

# 《非线性最优化理论与方法》

习题解答

# 目 录

第1章 引论 . . . . .	1
第2章 凸分析基础 . . . . .	4
第3章 无约束优化最优性条件 . . . . .	12
第4章 线搜索方法 . . . . .	17
第5章 最速下降法与牛顿方法 . . . . .	19
第6章 共轭梯度法 . . . . .	20
第7章 最小二乘问题 . . . . .	23
第8章 临近点算法 . . . . .	26
第9章 交替极小化方法 . . . . .	27
第10章 约束优化最优性条件 . . . . .	28
第11章 二次规划 . . . . .	41
第12章 约束优化的可行方法 . . . . .	44

# 第1章 引论

## 习题

### 1. 证明

$$\min_{\mathbf{x} \in \mathbb{R}^n} f(\mathbf{x}) \iff \min_{\mathbf{x} \in \mathbb{R}^n, t \in \mathbb{R}} \{t \mid t - f(\mathbf{x}) \geq 0\}.$$

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

0. 所以优化问题

$$\min\{t \mid t - f(\mathbf{x}) \geq 0\}$$

的最优值为  $f(\mathbf{x}^*)$ . 即

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

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

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

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

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

从而由三角不等式,

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} &\leq \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\| + \|\mathbf{x}_k - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} \\ &= \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_k - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = 1. \end{aligned}$$

另一方面,

$$\begin{aligned} \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}_k\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} &\geq \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_{k+1} - \mathbf{x}^*\| - \|\mathbf{x}_k - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} \\ &= \lim_{k \rightarrow \infty} \frac{\|\mathbf{x}_k - \mathbf{x}^*\|}{\|\mathbf{x}_k - \mathbf{x}^*\|} = 1. \end{aligned}$$

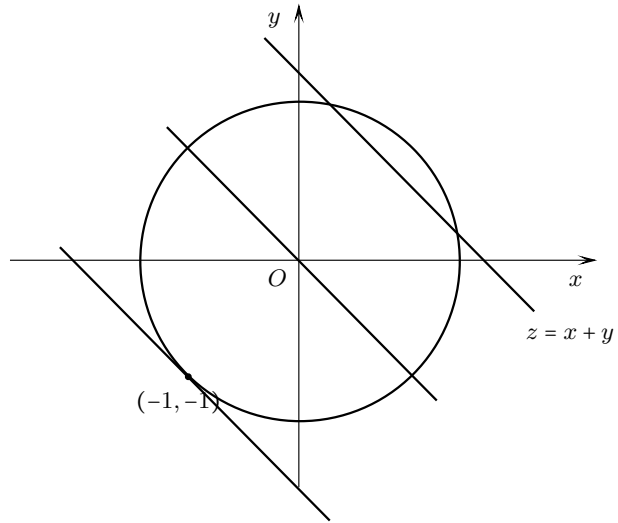
两式结合得命题结论.

证毕

3. 用图解法求下述优化问题的最优解:

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

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



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

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

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

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

即

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

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

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

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

$$\begin{aligned} \max_{\Sigma > \mathbf{0}} \quad & \mathbf{a}^\top \Sigma \mathbf{a} \\ \text{s.t.} \quad & \|\Sigma - \Sigma_0\|_F \leq \rho \end{aligned}$$

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

$$\begin{aligned} \max \quad & \mathbf{a}^\top \Sigma_0 \mathbf{a} + \mathbf{a}^\top \Delta\Sigma \mathbf{a} \\ \text{s.t.} \quad & \|\Delta\Sigma\|_F \leq \rho \end{aligned}$$

略去常数项得

$$\begin{aligned} \max \quad & \mathbf{a}^T \Delta \Sigma \mathbf{a} \\ \text{s.t.} \quad & \|\Delta \Sigma\|_F \leq \rho \end{aligned}$$

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

$$\mathbf{a}^T \Delta \Sigma \mathbf{a} = \langle \Delta \Sigma, \mathbf{a} \mathbf{a}^T \rangle \leq \|\Delta \Sigma\|_F \|\mathbf{a} \mathbf{a}^T\|_F.$$

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

由于  $\|\Delta \Sigma\|_F \leq \rho$ . 显然, 在  $\Delta \Sigma$  与矩阵  $\mathbf{a} \mathbf{a}^T$  同向时, 不等式取等号, 且在  $\|\Delta \Sigma\|_F = \rho$  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

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

进而上述优化问题的最优值为

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

## 第2章 凸分析基础

### 习题

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

(2) 设  $x_i > 0, i = 1, 2, \dots, n$ . 证明

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

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  $\alpha \in (0, 1)$ ,

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

而这依次等价于

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

利用对数函数的单调性得,

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

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  $x_1, x_2, \dots, x_n \in (0, \infty)$ ,

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

而这依次等价于

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

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

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

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,

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

证毕

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

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

为凸集.

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

$$f(x_1) \leq \alpha_1, \quad f(x_2) \leq \alpha_2.$$

由  $f(x)$  为凸函数知

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

从而,

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

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

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

由上图是凸集, 则

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

也就是

$$f(\lambda x_1 + (1 - \lambda)x_2) \leq \lambda f(x_1) + (1 - \lambda)f(x_2). \quad \text{证毕}$$

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

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

$$A_1 = \text{co} \{a_1, a_2, \dots, a_m\} = \left\{ \sum_{j=1}^m \lambda_j a_j \mid \sum_{j=1}^m \lambda_j = 1, \lambda_j \geq 0, j = 1, 2, \dots, m \right\}.$$

仿射集为

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

子空间为

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

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

$$(1) \mathbf{A}\mathbf{x} \geq \mathbf{0}, \mathbf{b}^T \mathbf{x} > 0, \mathbf{x} \geq \mathbf{0};$$

$$(2) \mathbf{A}^T \mathbf{y} \geq \mathbf{b}, \mathbf{y} \leq \mathbf{0}.$$

解: 显然, 系统(1)式等价于

$$\begin{pmatrix} \mathbf{A} \\ \mathbf{I} \end{pmatrix} \mathbf{x} \geq \mathbf{0}, \quad \mathbf{b}^T \mathbf{x} > 0.$$



由Farkas引理,

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

$$(1) \mathbf{A}^T \mathbf{x} \geq \mathbf{0}, \mathbf{b}^T \mathbf{x} < 0;$$

$$(2) \mathbf{A} \mathbf{y} = \mathbf{b}, \mathbf{y} \geq \mathbf{0}.$$

上述系统有解等价于下述系统无解

$$(\mathbf{A}^T \quad \mathbf{I}) \mathbf{y} = -\mathbf{b}, \mathbf{y} \geq \mathbf{0}.$$

也就是系统  $\mathbf{A}^T \mathbf{y} \leq -\mathbf{b}, \mathbf{y} \geq \mathbf{0}$  无解. 也即系统  $\mathbf{A}^T \mathbf{y} \geq \mathbf{b}, \mathbf{y} \leq \mathbf{0}$  无解. 证毕

6. 设  $S$  为  $\mathbb{R}^n$  的闭凸集. 若存在  $\mathbf{x}_0 \in S$  和  $\mathbf{d} \in \mathbb{R}^n$ , 使对任意的  $\alpha \geq 0$ , 均有  $\mathbf{x}_0 + \alpha \mathbf{d} \in S$ , 则对任意的  $\mathbf{x} \in S$  和  $\beta \geq 0$ , 均有  $\mathbf{x} + \beta \mathbf{d} \in S$ .

