梯度法下的迭代格式

$$\boldsymbol{d}_{k+1} = -\boldsymbol{g}_{k+1} + \beta_k \boldsymbol{d}_k, \qquad \beta_k = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{g}_{k+1}}{\boldsymbol{d}_k^{\mathrm{T}} (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k)}.$$

显然,

$$\boldsymbol{g}_{k+1} = \beta_k \boldsymbol{d}_k - \boldsymbol{d}_{k+1}.$$

从而由β_k的迭代公式,

$$\beta_k \boldsymbol{d}_k^{\mathrm{T}} (\boldsymbol{g}_{k+1} - \boldsymbol{g}_k) = \boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{g}_{k+1} = (\beta_k \boldsymbol{d}_k - \boldsymbol{d}_{k+1})^{\mathrm{T}} \boldsymbol{g}_{k+1}.$$

展开得

$$-\beta_k \boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{g}_k = -\boldsymbol{d}_{k+1}^{\mathrm{T}} \boldsymbol{g}_{k+1}.$$

整理得

$$\beta_k = \frac{\boldsymbol{d}_{k+1}^{\mathrm{T}} \boldsymbol{g}_{k+1}}{\boldsymbol{d}_k^{\mathrm{T}} \boldsymbol{g}_k}.$$

证毕

3. 对强Wolfe步长规则下的共轭梯度法($0 < \sigma_1 < \sigma_2 < \frac{1}{2}$), 若$|\beta_k| \leqslant \beta_k^{FR}$, 则

$$-\frac{1}{1-\sigma_2} \leqslant \frac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2} \leqslant \frac{2\sigma_2 - 1}{1-\sigma_2}.$$

解: 先给出参数β的F-R公式

$$\beta_k^{FR} = \frac{\boldsymbol{g}_{k+1}^{\mathrm{T}} \boldsymbol{g}_{k+1}}{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{g}_k}.$$

下用数学归纳法证明.

当$k = 0$时, $\boldsymbol{d}_0 = -\boldsymbol{g}_0$. 则

$$\frac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2} = -1.$$

从而由$\sigma \in (0, 1/2)$得,

$$-\frac{1}{1-\sigma_2} \leqslant \frac{\boldsymbol{g}_k^{\mathrm{T}} \boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2} = -1 \leqslant \frac{2\sigma_2 - 1}{1-\sigma_2}.$$

设结论对k成立, 即
$$-\frac{1}{1-\sigma_2}\leqslant\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}\leqslant\frac{2\sigma_2-1}{1-\sigma_2}.$$
则对任意的 $|\beta_k|\le\beta_k^{FR}=\|\boldsymbol{g}_{k+1}\|^2/\|\boldsymbol{g}_k\|^2$, 由$\boldsymbol{d}_{k+1}=-\boldsymbol{g}_{k+1}+\beta_k\boldsymbol{d}_k$得
$$\frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_{k+1}}{\|\boldsymbol{g}_{k+1}\|^2}=-1+\beta_k\frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_{k+1}\|^2}.$$
从而
$$\left|\frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_{k+1}}{\|\boldsymbol{g}_{k+1}\|^2}+1\right|=|\beta_k|\frac{\left|\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_k\right|}{\|\boldsymbol{g}_{k+1}\|^2}\leqslant\beta_k^{FR}\frac{\left|\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_k\right|}{\|\boldsymbol{g}_{k+1}\|^2}=\frac{\left|\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_k\right|}{\|\boldsymbol{g}_k\|^2}.$$
另一方面, 由强 Wolfe 步长规则
$$f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)\leqslant f(\boldsymbol{x}_k)+\sigma_1\alpha\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k,$$
$$|\nabla f(\boldsymbol{x}_k+\alpha\boldsymbol{d}_k)^{\mathrm{T}}\boldsymbol{d}_k|\leqslant\sigma_2|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k|$$
其中, $0<\sigma_1<\sigma_2<1$, 及 $\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k<0$得,
$$\left|\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_k\right|\leqslant\sigma_2\left|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k\right|.$$
从而,
$$\left|\frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_{k+1}}{\|\boldsymbol{g}_{k+1}\|^2}+1\right|\leqslant\sigma_2\frac{\left|\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k\right|}{\|\boldsymbol{g}_k\|^2}.$$
整理得
$$-1+\sigma_2\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}\leqslant\frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_{k+1}}{\|\boldsymbol{g}_{k+1}\|^2}\leqslant-1-\sigma_2\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}.$$
由结论
$$-\frac{1}{1-\sigma_2}\leqslant\frac{\boldsymbol{g}_k^{\mathrm{T}}\boldsymbol{d}_k}{\|\boldsymbol{g}_k\|^2}$$
得
$$-\frac{1}{1-\sigma_2}\leqslant\frac{\boldsymbol{g}_{k+1}^{\mathrm{T}}\boldsymbol{d}_{k+1}}{\|\boldsymbol{g}_{k+1}\|^2}\leqslant\frac{2\sigma_2-1}{1-\sigma_2}.$$
这说明, 结论对$k+1$成立. 证毕

第7章 最小二乘问题

习题

1. 设$\boldsymbol{A} \in \mathbb{R}^{m_1 \times n}$, $\boldsymbol{B} \in \mathbb{R}^{m_2 \times n}$, $\boldsymbol{b} \in \mathbb{R}^{m_1}$. 试证明线性最小二乘问题

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} \left\| \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \boldsymbol{x} - \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix} \right\|^2$$

的最优解满足方程

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{A} \\ \boldsymbol{A}^{\mathrm{T}} & -\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix}.$$

解: 由无约束优化问题的最优性条件, 线性最小二乘问题的最优解满足

$$\nabla_{\boldsymbol{x}} \left\| \begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \boldsymbol{x} - \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix} \right\|^2 = \begin{pmatrix} \boldsymbol{A}^{\mathrm{T}} & \boldsymbol{B}^{\mathrm{T}} \end{pmatrix} \left[\begin{pmatrix} \boldsymbol{A} \\ \boldsymbol{B} \end{pmatrix} \boldsymbol{x} - \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix} \right] = \boldsymbol{0}.$$

整理得

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \right) \boldsymbol{x}.$$

另一方面, 线性方程组

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{A} \\ \boldsymbol{A}^{\mathrm{T}} & -\boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \end{pmatrix} \begin{pmatrix} \boldsymbol{y} \\ \boldsymbol{x} \end{pmatrix} = \begin{pmatrix} \boldsymbol{b} \\ \boldsymbol{0} \end{pmatrix}$$

等价于如下方程组问题

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

而该方程组关于$\boldsymbol{x}$的解满足

$$\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b} = \left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \boldsymbol{B}^{\mathrm{T}}\boldsymbol{B} \right) \boldsymbol{x}.$$

结合前一结论得证. 证毕

2. 对非线性最小二乘问题

$$\min_{\boldsymbol{x} \in \mathbb{R}^n} f(\boldsymbol{x}) = \frac{1}{2} \boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}} \boldsymbol{r}(\boldsymbol{x}) = \frac{1}{2} \sum_{i=1}^{m} r_i^2(\boldsymbol{x})$$

设对任意的$\boldsymbol{x} \in \mathbb{R}^n$, 向量值函数$\boldsymbol{r}(\boldsymbol{x})$的Jacobi矩阵$\boldsymbol{J}(\boldsymbol{x})$列满秩. 则

$$\lim_{\mu \to 0} \boldsymbol{d}^{\mathrm{LM}}(\mu) = \boldsymbol{d}^{\mathrm{GN}}, \quad \lim_{\mu \to \infty} \frac{\boldsymbol{d}^{\mathrm{LM}}(\mu)}{\|\boldsymbol{d}^{\mathrm{LM}}(\mu)\|} = \frac{\boldsymbol{d}^{s}}{\|\boldsymbol{d}^{s}\|},$$

其中, $\boldsymbol{d}^s$为目标函数在$\boldsymbol{x}$点的最速下降方向.

解: 对非线性最小二乘问题, 目标函数的梯度为

$$\boldsymbol{g}(\boldsymbol{x}) = \nabla\Big(\frac{1}{2}\boldsymbol{r}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x})\Big) = \boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}).$$

目标函数的L-M方向为

$$\boldsymbol{d}^{\mathrm{LM}} = -\Big(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu\boldsymbol{I}\Big)^{-1}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}).$$

显然,

$$\lim_{\mu\to 0}\boldsymbol{d}^{\mathrm{LM}}(\mu) = -\Big(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x})\Big)^{-1}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x}) = \boldsymbol{d}^{\mathrm{GN}}.$$

下讨论$\mu\to\infty$的情况.

对任意发散到∞的正数列$\{\mu_k\}$, 数列$\Big\{\dfrac{\boldsymbol{d}(\mu_k)}{\|\boldsymbol{d}(\mu_k)\|}\Big\}$有界. 故有收敛子列, 不妨设为其本身, 即$\{\dfrac{\boldsymbol{d}(\mu_k)}{\|\boldsymbol{d}(\mu_k)\|}\}$收敛.

由$\boldsymbol{d}(\mu_k)$的定义,

$$\begin{aligned}\boldsymbol{d}(\mu_k) &= -(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu\boldsymbol{I})^{-1}\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{r}(\boldsymbol{x})\\ &= -(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu\boldsymbol{I})^{-1}\boldsymbol{g}(\boldsymbol{x}_k).\end{aligned}$$

所以,

$$\begin{aligned}-\frac{\boldsymbol{g}(\boldsymbol{x})}{\|\boldsymbol{g}(\boldsymbol{x})\|} &= \frac{(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu_k\boldsymbol{I})\boldsymbol{d}(\mu_k)}{\|(\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x}) + \mu_k\boldsymbol{I})\boldsymbol{d}(\mu_k)\|}\\ &= \frac{\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x})\dfrac{\boldsymbol{d}(\mu_k)}{\mu_k\|\boldsymbol{d}(\mu_k)\|} + \dfrac{\boldsymbol{d}(\mu_k)}{\|\boldsymbol{d}(\mu_k)\|}}{\Big\|\boldsymbol{J}(\boldsymbol{x})^{\mathrm{T}}\boldsymbol{J}(\boldsymbol{x})\dfrac{\boldsymbol{d}(\mu_k)}{\mu_k\|\boldsymbol{d}(\mu_k)\|} + \dfrac{\boldsymbol{d}(\mu_k)}{\|\boldsymbol{d}(\mu_k)\|}\Big\|}.\end{aligned}$$

令$k\to\infty$得

$$\lim_{k\to\infty}\frac{\boldsymbol{d}(\mu_k)}{\|\boldsymbol{d}(\mu_k)\|} = -\frac{\boldsymbol{g}(\boldsymbol{x})}{\|\boldsymbol{g}(\boldsymbol{x})\|}.$$

结合(7.2.12)得,

$$\lim_{k\to\infty}\Big\langle\frac{-\boldsymbol{g}(\boldsymbol{x})}{\|\boldsymbol{g}(\boldsymbol{x})\|}, \frac{\boldsymbol{d}(\mu_k)}{\|\boldsymbol{d}(\mu_k)\|}\Big\rangle = 1.$$

这说明, 在$\mu_k\to\infty$时, LM方向收敛到该点的最速下降方向. 证毕

3. 设$\boldsymbol{A}\in\mathbb{R}^{m\times n}$, $\boldsymbol{b}\in\mathbb{R}^m$, $\mu_1>\mu_2>0$, 并设$\boldsymbol{x}_1, \boldsymbol{x}_2$分别是方程

$$(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \mu\boldsymbol{I})\boldsymbol{x} = -\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{b}$$

在μ取μ_1, μ_2时的解, 则$\|\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}\|^2 < \|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}\|^2$.

