无约束优化习题解答

2021 年 5 月 23 日

Exercise 1　写出基于Wolfe-Powell准则的非精确一维搜索算法中插值多项式$p^{(1)}(t), p^{(2)}(t)$的具体表达式。

$p(t) = At^2 + Bt + C$

由

$$\begin{cases} p^{(1)}(a_1) = \varphi_1 \\ p^{(1)\prime}(a_1) = \varphi_1' \\ p^{(1)}(\alpha) = \varphi \end{cases}$$

可设

$$p^{(1)}(t) = \varphi_1 + \varphi_1'(t - a_1) + A(t - a_1)^2$$

向上式代入$p^{(1)}(\alpha) = \varphi$得

$$\varphi_1 + \varphi_1'(\alpha - a_1) + A(\alpha - a_1)^2 = \varphi \;\Rightarrow\; A = \frac{\varphi - \varphi_1 + \varphi_1'(a_1 - \alpha)}{(\alpha - a_1)^2}$$

由

$$\begin{cases} p^{(2)\prime}(a_1) = \varphi_1' \\ p^{(2)}(\alpha) = \varphi \\ p^{(2)\prime}(\alpha) = \varphi' \end{cases}$$

可设

$$p^{(2)}(t) = \varphi + \varphi'(t - \alpha) + C(t - \alpha)^2$$

于是

$$p^{(2)\prime}(a_1) = \varphi' + 2C(a_1 - \alpha) = \varphi_1' \;\Rightarrow\; C = \frac{\varphi_1' - \varphi'}{2(a_1 - \alpha)}$$

Excercise 2　证明基于Goldstein准则的非精确一维搜索算法的全局收敛性。　Goldstein准则：

$$\varphi(\alpha) \leqslant \varphi(0) + \rho\alpha\varphi'(0) \tag{1}$$

$$\varphi(\alpha) \geqslant \varphi(0) + (1 - \rho)\alpha\varphi'(0) \tag{2}$$

$\varphi'(0) = \nabla f(x^{(k)})^T d$

设$\forall k$，$g^{(k)} = \nabla f(x^{(k)}) \neq 0$和$f(x^{(k)})$有下界，则$f(x^{(k)}) - f(x^{(k+1)}) \to 0$，由(1)，$-g^{(k)^\top} s^{(k)} \to 0$。

（反证）若$g^{(k)} \to 0$不成立，则$\exists \varepsilon > 0$和子列$\left\{x^{(k)}\right\}_{k \in K}$使得$\left\|g^{(k)}\right\| \geqslant \varepsilon$，从而由

$$-g^{(k)^\top} s^{(k)} = \left\|g^{(k)}\right\| \left\|s^{(k)}\right\| \cos\theta_k \geqslant \varepsilon \left\|s^{(k)}\right\| \sin\mu$$

以及对$\forall k$, $\theta_k \leqslant \frac{\pi}{2} - \mu$，得$\left\|s^{(k)}\right\| \to 0$。
又由

$$f(x^{(k)} + s^{(k)}) = f(x^{(k)}) + \nabla f(x^{(k)})^\top s^{(k)} + o\left(\left\|s^{(k)}\right\|\right)$$

得

$$\lim_{k \to \infty} \frac{f(x^{(k)}) - f(x^{(k+1)})}{-\nabla f(x^{(k)})^\top s^{(k)}} = 1$$

由式(2)得

$$\frac{f(x^{(k)}) - f(x^{(k+1)})}{-\nabla f(x^{(k)})^\top s^{(k)}} \leqslant 1 - \rho < 1$$

矛盾！所以$g^{(k)} = \nabla f(x^{(k)}) \to 0, k \to \infty$。

Exercise 3　将非线性方程组求根$F(x) = 0$的牛顿迭代，用于求最优化问题$\min_{x \in \mathbb{R}^n} f(x)$，给出相应的迭代格式并说明理由。 牛顿迭代的原理是取$F(x)$在$x^{(k)}$处的一阶Taylor展开做近似并令其为0

$$F(x^{(k)}) + J_F(x^{(k)})(x - x^{(k)}) = 0$$

$F(x^{(k)}) + J_F \Delta X = 0$

解的迭代格式为

$$x^{(k+1)} = x^{(k)} - \left(J_F(x^{(k)})\right)^{-1} F(x^{(k)})$$

对无约束问题，求解$\nabla f(x) = 0$，即有迭代格式

$$x^{(k+1)} = x^{(k)} - \left(\nabla^2 f(x^{(k)})\right)^{-1} \nabla f(x^{(k)})$$

Exercise 4　证明对称秩一拟牛顿法具有遗传性和二次终止性。 对于二次函数$f(x) = \frac{1}{2} x^\top G x + c^\top x$，

$$\nabla f(x) = Gx + c,\ \nabla^2 f(x) = G,$$

拟牛顿法中$\boxed{y^{(k)} = g^{(k+1)} - g^{(k)} = G_k s^{(k)}}$，正割条件为$H_{k+1} y^{(k)} = s^{(k)}$。

遗传性 使用归纳法。

1. $k = 1$时，由正割条件，直接成立。

2. 假设遗传性对于k成立，即$H_k y^{(l)} = s^{(l)}, l = 0, 1, \ldots, k-1$。对于$k+1$和$l = 0, 1, \ldots, k-1$，由对称秩一校正公式

$$H_{k+1} y^{(l)} = H_k y^{(l)} + \frac{(s^{(k)} - H_k y^{(k)})(s^{(k)} - H_k y^{(k)})^\top y^{(l)}}{(s^{(k)} - H_k y^{(k)})^\top y^{(k)}}$$

其中

$$\begin{aligned}(s^{(k)} - H_k y^{(k)})^\top y^{(l)} =& {s^{(k)}}^\top y^{(l)} - {y^{(k)}}^\top H_k y^{(l)} \\ =& {s^{(k)}}^\top y^{(l)} - {s^{(l)}}^\top y^{(k)} \\ =& {s^{(k)}}^\top G s^{(l)} - {s^{(l)}}^\top G s^{(k)} \\ =& 0\end{aligned}$$

二次终止性 （这里需要假设$s_0, \ldots, s_{n-1}$线性无关）由于

$$H_n y^{(l)} = \boxed{H_n G} s^{(l)} = s^{(l)},\ l = 0, 1, \ldots, n-1$$

而$s_0, \ldots, s_{n-1}$是$\mathbb{R}^n$的一组基，因此 $H_n G = I \Rightarrow H_n = G^{-1}$。而对二次函数，$x - G^{-1}\nabla f(x) = x - G^{-1}(Gx + c) = -G^{-1}c$直接得到全局最优解，因此

$$x^{(n+1)} = x^{(n)} - H_n g^{(k)}$$

为全局最优解。

Exercise 5 利用秩一校正的求逆公式（Sherman-Morrison定理），由$H_{k+1}^{(DFP)}$推导$B_{k+1}^{(DFP)}$。

$$(A + uv^\top)^{-1} = A^{-1} - \frac{A^{-1}uv^\top A^{-1}}{1 + v^\top A^{-1}u}$$

$H \Longleftrightarrow B$

$$H_{k+1}^{(DFP)} = H_k + \frac{s^{(k)}{s^{(k)}}^\top}{{s^{(k)}}^\top y^{(k)}} - \frac{H_k y^{(k)} {y^{(k)}}^\top H_k}{{y^{(k)}}^\top H_k y^{(k)}}$$