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

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

令  $\lambda = \frac{1}{\alpha} \rightarrow 0$ , 并利用集合的闭性得,  $\mathbf{x} + t \mathbf{d} \in S$ . 证毕

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

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

的次梯度.

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

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

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

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

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

$$\Delta = d_S^2(\mathbf{x} + \mathbf{h}) - d_S^2(\mathbf{x}).$$

根据定义,

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

所以,

$$\begin{aligned} \Delta &= d_S^2(\mathbf{x} + \mathbf{h}) - d_S^2(\mathbf{x}) \\ &\geq \|\mathbf{x} + \mathbf{h} - P_S(\mathbf{x} + \mathbf{h})\|^2 - \|\mathbf{x} - P_S(\mathbf{x} + \mathbf{h})\|^2 \\ &= \|\mathbf{h}\|^2 + 2\mathbf{h}^\top(\mathbf{x} - P_S(\mathbf{x} + \mathbf{h})). \end{aligned}$$

利用投影算子的非扩张性

$$\|P_S(\mathbf{x} + \mathbf{h}) - P_S(\mathbf{x})\| \leq \|\mathbf{h}\|$$

得

$$\begin{aligned} \Delta &\geq \|\mathbf{h}\|^2 + 2\mathbf{h}^\top(\mathbf{x} - P_S(\mathbf{x} + \mathbf{h})) \geq \|\mathbf{h}\|^2 + 2\mathbf{h}^\top(\mathbf{x} - P_S(\mathbf{x})) - o(\|\mathbf{h}\|) \\ &= 2\mathbf{h}^\top(\mathbf{x} - P_S(\mathbf{x})) + o(\|\mathbf{h}\|). \end{aligned} \quad (2)$$

另一方面, 由

$$d_S^2(\mathbf{x} + \mathbf{h}) \leq \|\mathbf{x} + \mathbf{h} - P_S(\mathbf{x})\|^2$$

得

$$\begin{aligned} \Delta &= d_S^2(\mathbf{x} + \mathbf{h}) - d_S^2(\mathbf{x}) \\ &\leq \|\mathbf{x} + \mathbf{h} - P_S(\mathbf{x})\|^2 - \|\mathbf{x} - P_S(\mathbf{x})\|^2 \\ &= \|\mathbf{h}\|^2 + 2\mathbf{h}^\top(\mathbf{x} - P_S(\mathbf{x})). \end{aligned}$$

结合(2)得

$$\Delta = 2\mathbf{h}^\top(\mathbf{x} - P_S(\mathbf{x})) + o(\|\mathbf{h}\|).$$

从而

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

代入(1)得

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

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

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

$$d_S(\mathbf{y}) - d_S(\mathbf{x}) = d_S(\mathbf{y}) \geq \xi^\top(\mathbf{y} - \mathbf{x}), \quad \forall \mathbf{y} \in \mathbb{R}^n.$$

取  $\mathbf{y} \in S$ , 则

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

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

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

$$\|\xi\|^2 = \xi^T(\mathbf{y} - \mathbf{x}) \leq d_S(\mathbf{x} + \xi) \leq \|\mathbf{x} + \xi - \mathbf{x}\| = \|\xi\|.$$

这说明  $\|\xi\| \leq 1$ .

结合前面的讨论得,

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

下证反向包含式成立.

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

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

另一方面, 由极锥  $N_S(\mathbf{x})$  的定义,

$$\xi^T(P_S(\mathbf{y}) - \mathbf{x}) \leq 0.$$

即

$$\xi^T P_S(\mathbf{y}) \leq \xi^T \mathbf{x}.$$

所以, 由  $\|\xi\| \leq 1$  得

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

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

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

结合(3)式得

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

综上所述得,

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

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

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

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

$$\Delta = \delta_S(\mathbf{x} + \mathbf{h}) - \delta_S(\mathbf{x}) \equiv 0.$$

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

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

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

令  $\mathbf{y} \in S$  得

$$0 = \delta_S(\mathbf{y}) \geq \xi^T(\mathbf{y} - \mathbf{x}).$$

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

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

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

令  $\mathbf{y} \in S$  得

$$\delta_S(\mathbf{y}) - \delta_S(\mathbf{x}) = -\infty \geq \xi^T(\mathbf{y} - \mathbf{x}).$$

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

综上所述得,

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

证毕

### 第3章 无约束优化最优性条件

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

$$\max\{\mathbf{b}^T \mathbf{x} \mid \mathbf{x}^T \mathbf{Q} \mathbf{x} \leq 1\}$$

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

$$(\mathbf{x}^T \mathbf{y})^2 \leq (\mathbf{x}^T \mathbf{Q} \mathbf{x})(\mathbf{y}^T \mathbf{Q}^{-1} \mathbf{y}).$$

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

$$\max\{\mathbf{b}^T \mathbf{B}^{-1} \mathbf{y} \mid \mathbf{y}^T \mathbf{y} \leq 1\}$$

也就是

$$\max\{(\mathbf{B}^{-T} \mathbf{b})^T \mathbf{y} \mid \mathbf{y}^T \mathbf{y} \leq 1\}$$