解: 考虑如下凸优化问题

$$\min_{\boldsymbol{x}\in\mathbb{R}^n} \quad \|\boldsymbol{A}\boldsymbol{x} + \boldsymbol{b}\|^2 + \mu\|\boldsymbol{x}\|^2$$

其最优解满足

$$\left(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A} + \mu\boldsymbol{I}\right)\boldsymbol{x} = -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}.$$

则对$\mu = \mu_1$, μ_2的最优解$\boldsymbol{x}_1$, $\boldsymbol{x}_2$, 由最优解的定义得

$$\|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}\|^2 + \mu_1\|\boldsymbol{x}_1\|^2 \leqslant \|\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}\|^2 + \mu_1\|\boldsymbol{x}_2\|^2,$$
$$\|\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}\|^2 + \mu_2\|\boldsymbol{x}_2\|^2 \leqslant \|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}\|^2 + \mu_2\|\boldsymbol{x}_1\|^2.$$

前一式乘以μ_2, 后一式乘以μ_1, 并将两式相加得

$$(\mu_1 - \mu_2)\|\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}\|^2 \leqslant (\mu_1 - \mu_2)\|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}\|^2.$$

由$\mu_1 > \mu_2$得,

$$\|\boldsymbol{A}\boldsymbol{x}_2 + \boldsymbol{b}\|^2 \leqslant \|\boldsymbol{A}\boldsymbol{x}_1 + \boldsymbol{b}\|^2.$$

结论得证. 证毕

第8章 临近点算法

习题

1. 设$f,g:\mathbb{R}^n\to\mathbb{R}$为凸函数, f连续可微. 则对任意的$L>0, \boldsymbol{y}\in\mathbb{R}^n$和

$$\boldsymbol{x}^+=\arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\langle\nabla f(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle+\frac{L}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2+g(\boldsymbol{x}),$$

若

$$f(\boldsymbol{x}^+)\leqslant f(\boldsymbol{y})+\langle\nabla f(\boldsymbol{y}),\boldsymbol{x}^+-\boldsymbol{y}\rangle+\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2,$$

则

$$f(\boldsymbol{x})+g(\boldsymbol{x})\geqslant f(\boldsymbol{x}^+)+g(\boldsymbol{x}^+)+\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2+L\langle\boldsymbol{y}-\boldsymbol{x},\boldsymbol{x}^+-\boldsymbol{y}\rangle,\quad\forall\boldsymbol{x}\in\mathbb{R}^n.$$

解: 由$\boldsymbol{x}^+=\arg\min_{\boldsymbol{x}\in\mathbb{R}^n}\langle\nabla f(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle+\frac{L}{2}\|\boldsymbol{x}-\boldsymbol{y}\|^2+g(\boldsymbol{x})$得,

$$\nabla f(\boldsymbol{y})+L(\boldsymbol{x}^+-\boldsymbol{y})+\xi_2(\boldsymbol{x}^+)=0,\quad \xi_2\in\partial g(\boldsymbol{x}^+).$$

所以, 对任意的$\boldsymbol{x}\in\mathbb{R}^n$, 由$f(\boldsymbol{x}^+)\leqslant f(\boldsymbol{y})+\langle\nabla f(\boldsymbol{y}),\boldsymbol{x}^+-\boldsymbol{y}\rangle+\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2$得,

$$\begin{aligned}
f(\boldsymbol{x})-f(\boldsymbol{x}^+)\geqslant{}& f(\boldsymbol{x})-f(\boldsymbol{y})-\langle\nabla f(\boldsymbol{y}),\boldsymbol{x}^+-\boldsymbol{y}\rangle-\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2-g(\boldsymbol{x}^+)\\
={}& f(\boldsymbol{x})+g(\boldsymbol{x})-f(\boldsymbol{y})-\langle\nabla f(\boldsymbol{y}),\boldsymbol{x}-\boldsymbol{y}\rangle-\langle\nabla f(\boldsymbol{y}),\boldsymbol{x}^+-\boldsymbol{x}\rangle-\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2-g(\boldsymbol{x}^+)\\
\geqslant{}& g(\boldsymbol{x})-g(\boldsymbol{x}^+)-\langle\nabla f(\boldsymbol{y}),\boldsymbol{x}^+-\boldsymbol{x}\rangle-\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2\\
={}& g(\boldsymbol{x})-g(\boldsymbol{x}^+)+\langle L(\boldsymbol{x}^+-\boldsymbol{y})+\xi_2(\boldsymbol{x}^+),\boldsymbol{x}^+-\boldsymbol{x}\rangle-\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2\\
={}& g(\boldsymbol{x})-g(\boldsymbol{x}^+)+L\langle\boldsymbol{x}^+-\boldsymbol{y},\boldsymbol{x}^+-\boldsymbol{x}\rangle-\langle\xi_2(\boldsymbol{x}^+),\boldsymbol{x}-\boldsymbol{x}^+\rangle-\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2\\
\geqslant{}& L\langle\boldsymbol{x}^+-\boldsymbol{y},\boldsymbol{y}-\boldsymbol{x}\rangle+\frac{L}{2}\|\boldsymbol{x}^+-\boldsymbol{y}\|^2
\end{aligned}$$

证毕

第9章 交替极小化方法

习题

1. 用交替极小化方法求下述优化问题的最优解

$$\min_{x>0,y\geqslant 0} f(x,y) = \frac{10}{x} + \frac{(x-y)^2}{2x} + \frac{3y^2}{2x}$$

解: 对任意的$x > 0$, 计算关于y的优化问题. 此时, 利用目标函数关于y为凸函数得内层优化问题关于y的最优值点$y = \frac{2}{5}x$. 这样上述优化问题化为

$$\min f(x) = \frac{10}{x} + \frac{18x}{25}$$

该优化问题关于x为凸函数. 令其关于x的导数为零得$x = \frac{5}{3}\sqrt{5}$. 相应地, $y = \frac{2}{5}x = \frac{2}{3}\sqrt{5}$. 证毕

第10章 约束优化最优性条件

习题

1. 设$\boldsymbol{x}^*$为下述优化问题的最优解

$$\min\{f(\boldsymbol{x}) \mid \boldsymbol{l} \leqslant \boldsymbol{x} \leqslant \boldsymbol{u}\},$$

其中, $f:\mathbb{R}^n \to \mathbb{R}$连续可微, 向量$\boldsymbol{l}, \boldsymbol{u} \in \mathbb{R}^n$ 满足$\boldsymbol{l} \leqslant \boldsymbol{u}$. 试证明

$$\frac{\partial f_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}_i} = \begin{cases} \geqslant 0, & \boldsymbol{x}_i^* = \boldsymbol{l}_i, \\ \leqslant 0, & \boldsymbol{x}_i^* = \boldsymbol{u}_i, \\ 0, & \boldsymbol{l}_i < \boldsymbol{x}_i^* < \boldsymbol{u}_i. \end{cases}$$

解: 设$\boldsymbol{x}^*$为凸约束优化问题

$$\min\{f(\boldsymbol{x}) \mid \boldsymbol{l} \leqslant \boldsymbol{x} \leqslant \boldsymbol{u}\},$$

的最优解, 则$\boldsymbol{x}^*$为该优化问题的稳定点, 即

$$\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \geqslant 0, \quad \forall \boldsymbol{x} \in \Omega.$$

由此, 对任意的$i = 1, 2, \cdots, n$,

$$\nabla_i f(\boldsymbol{x}^*)(\boldsymbol{x}_i - \boldsymbol{x}_i^*) \geqslant 0 \quad \forall\ \boldsymbol{l}_i \leqslant \boldsymbol{x}_i \leqslant \boldsymbol{u}_i.$$

显然, 若$\boldsymbol{x}_i^* = \boldsymbol{l}_i$, 则对任意的$\boldsymbol{l}_i \leqslant \boldsymbol{x}_i \leqslant \boldsymbol{u}_i$, 均有$\boldsymbol{x}_i - \boldsymbol{x}_i^* \geqslant 0$. 从而$\dfrac{\partial f_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}_i} \geqslant 0$.

若$\boldsymbol{x}_i^* = \boldsymbol{u}_i$, 则对任意的$\boldsymbol{l}_i \leqslant \boldsymbol{x}_i \leqslant \boldsymbol{u}_i$, 均有$\boldsymbol{x}_i - \boldsymbol{x}_i^* \leqslant 0$. 从而$\dfrac{\partial f_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}_i} \leqslant 0$.

若$\boldsymbol{l}_i \leqslant \boldsymbol{x}_i^* \leqslant \boldsymbol{u}_i$, 分别取$\boldsymbol{x}_i = \boldsymbol{l}_i$和$\boldsymbol{x}_i = \boldsymbol{u}_i$得, $\dfrac{\partial f_i(\boldsymbol{x}^*)}{\partial \boldsymbol{x}_i} = 0$. 证毕

2. 设$\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{b} \in \mathbb{R}^m$. 试给出下述可行域点的可行方向所满足的条件.

$$\Omega_1 = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \boldsymbol{x} \geqslant \boldsymbol{0}\};$$
$$\Omega_2 = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} \geqslant \boldsymbol{b},\ \boldsymbol{x} \geqslant \boldsymbol{0}\}.$$

解: 根据定义, Ω_1 的可行方向 $\boldsymbol{d}$ 满足

$$\begin{cases} \boldsymbol{A}(\boldsymbol{x} + \alpha\boldsymbol{d}) = \boldsymbol{b}, \\ \boldsymbol{x} + \alpha\boldsymbol{d} \geqslant 0. \end{cases}$$

其中$\alpha > 0$充分小.

