《非线性最优化理论与方法》

(王宜举, 修乃华, 科学出版社, 第三版, 2019)

习题解答

目　录

第1章 引论

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.

证明 (1)设 $f(x) = -\ln x, (x > 0)$, 容易验证 $f(x)$是$(0, \infty)$上的凸函数. 则对任意的 $x, y \in (0, \infty)$和任意的$\alpha \in (0,1)$ 有

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

$$\Leftrightarrow \ -\ln(\alpha x + (1-\alpha) y) \leq -\alpha \ln x - (1-\alpha) \ln y$$

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

(2)设 $f(x) = x + \frac{1}{x}, (x > 0)$, 容易验证$f(x)$是$(0, \infty)$上的凸函数. 则对任意的$x_1, x_2, \cdots, x_n \in \mathrm{R}^{++}$,

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

$$\Leftrightarrow \frac{\sum\limits_{j=1}^{n} x_j}{n} + \frac{n}{\sum\limits_{j=1}^{n} x_j} \leq \frac{\sum\limits_{j=1}^{n} \left(x_j + \frac{1}{x_j}\right)}{n}$$

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

而等号成立的条件: $x_1 = x_2 = \cdots = x_n$. 证毕

2. 试用Cauchy-Schwarz不等式证明

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

证明 设 $x = (x_1, x_2, \cdots, x_n)^{\top}$, 记 $\mathrm{e} = (1; 1; \cdots; 1)$ 和 $|x| = (|x_1|; |x_2|; \cdots; |x_n|)$.
则由Cauchy-Schwarz不等式,

$$\|x\|_1 = \mathrm{e}^{\top} |x| \leq \|x\| \|\mathrm{e}\|_2 = \sqrt{n} \|x\|.$$

证毕

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

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

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

$$(x^{\top} y)^2 \leqslant (x^{\top} Q x)(y^{\top} Q^{-1} y).$$

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

$$\max\{b^\top B^{-1}y \mid y^\top y \le 1\}$$

也就是

$$\max\{(B^{-\top}b)^\top y \mid y^\top y \le 1\}$$

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

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

另外,

$$\begin{aligned}(x^\top y)^2 &= y^\top x x^\top y = \left\langle xx^\top, yy^\top\right\rangle,\\ &= \left\langle xx^\top, QQ^{-1}yy^\top\right\rangle = \left\langle xx^\top Q, Q^{-1}yy^\top\right\rangle,\\ &\le \|xx^\top Q\|_{\mathrm{F}}\|Q^{-1}yy^\top\|_{\mathrm{F}} = \mathrm{tr}\left(xx^\top Q\right)\mathrm{tr}\left(Q^{-1}yy^\top\right),\\ &= \left(x^\top Qx\right)\left(y^\top Q^{-1}y\right).\end{aligned}$$

第二部分得证. 证毕

4. 利用$\min\limits_{x,y} f(x,y)=\min\limits_{x}\min\limits_{y} f(x,y)$求如下优化问题的解析解

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

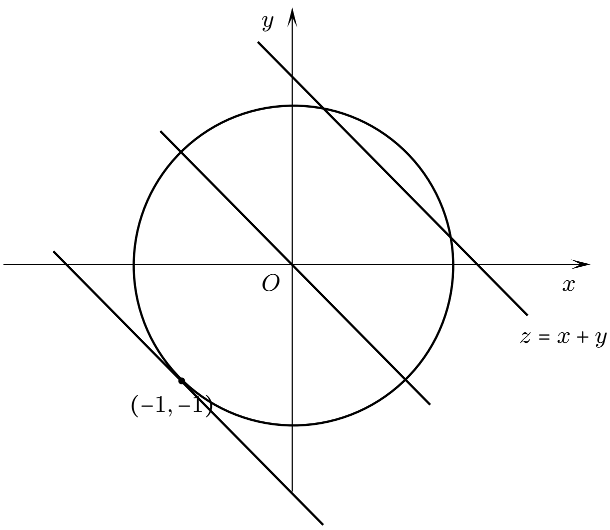
证明 对任意的$x>0$, 计算关于y的优化问题. 此时, 利用目标函数关于y为凸函数得内层优化问题关于y的最优值点$y=\frac{2}{5}x$.这样上述优化问题化为

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

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

5. 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: 图形见左图. 易知最优解为 $x_1=x_2=-1$.

6. 证明函数$f:\mathrm{R}^n\to\mathrm{R}$为凸函数的充分必要条件是函数$f$的上图为凸集, 其中上图定义为

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

证明 必要性：对 $\forall\ x_1, x_2 \in \mathrm{R}^n$，$\lambda \in [0,1]$, 显然有

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

由上图是凸集, 则

$$\begin{aligned}&(\lambda x_1 + (1-\lambda)x_2, \lambda f(x_1) + (1-\lambda)f(x_2))\\ =\ &\lambda(x_1, f(x_1)) + (1-\lambda)(x_2, f(x_2)) \in \mathrm{epi}f\\ \Rightarrow\ &f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2)\end{aligned}$$

充分性：对 $\forall\ (x_1, y_1), (x_2, y_2) \in \mathrm{epi}f$，$\lambda \in [0,1]$, 有

$$f(x_1) \le y_1, f(x_2) \le y_2,$$

由 $f(x)$ 是凸函数, 则

$$f(\lambda x_1 + (1-\lambda)x_2) \le \lambda f(x_1) + (1-\lambda)f(x_2) \le \lambda y_1 + (1-\lambda)y_2$$

$$\Rightarrow \lambda(x_1, y_1) + (1-\lambda)(x_2, y_2) = (\lambda x_1 + (1-\lambda)x_2, \lambda y_1 + (1-\lambda)y_2) \in \mathrm{epi}f.$$ 证毕

7. 讨论由$a_1, a_2, \cdots, a_m \in \mathrm{R}^n$所生成的多面胞、仿射集、子空间及这些点和原点所生成的仿射包之间的关系.

解:

多面胞 $$\begin{aligned}A_1 &:= \mathrm{co}\ \{a_1, a_2, \cdots, a_m\}\\ &= \{\sum_{j=1}^m \lambda_j a_j \mid \sum_{j=1}^m \lambda_j = 1, \lambda_j \in [0,1], j = 1, 2, \cdots, m\}.\end{aligned}$$

仿射集 $$\begin{aligned}A_2 &:= \mathrm{aff}\ \{a_1, a_2, \cdots, a_m\}\\ &= \{\sum_{j=1}^m \lambda_j a_j \mid \sum_{j=1}^m \lambda_j = 1, \lambda_j \in \mathrm{R}, j = 1, 2, \cdots, m\}.\end{aligned}$$

子空间 $$\begin{aligned}A_3 &:= \mathrm{span}\ \{a_1, a_2, \cdots, a_m\}\\ &= \{\sum_{j=1}^m \lambda_j a_j \mid \lambda_j \in \mathrm{R}, j = 1, 2, \cdots, m\}.\end{aligned}$$

仿射胞 $$\begin{aligned}A_4 &:= \mathrm{co\ aff}\ \{0, a_1, a_2, \cdots, a_m\}\\ &= \mathrm{co}\ \{\lambda_0 0 + \sum_{j=1}^m \lambda_j a_j \mid \sum_{j=0}^m \lambda_j = 1, \lambda_j \in \mathrm{R}, j = 0, 1, 2, \cdots, m\}\\ &= \mathrm{co}\ \{\sum_{j=1}^m \lambda_j a_j \mid \lambda_j \in \mathrm{R}, j = 1, 2, \cdots, m\}\\ &= \{\sum_{j=1}^m \lambda_j a_j \mid \lambda_j \in \mathrm{R}, j = 1, 2, \cdots, m\}.\end{aligned}$$

由以上定义可以看出A_1, A_2, A_3, A_4之间的关系: $A_1 \subseteq A_2 \subseteq A_3 = A_4$. 证毕

8. 设$d \in \mathrm{R}^n$是连续可微函数$f : \mathrm{R}^n \to \mathrm{R}$在点$x \in \mathrm{R}^n$的下降方向. 试建立$\alpha$为单元函数$f(x+\alpha d)$在$\alpha \geqslant 0$上的最小值点的一个必要条件, 并讨论在什么条件下该条件是充分的.

解: 记 $g(\alpha) := f(x+\alpha d)$, 则 $g(\alpha)$ 在 $\alpha \geq 0$ 上的最小值点的一个必要条件为

$$g'(\alpha) = \nabla_x f(x+\alpha d)^\top d = 0.$$

若 $g(\alpha) = f(x+\alpha d)$ 是关于α的凸函数, 则 $\nabla_x f(x+\alpha d)^\top d = 0$ 是充分条件. 证毕

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

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

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: 一阶必要条件为

$$\nabla_x \|Ax-b\|^2 = A^\top(Ax-b) = 0.$$

由于

$$\nabla^2 \|Ax-b\|^2 = A^\top A \geq 0,$$

知 $\|Ax-b\|^2$关于$x \in \mathrm{R}^n$为凸函数, 所以 $A^\top(Ax-b) = 0$ 也是充分条件. 在$A^\top A$正定时,最优解唯一. 证毕

10. 试在3维欧式空间中确定一通过点$(3;4;5)$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

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

根据四面体的提及公式, 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$abc \neq 0$. 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

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

满足上述条件的一法向量为$(1/3; 1/4; 1/5)$. 代入体积公式得体积=270. 证毕

11. 设$A \in \mathrm{R}^{n\times n}$, $b \in \mathrm{R}^n$. 试给出下述优化问题的最优性条件

$$\min \frac{1}{2} x^\top A x + b^\top x$$

解: 一阶必要条件：$\nabla_x \left(\frac{1}{2} x^\top A x + b^\top x\right) = \frac{A+A^\top}{2} x + b = 0$,

二阶必要条件：$\nabla^2 \left(\frac{1}{2} x^\top A x + b^\top x\right) = \frac{A+A^\top}{2} \succeq 0$,

二阶充分条件：$\nabla^2 \left(\frac{1}{2} x^\top A x + b^\top x\right) = \frac{A+A^\top}{2} \succ 0$. 证毕

12. 设$a \in \mathrm{R}^n, \lambda > 0$. 求如下问题的最优解

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

解: 该问题等价于

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

这样, 要求解原问题, 只需求解如下子问题

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

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

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

利用定理1.6.4, 子问题(1.6.2)的最优解为

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

(2) 若$a_i < 0$, 则子问题(1.6.2)的最优解非正. 从而子问题(1.6.2)可化为

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

利用无约束优化问题的最优性条件, 该子问题的最优解为

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

引入符号函数

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

则子问题(1.6.2)的最优解为

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

由此得原问题的最优解.

13. 设$a \in \mathrm{R}, \lambda > 0$. 求下述优化问题的最优解

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

并在此基础上, 给出如下优化问题的最优解

$$\min_{x \in \mathrm{R}^n} \lambda |x|_1 - a^\top x$$

其中, $a \in \mathrm{R}^n, \lambda > 0$.

解: 对优化问题

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

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

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

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

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

综上所述, 优化问题

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

第2章 线搜索方法与信赖域方法

1. 对n元严格凸二次函数$f(x)=\dfrac{1}{2}x^\top Ax+b^\top x$, 试求在$x_k$点沿下降方向$d_k$的最优步长. 若取$d_k=-g_k$, 试计算目标函数在每一迭代步的下降量.

解: 设 $\alpha_k=\arg\min\limits_{\alpha\geq 0} f(x_k+\alpha d_k)$.则由

$$d_k^\top\nabla f(x_k+\alpha d_k)=d_k^\top[A(x_k+\alpha d_k)+b]=0$$

知

$$\alpha_k=\frac{-g_k^\top d_k}{d_k^\top Ad_k}, g_k=Ax_k+b.$$

若 $d_k=-g_k$, 则 $\alpha_k=\frac{g_k^\top g_k}{g_k^\top Ag_k}$, 且

$$\begin{aligned}
f(x_k+\alpha_k d_k) &= f(x_k-\alpha_k g_k)\\
&= \frac{1}{2}(x_k-\alpha_k g_k)^\top A(x_k-\alpha_k g_k)+b^\top(x_k-\alpha_k g_k)\\
&= \frac{1}{2}x_k^\top Ax_k+b^\top x_k-\alpha_k g_k^\top Ax_k+\frac{1}{2}\alpha_k^2 g_k^\top Ag_k-\alpha_k b^\top g_k\\
&= f(x_k)-\alpha_k g_k^\top Ax_k+\frac{1}{2}\alpha_k^2 g_k^\top Ag_k-\alpha_k b^\top g_k
\end{aligned}$$

那么,目标函数在每一次迭代步的下降量为

$$\begin{aligned}
f(x_k+\alpha_k d_k)-f(x_k) &= -\alpha_k g_k^\top Ax_k+\frac{1}{2}\alpha_k^2 g_k^\top Ag_k-\alpha_k b^\top g_k\\
&= -\frac{g_k^\top g_k}{g_k^\top Ag_k}g_k^\top Ax_k+\frac{1}{2}\left(\frac{g_k^\top g_k}{g_k^\top Ag_k}\right)^2 g_k^\top Ag_k-\frac{g_k^\top g_k}{g_k^\top Ag_k}b^\top g_k\\
&= -\frac{g_k^\top g_k}{g_k^\top Ag_k}\left[g_k^\top Ax_k-\frac{1}{2}g_k^\top g_k+b^\top g_k\right]\\
&= -\frac{g_k^\top g_k}{g_k^\top Ag_k}\frac{1}{2}g_k^\top g_k=-\frac{1}{2}\frac{\|g_k\|^4}{g_k^\top Ag_k}
\end{aligned}$$

证毕

2. 对上题中的凸二次函数, 试求在x_k点沿下降方向d_k的Goldstein步长α_k, 并证明:

$$2\sigma\bar{\alpha}_k\leqslant\alpha_k\leqslant 2(1-\sigma)\bar{\alpha}_k,$$

其中$\bar{\alpha}_k$为最优步长.