$y \Longleftrightarrow s$

为方便书写，忽视所有角标k。记$M = H + \frac{ss^\top}{s^\top y}$，则有

$$M^{-1} = H^{-1} - \frac{H^{-1}ss^\top H^{-1}}{s^\top y + s^\top H^{-1}s} = B - \frac{Bss^\top B}{s^\top y + s^\top Bs} \tag{3}$$

$$B_{k+1}^{(DFP)} = \left(H_{k+1}^{(DFP)}\right)^{-1} = M^{-1} + \frac{M^{-1}Hyy^\top HM^{-1}}{y^\top Hy - y^\top HM^{-1}Hy} \tag{4}$$

$$B^{(DFP)}_{k+1} = \left(M - \frac{\overset{u}{Hy}\,\overset{v^T}{y^T H}}{y^T H y}\right)^{-1}$$

将(3)代入(4)的RHS第二项并展开，得

$$\frac{M^{-1}Hyy^\top HM^{-1}}{y^\top Hy-y^\top HM^{-1}Hy}=\frac{\left(B-\frac{Bss^\top B}{s^\top y+s^\top Bs}\right)Hyy^\top H\left(B-\frac{Bss^\top B}{s^\top y+s^\top Bs}\right)}{y^\top Hy-y^\top H\left(B-\frac{Bss^\top B}{s^\top y+s^\top Bs}\right)Hy}$$

利用$BH=I$，

$$y^\top Hy-y^\top HM^{-1}Hy=y^\top Hy-y^\top HBHy+\frac{y^\top HBss^\top BHy}{s^\top y+s^\top Bs}=\frac{\left(y^\top s\right)^2}{s^\top y+s^\top Bs}$$

$$\begin{aligned}\frac{M^{-1}Hyy^\top HM^{-1}}{y^\top Hy-y^\top HM^{-1}Hy}&=\frac{yy^\top(s^\top y+s^\top Bs)}{(y^\top s)^2}-\frac{ys^\top B}{y^\top s}-\frac{Bsy^\top}{y^\top s}+\frac{Bss^\top B}{s^\top y+s^\top Bs}\\&=\left(1+\frac{s^\top Bs}{y^\top s}\right)\frac{yy^\top}{y^\top s}-\frac{ys^\top B+Bsy^\top}{y^\top s}+\frac{Bss^\top B}{s^\top y+s^\top Bs}\end{aligned}$$

将上式和式(3)代入(4)即有

$$B_{k+1}^{(DFP)}=B_k+\left(1+\frac{s^{(k)^\top}B_ks^{(k)}}{y^{(k)^\top}s^{(k)}}\right)\frac{y^{(k)}y^{(k)^\top}}{y^{(k)^\top}s^{(k)}}-\frac{B_ks^{(k)}y^{(k)^\top}+y^{(k)}s^{(k)^\top}B_k}{y^{(k)^\top}s^{(k)}}$$

Exercise 6 共轭梯度法性质定理：设目标函数$f(x)=\frac{1}{2}x^\top Gx+c^\top x$，则采用精确一维搜索的共轭梯度法经$m\leqslant n$步迭代后终止，且对所有的$1\leqslant k\leqslant m$成立下列关系：

$$d^{(k)^\top}Gd^{(j)}=0,\ j=0,1,\dots,k-1 \tag{6.1}$$

$$g^{(k)^\top}g^{(j)}=0,\ j=0,1,\dots,k-1 \tag{6.2}$$

$$d^{(k)^\top}g^{(k)}=-g^{(k)^\top}g^{(k)} \tag{6.3}$$

$$\operatorname{span}\{g^{(0)},g^{(1)},\dots,g^{(k)}\}=\operatorname{span}\{g^{(0)},Gg^{(0)},\dots,G^kg^{(0)}\} \tag{6.4}$$

$$\operatorname{span}\{d^{(0)},d^{(1)},\dots,d^{(k)}\}=\operatorname{span}\{g^{(0)},Gg^{(0)},\dots,G^kg^{(0)}\} \tag{6.5}$$

共轭梯度法步骤中得到的条件：

二次

$$g^{(k+1)^\top}d^{(k)}=0 \quad \min_\alpha f(x^{(k)}+\alpha d^{(k)}) \quad \text{(精确一维搜索)} \tag{6.6}$$

$$\alpha_k=-\frac{d^{(k)^\top}g^{(k)}}{d^{(k)^\top}Gd^{(k)}} \quad \nabla_\alpha f(\cdots)=0 \quad \text{(精确一维搜索)} \tag{6.7}$$

$$g^{(k+1)}=g^{(k)}+\alpha_kGd^{(k)} \quad \text{(直接展开}\nabla f(x^{(k+1)})\text{)} \tag{6.8}$$

$$\beta_k=\frac{g^{(k+1)^\top}g^{(k+1)}}{g^{(k)^\top}g^{(k)}} \tag{6.9}$$

$$d^{(k+1)}=-g^{(k+1)}+\beta_kd^{(k)} \tag{6.10}$$

$$\begin{aligned}g^{(k+1)}=\nabla f(x^{(k+1)})&=\nabla f(x^{(k)}+\alpha_kd^{(k)})\\&=G(x^{(k)}+\alpha_kd^{(k)})+c\\&=\underbrace{Gx^{(k)}+c}_{g^{(k)}}+\alpha_kGd^{(k)}\end{aligned}$$

证明. (6.3)**的证明**

$$
\begin{aligned}
d^{(k)^\top} g^{(k)} &= - g^{(k)^\top} g^{(k)} + \beta_k d^{(k-1)^\top} g^{(k)} && \text{(by (6.10))} \\
&= - g^{(k)^\top} g^{(k)} + 0 && \text{(by (6.6))} \\
&= - g^{(k)^\top} g^{(k)}
\end{aligned}
$$

(6.1)**与**(6.2)**的证明** k=1时直接验证可得结论成立若(6.1)与(6.2)对k成立，则对于$k+1$

1) 对(6.8)式左右两边转置后右乘$g^{(j)}$得

$$
\begin{aligned}
g^{(k+1)} g^{(j)} =& g^{(k)^\top} g^{(j)} + \alpha_k d^{(k)^\top} G g^{(j)} \\
=& g^{(k)^\top} g^{(j)} + \alpha_k d^{(k)^\top} G (d^{(j)} - \beta_{j-1} d^{(j-1)}) \\
=& g^{(k)^\top} g^{(j)} + \boxed{\alpha_k d^{(k)^\top} G d^{(j)}}
\end{aligned}
$$

(6,1) j=k, j<k

$j = k$时，将(6.7)代入即可得上式为0；$j < k$时，由归纳假设得上式为0。综上，(6.2)成立。

2) 由(6.10)式，

$$
\begin{aligned}
d^{(k+1)^\top} G =& (-g^{(k+1)} + \beta_k d^{(k)}) G d^{(j)} \\
=& - g^{(k+1)} G d^{(j)} + \beta_k d^{(k)} G d^{(j)} \\
=& g^{(k+1)^\top} (g^{(j)} - g^{(j+1)}) / \alpha_k + \beta_k d^{(k)} G d^{(j)}
\end{aligned}
$$

$j = k$时，由(6.2)(6.7)(6.9)得上式为0；$j < k$时，由归纳假设得上式为0。综上，(6.1)成立。

(6.4)**与**(6.5)**的证明** 由(6.10)式知，存在可逆方阵

$$
Q = \begin{bmatrix}
-1 & \beta_0 & 0 & \dots & 0 & 0 \\
0 & -1 & \beta_1 & \dots & 0 & 0 \\
0 & 0 & -1 & \dots & 0 & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
0 & 0 & 0 & \dots & -1 & \beta_{k-1} \\
0 & 0 & 0 & \dots & 0 & -1
\end{bmatrix}
$$