由于$\boldsymbol{x}\in\Omega$, 即满足

$$\boldsymbol{A}\boldsymbol{x}=\boldsymbol{b},\quad \boldsymbol{x}\geqslant 0.$$

所以$\boldsymbol{x}$点的可行方向满足

$$\boldsymbol{A}\boldsymbol{d}=\boldsymbol{0}.$$

进一步, $\boldsymbol{x}_i=0$, 则 $\boldsymbol{d}_i\geqslant 0$; 若 $\boldsymbol{x}_i>0$, 则$\boldsymbol{d}_i\in\mathbb{R}$.

综上, 可行点$\boldsymbol{x}\in\Omega_1$ 的可行方向$\boldsymbol{d}$满足

$$\begin{cases}\boldsymbol{A}\boldsymbol{d}=\boldsymbol{0}\\ \boldsymbol{d}_i\geqslant\boldsymbol{0},\ 若\boldsymbol{x}_i=0.\end{cases}$$

对$\boldsymbol{x}\in\Omega_2=\{\boldsymbol{x}\in\mathbb{R}^n\mid\boldsymbol{A}\boldsymbol{x}\geqslant\boldsymbol{b},\ \boldsymbol{x}\geqslant\boldsymbol{0}\}$. 记

$$\boldsymbol{A}=\begin{pmatrix}\boldsymbol{a}_1^{\mathrm{T}}\\ \boldsymbol{a}_2^{\mathrm{T}}\\ \vdots\\ \boldsymbol{a}_m^{\mathrm{T}}\end{pmatrix}.$$

则

$$\boldsymbol{A}\boldsymbol{x}\geqslant\boldsymbol{b}\Longleftrightarrow\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}\geqslant\boldsymbol{b}_i,\ i=1,2,\cdots,m.$$

由此, 若$\boldsymbol{x}\in\Omega_2$满足$\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{b}_i$, 则可行方向$\boldsymbol{d}$满足$\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}\geqslant 0$; 若满足$\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}>\boldsymbol{b}_i$, 则可行方向$\boldsymbol{d}$不受约束.

类似地, 若$\boldsymbol{x}\in\Omega_2$满足$\boldsymbol{x}_i=0$, 则可行方向$\boldsymbol{d}\geqslant 0$; 否则, 可行方向$\boldsymbol{d}$不受约束.

综上, 可行点$\boldsymbol{x}\in\Omega_1$ 的可行方向$\boldsymbol{d}$满足

$$\begin{cases}\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{d}\geqslant\boldsymbol{0}, 若\boldsymbol{a}_i^{\mathrm{T}}\boldsymbol{x}=\boldsymbol{b}_i,\\ \boldsymbol{d}_i\geqslant\boldsymbol{0},\ 若\boldsymbol{x}_i=0.\end{cases}$$

3. 设$f:\mathbb{R}^n\to\mathbb{R}$连续可微, $\boldsymbol{x}_0\in\mathbb{R}^n$, $\mathcal{K}$为$\mathbb{R}^n$的一个子空间. 若$\boldsymbol{x}^*\in\mathbb{R}^n$为下述优化问题的最优解

$$\min\{f(\boldsymbol{x})\mid\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}\}$$

则$\nabla f(\boldsymbol{x}^*)$与子空间$\mathcal{K}$正交, 也就是

$$\langle\nabla f(\boldsymbol{x}^*),\boldsymbol{y}\rangle=0,\quad\forall\ \boldsymbol{y}\in\mathcal{K}.$$

解: 由

$$\boldsymbol{x}^*=\arg\min\{f(\boldsymbol{x})\mid\boldsymbol{x}\in\boldsymbol{x}_0+\mathcal{K}\}$$

知

$$\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{x} - \boldsymbol{x}^* \rangle \geqslant 0, \quad \forall\ \boldsymbol{x} \in \boldsymbol{x}_0 + \mathcal{K}.$$

由于$\boldsymbol{x}^* \in \boldsymbol{x}_0 + \mathcal{K}$, 所以存在$\boldsymbol{y}^* \in \mathcal{K}$, 使得

$$\langle \nabla f(\boldsymbol{x}^*), \bar{\boldsymbol{y}} - \boldsymbol{y}^* \rangle \geqslant 0, \quad \forall\ \bar{\boldsymbol{y}} \in \mathcal{K}.$$

由于$\mathcal{K} \subset \mathbb{R}^n$为子空间, 故分别取$\bar{\boldsymbol{y}} = \boldsymbol{y} + \boldsymbol{y}^*, -\boldsymbol{y} + \boldsymbol{y}^*$得,

$$\langle \nabla f(\boldsymbol{x}^*), \boldsymbol{y} \rangle = 0, \quad \forall \boldsymbol{y} \in \mathcal{K}.$$

由此得结论.

4. 设$\boldsymbol{A} = (a_{ij})_{n\times n}$对称. 证明如下优化问题

$$\begin{aligned} \min_{\mu\in\mathbb{R}, \boldsymbol{x}\in\mathbb{R}^n} \ & \|\boldsymbol{A} - \mu \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\|_{\mathrm{F}}^2 = \sum_{i,j=1}^{n} (a_{ij} - \mu \boldsymbol{x}_i \boldsymbol{x}_j)^2 \\ \text{s.t.}\ & \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1 \end{aligned}$$

的最优解为$\boldsymbol{A}$在绝对值意义下的最大特征值及其对应的特征向量.

解: 显然, 目标函数

$$\begin{aligned} \left\|\boldsymbol{A} - \mu \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right\|_F^2 &= \|\boldsymbol{A}\|_F^2 - 2\mu \left\langle \boldsymbol{A}, \boldsymbol{x}\boldsymbol{x}^{\mathrm{T}} \right\rangle + \mu \left\|\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right\|_F^2 \\ &= \|\boldsymbol{A}\|_F^2 - 2\mu \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \mu \|\boldsymbol{x}\|^4. \end{aligned}$$

故原优化问题的Lagrange 函数为

$$L(\mu, \boldsymbol{x}, \lambda) = \|\boldsymbol{A}\|_F^2 - 2\mu \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \mu^2 \|\boldsymbol{x}\|^4 - \lambda(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} - 1).$$

由此, 原优化问题的一阶最优性条件(KKT条件)为

$$\begin{cases} \nabla_\mu L(\mu, \boldsymbol{x}, \lambda) = -2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + 2\mu\|\boldsymbol{x}\|^4 = 0, \\ \nabla_{\boldsymbol{x}} L(\mu, \boldsymbol{x}, \lambda) = -4\mu \boldsymbol{A}\boldsymbol{x} + 4\mu^2\|\boldsymbol{x}\|^2\boldsymbol{x} - 2\lambda\boldsymbol{x} = 0, \\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x} = 1 \Leftrightarrow \|\boldsymbol{x}\|^2 = 1. \end{cases}$$

其中, $\nabla_{\boldsymbol{x}}\|\boldsymbol{x}\|^4 = \nabla_{\boldsymbol{x}}\left(\|\boldsymbol{x}\|^2\|\boldsymbol{x}\|^2\right) = 2\boldsymbol{x}\|\boldsymbol{x}\|^2 + \|\boldsymbol{x}\|^2 2\boldsymbol{x} = 4\|\boldsymbol{x}\|^2\boldsymbol{x}$. 整理得,

$$\begin{cases} -2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + 2\mu = 0, \\ -4\mu\boldsymbol{A}\boldsymbol{x} + \left(4\mu^2 - 2\lambda\right)\boldsymbol{x} = 0, \\ \|\boldsymbol{x}\|^2 = 1. \end{cases}$$

进一步有,

$$\begin{cases}\lambda = 0, \\ \mu = \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}, \\ \mu\left(\mu\boldsymbol{x} - \boldsymbol{A}\boldsymbol{x}\right) = 0.\end{cases}$$

若$\mu \neq 0$, 则$\mu\boldsymbol{x} = \boldsymbol{A}\boldsymbol{x}$, 即 $(\mu, \boldsymbol{x})$ 为矩阵$\boldsymbol{A}$ 的特征对. 则由

$$\begin{aligned}\left\|\boldsymbol{A} - \mu\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right\|_F^2 &= \|\boldsymbol{A}\|_F^2 - 2\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x} + \mu\|\boldsymbol{x}\|^4 \\ &= \|\boldsymbol{A}\|_F^2 - 2\mu\boldsymbol{x}^{\mathrm{T}}\mu\boldsymbol{x} + \mu\|\boldsymbol{x}\|^4 \\ &= \|\boldsymbol{A}\|_F^2 - \mu^2\end{aligned}$$

所以, 要使目标函数值达到最小, μ为矩阵$\boldsymbol{A}$的绝对值意义下的最大特征根.$|\lambda_{\max}(\boldsymbol{A})|$表示 $\boldsymbol{A}$ 在绝对值意义下的最大特征值.

另一方面, 由于

$$\|\boldsymbol{A}\|_F^2 - \mu^2 \leq \|\boldsymbol{A}\|_F^2 = \left\|\boldsymbol{A} - 0\boldsymbol{x}\boldsymbol{x}^{\mathrm{T}}\right\|_F^2,$$

因此, $\mu = 0$肯定不是最优化问题的最优解. 由此得命题结论. 证毕

5. 设$\boldsymbol{a}, \boldsymbol{c} \in \mathbb{R}^n$, $b > 0$. 则下述优化问题

$$\begin{aligned}\min\ \ & f(\boldsymbol{x}) = \sum_{i=1}^n \frac{\boldsymbol{c}_i}{\boldsymbol{x}_i} \\ \text{s.t.}\ \ & \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = b \\ & \boldsymbol{x} \geqslant 0\end{aligned}$$

的最优值为$f^* = \frac{1}{b}\left(\sum_{i=1}^n (\boldsymbol{a}_i\boldsymbol{c}_i)^{1/2}\right)^2$.

