(4) 用增广拉格朗日函数方法求解如下优化问题：

$$\min x_1 + \frac{1}{3}(x_2+1)^2$$
$$s.t. x_1 \ge 0, x_2 \ge 1$$

$C_1(x) = x_1 \ge 0, \quad C_2(x) = x_2 - 1 \ge 0,$

$$\mathcal{L} = x_1 + \frac{1}{3}(x_2+1)^2 + \frac{1}{2\sigma}\left(\max\{\lambda_1 - \sigma x_1, 0\}^2 - \lambda_1^2\right) + \frac{1}{2\sigma}\left(\max\{\lambda_2 - \sigma(x_2-1), 0\}^2 - \lambda_2^2\right)$$

$$\frac{\partial \mathcal{L}}{\partial x_1} = \begin{cases} 1 + \cancel{\frac{1}{2\sigma}} \cdot \cancel{2}(\lambda_1 - \sigma x_1)\cdot(-\cancel{\sigma}), & x_1 \le \frac{\lambda_1}{\sigma} \\ 1, & x_1 > \frac{\lambda_1}{\sigma} \end{cases} = \begin{cases} 1 + \sigma x_1 - \lambda_1, & x_1 \le \frac{\lambda_1}{\sigma} \\ 1 & x_1 > \frac{\lambda_1}{\sigma} \end{cases}$$

$$\frac{\partial \mathcal{L}}{\partial x_2} = \begin{cases} \frac{2}{3}(x_2+1) + \cancel{\frac{1}{2\sigma}} \cdot \cancel{2}(\lambda_2 - \sigma(x_2-1))\cdot(-\cancel{\sigma}), & x_2 - 1 \le \frac{\lambda_2}{\sigma} \\ \frac{2}{3}(x_2+1), & x_2 - 1 > \frac{\lambda_2}{\sigma} \end{cases}$$

$$= \begin{cases} \left(\frac{2}{3}+\sigma\right)x_2 + \left(\frac{2}{3} - \sigma - \lambda_2\right), & x_2 - 1 \le \frac{\lambda_2}{\sigma} \\ \frac{2}{3}(x_2+1), & x_2 - 1 > \frac{\lambda_2}{\sigma} \end{cases}$$

$\frac{\partial \mathcal{L}}{\partial x_1} = 0 \Rightarrow$ 只能 $x_1 \le \frac{\lambda_1}{\sigma}$ 且 $x_1 = \frac{\lambda_1 - 1}{\sigma} \ge 0 \left(\Rightarrow \lambda_1 \ge 1\right)$

$\frac{\partial \mathcal{L}}{\partial x_2} = 0 \Rightarrow$ 只能 $x_2 - 1 \le \frac{\lambda_2}{\sigma}$ 且 $x_2 = \frac{\lambda_2 + \sigma - \frac{2}{3}}{\frac{2}{3}+\sigma} = \frac{3\lambda_2 - 2 + 3\sigma}{2 + 3\sigma} \ge 1 \Rightarrow \left(\lambda_2 \ge \frac{4}{3}\right)$

Lagrangian 乘子的更新：

$$\lambda_1^{(k+1)} = \max\{\lambda_1^{(k)} - \sigma C_1(x), 0\} = \max\{\lambda_1^{(k)} - \sigma \cdot \frac{\lambda_1^{(k)} - 1}{\sigma}, 0\} = 1$$

$$\lambda_2^{(k+1)} = \max\{\lambda_2^{(k)} - \sigma C_2(x), 0\} = \max\{\lambda_2^{(k)} - \sigma \cdot \frac{3\lambda_2 - 4}{2+3\sigma}, 0\}$$
$$= \max\left\{\frac{2(\lambda_2^{(k)} + 2\sigma)}{2+3\sigma}, 0\right\} = \frac{2(\lambda_2^{(k)} + 2\sigma)}{2+3\sigma}$$

对 λ_2，求解不动点方程知 $\lambda_2 = \frac{2(\lambda_2 + 2\sigma)}{2+3\sigma} \Rightarrow \lambda_2 = \frac{4}{3}$

故而最优点 $(x_1^*, x_2^*) = (0, 1)$

2023 Homework 1

1. 证明：$f(x)$为凸函数的充要条件为$epi(f)$为凸集.

$\Rightarrow$ f 凸，$\forall (x_1,t_1),(x_2,t_2)\in epi(f)$，有

$$f(x_1)=t_1\le t,\quad f(x_2)=t_2\le t$$

注意到 $f(\lambda x_1+(1-\lambda)x_2)\le \lambda t_1+(1-\lambda)t_2\le t$

$\therefore$ $\lambda x_1+(1-\lambda)x_2\in epi(f)$.

$\Leftarrow$ $\because$ $epi(f)$ 凸 且 $(x_1,f(x_1)),(x_2,f(x_2))\in epi(f)$.

$\therefore$ $\left(\lambda x_1+(1-\lambda)x_2,\ \lambda f(x_1)+(1-\lambda)f(x_2)\right)\in epi(f)$.

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

$\square$

2. 设凸函数$f(x)$为$R^n\to R$的一阶连续可微函数。证明： f(x)的任意局部极小点必为全局极小值点；若f(x)是严格凸函数，其极小值点是唯一的。

设 x^* 为 $f(x)$ 在 $O(x^*,\delta)$ 内的局部极小点，则

$$f(x^*)\le f(x)\quad \forall x\in O(x^*,\delta)$$

任取 $x'\in\mathbb{R}^n$，则 $\exists\lambda\in(0,1)$ 使得 $\lambda x^*+(1-\lambda)x'\in O(x^*,\delta)$，从而

$$f(x^*)\le f(\lambda x^*+(1-\lambda)x')\le \lambda f(x^*)+(1-\lambda)f(x')$$

$$\Rightarrow\quad (1-\lambda)f(x^*)\le(1-\lambda)f(x')\Rightarrow f(x^*)\le f(x')$$

由 x' 的任意性知 x^* 为 $\mathbb{R}^n$ 上全局极小值点

若 f 严格凸，假设存在两极小值点 $x_1\ne x_2$，则 $f(x_1)=f(x_2)\le f(x)\ (\forall x\in\mathbb{R}^n)$

$$f(\lambda x_1+(1-\lambda)x_2)<\lambda f(x_1)+(1-\lambda)f(x_2)=f(x_1)$$

这与 x_1 为极小值点矛盾！假设不成立 原命题得证

4. 证明：设α_k是$\min\limits_{\alpha>0} f(x_k+\alpha d_k)$的解，$\|\nabla^2 f(x_k+\alpha d_k)\| \le M$对一切$\alpha>0$均成立，其中$M$为一正常数，则有

$\|G_{k+1}\| \le M$ $\quad \frac{1}{\|G_{k+1}\|} \ge \frac{1}{M}$

$$f(x_k)-f(x_k+\alpha_k d_k) \ge \frac{1}{2M}\|g_k\|^2 cos^2\langle d_k,-g_k\rangle$$

$$f(x_k+\alpha_k d_k) = f(x_k)+\alpha_k g_k^T d_k + \frac{1}{2}\alpha_k^2 d_k^T G_k d_k + o(\|d_k\|^2)$$