使得$(d^{(0)}, d^{(1)}, \dots, d^{(k)}) Q = (g^{(0)}, g^{(1)}, \dots, g^{(k)})$，所以

$$
\mathrm{span}\{d^{(0)}, d^{(1)}, \dots, d^{(k)}\} = \mathrm{span}\{g^{(0)}, g^{(1)}, \dots, g^{(k)}\}
$$

下面只需要使用数学归纳法证明(6.4)。

$d^{(k)}, g^{(k)} \in \text{span}\{g^{(0)}, \dots, G^k g^{(0)}\}$

1) $k=0$时，由定义直接得到结论成立。

2) 假设结论对k成立。对于$k+1$，由(6.8)式和归纳假设，

$$g^{(k+1)} = g^{(k)} + \alpha_k G d^{(k)} \in \text{span}\{g^{(0)}, Gg^{(0)}, \dots, G^{\boxed{k+1}} g^{(0)}\}$$

即得

$$\text{span}\{g^{(0)}, g^{(1)}, \dots, g^{(k+1)}\} \subseteq \text{span}\{g^{(0)}, Gg^{(0)}, \dots, G^{k+1} g^{(0)}\}$$

前面已证$g^{(l)} \perp g^{(j)}$, $0 \leqslant j < l \leqslant k+1$，因此 $\{g^{(0)}, g^{(1)}, \dots, g^{(k+1)}\}$线性无关，

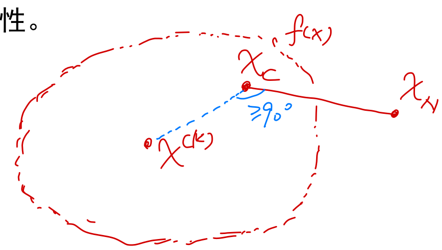
$$\dim\left(\text{span}\{g^{(0)}, g^{(1)}, \dots, g^{(k+1)}\}\right) = k+2$$

因此结论对$k+1$成立，即

$$\text{span}\{g^{(0)}, g^{(1)}, \dots, g^{(k+1)}\} = \text{span}\{g^{(0)}, Gg^{(0)}, \dots, G^{k+1} g^{(0)}\}.$$

□

Exercise 7　证明折线法（信赖域方法）子问题模型的函数单调性。



$$
\begin{aligned}
s_C^{(k)} &= -\frac{\left\|g^{(k)}\right\|^2}{g^{(k)\top} B_k g^{(k)}} g^{(k)} \\
s_N^{(k)} &= -B_k^{-1} g^{(k)} \\
x^{(k+1)} &= x^{(k)} + s_C^{(k)} + \lambda(s_N^{(k)} - s_C^{(k)})
\end{aligned}
$$

(1) 证明沿着Cauchy点$x_C^{(k+1)}$和牛顿点$x_N^{(k+1)}$的连线，到$x^{(k)}$的距离单调增加。

只要证明$\left(x_C^{(k+1)} - x^{(k)}\right)^\top \left(x_C^{(k+1)} - x_N^{(k+1)}\right) \leqslant 0$，即证明

$$\left(-\frac{g^{(k)\top} g^{(k)}}{g^{(k)\top} B_k g^{(k)}} g^{(k)}\right)^\top \left(B_k^{-1} g^{(k)} - \frac{g^{(k)\top} g^{(k)}}{g^{(k)\top} B_k g^{(k)}} g^{(k)}\right) \leqslant 0$$

只要证明

$$g^{(k)\top} B_k^{-1} g^{(k)} - \frac{g^{(k)\top} g^{(k)}}{g^{(k)\top} B_k g^{(k)}} g^{(k)\top} g^{(k)} \geqslant 0$$

$$\iff \quad g^{(k)\top} B_k^{-1} g^{(k)} g^{(k)\top} B_k g^{(k)} - \left(g^{(k)\top} g^{(k)}\right)^2 \geqslant 0 \tag{$*$}$$

即可。

记$S_k = \sqrt{B_k}$, $a^{(k)} = S_k g^{(k)}$, $b^{(k)} = S_k^{-1} g^{(k)}$，则由Cauchy不等式

$$\left(a^{(k)\top} b^{(k)}\right)^2 \leqslant \left(a^{(k)\top} a^{(k)}\right)\left(b^{(k)\top} b^{(k)}\right)$$

式($*$)得证。

(2)

$\lambda\ 0 \to 1$

$$
\begin{aligned}
h(\lambda) =& q^{(k)}\left(s_C + \lambda(s_N - s_C)\right) \\
=& \underline{f(x^{(k)}) + g^\top\left(s_C + \lambda(s_N - s_C)\right) + \frac{1}{2}\left(s_C + \lambda(s_N - s_C)\right)^\top B_k\left(s_C + \lambda(s_N - s_C)\right)} \\
h'(\lambda) =& \underline{g^\top(s_N - s_C) + s_C^\top B_k(s_N - s_C) + \boxed{\lambda}(s_N - s_C)^\top B_k(s_N - s_C)} \\
\leqslant& g^\top(s_N - s_C) + s_C^\top B_k(s_N - s_C) + (s_N - s_C)^\top B_k(s_N - s_C) \\
=& \left(g^\top + s_C^\top B_k + (s_N - s_C)^\top B_k\right)(s_N - s_C) \\
=& \underline{(g^\top + s_N^\top B_k)(s_N - s_C)} \\
=& 0 \cdot (s_N - s_C) \\
=& 0
\end{aligned}
$$

约束优化习题讲义

2021 年 5 月 24 日

1 二次规划

Exercise 1 设 x^* 是一般的二次规划问题(122)的局部极小点，则 x^* 也必是等式约束问题

$$(EQ)\begin{cases} \min & Q(x)=\frac{1}{2}x^\top Gx+c^\top x \\ \text{s.t.} & a_i^\top x=b_i, `i\in\mathcal{E}\cup\mathcal{I}(x^*) \end{cases}$$

的局部极小点。反之，如果 x^* 是一般问题(122)的可行点，同时是 (EQ) 的 K-T 点，且相应的 Lagrange 乘子 λ^* 满足 $\lambda_i^*\geqslant 0$，$i\in\mathcal{I}(x^*)$，则 x^* 必是原问题(122)的 K-T 点。

一般的二次规划为

$$\begin{aligned} &\min\ Q(x)=\frac{1}{2}x^\top Gx+c^\top x \\ &\text{ s.t. } a_i^\top x=b_i,\ i\in\mathcal{E}=\{1,\ldots,m_e\} \\ &\qquad a_i^\top x\geqslant b_i,\ i\in\mathcal{I}=\{m_e+1,\ldots,m\} \end{aligned} \tag{122}$$

证明. x^* 是(122)的局部极小点，即存在 x^* 的邻域 U，使得

$$\begin{aligned} x^* &=\arg\min_x\left\{Q(x)\ :\ \begin{array}{l} a_i^\top x=b_i, i\in\mathcal{E} \\ a_i^\top x\geqslant b_i, i\in\mathcal{I} \\ x\in U \end{array}\right\}. \\ &=\arg\min_x\left\{Q(x)\ :\ \begin{array}{l} a_i^\top x=b_i, i\in\mathcal{E}\cup\mathcal{I}(x^*) \\ a_i^\top x>b_i, i\in\mathcal{I}\backslash\mathcal{I}(x^*) \\ x\in U \end{array}\right\}. \end{aligned}$$