解: 该问题的 Lagrange 函数为

$$L(\boldsymbol{x}, \lambda, \mu) = \sum_{i=1}^n \frac{\boldsymbol{c}_i}{\boldsymbol{x}_i} - \lambda\left(\sum_{i=1}^n \boldsymbol{a}_i\boldsymbol{x}_i - b\right) - \mu^{\mathrm{T}}\boldsymbol{x}.$$

得最优性条件

$$\begin{cases}\nabla_{\boldsymbol{x}_i}L(\boldsymbol{x}, \lambda, \mu) = -\frac{\boldsymbol{c}_i}{\boldsymbol{x}_i^2} - \lambda\boldsymbol{a}_i - \mu_i\boldsymbol{x}_i = 0, \quad i = 1, 2, \cdots, n, \\ \sum\limits_{i=1}^n \boldsymbol{a}_i\boldsymbol{x}_i = b; \quad \mu^{\mathrm{T}}\boldsymbol{x} = 0, \quad \boldsymbol{x} \geq 0,\ \ \mu \geq 0.\end{cases}$$

由目标函数知 $\boldsymbol{x}_i^* > 0$, $i = 1, 2, \cdots, n$. 所以$\mu^* = 0$. 由此

$$\frac{\boldsymbol{c}_i}{\boldsymbol{x}_i^*} = -\lambda^*\boldsymbol{a}_i\boldsymbol{x}_i^*,\ i = 1, 2, \cdots, n.$$

所以,

$$\sum_{i=1}^{n}\frac{\boldsymbol{c}_i}{\boldsymbol{x}_i^*}=-\lambda^*\sum_{i=1}^{n}\boldsymbol{a}_i\boldsymbol{x}_i^*=-\lambda^* b.$$

进而有

$$\lambda^*=-\frac{f(\boldsymbol{x}^*)}{b}=-\frac{f^*}{b}.$$

另一方面, 由KKT条件的第一式

$$\frac{\boldsymbol{c}_i}{\boldsymbol{x}_i^{*2}}=-\lambda^*\boldsymbol{a}_i,\ i=1,2,\cdots,n.$$

所以,

$$\frac{\boldsymbol{c}_i}{\boldsymbol{x}_i^*}=\sqrt{-\lambda^*\boldsymbol{a}_i\boldsymbol{c}_i}=\sqrt{\frac{f^*}{b}}\sqrt{\boldsymbol{a}_i\boldsymbol{c}_i},\ i=1,2,\cdots,n.$$

由此,

$$f^*=\sum_{i=1}^{n}\frac{\boldsymbol{c}_i}{\boldsymbol{x}_i^*}=\sqrt{\frac{f^*}{b}}\sum_{i=1}^{n}\sqrt{\boldsymbol{a}_i\boldsymbol{c}_i},$$

解之得

$$f^*=\frac{1}{b}\left[\sum_{i=1}^{n}(\boldsymbol{a}_i\boldsymbol{c}_i)^{1/2}\right]^2.$$

结论得证. 证毕

6. 写出下述优化问题

$$\begin{aligned}\min\ &\left(\boldsymbol{x}_1-\frac{2}{3}\right)^2+(\boldsymbol{x}_2-2)^2\\ \text{s.t.}\ &-\boldsymbol{x}_1^2+\boldsymbol{x}_2\geqslant 0\\ &\boldsymbol{x}_1+\boldsymbol{x}_2\leqslant 6\\ &\boldsymbol{x}_1,\boldsymbol{x}_2\geqslant 0\end{aligned}$$

的KKT条件, 并验证(2/3;2)为该优化问题的唯一全局最优值点.

解: 容易验证, 该优化问题为严格凸规划问题. 故其KKT点为其全局最优值点.

该优化问题的KKT条件为

$$\begin{cases}2(\boldsymbol{x}_1-\dfrac{2}{3})+2\lambda_1\boldsymbol{x}_1+\lambda_2-\mu_1=0\\ 2(\boldsymbol{x}_2-2)-\lambda_1+\lambda_2-\mu_2=0\\ -\boldsymbol{x}_1^2+\boldsymbol{x}_2\geqslant 0,\lambda_1\geqslant 0,\lambda_1(-\boldsymbol{x}_1^2+\boldsymbol{x}_2)=0\\ \boldsymbol{x}_1+\boldsymbol{x}_2-6\leqslant 0,\lambda_2\geqslant 0,\lambda_2(\boldsymbol{x}_1+\boldsymbol{x}_2-6)=0\\ \boldsymbol{x}_1\geqslant 0,\mu_1\geqslant 0,\mu_1\boldsymbol{x}_1=0\\ \boldsymbol{x}_2\geqslant 0,\mu_2\geqslant 0,\mu_2\boldsymbol{x}_2=0\end{cases}$$

容易验证, $\boldsymbol{x}=(2/3;2),\lambda_1=\lambda_2=0,\mu_1=\mu_2=0$为上述系统的解. 故由凸规划问题的最优性条件知, $(2/3;2)$为该优化问题的唯一全局最优值点.

7. 设矩阵$\boldsymbol{A}\in\mathbb{R}^{n\times n}$. 对无约束优化问题

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-\|\boldsymbol{A}\boldsymbol{x}\|$$

借助2-范数等价条件

$$\|\boldsymbol{u}\|=\max\{\boldsymbol{u}^{\mathrm{T}}\boldsymbol{v}\mid\|\boldsymbol{v}\|\leqslant 1\},$$

它可化为如下约束优化问题

$$\begin{aligned}\min\ &\ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\\ \text{s.t.}\ &\ \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\leqslant 1\end{aligned}$$

试借助KKT条件给出原优化问题的最优解和最优值.

解: 由于可行域为单位球, 故约束优化问题

$$\begin{aligned}\min\ &\ \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\\ \text{s.t.}\ &\ \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\leqslant 1\end{aligned}$$

满足M-F约束规范. 从而对其最优解$(\boldsymbol{x},\boldsymbol{y})$, 存在Lagrange乘子$\lambda\geqslant 0$使得

$$\begin{cases}\boldsymbol{x}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}\\ \boldsymbol{A}\boldsymbol{x}=\lambda\boldsymbol{y}\\ \boldsymbol{y}^{\mathrm{T}}\boldsymbol{y}\leqslant 1\end{cases}$$

由于$(\boldsymbol{x},\boldsymbol{y})$为优化问题的最优解, 故

$$\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$$

取最大值. 这说明最优解$\boldsymbol{y}^*$为矩阵$\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}$的最大特征值对应的特征向量. 从而, 原问题的最优解为$\boldsymbol{x}^*=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}^*$.

原问题的最优值为

$$\frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}-\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=-\frac{1}{2}\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}=-\frac{1}{2}\lambda_{\max}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}).$$

8. 对n阶实对称阵$\boldsymbol{A},\boldsymbol{B}$, 试证明下述结论等价.

(1) 对任意的非零向量$\boldsymbol{x}\in\mathbb{R}^n$, $\max\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x},\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}\}\geqslant 0(>0)$;

(2) 存在非负常数λ和μ满足$\lambda+\mu=1$使得$\lambda\boldsymbol{A}+\mu\boldsymbol{B}$半正定(正定).

解: 考虑优化问题

$$\begin{aligned}&\min t\\ &\text{s.t. } t-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}\geqslant 0\\ &\qquad t-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}\geqslant 0\\ &\qquad \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\geqslant 1\end{aligned}$$

显然, (1)成立, 即$\max\{\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x},\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x}\}\geqslant 0$的充分必要条件是该优化问题的最优值$t^*\geqslant 0$, 而(1)不成立的充要条件是该优化问题的最优值$t^*=-\infty$.

该优化问题的Lagrange对偶规划为

$$\begin{aligned}\max_{\mu\geqslant 0}\min_{t,\boldsymbol{x}}L(t,\mu_1,\mu_2,\mu_3)&=\max_{\mu\geqslant 0}\min_{t,\boldsymbol{x}}\left\{t-\mu_1(t-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x})-\mu_2(t-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{B}\boldsymbol{x})-\mu_3\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\right\}\\ &=\max_{\mu\geqslant 0}\min_{t,\boldsymbol{x}}\left\{t-\mu_1 t-\mu_2 t+\boldsymbol{x}^{\mathrm{T}}(\mu_1\boldsymbol{A}+\mu_2\boldsymbol{B}-\mu_3\boldsymbol{I})\boldsymbol{x}\right\}\end{aligned}$$

要使内层优化问题的最小值达到最大, 参数μ需满足

$$\begin{cases}1-\mu_1-\mu_2=0\\ \mu_1\boldsymbol{A}+\mu_2\boldsymbol{B}-\mu_3 I\text{半正定.}\end{cases}$$

从而利用内层优化问题的KKT条件可将对偶规划问题写成

$$\begin{aligned}&\max\nolimits_{\mu\geqslant 0}\ 0\\ &\text{s.t. } 1-\mu_1-\mu_2=0\\ &\qquad \mu_1\boldsymbol{A}+\mu_2\boldsymbol{B}-\mu_3 I\text{半正定.}\end{aligned}$$

由弱对偶定理, 原问题有解的充分必要条件是对偶问题相容, 即存在$\mu_1,\mu_2\geqslant 0,\mu_1+\mu_2=1$使得$\mu_1\boldsymbol{A}_1+\mu_2\boldsymbol{A}_2$半正定. 证毕

9. 设$\boldsymbol{a}\in\mathbb{R}^n_+$. 求下述优化问题的最优解

$$\begin{aligned}&\max\ \min_{1\leqslant i\leqslant n}\boldsymbol{x}_i\\ &\ \text{s.t. } \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}\leqslant 1\end{aligned}$$

解: 不妨设向量$\boldsymbol{a}$的前m个元素非零. 则

$$\begin{aligned}&\max\ \min_{1\leqslant i\leqslant n}\boldsymbol{x}_i\\ &\ \text{s.t. } \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}\leqslant 1\end{aligned}\quad\Leftrightarrow\quad\begin{aligned}&\max\ \min_{1\leqslant i\leqslant m}\boldsymbol{x}_i\\ &\ \text{s.t. } \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}\leqslant 1\end{aligned}$$

对后者, 由于$\boldsymbol{x}=\boldsymbol{0}$为其可行解, 且最优值为零, 故上述优化问题的最优解非负. 进一步, 对该优化问题, 不等式约束取等号.

对该优化问题, 我们说其最优解所有的分量都相等. 否则, 不妨设$\boldsymbol{x}_1>\boldsymbol{x}_2\geqslant\boldsymbol{x}_3\geqslant\cdots\geqslant\boldsymbol{x}_{m-1}>\boldsymbol{x}_m\geqslant 0$为该优化问题的最优解. 显然, 其最优值为$\boldsymbol{x}_m$.

定义向量$\boldsymbol{x}'$如下,

$$\begin{cases} \boldsymbol{x}_1' = \boldsymbol{x}_2 \\ \boldsymbol{x}_i' = \boldsymbol{x}_i, \quad 2 \leqslant i \leqslant m-1, \\ \boldsymbol{x}_m' = \boldsymbol{x}_m + \boldsymbol{a}_1(\boldsymbol{x}_1 - \boldsymbol{x}_2)/\boldsymbol{a}_m \end{cases}$$

则$\boldsymbol{x}'$仍为该优化问题的可行解, 而最优值为$\boldsymbol{x}_m' = \boldsymbol{x}_m + \boldsymbol{a}_1(\boldsymbol{x}_1 - \boldsymbol{x}_2)/\boldsymbol{a}_m > \boldsymbol{x}_m$. 这与假设矛盾. 所以, 该优化问题最优解的所有分量都相等. 由此, 该优化问题化为

$$\begin{aligned} &\max x \\ &\text{s.t. } x \sum_{i=1}^{m} \boldsymbol{a}_i = 1 \end{aligned}$$

容易看出, 其最优解为$x^* = 1/\sum_{i=1}^{m} \boldsymbol{a}_i$.