由题设知 $f(x_k+\alpha d_k) \le f(x_k)+\alpha g_k^T d_k + \frac{\alpha^2}{2}\cdot M\|d_k\|^2$ 对$\forall \alpha>0$成立

不妨取 $\bar{\alpha} = \frac{-g_k^T d_k}{M\|d_k\|^2} > 0$，则

$$f(x_k)-f(x_k+\alpha_k d_k) \ge f(x_k)-f(x_k+\bar{\alpha} d_k)$$

$$\ge -\bar{\alpha} g_k^T d_k - \frac{\bar{\alpha}^2}{2} M\|d_k\|^2$$

$$= \frac{(g_k^T d_k)^2}{M\|d_k\|^2} - \frac{(g_k^T d_k)^2}{2M\|d_k\|^2} = \frac{(g_k^T d_k)^2}{2M\|d_k\|^2}$$

$$= \frac{(\|g_k\|\|d_k\|\cdot\cos\langle -g_k,d_k\rangle)^2}{2M\|d_k\|^2} = \frac{\|g_k\|^2\cos^2\langle -g_k,d_k\rangle}{2M}$$

$$\cancel{= -\alpha_k g_k^T d_k - \frac{1}{2}\alpha_k^2 d_k^T G_k d_k}$$

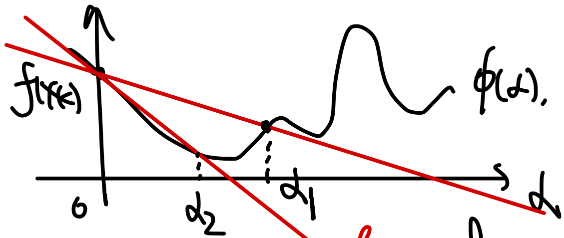
$$\cancel{= \frac{(g_k^T d_k)^2}{d_k^T G_k d_k} - \frac{1}{2}\cdot\frac{(g_k^T d_k)^2}{d_k^T G_k d_k}}$$

$$\cancel{= \frac{(g_k^T d_k)^2}{2(d_k^T G_k d_k)} \ge \frac{(g_k^T d_k)^2}{2\|d_k\|^2\|G_k\|}}$$

$$\cancel{\ge \frac{(\|g_k\|\|d_k\|\cdot\cos\langle -g_k,d_k\rangle)^2}{2M\|d_k\|^2}}$$

$$\cancel{= \frac{\|g_k\|^2\cdot\cos^2\langle -g_k,d_k\rangle}{2M}}$$

5. 证明：设$f(x_k+\alpha d_k)$在$\alpha>0$时有下界，且$g_k^T d_k<0$，则必存在α_k，在点$x_k+\alpha_k d_k$处满足Wolfe准则或Goldstein准则。



$l_1(\alpha)=f(x_k)+\rho\alpha_k g_k^T d_k.$

$l_2(\alpha)=f(x_k)+(1-\rho)\alpha_k g_k^T d_k.$ $\quad \rho\in(0,\frac{1}{2})$

Wolfe：设 $l_1(\alpha)$ 与 $\phi(\alpha)$ 第1次相交在 α_1 则

$$f(x_k+\alpha_1 d_k)=f(x_k)+\rho\alpha_1 g_k^T d_k$$
$$f(x_k+0d_k)=f(x_k)+\rho\cdot 0\cdot g_k^T d_k.$$

则 $\exists\, \alpha^*\in(0,\alpha_1)$，使得 $f(x_k+\alpha_1 d_k)-f(x_k)=g(x_k+\alpha^* d_k)^T d_k\cdot\alpha_1$

$$\Rightarrow \rho\cancel{\alpha_1} g_k^T d_k = g_{k+1}^T d_k\cdot\cancel{\alpha_1}$$

$$\Rightarrow g_{k+1}^T d_k=\rho g_k^T d_k>\sigma g_k^T d_k \quad (\text{因为 } 0<\rho<\sigma<1)$$

$\Rightarrow$ 满足 Wolfe 准则

Goldstein：设 $l_1(\alpha)$ 与 $\phi(\alpha)$ 第一次相交在 α_1，$l_2(\alpha)$ 与 $\phi(\alpha)$ 第一次相交在 α_2，则

$$f(x_k)=f(x_k)+\rho\alpha_1 g_k^T d_k=f(x_k)+(1-\rho)\alpha_2 g_k^T d_k.$$

$$\Rightarrow \rho\alpha_1=(1-\rho)\alpha_2\Rightarrow \alpha_1=\frac{1-\rho}{\rho}\alpha_2=(\frac{1}{\rho}-1)\alpha_2>\alpha_2.$$

$(0<\rho<1)$

$\therefore\ \forall\, \alpha^{**}\in(\alpha_2,\alpha_1)$ 满足 Goldstein 准则

6. 请简要说明最速下降方法的计算步骤，并解决下面这个问题:

取初值点$x^{(0)} = (1,1)^T$，采用精确线性搜索的最速下降方法求解如下无约束问题：

$$minf(x) = x_1^2 + 2x_2^2$$

最优点为 $(x_1^*, x_2^*) = (0,0)$

$g(x) = \begin{pmatrix} 2x_1 \\ 4x_2 \end{pmatrix}$, $g_0 = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$. $d_0 = \begin{pmatrix} -1 \\ -2 \end{pmatrix}$.

$\phi(\alpha) = f(x_0 + \alpha d_0) = f(1-\alpha, 1-2\alpha) = (1-\alpha)^2 + 2(1-2\alpha)^2 \Rightarrow \alpha_0 = \frac{5}{9}$

$\Rightarrow x_1 = \begin{pmatrix} 1-\alpha \\ 1-2\alpha \end{pmatrix} = \begin{pmatrix} 4/9 \\ -1/9 \end{pmatrix}$ $g_1 = \begin{pmatrix} 8/9 \\ -4/9 \end{pmatrix}$ $d_1 = \begin{pmatrix} -2 \\ 1 \end{pmatrix}$

$\phi(\alpha) = f(x_1 + \alpha d_1) = f(\frac{4}{9} - 2\alpha, -\frac{1}{9} + \alpha) = (\frac{4}{9} - 2\alpha)^2 + 2(-\frac{1}{9} + \alpha)^2 \Rightarrow \alpha_1 = \frac{5}{27}$

$\Rightarrow x_2 = \begin{pmatrix} \frac{4}{9} - 2\alpha \\ -\frac{1}{9} + \alpha \end{pmatrix} = \begin{pmatrix} 2/27 \\ 2/27 \end{pmatrix}$.