约束

邻域

$\{x \mid a_i^\top x\geq b_i\}$ $\supseteq$ $\{x \mid a_i^\top x=b_i\}$ $\cup$ $\{x \mid a_i^\top x> b_i\}$

由连续性，存在 x^* 的邻域 V 使得对 $\forall x\in V, i\in\mathcal{I}\backslash\mathcal{I}(x^*)$，$a_i^\top x>b_i$，有

$$x^* = \arg\min_x \left\{ Q(x) \ : \ \begin{array}{l} a_i^\top x = b_i, i\in\mathcal{E}\cup\mathcal{I}(x^*) \\ a_i^\top x > b_i, i\in\mathcal{I}\backslash\mathcal{I}(x^*) \\ x\in U\cap \boxed{V} \end{array} \right\}$$

$$= \arg\min_x \left\{ Q(x) \ : \ \begin{array}{l} a_i^\top x = b_i, i\in\mathcal{E}\cup\mathcal{I}(x^*) \\ x\in U\cap V \end{array} \right\}$$

自动满足

即 x^* 为 (EQ) 的局部极小点。

反之，x^* 是 (EQ) 的 K-T 点，即

$$\nabla Q(x^*) + \sum_{i\in\mathcal{E}\cup\mathcal{I}(x^*)} \lambda_i^* a_i^\top = 0$$

且 $\lambda_i^*\geqslant 0, i\in\mathcal{I}(x^*)$，则令

$$\bar\lambda_i = \begin{cases} \lambda_i^* & i\in\mathcal{E}\cup\mathcal{I}(x^*) \\ 0 & i\in\mathcal{I}\backslash\mathcal{I}(x^*) \end{cases}$$

加上 x^* 为(122)的可行点，即有(122)的 K-T 条件:

$$\begin{cases} \nabla Q(x^*) - \sum_{\mathcal{E}\cup\mathcal{I}} \bar\lambda_i a_i = 0 \\ a_i x^* = 0, \ i\in\mathcal{E} \\ a_i x^* \geqslant 0, \ i\in\mathcal{I} \\ \bar\lambda_i\geqslant 0, \ i\in\mathcal{I} \\ \bar\lambda_i(a_i^\top x - b_i) = 0, \ i\in\mathcal{I} \end{cases}$$

□

Excercise 2 考虑等式约束问题

$$(EQ1)\begin{cases} \min & \frac12 s^\top G s + (Gx^{(k)}+c)^\top s \\ \text{s.t.} & a_i^\top s = 0, \ i\in\mathcal{E}_k \end{cases}$$

求得其解为 $s^{(k)}$，及其相应的 Lagrange 乘子 $\lambda_i^{(k)}$，$i\in\mathcal{E}_k$。
若 $s^{(k)}=0$，且 $\lambda_i^{(k)}\geqslant 0$, $i\in\mathcal{E}_k$ 不成立，则由 $\lambda_{i_q}^{(k)}=\min_{i\in\mathcal{I}(x^{(k)})}\lambda_i^{(k)}<0$ 确定 i_q，那么如下问题

$$(EQ3)\begin{cases} \min & \frac12 s^\top G s + (Gx^{(k)}+c)^\top s \\ \text{s.t.} & a_i^\top s = 0, \ i\in\hat{\mathcal{E}}=\mathcal{E}_k\backslash\{i_q\}. \end{cases}$$

的解 $\hat s$ 是原问题在当前点 $x^{(k)}$ 处的可行方向，即 $a_{i_q}^\top\hat s\geqslant 0$.

$s^{(k)} \neq 0$

(3) $s^{(k)} = 0$

证明. (EQ1) 的 K-T 条件为

$$\begin{cases} Gs^{(k)} + Gx^{(k)} + c - \sum\limits_{i \in \mathcal{E}_k} \lambda_i^{(k)} a_i = 0 \\ a_i^\top s^{(k)} = 0, \ i \in \mathcal{E}_k \end{cases}$$

对第一式，由于 $s^{(k)} = 0$，等价于

$$Gx^{(k)} + c - \sum_{i \in \mathcal{E}_k} \lambda_i^{(k)} a_i = 0 \tag{1}$$

左乘上 $\hat{s}^\top$，得

$$\hat{s}^\top (Gx^{(k)} + c) - \sum_{i \in \mathcal{E}_k} \lambda_i^{(k)} \hat{s}^\top a_i = 0$$

$\mathcal{E}_k \setminus \{i_q\}$ (EQ3) $=0$

$\{i_q\}$

$$\Rightarrow \ \hat{s}^\top (Gx^{(k)} + c) = \lambda_{i_q}^{(k)} a_{i_q}^\top \hat{s}$$ (由 (EQ3) 的约~~束条件~~)

只要证明 $\hat{s}^\top (Gx^{(k)} + c) \leqslant 0$ 即可。

反证：若 $\hat{s}^\top (Gx^{(k)} + c) > 0$，则取 $\tilde{s} = -\hat{s}$。$\tilde{s}$ 满足 (EQ3) 的约束条件，且

$$\begin{aligned} & \frac{1}{2} \tilde{s}^\top G \tilde{s} + (Gx^{(k)} + c)^\top \tilde{s} \\ = & \frac{1}{2} \hat{s}^\top G \hat{s} - (Gx^{(k)} + c)^\top \hat{s} \\ < & \frac{1}{2} \hat{s}^\top G \hat{s} + (Gx^{(k)} + c)^\top \hat{s} \end{aligned}$$

与 $\hat{s}$ 为 (EQ3) 的解矛盾。证毕。 □

另一个证明. (EQ3) 的 K-T 条件为

$$\begin{cases} G\hat{s} + (Gx^{(k)} + c) - \sum\limits_{i \in \hat{\mathcal{E}}} \hat{\lambda}_i a_i = 0 \\ a_i^\top \hat{s} = 0, \ i \in \hat{\mathcal{E}} \end{cases} \tag{2}$$

(1)与(2)的第一式作差，得

$$G\hat{s} + \lambda_{i_q}^{(k)} a_{i_q} + \sum_{i \in \hat{\mathcal{E}}} (\lambda_i^{(k)} - \hat{\lambda}_i) a_i = 0$$

$$\Rightarrow \ \hat{s}^\top G \hat{s} + \lambda_{i_q}^{(k)} a_{i_q}^\top \hat{s} = 0$$

只要证明 $\hat{s}^\top G \hat{s} \geqslant 0$。

若 $\hat{s}^\top G \hat{s} < 0$，则 (EQ3) 无下界，与 $\hat{s}$ 为其解矛盾，故得证。 □

2 非线性约束最优化

Exercise 4 **证明 (125) 中定义的 $\psi(x,\lambda)$ 是关于 Lagrange-Newton 法的下降函数。**

证明.

$$
\begin{aligned}
\nabla\psi(x,\lambda) &= \begin{pmatrix} \nabla_x\psi(x,\lambda) \\ \nabla_\lambda\psi(x,\lambda) \end{pmatrix} \\
&= \begin{pmatrix} 2W(x,\lambda)(\nabla f(x) - A(x)^\top\lambda) + A(x)^\top c(x) \\ -2A(x)(\nabla f(x) - A(x)^\top\lambda) \end{pmatrix} \\
&= 2\begin{pmatrix} W(x,\lambda) & -A(x)^\top \\ -A(x) & \end{pmatrix}\begin{pmatrix} \nabla f(x) - A(x)^\top\lambda \\ -c(x) \end{pmatrix}
\end{aligned}
$$