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

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

另外,

$$\begin{aligned} (\mathbf{x}^T \mathbf{y})^2 &= \mathbf{y}^T \mathbf{x} \mathbf{x}^T \mathbf{y} = \langle \mathbf{x} \mathbf{x}^T, \mathbf{y} \mathbf{y}^T \rangle, \\ &= \langle \mathbf{x} \mathbf{x}^T, \mathbf{Q} \mathbf{Q}^{-1} \mathbf{y} \mathbf{y}^T \rangle = \langle \mathbf{x} \mathbf{x}^T \mathbf{Q}, \mathbf{Q}^{-1} \mathbf{y} \mathbf{y}^T \rangle, \\ &\leq \|\mathbf{x} \mathbf{x}^T \mathbf{Q}\|_F \|\mathbf{Q}^{-1} \mathbf{y} \mathbf{y}^T\|_F = \text{tr}(\mathbf{x} \mathbf{x}^T \mathbf{Q}) \text{tr}(\mathbf{Q}^{-1} \mathbf{y} \mathbf{y}^T), \\ &= (\mathbf{x}^T \mathbf{Q} \mathbf{x})(\mathbf{y}^T \mathbf{Q}^{-1} \mathbf{y}). \end{aligned}$$

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

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

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

$$\min_{\mathbf{x} \in \mathbb{R}^n} \|\mathbf{Ax} - \mathbf{b}\|^2$$

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

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

由于 $\nabla^2\|\mathbf{Ax}-\mathbf{b}\|^2 = \mathbf{A}^T\mathbf{A}$ 半正定, 故 $\|\mathbf{Ax}-\mathbf{b}\|^2$ 关于 $\mathbf{x} \in \mathbb{R}^n$ 为凸函数, 所以满足 $\mathbf{A}^T(\mathbf{Ax}-\mathbf{b}) = 0$ 的 $\mathbf{x}$ 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 $\mathbf{A}^T\mathbf{A}$ 正定时, 最优解唯一. 证毕

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

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

解: 对该优化问题, 由一阶最优性条件, 最优解满足

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

此即最优解为

$$\mathbf{x}^* = \frac{1}{m} \sum_{i=1}^m \mathbf{a}_i. \quad \text{证毕}$$

5. 设 $\mathbf{a} \in \mathbb{R}^n, \lambda > 0$ . 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

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

解: 前一优化问题等价于

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

这样, 要求解原问题, 只需求解如下子问题

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

对此, 我们分情况讨论.

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

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

利用优化问题的最优性条件, 上述问题的最优解为

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

(2) 若 $a_i < 0$ , 则子问题的最优解非正. 该子问题可化为

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



利用优化问题的最优性条件, 该子问题的最优解为

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

引入符号函数

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

则问题的最优解为

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

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

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

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

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

从而, 最优值

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

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

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

从而, 最优值

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

综上所述, 优化问题

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

## 第4章 线搜索方法

### 习题

1. 设函数  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  连续可微,  $\mathbf{d} \in \mathbb{R}^n$  为该函数在点  $\mathbf{x} \in \mathbb{R}^n$  的下降方向. 试建立函数  $\phi(\alpha) = f(\mathbf{x} + \alpha\mathbf{d})$  在  $\alpha \geq 0$  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  $\mathbf{d}$  为点  $\mathbf{x} \in \mathbb{R}^n$  的下降方向, 故

$$f'_\alpha(\mathbf{x} + \alpha\mathbf{d})|_{\alpha=0} = \mathbf{d}^\top \nabla f(\mathbf{x}) < 0.$$

这说明,  $f(\mathbf{x} + \alpha\mathbf{d})$  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  $\phi(\alpha) = f(\mathbf{x} + \alpha\mathbf{d})$  关于  $\alpha > 0$  的最小值点满足

$$f'_\alpha(\mathbf{x} + \alpha\mathbf{d}) = \mathbf{d}^\top \nabla f(\mathbf{x} + \alpha\mathbf{d}) = 0.$$

如果函数  $f(\mathbf{x} + \alpha\mathbf{d})$  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  $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^\top \mathbf{A}\mathbf{x} + \mathbf{b}^\top \mathbf{x}$  在  $\mathbf{x}_k$  点沿下降方向  $\mathbf{d}_k$  的最优步长. 若取  $\mathbf{d}_k = -\mathbf{g}_k$ , 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,

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

则由

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

知最优步长

$$\alpha_k = \frac{-\mathbf{g}_k^\top \mathbf{d}_k}{\mathbf{d}_k^\top \mathbf{A}\mathbf{d}_k},$$

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

若  $\mathbf{d}_k = -\mathbf{g}_k$ , 则  $\alpha_k = \frac{\mathbf{g}_k^\top \mathbf{g}_k}{\mathbf{g}_k^\top \mathbf{A}\mathbf{g}_k}$ . 从而,

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

由此,目标函数在该迭代步的下降量为

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

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  $f(\mathbf{x}) = \frac{1}{2}\mathbf{x}^T \mathbf{A}\mathbf{x} + \mathbf{b}^T \mathbf{x} + c$ , 其中  $\mathbf{A}$  对称正定,

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

对初始点  $\mathbf{x}_0$ , 迭代过程为

$$\mathbf{d}_1 = -\mathbf{A}^{-1}\mathbf{g}_0 = -\mathbf{A}^{-1}(\mathbf{A}\mathbf{x}_0 + \mathbf{b}) = -\mathbf{x}_0 - \mathbf{A}^{-1}\mathbf{b};$$

$$\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{d}_1 = -\mathbf{A}^{-1}\mathbf{b}.$$

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

证毕

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

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

的下降算法. 分析  $\mathbf{B}_k$  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

$$\begin{aligned} f(\mathbf{x}_k + \alpha \mathbf{d}_k) &= f(\mathbf{x}_k) + \alpha \nabla f(\mathbf{x}_k)^T \mathbf{d}_k + o(\alpha \|\mathbf{d}_k\|) \\ &= f(\mathbf{x}_k) + \alpha \mathbf{g}_k^T \mathbf{B}_k^{-1} \mathbf{g}_k + o(\alpha \|\mathbf{d}_k\|) \leq f(\mathbf{x}_k). \end{aligned}$$

若  $\mathbf{B}_k = \mathbf{I}$ , 则该搜索方向为最速下降方向, 收敛速度至多为线性的; 若  $\mathbf{B}_k = \mathbf{G}_k$ , 目标函数  $f(\mathbf{x})$  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  $\mathbf{A} \in \mathbb{R}^{n \times n}$  对称正定, 向量组  $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$  关于矩阵  $\mathbf{A}$  共轭. 试证明:

$$(1) \text{ 对任意的 } \mathbf{x} \in \mathbb{R}^n, \mathbf{x} = \sum_{i=1}^n \frac{\mathbf{d}_i^\top \mathbf{A} \mathbf{x}}{\mathbf{d}_i^\top \mathbf{A} \mathbf{d}_i} \mathbf{d}_i;$$

$$(2) \mathbf{A}^{-1} = \sum_{i=1}^n \frac{\mathbf{d}_i \mathbf{d}_i^\top}{\mathbf{d}_i^\top \mathbf{A} \mathbf{d}_i}.$$

解: (1) 由向量组  $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n$  关于矩阵  $\mathbf{A}$  共轭知,  $\mathbf{d}_1, \mathbf{d}_2, \dots, \mathbf{d}_n \in \mathbb{R}^n$  线性无关.