以此类推…

2023 Homework 2

1. 对问题

$$min f(x) = 10x_1^2 + x_2^2,$$

选择初始点为$(0.1, 1)^T$,证明最速下降法线性收敛。

$$G(x) = \begin{pmatrix} 10 & \\ & 1 \end{pmatrix}$$

$$(x_1^*, x_2^*) = (0,0) \qquad g(x) = \begin{pmatrix} 20x_1 \\ 2x_2 \end{pmatrix}, \ \parallel \begin{pmatrix} 10x_1 \\ x_2 \end{pmatrix},$$

$$x_{k+1} = x_k - \alpha_k g_k = \begin{pmatrix} x_k^1 - \alpha_k \cdot 10 x_k^1 \\ x_k^2 - \alpha_k \cdot x_k^2 \end{pmatrix}$$

$$\|x_{k+1}\|^2 = (x_k^{(1)})^2 (1-10\alpha_k)^2 + (x_k^{(2)})^2 (1-\alpha_k)^2.$$

$$0 < \alpha_k = \frac{g_k^T g_k}{g_k^T G g_k} = \frac{100[x_k^{(1)}]^2 + (x_k^{(2)})^2}{1000(x_k^{(1)})^2 + (x_k^{(2)})^2} < 1$$

$$\therefore 1 - 10\alpha_k < 1 - \alpha_k.$$

$$\therefore \|x_{k+1}\|^2 < \left[(x_k^{(1)})^2 + (x_k^{(2)})^2\right](1-\alpha_k)^2 = \|x_k\|^2 (1-\alpha_k)^2$$

$$\Rightarrow \frac{\|x_{k+1}\|}{\|x_k\|} < 1 - \alpha_k < 1. \qquad \therefore \text{线性收敛}$$

2. 设函数$f(x)$为凸的梯度L-利普希兹连续函数，$f^* = f(x^*) = \min\limits_{x} f(x)$存在且可达，如果步长$\alpha_k$ 取为常数α且满足$0 < \alpha < \frac{1}{L}$,那么由最速下降法得到的点列$\{x^k\}$的函数值收敛到最优值，且在函数值的意义下收敛速度为$O(\frac{1}{k})$。(利普希兹连续函数性质：$f(y) \le f(x) + \nabla f(x)^T(y - x) + \frac{L}{2}\|y - x\|^2$)

$x_{k+1} = x_k - \alpha g_k$，我要证明 $|f_{k+1} - f^*| = O(\frac{1}{k})$

$$
\begin{aligned}
f(x_{k+1}) &\le f(x_k) + g_k^T(-\alpha g_k) + \frac{L}{2}\|-\alpha g_k\|^2 \\
&= f(x_k) - \alpha g_k^T g_k + \frac{L}{2}\alpha^2 g_k^T g_k \\
&= f(x_k) + g_k^T g_k\left(\frac{L}{2}\alpha^2 - \alpha\right). \qquad (*)
\end{aligned}
$$

f凸函数 $\Leftrightarrow f(x^*) \ge f(x_k) + g_k^T(x^* - x_k)$

$\Leftrightarrow f(x_k) \le f(x^*) + g_k^T(x_k - x^*)$

于是$(*)$：$f(x_{k+1}) \le f^* + g_k^T(x_k - x^*) + g_k^T g_k\left(\frac{L}{2}\alpha^2 - \alpha\right)$

$\left(L < \frac{1}{\alpha} \Rightarrow \frac{L}{2}\alpha^2 - \alpha < -\frac{\alpha}{2}\right).$

$< f^* + g_k^T(x_k - x^*) - \frac{\alpha}{2} g_k^T g_k$

注意到 $\|x_k - x^* - \alpha g_k\|^2 = \|x_k - x^*\|^2 + \alpha^2\|g_k\|^2 - 2\alpha g_k^T(x_k - x^*)$

因此 $f_{k+1} \le f^* + \dfrac{\|x_k - x^*\|^2 + \alpha^2\|g_k\|^2 - \|x_k - x^* - \alpha g_k\|^2}{2\alpha} - \frac{\alpha}{2}\|g_k\|^2$

$= f^* + \frac{1}{2\alpha}\left(\|x_k - x^*\|^2 - \|x_k - x^* - \alpha g_k\|^2\right)$

$\Rightarrow f_{k+1} - f^* \le \frac{1}{2\alpha}\left(\|x_k - x^*\|^2 - \|x_{k+1} - x^*\|^2\right)$

对k进行求和

$$\sum_{i=0}^{k-1}(f_{i+1}-f^*) \le \frac{1}{2\alpha}(\|x_0-x^*\|-\|x_k-x^*\|^2) \le \frac{1}{2\alpha}\|x_0-x^*\|^2$$

由于 $\{f_k\}\downarrow$，因此

$$f_k-f^* \le \frac{1}{k}\sum_{i=0}^{k-1}(f_{i+1}-f^*) \le \frac{1}{2\alpha k}\|x_0-x^*\|^2 = \frac{C}{2\alpha k}$$

$\therefore f_k-f^* = o(\frac{1}{k})$

4. 设$f(x) = x_1^4 + x_1x_2 + (1 + x_2)^2$,$x^0 = (0,0)^T$,确定$v$的一个下界$\overline{v}$使得$G_0 + vI$在$v > \overline{v}$时正定，令$v_0 = 1$，此时由LM方法产生了$d_0$,验证此时$f(x^0 + d_0) < f(x^0)$,再验证只有当$v \le 0.9$时得到的$d_0$才能使$f(x^0 + d_0) < f(x^0)$。

$$g=\begin{pmatrix} 4x_1^3+x_2 \\ x_1+2(x_2+1) \end{pmatrix} \Rightarrow G(x_1,x_2)=\begin{pmatrix} 12x_1 & 1 \\ 1 & 2 \end{pmatrix} \quad g_0=\begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

$$G_0+vI=\begin{pmatrix} v & 1 \\ 1 & v+2 \end{pmatrix}$$ 顺序主子式全大于0 $\Rightarrow \begin{cases} v>0 \\ v(v+2)>1 \end{cases}$

$\Rightarrow v>-1+\sqrt{2}=\overline{v}$

取 $v_0=1$ 则 $G_0+I=\begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}$

$$(G_0+I)d_0=-g_0 \Rightarrow \begin{pmatrix} 1 & 1 \\ 1 & 3 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 0 \\ -2 \end{pmatrix} \Rightarrow d_0=\begin{pmatrix} 1 \\ -1 \end{pmatrix}$$

此时 $f(x^0+d_0)=f(1,-1)=1-1+0=0<f(x^0)=1$.

$$(G_0+vI)d_0=-g_0 \Rightarrow \begin{pmatrix} v & 1 \\ 1 & v+2 \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}=\begin{pmatrix} 0 \\ -2 \end{pmatrix} \Rightarrow d_0=\begin{pmatrix} \frac{2}{v^2+2v-1} \\ \frac{-2v}{v^2+2v-1} \end{pmatrix}$$

$$f(x_0+d_0)=\frac{16}{(v^2+2v-1)^4}+\frac{-4v}{(v^2+2v-1)^2}+\frac{(v^2-1)^2}{(v^2+2v-1)^2}$$

$$=\frac{16-4v(v^2+2v-1)^2+(v^2-1)^2(v^2+2v-1)^2}{(v^2+2v-1)^4}<f(x_0)=1.$$

$\Rightarrow$

编程计算可得 $v<0.9004$

7. 如果α_k由不精确线搜索的Wolfe-Powell准则产生，那么FR算法具有下降性质$g_k^T d_k < 0$.

即强Wolfe准则.