证明

$$\begin{aligned}
f(x_k+\alpha_k d_k) &= \frac{1}{2}(x_k+\alpha_k d_k)^\top A(x_k+\alpha_k d_k)+b^\top(x_k+\alpha_k d_k)\\
&= f(x_k)+\frac{1}{2}\alpha_k^2 d_k^\top Ad_k+\alpha_k g_k^\top d_k
\end{aligned}$$

由 Goldstein 步长知

$$(1-\sigma)\alpha_k g_k^\top d_k \leq f(x_k+\alpha_k d_k)-f(x_k) \leq \sigma\alpha_k g_k^\top d_k.$$

从而,

$$(1-\sigma)\alpha_k g_k^\top d_k \leq \frac{1}{2}\alpha_k^2 d_k^\top A d_k + \alpha_k g_k^\top d_k \leq \sigma\alpha_k g_k^\top d_k.$$

进一步有,

$$2\sigma\frac{-g_k^\top d_k}{d_k^\top A d_k} \leq \alpha_k \leq 2(1-\sigma)\frac{-g_k^\top d_k}{d_k^\top A d_k}.$$

从而,

$$2\sigma\bar{\alpha}_k \leq \alpha_k \leq 2(1-\sigma)\bar{\alpha}_k.$$

证毕

3. 考虑Armijo步长规则下的下降算法

$$x_{k+1} = x_k + \alpha_k d_k,$$

其中搜索方向由下式确定:

$$(d_k)_i = \begin{cases} -\dfrac{\partial f(x_k)}{\partial x_i}, & \text{如果}\left|\dfrac{\partial f(x_k)}{\partial x_i}\right|\text{取最大值}, \\ 0, & \text{其他}. \end{cases}$$

试证明算法产生的迭代点列的任一聚点都为原问题的稳定点.

证明 考虑

$$\langle d_k, -g_k\rangle = \sum_{i\in I}\left(\frac{\partial f(x_k)}{\partial x_i}\right)^2 > 0, \quad I = \{i \mid (d_k)_i \neq 0\}.$$

所以, d_k 与 $-g_k$ 的夹角 $\theta \in [0, \frac{\pi}{2}-\mu]$, $\mu \in (0, \frac{\pi}{2})$.
另一方面

$$\left|\frac{\partial f(x_k)}{\partial x_i}\right| \leq \|d_k\| = \sqrt{\sum_{i\in I}\left(\frac{\partial f(x_k)}{\partial x_i}\right)^2} \leq \sqrt{\sum_{i=1}^{n}\left(\frac{\partial f(x_k)}{\partial x_i}\right)^2} = \|g_k\|.$$

取 $\Gamma := \max_k\{\|g_k\|\}$, $\nu := \min_{k,i}\left\{\left|\frac{\partial f(x_k)}{\partial x_i}\right|\right\}$. 从而有 $\nu \leq \|d_k\| \leq \Gamma$. 再由定理2.2.3得证. 证毕

4. 设n阶实矩阵A对称正定. 试证明对于任意的$x \in \mathrm{R}^n$,

$$\frac{4\kappa(A)}{(1+\kappa(A))^2}(x^\top A x)(x^\top A^{-1}x) \leq \|x\|^4 \leqslant (x^\top A x)(x^\top A^{-1}x).$$

证明 首先证明右边不等式

$$\begin{aligned}\|x\|^4 &= x^\top x\cdot x^\top x = \left\langle x, xx^\top x\right\rangle = \left\langle xx^\top, xx^\top\right\rangle \\ &= \left\langle xx^\top, AA^{-1}xx^\top\right\rangle = \left\langle xx^\top A, A^{-1}xx^\top\right\rangle \\ &\le \|xx^\top A\|_{\mathrm F}\|A^{-1}xx^\top\|_{\mathrm F} \le \operatorname{tr}\left(xx^\top A\right)\operatorname{tr}\left(A^{-1}xx^\top\right) \\ &= \operatorname{tr}\left(x^\top Ax\right)\operatorname{tr}\left(x^\top A^{-1}x\right) = \left(x^\top Ax\right)\left(x^\top A^{-1}x\right).\end{aligned}$$

证毕

5. 对二维子空间子问题(2.3.4), 设B_k对称正定. 试利用上题中的不等式求问题(2.3.4)的最优解.

6. 设$f:\mathrm{R}^n\to\mathrm{R}$的梯度函数$\nabla f(x)$Lipschitz连续, 常数为$L$. 对常数$\mu>L$和$x\in\mathrm{R}^n$, 令

$$x^+ = \arg\min_y\{f(x)+\langle\nabla f(x), y-x\rangle+\frac{\mu}{2}\|y-x\|^2\}.$$

则

$$f(x^+)-f(x)\leqslant\frac{L-\mu}{2}\|x^+-x\|^2.$$

再设函数$g:\mathrm{R}^n\to\mathrm{R}$为非光滑函数. 对常数$\mu>L$和$x\in\mathrm{R}^n$, 令

$$x^+ = \arg\min\{f(x)+\langle\nabla f(x), y-x\rangle+\frac{\mu}{2}\|y-x\|^2+g(y)\}.$$

则

$$f(x^+)+g(x^+)\leqslant f(x)+g(x)+\frac{L-\mu}{2}\|x^+-x\|^2.$$

证明 首先, 对任意的$x,y\in\mathrm{R}^n$, 利用微分中值定理, 存在$\xi\in(x,y)$使得

$$f(y)-f(x)=\langle\nabla f(\xi), y-x\rangle.$$

再利用梯度函数的Lipshcitz连续性, 由该式得

$$\begin{aligned}f(y)-f(x) &= \langle\nabla f(x), y-x\rangle+\langle\nabla f(\xi)-\nabla f(x), y-x\rangle \\ &\leqslant \langle\nabla f(x), y-x\rangle+\|\nabla f(\xi)-\nabla f(x)\|\|y-x\| \\ &\leqslant \langle\nabla f(x), y-x\rangle+\frac{L}{2}\|y-x\|^2.\end{aligned}$$

其次, 对常数$\mu>L$和$x\in\mathrm{R}^n$, 由

$$x^+ = \arg\min_y\{f(x)+\langle\nabla f(x), y-x\rangle+\frac{\mu}{2}\|y-x\|^2\}$$

得

$$\langle\nabla f(x), x^+-x\rangle+\frac{\mu}{2}\|x^+-x\|^2\leqslant 0.$$

则由前式得

$$\begin{aligned} f(x^+) - f(x) &\leqslant \langle \nabla f(x), x^+ - x \rangle + \frac{L}{2}\|x^+ - x\|^2 \\ &= \langle \nabla f(x), x^+ - x \rangle + \frac{\mu}{2}\|x^+ - x\|^2 + \frac{L-\mu}{2}\|x^+ - x\|^2 \\ &= \frac{L-\mu}{2}\|x^+ - x\|^2. \end{aligned}$$

剩下的结论可类似证明. 证毕

第3章 最速下降法与牛顿方法

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.

证明 设严格凸二次函数为 $f(x)=\frac{1}{2}x^\top Ax+b^\top x+c$, 其中 $A\succ 0$, 则有

$$g_k=\nabla f(x_k)=Ax+b,\ d_k=-A^{-1}g_k.$$

设初始点为 x_0, 一步迭代有：

$$d_1=-A^{-1}g_0=-A^{-1}\left(Ax_0+b\right)=-x_0-A^{-1}b\ ;$$

$$x_1=x_0+d_1=-A^{-1}b.$$

从而, $g_1=-AA^{-1}b+b=0$, 算法终止. 证毕

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

$$B_kd_k=-g_k$$

的下降算法. 分析B_k的不同取法对算法收敛速度的影响.

证明 考虑搜索方向的下降性

$$\begin{aligned}f(x_k+\alpha d_k)&=f(x_k)+\alpha\nabla f(x_k)^\top d_k+o(\|\alpha d_k\|)\\&=f(x_k)+\alpha g_k^\top B_k^{-1}g_k+o(\|\alpha d_k\|)\le f(x_k).\end{aligned}$$

当 $B_k=I$, 该搜索方向产生最速下降算法, 收敛速度至多为线性的;

当 $B_k=G_k$, 且 f 为二次连续可微以及 G_k 在极小点处 Lipschitz 连续, 则收敛速度为二阶收敛. 证毕

3. 设用牛顿算法求函数的极小值点时产生迭代点列$\{x_k\}$. 现对变量x做一变换$x=Qy$,其中Q为n阶非奇异矩阵,并用牛顿算法求变换后的函数的极小值点,设产生的迭代点列为$\{y_k\}$. 则$y_k=Q^{-1}x_k$.

证明 对于变换后的点列 $\{y_k\}$, 我们有

$$\begin{aligned}s_k(y_k)&=-G_k(y_k)^{-1}\nabla f(y_k)\\&=-G_k(y_k)^{-1}Q^\top\nabla f(x_k)\\&=-\left(Q^\top\nabla^2 f(x_k)Q\right)^{-1}Q^\top\nabla f(x_k)\\&=-Q^{-1}G_k(x_k)^{-1}\nabla f(x_k)\\&=Q^{-1}s_k(x_k).\end{aligned}$$

利用归纳法, 若 $y_k=Q^{-1}x_k$, 则

$$\begin{aligned}y_{k+1}&=y_k-s_k(y_k)\\&=Q^{-1}x_k-Q^{-1}s_k(x_k)\\&=Q^{-1}x_{k+1}.\end{aligned}$$

从而, 若迭代点列 $\{x_k\}$ 产生的极小值点为 x^*, 则变换后的极小值点为

$$\lim_{k\to\infty} y_k = \lim_{k\to\infty} Q^{-1}x_k = Q^{-1}x^*.$$

证毕

4. 设凸二次函数$f(x)$的最小值点为x^*, 并设$x_0 = x^* + ts$, 其中, s为目标函数的Hesse矩阵某特征值的特征向量. 若以$x_0 \neq x^*$为初始点,则最优步长规则下的最速下降算法一步就可到达最小值点.

证明 设凸二次函数为 $f(x) = \frac{1}{2}x^\top Ax + b^\top x$, 由第2章第1题有, 最速下降算法的最优步长满足

$$\alpha_k = \frac{-g_k^\top d_k}{d_k^\top A g_k}, \quad (g_k = Ax_k + b).$$

对于第一步迭代,

$$\begin{aligned} g_0 &= Ax_0 + b = A\left(x^* + ts\right) + b = tAs = t\lambda s, \\ \alpha_0 &= \frac{(t\lambda)^2 s^\top s}{(t\lambda)^2 s^\top As} = \frac{s^\top s}{s^\top \lambda s} = \frac{1}{\lambda}. \end{aligned}$$

其中, λ 是目标函数的 Hesse 矩阵的特征向量 s 对应的特征值, 即 $As = \lambda s$, 另外,在最优点处满足 $Ax^* + b = 0$. 从而,第一步迭代后有

$$x_1 = x_0 - \alpha_0 g_0 = x_0 - \frac{1}{\lambda}t\lambda s = x_0 - ts = x^*.$$

证毕

第4章 共轭梯度法

1.设 $f:\mathrm{R}^n\to\mathrm{R}$ 连续可微, 向量组 $d_1,d_2,\cdots,d_n\in\mathrm{R}^n$ 线性无关. 对 $\bar{x}\in\mathrm{R}^n$, 若对 $j=1,2,\cdots,n$, 函数$f(\bar{x}+\lambda d_j)$在$\lambda=0$时取到最小值, 试证明 $\nabla f(\bar{x})=0$. 此时 $\bar{x}$ 是函数 $f(x)$ 的局部最小值点吗?

证明 由题设条件

$$
\begin{aligned}
f(\bar{x}) &\le \lim_{\lambda\to 0} f(\bar{x}+\lambda d_j)\\
&= f(\bar{x})+\lim_{\lambda\to 0}\left(\lambda\nabla f(\bar{x})^\top d_j+o(\|\lambda d_j\|)\right),\ j=1,2,\cdots,n.
\end{aligned}
$$

从而, 对$j=1,2,\cdots,n,$

$$\lim_{\lambda\to 0}\lambda\nabla f(\bar{x})^\top d_j\ge 0.$$

进而存在$\delta>0$使得

$$\lambda\in[-\delta,\delta],\ \lambda\nabla f(\bar{x})^\top d_j\ge 0,\ j=1,2,\cdots,n.$$

取 $\lambda=-t\ \max_j\{|\nabla f(\bar{x})^\top d_j|\},t\ge 0.$ 从而, 当 t 充分小时

$$
\begin{aligned}
0\le\lambda\nabla f(\bar{x})^\top d_j &= -t\ \max_j\left\{|\nabla f(\bar{x})^\top d_j|\right\}\nabla f(\bar{x})^\top d_j\\
&\le -t\left\|\nabla f(\bar{x})^\top d_j\right\|^2,\ j=1,2,\cdots,n\\
&\Rightarrow \nabla f(\bar{x})^\top d_j=0,\ j=1,2,\cdots,n.
\end{aligned}
$$

再由向量组 $d_1,d_2,\cdots,d_n\in\mathrm{R}^n$ 线性无关知,

$$\nabla f(\bar{x})=0.$$

对于任意方向 $d\in\mathrm{R}^n$ 可以表示为 $d=\sum_{j=1}^n\alpha_j d_j$, 不失一般性假设 $\sum_{j=1}^n\alpha_j=1$ (否则,可以令 $\beta=\sum_{j=1}^n\alpha_j$, 则 $d=\beta\sum_{j=1}^n\frac{\alpha_j}{\beta}d_j$, 这与方向 d 只差一个倍数, 其中 $\sum_{j=1}^n\frac{\alpha_j}{\beta}=1$), 那么

$$
\begin{aligned}
f(\bar{x}+\lambda d) &= f(\bar{x})+\lambda\sum_{j=1}^n\alpha_j\nabla f(\bar{x})^\top d_j+o(\|\lambda d\|)\\
&= \sum_{j=1}^n\alpha_j\left[f(\bar{x})+\lambda\nabla f(\bar{x})^\top d_j+o(\|\lambda d\|)\right]\\
&= \sum_{j=1}^n\alpha_j f(\bar{x}+\lambda d_j)\\
&\ge \sum_{j=1}^n\alpha_j f(\bar{x})=f(\bar{x})
\end{aligned}
$$

上式说明 $\bar{x}$ 是函数 $f(x)$ 的局部最小值点. 证毕

2.设 $G\in \mathrm{R}^{n\times n}$ 对称正定, $s_1,s_2,\cdots,s_n\in \mathrm{R}^n$ 线性无关. 试证明由下述子式定义的向量组关于矩阵 G 共轭:

$$d_k=\begin{cases} s_k, & \text{如果 } k=1,\\ s_k-\sum\limits_{i=1}^{k-1}\frac{d_i^\top G s_k}{d_i^\top G d_i}d_i, & \text{如果 } k\ge 2.\end{cases}$$

3.设矩阵 $A\in \mathrm{R}^{n\times n}$ 对称正定,向量组 $d_1,d_2,\cdots,d_n\in \mathrm{R}^n$ 关于矩阵 A 共轭. 试证明:

(1) 对任意$x\in \mathrm{R}^n$, $x=\sum\limits_{i=1}^{n}\frac{d_i^\top Ax}{d_i^\top Ad_i}d_i$; (2) $A^{-1}=\sum\limits_{i=1}^{n}\frac{d_id_i^\top}{d_i^\top Ad_i}$.