由式 (124)，

$$
\begin{pmatrix} W(x,\lambda) & -A(x)^\top \\ -A(x) & \end{pmatrix}\begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} = -\begin{pmatrix} \nabla f(x) - A(x)^\top\lambda \\ -c(x) \end{pmatrix}
$$

有

$$
\begin{aligned}
&\nabla\psi(x,\lambda)^\top\begin{pmatrix} \delta_x \\ \delta_y \end{pmatrix} \leq 0 \\
=&-2\begin{pmatrix} \nabla f(x) - A(x)^\top\lambda \\ -c(x) \end{pmatrix}^\top\begin{pmatrix} \nabla f(x) - A(x)^\top\lambda \\ -c(x) \end{pmatrix} \\
=&-2\psi(x,\lambda). \leq 0
\end{aligned}
$$

✓

□

$\min \|C(x)\| < \varepsilon$

Excercise 5 **证明罚函数法求解带误差界近似问题的算法有限终止性。**

$\min f(x) + \rho\|C(x)\|^2$

证明. 反证。假设对所有的 σ_k 都有 $\|c(x(\sigma_k))_-\| \geq \varepsilon$。由题意，存在 $\bar{x}$ 满足 $\|c(\bar{x})_-\| < \varepsilon$。由 $x(\sigma_k)$ 的定义有

停止条件 $\|C(x)\| < \varepsilon$

$$f(\bar{x}) + \sigma_k\|c(\bar{x})_-\|^2 \geq f(x(\sigma_k)) + \sigma_k\|c(x(\sigma_k))_-\|^2.$$

由引理 1(3)，$f(x(\sigma_k)) \geq f(x(\sigma_1))$，故

$$f(\bar{x}) + \sigma_k\|c(\bar{x})_-\|^2 \geq f(x(\sigma_k)) + \sigma_k\|c(x(\sigma_k))_-\|^2 \geq f(x(\sigma_1)) + \sigma_k\|c(x(\sigma_k))_-\|^2,$$

整理可得

$$0 > \|c(\bar{x})_-\|^2 - \varepsilon^2 \geq \|c(\bar{x})_-\|^2 - \|c(x(\sigma_k))_-\|^2 \geq \frac{1}{\sigma_k}(f(x(\sigma_1)) - f(\bar{x})).$$

$k \to \infty$ 时 σ_k 也趋于无穷，故上式取极限得 $0 > 0$，矛盾。 □

引理 1 考虑简单罚函数

$$P_\sigma(x)=f(x)+\sigma\left\|c(x)_-\right\|^2$$

记 $x(\sigma)$ 是无约束问题 $\min_{x\in\mathbb{R}^n}P_\sigma(x)$ 的最优解，设 $\sigma_{k+1}>\sigma_k>0$，则有

$$P_{\sigma_k}(x(\sigma_k))\leqslant P_{\sigma_{k+1}}(x(\sigma_{k+1})), \tag{3}$$

$$\|c(x(\sigma_k))_-\|\geqslant\|c(x(\sigma_{k+1}))_-\|, \tag{4}$$

$$f(x(\sigma_k))\leqslant f(x(\sigma_{k+1})) \tag{5}$$

证明. (1) 由 $x(\sigma_k)$ 的定义，

$$\begin{aligned}P_{\sigma_k}(x(\sigma_k))= f(x(\sigma_k))+\sigma_k\|c(x(\sigma_k))_-\|^2 &\le f(x(\sigma_{k+1}))+\sigma_k\|c(x(\sigma_{k+1}))_-\|^2\\ &\le \underbrace{f(x(\sigma_{k+1}))+\sigma_{k+1}\|c(x(\sigma_{k+1}))_-\|^2}_{=P_{\sigma_{k+1}}(x(\sigma_{k+1}))}\end{aligned}$$

(2) 由 $x(\sigma_k)$ 和 $x(\sigma_{k+1})$ 的定义：

$$f(x(\sigma_k))+\sigma_k\|c(x(\sigma_k))_-\|^2\leqslant f(x(\sigma_{k+1}))+\sigma_k\|c(x(\sigma_{k+1}))_-\|^2$$

$$f(x(\sigma_{k+1}))+\sigma_{k+1}\|c(x(\sigma_{k+1}))_-\|^2\leqslant f(x(\sigma_k))+\sigma_{k+1}\|c(x(\sigma_k))_-\|^2$$

两式相加，得

$$\begin{aligned}\sigma_k\|c(x(\sigma_k))_-\|^2+\sigma_{k+1}\|c(x(\sigma_{k+1}))_-\|^2 &\leqslant\sigma_k\|c(x(\sigma_{k+1}))_-\|^2+\sigma_{k+1}\|c(x(\sigma_k))_-\|^2\\ (\sigma_{k+1}-\sigma_k)\left\|c(x(\sigma_{k+1}))_-\right\|^2 &\leqslant(\sigma_{k+1}-\sigma_k)\left\|c(x(\sigma_k))_-\right\|^2\\ \Rightarrow\ \|c(x(\sigma_{k+1}))_-\|^2 &\leqslant\|c(x(\sigma_k))_-\|^2\end{aligned}$$

(2) 式得证。

(3) 由 (1)、(2) 立得。

□

引理 3 令 $\delta=\|c(x(\sigma))_-\|$，则 $x(\sigma)$ 也是约束问题

$$\begin{aligned}\min\quad & f(x)\\ \text{s.t.}\quad & \|c(x)_-\|\leqslant\delta\end{aligned}$$

的最优解。

证明. 对任意 x 满足 $\|c(x)_-\|\leqslant\delta$， $f(x(\sigma))\leq f(x)$, $\forall x$ satisfy $\|c(x)_-\|\leq\delta$

$$\begin{aligned}f(x)+\sigma\|c(x)_-\|^2 &\geqslant f(x(\sigma))+\sigma\|c(x(\sigma))_-\|^2\\ f(x) &\geqslant f(x(\sigma))+\sigma\left(\delta^2-\|c(x)_-\|^2\right)\geqslant f(x(\sigma)).\end{aligned}$$

□

Exercise 6 给出约束最优化问题的二阶充分最优性条件，并用于说明增广 Lagrange 函数的极小点与原问题最优解的等价性。

考虑等式约束问题

$$\begin{aligned} \min \quad & f(x) \\ \text{s.t.} \quad & c(x) = 0 \end{aligned}$$

其中 $c(x) = (x_1(x), \ldots, c_{m_e}(x))^\top$。定义增广 Lagrange 函数

$$P(x, \lambda, \sigma) = L(x, \lambda) + \frac{\sigma}{2}\|c(x)\|^2.$$

本题包括两个命题的证明： $L(x,\lambda)= f(x)+ \lambda^T c(x)$

(1) 设 $\bar{x}$ 是等式约束问题的可行解，且对某个 $\bar{\lambda}, \bar{x}$ 满足 $P(x, \bar{\lambda}, \sigma)$ 的极小点二阶充分条件，则 $\bar{x}$ 是该等式约束问题的严格局部最优解。

(2) 设 x^* 和 λ^* 满足等式约束问题局部最优解的二阶充分条件，则存在 σ_0 使得当 $\sigma > \sigma_0$ 时，x^* 是函数 $P(x, \lambda^*, \sigma)$ 的严格局部极小点。

证明. (1) 设 $\bar{x}$ 满足 $P(x, \bar{\lambda}, \sigma)$ 的极小点二阶充分条件，故存在 $\delta > 0$，对任意 x 满足 $0 < \|x - \bar{x}\| < \delta$ 都有