故对任意的  $\mathbf{x} \in \mathbb{R}^n$ , 存在数组  $\alpha_1, \alpha_2, \dots, \alpha_n$  使得

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

从而

$$\langle \mathbf{x}, \mathbf{A} \mathbf{d}_j \rangle = \left\langle \sum_{i=1}^n \alpha_i \mathbf{d}_i, \mathbf{A} \mathbf{d}_j \right\rangle = \sum_{i=1}^n \alpha_i \langle \mathbf{d}_i, \mathbf{A} \mathbf{d}_j \rangle = \alpha_j \mathbf{d}_j^\top \mathbf{A} \mathbf{d}_j, \quad j = 1, 2, \dots, n.$$

由此,

$$\alpha_j = \frac{\langle \mathbf{x}, \mathbf{A} \mathbf{d}_j \rangle}{\mathbf{d}_j^\top \mathbf{A} \mathbf{d}_j} = \frac{\mathbf{d}_j^\top \mathbf{A} \mathbf{x}}{\mathbf{d}_j^\top \mathbf{A} \mathbf{d}_j}, \quad j = 1, 2, \dots, n.$$

也即

$$\mathbf{x} = \sum_{i=1}^n \frac{\mathbf{d}_i^\top \mathbf{A} \mathbf{x}}{\mathbf{d}_i^\top \mathbf{A} \mathbf{d}_i} \mathbf{d}_i.$$

(2) 设  $\mathbf{A}^{-1} = (\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n)$ ,  $\mathbf{a}_j \in \mathbb{R}^n$ ,  $j = 1, 2, \dots, n$ . 则由  $\mathbf{I} = \mathbf{A} \mathbf{A}^{-1} = (\mathbf{A} \mathbf{a}_1, \mathbf{A} \mathbf{a}_2, \dots, \mathbf{A} \mathbf{a}_n)$

知  $\mathbf{A} \mathbf{a}_j = \mathbf{e}_j$ ,  $j = 1, 2, \dots, n$ , 其中  $\mathbf{e}_j \in \mathbb{R}^n$  表示第  $j$  个单位向量.

由(1)知,

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

从而

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

证毕

2. 在共轭梯度法的迭代格式中, 若参数 $\beta$ 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

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

解: 先给出DY共轭梯度法下的迭代格式

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

显然,

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

从而由 $\beta_k$ 的迭代公式,

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

展开得

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

整理得

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

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 $|\beta_k| \leq \beta_k^{FR}$ , 则

$$-\frac{1}{1-\sigma_2} \leq \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} \leq \frac{2\sigma_2-1}{1-\sigma_2}.$$

解: 先给出参数 $\beta$ 的F-R公式

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

下用数学归纳法证明.

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

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

从而由 $\sigma \in (0, 1/2)$ 得,

$$-\frac{1}{1-\sigma_2} \leq \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} = -1 \leq \frac{2\sigma_2-1}{1-\sigma_2}.$$

设结论对 $k$ 成立, 即

$$-\frac{1}{1-\sigma_2} \leq \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} \leq \frac{2\sigma_2-1}{1-\sigma_2}.$$

则对任意的  $|\beta_k| \leq \beta_k^{FR} = \|\mathbf{g}_{k+1}\|^2 / \|\mathbf{g}_k\|^2$ , 由  $\mathbf{d}_{k+1} = -\mathbf{g}_{k+1} + \beta_k \mathbf{d}_k$  得

$$\frac{\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}}{\|\mathbf{g}_{k+1}\|^2} = -1 + \beta_k \frac{\mathbf{g}_{k+1}^T \mathbf{d}_k}{\|\mathbf{g}_{k+1}\|^2}.$$

从而

$$\left| \frac{\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}}{\|\mathbf{g}_{k+1}\|^2} + 1 \right| = |\beta_k| \frac{|\mathbf{g}_{k+1}^T \mathbf{d}_k|}{\|\mathbf{g}_{k+1}\|^2} \leq \beta_k^{FR} \frac{|\mathbf{g}_{k+1}^T \mathbf{d}_k|}{\|\mathbf{g}_{k+1}\|^2} = \frac{|\mathbf{g}_{k+1}^T \mathbf{d}_k|}{\|\mathbf{g}_k\|^2}.$$

另一方面, 由强 Wolfe 步长规则

$$\begin{aligned} f(\mathbf{x}_k + \alpha \mathbf{d}_k) &\leq f(\mathbf{x}_k) + \sigma_1 \alpha \mathbf{g}_k^T \mathbf{d}_k, \\ |\nabla f(\mathbf{x}_k + \alpha \mathbf{d}_k)^T \mathbf{d}_k| &\leq \sigma_2 |\mathbf{g}_k^T \mathbf{d}_k| \end{aligned}$$

其中,  $0 < \sigma_1 < \sigma_2 < 1$ , 及  $\mathbf{g}_k^T \mathbf{d}_k < 0$  得,

$$|\mathbf{g}_{k+1}^T \mathbf{d}_k| \leq \sigma_2 |\mathbf{g}_k^T \mathbf{d}_k|.$$

从而,

$$\left| \frac{\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}}{\|\mathbf{g}_{k+1}\|^2} + 1 \right| \leq \sigma_2 \frac{|\mathbf{g}_k^T \mathbf{d}_k|}{\|\mathbf{g}_k\|^2}.$$

整理得

$$-1 + \sigma_2 \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2} \leq \frac{\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}}{\|\mathbf{g}_{k+1}\|^2} \leq -1 - \sigma_2 \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2}.$$

由结论

$$-\frac{1}{1-\sigma_2} \leq \frac{\mathbf{g}_k^T \mathbf{d}_k}{\|\mathbf{g}_k\|^2}$$

得

$$-\frac{1}{1-\sigma_2} \leq \frac{\mathbf{g}_{k+1}^T \mathbf{d}_{k+1}}{\|\mathbf{g}_{k+1}\|^2} \leq \frac{2\sigma_2-1}{1-\sigma_2}.$$

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

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

的最优解满足方程

$$\begin{pmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A}^\top & -\mathbf{B}^\top \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}.$$

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

$$\nabla_{\mathbf{x}} \left\| \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \right\|^2 = (\mathbf{A}^\top \quad \mathbf{B}^\top) \left[ \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \mathbf{x} - \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix} \right] = \mathbf{0}.$$