自然能证明

我们证明 $-\frac{1}{1-\sigma}<\frac{g_k^T d_k}{\|g_k\|^2}<\frac{2\sigma-1}{1-\sigma}$ 当 $\sigma\in(0,\frac{1}{2})$ 时 <0

① $k=0$, $\frac{g_0^T d_0}{\|g_0\|^2}=\frac{g_0^T(-g_0)}{\|g_0\|^2}=-1\in\left(-\frac{1}{1-\sigma},\frac{2\sigma-1}{1-\sigma}\right)$,

② $k=K$时成立 则$k+1$时

$$\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2}=\frac{g_{k+1}^T\left(-g_{k+1}+\frac{g_{k+1}^T g_{k+1}}{g_k^T g_k}d_k\right)}{\|g_{k+1}\|^2}$$

$$=\frac{-\|g_{k+1}\|^2+\frac{\|g_{k+1}\|^2}{\|g_k\|^2}g_{k+1}^T d_k}{\|g_{k+1}\|^2}=-1+\frac{g_{k+1}^T d_k}{\|g_k\|^2}$$

强Wolfe准则 $\Rightarrow |g_{k+1}^T d_k|<-\sigma g_k^T d_k$ $\therefore$

$$-\frac{1}{1-\sigma}<-1+\frac{\sigma g_k^T d_k}{\|g_k\|^2}<\frac{g_{k+1}^T d_{k+1}}{\|g_{k+1}\|^2}<-1-\sigma\frac{g_k^T d_k}{\|g_k\|^2}<-1+\frac{\sigma}{1-\sigma}=\frac{2\sigma-1}{1-\sigma}$$

2023 Homework 3

1. 对最小二乘问题

$$minf(x) = 1/2\sum_{i=1}^{2} r^2{}_i(x).$$

$$f(x) = \frac{1}{2}\sum_{i=1}^{2} r_i^2(x)$$

其中

$$r_1(x) = x_1^3 - x_2 - 1, r_2(x) = x_1^2 - x_2.$$

$$r(x) = [x_1^3 - x_2 - 1, x_1^2 - x_2]^T$$

$$g(x) = \sum_{i=1}^{2} r_i(x) \nabla r_i(x)$$

写出$J(x)$,$\nabla f(x)$, $S(x)$.

$$J(x) = [\nabla r_1(x), \nabla r_2(x)]^T = \begin{pmatrix} 3x_1^2, & -1 \\ 2x_1, & -1 \end{pmatrix}$$

$$G(x) = \sum_{i=1}^{2} (\nabla r_i(x))^2 + r_i(x) \nabla^2 r_i(x)$$

$$\nabla f(x) = J(x)^T r(x) = \begin{pmatrix} 3x_1^2 & 2x_1 \\ -1 & -1 \end{pmatrix} \begin{pmatrix} x_1^3 - x_2 - 1 \\ x_1^2 - x_2 \end{pmatrix}$$

$$= \begin{pmatrix} 3x_1^2(x_1^3 - x_2 - 1) + 2x_1(x_1^2 - x_2) \\ -x_1^3 + x_2 + 1, -x_1^2 + x_2. \end{pmatrix}$$

$$S(x) = \sum_{i=1}^{2} r_i(x) \nabla^2 r_i(x)$$

$$\nabla^2 r_1(x) = \begin{pmatrix} 6x_1 & 0 \\ 0 & 0 \end{pmatrix} \quad \nabla^2 r_2(x) = \begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}.$$

$$S(x) = \begin{pmatrix} 6x_1(x_1^3 - x_2 - 1) + 2(x_1^2 - x_2) & 0 \\ 0 & 0 \end{pmatrix}.$$

$$= \begin{pmatrix} 6x_1^4 - 6x_1x_2 + 2x_1^2 - 6x_1 - 2x_2 & 0 \\ 0 & 0 \end{pmatrix}$$

$$d = -(J^TJ+\nu_i I)^{-1}(J^Tr)$$

2. 设d_i是方程组

$$\left(J^{\mathrm{T}} J+\nu_i I\right) d=-J^{\mathrm{T}} r \quad i=1,2$$

的解,其中$\nu_1>\nu_2>0$.证明:$q\left(d_2\right)<q\left(d_1\right)$,其中$q(d)=\frac{1}{2}\|J d+r\|^2$.

$$q(d) = \frac{1}{2}(Jd+r)^T(Jd+r) = \frac{1}{2}\left(d^TJ^TJd + 2d^TJ^Tr + r^Tr\right)$$

而 $J^TJd + \nu_i d = -J^Tr$

$$\therefore q(d) = \frac{1}{2}\left(d^T(-J^Tr - \nu_i d) + 2d^TJ^Tr + r^Tr\right)$$

$$= \frac{1}{2}\left(d^TJ^Tr - \nu_i d^Td + r^Tr\right)$$

$$= \frac{1}{2}\left(\left[-(J^TJ+\nu_i I)^{-1}J^Tr\right]^TJ^Tr - \nu_i\left[(J^TJ+\nu_i I)^{-1}J^Tr\right]^T(J^TJ+\nu_i I)^{-1}J^Tr + r^Tr\right]$$

$$= \frac{1}{2}\left(-r^TJ(J^TJ+\nu_i I)^{-1}J^Tr - \nu_i r^TJ(J^TJ+\nu_i I)^{-2}J^Tr + r^Tr\right)$$

对$(J^TJ+\nu_i I)^{-1}$作特征值分解,即$(J^TJ+\nu_i I)^{-1} = Q^T\Lambda_i Q$.

其中Λ_i的对角元素为$\frac{1}{\lambda_i+\nu_i}$,其中$\lambda_i$为$J^TJ$的特征值,那么

$$= \frac{1}{2}\left(-r^TJQ^T\Lambda_i QJ^Tr - \nu_i r^TJQ^T\Lambda_i^2 QJ^Tr + r^Tr\right),$$

$$= \frac{1}{2}\left(-(QJ^Tr)^T\Lambda_i(QJ^Tr) - \nu_i(QJ^Tr)\Lambda_i^2(QJ^Tr) + r^Tr\right),$$

$$\boxed{= \frac{1}{2}\left(r^Tr - \boxed{(QJ^Tr)^T(\Lambda_i + \nu_i\Lambda_i^2)(QJ^Tr)}\right)} \downarrow \uparrow$$

$\Downarrow$

特征值为 $\frac{1}{\lambda_i+v_i}+\frac{v_i}{(\lambda_i+v_i)^2}=\frac{\lambda_i+2v_i}{(\lambda_i+v_i)^2}$

$$f'(v_i)=\frac{2(\lambda_i+v_i)^2-(\lambda_i+2v_i)(2(\lambda_i+v_i))}{(\lambda_i+v_i)^4}$$

$$=\frac{2(\lambda_i^2+2\lambda_i v_i+v_i^2)-2(\lambda_i^2+3\lambda_i v_i+2v_i^2)}{\square}$$

$$=\frac{-2\lambda_i v_i-2v_i^2}{\square}$$

$$=\frac{-2v_i(\lambda_i+v_i)}{(\lambda_i+v_i)^4}=\frac{-2v_i}{(\lambda_i+v_i)^3}$$

$\left.\begin{matrix}\lambda_i+v_i>0\\ v_i>0\end{matrix}\right\}\Rightarrow f'(v_i)<0\Rightarrow v_i\uparrow$  $\downarrow$  $\uparrow$

$\therefore v_1>v_2\Rightarrow g(d_1)>g(d_2)$

3. 求解非线性优化问题

$$\begin{cases} \min f(x) = x_1 - x_2^2 \\ \text{ s.t. } x_1 \geq 1, \end{cases}$$

$a_1(x) = (1, 0)$

$c_1(x) = x_1 - 1 \geq 0$

的Kuhn-Tucker点,并验证该点是否为极小值点.