证明 (1)由性质4.1.1 $\Rightarrow$ 向量组 $d_1,d_2,\cdots,d_n\in \mathrm{R}^n$ 线性无关.

$\Rightarrow$ 对 $\forall x\in \mathrm{R}^n$, 存在实数组 $\alpha_1,\alpha_2,\cdots,\alpha_n$ 使得

$$x=\sum_{i=1}^{n}\alpha_i d_i.$$

从而

$$\langle x,Ad_j\rangle=\left\langle\sum_{i=1}^{n}\alpha_i d_i,Ad_j\right\rangle=\sum_{i=1}^{n}\langle\alpha_i d_i,Ad_j\rangle=\alpha_j d_j^\top Ad_j,\ j=1,2,\cdots,n.$$

这样,

$$\alpha_j=\frac{\langle x,Ad_j\rangle}{d_j^\top Ad_j}=\frac{d_j^\top Ax}{d_j^\top Ad_j},\ j=1,2,\cdots,n.$$

也即

$$x=\sum_{i=1}^{n}\frac{d_i^\top Ax}{d_i^\top Ad_i}d_i.$$

(2)设 $A^{-1}=(a_1,a_2,\cdots,a_n)$, $a_j\in \mathrm{R}^n$, $j=1,2,\cdots,n$. 由 $I=AA^{-1}=(Aa_1,Aa_2,\cdots,Aa_n)$ $\Rightarrow$ $Aa_j=e_j$, $j=1,2,\cdots,n$, 其中 $e_j\in \mathrm{R}^n$ 表示第 j 个分量为 1 其它全为 0 的向量.

由(1)可知

$$a_j=\sum_{i=1}^{n}\frac{d_i^\top Aa_j}{d_i^\top Ad_i}d_i=\sum_{i=1}^{n}\frac{d_i^\top e_j}{d_i^\top Ad_i}d_i=\sum_{i=1}^{n}\frac{d_{ij}}{d_i^\top Ad_i}d_i.$$

其中, d_{ij} 表示向量 d_i 的第 j 个分量.

从而有

$$\begin{aligned}A^{-1}&=(a_1,a_2,\cdots,a_n)\\&=\left(\sum_{i=1}^{n}\frac{d_{i1}}{d_i^\top Ad_i}d_i,\sum_{i=1}^{n}\frac{d_{i2}}{d_i^\top Ad_i}d_i,\cdots,\sum_{i=1}^{n}\frac{d_{in}}{d_i^\top Ad_i}d_i\right)\\&=\sum_{i=1}^{n}\frac{d_i}{d_i^\top Ad_i}(d_{i1},d_{i2},\cdots,d_{in})=\sum_{i=1}^{n}\frac{d_id_i^\top}{d_i^\top Ad_i}.\end{aligned}$$

证毕

4.设矩阵 $A\in \mathrm{R}^{n\times n}$ 对称正定, $b\in \mathrm{R}^n$, 又 $x^*\in \mathrm{R}^n$ 为非线性最优化问题

$$\min\ \ \frac{1}{2}x^\top Ax\qquad \text{s.t.}\ \ x\ge b$$

的最优解. 试证明 x^* 和向量 (x^*-b) 关于矩阵 A 共轭.

证明 该问题的 Lagrange 函数为

$$L(x,\lambda)=\frac{1}{2}x^\top Ax-\lambda^\top(x-b).$$

从而最优解 x^* 满足下述最优性条件

$$\begin{cases}\nabla_x L(x^*,\lambda)=Ax^*-\lambda=0,\\ x^*\lambda\geq b,\ \ \lambda^\top(x^*-b)=0.\end{cases}$$

从而

$$0=(Ax^*)^\top(x^*-b)=(x^*)^\top A(x^*-b).$$

证毕

5.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

$$\beta_k=\frac{d_{k+1}^\top g_{k+1}}{d_k^\top g_k}.$$

证明 由步长规则知,

$$d_{k+1}=-g_{k+1}+\beta_k d_k,\quad g_{k+1}=\beta_k d_k-d_{k+1}.$$

从而由 D-Y 公式

$$\beta_k=\frac{g_{k+1}^\top g_{k+1}}{d_k^\top(g_{k+1}-g_k)}$$

知

$$\beta_k d_k^\top(g_{k+1}-g_k)=g_{k+1}^\top g_{k+1}=(\beta_k d_k-d_{k+1})^\top g_{k+1}.$$

整理得

$$-\beta_k d_k^\top g_k=-d_{k+1}^\top g_{k+1},\ \beta_k=\frac{d_{k+1}^\top g_{k+1}}{d_k^\top g_k}.$$

证毕

6.在强 Wolfe 步长规则下的共轭梯度法中 $\left(0<\sigma_1<\sigma_2<\frac{1}{2}\right)$, 对任意的 $|\beta_k|\leq\beta_k^{\mathrm{FR}}$, 均有

$$-\frac{1}{1-\sigma_2}\leq\frac{g_k^\top d_k}{\|g_k\|^2}\leq\frac{2\sigma_2-1}{1-\sigma_2}.$$

证明 证明思路跟引理4.4.1相似, 我们仅给出部分内容的证明, 其它仿照引理即可.

对任意的 $|\beta_k|\leq\beta_k^{\mathrm{FR}}=\|g_{k+1}\|^2/\|g_k\|^2$, 由$d_{k+1}=-g_{k+1}+\beta_k d_k$知,

$$\frac{g_{k+1}^\top d_{k+1}}{\|g_{k+1}\|^2}=-1+\beta_k\frac{g_{k+1}^\top d_k}{\|g_{k+1}\|^2}.$$

从而

$$\left|\frac{g_{k+1}^{\top}d_{k+1}}{\|g_{k+1}\|^2}+1\right| = |\beta_k|\,\frac{\left|g_{k+1}^{\top}d_k\right|}{\|g_{k+1}\|^2} \le \beta_k^{\mathrm{FR}}\frac{\left|g_{k+1}^{\top}d_k\right|}{\|g_{k+1}\|^2} = \frac{\left|g_{k+1}^{\top}d_k\right|}{\|g_k\|^2}.$$

由强Wolfe步长规则及 $g_k^{\top}d_k < 0$知

$$\left|g_{k+1}^{\top}d_k\right| \le \sigma_2\left|g_k^{\top}d_k\right|.$$

从而

$$\left|\frac{g_{k+1}^{\top}d_{k+1}}{\|g_{k+1}\|^2}+1\right| \le \sigma_2\frac{\left|g_k^{\top}d_k\right|}{\|g_k\|^2}.$$

进一步有,

$$-1+\sigma_2\frac{g_k^{\top}d_k}{\|g_k\|^2} \le \frac{g_{k+1}^{\top}d_{k+1}}{\|g_{k+1}\|^2} \le -1-\sigma_2\frac{g_k^{\top}d_k}{\|g_k\|^2}.$$

证毕

第5章 拟牛顿方法

1.设对称矩阵 A 满足 $\|A\|_{\mathrm{F}} \le 1$, 试证明 $(I-A)$ 为半正定矩阵.

证明 对任意向量x

$$\begin{aligned} x^\top(I-A)x &= \mathrm{tr}\left(x^\top(I-A)x\right) = \left\langle I-A, xx^\top\right\rangle \\ &= \left\|(I-A)xx^\top\right\|_{\mathrm{F}} = \left\|xx^\top - Axx^\top\right\|_{\mathrm{F}} \\ &\ge \left\|xx^\top\right\|_{\mathrm{F}} - \left\|Axx^\top\right\|_{\mathrm{F}} = \|x\|_{\mathrm{F}}^2 - \|A\|_{\mathrm{F}}\|x\|_{\mathrm{F}}^2 \\ &= \|x\|_{\mathrm{F}}^2(1-\|A\|_{\mathrm{F}}) \ge 0. \end{aligned}$$

证毕

2. 设矩阵 $P \in \mathrm{R}^{n\times n}$ 非奇异, $A \in \mathrm{R}^{n\times n}$. 对称. 则

$$\|A\|_{\mathrm{F}} \le \|PAP^{-1} + (PAP^{-1})^\top\|_{\mathrm{F}}.$$

证明 考虑右端项的平方

$$\begin{aligned} &\|PAP^{-1} + (PAP^{-1})^\top\|_{\mathrm{F}}^2 \\ =\ & 4\|PAP^{-1}\|_{\mathrm{F}}^2 = 4\mathrm{tr}(PAP^{-1}PAP^{-1}) \\ =\ & 4\mathrm{tr}(PA^2P^{-1}) = 4\mathrm{tr}(P^{-1}PA^2) \\ =\ & 4\mathrm{tr}(A^2) = 4\|A^2\|_{\mathrm{F}} = 4\|A\|_{\mathrm{F}}^2 \ge \|A\|_{\mathrm{F}}^2. \end{aligned}$$

证毕

3.设 $H_k \in \mathrm{R}^{n\times n}$ 为奇异的对称半正定矩阵. 试证明矩阵 H_{k+1}^{DFP} 奇异.

证明 DFP 校正公式为

$$H_{k+1}^{\mathrm{DFP}} = H_k + \frac{s_k s_k^\top}{s_k^\top y_k} - \frac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k}.$$

由于 $H_k \succ 0$, 则 $\exists\ x \ne 0,\ H_k x = 0\ \Rightarrow$

$$\begin{aligned} H_{k+1}^{\mathrm{DFP}} x &= H_k x + \frac{s_k s_k^\top}{s_k^\top y_k} x - \frac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k} x \\ &= \frac{s_k s_k^\top}{s_k^\top y_k} x = \frac{s_k g_k^\top H_k x}{s_k^\top y_k} = 0. \end{aligned}$$

其中, $s_k = x_{k+1} - x_k = -H_k g_k$. 从而有矩阵 H_{k+1}^{DFP} 奇异. 证毕

4.设 H_k阵对称正定, 且 $s_k^\top y_k > 0$.试证明使 $H_{k+1}^{\mathrm{Broyden}}$ 为秩-1校正公式的参数 ϕ 不在区间 $[0,1]$ 内.

证明 若 $\phi \in [0,1]$, 则

$$0 \le \frac{s_k^\top y_k}{s_k^\top y_k - y_k^\top H_k y_k} \le 1.$$

而$s_k^\top y_k > 0$可以从上式推出$y_k^\top H_k y_k < 0$, 这与 H_k 为对称正定矩阵矛盾. 证毕

5.试证明 BFGS 校正公式可以写成

$$H_{k+1} = H_k + \left(1 + \frac{y_k^\top H_k y_k}{y_k^\top s_k}\right)\frac{s_k s_k^\top}{s_k^\top y_k} - \frac{H_k y_k s_k^\top + s_k y_k^\top H_k}{s_k^\top y_k},$$

并进一步可以写成

$$H_{k+1} = V_k^\top H_k V_k + \frac{s_k s_k^\top}{s_k^\top y_k},$$

其中 $V_k = I - \rho_k y_k s_k^\top$.

证明 由 BFGS 校正公式得

$$\begin{aligned} H_{k+1} &= \left(I - \frac{s_k y_k^\top}{s_k^\top y_k}\right) H_k \left(I - \frac{y_k s_k^\top}{s_k^\top y_k}\right) + \frac{s_k s_k^\top}{s_k^\top y_k} \\ &= H_k - \frac{s_k y_k^\top}{s_k^\top y_k} H_k - H_k \frac{y_k s_k^\top}{s_k^\top y_k} + \frac{s_k y_k^\top H_k y_k s_k^\top}{\left(s_k^\top y_k\right)^2} + \frac{s_k s_k^\top}{s_k^\top y_k} \\ &= H_k + \left(1 + \frac{y_k^\top H_k y_k}{y_k^\top s_k}\right)\frac{s_k s_k^\top}{s_k^\top y_k} - \frac{H_k y_k s_k^\top + s_k y_k^\top H_k}{s_k^\top y_k}. \end{aligned}$$

进一步, 记 $\rho_k := 1/\left(s_k^\top y_k\right)$, $V_k := I - \rho_k y_k s_k^\top$ 有

$$\begin{aligned} H_{k+1} &= \left(I - \frac{s_k y_k^\top}{s_k^\top y_k}\right) H_k \left(I - \frac{y_k s_k^\top}{s_k^\top y_k}\right) + \frac{s_k s_k^\top}{s_k^\top y_k} \\ &= V_k^\top H_k V_k + \frac{s_k s_k^\top}{s_k^\top y_k}. \end{aligned}$$

证毕

6. 在 DFP 校正公式

$$H_{k+1}^{\mathrm{DFP}} = H_k + \frac{s_k s_k^\top}{y_k^\top s_k} - \frac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k}$$

中, 记

$$A_k = \frac{s_k s_k^\top}{y_k^\top s_k}, \qquad B_k = -\frac{H_k y_k y_k^\top H_k}{y_k^\top H_k y_k},$$

假设 H_1 对称正定, $\nabla f(x_k) \neq 0$, $k = 1, 2, \cdots, n$. 证明当用 DFP 方法极小化严格凸二次函数 $f(x) = \frac{1}{2}x^\top G x + g^\top x$时有

$$\sum_{k=1}^{n} A_k = G^{-1}, \quad \sum_{k=1}^{n} B_k = -H_1.$$

证明 二次函数 $f(x)$严格凸 $\Rightarrow G \succ 0$, 由定理5.1.1知, $H_{n+1} = G^{-1}$.