整理得

$$\mathbf{A}^\top \mathbf{b} = (\mathbf{A}^\top \mathbf{A} + \mathbf{B}^\top \mathbf{B}) \mathbf{x}.$$

另一方面, 线性方程组

$$\begin{pmatrix} \mathbf{I} & \mathbf{A} \\ \mathbf{A}^\top & -\mathbf{B}^\top \mathbf{B} \end{pmatrix} \begin{pmatrix} \mathbf{y} \\ \mathbf{x} \end{pmatrix} = \begin{pmatrix} \mathbf{b} \\ \mathbf{0} \end{pmatrix}$$

等价于如下方程组问题

$$\begin{cases} \mathbf{y} + \mathbf{A}\mathbf{x} = \mathbf{b}, \\ \mathbf{A}^\top \mathbf{y} - \mathbf{B}^\top \mathbf{B}\mathbf{x} = \mathbf{0}. \end{cases}$$

而该方程组关于  $\mathbf{x}$  的解满足

$$\mathbf{A}^\top \mathbf{b} = (\mathbf{A}^\top \mathbf{A} + \mathbf{B}^\top \mathbf{B}) \mathbf{x}.$$

结合前一结论得证.

证毕

2. 对非线性最小二乘问题

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

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

$$\lim_{\mu \rightarrow 0} \mathbf{d}^{\text{LM}}(\mu) = \mathbf{d}^{\text{GN}}, \quad \lim_{\mu \rightarrow \infty} \frac{\mathbf{d}^{\text{LM}}(\mu)}{\|\mathbf{d}^{\text{LM}}(\mu)\|} = \frac{\mathbf{d}^s}{\|\mathbf{d}^s\|},$$

其中,  $\mathbf{d}^s$  为目标函数在  $\mathbf{x}$  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

$$\mathbf{g}(\mathbf{x}) = \nabla \left( \frac{1}{2} \mathbf{r}(\mathbf{x})^T \mathbf{r}(\mathbf{x}) \right) = \mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x}).$$

目标函数的L-M方向为

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

显然,

$$\lim_{\mu \rightarrow 0} \mathbf{d}^{\text{LM}}(\mu) = -(\mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}))^{-1} \mathbf{J}(\mathbf{x})^T \mathbf{r}(\mathbf{x}) = \mathbf{d}^{\text{GN}}.$$

下讨论  $\mu \rightarrow \infty$  的情况.

对任意发散到  $\infty$  的正数列  $\{\mu_k\}$ , 数列  $\left\{ \frac{\mathbf{d}(\mu_k)}{\|\mathbf{d}(\mu_k)\|} \right\}$  有界. 故有收敛子列, 不妨设为其本身, 即  $\left\{ \frac{\mathbf{d}(\mu_k)}{\|\mathbf{d}(\mu_k)\|} \right\}$  收敛.  
由  $\mathbf{d}(\mu_k)$  的定义,

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

所以,

$$\begin{aligned} -\frac{\mathbf{g}(\mathbf{x})}{\|\mathbf{g}(\mathbf{x})\|} &= \frac{(\mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) + \mu_k \mathbf{I}) \mathbf{d}(\mu_k)}{\|(\mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) + \mu_k \mathbf{I}) \mathbf{d}(\mu_k)\|} \\ &= \frac{\mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) \frac{\mathbf{d}(\mu_k)}{\mu_k \|\mathbf{d}(\mu_k)\|} + \frac{\mathbf{d}(\mu_k)}{\|\mathbf{d}(\mu_k)\|}}{\left\| \mathbf{J}(\mathbf{x})^T \mathbf{J}(\mathbf{x}) \frac{\mathbf{d}(\mu_k)}{\mu_k \|\mathbf{d}(\mu_k)\|} + \frac{\mathbf{d}(\mu_k)}{\|\mathbf{d}(\mu_k)\|} \right\|}. \end{aligned}$$

令  $k \rightarrow \infty$  得

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

结合(7.2.12)得,

$$\lim_{k \rightarrow \infty} \left\langle \frac{-\mathbf{g}(\mathbf{x})}{\|\mathbf{g}(\mathbf{x})\|}, \frac{\mathbf{d}(\mu_k)}{\|\mathbf{d}(\mu_k)\|} \right\rangle = 1.$$

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

证毕

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

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

在 $\mu$ 取 $\mu_1, \mu_2$ 时的解, 则 $\|\mathbf{A}\mathbf{x}_2 + \mathbf{b}\|^2 < \|\mathbf{A}\mathbf{x}_1 + \mathbf{b}\|^2$ .

解: 考虑如下凸优化问题

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

其最优解满足

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

则对 $\mu = \mu_1, \mu_2$ 的最优解 $\mathbf{x}_1, \mathbf{x}_2$ , 由最优解的定义得

$$\|\mathbf{A}\mathbf{x}_1 + \mathbf{b}\|^2 + \mu_1\|\mathbf{x}_1\|^2 \leq \|\mathbf{A}\mathbf{x}_2 + \mathbf{b}\|^2 + \mu_1\|\mathbf{x}_2\|^2,$$

$$\|\mathbf{A}\mathbf{x}_2 + \mathbf{b}\|^2 + \mu_2\|\mathbf{x}_2\|^2 \leq \|\mathbf{A}\mathbf{x}_1 + \mathbf{b}\|^2 + \mu_2\|\mathbf{x}_1\|^2.$$

前一式乘以 $\mu_2$ , 后一式乘以 $\mu_1$ , 并将两式相加得

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

由 $\mu_1 > \mu_2$ 得,

$$\|\mathbf{A}\mathbf{x}_2 + \mathbf{b}\|^2 \leq \|\mathbf{A}\mathbf{x}_1 + \mathbf{b}\|^2.$$

结论得证.

证毕

## 第8章 临近点算法

### 习题

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

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

若

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

则

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

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

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

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

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

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

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

**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  $\mathbf{x}^*$  为下述优化问题的最优解

$$\min\{f(\mathbf{x}) \mid \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\},$$

其中,  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  连续可微, 向量  $\mathbf{l}, \mathbf{u} \in \mathbb{R}^n$  满足  $\mathbf{l} \leq \mathbf{u}$ . 试证明

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

解: 设  $\mathbf{x}^*$  为凸约束优化问题

$$\min\{f(\mathbf{x}) \mid \mathbf{l} \leq \mathbf{x} \leq \mathbf{u}\},$$

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

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

由此, 对任意的  $i = 1, 2, \dots, n$ ,

$$\nabla_i f(\mathbf{x}^*)(\mathbf{x}_i - \mathbf{x}_i^*) \geq 0 \quad \forall \mathbf{l}_i \leq \mathbf{x}_i \leq \mathbf{u}_i.$$

显然, 若  $\mathbf{x}_i^* = \mathbf{l}_i$ , 则对任意的  $\mathbf{l}_i \leq \mathbf{x}_i \leq \mathbf{u}_i$ , 均有  $\mathbf{x}_i - \mathbf{x}_i^* \geq 0$ . 从而  $\frac{\partial f_i(\mathbf{x}^*)}{\partial \mathbf{x}_i} \geq 0$ .