$L(x,\lambda) = x_1 - x_2^2 - \lambda(x_1 - 1)$

$$\begin{cases} \frac{\partial L}{\partial x_1} = 1 - \lambda = 0 \\ \frac{\partial L}{\partial x_2} = -2x_2 = 0 \\ \lambda(x_1 - 1) = 0 \\ \lambda \geq 0 \end{cases} \Rightarrow \begin{cases} x_1 = 1 \\ x_2 = 0 \\ \lambda = 1 \end{cases}$$

$\therefore$ KKT点为 $((1, 0), 1)$

$\nabla^2 L = \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix}$

$F^* = \{d : a^T d \geq 0\} = \{d : d = (0, d_2), d_2 \in \mathbb{R}\}$

$F_1^* = \{d : (g^*)^T d = 0, d \in F^*\} = \{d : d \in F^*\}$

$\forall d \in F_1^*$, $d^T \nabla^2 L d = \begin{pmatrix} 0 \\ d_2 \end{pmatrix}^T \begin{pmatrix} 0 & 0 \\ 0 & -2 \end{pmatrix} \begin{pmatrix} 0 \\ d_2 \end{pmatrix} = -2|d_2| \leq 0$

$\therefore$ 不是极小值点.

4. 叙述约束优化问题取严格极小值的二阶充分条件,并对于如下优化问题:

$$min \quad x_1^2+x_2^2, \quad st.\frac{x_1^2}{4}+x_2^2=1.$$

求其严格局部极小点。

二阶充分条件:若 $\forall d\in F_1^*$, $d^T\nabla^2 L d>0$, 则 x^* 为严格极小值点, 其中 $F_1^*=\{d:(g^*)^T d=0,\ d\in F^*\}$, $F^*=\{d:\ a_i^T d=0,\ i\in E,\ a_i^T d\geq 0,\ i\in I(x^*)\}$.

$$L=x_1^2+x_2^2-\lambda\left(\frac{x_1^2}{4}+x_2^2-1\right) \quad a_1(x)=\begin{pmatrix}\frac{x_1}{2}\\2x_2\end{pmatrix} \quad g(x)=\begin{pmatrix}(2-\frac{\lambda}{2})x_1\\(2-2\lambda)x_2\end{pmatrix}$$

$$\begin{cases}\frac{\partial L}{\partial x_1}=2x_1-\lambda\cdot\frac{x_1}{2}=0\\ \frac{\partial L}{\partial x_2}=2x_2-2\lambda x_2=0\\ \lambda\left(\frac{x_1^2}{4}+x_2^2-1\right)=0\\ \lambda\geq 0\\ \frac{x_1^2}{4}+x_2^2=1\end{cases} \Rightarrow \begin{cases}x_1=0\\x_2=\pm 1\\ \lambda=1\end{cases} \text{或} \begin{cases}x_1=\pm 2\\x_2=0\\ \lambda=4\end{cases}$$

$$\nabla^2 L=\begin{pmatrix}2-\frac{\lambda}{2} & 0\\ 0 & 2-2\lambda\end{pmatrix}$$

(i) $\lambda=1$ 时 $\nabla^2 L=\begin{pmatrix}3/2 & \\ & 0\end{pmatrix}$

$F^*=\{d:\ a_1(x)^T d=0\}=\{d:\ d=\begin{pmatrix}d_1\\0\end{pmatrix},\ d_1\in R\}$

$F_1^*=\{d:\ (g^*)^T d=0,\ d\in F^*\}=F^*$.

$d^T\nabla^2 L d=\frac{3}{2}d_1^2>0$, ∴ 为严格极小点

(ii) $\lambda=4$ $\nabla^2 L=\begin{pmatrix}-2 & \\ & -6\end{pmatrix}$

$F^*=\{d:\ d=(0,d_2),\ d_2\in R\}$

$F_1^*=\{d:(g^*)^T d=0,\ d\in F^*\}=F^*$

$d^T\nabla^2 L d=-6d_2^2<0$ ∴ 不是严格极小点

∴ 严格局部极小点为 $x^*=(0,\pm 1)$

5. 假设可行点x^*是一般约束优化问题的局部极小点.证明:如果$f(x)$和$c_i(x), i \in \mathcal{E} \cup \mathcal{I}$在点$x^*$ 处是可微的，那么($\mathcal{T}_x$表示可行方向构成的集合)

$$d^\top \nabla f(x^*) \geq 0, \underline{d \in \mathcal{T}_x(x^*)}, \text{可行方向集.}$$

等价于

$$\underbrace{\mathcal{T}_x(x^*)}_{\text{可行方向集}} \cap \{d | \underbrace{\nabla f(x^*)^\top d < 0}_{\text{下降方向}}\} = \emptyset.$$

$\Rightarrow$ 若 $T_x(x^*) \cap \{d | (g^*)^T d < 0\} \neq \phi$. 不妨设 $d' \in T_x(x^*) \cap \{d | (g^*)^T d < 0\}$,

则 $g(x^*)^T (d') < 0$ 矛盾!

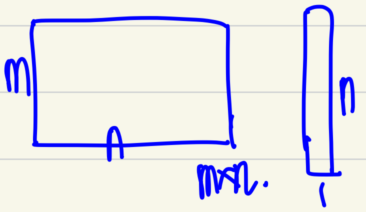
$g(x^*)^T (d') \geq 0$,

$\Leftarrow$ 假设 $\exists d'' \in T_x(x^*)$ 使得 $(d'')^T g(x^*) < 0$. 则 $d'' \in T_x(x^*) \cap \{d | (g^*)^T d < 0\}$.

这与 $T_x(x^*) \cap \{d | (g^*)^T d < 0\} \neq \phi$ 矛盾!

2024 Homework 3

(1) (a) 最小二乘问题. $f=r^Tr$, $J=[\nabla r_1,\dots \nabla r_m]^T\in\mathbb{R}^{m\times n}$.



$\Rightarrow$ J列满秩, 假设 J^TJ 奇异, 则 $\exists x\neq 0$ 使得 $(J^TJ)x=0$

那么 $x^T(J^TJ)x=(Jx)^TJx=\|Jx\|^2=0$, $\Rightarrow Jx=0$

但J列满秩, 若 $Jx=0$ 只能 $x=0$. 矛盾! 假设不成立.