$U= \{x \mid 0< \|x-\bar{x}\| <\delta\}$

$$P(x, \bar{\lambda}, \sigma) = f(x) + \bar{\lambda}^T c(x) + \frac{\sigma}{2}\|c(x)\|^2 > P(\bar{x}, \bar{\lambda}, \sigma) = f(\bar{x}),$$

因而对满足 $0 < \|x - \bar{x}\| < \delta$ 的可行点 x 均有 $V= \{x\in U \mid c(x)=0\}$

$$f(x) = P(x, \bar{\lambda}, \sigma) > P(\bar{x}, \bar{\lambda}, \sigma) = f(\bar{x}),$$

即 $\bar{x}$ 是等式约束问题的严格局部最优解

(2) 原问题局部最优解的二阶充分条件为

$$\nabla_x L(x^*, \lambda^*) = 0, \quad c(x^*) = 0$$

且对所有 $d \in Ker\, A(x^*)$ 均有 J_c

$$d^\top \nabla^2_{xx} L(x^*, \lambda^*) d > 0$$

则 $\nabla_x P(x^*, \lambda^*, \sigma) = \nabla_x L(x^*, \lambda^*) + \sigma A(x^*)^\top c(x^*) = 0$ 对所有 $\sigma > 0$ 成立。

下证存在 σ_0 使得当 $\sigma > \sigma_0$ 时，有 $\nabla^2_{xx} P(x^*, \lambda^*, \sigma) > 0$。

Shrinkage( ) $= \underset{u}{\operatorname{argmin}}\ \|x\|_2 + \|x-u\|^2$

FISTA Prox-Linear $\|x\|_0$

OMP ADMM $\|x\|_1$

$\|x\|_2$

$\|x\|_2^2$

反证，假设对任意正整数 k，都有存在方向 d_k，满足

$$\|d_k\| = 1\text{且}d_k^T\nabla_x^2 P(x^*,\lambda^*,k)d_k \le 0, \quad (\text{即取}\sigma_k = k)$$

将 $\nabla_{xx}^2 P(x^*,\lambda^*,\sigma)$ 展开得

$$d_k^T\left(\nabla_{xx}^2 L(x^*,\lambda^*) + kA(x^*)^T A(x^*)\right)d_k \le 0$$

$$d_k^T\nabla_{xx}^2 L(x^*,\lambda^*)d_k \le -k\|A(x^*)d_k\|^2$$

$$-\frac{1}{k}d_k^T\nabla_{xx}^2 L(x^*,\lambda^*)d_k \ge \|A(x^*)d_k\|^2$$

$$\frac{1}{k}\|\nabla_{xx}^2 L(x^*,\lambda^*)\|_2 \ge \|A(x^*)d_k\|^2$$

$\|x\|_2$

$\|A\|_2 = \max_{d\in\mathbb{R}^n\setminus\{0\}} \frac{\|Ad\|_2}{\|d\|_2}$

$= \max_{\|d\|_2=1} \|Ad\|_2$

其中 $\|\nabla_{xx}^2 L(x^*,\lambda^*)\|_2$ 为 $\nabla_{xx}^2 L(x^*,\lambda^*)$ 的谱范数。由于 d_k 属于 $\mathbb{R}^n$ 中的紧集

$$\{d\in\mathbb{R}^n : \|d\| = 1\},$$

因此有聚点 $\bar{d}$，满足

$A(x^*)\bar{d} = 0$

$$\lim_{k\to\infty}\frac{1}{k}\|\nabla_{xx}^2 L(x^*,\lambda^*)\|_2 = 0 \ge \|A(x^*)\bar{d}\|^2$$

因此 $\bar{d}\in Ker\,A(x^*)$，但 $\bar{d}^T\nabla_{xx}^2 L(x^*,\lambda^*)\bar{d} \le 0$ 与原问题二阶充分条件矛盾。

□

凸优化习题讲义

2021 年 5 月 24 日

Ex 1　Let $C \subset \mathbb{R}^n$ be the solution set of a quadratic inequality,

$$C = \left\{ x \in \mathbb{R}^n | x^\top A x + b^\top x + c \leqslant 0 \right\},$$

with $A \in \mathbf{S}^n, b \in \mathbb{R}^n$, and $c \in \mathbb{R}$.

(a) Show that C is convex if $A \succeq 0$.

(b) Show that the intersection of C and the hyperplane defined by $g^\top x + h = 0$ (where $g \neq 0$) is convex if $A + \lambda g g^\top \succeq 0$ for some $\lambda \in \mathbb{R}$.

证明. (a) 令 $x, y \in C$，$\theta \in [0,1]$ 则

$$\begin{aligned}
&(\boxed{\theta x + (1-\theta) y})^\top A(\theta x + (1-\theta) y) + b^\top (\theta x + (1-\theta) y) + c \\
=&\theta \left(x^\top A x + b^\top x + c \right) + (1-\theta)(y^\top A y + b^\top y + c) \\
&+ \underline{(\theta x + (1-\theta) y)^\top A(\theta x + (1-\theta) y) - \theta x^\top A x - (1-\theta) y^\top A y}
\end{aligned}$$

只需要证明 $(\theta x + (1-\theta) y)^\top A(\theta x + (1-\theta) y) - \theta x^\top A x - (1-\theta) y^\top A y \leqslant 0$。

$$\begin{aligned}
&(\theta x + (1-\theta) y)^\top A(\theta x + (1-\theta) y) - \theta x^\top A x - (1-\theta) y^\top A y \\
=&\underbrace{(\theta^2 - \theta)}_{\leqslant 0}\underline{(x^\top A x + y^\top A y - 2x^\top A y)} \quad 要 \geqslant 0 \\
=&(\theta^2 - \theta) \left(\sqrt{A} x - \sqrt{A} y \right)^\top \left(\sqrt{A} x - \sqrt{A} y \right)
\end{aligned}$$

又因为当 $\theta \in [0,1]$ 时，$\theta^2 - \theta \leqslant 0$，故上式 $\leqslant 0$。

(b) 假设 $x, y \in C$，且 $g^\top x + h = 0, g^\top y + h = 0$，及 $\theta \in [0,1]$。显然，

$$g^\top (\boxed{\theta x + (1-\theta) y}) + h = \theta (g^\top x + h) + (1-\theta)(g^\top y + h) = 0.$$

至于证明 $(\theta x+(1-\theta)y)^\top A(\theta x+(1-\theta)y)+b^\top(\theta x+(1-\theta)y)+c\leqslant 0$，由上一问可知，只需要证明

$$x^\top Ax+y^\top Ay-2x^\top Ay\geqslant 0$$

注意到 $h^2=(-h)(-h)=(g^\top x)(g^\top x)=(g^\top y)(g^\top y)=(g^\top x)(g^\top y)$，故令 $S=\sqrt{A+\lambda gg^\top}$，我们有

$g^\top x+h=0$　　$g^\top y+h=0$

$$\begin{aligned}
&x^\top Ax+y^\top Ay-2x^\top Ay\\
=&x^\top Ax+y^\top Ay-2x^\top Ay+\lambda h^2+\lambda h^2-2\lambda h^2\\
=&x^\top Ax+y^\top Ay-2x^\top Ay+\boxed{\lambda x^\top gg^\top x}+\lambda y^\top gg^\top y-2\lambda x^\top gg^\top y\\
=&x^\top(A+\lambda gg^\top)x+y^\top(A+\lambda gg^\top)y-2x^\top(A+\lambda gg^\top)y\\
=&(Sx-Sy)^\top(Sx-Sy)\geqslant 0
\end{aligned}$$

□

Ex 2 Let $\lambda_1(X)\geqslant\lambda_2\geqslant\ldots\geqslant\lambda_n(X)$ denote the eigenvalues of a matrix $X\in\mathbf{S}^n$. Prove that the maximum eigenvalue $\lambda_1(X)$ is convex. Moreover, show that $\sum_{i=1}^k\lambda_i(X)$ is convex on $\mathbf{S}^n$.