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

若  $\mathbf{l}_i \leq \mathbf{x}_i^* \leq \mathbf{u}_i$ , 分别取  $\mathbf{x}_i = \mathbf{l}_i$  和  $\mathbf{x}_i = \mathbf{u}_i$  得,  $\frac{\partial f_i(\mathbf{x}^*)}{\partial \mathbf{x}_i} = 0$ .

证毕

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

$$\Omega_1 = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0}\};$$

$$\Omega_2 = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{A}\mathbf{x} \geq \mathbf{b}, \mathbf{x} \geq \mathbf{0}\}.$$

解: 根据定义,  $\Omega_1$  的可行方向  $\mathbf{d}$  满足

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

其中  $\alpha > 0$  充分小.

由于  $\mathbf{x} \in \Omega$ , 即满足

$$\mathbf{Ax} = \mathbf{b}, \quad \mathbf{x} \geq 0.$$

所以  $\mathbf{x}$  点的可行方向满足

$$\mathbf{Ad} = \mathbf{0}.$$

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

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

$$\begin{cases} \mathbf{Ad} = \mathbf{0} \\ \mathbf{d}_i \geq 0, \text{ 若 } \mathbf{x}_i = 0. \end{cases}$$

对  $\mathbf{x} \in \Omega_2 = \{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{Ax} \geq \mathbf{b}, \mathbf{x} \geq 0\}$ . 记

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

则

$$\mathbf{Ax} \geq \mathbf{b} \iff \mathbf{a}_i^T \mathbf{x} \geq \mathbf{b}_i, \quad i = 1, 2, \dots, m.$$

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

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

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

$$\begin{cases} \mathbf{a}_i^T \mathbf{d} \geq 0, \text{ 若 } \mathbf{a}_i^T \mathbf{x} = \mathbf{b}_i, \\ \mathbf{d}_i \geq 0, \text{ 若 } \mathbf{x}_i = 0. \end{cases}$$

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

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

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

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

解: 由

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

知

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

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

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

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

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

由此得结论.

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

$$\begin{aligned} \min_{\mu \in \mathbb{R}, \mathbf{x} \in \mathbb{R}^n} \quad & \|\mathbf{A} - \mu \mathbf{x} \mathbf{x}^T\|_F^2 = \sum_{i,j=1}^n (a_{ij} - \mu x_i x_j)^2 \\ \text{s.t.} \quad & \mathbf{x}^T \mathbf{x} = 1 \end{aligned}$$

的最优解为  $\mathbf{A}$  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

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

故原优化问题的Lagrange 函数为

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

由此, 原优化问题的一阶最优性条件(KKT条件)为

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

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

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



进一步有,

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

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

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

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

另一方面, 由于

$$\|\mathbf{A}\|_F^2 - \mu^2 \leq \|\mathbf{A}\|_F^2 = \|\mathbf{A} - 0 \mathbf{x} \mathbf{x}^T\|_F^2,$$

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

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

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

解: 该问题的 Lagrange 函数为

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

得最优性条件

$$\begin{cases} \nabla_{\mathbf{x}_i} L(\mathbf{x}, \lambda, \mu) = -\frac{\mathbf{c}_i}{\mathbf{x}_i^2} - \lambda \mathbf{a}_i - \mu_i \mathbf{x}_i = 0, & i = 1, 2, \dots, n, \\ \sum_{i=1}^n \mathbf{a}_i \mathbf{x}_i = b; \quad \mu^T \mathbf{x} = 0, \quad \mathbf{x} \geq 0, \quad \mu \geq 0. \end{cases}$$

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

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

所以,

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

进而有

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

另一方面, 由KKT条件的第一式

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

所以,

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

由此,

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

解之得

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

结论得证.

证毕

6. 写出下述优化问题

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

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

$$\begin{cases} 2\left(\mathbf{x}_1 - \frac{2}{3}\right) + 2\lambda_1 \mathbf{x}_1 + \lambda_2 - \mu_1 = 0 \\ 2(\mathbf{x}_2 - 2) - \lambda_1 + \lambda_2 - \mu_2 = 0 \\ -\mathbf{x}_1^2 + \mathbf{x}_2 \geq 0, \lambda_1 \geq 0, \lambda_1(-\mathbf{x}_1^2 + \mathbf{x}_2) = 0 \\ \mathbf{x}_1 + \mathbf{x}_2 - 6 \leq 0, \lambda_2 \geq 0, \lambda_2(\mathbf{x}_1 + \mathbf{x}_2 - 6) = 0 \\ \mathbf{x}_1 \geq 0, \mu_1 \geq 0, \mu_1 \mathbf{x}_1 = 0 \\ \mathbf{x}_2 \geq 0, \mu_2 \geq 0, \mu_2 \mathbf{x}_2 = 0 \end{cases}$$

容易验证,  $\mathbf{x} = (2/3; 2)$ ,  $\lambda_1 = \lambda_2 = 0$ ,  $\mu_1 = \mu_2 = 0$  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. 对  $n$  阶实对称阵  $\mathbf{A}, \mathbf{B}$ , 试证明下述结论等价.

- (1) 对任意的非零向量  $\mathbf{x} \in \mathbb{R}^n$ ,  $\max\{\mathbf{x}^T \mathbf{A} \mathbf{x}, \mathbf{x}^T \mathbf{B} \mathbf{x}\} \geq 0 (> 0)$ ;
- (2) 存在非负常数  $\lambda$  和  $\mu$  满足  $\lambda + \mu = 1$  使得  $\lambda \mathbf{A} + \mu \mathbf{B}$  半正定(正定).