$\Leftarrow$ J^TJ 非奇异时, 假设J不列满秩, 则 $\exists x\neq 0$ 使得 $Jx=0$, 从而 $J^T(Jx)=(J^TJ)x=0$. 但 J^TJ 非奇异 若 $(J^TJ)x=0$ 只能 $x=0$. 矛盾! 假设不成立.

(b) $\Rightarrow$ $\forall x\neq 0$, $x^T(J^TJ)x=(Jx)^TJx=\|Jx\|^2$, 因为J列满秩, 所以 $Jx\neq 0$, 那么 $x^T(J^TJ)x=\|Jx\|^2>0$,

$\Leftarrow$ J^TJ 正定, $\forall x\neq 0$, $x^T(J^TJ)x=(Jx)^TJx>0$, 若J不列满秩, 则可找 $\tilde{x}$ 使得 $J\tilde{x}=0$, 这与 $(J\tilde{x})^T(J\tilde{x})>0$ 矛盾!

(2) $f=\frac{1}{2}\sum_{i=1}^{m} r(x)_i^T r(x)_i \Rightarrow g=\sum_{i=1}^{m} r_i(x)^T\nabla r_i(x)=J(x)^T r(x)$

$J(x)=[\nabla r_1(x),\dots \nabla r_m(x)]^T\in\mathbb{R}^{m\times n}$. $\forall x, y\in D$.

$$\|J(x)-J(y)\|^2=(J(x)-J(y))^T(J(x)-J(y))$$

$$=J(x)^TJ(x)-2J(x)^TJ(y)+J(y)^TJ(y)$$

$$=\sum_{i=1}^{m}\left[(\nabla r_i(x))^2-2\nabla r_i(x)^T\nabla r_i(y)+(\nabla r_i(y))^2\right]$$

$$=\sum_{i=1}^{m}\|\nabla r_i(x)-\nabla r_i(y)\|^2\le mL^2\|x-y\|^2$$

$\therefore$ $J(x)$ 的 Lipschitz 常数为 $\sqrt{m}L$

$$\|g(x)-g(y)\| = \sum_{i=1}^{m} \left| r_i(x)^T \nabla r_i(x) - r_i(y)^T \nabla r_i(y) \right|$$

$$= \sum_{i=1}^{m} \left| r_i^T(x)\left(\nabla r_i(x) - \nabla r_i(y)\right) + \nabla r_i^T(y)\left(r_i^T(x) - r_i^T(y)\right) \right|$$

$$\le \sum_{i=1}^{m} \|r_i(x)\| \, \|\nabla r_i(x) - \nabla r_i(y)\| + \sum_{i=1}^{m} \|\nabla r_i(y)\| \, \|r_i^T(x) - r_i^T(y)\|$$

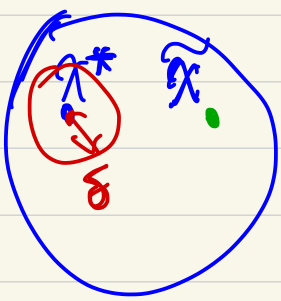
$$\le \left(m(ML) + \sum_{i=1}^{m} \|\nabla r_i(y)\| \cdot L \right) \|x-y\|$$

注意到 $\|r_i(x) - r_i(y)\| \le L\|x-y\| \Rightarrow \|\nabla r_i(x)\| < L$

$$\therefore \|g(x)-g(y)\| \le (mML + mL^2)\|x-y\|$$

Lipschitz 常数为 $mML + mL^2$

(3)



Ω.

不妨设 x^* 为 $f(x)$ 在 $O(x^*, \delta)$ 内的局部解.
则 $f(x^*) \le f(x)$, $x \in O(x^*, \delta)$.

$\forall \tilde{x} \in \Omega$, 取 λ s.t. $\lambda x^* + (1-\lambda)\tilde{x} \in O(x^*, \delta)$.
则

$$f(x^*) \le f(\lambda x^* + (1-\lambda)\tilde{x}) \le \lambda f(x^*) + (1-\lambda) f(\tilde{x})$$

$$\Rightarrow (1-\lambda) f(x^*) \le (1-\lambda) f(\tilde{x}) \Rightarrow f(x^*) \le f(\tilde{x})$$

由 $\tilde{x}$ 的任意性知 $f(x^*)$ 也为 Ω 上的最优解. 即全局最优解.

再记 $S = \{x^* : f(x^*) \le f(x), \forall x \in \Omega\}$. 下证明其为凸集

① $|S| = 0$ 或 1 $\Rightarrow$ S 必为凸集

② $|S| \ge 2$ 时

$\forall x_1^*, x_2^* \in S$,

$f(\lambda x_1^* + (1-\lambda) x_2^*) \le \lambda f(x_1^*) + (1-\lambda) f(x_2^*) \le \lambda f(x) + (1-\lambda) f(x) = f(x)$

$\therefore \lambda x_1^* + (1-\lambda) x_2^* \in S \Rightarrow S$ convex

$a_1(x) = a \in \mathbb{R}^n$.

(4) $\min f = x^T x \quad$ s.t. $a^T x + \alpha \ge 0$.

$L = x^T x - \lambda(a^T x + \alpha)$

KKT: $\begin{cases} \frac{\partial L}{\partial x} = 2x - \lambda a = 0 \Rightarrow x = \frac{\lambda}{2} a \\ \lambda(a^T x + \alpha) = 0 \Rightarrow \lambda(\frac{\lambda}{2} a^T a + \alpha) = 0 \\ \lambda \ge 0,\ a^T x + \alpha \ge 0 \end{cases}$

① $\lambda = 0$, KKT对 $(x, \lambda) = (0, 0)$, 此时 $\alpha \ge 0$

② $\lambda > 0$, $a^T x + \alpha = \frac{\lambda}{2} a^T a + \alpha = 0 \Rightarrow \lambda = \frac{-2\alpha}{a^T a}$ $(\alpha < 0)$

KKT对 $(x, \lambda) = \left(\frac{-\alpha \cdot a}{a^T a}, \frac{-2\alpha}{a^T a}\right)$

$\nabla^2 L = 2I$ 正定, $a_1(x) = a \in \mathbb{R}^n$ $a_1(x)^T d = a^T d$.

① 当 $\alpha \ge 0$ 时 对 $\forall d \in F_1(0) = \{d : a^T d = 0, \lambda^* > 0\}$

$d^T \nabla^2 L d = 2 d^T d > 0$, $\therefore x = 0$ 为最优解.

② 当 $\alpha < 0$ 时 对 $\forall d \in F_1(x^*) = \{d : a^T d \ge 0, \lambda^* = 0\}$.

$d^T \nabla^2 L d = 2 d^T d > 0$. $\therefore x = \frac{-\alpha \cdot a}{a^T a}$ 为最优解.