证明. 令 $X=O\Lambda O^\top$ 为 X 的特征值分解，其中 $O=[o_1,\ldots,o_n]$ 为正交阵，

$$\Lambda=diag(\lambda_1(X),\lambda_2(X),\ldots,\lambda_n(X))$$

为对角阵。注意到取 $V\in\mathbb{R}^{n\times k}$ 为

$$V=\begin{bmatrix} o_1 & o_2 & \ldots & o_k\end{bmatrix}$$

则 $\boxed{tr(V^\top XV)=\sum_{i=1}^k\lambda_i(X)}$，且 $V^\top V=I$。另一方面，对任意 $V\in\mathbb{R}^{n\times k}$ 使得 $V^\top V=I$，记 $V=[v_1,\ldots,v_k]$，

$tr(V^\top XV)\leq\sum_{i=1}^k\lambda_i(X)$

$$tr\left(V^\top XV\right)=\sum_{j=1}^n\sum_{i=1}^k\lambda_j(X)\left(v_i^\top o_j\right)^2=\sum_{j=1}^n\lambda_j(X)\sum_{i=1}^k\left(v_i^\top o_j\right)^2$$

注意到，$v_1,\ldots,v_k$ 可以通过添加另外 $n-k$ 个列向量 $v_{k+1},\ldots,v_n$，成为 $\mathbb{R}^n$ 上的一组标准正交基，此时

$$\sum_{i=1}^n\left(v_i^\top o_j\right)^2=1\quad\Rightarrow\quad\sum_{i=1}^k\left(v_i^\top o_j\right)^2\leqslant 1.$$

另一方面，

$$\sum_{i=1}^k\boxed{\sum_{j=1}^n(v_i^\top o_j)}=\sum_{j=1}^n\sum_{i=1}^k\left(v_i^\top o_j\right)^2=k.$$

因此

$$
\begin{aligned}
tr\left(V^\top X V\right) \leqslant \max \quad & \sum_{j=1}^{n} a_j \lambda_j(X) \\
\text{s.t.} \quad & \sum_{j=1}^{n} a_j = k \\
& a_j \in [0,1]
\end{aligned}
$$

$C = \max f(x)$

$\left\{ \begin{array}{l} f(x) \leqslant C \\ \exists\, x_0 \text{ 使 } f(x_0) = C \end{array} \right.$

RHS 这个简单的优化问题，其最大值为 $\sum_{i=1}^{k} \lambda_i(X)$。

综上，$\boxed{\sum_{i=1}^{k} \lambda_i(X) = \max\{tr\left(V^\top X V\right) \ : \ V \in \mathbb{R}^{n\times k},\ V^\top V = I\}}$。由于 $tr\left(V^\top X V\right)$ 是关于 X 的线性函数，$\sum_{i=1}^{k} \lambda_i(X)$ 是一族关于 X 的线性函数的上确界，因此是关于 X 的凸函数。 □

Ex 3 Find the dual function of the LP

$$
\begin{aligned}
\min \quad & c^\top x \\
\text{s.t.} \quad & Gx \leqslant h \\
& Ax = b.
\end{aligned}
$$

Give the dual problem, and make the implicit equality constraints explicit.

解. Lagrange 函数为

$$
L(x,\lambda,\mu) = c^\top x + \lambda^\top (Gx - h) + \mu^\top (Ax - b)
$$

其中对偶变量 $\lambda \geqslant 0$。其对偶函数为

$$
\begin{aligned}
G(\lambda,\mu) &= \min_x c^\top x + \lambda^\top (Gx - h) + \mu^\top (Ax - b) \\
&= \min_x \underbrace{\boxed{(c + G^\top \lambda + A^\top \mu)^\top x}}_{\text{一次}} - h^\top \lambda - b^\top \mu \\
&= \begin{cases} -h^\top \lambda - b^\top \mu & \text{if } c + G^\top \lambda + A^\top \mu = 0 \\ -\infty & \text{otherwise.} \end{cases}
\end{aligned}
$$

因此对偶问题为

$\max\limits_{\lambda,\mu} \ G(\lambda,\mu)$

$\text{s.t.} \ \lambda \geqslant 0$

$(G(\lambda,\mu) > -\infty)$

$$
\begin{aligned}
\max \quad & -h^\top \lambda - b^\top \mu \\
\text{s.t.} \quad & c + G^\top \lambda + A^\top \mu = 0 \\
& \lambda \geqslant 0.
\end{aligned}
$$

□

Ex 4 Consider the equality constrained least-squares problem

$$\begin{aligned} \min \quad & \|Ax-b\|_2^2 \\ \text{s.t.} \quad & Gx = h \end{aligned}$$

$A^\top A$ 可逆

where $A \in \mathbb{R}^{m\times n}$ with $\mathbf{rank}A = n$, and $G \in \mathbb{R}^{p\times n}$ with $\mathbf{rank}G = p$. Give the KKT conditions, and derive expressions for the primal solution x^* and the dual solution ν^*.

$GG^\top$ 可逆

解. Lagrange 函数

$$L(x,\mu) = \|Ax-b\|_2^2 + \mu^\top(Gx-h)$$

$$\nabla_x L(x,\mu) = 2A^\top(Ax-b) + G^\top\mu$$

KKT 条件为

$$\begin{cases} 2A^\top(Ax-b) + G^\top\mu = 0 \\ Gx = h \end{cases}$$

由 $\mathbf{rank}A = n$，$A^\top A$ 可逆，故

$$\begin{aligned} 2A^\top Ax =& 2A^\top b - G^\top\mu \\ x =& \left(A^\top A\right)^{-1}\left(A^\top b - \frac{G^\top\mu}{2}\right) \end{aligned}$$

$p \leq n$

将该式代入到 KKT 的第二式中，得

$\mathrm{Rank}(G(A^\top A)^{-1}G^\top) = p.$

$$\begin{aligned} G\left(A^\top A\right)^{-1}\left(A^\top b - \frac{G^\top\mu}{2}\right) =& h \\ G\left(A^\top A\right)^{-1}G^\top\mu =& 2G\left(A^\top A\right)^{-1}A^\top b - 2h \end{aligned}$$

再由 $\mathbf{rank}G = p$，$G\left(A^\top A\right)^{-1}G^\top$ 可逆，故

$$\mu = \left[G\left(A^\top A\right)^{-1}G^\top\right]^{-1}\left(2G\left(A^\top A\right)^{-1}A^\top b - 2h\right)$$

$$x = \left(A^\top A\right)^{-1}\left(A^\top b - G^\top\left[G\left(A^\top A\right)^{-1}G^\top\right]^{-1}\left(G\left(A^\top A\right)^{-1}A^\top b - h\right)\right)$$

□

min f(x)
s.t. Ax=b

Ex 5 Suppose $Q \succeq 0$. The problem

$$\begin{aligned} \min \quad & f(x) + (Ax-b)^\top Q(Ax-b) \\ \text{s.t.} \quad & Ax = b \end{aligned}$$

is equaivalent to the primal equality constrained optimization problem. What is the Newton step for this problem? Is it the same sa that for the primal problem?