解: 考虑优化问题

$$\begin{aligned} \min t \\ \text{s.t. } t - \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0 \\ t - \mathbf{x}^T \mathbf{B} \mathbf{x} \geq 0 \\ \mathbf{x}^T \mathbf{x} \geq 1 \end{aligned}$$

显然, (1) 成立, 即  $\max\{\mathbf{x}^T \mathbf{A} \mathbf{x}, \mathbf{x}^T \mathbf{B} \mathbf{x}\} \geq 0$  的充分必要条件是该优化问题的最优值  $t^* \geq 0$ , 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

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

要使内层优化问题的最小值达到最大, 参数  $\mu$  需满足

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

从而利用内层优化问题的KKT条件可将对偶规划问题写成

$$\begin{aligned} \max_{\mu \geq 0} 0 \\ \text{s.t. } 1 - \mu_1 - \mu_2 = 0 \\ \mu_1 \mathbf{A} + \mu_2 \mathbf{B} - \mu_3 \mathbf{I} \text{ 半正定.} \end{aligned}$$

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  $\mu_1, \mu_2 \geq 0$ ,  $\mu_1 + \mu_2 = 1$  使得  $\mu_1 \mathbf{A}_1 + \mu_2 \mathbf{A}_2$  半正定. 证毕

8. 设  $\mathbf{a} \in \mathbb{R}_+^n$ . 求下述优化问题的最优解

$$\begin{aligned} \max \min_{1 \leq i \leq n} \mathbf{x}_i \\ \text{s.t. } \mathbf{a}^T \mathbf{x} \leq 1 \end{aligned}$$

解: 不妨设向量  $\mathbf{a}$  的前  $m$  个元素非零. 则

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

对后者, 由于  $\mathbf{x} = \mathbf{0}$  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  $\mathbf{x}_1 > \mathbf{x}_2 \geq \mathbf{x}_3 \geq \dots \geq \mathbf{x}_{m-1} > \mathbf{x}_m \geq 0$  为该优化问题的最优解. 显然, 其最优值为  $\mathbf{x}_m$ .

定义向量  $\mathbf{x}'$  如下,

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

则  $\mathbf{x}'$  仍为该优化问题的可行解, 而最优值为  $\mathbf{x}'_m = \mathbf{x}_m + \mathbf{a}_1(\mathbf{x}_1 - \mathbf{x}_2)/\mathbf{a}_m > \mathbf{x}_m$ . 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

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

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

由此, 原优化问题

$$\begin{aligned} \max \min_{1 \leq i \leq n} \mathbf{x}_i \\ \text{s.t. } \mathbf{a}^T \mathbf{x} \leq 1 \end{aligned}$$

的最优解为

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

最优值为  $1/\sum_{i=1}^m \mathbf{a}_i$ .

证毕

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

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

解: 该优化问题的 Lagrange 函数为

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

它关于  $\mathbf{x}$  为凸函数.

下求其关于  $\mathbf{x}$  的最小值解. 对此, 令

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

得

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

从而

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

由此, 优化问题的Lagrange 对偶为

$$\max_{\lambda \in \mathbb{R}^m} -\frac{1}{4} \left\| \mathbf{A}^T \lambda \right\|^2 - \lambda^T \mathbf{A} \mathbf{b}. \quad \text{证毕}$$

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

$$\min \mathbf{c}^T \mathbf{x} \quad \text{s.t.} \quad \mathbf{A} \mathbf{x} = \mathbf{0}, \mathbf{x}^T \mathbf{x} \leq 1$$

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

$$\begin{cases} \mathbf{c} - \mathbf{A}^T \lambda + 2\mu \mathbf{x} = \mathbf{0}, \\ \mathbf{A} \mathbf{x} = \mathbf{0}, \mathbf{x}^T \mathbf{x} \leq 1, \\ \mu(1 - \mathbf{x}^T \mathbf{x}) = 0, \mu \geq 0. \end{cases}$$

第一式左乘  $\mathbf{A}$ , 则由第二式得

$$\lambda = (\mathbf{A} \mathbf{A}^T)^{-1} \mathbf{A} \mathbf{c}.$$

从而

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

再由  $\mu = \mu \mathbf{x}^T \mathbf{x}$  得

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

从而

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

这样,

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

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

11. 建立下述优化问题的Lagrange对偶规划

$$\begin{aligned} \min \quad & \mathbf{x}_1^2 + \mathbf{x}_2^2 \\ \text{s.t.} \quad & \mathbf{x}_1 + \mathbf{x}_2 \geq 4 \\ & \mathbf{x}_1, \mathbf{x}_2 \geq 0 \end{aligned}$$

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

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

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

基于该优化问题的Lagrange函数

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

可得其Lagrange对偶函数

$$\theta(\mu) = \inf \{L(\mathbf{x}, \mu) \mid \mathbf{x}_1, \mathbf{x}_2 \geq 0\}, \quad \mu \geq 0.$$

由于

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

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

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

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

12. 设  $f: \mathbb{R}^n \rightarrow \mathbb{R}$  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  $\min_{\mathbf{x} \in \Theta} \{f(\mathbf{x}) \mid \mathbf{A}\mathbf{x} = \mathbf{b}\}$  的鞍点满足如下变分不等式问题, 即求  $\mathbf{w}^* \in \mathcal{W}$  使满足

$$\mathbf{F}(\mathbf{w}^*)^\top (\mathbf{w} - \mathbf{w}^*) \geq 0, \quad \forall \mathbf{w} \in \mathcal{W},$$

其中,

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

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

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

其中

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

由鞍点条件的后一式,

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

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

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

由鞍点条件的前一式,

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

再由稳定点条件得

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

两式结合得命题结论.

证毕

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

$$\min \left\{ \sum_{i=1}^n \frac{\mathbf{a}_i^2}{\mathbf{x}_i} \mid \mathbf{e}^\top \mathbf{x} \leq 1, \mathbf{x} > \mathbf{0} \right\}$$

解: 显然, 该优化问题等价于

$$\min \left\{ \sum_{i=1}^n \frac{\mathbf{a}_i^2}{\mathbf{x}_i} \mid \mathbf{e}^\top \mathbf{x} = 1, \mathbf{x} > \mathbf{0} \right\}, \quad \mathbf{a} \geq \mathbf{0}.$$

利用KKT条件

$$\begin{cases} \frac{a_i^2}{x_i^2} = \mu, & i = 1, 2, \dots, n \\ \sum_{i=1}^n x_i = 1 \end{cases}$$

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

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

$$\begin{aligned} \min \quad & \mathbf{a}^T \mathbf{x} \\ \text{s.t.} \quad & (\mathbf{x} - \mathbf{x}_0)^T \mathbf{G} (\mathbf{x} - \mathbf{x}_0) \leq \rho \end{aligned}$$