由此, 原优化问题

$$\begin{aligned} &\max \min_{1 \leqslant i \leqslant n} \boldsymbol{x}_i \\ &\text{s.t. } \boldsymbol{a}^{\mathrm{T}} \boldsymbol{x} \leqslant 1 \end{aligned}$$

的最优解为

$$\begin{cases} \boldsymbol{x}_i^* = 1/\sum_{i=1}^{m} \boldsymbol{a}_i, \text{若} \boldsymbol{a}_i > 0, \\ \boldsymbol{x}_i^* > 1/\sum_{i=1}^{m} \boldsymbol{a}_i, \text{若} \boldsymbol{a}_i = 0. \end{cases}$$

最优值为$1/\sum_{i=1}^{m} \boldsymbol{a}_i$. 证毕

10. 设$\boldsymbol{A} \in \mathbb{R}^{m \times n}$, $\boldsymbol{b} \in \mathbb{R}^n$. 试给出下述二次规划问题的Lagrange对偶.

$$\min \ \|\boldsymbol{x} - \boldsymbol{b}\|^2 \qquad \text{s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}$$

解: 该优化问题的 Lagrange 函数为

$$L(\boldsymbol{x}, \lambda) = \|\boldsymbol{x} - \boldsymbol{b}\|^2 - \lambda^{\mathrm{T}} \boldsymbol{A}\boldsymbol{x}, \quad \lambda \in \mathbb{R}^m, \ \boldsymbol{x} \in \mathbb{R}^n.$$

它关于$\boldsymbol{x}$为凸函数.

下求其关于$\boldsymbol{x}$的最小值解. 对此, 令

$$\nabla_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda) = 2(\boldsymbol{x} - \boldsymbol{b}) - \boldsymbol{A}^{\mathrm{T}} \lambda = 0$$

得

$$\boldsymbol{x} = \frac{1}{2} \boldsymbol{A}^{\mathrm{T}} \lambda + \boldsymbol{b}.$$

从而

$$\begin{aligned}\theta(\lambda) &= \|\frac{1}{2}\boldsymbol{A}^{\mathrm{T}}\lambda + \boldsymbol{b} - \boldsymbol{b}\|^2 - \lambda^{\mathrm{T}}\boldsymbol{A}\left(\frac{1}{2}\boldsymbol{A}^{\mathrm{T}}\lambda + \boldsymbol{b}\right) \\ &= -\frac{1}{4}\|\boldsymbol{A}^{\mathrm{T}}\lambda\|^2 - \lambda^{\mathrm{T}}\boldsymbol{A}\boldsymbol{b},\end{aligned}$$

由此, 优化问题的Lagrange 对偶为

$$\max_{\lambda\in\mathbb{R}^m} -\frac{1}{4}\|\boldsymbol{A}^{\mathrm{T}}\lambda\|^2 - \lambda^{\mathrm{T}}\boldsymbol{A}\boldsymbol{b}.$$

证毕

11. 设矩阵$\boldsymbol{A}\in\mathbb{R}^{m\times n}$行满秩, $\boldsymbol{c}\in\mathbb{R}^n$非零, 且$\boldsymbol{A}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}})^{-1}\boldsymbol{A}\boldsymbol{c}-\boldsymbol{c}\neq\boldsymbol{0}$. 求下述优化问题的最优解和最优值

$$\min\ \boldsymbol{c}^{\mathrm{T}}\boldsymbol{x} \quad \text{s.t. } \boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}, \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\leqslant 1$$

解: 显然该问题凸且 $\boldsymbol{x}=0$ 满足 Slater 条件. 从而由其KKT 条件

$$\begin{cases}\boldsymbol{c}-\boldsymbol{A}^{\mathrm{T}}\lambda+2\mu\boldsymbol{x}=0,\\ \boldsymbol{A}\boldsymbol{x}=0,\ \boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}\leqslant 1,\\ \mu(1-\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x})=0,\ \mu\geqslant 0.\end{cases}$$

第一式左乘$\boldsymbol{A}$, 则由第二式得

$$\lambda=\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{c}.$$

从而

$$\mu\boldsymbol{x}=\frac{\boldsymbol{A}^{\mathrm{T}}\lambda-\boldsymbol{c}}{2}.$$

再由$\mu=\mu\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}$得

$$\mu^2=(\mu\boldsymbol{x})^{\mathrm{T}}(\mu\boldsymbol{x}).$$

从而

$$\mu=\|\mu\boldsymbol{x}\|_2=\frac{\|\boldsymbol{A}^{\mathrm{T}}\lambda-\boldsymbol{c}\|_2}{2}=\frac{\|\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{c}-\boldsymbol{c}\|_2}{2}.$$

这样,

$$\boldsymbol{x}=\frac{\boldsymbol{A}^{\mathrm{T}}\lambda-\boldsymbol{c}}{2\mu}=\frac{\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{c}-\boldsymbol{c}}{\|\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{c}-\boldsymbol{c}\|_2}.$$

从而最优值为 $\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}=-2\mu\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}=-2\mu=-\|\boldsymbol{A}^{\mathrm{T}}\left(\boldsymbol{A}\boldsymbol{A}^{\mathrm{T}}\right)^{-1}\boldsymbol{A}\boldsymbol{c}-\boldsymbol{c}\|_2.$ 证毕

12. 建立下述优化问题的Lagrange对偶规划

$$\min\ \boldsymbol{x}_1^2+\boldsymbol{x}_2^2$$
$$\text{s.t. } \boldsymbol{x}_1+\boldsymbol{x}_2\geqslant 4$$
$$\boldsymbol{x}_1,\boldsymbol{x}_2\geqslant 0$$

验证对偶规划问题的目标函数为$\theta(u)=-u^2/2-4u$, 并且对偶间隙为零.

解: 对该优化问题, 由于

$$\boldsymbol{x}_1^2+\boldsymbol{x}_2^2\geqslant\frac{(\boldsymbol{x}_1+\boldsymbol{x}_2)^2}{2}\geqslant\frac{4^2}{2}=8,$$

且$\boldsymbol{x}_1=\boldsymbol{x}_2=2$时, 上式取等号, 所以该优化问题的最优解为 $\boldsymbol{x}_1=\boldsymbol{x}_2=2$, 最优值为 8.

基于该优化问题的Lagrange函数

$$L(\boldsymbol{x},\mu)=\boldsymbol{x}_1^2+\boldsymbol{x}_2^2-\mu\left(\boldsymbol{x}_1+\boldsymbol{x}_2-4\right)$$

可得其Lagrange对偶函数

$$\theta(\mu)=\inf\left\{L(\boldsymbol{x},\mu)\mid\boldsymbol{x}_1,\boldsymbol{x}_2\geqslant 0\right\},\quad \mu\geqslant 0.$$

由于

$$\begin{aligned}L(\boldsymbol{x},\mu)=&\ \boldsymbol{x}_1^2+\boldsymbol{x}_2^2-\mu\left(\boldsymbol{x}_1+\boldsymbol{x}_2-4\right),\\ &=\left(\boldsymbol{x}_1-\frac{\mu}{2}\right)^2+\left(\boldsymbol{x}_2-\frac{\mu}{2}\right)^2-\frac{\mu^2}{2}+4\mu,\\ &\geqslant-\frac{\mu^2}{2}+4\mu,\end{aligned}$$

且等号成立的条件是 $\boldsymbol{x}_1=\boldsymbol{x}_2=\mu/2\geqslant 0$, 所以该优化问题的Lagrange对偶函数为

$$\theta(\mu)=\ -\frac{\mu^2}{2}+4\mu=-\frac{1}{2}(\mu-4)^2+8,\quad \mu\geqslant 0.$$

从而该优化问题的Lagrange 对偶为

$$\max_{\mu\geq 0}\ -\frac{1}{2}(\mu-4)^2+8$$

显然, 该优化问题的最优解为4, 最优值为8. 结合原优化问题的讨论得, 对偶间隙为零. 证毕

13. 设$f:\mathbb{R}^n\to\mathbb{R}$为连续可微的凸函数, $\Theta\subset\mathbb{R}^n$为非空闭凸集. 试证明优化问题$\min\limits_{\boldsymbol{x}\in\Theta}\{f(\boldsymbol{x})\mid\ \boldsymbol{A}\boldsymbol{x}=\boldsymbol{b}\}$的鞍点满足如下变分不等式问题, 即求$\boldsymbol{w}^*\in\mathcal{W}$使满足

$$\boldsymbol{F}(\boldsymbol{w}^*)^{\mathrm{T}}(\boldsymbol{w}-\boldsymbol{w}^*)\geqslant 0,\quad \forall\boldsymbol{w}\in\mathcal{W},$$

其中,

$$\boldsymbol{w}=\begin{pmatrix}\boldsymbol{x}\\ \boldsymbol{\lambda}\end{pmatrix},\ F(\boldsymbol{w})=\begin{pmatrix}\nabla f(\boldsymbol{x})-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda}\\ \boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\end{pmatrix},\ \mathcal{W}=\Theta\times\mathbb{R}^m.$$

解: 设$(\boldsymbol{x}^*,\boldsymbol{\lambda}^*)$为优化问题(14.4.4)的Lagrange函数的鞍点, 即满足

$$L(\boldsymbol{x}^*;\boldsymbol{\lambda})\leqslant L(\boldsymbol{x}^*;\boldsymbol{\lambda}^*)\leqslant L(\boldsymbol{x};\boldsymbol{\lambda}^*),\quad \forall\ \boldsymbol{x}\in\Omega,\lambda\in\mathbb{R}^n,$$

其中

$$L(\boldsymbol{x};\boldsymbol{\lambda})=f(\boldsymbol{x})-\boldsymbol{\lambda}^{\mathrm{T}}(\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}).$$

由鞍点条件的后一式,

$$\boldsymbol{x}^*=\arg\min_{\boldsymbol{x}\in\Theta}L(\boldsymbol{x};\boldsymbol{\lambda}^*)$$

由Θ为闭凸集, 从而由稳定点条件得, $\boldsymbol{x}^*$满足

$$\nabla_{\boldsymbol{x}}L(\boldsymbol{x};\boldsymbol{\lambda}^*)^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*)=(\nabla f(\boldsymbol{x}^*)-\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda})^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{x}^*)\geqslant 0,\quad \forall\ \boldsymbol{x}\in\Theta.$$