解. KKT 条件为

$$\begin{cases} \nabla f(x) + 2A^\top Q(Ax-b) + A^\top \mu = 0 \\ Ax - b = 0 \end{cases}$$

牛顿迭代步 $\boxed{\delta}$ 满足

$$\begin{bmatrix} \nabla^2 f(x) + 2A^\top QA & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \delta \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}.$$

注意到该方程组的第二行 $A\delta = 0$，这意味着该方程组的解 δ 满足

$$(\nabla^2 f(x) + \underbrace{2A^\top QA)\,\delta}_{\to 0} = \boxed{\nabla^2 f(x)\delta,}$$

$$\Rightarrow \begin{cases} \nabla^2 f(x)\,\delta + A^\top w = -\nabla f(x) \\ A\delta = 0. \end{cases}$$

即 δ 满足

$$\boxed{\begin{bmatrix} \nabla^2 f(x) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \delta \\ w \end{bmatrix} = \begin{bmatrix} -\nabla f(x) \\ 0 \end{bmatrix}}.$$

即与原问题的牛顿步相同。 □

Ex 6 Suppose we use the infeasible start Newton method to minimize $f(x)$ subject to $a_i^\top x = b_i$, $i = 1, \ldots, p$.

(a) Suppose the initial point $x^{(0)}$ satisfies the linear equality $a_i^\top x^{(0)} = b_i$. Show that the linear equality will remain satisfied for future iterates, i.e., $a_i^\top x^{(k)} = b_i$ for all k.

(b) Suppose that one of the equality constraints becomes satisfied at iteration k, i.e., we have $a_i^\top x^{(k-1)} \neq b_i$, $a_i^\top x^{(k)} = b_i$. Show that at iteration k, *all* the equality constraints are satisfied.

证明. (a) 因为对任意 k，其牛顿步 $\delta^{(k)}$ 满足

$$\begin{bmatrix} \nabla^2 f(x^{(k)}) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \delta^{(k)} \\ w \end{bmatrix} = -\begin{bmatrix} \nabla f(x^{(k)}) \\ Ax^{(k)} - b \end{bmatrix}.$$

if $a_i^\top x^{(k)} - b_i = 0$ 满足

其中 $A = [a_1, a_2, \ldots, a_p]^\top$. 由 $a_i^\top x^{(k)} - b_i = 0$ 可推出 $a_i^\top \delta^{(k)} = 0$, 即 $a_i^\top (x^{(k)} + \alpha\delta^{(k)}) - b_i = 0$ 对任意的实数 α 成立，因此 $a_i^\top x^{(k+1)} - b_i = 0$。由题设，$k = 0$ 时，$a_i^\top x^{(0)} - b_i = 0$，故对任意 $k \geqslant 0$ 均有 $a_i^\top x^{(k)} - b_i = 0$。

(b) 考虑第 $k-1$ 步的更新量 $\delta^{(k-1)}$，满足

$$\begin{bmatrix} \nabla^2 f(x^{(k-1)}) & A^\top \\ A & 0 \end{bmatrix} \begin{bmatrix} \delta^{(k-1)} \\ w \end{bmatrix} = - \begin{bmatrix} \nabla f(x^{(k-1)}) \\ Ax^{(k-1)} - b \end{bmatrix}.$$

因此必有 $A(x^{(k-1)} + \delta^{(k-1)}) - b = 0$。

断言： $x^{(k)} = x^{(k-1)} + \delta^{(k-1)}$。$(\alpha = 1)$

反证。如果 $x^{(k)} = x^{(k-1)} + \alpha\delta^{(k-1)}$，其中 $\alpha \neq 1$，则

$a_i^\top \delta^{(k-1)} = -\underbrace{(a_i^\top x^{(k-1)} - b)}_{=0}$

$$\left.\begin{aligned} a_i^\top (x^{(k-1)} + \alpha\delta^{(k-1)}) &= b_i \\ a_i^\top (x^{(k-1)} + \delta^{(k-1)}) &= b_i \end{aligned}\right\} \Rightarrow (\alpha - 1) a_i^\top \delta^{(k-1)} = 0$$

这会得到 $a_i^\top \delta^{(k-1)} = 0$ 且 $a_i^\top x^{(k-1)} - b_i = 0$ 的结论，与题设矛盾。

因此 $Ax^{(k)} = A(x^{(k-1)} + \delta^{(k-1)}) = b$，所有等式约束均被满足。

□

Ex 7 Suppose we add the constraint $x^\top x \leqslant R^2$ to the problem (106):

$$\begin{aligned} \min \quad & f_0(x) \\ \text{s.t.} \quad & f_i(x) \leqslant 0, \quad i = 1, \ldots, m \\ & Ax = b \\ & x^\top x \leqslant R^2 \end{aligned}$$

Let $\tilde{\phi}$ denote the logarithmic barrier function for this modified problem. Find $a > 0$ for which $\nabla^2(tf_0(x) + \tilde{\phi}) \succeq aI$ holds, for all feasible x.

解.

$$\tilde{\phi}(x) = -\sum_{i=1}^m \log(-f_i(x)) - \log(R^2 - x^\top x)$$

$$\nabla\tilde{\phi}(x) = \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla f_i(x) + \frac{1}{R^2 - x^\top x} 2x$$

$$\nabla^2\tilde{\phi}(x) = \sum_{i=1}^m \frac{1}{f_i(x)^2} \nabla f_i(x) \nabla f_i(x)^\top + \sum_{i=1}^m \frac{1}{-f_i(x)} \nabla^2 f_i(x) + \frac{1}{(R^2 - x^\top x)^2} xx^\top + \frac{1}{R^2 - x^\top x} I$$

前三项都是半正定的，而且 $\nabla^2(tf_0(x))$ 也是半正定的，所以令 $a=\frac{1}{R^2}$ 即有 $\nabla^2(tf_0(x)+\tilde{\phi})\succeq aI$ 成立。 □

Ex 8 Consider the problem (106), with central path $x^*(t)$ for $t>0$, defined as the solution of (111).

For $u>p^*$, let $z^*(u)$ denote the solution of

$$\begin{aligned}\min\quad & -\log(u-f_0(x))-\sum_{i=1}^m\log(-f_i(x))\\ \text{s.t.}\quad & Ax=b\end{aligned}$$

Show that the curve defined by $z^*(u)$, for $u>p^*$, is the central path. (In other words, for each $u>p^*$, there is a $t>0$ for which $x^*(t)=z^*(u)$, and conversely, for each $t>0$, there is a $u>p^*$ for which $z^*(u)=x^*(t)$).

证明. 对任意 $u>p^*$，$z^*(u)$ 满足

$$\begin{cases}\dfrac{1}{u-f_0(z^*(u))}\nabla f(z^*(u))+\displaystyle\sum_{i=1}^m\frac{1}{-f_i(z^*(u))}\nabla f_i(z^*(u))+A^\top\nu=0\\ Az^*(u)=b\end{cases}$$

令 $t=\frac{1}{u-f_0(z^*(u))}$，则 $z^*(u)$ 正好满足 $x^*(t)$ 对应的 KKT 系统。

反之，对任意一个 $t>0$，$x^*(t)$ 满足

$$\begin{cases}t\nabla f(x^*(t))+\displaystyle\sum_{i=1}^m\frac{1}{-f_i(x^*(t))}\nabla f_i(x^*(t))+A^\top\nu=0\\ Ax^*(t)=b\end{cases}$$

令 $u=\frac{1}{t}+f_0(x^*(t))$，则 $x^*(t)$ 满足 $z^*(u)$ 对应的 KKT 系统。 □