由题设条件

$$
\begin{aligned}
& H_{k+1} = H_k + A_k + B_k, \\
\Rightarrow & \sum_{k=1}^{n} H_{k+1} = \sum_{k=1}^{n} H_k + \sum_{k=1}^{n} A_k + \sum_{k=1}^{n} B_k, \\
\Rightarrow & H_{n+1} = H_1 + \sum_{k=1}^{n} A_k + \sum_{k=1}^{n} B_k, \\
\Rightarrow & G^{-1} = H_1 + \sum_{k=1}^{n} A_k + \sum_{k=1}^{n} B_k.
\end{aligned}
$$

而由假设,

$$
\begin{aligned}
& \sum_{k=1}^{n} B_k = -\overline{H}_1 \neq -H_1 \\
\Longrightarrow & \sum_{k=1}^{n} A_k = G^{-1} + \overline{H}_1 - H_1, \\
\Longrightarrow & \sum_{k=1}^{n} \frac{s_k s_k^\top}{y_k^\top s_k} = G^{-1} + \overline{H}_1 - H_1, \\
\xRightarrow{y_k = G s_k} & \sum_{k=1}^{n} \frac{s_k s_k^\top}{s_k^\top G s_k} = G^{-1} + \overline{H}_1 - H_1, \\
\Longrightarrow & \sum_{k=1}^{n} \frac{s_k s_k^\top G}{s_k^\top G s_k} = I + \left(\overline{H}_1 - H_1\right) G, \\
\Longrightarrow & \operatorname{tr}\left(\sum_{k=1}^{n} \frac{s_k s_k^\top G}{s_k^\top G s_k}\right) = \operatorname{tr}\left(I + \left(\overline{H}_1 - H_1\right) G\right), \\
\Longrightarrow & n = n + \operatorname{tr}\left(\left(\overline{H}_1 - H_1\right) G\right), \\
\Longrightarrow & \operatorname{tr}\left(\left(\overline{H}_1 - H_1\right) G\right) = \left\langle \overline{H}_1 - H_1, G \right\rangle = 0, \\
\xRightarrow{G \succ 0} & \overline{H}_1 - H_1 = 0,
\end{aligned}
$$

矛盾, 从而有

$$
\sum_{k=1}^{n} B_k = -H_1, \quad \sum_{k=1}^{n} A_k = G^{-1}.
$$

证毕

第6章 最小二乘问题

1.考虑如下数据拟合问题: 设拟合函数为

$$\phi(t) = \left(1 - \frac{x_1}{x_2} t\right)^{\frac{1}{cx_1} - 1},$$

其中 $c = 96.05$. 通过观察得到如下数据

t_i	2000	5000	10000	20000	30000	50000
ϕ_i	0.9427	0.8616	0.7384	0.5362	0.3739	0.3096

试用 Gauss-Newdon 算法求参数 x_1, x_2 , 使得残量函数的 2-范数最小.

解: 利用 MATLAB 软件解得 $x_1 = -0.01604,\ x_2 = 671.0745.$ 证毕

2.设 $A \in \mathrm{R}^{m_1 \times n}, B \in \mathrm{R}^{m_2 \times n}, b \in \mathrm{R}^{m_1}$. 试证明下述线性方程组

$$\begin{pmatrix} I & A \\ A^\top & -B^\top B \end{pmatrix} \begin{pmatrix} y \\ x \end{pmatrix} = \begin{pmatrix} b \\ 0 \end{pmatrix} \tag{6.1}$$

关于 x 的解为如下线性最小二乘问题

$$\min_{x \in \mathrm{R}^n} \left\| \begin{pmatrix} A \\ B \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|^2 \tag{6.2}$$

的最优解.

证明 线性方程组 (6.1) 等价于下述方程组问题

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

而该问题的解满足

$$A^\top b = (A^\top A + B^\top B)\, x,$$

问题 (6.2) 的最优解满足

$$\nabla_x \left\| \begin{pmatrix} A \\ B \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right\|^2 = \begin{pmatrix} A \\ B \end{pmatrix}^\top \left[\begin{pmatrix} A \\ B \end{pmatrix} x - \begin{pmatrix} b \\ 0 \end{pmatrix} \right] = 0$$

也即

$$A^\top b = (A^\top A + B^\top B)\, x,$$

结果得证. 证毕

3.对非线性最小二乘问题 (6.2.1), 设对任意 $x \in \mathrm{R}^n$, 向量函数 $r(x)$ 的 Jacobi 矩阵 $J(x)$ 列满秩. 则

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

证明 由式 (6.2.7) 知

$$\begin{aligned} d^{\mathrm{LM}} &= -\left(J(x)^\top J(x) + \mu I\right)^{-1} J(x)^\top r(x), \\ \Rightarrow \lim_{\mu\to 0} d^{\mathrm{LM}}(\mu) &= -\left(J(x)^\top J(x)\right)^{-1} J(x)^\top r(x) = d^{\mathrm{GN}}. \end{aligned}$$

显然, $\left\{\frac{d^{\mathrm{LM}}(\mu)}{\|d^{\mathrm{LM}}(\mu)\|}\right\}$ 有界. 所以存在收敛子列, 不妨设是其本身 (有待完善). 证毕

4.已知某物理量 y 与另外两个物理量 t_1, t_2 的关系为

$$y = \frac{x_1 x_3 t_1}{1 + x_1 t_1 + x_2 t_2},$$

其中 x_1, x_2, x_3 为待定参数. 为确定这三个参数, 测得 t_1, t_2 和 y 的如下一组数据:

t_1	1.0	2.0	1.0	2.0	0.1
t_2	1.0	1.0	2.0	2.0	0.0
y	0.13	0.22	0.08	0.13	0.19

(1)用最小二乘法确立关于 x_1, x_2, x_3 的数学模型;

(2)对列出的非线性最小二乘问题写出 Gauss-Newton 公式的迭代形式.

证明 (1)最小二乘得其数学模型为

$$\min \sum_{i=1}^{5}\left(y_i - \frac{x_1 x_3 t_{1i}}{1 + x_1 t_{1i} + x_2 t_{2i}}\right)^2$$

其中, $y = \{y_1, y_2, \cdots, y_5\}^\top$, $t1 = \{t_{11}, t_{12}, \cdots, t_{15}\}^\top$, $t2 = \{t_{21}, t_{22}, \cdots, t_{25}\}^\top$. 利用 MATLAB 软件可以求得 $x_1 = 0.9639$, $x_2 = 1.3415$, $x_3 = 0.5426$, 关系式为

$$y = \frac{0.5230 t_1}{1 + 0.9639 t_1 + 1.3415 t_2}.$$

(2)令 $r(x) = (r_1(x), r_2(x), \cdots, r_5(x))^\top$

$$r_i(x) = y_i - \frac{x_1 x_3 t_{1i}}{1 + x_1 t_{1i} + x_2 t_{2i}}, \quad i = 1, 2, \cdots, 5$$

记 $J(x)=D_x r(x)$, $S(x)=\sum_{i=1}^{5} r_i(x)\nabla^2 r_i(x)$, 则该问题的 Gauss-Newton 公式的迭代形式为

$$x_{k+1}=x_k-\left(J(x_k)^\top J(x_k)+S(x_k)\right)^{-1} J(x_k)^\top r(x_k).$$

具体计算过程此处略. 证毕

5.设 $A\in \mathrm{R}^{m\times n}$, $r\in \mathrm{R}^m$, $\mu_1>\mu_2>0$, 并设 x_1, x_2 分别是方程

$$\left(A^\top A+\mu I\right)x=-A^\top r,$$

在 μ 取 μ_1, μ_2 时的解, 则 $\|Ax_2+r\|^2<\|Ax_1+r\|^2$.

证明 考虑如下凸优化模型

$$\min_{x\in \mathrm{R}^n} \quad \|Ax+r\|^2+\mu\|x\|^2$$

该优化模型在最优解处满足的充要条件为

$$\left(A^\top A+\mu I\right)x=-A^\top r,$$

而, 当 μ 取 μ_1, μ_2 时,其最优解为 x_1, x_2. 又由性质 6.2.1 知, 该模型的最优解是关于 $\mu>0$ 单调不增函数, 所以

$$\mu_1>\mu_2>0\Rightarrow \|x_1\|<\|x_2\|,$$

且

$$\begin{aligned}
&\|Ax_2+r\|^2+\mu_2\|x_2\|^2\le\|Ax_1+r\|^2+\mu_2\|x_1\|^2,\\
\Rightarrow\ &\|Ax_2+r\|^2-\|Ax_1+r\|^2\le\mu_2\left(\|x_1\|^2-\|x_2\|^2\right)<0,\\
\Rightarrow\ &\|Ax_2+r\|^2<\|Ax_1+r\|^2.
\end{aligned}$$

证毕

第7章 约束优化最优性条件

1.设 $A \in \mathrm{R}^{m\times n}$, $b \in \mathrm{R}^m$. 试给出下述区域中点的可行方向所满足的条件:

$$\Omega_1 = \{x \in \mathrm{R}^n \mid Ax = b, x \geq 0\}, \qquad \Omega_2 = \{x \in \mathrm{R}^n \mid Ax \geq b, x \geq 0\}.$$

解: 根据定义 7.1.1 知, Ω_1 的可行方向 d 满足

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

$\Rightarrow$ (1)$Ad = 0$, 且 (2) 若 x 存在分量为零则 $d \geq 0$; 若 x 不存在分量为零则 $d \in \mathrm{R}^n$.

Ω_2 的可行方向 d 满足

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

$\Rightarrow$ (1)若存在 i 使 $(Ax)_i = b_i$ 则 $Ad \geq 0$; 若 $Ax > b$ 则 $Ad \in \mathrm{R}^m$, 且 (2) 若 x 存在分量为零则 $d \geq 0$; 若 x 不存在分量为零则 $d \in \mathrm{R}^n$. 证毕

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

(1) $Ax \geq 0,\ x \geq 0,\ b^\top x > 0$; (2) $A^\top y \geq b,\ y \leq 0$.

证明 记

$$B := \begin{pmatrix} A \\ I \end{pmatrix}_{2m\times n}, \qquad \bar{b} := -b, \qquad \bar{y} := -y.$$

则

$$(1) \Leftrightarrow Bx = \begin{pmatrix} A \\ I \end{pmatrix} x \geq 0,\ c^\top x < 0;$$

而

$$\begin{aligned} \text{系统(2)有解} &\Leftrightarrow A^\top \bar{y} \leq c,\ \bar{y} \geq 0, \\ &\Leftrightarrow A^\top \bar{y} + z = c,\ \bar{y} \geq 0, z \geq 0, \\ &\Leftrightarrow Bw = c,\ w := \begin{pmatrix} \bar{y} \\ z \end{pmatrix} \geq 0. \end{aligned}$$

由引理 7.2.2$'$ 知上述系统之一有解, 所以 (1) 和 (2) 恰好只有一个有解. 证毕

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

$$\min_{\mu\in\mathrm{R},\ x\in\mathrm{R}^n} \left\| A - \mu x x^\top \right\|_{\mathrm{F}}^2 = \sum_{i,j=1}^n (a_{ij} - \mu x_i x_j)^2, \quad \text{s.t. } x^\top x = 1$$

的解为 A 在绝对值意义下的最大特征值及其对应的特征向量.

证明　考虑目标函数

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

原优化问题的 Lagrange 函数为

$$L(\mu, x, \lambda)=\|A\|_{\mathrm{F}}^{2}-2\mu x^{\top} A x+\mu^{2}\|x\|^{4}-\lambda(x^{\top} x-1).$$

所以最优性条件为

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

其中, $\nabla_{x}\|x\|^{4}=\nabla_{x}\left(\|x\|^{2}\|x\|^{2}\right)=2x\|x\|^{2}+\|x\|^{2} 2x=4\|x\|^{2} x$, 得到

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

当 $\mu \neq 0$ 时, 有 $\mu x=A x$, 即 μ 是 A 的特征值, x 是 A 关于特征值 μ 对应的特征向量. 又

$$\begin{aligned}\left\|A-\mu x x^{\top}\right\|_{\mathrm{F}}^{2} &= \|A\|_{\mathrm{F}}^{2}-2 x^{\top} A x+\mu\|x\|^{4} \\ &= \|A\|_{\mathrm{F}}^{2}-2\mu x^{\top} \mu x+\mu\|x\|^{4} \\ &= \|A\|_{\mathrm{F}}^{2}-\mu^{2}\end{aligned}$$

所以, 要使目标函数值达到最小, 只需要使

$$\mu_{0}=\left|\lambda_{\max }(A)\right|.$$

其中, $\left|\lambda_{\max }(A)\right|$ 表示 A 在绝对值意义下的最大特征值. 另外, 显然

$$\|A\|_{\mathrm{F}}^{2}-\mu_{0}^{2} \leq\|A\|_{\mathrm{F}}^{2}=\left\|A-0 x x^{\top}\right\|_{\mathrm{F}}^{2},$$

因此, 无需再考虑 $\mu=0$ 时的情况. 证毕

4.设约束优化问题 (7.1.1) 中的函数连续可微, x^{*} 为其一局部最优值点. 记该问题在 x^{*} 点的可行方向所构成的集合的闭, 即可行域在 x^{*} 点的切锥为 $T\left(x^{*}\right)$, 约束优化问题在 x^{*} 点的线性化锥的定义为

$$\mathcal{F}\left(x^{*}\right)=\left\{s \in \mathrm{R}^{n} \mid s^{\top} \nabla c_{i}\left(x^{*}\right)=0,\ i \in \mathcal{E};\ \ s^{\top} \nabla c_{i}\left(x^{*}\right) \geq 0,\ i \in \mathcal{I}\left(x^{*}\right)\right\}.$$

若 $T(x^*) = \mathcal{F}(x^*)$, 则约束优化问题在 x^* 点 KKT 条件成立.

证明 首先 A 的极锥满足

$$A^\circ = \{z \in \mathrm{R}^n \mid \langle z, y\rangle \le 0,\ y \in A\}.$$

其次, 我们说, 若 x^* 为局部最优值点 $\Rightarrow -\nabla f(x^*) \in T^\circ(x^*)$. 事实上, 记可行域为 S, 则

$$T(x^*) = \left\{y \in \mathrm{R}^n \mid \lim_{k\to\infty} \alpha_k \left(x^k - x^*\right) = y, \lim_{k\to\infty} x^k = x^*, x^k \in S, \alpha_k \ge 0, k = 1, 2, 3, \cdots\right\}.$$

那么, $\forall\ y \in T(x^*)$, $\exists\ x^k \in S, x^k \to x^*$ 的 $\left\{x^k\right\}$ 以及 $\alpha_k \ge 0$ 使

$$\alpha_k(x^k - x^*) \to y,$$

又 f 在 x^* 可微, 以及 x^* 局部极小 $\Rightarrow$

$$\begin{aligned}
& f(x^k) - f(x^*) = \nabla f(x^*)(x^k - x^*) + o(\|x^k - x^*\|) \ge 0,\\
\Rightarrow\ & = \left\langle \nabla f(x^*), \alpha_k(x^k - x^*)\right\rangle + \frac{o(\|x^k - x^*\|)}{\|x^k - x^*\|}\alpha_k\|x^k - x^*\| \ge 0,\\
\Rightarrow\ & \langle \nabla f(x^*), y\rangle + 0\cdot y \ge 0,\ k \to \infty,\\
\Rightarrow\ & \langle \nabla f(x^*), y\rangle \ge 0,\\
\Rightarrow\ & -\nabla f(x^*) \in T^\circ(x^*) = \{z \in \mathrm{R}^n \mid \langle z, y\rangle \le 0,\ y \in T(x^*)\}.
\end{aligned}$$

下证明最后结果. 令

$$K = \left\{\ y \in \mathrm{R}^n \middle|\ y^\top a_i \le 0, i = 1, 2, \cdots, m;\ y^\top b_j = 0, j = 1, 2, \cdots, l\right\},$$

可以证明 K 的极锥为

$$K^\circ = \left\{\ x \in \mathrm{R}^n \middle|\ x = \sum_{i=1}^m \alpha_i a_i + \sum_{j=1}^l \beta_j b_j;\ \alpha_i \ge 0, i = 1, 2, \cdots, m\right\}.$$

从而有

$$\mathcal{F}^\circ(x^*) = \left\{\ x \in \mathrm{R}^n \middle|\ x = \sum_{i\in\mathcal{I}(x^*)} \lambda_i \nabla c_i(x^*) + \sum_{i\in\mathcal{E}} \mu_i \nabla c_i(x^*);\ \lambda_i \ge 0, i \in \mathcal{I}(x^*)\right\}.$$

由已知条件, 若 $T(x^*) = \mathcal{F}(x^*)\ \Rightarrow\ -\nabla f(x^*) \in T^\circ(x^*) = \mathcal{F}^\circ(x^*)$.