(5). $L = (x-1)^2 + (y-2)^2 - \lambda\left((x-1)^2 - 5y\right)$.

$$\frac{\partial L}{\partial x} = 2(x-1) - 2\lambda(x-1) = 0$$

$$\frac{\partial L}{\partial y} = 2(y-2) + 5\lambda = 0 \quad \Rightarrow \quad \begin{cases} x = 1 \\ y = 2 \\ \lambda = 0 \end{cases} \text{和} \begin{cases} x = \text{任意} \\ y = -\frac{1}{2} \\ \lambda = 1 \end{cases}$$

$$\lambda\left((x-1)^2 - 5y\right) = 0$$

∴ KKT点为 $((1,2), 0)$

$$\nabla^2 L = \begin{pmatrix} 2-2\lambda & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 2 & \\ & 2 \end{pmatrix} \text{ 正定}$$

∴ $(1,2)$ 为极小点

(6) $P_E(x, \sigma) = x_1 + x_2 + \frac{\sigma}{2}\left(x_2 - x_1^2\right)^2$

$$\frac{\partial P}{\partial x_1} = 1 + \sigma(x_2 - x_1^2)(-2x_1) = 0$$

$$\frac{\partial P}{\partial x_2} = 1 + \sigma(x_2 - x_1^2)(1) = 0 \quad \Rightarrow \quad \begin{cases} x_1 = -\frac{1}{2} \\ x_2 = -\frac{1}{\sigma} + \frac{1}{4} \end{cases}$$

令 $\sigma \to \infty$，则 $(x_1^*, x_2^*) = \left(-\frac{1}{2}, \frac{1}{4}\right)$

(7) $L = x_1^2 + 2x_2^2 - \mu \ln(x_1 + x_2 - 1)$

$$\left.\begin{aligned} \frac{\partial L}{\partial x_1} &= 2x_1 - \frac{\mu}{x_1+x_2-1} = 0 \\ \frac{\partial L}{\partial x_2} &= 4x_2 - \frac{\mu}{x_1+x_2-1} = 0 \end{aligned}\right\} \Rightarrow \left| \begin{aligned} x_1 &= \frac{1 \pm \sqrt{1+3\mu}}{3} \\ x_2 &= \frac{1 \pm \sqrt{1+3\mu}}{6}, \end{aligned}\right.$$

$$x_1 + x_2 \geq 1 \Rightarrow \left| \begin{aligned} x_1 &= \frac{1+\sqrt{1+3\mu}}{3} \\ x_2 &= \frac{1+\sqrt{1+3\mu}}{6} \end{aligned}\right.$$

令 $\mu \to 0$. $(x_1^*, x_2^*) = \left(\frac{2}{3}, \frac{1}{3}\right)$.

(8) $L = 2x_1^2 + x_2^2 - 2x_1x_2 - \lambda(x_1+x_2-1) + \frac{\sigma}{2}(x_1+x_2-1)^2$

$= 2x_1^2 + x_2^2 - 2x_1x_2 - \lambda(x_1+x_2-1) + (x_1+x_2-1)^2$

$$\frac{\partial L}{\partial x_1} = 4x_1 - 2x_2 - \lambda_k + 2(x_1+x_2-1) = 0$$

$$\frac{\partial L}{\partial x_2} = 2x_2 - 2x_1 - \lambda_k + 2(x_1+x_2-1) = 0$$

$$\Rightarrow \left| \begin{aligned} x_1 &= \frac{\lambda_k+2}{6} \\ x_2 &= \frac{\lambda_k+2}{4} \end{aligned}\right. \qquad 2(\lambda_k+2) + 3(\lambda_k+2)$$

$$\lambda_{k+1} = \lambda_k - \sigma_k C(x_k) = \lambda_k - 2\left(\frac{\lambda_k+2}{6} + \frac{\lambda_k+2}{4} - 1\right)$$

$$= \lambda_k - 2\left(\frac{5(\lambda_k+2)}{12} - 1\right)$$

$$= \lambda_k - \frac{5}{6}(\lambda_k+2) + 2$$

$$= \frac{1}{6}\lambda_k + \frac{1}{3}$$

$x = \frac{1}{6}x + \frac{1}{3} \qquad \frac{5}{6}x = \frac{1}{3} \qquad x = \frac{2}{5}.$

$\therefore \lambda_k \to \frac{2}{5} \Rightarrow (x_1^*, x_2^*) = \left(\frac{2}{5}, \frac{3}{5}\right)$

2024 Homework 1

(1) $S=\{x^*: f(x^*)\le f(x),\ \forall x\in \mathrm{dom} f\}$.

$\forall x_1^*, x_2^* \in S$,

$f(\lambda x_1^* + (1-\lambda)x_2^*) \le \lambda f(x_1^*) + (1-\lambda) f(x_2^*) \le f(x)$

$\therefore \lambda x_1^* + (1-\lambda) x_2^* \in S \Rightarrow S$ convex

(2) $g(x)=\begin{pmatrix} 8+2x_1 \\ 12-4x_2 \end{pmatrix} = 0 \Rightarrow (x_1, x_2) = (-4, 3)$,

$G(x)=\begin{pmatrix} 2 & \\ & -4 \end{pmatrix}$ 不定阵，在 (x_1, x_2) 上为 saddle point

(3) $g(x)=\begin{pmatrix} 2(x_1+x_2^2) \\ 2(x_1+x_2^2)\cdot 2x_2 \end{pmatrix} = \left(2(x_1+x_2^2),\ 4x_2(x_1+x_2^2)\right)^T$

$g(1,0)=(2,0)^T$，则 $g^T p = -2 < 0$ $\therefore$ 为下降方向

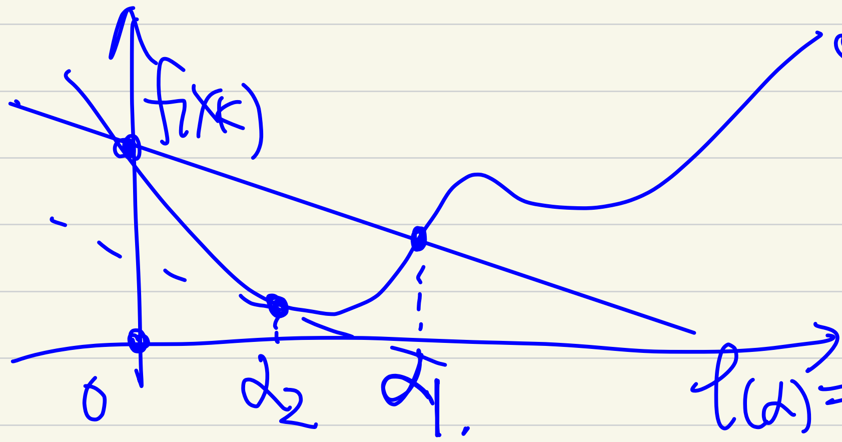
(4) $\phi(\alpha)\ f(x+\alpha(-1,1)) = f(1-\alpha, \alpha) = (1-\alpha+\alpha^2)^2$
$(1,0)$

$\frac{\partial \phi(\alpha)}{\partial \alpha} = 2(1-\alpha+\alpha^2)\cdot(2\alpha-1) = 0 \Rightarrow \alpha = 1/2$

(4). $\dfrac{\|x_{k+1}-0\|}{\|x_k-0\|} = \dfrac{\frac{1}{(k+1)!}}{\frac{1}{k!}} = \dfrac{k!}{(k+1)!} = \dfrac{1}{k+1} \to 0.$ 超线性

$\dfrac{\|x_{k+1}-0\|}{\|x_k-0\|^2} = \dfrac{(k!)^2}{(k+1)!} = \dfrac{k!}{k+1} \to \infty.$ ✗(不二次)

(5).