由鞍点条件的前一式,

$$\boldsymbol{\lambda}^*=\arg\min_{\lambda\in\mathbb{R}^m}L(\boldsymbol{x}^*;\boldsymbol{\lambda})$$

再由稳定点条件得

$$\nabla_{\lambda}L(\boldsymbol{x}^*;\boldsymbol{\lambda}^*)^{\mathrm{T}}(\boldsymbol{\lambda}-\boldsymbol{\lambda}^*)=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}^*-\boldsymbol{b})^{\mathrm{T}}(\boldsymbol{\lambda}-\boldsymbol{\lambda}^*)\geqslant 0,\quad \forall\ \boldsymbol{\lambda}\in\mathbb{R}^m.$$

两式结合得命题结论. 证毕

14. 设$\boldsymbol{a}\in\mathbb{R}^n$. 试用KKT条件求下述优化问题的最优解和最优值

$$\min\{\sum_{i=1}^n\frac{\boldsymbol{a}_i^2}{\boldsymbol{x}_i}\mid\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}\leqslant 1,\ \boldsymbol{x}>\boldsymbol{0}\}$$

解: 显然, 该优化问题等价于

$$\min\{\sum_{i=1}^n\frac{\boldsymbol{a}_i^2}{\boldsymbol{x}_i}\mid\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x}=1,\ \boldsymbol{x}>\boldsymbol{0}\},\quad \boldsymbol{a}\geqslant 0.$$

利用KKT条件

$$\begin{cases} \frac{\boldsymbol{a}_i^2}{\boldsymbol{x}_i^2} = \mu, \quad i = 1, 2, \cdots, n \\ \sum\limits_{i=1}^{n} \boldsymbol{x}_i = 1 \end{cases}$$

的第一式得$\boldsymbol{a}_i$与$\boldsymbol{x}_i$成固定比例, $i = 1, 2, \cdots, n$. 利用第二式得$\boldsymbol{x} = \frac{\boldsymbol{a}}{\|\boldsymbol{a}\|_1}$. 代入得最优值$\|\boldsymbol{a}\|_1^2$.

15. 设$\boldsymbol{a}, \boldsymbol{x}_0 \in \mathbb{R}^n$, $\rho > 0$, 矩阵$\boldsymbol{G}$对称正定. 试借助Lagrange对偶求解如下优化问题

$$\begin{aligned} &\min\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \\ &\ \text{s.t.}\ (\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{x} - \boldsymbol{x}_0) \leqslant \rho \end{aligned}$$

解: 该优化问题的Lagrange函数为

$$L(\boldsymbol{x}, \lambda) = \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} + \lambda(\boldsymbol{x} - \boldsymbol{x}_0)^{\mathrm{T}}\boldsymbol{G}(\boldsymbol{x} - \boldsymbol{x}_0) - \lambda\rho, \quad \lambda \geqslant 0.$$

其对偶规划为

$$\max_{\lambda \geqslant 0} \min_{\boldsymbol{x}} L(\boldsymbol{x}, \lambda).$$

由于$\boldsymbol{G}$对称正定, 故内层优化为凸优化问题, 其关于$\boldsymbol{x}$的最优值点满足

$$\boldsymbol{a} + 2\lambda\boldsymbol{G}(\boldsymbol{x} - \boldsymbol{x}_0) = 0.$$

由此得

$$\boldsymbol{x} = \boldsymbol{x}_0 - \frac{1}{2\lambda}\boldsymbol{G}^{-1}\boldsymbol{a}.$$

代入Lagrange函数得对偶规划

$$\max_{\lambda \geqslant 0} \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_0 - \frac{1}{4\lambda}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{a} - \lambda\rho$$

即

$$\min_{\lambda \geqslant 0} \frac{1}{4\lambda}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{a} + \lambda\rho - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_0$$

同样, 目标函数关于$\lambda \geqslant 0$为凸函数, 故其最优值点满足

$$-\frac{1}{4\lambda^2}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{a} + \rho = 0.$$

从而, 对偶规划问题的最优解为

$$\lambda = \sqrt{\frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{a}}{4\rho}}.$$

由于原规划问题为凸规划问题, 故强对偶定理成立. 从而原规划问题的最优解为

$$\boldsymbol{x}^* = \boldsymbol{x}_0 - \sqrt{\frac{\rho}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{a}}}\boldsymbol{G}^{-1}\boldsymbol{a}.$$

进一步, 该优化问题的最优值为

$$\begin{aligned}\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}^* &= \boldsymbol{a}^{\mathrm{T}}(\boldsymbol{x}_0 - \sqrt{\frac{\rho}{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{a}}}\boldsymbol{G}^{-1}\boldsymbol{a}) \\ &= \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_0 - \sqrt{\rho\boldsymbol{a}^{\mathrm{T}}\boldsymbol{G}^{-1}\boldsymbol{a}}.\end{aligned}$$

16. 对无约束优化问题

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2+\lambda\|\boldsymbol{x}\|_1$$

引入变量$\boldsymbol{y}=\boldsymbol{x}$, 得约束优化问题

$$\begin{aligned}&\min\ \tfrac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2+\lambda\|\boldsymbol{y}\|_1 \\ &\ \text{s.t.}\ \boldsymbol{y}-\boldsymbol{x}=\boldsymbol{0}\end{aligned}$$

试借助后者的对偶给出原规划问题的一个易于求解的转化形式.

解: 该问题的Lagrange对偶为

$$\max_{\boldsymbol{z}\in\mathbb{R}^n}\min_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})$$

其中，

$$L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})=\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2+\lambda\|\boldsymbol{y}\|_1+\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{x}-\boldsymbol{y}).$$

由于内层优化问题的目标函数关于$\boldsymbol{x},\boldsymbol{y}$可分离, 所以内层优化问题可写成

$$\min_{\boldsymbol{x}\in\mathbb{R}^n}\Big(\frac{1}{2}\|\boldsymbol{A}\boldsymbol{x}-\boldsymbol{b}\|^2+\boldsymbol{z}^{\mathrm{T}}\boldsymbol{x}\Big)+\min_{\boldsymbol{y}\in\mathbb{R}^n}\Big(\lambda\|\boldsymbol{y}\|_1-\boldsymbol{z}^{\mathrm{T}}\boldsymbol{y}\Big)$$

对前一优化问题, 由于目标函数为凸函数, 故由一阶最优性条件得

$$\boldsymbol{x}^*=(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}-\boldsymbol{z}).$$

而对后一优化问题, 利用目标函数的可分离性得

$$\min_{\boldsymbol{y}\in\mathbb{R}^n}\lambda|\boldsymbol{y}|_1-\boldsymbol{z}^{\mathrm{T}}\boldsymbol{y}=\begin{cases}0, & 若\|\boldsymbol{z}\|_\infty\leqslant\lambda,\\ -\infty, & 若\|\boldsymbol{z}\|_\infty>\lambda.\end{cases}$$

这样, 对偶问题可写成

$$
\begin{aligned}
&\max_{\boldsymbol{z}\in\mathbb{R}^n}\min_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{z})\\
=&\max_{\|\boldsymbol{z}\|_\infty\leqslant\lambda}\frac{1}{2}\|\boldsymbol{A}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}-\boldsymbol{z})-\boldsymbol{b}\|^2+\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}-\boldsymbol{z})\\
=&\max_{\|\boldsymbol{z}\|_\infty\leqslant\lambda}-\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{z}+\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}\\
=&\min_{\|\boldsymbol{z}\|_\infty\leqslant\lambda}\frac{1}{2}\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{z}-\boldsymbol{z}^{\mathrm{T}}(\boldsymbol{A}^{\mathrm{T}}\boldsymbol{A})^{-1}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{b}
\end{aligned}
$$

由于原问题为线性约束的凸优化问题, 所以强对偶定理成立. 这样, 原非光滑优化问题就等价地化为一带简单约束的光滑优化问题.

第11章 二次规划

习题

1. 设$\boldsymbol{a}, \boldsymbol{x}_0 \in \mathbb{R}^n$. 试计算下述二次规划问题的最优解

$$\begin{aligned} &\min \|\boldsymbol{x}-\boldsymbol{x}_0\|^2 \\ &\text{s.t. } \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = \boldsymbol{0} \end{aligned}$$

解: 由该优化问题的KKT条件

$$\begin{cases} 2(\boldsymbol{x}-\boldsymbol{x}_0) - \lambda\boldsymbol{a} = \boldsymbol{0}, \\ \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} = 0. \end{cases}$$

解之得

$$\begin{cases} \lambda^* = \frac{-2\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_0}{\|\boldsymbol{a}\|^2}, \\ \boldsymbol{x}^* = \boldsymbol{x}_0 - \frac{\boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}_0}{\|\boldsymbol{a}\|^2}\boldsymbol{a}. \end{cases}$$

由于该规划为凸规划, 所以$\boldsymbol{x}^*$ 为最优解. 证毕

2. 设$a > b > 0$,

$$\boldsymbol{G} = \begin{pmatrix} a & b & b & \cdots & b \\ b & a & b & \cdots & b \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ b & b & b & \cdots & a \end{pmatrix}.$$

证明如下二次规划问题

$$\begin{aligned} &\min_{\boldsymbol{x}>\boldsymbol{0}} \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x} + c\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} \\ &\text{s.t. } \sum_{i=1}^{n-1} x_i < d, \quad \sum_{i=1}^{n} x_i \geqslant d \end{aligned}$$

最优解的所有分量都相等.

解: 容易计算, 矩阵$\boldsymbol{G}$的行列式

$$|\boldsymbol{G}| = [a + (n-1)b](a-b)^{n-1}.$$

由此, 矩阵G的所有顺序主子式都大于零. 故矩阵$\boldsymbol{G}$正定.