则 $\exists\ \bar\lambda_i \ge 0, i \in \mathcal{I}(x^*);\ \bar\mu_i \in \mathrm{R}, i \in \mathcal{E}$ 使

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

令 $\bar{\lambda}_i = 0, i \in \mathcal{I} \backslash \mathcal{I}(x^*)$, 从而有

$$0 = \nabla f(x^*) + \sum_{i \in \mathcal{I}} \bar{\lambda}_i \nabla c_i(x^*) + \sum_{i \in \mathcal{E}} \bar{\mu}_i \nabla c_i(x^*).$$

证毕

5. 设$A_1 \in \mathrm{R}^{m_1 \times n}$, $A_2 \in \mathrm{R}^{m_2 \times n}$, $b_1 \in \mathrm{R}^{m_1}$, $b_2 \in \mathrm{R}^{m_2}$, $c \in \mathrm{R}^n$, $d \in \mathrm{R}$. 若对任意的$x \in \mathrm{R}^n$都有

$$\left.\begin{array}{l} A_1 x = b_1 \\ A_2 x \leqslant b_2 \end{array}\right\} \Rightarrow c^\top x \leqslant d,$$

则存在$y_1 \in \mathrm{R}^{m_1}, y_2 \in \mathrm{R}_+^{m_2}$, 满足

$$\left\{\begin{array}{l} c = A_1^\top y_1 + A_2^\top y_2, \\ d \geqslant b_1^\top y_1 + b_2^\top y_2. \end{array}\right.$$

证明 考虑线性规划问题

$$\max \quad c^\top x$$

$$\text{s.t. } A_1 x = b_1$$

$$A_2 x \leqslant b_2$$

根据题设, 该线性规划问题有最优解, 设为x^*. 则$c^\top x^* \leqslant d$. 由定理7.2.1, 存在$y_1 \in \mathrm{R}^{m_1}, y_2 \in \mathrm{R}_+^{m_2}$使得

$$\begin{cases} c = A_1^\top y_1 + A_2^\top y_2, \\ y_1^\top (A_1 x^* - b_1) + y_2^\top (A_2 x^* - b_2) = 0. \end{cases}$$

利用前一式可将后一式整理得,

$$\begin{aligned} 0 &= (y_1^\top A_1 + y_2^\top A_2) x^* - (y_1^\top b_1 + y_2^\top b_2) \\ &= c^\top x^* - (y_1^\top b_1 + y_2^\top b_2) \\ &\leqslant d - (y_1^\top b_1 + y_2^\top b_2). \end{aligned}$$

证毕

6. 对$A_1 \in \mathrm{R}^{m_1 \times n}$, $A_2 \in \mathrm{R}^{m_2 \times n}$, $b_1 \in \mathrm{R}^{m_1}$, $b_2 \in \mathrm{R}^{m_2}$, $c \in \mathrm{R}^n$, $d \in \mathrm{R}$, 设下述线性系统有解

$$\begin{cases} A_1 x = b_1, \\ A_2 x \leqslant b_2. \end{cases}$$

则下述两系统恰有一个有解:

$$(1) \quad \begin{cases} A_1 x = b_1, \\ A_2 x \leqslant b_2, \\ c^\top x > d, \end{cases} \qquad\qquad (2) \quad \begin{cases} c = A_1^\top y_1 + A_2^\top y_2, \\ d \geqslant b_1^\top y_1 + b_2^\top y_2, \\ y_1 \in \mathrm{R}^{m_1}, y_2 \in \mathrm{R}_+^{m_2}. \end{cases}$$

证明 若系统(1)无解, 利用前一命题结论知系统(2)有解.

反过来, 设系统(1)有解. 反设系统(2)也有解, 则存在$x \in \mathrm{R}^n$ 和$y_1 \in \mathrm{R}^{m_1}, y_2 \in \mathrm{R}_+^{m_2}$ 分别满足系统(1)(2). 则

$$d < x^\top c = x^\top A_1^\top y_1 + x^\top A_2^\top y_2 \leqslant b_1^\top y_1 + b_2^\top y_2 \leqslant d.$$

此矛盾说明系统(2)无解. 证毕

7.建立下述优化问题的 KKT 条件和二阶最优性条件, 并讨论为什么最优值无下界

$$\begin{aligned} \min \quad & -3x_1 + x_2 - x_3^2, \\ \text{s.t.} \quad & x_1 + x_2 + x_3 \le 0, \\ & -x_1 + 2x_2 + x_3^2 = 0. \end{aligned}$$

解: 记

$$c_1(x) := -x_1 - x_2 - x_3, \quad c_2(x) := -x_1 + 2x_2 + x_3^2.$$

该问题的 Lagrange 函数为

$$L(x,\lambda) = -3x_1 + x_2 - x_3^2 + \lambda_1\left(x_1 + x_2 + x_3\right) + \lambda_2\left(-x_1 + 2x_2 + x_3^2\right).$$

得该优化问题的 KKT 条件

$$\begin{cases} \nabla_{x_1} L(x,\lambda) = -3 + \lambda_1 - \lambda_2 = 0, \\ \nabla_{x_2} L(x,\lambda) = 1 + \lambda_1 + 2\lambda_2 = 0, \\ \nabla_{x_3} L(x,\lambda) = -2x_3 + \lambda_1 + 2\lambda_2 x_3 = 0, \\ \lambda_1\left(x_1 + x_2 + x_3\right) = 0,\ x_1 + x_2 + x_3 \le 0, \\ -x_1 + 2x_2 + x_3^2 = 0,\ \lambda_1 \ge 0. \end{cases}$$

解得

$$\lambda_1^* = \frac{5}{3},\ \lambda_2^* = -\frac{4}{3},\ x_1^* = -\frac{115}{588},\ x_2^* = -\frac{95}{588},\ x_3^* = \frac{5}{14}.$$

下面建立二阶必要条件, 计算知

$$\nabla c_1(x^*) = \begin{pmatrix} -1 \\ -1 \\ -1 \end{pmatrix}, \quad \nabla c_2(x^*) = \begin{pmatrix} -1 \\ 2 \\ 1 \end{pmatrix}.$$

显然, $\nabla c_1(x^*)$ 与 $\nabla c_2(x^*)$ 线性无关. 对于满足 $s^\top \nabla c_1(x^*)$, $s^\top \nabla c_2(x^*)$ 的 s, 即

$$\begin{cases} -s_1 - s_2 - s_3 = 0, \\ -s_1 + 2s_2 + s_3 = 0. \end{cases} \Rightarrow s = (t, 2t, -3t)^\top, (t \ne 0).$$

故

$$s^\top \nabla_{xx} L(x^*,\lambda^*) s = (t, 2t, -3t)\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 2\lambda_2^* - 2 \end{pmatrix}\begin{pmatrix} t \\ 2t \\ -3t \end{pmatrix} = -42t^2 < 0.$$

所以, 知 $x^* = \left(-\frac{115}{588}, -\frac{95}{588}, \frac{5}{14}\right)$ 不是该优化问题的局部最优解.

另外, 原问题等价于

$$\begin{aligned} \min \quad & -4x_1 + 3x_2, \\ \text{s.t.} \quad & x_1 + x_2 + x_3 \le 0. \end{aligned}$$

显然, 问题 (7) 无下界 (可取 $x_1 = x_3 = 0, x_2 = -\infty$). 证毕

8.设 $A \in \mathrm{R}^{l\times m}$, $B \in \mathrm{R}^{l\times n}$, $c \in \mathrm{R}^l$, $f : \mathrm{R}^m \to \mathrm{R}$ 和 $g : \mathrm{R}^n \to \mathrm{R}$ 连续可微, $\mathcal{X}$ 和 $\mathcal{Y}$ 分别为 R^m 和 R^n 的非空闭集. 对约束优化问题

$$\min \ f(x) + g(y), \quad \text{s.t.} \ \ Ax + By = c, \ x \in \mathcal{X}, y \in \mathcal{Y},$$

引入 Lagrange 函数

$$L(x, y, \lambda) = f(x) + g(y) - \lambda^\top (Ax + By - c).$$

试证明原约束优化问题的稳定点等价于满足下述条件的 $z^* \in \mathcal{Z}$:

$$\langle Q(z^*), z - z^* \rangle \ge 0, \quad \forall z \in \mathcal{Z},$$

其中

$$\mathcal{Z} = \mathcal{X} \times \mathcal{Y} \times \mathrm{R}^l, \ z = \begin{pmatrix} x \\ y \\ \lambda \end{pmatrix}, \ Q(z) = \begin{pmatrix} \nabla_x L(x, y, \lambda) \\ \nabla_y L(x, y, \lambda) \\ Ax + By - c \end{pmatrix}.$$

证明 显然 Lagrange 函数连续可微, 其在点 $z^* = (x^*, y^*, \lambda^*)^\top$ 处的泰勒展开为

$$L(x, y, \lambda) = L(x^*, y^*, \lambda^*) + \langle Q(z^*), z - z^* \rangle + o(\|z - z^*\|), \ \forall x \in \mathcal{X}, y \in \mathcal{Y}, \lambda \in \mathrm{R}.$$

若 $\langle Q(z^*), z - z^* \rangle \ge 0$, 则对 $\forall x \in \mathcal{X}, y \in \mathcal{Y}, \lambda \in \mathrm{R}$ 有

$$\begin{aligned} & L(x, y, \lambda) \ge L(x^*, y^*, \lambda^*) + o(\|z - z^*\|), \\ \Rightarrow\ & f(x) + g(y) - \lambda^\top (Ax + By - c) \ge L(x^*, y^*, \lambda^*) = f(x^*) + g(y^*), \\ \overset{\lambda\to 0}{\Rightarrow}\ & f(x) + g(y) \ge f(x^*) + g(y^*), \ \forall x \in \mathcal{X}, y \in \mathcal{Y}. \end{aligned}$$

即 $(x^*, y^*)^\top$ 为原问题的稳定点. 另一方面, 若 $(x^*, y^*)^\top$ 为原问题的稳定点, 对 $\forall x \in \mathcal{X}, y \in \mathcal{Y}, \lambda \in \mathrm{R}$ 有

$$\begin{aligned} L(x, y, \lambda) &= L(x^*, y^*, \lambda^*) + \langle Q(z^*), z - z^* \rangle + o(\|z - z^*\|), \\ &= f(x^*) + g(y^*) + \langle Q(z^*), z - z^* \rangle + o(\|z - z^*\|), \\ &\le f(x) + g(y) + \langle Q(z^*), z - z^* \rangle + o(\|z - z^*\|). \end{aligned}$$

$$\Rightarrow$$

$$\begin{aligned}
&-\lambda^{\top}(Ax+By-c)\leq\langle Q(z^*),z-z^*\rangle+o(\|z-z^*\|),\\
\overset{\lambda\to0}{\Rightarrow}\ &\langle Q(z^*),z-z^*\rangle+o(\|z-z^*\|)\geq0,\\
\Rightarrow\ &\langle Q(z^*),z-z^*\rangle\geq0,\forall z\in\mathcal{Z}.
\end{aligned}$$

证毕

9.设 $a=(a_1,a_2,\cdots,a_n)$, $c=(c_1,c_2,\cdots,c_n)>0$, b 为正整数, 则对下述优化问题

$$\min\ f(x)=\sum_{i=1}^n\frac{c_i}{x_i},\qquad \text{s.t.}\ \ \sum_{i=1}^n a_ix_i=b,x\geq0$$

最优值 $f^*=\frac{1}{b}\left[\sum_{i=1}^n(a_ic_i)^{1/2}\right]^2$.

证明 该问题的 Lagrange 函数为

$$L(x,\lambda,\mu)=\sum_{i=1}^n\frac{c_i}{x_i}-\lambda\left(\sum_{i=1}^n a_ix_i-b\right)-\mu^{\top}x.$$

得最优性条件

$$\begin{cases}\nabla_{x_i}L(x,\lambda,\mu)=-\frac{c_i}{x_i^2}-\lambda a_i-\mu_ix_i=0,\quad i=1,2,\cdots,n,\\ \sum_{i=1}^n a_ix_i=b;\quad \mu^{\top}x=0,\quad x\geq0,\ \mu\geq0.\end{cases}$$

显然由目标函数知 $x_i^*>0,\ i=1,2,\cdots,n\quad\Rightarrow\mu^*=0\quad\Rightarrow$

$$\begin{aligned}
&\frac{c_i}{x_i^*}=-\lambda^*a_ix_i^*,\ i=1,2,\cdots,n,\\
\Rightarrow\ &\sum_{i=1}^n\frac{c_i}{x_i^*}=-\lambda^*\sum_{i=1}^n a_ix_i^*=-\lambda^*b,\\
\Rightarrow\ &\lambda^*=-\frac{f(x^*)}{b}=-\frac{f^*}{b}.
\end{aligned}$$

从而

$$\begin{aligned}
&\frac{c_i}{x_i^{*2}}=-\lambda^*a_i,\ i=1,2,\cdots,n,\\
\Rightarrow\ &\frac{c_i}{x_i^*}=\sqrt{-\lambda^*a_ic_i}=\sqrt{\frac{f^*}{b}}\sqrt{a_ic_i},\ i=1,2,\cdots,n,\\
\Rightarrow\ &f^*=\sum_{i=1}^n\frac{c_i}{x_i^*}=\sqrt{\frac{f^*}{b}}\sum_{i=1}^n\sqrt{a_ic_i},\\
\Rightarrow\ &f^*=\frac{1}{b}\left[\sum_{i=1}^n(a_ic_i)^{1/2}\right]^2.
\end{aligned}$$

证毕

10.写出下述优化问题

$$\begin{aligned} \min \quad & \left(x_1-\frac{4}{9}\right)^2+(x_2-2)^2, \\ \text{s.t.} \quad & -x_1^2+x_2 \geq 0, \\ & x_1+x_2 \leq 6, \\ & x_1, x_2 \geq 0. \end{aligned}$$

的 KKT 条件, 并验证 $\left(\frac{4}{9}, 2\right)$ 为该优化问题的唯一全局最优值点.