解: 该优化问题的Lagrange函数为

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

其对偶规划为

$$\max_{\lambda \geq 0} \min_{\mathbf{x}} L(\mathbf{x}, \lambda).$$

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

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

由此得

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

代入Lagrange函数得对偶规划

$$\max_{\lambda \geq 0} \mathbf{a}^T \mathbf{x}_0 - \frac{1}{4\lambda} \mathbf{a}^T \mathbf{G}^{-1} \mathbf{a} - \lambda \rho$$

即

$$\min_{\lambda \geq 0} \frac{1}{4\lambda} \mathbf{a}^T \mathbf{G}^{-1} \mathbf{a} + \lambda \rho - \mathbf{a}^T \mathbf{x}_0$$

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

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

从而, 对偶规划问题的最优解为

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



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

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

进一步, 该优化问题的最优值为

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

### 15. 对无约束优化问题

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

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

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

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

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

其中,

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

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

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

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

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

而对后一优化问题, 利用目标函数的可分离性得

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

这样, 对偶问题可写成

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

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

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

解: 由该优化问题的KKT条件

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

解之得

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

由于该规划为凸规划, 所以  $\mathbf{x}^*$  为最优解.

证毕

2. 设  $a > b > 0$ ,

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

证明如下二次规划问题

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

最优解的所有分量都相等.

解: 容易计算, 矩阵  $\mathbf{G}$  的行列式

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

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

用反证法证明. 设  $\mathbf{x} \in \mathbb{R}^n$  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

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

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

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

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

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

这与  $\mathbf{x}$  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  $\sigma_1 > \sigma_2 > \dots > \sigma_r > 0$ ,  $s < r$ . 求下述二次规划问题的最优解

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

解: 定义矩阵

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

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  $\sum_{i=1}^r \sum_{j=1}^s \sigma_i x_{ij}^2$  取到最大值,  $x_{1j}, x_{2j}, \dots, x_{rj}$  应依次取尽可能靠近 1 的值. 考虑到约束条件  $\sum_{j=1}^s x_{ij} \leq 1, \quad i = 1, 2, \dots, r,$  和  $\sum_{i=1}^r x_{ij} \leq 1, \quad j = 1, 2, \dots, s,$  该优化问题的最优解为

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

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

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

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

**解:** 根据题设,  $Q(\mathbf{x}_k + \alpha \mathbf{d}_k)$  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,  $Q(\mathbf{x}_k + \alpha \mathbf{d}_k)$  关于  $\alpha \leq 1$  严格单调递减, 关于  $\alpha \geq 1$  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  $\bar{\mathbf{x}}$  为下述连续可微约束优化问题

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

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

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

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

解: 线性规划子问题

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

的最优值为零当且仅当系统

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

无解. 由Gordan定理

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

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

(2)  $\mathbf{A} \mathbf{y} = \mathbf{0}, \mathbf{y} \not\geq \mathbf{0}$ ,

知存在非负但不全为零的乘子  $\boldsymbol{\lambda}$  使得

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

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

$$\begin{cases} \lambda_0 \nabla f(\bar{\mathbf{x}}) = \sum_{i \in \mathcal{I}} \lambda_i \nabla c_i(\bar{\mathbf{x}}) \\ \lambda_i \geq 0, c_i(\bar{\mathbf{x}}) \geq 0, \lambda_i c_i(\bar{\mathbf{x}}) = 0, i \in \mathcal{I}. \end{cases}$$

证毕

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

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

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

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

其KT条件为

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

整理得

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

由此得命题结论.

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

解: 根据投影的定义,

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

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

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

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

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

由§11.3的讨论, 其对偶规划问题为

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

显然, 其最优解为

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

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  $\mathbf{u} \in \mathbb{R}^n$  到半空间  $\{\mathbf{x} \in \mathbb{R}^n \mid \mathbf{a}^T \mathbf{x} \geq b\}$  上的投影为

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

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

$$\min\{1, \alpha\} \|\mathbf{r}(1)\| \leq \|\mathbf{r}(\alpha)\| \leq \max\{1, \alpha\} \|\mathbf{r}(1)\|.$$

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

$$\|\mathbf{r}(\alpha)\| = \begin{cases} \geq \|\mathbf{r}(1)\|, & \alpha > 1 \\ \leq \|\mathbf{r}(1)\|, & 0 < \alpha < 1. \end{cases}$$

由性质12.5.3, 函数  $\frac{\|\mathbf{r}(\alpha)\|}{\alpha}$  关于  $\alpha \geq 0$  单调不增. 故

$$\frac{\|\mathbf{r}(\alpha)\|}{\alpha} = \begin{cases} \leq \|\mathbf{r}(1)\|, & \alpha > 1 \\ \geq \|\mathbf{r}(1)\|, & 0 < \alpha < 1. \end{cases} \implies \|\mathbf{r}(\alpha)\| = \begin{cases} \leq \alpha \|\mathbf{r}(1)\|, & \alpha > 1 \\ \geq \alpha \|\mathbf{r}(1)\|, & 0 < \alpha < 1. \end{cases}$$

两者结合得

$$\min\{1, \alpha\} \|\mathbf{r}(1)\| \leq \|\mathbf{r}(\alpha)\| \leq \max\{1, \alpha\} \|\mathbf{r}(1)\|.$$

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

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

$$\begin{aligned} \min \quad & \frac{1}{2} \|\mathbf{A} - \alpha \mathbf{I}\|_{\mathbb{F}}^2 \\ \text{s.t.} \quad & \alpha \geq 0 \end{aligned}$$



显然,

$$\begin{aligned}\frac{1}{2}\|\mathbf{A} - \alpha\mathbf{I}\|_{\text{F}}^2 &= \frac{1}{2}\|\mathbf{A}\|_{\text{F}}^2 - \alpha\text{tr}(\mathbf{A}) + \frac{n}{2}\alpha^2 \\ &= \frac{1}{2}\|\mathbf{A}\|_{\text{F}}^2 + \frac{n}{2}\left(-\frac{2}{n}\alpha\text{tr}(\mathbf{A}) + \alpha^2\right) \\ &= \frac{1}{2}\|\mathbf{A}\|_{\text{F}}^2 + \frac{n}{2}\left(\alpha - \frac{1}{n}\text{tr}(\mathbf{A})\right)^2 - \frac{1}{2n}\text{tr}^2(\mathbf{A}).\end{aligned}$$

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

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

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

的最优解为 $\frac{1}{m} \sum_{i=1}^m \mathbf{a}_i$ 到 $\Omega$ 上的投影.

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

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

其稳定点条件为

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

严格凸二次规划

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

的稳定点条件为

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

由于严格凸二次规划问题的解与稳定点等价, 故结论成立.