用反证法证明. 设$\boldsymbol{x} \in \mathbb{R}^n$为该二次规划的最优解, 且其分量不全相同, 不妨设$x_1 \neq x_2$. 由题设, 目标函数的最优值

$$f(\boldsymbol{x}) = \frac{1}{2}\boldsymbol{x}^{\mathrm{T}}\boldsymbol{G}\boldsymbol{x} + c\boldsymbol{e}^{\mathrm{T}}\boldsymbol{x} = \frac{1}{2}a\sum_{i=1}^{n} x_i^2 + b\sum_{\substack{i,j=1 \\ i\neq j}}^{n} x_i x_j + c\sum_{i=1}^{n} x_i$$

令$\bar{x}_1 = \bar{x}_2 = \frac{x_1+x_2}{2}$. 则$\bar{\boldsymbol{x}} = (\bar{x}_1; \bar{x}_2; x_3; \cdots; x_n)$为该问题的可行解, 对应的目标函数值为

$$f(\bar{\boldsymbol{x}}) = \frac{1}{2}\bar{\boldsymbol{x}}^{\mathrm{T}}\boldsymbol{G}\bar{\boldsymbol{x}} + c\boldsymbol{e}^{\mathrm{T}}\bar{\boldsymbol{x}} = \frac{1}{2}a\sum_{i=1}^{n}\bar{x}_i^2 + b\sum_{\substack{i,j=1\\ i\neq j}}^{n}\bar{x}_i\bar{x}_j + c\sum_{i=1}^{n}\bar{x}_i.$$

则由$\bar{x}_1 = \bar{x}_2 = \frac{x_1+x_2}{2}$知,

$$\begin{aligned}
f(\boldsymbol{x}) - f(\bar{\boldsymbol{x}}) &= (\frac{1}{2}a\sum_{i=1}^{n}x_i^2 + b\sum_{\substack{i,j=1\\ i\neq j}}^{n}x_ix_j + c\sum_{i=1}^{n}x_i) - (\frac{1}{2}a\sum_{i=1}^{n}\bar{x}_i^2 + b\sum_{\substack{i,j=1\\ i\neq j}}^{n}\bar{x}_i\bar{x}_j + c\sum_{i=1}^{n}\bar{x}_i)\\
&= [\frac{1}{2}a(x_1^2+x_2^2) + bx_1x_2 + b(x_1+x_2)\sum_{i=3}^{n}x_i + c(x_1+x_2)]\\
&\quad - [\frac{1}{2}a(\bar{x}_1^2+\bar{x}_2^2) + b\bar{x}_1\bar{x}_2 + b(\bar{x}_1+\bar{x}_2)\sum_{i=3}^{n}x_i + c(\bar{x}_1+\bar{x}_2)]\\
&= [\frac{1}{2}a(x_1+x_2)^2 + (b-a)x_1x_2] - [\frac{1}{2}a(\bar{x}_1+\bar{x}_2)^2 + (b-a)\bar{x}_1\bar{x}_2]\\
&= (b-a)(x_1x_2 - \bar{x}_1^2)\\
&= (b-a)[x_1x_2 - (\frac{x_1+x_2}{2})^2]\\
&= (a-b)\frac{(x_1-x_2)^2}{4} > 0.
\end{aligned}$$

这与$\boldsymbol{x}$为优化问题的最优解矛盾. 假设不成立. 故该优化问题的最优解的所有分量都相等.

3. 设$\sigma_1 > \sigma_2 > \cdots > \sigma_r > 0$, $s < r$. 求下述二次规划问题的最优解

$$\begin{aligned}
\max\ & \sum_{i=1}^{r}\sum_{j=1}^{s}\sigma_i x_{ij}^2\\
\text{s.t.}\ & \sum_{j=1}^{s}x_{ij} \leqslant 1, \quad i = 1,2,\cdots,r\\
& \sum_{i=1}^{r}x_{ij} \leqslant 1, \quad j = 1,2,\cdots,s\\
& x_{ij} \geqslant 0, \quad i = 1,2,\cdots,r,\ j = 1,2,\cdots,s
\end{aligned}$$

解: 定义矩阵

$$\boldsymbol{X} = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1s}\\ x_{21} & x_{22} & \cdots & x_{2s}\\ \vdots & \vdots & \vdots & \vdots\\ x_{s1} & x_{s2} & \cdots & x_{ss}\\ \vdots & \vdots & \vdots & \vdots\\ x_{r1} & x_{r2} & \cdots & x_{rs} \end{pmatrix}.$$

则上述优化问题的目标函数可表示成如下矩阵所有元素的和

$$\begin{pmatrix} \sigma_1 x_{11}^2 & \sigma_1 x_{12}^2 & \cdots & \sigma_1 x_{1r}^2 \\ \sigma_2 x_{21}^2 & \sigma_2 x_{22}^2 & \cdots & \sigma_2 x_{2s}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_s x_{s1}^2 & \sigma_s x_{s2}^2 & \cdots & \sigma_s x_{ss}^2 \\ \vdots & \vdots & \vdots & \vdots \\ \sigma_r x_{r1}^2 & \sigma_r x_{r2}^2 & \cdots & \sigma_r x_{rs}^2 \end{pmatrix}.$$

由于对任意的$x \in (0,1)$, $1 > x > x^2 > 0$, 及$\sigma_1 > \sigma_2 > \cdots > \sigma_r > 0$, 欲使目标函数$\sum\limits_{i=1}^{r}\sum\limits_{j=1}^{s}\sigma_i x_{ij}^2$取到最大值, $x_{1j}, x_{1j}, \cdots, x_{rj}$应依次取尽可能靠近1的值. 考虑到约束条件$\sum\limits_{j=1}^{s} x_{ij} \leqslant 1, \quad i = 1, 2, \cdots, r$, 和$\sum\limits_{i=1}^{r} x_{ij} \leqslant 1, \quad j = 1, 2, \cdots, s$, 该优化问题的最优解为

$$x_{ij} = \begin{cases} 1, \ i = j \leqslant r, \\ 0, \text{否则}. \end{cases}$$

4. 设矩阵$\boldsymbol{G}$对称正定, $\boldsymbol{d}_k$为如下二次规划

$$\begin{aligned} \min \quad & Q(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k) = \frac{1}{2}(\boldsymbol{x}_k + \boldsymbol{d})^{\mathrm{T}} \boldsymbol{G} (\boldsymbol{x}_k + \boldsymbol{d}) + \boldsymbol{g}^{\mathrm{T}} (\boldsymbol{x}_k + \boldsymbol{d}) \\ \text{s.t.} \quad & \boldsymbol{a}_i^{\mathrm{T}} \boldsymbol{d} = 0, \quad i \in \mathcal{S}_k \end{aligned}$$

的最优解. 证明: $Q(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$关于$\alpha \in [0,1]$严格单调递减.

解: 根据题设, $Q(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$为关于$\alpha \in \mathbb{R}$的严格凸二次函数, 且$\alpha = 1$为该二次函数的最小值点. 由二次函数的性质, $Q(\boldsymbol{x}_k + \alpha \boldsymbol{d}_k)$关于$\alpha \leqslant 1$严格单调递减, 关于$\alpha \geqslant 1$严格单调递增. 命题结论得证.

第12章 约束优化的可行方法

习题

1. 设$\bar{\boldsymbol{x}}$为下述连续可微约束优化问题

$$\min\{f(\boldsymbol{x}) \mid c_i(\boldsymbol{x}) \leqslant 0,\ i \in \mathcal{I}\}$$

的可行点, 并设$(\bar{z}, \bar{\boldsymbol{d}})$为下述线性规划子问题的最优解

$$\begin{aligned} &\min\ z \\ &\text{s.t.}\ \nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d} - z \leqslant 0 \\ &\qquad \nabla c_i(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d} - z \leqslant 0,\ i \in \mathcal{I}(\bar{\boldsymbol{x}}) \\ &\qquad -1 \leqslant \boldsymbol{d}_j \leqslant 1,\ j = 1, 2, \cdots, n \end{aligned}$$

证明: 如果$z = 0$, 则$\bar{\boldsymbol{x}}$为原规划问题的FJ点.

解: 线性规划子问题

$$\begin{aligned} &\min\ z \\ &\text{s.t.}\ \nabla f(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d} - z \leqslant 0 \\ &\qquad \nabla c_i(\bar{\boldsymbol{x}})^{\mathrm{T}}\boldsymbol{d} - z \leqslant 0,\ i \in \mathcal{I}(\bar{\boldsymbol{x}}) \\ &\qquad -1 \leqslant \boldsymbol{d}_j \leqslant 1,\ j = 1, 2, \cdots, n \end{aligned}$$

的最优值为零当且仅当系统

$$\boldsymbol{d}^{\mathrm{T}}\nabla f(\bar{\boldsymbol{x}}) < 0, \quad \boldsymbol{d}^{\mathrm{T}}\nabla c_i(\bar{\boldsymbol{x}}) > 0,\ i \in \mathcal{I}(\bar{\boldsymbol{x}})$$

无解. 由Gordan定理

设$\boldsymbol{A} \in \mathbb{R}^{n\times m}$, 则下述两系统恰有一个有解.

(1) $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} < \boldsymbol{0}$;

(2) $\boldsymbol{A}\boldsymbol{y} = \boldsymbol{0},\ \boldsymbol{y} \gneqq \boldsymbol{0}$,

知存在非负但不全为零的乘子$\boldsymbol{\lambda}$使得

$$\lambda_0\nabla f(\bar{\boldsymbol{x}}) = \sum_{i\in\mathcal{I}(\bar{\boldsymbol{x}})} \lambda_i\nabla c_i(\bar{\boldsymbol{x}}).$$

令$\lambda_i = 0, i \in \mathcal{I}\backslash\mathcal{I}(\bar{\boldsymbol{x}})$得

$$\begin{cases} \lambda_0\nabla f(\bar{\boldsymbol{x}}) = \sum\limits_{i\in\mathcal{I}} \lambda_i\nabla c_i(\bar{\boldsymbol{x}}) \\ \lambda_i \geqslant 0, c_i(\bar{\boldsymbol{x}}) \geqslant 0, \lambda_i c_i(\bar{\boldsymbol{x}}) = 0,\ i \in \mathcal{I}. \end{cases}.$$

证毕

2. 设$\boldsymbol{A} \in \mathbb{R}^{m\times n}, \boldsymbol{b} \in \mathbb{R}^m$. 证明$\boldsymbol{x}_0 \in \mathbb{R}^n$到$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$上的投影为下述关于$\boldsymbol{u}$ 的解

$$\begin{pmatrix} \boldsymbol{I} & \boldsymbol{A}^{\mathrm{T}} \\ \boldsymbol{A} & \boldsymbol{0} \end{pmatrix} \begin{pmatrix} \boldsymbol{u} \\ \boldsymbol{v} \end{pmatrix} = \begin{pmatrix} \boldsymbol{x}_0 \\ \boldsymbol{b} \end{pmatrix}.$$

解: 由投影的定义, $\boldsymbol{x}_0 \in \mathbb{R}^n$到$\Omega = \{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\}$上的投影为如下凸二次规划问题的解

$$\begin{aligned} &\min \ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{x}_0\|^2 \\ &\text{s.t. } \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b} \end{aligned}$$

其KT条件为

$$\begin{cases} \boldsymbol{x} - \boldsymbol{x}_0 = -\boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} \\ \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}. \end{cases}$$

整理得

$$\begin{cases} \boldsymbol{I}\boldsymbol{x} + \boldsymbol{A}^{\mathrm{T}}\boldsymbol{\lambda} = \boldsymbol{x}_0 \\ \boldsymbol{A}\boldsymbol{x} + \boldsymbol{0}\boldsymbol{\lambda} = \boldsymbol{b}. \end{cases}$$

由此得命题结论.