解: 该问题的 Lagrange 函数为

$$L(x,\lambda)=\left(x_1-\frac{4}{9}\right)^2+(x_2-2)^2-\lambda_1\left(-x_1^2+x_2\right)-\lambda_2\left(6-x_1-x_2\right)-\lambda_3 x_1-\lambda_4 x_2.$$

得该优化问题的 KKT 条件

$$\begin{cases} \nabla_{x_1} L(x,\lambda)=2\left(x_1-\frac{4}{9}\right)+2\lambda_1 x_1+\lambda_2-\lambda_3=0, \\ \nabla_{x_2} L(x,\lambda)=2\left(x_2-2\right)-\lambda_1+\lambda_2-\lambda_4=0, \\ \lambda_1\left(-x_1^2+x_2\right)=0, -x_1^2+x_2 \geq 0, \\ \lambda_2\left(6-x_1-x_2\right)=0, x_1+x_2 \leq 6, \\ \lambda_3 x_1=0,\ \lambda_4 x_2=0,\ x_1, x_2 \geq 0,\ \lambda \geq 0. \end{cases}$$

容易验证 $x_1=\frac{4}{9},\ x_2=2, \lambda_1=\lambda_2=\lambda_3=\lambda_4=0$ 满足上述 KKT 条件, 又容易验证该优化为题为凸优化, 所以 $\left(\frac{4}{9}, 2\right)$ 为该优化问题的唯一全局最优值点. 证毕

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

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

解: 问题的 Lagrange 函数为

$$L(x,\lambda)=\|x-b\|^2-\lambda^\top Ax, \quad \lambda \in \mathrm{R}^m,\ x \in \mathrm{R}^n.$$

则 Lagrange 对偶函数为

$$\theta(\lambda)=\inf_{x\in\mathrm{R}^n}\{L(x,\lambda)\}, \quad \lambda \in \mathrm{R}^m.$$

由于

$$\begin{aligned} \nabla_x L(x,\lambda)=2(x-b)-A^\top\lambda=0 \ \Rightarrow\ x=\frac{1}{2}A^\top\lambda+b. \quad \Rightarrow \\ \theta(\lambda) &= \|\frac{1}{2}A^\top\lambda+b-b\|^2-\lambda^\top A\left(\frac{1}{2}A^\top\lambda+b\right), \\ &= \frac{1}{4}\|A^\top\lambda\|^2-\frac{1}{2}\|A^\top\lambda\|^2-\lambda^\top Ab, \\ &= -\frac{1}{4}\|A^\top\lambda\|^2-\lambda^\top Ab, \end{aligned}$$

从而 Lagrange 对偶为

$$\max_{\lambda\in\mathrm{R}^m} \ -\frac{1}{4}\|A^\top\lambda\|^2-\lambda^\top Ab.$$

证毕

12. 设 $A\in\mathrm{R}^{m\times n}$, $b\in\mathrm{R}^n$ 行满秩, $c\in\mathrm{R}^n$ 为非零向量, 且 $A^\top(AA^\top)^{-1}Ac-c$. 求下述优化问题的最优解和最优值

$$\min\ c^\top x,\qquad \text{s.t.}\quad Ax=0, x^\top x\le 1.$$

解: 显然该问题凸且 $x=0$ 满足 Slater 条件, 从而知最优解等价于 K-T 点.

问题的 Lagrange 函数为

$$L(x,\lambda,\mu)=c^\top x-\lambda^\top Ax-\mu(1-x^\top x).$$

得该优化问题的 KKT 条件

$$\begin{cases}\nabla_x L(x,\lambda,\mu)=c-A^\top\lambda+2\mu x=0,\\ Ax=0,\ x^\top x\le 1,\\ \mu(1-x^\top x)=0,\ \mu\ge 0.\end{cases}$$

由 (7) $\Rightarrow \lambda=(AA^\top)^{-1}Ac$ 且

$$\begin{aligned}&\mu x=\frac{A^\top\lambda-c}{2},\ \mu=\mu x^\top x,\\ \Rightarrow\ &\mu^2=\mu x^\top\mu x,\\ \Rightarrow\ &\mu=\|\mu x\|_2=\frac{\|A^\top\lambda-c\|_2}{2}=\frac{\|A^\top(AA^\top)^{-1}Ac-c\|_2}{2},\\ \Rightarrow\ &x=\frac{A^\top\lambda-c}{2\mu}=\frac{A^\top(AA^\top)^{-1}Ac-c}{\|A^\top(AA^\top)^{-1}Ac-c\|_2}.\end{aligned}$$

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

13.对映射 $\phi:\mathrm{R}^n\times\mathrm{R}^m$, 设 $\bar{x}$ 是极大值函数 $\max\{\phi(x,y)\mid y\in\mathrm{R}^m\}$ 的最小值点, $\bar{y}$ 是极小值函数 $\min\{\phi(x,y)\mid x\in\mathrm{R}^n\}$ 的最大值点. 试证明 $(\bar{x},\bar{y})$ 是函数 ϕ 的鞍点的充分必要条件是

$$\phi(\bar{x},\bar{y})=\min_{x\in\mathrm{R}^n}\max_{y\in\mathrm{R}^m}\phi(x,y).$$

证明 由鞍点以及 $\bar{x},\bar{y}$ 定义

$$\begin{aligned}&\phi(\bar{x},y)\le\phi(\bar{x},\bar{y})\le\phi(x,\bar{y}),\quad\forall\ x\in\mathrm{R}^n, y\in\mathrm{R}^m\\ \Leftrightarrow\ &\max_{y\in\mathrm{R}^m}\phi(\bar{x},y)\le\phi(\bar{x},\bar{y})\le\min_{x\in\mathrm{R}^n}\phi(x,\bar{y}),\\ \Leftrightarrow\ &\min_{x\in\mathrm{R}^n}\max_{y\in\mathrm{R}^m}\phi(x,y)\le\phi(\bar{x},\bar{y})\le\max_{y\in\mathrm{R}^m}\min_{x\in\mathrm{R}^n}\phi(x,y),\\ \Leftrightarrow\ &\phi(\bar{x},\bar{y})=\min_{x\in\mathrm{R}^n}\max_{y\in\mathrm{R}^m}\phi(x,y)=\max_{y\in\mathrm{R}^m}\min_{x\in\mathrm{R}^n}\phi(x,y).\end{aligned}$$

证毕

14.建立下述优化问题的 Lagrange 对偶规划

$$\min\ x_1^2+x_2^2,\qquad \text{s.t.}\ \ x_1+x_2\geq 4;\ x_1,x_2\geq 0.$$

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

证明 显然

$$x_1^2+x_2^2\geq\frac{(x_1+x_2)^2}{2}\geq\frac{4^2}{2}=8,$$

等号成立的条件是 $x_1=x_2=2$, 所以该优化问题的最优解为 $x_1=x_2=2$, 最优值为 8.(当然, 也可以利用 KKT 条件得到最优解和最优值.)

考虑 Lagrange 函数

$$L(x,\mu)=x_1^2+x_2^2-\mu\left(x_1+x_2-4\right).$$

Lagrange 对偶函数为

$$\theta(\mu)=\inf\left\{L(x,\mu)\mid x_1,x_2\geq 0\right\},\quad \mu\geq 0.$$

考虑

$$\begin{aligned}L(x,\mu)&=x_1^2+x_2^2-\mu\left(x_1+x_2-4\right),\\&=\left(x_1-\frac{\mu}{2}\right)^2+\left(x_2-\frac{\mu}{2}\right)^2-\frac{\mu^2}{2}+4\mu,\\&\geq -\frac{\mu^2}{2}+4\mu.\end{aligned}$$

最后一个不等式中等号成立的条件是 $x_1=x_2=\mu/2\geq 0$. 所以对偶规划的目标函数为

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

从而 Lagrange 对偶为

$$\max\ -\frac{1}{2}(\mu-4)^2+8,\qquad \text{s.t.}\quad \mu\geq 0.$$

显然, $\max\ \theta(\mu)=\theta(4)=8$, 即对偶间隙为零. 证毕

15. 考虑下述优化问题

$$\max\ \sum_{i=1}^n x_i^3,\qquad \text{s.t.}\ \ \sum_{i=1}^n x_i=0,\sum_{i=1}^n x_i^2=n.$$

试通过 KKT 条件计算该问题的 K-T 点, 并通过二阶条件验证这些 K-T 点为原问题的局部极大值点.

解: 原问题等价于下面优化问题

$$\min\ -\sum_{i=1}^n x_i^3,\qquad \text{s.t.}\ \ \sum_{i=1}^n x_i=0,\sum_{i=1}^n x_i^2=n.$$

考虑问题 (7) 的 Lagrange 函数

$$L(x,\lambda_1,\lambda_2) = -\sum_{i=1}^n x_i^3 + \lambda_1\sum_{i=1}^n x_i + \lambda_2\left(\sum_{i=1}^n x_i^2 - n\right).$$

其 KKT 条件为

$$\begin{cases} \nabla_{x_i} L(x,\lambda_1,\lambda_2) = -3x_i^2 + \lambda_1 + 2\lambda_2 x_i = 0,\ i = 1,2,\cdots,n, \\ \sum\limits_{i=1}^n x_i = 0, \sum\limits_{i=1}^n x_i^2 = n. \end{cases}$$

$$\Rightarrow \quad \sum_{i=1}^n \left(-3x_i^2 + \lambda_1 + 2\lambda_2 x_i\right) = -3n + n\lambda_1 = 0,\ \Rightarrow \quad \lambda_1 = 3.$$

$$\Rightarrow \quad \forall\ i,\ x_i \neq 0, \quad \Rightarrow$$

$$\lambda_2 = \frac{3}{2}\left(x_i - \frac{1}{x_i}\right),\ i = 1,2,\cdots,n,$$

$$\Rightarrow x_i = x_j \quad \text{or} \quad x_i = -\frac{1}{x_j}, \quad i \neq j.$$

并且上两式要同时存在, 即其中一式不可能对于所有的 $i \in \{1,2,\cdots,n\}$ 都成立. 所以, 不妨设

$$x_1 = x_2 = \cdots = x_k = a,$$

$$x_{k+1} = x_{k+2} = \cdots = x_n = -\frac{1}{a}, \quad k \geq 1,\ a > 0.$$

由

$$\sum_{i=1}^n x_i = 0, \sum_{i=1}^n x_i^2 = n, \quad \Rightarrow \quad \begin{cases} ka + (n-k)\frac{-1}{a} = 0, \\ ka^2 + (n-k)\frac{-1}{a^2} = n. \end{cases}$$

解得

$$k = \frac{n}{a^2+1} \text{ 且 } \lambda_2 = \frac{3}{2}\left(a - \frac{1}{a}\right).$$

说明: 这里的 a 的取值是一定能保证 k 的值为正整数的. 从而, 目标函数值为

$$f(x) = -ka^3 - (n-k)\frac{-1}{a^3} = -\frac{n}{a^2+1}a^3 - (n - \frac{n}{a^2+1})\frac{-1}{a^3} = n(\frac{1}{a} - a).$$

K-T 点 x 满足: $\frac{n}{a^2+1}$ 分量取值 a ; $\frac{na^2}{a^2+1}$ 分量取值 $\frac{1}{a}$.

下面考虑二阶充分条件, 不妨令 K-T 点 $x = \left(a,\cdots,a,\frac{1}{a},\cdots,\frac{1}{a}\right)$, 即前 $k = \frac{n}{a^2+1}$ 个分量取值 a. 记

$$c_1(x) = \sum_{i=1}^n x_i; \quad c_2(x) = \sum_{i=1}^n x_i^2 - n.$$

则由 $s^\top \nabla c_1(x) = 0,\ s^\top \nabla c_2(x) = 0 \quad \Rightarrow$

$$\begin{cases} s_1 + s_2 + \cdots + s_n = (s_1 + \cdots + s_k) + (s_{k+1} + \cdots + s_n) = 0, \\ 2x_1s_1 + 2x_2s_2 + \cdots + 2x_ns_n = 2a\left(s_1 + \cdots + s_k\right) - \frac{2}{a}\left(s_{k+1} + \cdots + s_n\right) = 0. \end{cases}$$

$$\Rightarrow \quad s_1 + \cdots + s_k = 0;\ s_{k+1} + \cdots + s_n = 0.$$

令

$$S := \{s \in \mathrm{R}^n |\ s_1 + \cdots + s_k = 0;\ s_{k+1} + \cdots + s_n = 0\}.$$

可以得到 S 的任意基向量都是形如下面的向量

$$\widetilde{s} = (\underbrace{0,\cdots,0,1,0,\cdots,0,-1,0\cdots,0}_{k个},\underbrace{0,\cdots,0,1,0,\cdots,0,-1,0\cdots,0}_{n-k个})^\top$$

容易计算

$$\begin{aligned}\nabla^2_{xx}L(x,\lambda_1,\lambda_2) &= \begin{pmatrix} -6x_1+2\lambda_2 & & \\ & \ddots & \\ & & -6x_n+2\lambda_2 \end{pmatrix} = -6\begin{pmatrix} x_1 & & \\ & \ddots & \\ & & x_n \end{pmatrix} + 2\lambda_2 I \\ &= -6\begin{pmatrix} a & & & & & \\ & \ddots & & & & \\ & & a & & & \\ & & & \frac{-1}{a} & & \\ & & & & \ddots & \\ & & & & & \frac{-1}{a} \end{pmatrix} + 3\left(a-\frac{1}{a}\right)I.\end{aligned}$$

从而有

$$\widetilde{s}^\top \nabla^2_{xx}L(x,\lambda_1,\lambda_2)\widetilde{s} = -6\left(2a-\frac{2}{a}\right)+3\left(a-\frac{1}{a}\right)4 = 0.$$

所以由二阶充分条件知这些 K-T 点为原问题的局部极大值点. 证毕

16.设 $a_1, a_2, \cdots, a_m \in \mathrm{R}^n$. 试用 KKT 条件给出下述优化问题的最优解

$$\min\ \sum_{i=1}^m \|x-a_i\|^2, \quad \text{s.t.}\ \ x^\top x = 1.$$

解: 该问题的 Lagrange 函数

$$L(x,\lambda) = \sum_{i=1}^m \|x-a_i\|^2 - \lambda\left(x^\top x - 1\right).$$

其 KKT 条件为

$$\begin{cases} \nabla_x L(x,\lambda) = 2\sum\limits_{i=1}^m (x-a_i) - \lambda x = 0, \\ x^\top x = 1. \end{cases}$$

$\Rightarrow$

$$\lambda^* = 2m - 2\|\sum_{i=1}^m a_i\|, \quad x^* = \frac{\sum\limits_{i=1}^m a_i}{\left\|\sum\limits_{i=1}^m a_i\right\|},$$

满足上述 KKT 条件, 又由该规划为凸, 知 x^* 为最优解. 证毕

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

$$\min\{\sum_{i=1}^n\frac{a_i^2}{x_i}\mid e^\top x\leqslant 1,\ x>0\}$$

证明 该问题显然等价于

$$\min\{\sum_{i=1}^n\frac{a_i^2}{x_i}\mid e^\top x=1,\ x>0\}$$

利用KKT条件,

$$\begin{cases}\frac{a_i^2}{x_i^2}=\mu, & i=1,2,\cdots,n\\ \sum\limits_{i=1}^n x_i=1\end{cases}$$

由第一式得a_i^2与x_i^2成比例, 即$|a_i|$与x_i成比例. 利用第二式得$x=\frac{|a|}{\|a\|_1}$. 代入得最优值$\|a\|_1^2$.

19. 对n阶实对称矩阵A,B, 利用对偶理论证明下述结论等价.

(1) 对任意非零向量$x\in\mathrm{R}^n$, $\max\{x^\top Ax,x^\top Bx\}\geqslant 0(>0)$;

(2) 存在非零常数λ,μ满足$\lambda+\mu=1$使得$\lambda A+\mu B$半正定(正定).

证明 考虑优化问题

$$\begin{aligned}\min\quad & t\\ \text{s.t.}\quad & t-x^\top Ax\geqslant 0\\ & t-x^\top Bx\geqslant 0\\ & x^\top x\geqslant 1\end{aligned}$$

显然, (1)成立的充分必要条件是上述优化问题的最优值$t^*\geqslant 0$, 而(1)不成立的充要条件是$t^*=-\infty$.

考虑上述优化问题的Lagrange函数

$$\max_{\mu\geqslant 0}\min_{t,x}L(t,\mu_1,\mu_2,\mu_3)=t-\mu_1(t-x^\top Ax)-\mu_2(t-x^\top Bx)-\mu_3x^\top x$$

对内层优化问题, 利用最优性条件

$$1-\mu_1-\mu_2=0,\quad \mu_1A+\mu_2B-\mu_3I\succcurlyeq 0$$

可将对偶规划问题写成

$$\begin{aligned}\max_{\mu\geqslant 0}\ & 0\\ \text{s.t.}\ & 1-\mu_1-\mu_2=0\\ & \mu_1A+\mu_2B-\mu_3I\succcurlyeq 0\end{aligned}$$

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, 即存在常数$\mu_1, \mu_2 \geqslant 0$满足$\mu_1+\mu_2 = 1$使得$\mu_1 A_1 + \mu_2 A_2 \geqslant 0$. 证毕

20. 考虑优化问题

$$\min_{x\in\mathcal{X}, y\in\mathcal{Y}} f(x,y)$$

其中, $\mathcal{X},\mathcal{Y}$分别为$\mathrm{R}^n, \mathrm{R}^m$中的非空闭凸集. 如果存在$x^* \in \mathcal{X}, y^* \in \mathcal{Y}$使得

$$x^* = \arg\min_{x\in\mathcal{X}} f(x,y^*), \quad y^* = \arg\min_{y\in\mathcal{Y}} f(x^*,y)$$

请问(x^*,y^*)是否为原问题的最优解?

解: 不是, 反例如下. 考虑定义在$[0,1]\times[0,1]$上的二元函数$f(x,y) = xy$. 显然, $(0,0)$点满足

$$x^* = \arg\min_{x\in\mathcal{X}} f(x,y^*), \quad y^* = \arg\min_{y\in\mathcal{Y}} f(x^*,y).$$

但它不是约束优化问题的最优解.

如果目标函数$f(x,y)$在(x^*,y^*)点附近为凸函数, 则结论成立. 因为对任意(x^*,y^*)附近的$(x,y)\in\mathcal{X}\times\mathcal{Y}$,

$$f(x,y) - f(x^*,y^*) \geqslant \nabla_x f(x^*,y^*)^\top (x-x^*) + \nabla_y f(x^*,y^*)^\top (y-y^*).$$

上式右端非负. 结论成立.

21. 对无约束优化问题

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

引入变量$y = x$, 得约束优化问题

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

试利用后者的对偶建立原优化问题的一个易于求解的转化形式.

解: 该问题的Lagrange对偶为

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

其中,

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

由于内层优化问题的目标函数关于x,y可分离, 所以内层优化问题可写成

$$\min_{x\in\mathrm{R}^n} \Big(\frac{1}{2}\|Ax-b\|^2 + z^\top x\Big) + \min_{y\in\mathrm{R}^n} \big(\lambda\|y\|_1 - z^\top y\big)$$

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

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

而对后一优化问题, 利用目标函数的可分离性易得(参第一章习题15)

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

这样, 对偶问题可写成

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

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8章 二次规划

1. 设 a 和 x_0 为 n 非零实向量. 试给出下述二次规划问题的最优解

$$\min\ \|x-x_0\|^2, \quad \text{s.t.}\ \ a^\top x=0.$$

解: 该问题的 Lagrange 函数

$$L(x,\lambda)=\|x-x_0\|^2-\lambda a^\top x.$$

其 KKT 条件为

$$\begin{cases}\nabla_x L(x,\lambda)=2(x-x_0)-\lambda a=0,\\ a^\top x=0.\end{cases}$$

$$\Rightarrow \lambda^*=\frac{-2a^\top x_0}{\|a\|^2},\quad x^*=x_0-\frac{a^\top x_0}{\|a\|^2}a,$$

又由该规划为凸, 知 x^* 为最优解. 证毕

2. 设n阶方阵G对称正定. 试问: $\max\{0,-G^{-1}g\}$是下述优化问题的最优解吗?

$$\min\{\frac{1}{2}x^\top Gx+g^\top x\mid x\geq 0\}$$

试比较

$$G=\begin{pmatrix}2 & -1\\ -1 & 3\end{pmatrix},\qquad g=(-3;2)$$

时, $\max\{0,-G^{-1}g\}$对应的目标函数值和$(3/2;0)$对应的目标函数值的大小.

解: 不一定是. 反例即为

$$G=\begin{pmatrix}2 & -1\\ -1 & 3\end{pmatrix},\qquad g=(-3;2)$$

时, $\max\{0,-G^{-1}g\}$时的二次规划问题. 此时的目标函数为$f(x)=x_1^2+\frac{3}{2}x_2^2-x_1x_2-3x_1+2x_2$

由于

$$G^{-1}=\begin{pmatrix}2 & -1\\ -1 & 3\end{pmatrix}^{-1}=\begin{pmatrix}\frac{3}{5} & \frac{1}{5}\\ \frac{1}{5} & \frac{2}{5}\end{pmatrix}$$

所以,

$$-G^{-1}g=(\frac{7}{5};\frac{-1}{5}).$$

这时$\max\{0,-G^{-1}g\}=(\frac{7}{5};0)$. 容易计算, 该点对应的目标函数值为$-\frac{56}{25}$, $(3/2;0)$对应的目标函数值为$-\frac{9}{4}$. 易知该值小于上一值, 其差为0.01.

3. 设$a,b,c,d\in\mathrm{R}$. 令

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

则对如下二次规划问题

$$\min\{\frac{1}{2}x^\top Gx+ce^\top x\mid e^\top x\geqslant d,x\geqslant 0\}$$

任意调换变量x_i,x_j的位置, 目标函数和约束函数都不变, 故称其为对称二次规划. 试给出矩阵G(半)正定的充要条件, 并用反证法证明若目标函数的Hesse阵G半正定, 则存在二次规划的最优解满足$x_1^*=\cdots=x_n^*$.

解: 首先, n阶实超对称矩阵A为正定矩阵当且仅当$a>-(n-1)b>0$ 或$a>b>0$.

事实上, 由矩阵超对称的定义得，若矩阵A为n阶实超对称矩阵，则A必为实对称矩阵。又因为n阶实对称矩阵正定当且仅当其n个顺序主子式均大于0, 因此上述问题转化为在矩阵A超对称的前提下讨论其顺序主子式的情况。由超对称的性质得A的顺序主子式形式如下：

$$A_k=\begin{vmatrix} a & b & b & \cdots & b\\ b & a & b & \cdots & b\\ b & b & a & \cdots & b\\ \vdots & \vdots & \vdots & & \vdots\\ b & b & b & \cdots & a\end{vmatrix}_{k\times k}=[a+(k-1)b](a-b)^{k-1},\quad k=1,2,\cdots,n$$

要使$A_k>0,\quad\forall\quad k=1,2,\cdots,n$, 即$[a+(k-1)b](a-b)^{k-1}>0,\quad\forall\quad k=1,2,\cdots,n$ 对于上述不等式的解我们分情况讨论：

当k为奇数时，有$(a-b)^{k-1}\geq 0$, 故解为$a>-(k-1)b,\quad\forall\quad k=1,3,5,\cdots$

当k为偶数时，若$b>0$, 则$a<-(k-1)b$ 或$a>b$. 若$b<0$, 则$a>-(k-1)b$ 或$a<b$.

由于一阶顺序主子式$A_1=a>0$, 故上述讨论中当k为偶数且$b>0$时仅有解$a>b$，当k为偶数且$b<0$时仅有解$a>-(k-1)b\quad\forall\quad k=2,4,6,\cdots$。

综上所述，

当$b<0$时，$a>-(k-1)b,\quad k=1,2,3,\cdots,n$ 此时仅要求$a>-(n-1)b$即可；

当$b>0$ k为奇数时，$a>-(k-1)b,\quad k=1,3,5,\cdots,n$ 此时仅要求$a>0$即可；

当$b>0$ k为偶数时，$a>b$。该结论可以简记为$a>-(n-1)b>0$ 或$a>b>0$.

其次, 设该二次规划的最优解为$x=(x_1,x_2,\cdots,x_n)^\top$, 且其分量不全相同, 不妨设$x_1\neq x_2$. 由假设可得目标函数的最优值

$$f(x)=\frac{1}{2}x^\top Gx+g^\top x=\frac{1}{2}a\sum_{i=1}^n x_i^2+b\sum_{\substack{i,j=1\\ i\neq j}}^n x_ix_j+c\sum_{i=1}^n x_i$$

现令$x_0=\frac{x_1+x_2}{2}$，则$\bar{x}=(x_0,x_0,\cdots,x_n)^\top$为该问题的可行解，将其带入目标函数可得

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

从而有

$$\begin{aligned}
f(x)-f(\bar{x})&=(\frac{1}{2}a\sum_{i=1}^n x_i^2+b\sum_{\substack{i,j=1\\i\neq j}}^n x_ix_j+c\sum_{i=1}^n x_i)-(\frac{1}{2}a\sum_{i=1}^n\bar{x}_i^2+b\sum_{\substack{i,j=1\\i\neq j}}^n\bar{x}_i\bar{x}_j+c\sum_{i=1}^n\bar{x}_i)\\
&=[\frac{1}{2}a(x_1^2+x_2^2)+bx_1x_2+b(x_1+x_2)\sum_{i=3}^n x_i+c(x_1+x_2)]\\
&\quad-[\frac{1}{2}a(x_0^2+x_0^2)+bx_0x_0+b(x_0+x_0)\sum_{i=3}^n x_i+c(x_0+x_0)]\\
&=[\frac{1}{2}a(x_1+x_2)^2+(b-a)x_1x_2]-[\frac{1}{2}a(x_0+x_0)^2+(b-a)x_0x_0]\\
&=(b-a)(x_1x_2-x_0^2)\\
&=(b-a)[x_1x_2-(\frac{x_1+x_2}{2})^2]\\
&=(b-a)\frac{(x_1-x_2)^2}{4}
\end{aligned}$$

因为矩阵G为实超对称半正定矩阵，故$a\geq b$, 从而可得$f(x)-f(\bar{x})\geq 0$. 这与$f(x)$为最优值矛盾，故最优解为$x_1=x_2=\cdots=x_n$.

4.设 $\sigma_1>\sigma_2>\cdots>\sigma_r>0,\ s<t$. 求述二次规划问题的最优解

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

5.设 d_k 为二次规划子问题 (8.4.3) 的最优解. 证明: 对任意 $\alpha\geq 0$, 有 $Q\left(x_k+\alpha d_k\right)\geq Q\left(x_k+d_k\right)$, 且 $Q\left(x_k+\alpha d_k\right)$ 关于 $\alpha\in[0,1]$ 严格单调递增.

7. 设矩阵$A\in\mathrm{R}^{n\times n}$. 试求下述优化问题的最优解

$$\begin{aligned}
&\max\ x^\top Ax\\
&\text{s.t.}\ \ \|x\|_2=\|y\|_2=1
\end{aligned}$$

解: 设$A = U\Sigma V^\top$为矩阵A的奇异值分解, 其中$\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \cdots, \sigma_n)$. 作线性变换$x = U^\top x, y = Vy$可将原问题化成

$$\begin{aligned} &\max \ \sum_{i=1}^n \sigma_i x_i y_i \\ &\text{s.t.} \ \ x^\top x = y^\top y = 1 \end{aligned}$$

由于根据Chaucy-Schwarz不等式,

$$\sum_{i=1}^n \sigma_i x_i y_i \le \sigma_1 \sum_{i=1}^n x_i y_i \le \sigma_1 \|x\|_2 \|y\|_2 = \sigma_1,$$

而在$x = y = e_1$时上式取等号, 故上述问题的最优解为$x = y = e_1$. 而原问题的最优解为(Ue_1, Ve_1).