3. 设$\boldsymbol{a} \in \mathbb{R}^n$, 证明点$\boldsymbol{x} \in \mathbb{R}^n$到闭凸集$\Omega$上的投影为$P_{\boldsymbol{a}+\Omega}(\boldsymbol{x}) = \boldsymbol{a} + P_\Omega(\boldsymbol{x} - \boldsymbol{a})$.

解: 根据投影的定义,

$$\begin{aligned} \boldsymbol{y} = P_{\boldsymbol{a}+\Omega}(\boldsymbol{x}) &\Longleftrightarrow \langle \boldsymbol{y} - \boldsymbol{x}, \boldsymbol{z} - \boldsymbol{y} \rangle \geqslant 0, \quad \forall\ \boldsymbol{z} - \boldsymbol{a} \in \Omega \\ &\Longleftrightarrow \langle (\boldsymbol{y} - \boldsymbol{a}) - (\boldsymbol{x} - \boldsymbol{a}), (\boldsymbol{z} - \boldsymbol{a}) - (\boldsymbol{y} - \boldsymbol{a}) \rangle \geqslant 0, \quad \forall\ \boldsymbol{z} - \boldsymbol{a} \in \Omega \\ &\Longleftrightarrow \boldsymbol{y} - \boldsymbol{a} = P_\Omega(\boldsymbol{x} - \boldsymbol{a}) \\ &\Longleftrightarrow \boldsymbol{y} = \boldsymbol{a} + P_\Omega(\boldsymbol{x} - \boldsymbol{a}). \end{aligned}$$

4. 设$\boldsymbol{a} \in \mathbb{R}^n, b \in \mathbb{R}$. 试证明$\boldsymbol{u} \in \mathbb{R}^n$到半空间$\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \geqslant b\}$ 上的投影为

$$P(\boldsymbol{u}) = \boldsymbol{u} + \frac{\max\{0, b - \boldsymbol{a}^{\mathrm{T}}\boldsymbol{u}\}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}.$$

解: 由投影算子的定义, $\boldsymbol{u} \in \mathbb{R}^n$到半空间$\{\boldsymbol{x} \in \mathbb{R}^n \mid \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \geqslant b\}$为如下严格凸二次规划问题的最优解

$$\begin{aligned} &\min \ \frac{1}{2}\|\boldsymbol{x} - \boldsymbol{u}\|^2 = \frac{1}{2}\|\boldsymbol{x}\|^2 - \boldsymbol{x}^{\mathrm{T}}\boldsymbol{u} + \frac{1}{2}\|\boldsymbol{u}\|^2 \\ &\text{s.t. } \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x} \geqslant b \end{aligned}$$

由§11.3的讨论, 其对偶规划问题为

$$\begin{aligned}\min\ & \frac{1}{2}\|\boldsymbol{a}\|^2\lambda^2+(b-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{u})\lambda\\ \text{s.t.}\ & \lambda\geqslant 0\end{aligned}$$

显然, 其最优解为

$$\lambda^*=\max\left\{0,\frac{(b-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{u})}{\|\boldsymbol{a}\|^2}\right\}=\frac{\max\{0,b-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{u}\}}{\|\boldsymbol{a}\|^2}.$$

由于强对偶定理对该优化问题处理, 故由定理11.3.1得上述二次规划问题的最优解, 即$\boldsymbol{u}\in\mathbb{R}^n$到半空间$\{\boldsymbol{x}\in\mathbb{R}^n\mid \boldsymbol{a}^{\mathrm{T}}\boldsymbol{x}\geqslant b\}$ 上的投影为

$$\boldsymbol{x}^*=\boldsymbol{u}+\lambda^*\boldsymbol{a}=\boldsymbol{u}+\frac{\max\{0,b-\boldsymbol{a}^{\mathrm{T}}\boldsymbol{u}\}}{\|\boldsymbol{a}\|^2}\boldsymbol{a}.$$

5. 设$\boldsymbol{F}:\mathbb{R}^n\to\mathbb{R}^n$为连续映射, Ω为$\mathbb{R}^n$中的非空闭凸集. 对任意的$\boldsymbol{x}\in\mathbb{R}^n$和$\alpha\in\mathbb{R}$,记$\boldsymbol{r}(\alpha)=\boldsymbol{x}-P_\Omega(\boldsymbol{x}-\alpha\boldsymbol{F}(\boldsymbol{x}))$. 则

$$\min\{1,\alpha\}\|\boldsymbol{r}(1)\|\leqslant\|\boldsymbol{r}(\alpha)\|\leqslant\max\{1,\alpha\}\|\boldsymbol{r}(1)\|.$$

解: 由性质12.5.2, 函数$\|\boldsymbol{r}(\alpha)\|$关于$\alpha\geqslant 0$单调不减. 故

$$\|\boldsymbol{r}(\alpha)\|=\begin{cases}\geqslant\|\boldsymbol{r}(1)\|, & \alpha>1\\ \leqslant\|\boldsymbol{r}(1)\|, & 0<\alpha<1.\end{cases}$$

由性质12.5.3, 函数$\dfrac{\|\boldsymbol{r}(\alpha)\|}{\alpha}$关于$\alpha\geqslant 0$单调不增. 故

$$\frac{\|\boldsymbol{r}(\alpha)\|}{\alpha}=\begin{cases}\leqslant\|\boldsymbol{r}(1)\|, & \alpha>1\\ \geqslant\|\boldsymbol{r}(1)\|, & 0<\alpha<1.\end{cases}\quad\Longrightarrow\|\boldsymbol{r}(\alpha)\|=\begin{cases}\leqslant\alpha\|\boldsymbol{r}(1)\|, & \alpha>1\\ \geqslant\alpha\|\boldsymbol{r}(1)\|, & 0<\alpha<1.\end{cases}$$

两者结合得

$$\min\{1,\alpha\}\|\boldsymbol{r}(1)\|\leqslant\|\boldsymbol{r}(\alpha)\|\leqslant\max\{1,\alpha\}\|\boldsymbol{r}(1)\|.$$

6. 证明$\dfrac{\langle\boldsymbol{A},\boldsymbol{I}\rangle}{n}\boldsymbol{I}$为$n$阶对称阵$\boldsymbol{A}$在F-范数意义下到矩阵子空间$\{\boldsymbol{X}\in\mathbb{S}^{n\times n}\mid\boldsymbol{X}=\alpha\boldsymbol{I},\alpha\in\mathbb{R}\}$上的投影.

解: 根据投影算子的定义, n阶对称阵$\boldsymbol{A}$到矩阵子空间$\{\boldsymbol{X}\in\mathbb{S}^{n\times n}\mid\boldsymbol{X}=\alpha\boldsymbol{I},\alpha\in\mathbb{R}\}$上的投影$\boldsymbol{Y}$为如下严格凸二次规划问题的最优解

$$\begin{aligned}\min\ & \frac{1}{2}\|\boldsymbol{A}-\alpha\boldsymbol{I}\|_{\mathrm{F}}^2\\ \text{s.t.}\ & \alpha\geqslant 0\end{aligned}$$

显然,

$$\begin{aligned}\frac{1}{2}\|\boldsymbol{A}-\alpha\boldsymbol{I}\|_{\mathrm{F}}^2&=\frac{1}{2}\|\boldsymbol{A}\|_{\mathrm{F}}^2-\alpha\mathrm{tr}(\boldsymbol{A})+\frac{\mathrm{n}}{2}\alpha^2\\&=\frac{1}{2}\|\boldsymbol{A}\|_{\mathrm{F}}^2+\frac{n}{2}\Big(-\frac{2}{n}\alpha\mathrm{tr}(\boldsymbol{A})+\alpha^2\Big)\\&=\frac{1}{2}\|\boldsymbol{A}\|_{\mathrm{F}}^2+\frac{n}{2}\Big(\alpha-\frac{1}{n}\mathrm{tr}(\boldsymbol{A})\Big)^2-\frac{1}{2\mathrm{n}}\mathrm{tr}^2(\boldsymbol{A}).\end{aligned}$$

由此, 上述优化问题的最优解为$\alpha^*=\frac{1}{n}\mathrm{tr}(\boldsymbol{A})$. 相应地, n阶对称阵$\boldsymbol{A}$到矩阵子空间$\{\boldsymbol{X}\in\mathbb{S}^{n\times n}\mid\boldsymbol{X}=\alpha\boldsymbol{I},\alpha\in\mathbb{R}\}$上的投影$\boldsymbol{Y}$为$\frac{1}{n}\mathrm{tr}(\boldsymbol{A})\boldsymbol{I}$.

7. 设$\boldsymbol{a}_1,\boldsymbol{a}_2,\cdots,\boldsymbol{a}_m\in\mathbb{R}^n$, Ω为$\mathbb{R}^n$中的闭凸集. 则优化问题

$$\begin{aligned}\min\ &\sum_{i=1}^m\|\boldsymbol{x}-\boldsymbol{a}_i\|^2\\ \text{s.t. }&\boldsymbol{x}\in\Omega\end{aligned}$$

的最优解为$\dfrac{1}{m}\sum\limits_{i=1}^m\boldsymbol{a}_i$到$\Omega$上的投影.

解: 根据投影算子的定义, $\dfrac{1}{m}\sum\limits_{i=1}^m\boldsymbol{a}_i$到$\Omega$上的投影为如下严格凸二次规划问题的最优解

$$\begin{aligned}\min\ &\|\boldsymbol{x}-\frac{1}{m}\sum_{i=1}^m\boldsymbol{a}_i\|^2\\ \text{s.t. }&\boldsymbol{x}\in\Omega\end{aligned}$$

其稳定点条件为

$$\langle\boldsymbol{x}^*-\frac{1}{m}\sum_{i=1}^m\boldsymbol{a}_i,\boldsymbol{x}-\boldsymbol{x}^*\rangle\geqslant 0,\ \ \forall\boldsymbol{x}\in\Omega.$$

严格凸二次规划

$$\begin{aligned}\min\ &\sum_{i=1}^m\|\boldsymbol{x}-\boldsymbol{a}_i\|^2\\ \text{s.t. }&\boldsymbol{x}\in\Omega\end{aligned}$$

的稳定点条件为

$$\langle m\boldsymbol{x}^*-\sum_{i=1}^m\boldsymbol{a}_i,\boldsymbol{x}-\boldsymbol{x}^*\rangle\geqslant 0,\ \ \forall\boldsymbol{x}\in\Omega.$$

由于严格凸二次规划问题的解与稳定点等价, 故结论成立.