8.考虑下述两规划问题

$$\begin{aligned} \min \ \ & \frac{1}{2}x^\top Gx + g^\top x, \\ \text{s.t.} \ \ & Ax = b \\ & x \ge 0, \end{aligned} \qquad \begin{aligned} \min \ \ & g^\top x - b^\top u \\ \text{s.t.} \ \ & Ax = b \\ & Gx + A^\top u - v = -g, \\ & x \ge 0, v \ge 0, v^\top x = 0. \end{aligned}$$

试证明由最后一问题得到的解 $\bar{x}$ 是前一问题的所有 K-T 点中目标函数值最小的点. 试问该点是第一个问题的最优值点吗?

证明 前一问题的 Lagrange 函数

$$L(x, u, v) = \frac{1}{2}x^\top Gx + g^\top x + u(Ax - b) - v^\top x.$$

其 KKT 条件为

$$\begin{cases} \nabla_x L(x, u, v) = Gx + g + A^\top u - v = 0, \\ Ax = b, x \ge 0, v \ge 0, v^\top x = 0. \end{cases}$$

可以看出, 其 KKT 条件就是第二个问题的约束条件, 从而后一问题得到的解 $\bar{x}$ 是前一问题的 K-T 点. 另外,由后一问题约束条件

$$\begin{aligned} & b^\top u = u^\top Ax, x^\top Gx + x^\top A^\top u - x^\top v = -x^\top g, \\ \Rightarrow \ & b^\top u = -x^\top Gx + -g^\top x, \\ \Rightarrow \ & g^\top x - b^\top u = x^\top Gx + 2g^\top x = 2\left(\frac{1}{2}x^\top Gx + g^\top x\right), \end{aligned}$$

所以后一问题得到的解 $\bar{x}$ 是前一问题的所有 K-T 点中使得 $g^\top x - b^\top u = 2\left(\frac{1}{2}x^\top Gx + g^\top x\right)$ 最小的点, 也即前一问题的所有 K-T 点中目标函数值最小的点.

显然, 由于 G 不一定半正定, 从而该问题不一定凸, 所以问题的 K-T 点不一定是最优值点. 证毕

9.多面胞

$$P=\left\{x \mid x=Ay, \mathrm{e}^{\top}y=1, y\geq 0\right\}$$

上到原点距离最近的点可以描述成下面的二次规划问题

$$\min\left\{\|Ay\|^2 \mid \mathrm{e}^{\top}y=1, y\geq 0\right\},$$

而该问题又可以通过下面的问题求解

$$\min\left\{\|Ay\|^2+\left(\mathrm{e}^{\top}y-1\right)^2 \mid y\geq 0\right\}.$$

试通过上述问题的对偶求解原问题.

证明 问题 (8) 的 Lagrange 函数

$$\begin{aligned}
L(x,\lambda) = &\ \|Ay\|^2+\left(\mathrm{e}^{\top}y-1\right)^2-\lambda^{\top}y,\\
= &\ y^{\top}A^{\top}Ay+y^{\top}\mathrm{e}\mathrm{e}^{\top}y-2\mathrm{e}^{\top}y-\lambda^{\top}y+1.
\end{aligned}$$

考虑 Lagrange 对偶函数

$$\theta(\lambda)=\inf\{L(y,\lambda) \mid y\in \mathrm{R}^n\},\quad \lambda\geq 0.$$

考虑

$$\begin{aligned}
\nabla_y L(y,\lambda)=2A^{\top}Ay+2\mathrm{e}\mathrm{e}^{\top}y-2\mathrm{e}-\lambda=0,\\
\Rightarrow \left(A^{\top}A+\mathrm{e}\mathrm{e}^{\top}\right)y=\mathrm{e}+\frac{\lambda}{2}.
\end{aligned}$$

从而

$$\begin{aligned}
\theta(\lambda) &= y^{\top}\left(A^{\top}A+\mathrm{e}\mathrm{e}^{\top}\right)y-2\mathrm{e}^{\top}y-\lambda^{\top}y+1,\\
&= y^{\top}\left(\mathrm{e}+\frac{\lambda}{2}\right)-2\mathrm{e}^{\top}y-\lambda^{\top}y+1,\\
&= -\left(\mathrm{e}+\frac{\lambda}{2}\right)^{\top}y+1,\\
&= -\left(\mathrm{e}+\frac{\lambda}{2}\right)^{\top}\left(A^{\top}A+\mathrm{e}\mathrm{e}^{\top}\right)^{-1}\left(\mathrm{e}+\frac{\lambda}{2}\right)+1.
\end{aligned}$$

此处, 为方便起见, 令 $A^{\top}A+\mathrm{e}\mathrm{e}^{\top}$ 可逆. 所以问题 (8) 的 Lagrange 对偶为

$$\begin{aligned}
&\max\ -\frac{1}{4}(2\mathrm{e}+\lambda)^{\top}\left(A^{\top}A+\mathrm{e}\mathrm{e}^{\top}\right)^{-1}(2\mathrm{e}+\lambda)+1,\\
&\text{s.t.}\ \ \lambda\geq 0.
\end{aligned}$$

记 $B:=(A^{\top}A+\mathrm{e}\mathrm{e}^{\top})^{-1}\succ 0$, 问题 (**??**) 等价于

$$\begin{aligned}
&\min\ \ (2\mathrm{e}+\lambda)^{\top}B(2\mathrm{e}+\lambda),\\
&\text{s.t.}\ \ \lambda\geq 0.
\end{aligned}$$

显然, 问题 (8) 的 最优解为 $\lambda^* = 0$, 所以原问题的最优解为

$$y^* = \left(A^\top A + ee^\top\right)^{-1}\left(e + \frac{\lambda^*}{2}\right) = \left(A^\top A + ee^\top\right)^{-1} e.$$

证毕

10. 用积极集方法求解下述二次规划问题

$$\begin{aligned} \min\ & -x_1^2 - x_2^2 - x_3^2 + 4x_1 + 6x_2, \\ \text{s.t. }\ & x_1 + x_2 \le 2, \\ & 2x_1 + 3x_2 \le 12, \\ & x_1, x_2, x_3 \ge 0. \end{aligned}$$

解: 具体过程略, 该问题无下界, 例如取 $x_1 = x_2 = 0,\ x_3 = +\infty$. 证毕

11.证明:在积极集方法中, 若 $\mathcal{A}(x_0)$ 对应的约束向量线性无关, 那么根据算法得到的 $\mathcal{A}(x_1)$ 对应的约束向量也线性无关. 进而通过数学归纳法得到积极集方法产生的集合 S_k 中的向量均线性无关.

第9章 约束优化的可行方法

1. 计算下述线性约束优化问题在点$\bar{x}=(1;1;0)$处的一个可行下降方向

$$
\begin{aligned}
\min\ & x_1^2+x_1x_2+2x_2^2-6x_1-2x_2-12x_3 \\
\text{s.t.}\ & x_1+x_2+x_3=2 \\
& x_1-2x_2\geqslant -3 \\
& x_1,x_2,x_3\geqslant 0
\end{aligned}
$$

2. 设$A\in \mathrm{R}^{m\times n}, b\in \mathrm{R}^m$. 证明下述线性方程组关于$u$的解是向量$x_0\in \mathrm{R}^n$到$\Omega=\{x\in \mathrm{R}^n \mid Ax=b\}$上的投影:

$$
\begin{pmatrix} I & A^\top \\ A & 0 \end{pmatrix}\begin{pmatrix} u \\ v \end{pmatrix}=\begin{pmatrix} x_0 \\ b \end{pmatrix}.
$$

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

4. 试分别通过子问题(9.1.2)和(9.2.1)计算下述约束优化问题在可行点$\bar{x}$的可行下降方向:

$$
\min\ \ f(x) \qquad \text{s.t. } a\leqslant x\leqslant b
$$

5. 设$\bar{x}$为下述连续可微约束优化问题

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

的可行点, 并设$(\bar{z},d)$为下述线性规划问题的最优解

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

证明: 如果$z=0$, 则$\bar{x}$为原规划问题的FJ点.

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

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

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

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

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

9. 设$a_1, a_2, \cdots, a_m \in \mathrm{R}^n$, Ω为R^n中的闭凸集. 则优化问题

$$\min \sum_{i=1}^{m} \|x - a_i\|^2 \qquad \text{s.t. } x \in \Omega$$

的最优解等价于向量$\dfrac{1}{m}\sum\limits_{i=1}^{m} a_i$到$\Omega$上的投影.

10. 设n阶矩阵A对称正定, $\lambda > 0, b \in \mathrm{R}$. 则$x \in \mathrm{R}^n$是约束优化问题

$$\begin{aligned} \min_{x\in\mathrm{R}^n} \quad & \frac{1}{2}x^\top Ax \\ \text{s.t.} \quad & -\lambda \le Ax - b \le \lambda \end{aligned}$$

的最优解当且仅当

$$P_\Omega(Ax - b - x) = Ax - b.$$

其中, $\Omega = \{x \in \mathrm{R} \mid |x| \le \lambda\}$.

证明: 记$z = Ax - b$. 则$x = A^{-1}(z + b)$. 从而原规划问题可化为

$$\begin{aligned} \min_{z\in\mathrm{R}^n} \quad & \frac{1}{2}(z+b)^\top A^{-1}(z+b) \\ \text{s.t.} \quad & -\lambda \le z \le \lambda \end{aligned}$$

利用约束优化问题的最优性条件, $z \in \mathrm{R}^n$为上述问题的最优解当且仅当

$$P_\Omega(z - A^{-1}(z + b)) = z.$$

利用$z = Ax - b$得

$$P_\Omega(Ax - b - x) = Ax - b.$$

11. 设$a \in \mathrm{R}^n$, $b_1 \ne b_2 \in \mathrm{R}$. 试求点$x_0 \in \mathrm{R}^n$到超平面$a^\top x = b_1$的距离，超平面$a^\top x = b_1$和$a^\top x = b_2$之间的距离.

解: 不失一般性，设$a \ne 0$. 令$\bar{a} = a/\|a\|, \bar{b}_i = b_i/\|a\|$. 则$\bar{a}^\top\bar{a} = 1$, 也就是$\bar{a}^+ = \bar{a}^\top$, 且超平面$a^\top x = b_i, i = 1, 2$可写成$\bar{a}^\top x = \bar{b}_i$.

由§9.3最后的讨论, R^n中任一点x到超平面$\bar{a}^\top x = \bar{b}_i$上的投影分别为

$$(I - \bar{a}^+\bar{a})x + \bar{a}^+\bar{b}_i.$$

两投影点之间的距离即为两超平面之间的距离, 即

$$\|\bar{a}^+\bar{b}_1 - \bar{a}^+\bar{b}_2\| = \|\bar{a}^+\||\bar{b}_1 - \bar{b}_2| = \|\bar{a}^\top\||\frac{b_1 - b_2}{\|a\|}| = \frac{|b_1 - b_2|}{\|a\|}.$$

12. 若$\Omega = \{x \mid a^\top x = b\}$为一超平面, 则由§1.5的结论, $a^+ = \frac{a^\top}{\|a\|^2}$. 从而$y \in \mathrm{R}^n$到$\Omega$上的投影为

$$\big(I - (a^\top)^+ a^\top\big)y + (a^\top)^+ b = \Big(I - \frac{aa^\top}{\|a\|^2}\Big)y + \frac{ab}{\|a\|^2}.$$

进而, $y \in \mathrm{R}^n$到超平面$\Omega = \{x \mid a^\top x = b\}$的距离为

$$\Big\|\frac{aa^\top y}{\|a\|^2} - \frac{ab}{\|a\|^2}\Big\| = \frac{|a^\top y - b|}{\|a\|}.$$

第10章 约束优化的罚函数方法

1. 对于不等式约束优化问题$\min\{f(x) \mid c_i(x) \leqslant 0, i \in \mathcal{I}\}$, 试分析下述函数在通过外点罚函数方法求上述优化问题的最小值点时的优势和劣势

$$P_\mu(x) = f(x) + \mu \sum_{i \in \mathcal{I}} \max\{0, c_i(x)\}$$

$$P_\mu(x) = f(x) + \mu \sum_{i \in \mathcal{I}} \max^2\{0, c_i(x)\}$$

$$P_\mu(x) = f(x) + \mu \max\{0, c_1(x), c_2(x), \cdots, c_{|\mathcal{I}|}(x)\}$$

$$P_\mu(x) = f(x) + \mu \max^2\{0, c_1(x), c_2(x), \cdots, c_{|\mathcal{I}|}(x)\}$$

2. 以$(2,6)^\top$为初始点, 用外点罚函数(罚因子$\mu = 10$)求下述规划问题的近似最小值点

$$\begin{aligned} \min \ & x_1^2 + x_2^2 \\ \text{s.t.} \ & 2x_1 + x_2 - 2 \leqslant 0 \\ & -x_2 + 1 \leqslant 0 \end{aligned}$$

3. 根据定理10.2.1, 优化问题(10.1.2)可以转化为

$$\sup_{\mu \geqslant 0} \inf_{x \in \mathrm{R}^n} P(x, \mu),$$

其中, $P(x,\mu)$为外点罚函数. 证明优化问题(10.1.2)等价于

$$\inf_{x \in \mathrm{R}^n} \sup_{\mu \geqslant 0} P(x, \mu).$$

4. 对于约束优化问题(10.1.2), 考虑(10.2.5)定义的混合罚函数, 试在适当条件下建立如下结论

$$\inf\{f(x) \mid c_i(x) \leqslant 0, i \in \mathcal{I}; c_i(x) = 0, i \in \mathcal{E}\} = \lim_{\mu \to \infty} P(x^\mu, \mu),$$

$$\lim_{\mu \to \infty} \mu B(x^\mu) = 0, \quad \lim_{\mu \to \infty} \frac{1}{\mu} \theta_c(x^\mu) \to 0.$$

第11章 序列二次规划方法

1. 以$x=(0,0)^\top$为初始点, 用Lagrange-Newton算法求解如下优化问题

$$\min\ x_1+x_2 \qquad \text{s.t. } x_2 \geqslant x_1^2$$

并考虑为什么在$\lambda=0$时算法会失败.

2. 设$H\in \mathrm{R}^{n\times n}$对称正定, 则$(x^*,\lambda^*)$为下述优化问题的K-T对

$$\min\ f(x) \qquad \text{s.t. } c_i(x)\leqslant 0, \quad i=1,2,\cdots,r$$

当且仅当$(0,\lambda^*)$为下述严格凸二次规划问题的K-T对

$$\begin{aligned}\min\ & \nabla f(x)^\top d+\frac{1}{2}d^\top H d\\ \text{s.t. }\ & c_i(x)+\nabla c_i(x)^\top d\leqslant 0, \quad i=1,2,\cdots,r\end{aligned}$$