$\phi(\alpha)=f(x_k+\alpha d_k)$.

$l(\alpha)=f(x_k)+\rho g_k^T d_k \alpha$.

$\phi(\alpha)$ 有下界 而 $l(\alpha)\to-\infty$. 不妨记第一次相交在 $\alpha_1>0$.

即 $\phi(\alpha_1)=f(x_k+\alpha_1 d_k)=f(x_k)+\rho g_k^T d_k\cdot\alpha_1$

$\phi(0)=\phi(\alpha_1) \Rightarrow \exists\ \alpha_2\in(0,\alpha_1)$ s.t.

$\phi(\alpha_1)-\phi(0)=\phi'(\alpha_2)^T\cdot\alpha_1 d_k$

$\rho g_k^T d_k\cdot\cancel{\alpha_1} = g(x_k+\alpha_2 d_k)^T\cdot\cancel{\alpha_1}\, d_k.$

$\underline{g(x_k+\alpha_2 d_k)^T d_k} = \rho \underbrace{g_k^T d_k}_{<0} > \sigma g_k^T d_k$

$0<\rho<\sigma<1$

∴满足 Wolfe 准则,

$|g(x_k+\alpha_2 d_k)^T d_k| = \rho|g_k^T d_k| < \sigma|g_k^T d_k| = -\sigma g_k^T d_k$

∴满足强 Wolfe 准则

2024 Homework 2

$$Qx = \lambda x$$

$$Q(x_0 - x^*) = \lambda (x_0 - x^*)$$

$$\Downarrow$$

$$g_0 - g^* = \lambda (x_0 - x^*)$$

$$g_0 = \lambda (x_0 - x^*)$$

(1)

$$x_1 = x_0 + \alpha_0 g_0$$

$$g_0 = Qx_0 - b$$

$$\alpha_0 = \frac{g_0^T g_0}{g_0^T Q g_0} = \frac{\lambda^2 (x_0 - x^*)^T (x_0 - x^*)}{\lambda^2 (x_0 - x^*)^T \underline{Q(x_0 - x^*)}} \quad = \lambda (x_0 - x^*)$$

$$= \frac{1}{\lambda}$$

$$\therefore x_1 = x_0 + \frac{1}{\lambda} \cdot \lambda (x_0 - x^*) = x^*$$

✓

(2) /

f convex $\Longleftrightarrow$ G 半正定

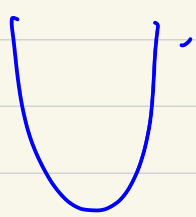
强 convex $\Longleftrightarrow$ G 正定

(3) f strongly convex, $\tilde{f} = f - \frac{m}{2} x^T x$ convex

$$= \alpha_k d_k.$$

$$s_k^T y_k = (x_{k+1} - x_k)^T (g_{k+1} - g_k)$$

$$f(x_{k+1}) = f(x_k + \alpha_k d_k) \geq f(x_k) + \alpha g_k^T d_k$$



$$\alpha_k d_k^T (g_{k+1} - g_k) > 0$$

$$d_k^T g_{k+1} > d_k^T g_k.$$

$$d_k^T g(x_k + \alpha_k d_k) > d_k^T g_k.$$

$$d_k^T g_{k+1} \geq d_k^T (g_k + G_k \alpha_k d_k)$$

$$= d_k^T g_k + \underline{\alpha_k d_k^T G_k d_k} > d_k^T g_k \quad \square$$

(4) $s_k^T y_k = \alpha_k d_k^T (g_{k+1} - g_k)$, 这里强 Wolfe.

By 强 Wolfe. $|g_{k+1}^T d_k| < -\sigma g_k^T d_k$,

$$\sigma g_k^T d_k < g_{k+1}^T d_k < -\sigma g_k^T d_k$$

$$\therefore s_k^T y_k = \alpha_k (g_{k+1}^T d_k - g_k^T d_k)$$

$$> \alpha_k \underbrace{(\sigma - 1)}_{<0} \underbrace{g_k^T d_k}_{<0} > 0.$$

(5)

$$SR_1: H_{k+1} = H_k + \frac{(s_k - H_k y_k)(s_k - H_k y_k)^T}{(s_k - H_k y_k)^T y_k}$$

$k=1$ 时 ✓

设 k 时，成立 $H_k y_j = s_j$ $(j=0,\dots,k-1)$，则
$k+1$ 时 ① $j=k$，拟 Newton 条件知成立

② $j=0,\dots,k-1$ 时.

$$H_{k+1} y_j = H_k y_j + \frac{(s_k - H_k y_k)^T (s_k^T y_j - y_k^T H_k y_j)}{(s_k - H_k y_k)^T y_k}$$

$$= H_k y_j + \frac{(s_k - H_k y_k)^T (s_k^T y_j - y_k^T s_j)}{(s_k - H_k y_k)^T y_k} = H_k y_j$$

$$s_k^T y_j = \alpha_k d_k^T G (\alpha_j d_j) = \alpha_j d_j^T G \alpha_k d_k = s_j^T G s_k = \underline{s_j^T y_k}.$$

③ $\therefore$ 对 $\forall n = 1 \dots$ 成立 $H_k y_j = s_j$. $(j=0 \dots k-1)$

(b) $\phi(x_k+\alpha_k d_k)=\frac{1}{2}(x_k+\alpha d_k)^T G(x_k+\alpha d_k)-b^T(x_k+\alpha d_k)$,

$$=\frac{1}{2}x_k^T G x_k+\alpha d_k^T G x_k+\frac{1}{2}\alpha^2 d_k^T G d_k-b^T d_k.$$

$$\frac{\partial\phi}{\partial\alpha}=d_k^T(Gx_k-b)+\alpha d_k^T G d_k=0$$

$$\alpha=\frac{-d_k^T g_k}{d_k^T G d_k}.$$

线性共轭梯度: $d_k=-g_k+\frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}d_{k-1}$.

$$d_k^T g_k=-g_k^T g_k+\frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}d_{k-1}^T g_k, \quad d_{k-1}^T g_k=0$$

$$=-g_k^T d_k$$

$$\therefore \alpha=\frac{g_k^T g_k}{d_k^T G d_k}.$$

2022期末

中国人民大学统计学院《最优化方法》（02班）期末试题

$\frac{x^{(k+1)}-0}{(x^{(k)}-0)^2}=(0.5)^{2^{k+1}-2^{k+1}}=0.5^0=1$

考试时间：2小时

一、填空问答题（共十题，每题4分）

(1) 序列$x^{(k)}=(0.5)^{2^k}$的收敛速度为<u>二次收敛</u>。

(2) 已知$x\in R^2$，有二次函数$f(x)=\frac{1}{2}x^TGx+b^Tx+c$。请绘制出当$G$为正定矩阵、半正定矩阵、负定矩阵和不定矩阵时，$f(x)$大致的函数图像。

$-g_k$

(3) 当使用负梯度方法进行寻优的过程中，我们发现相邻两个迭代步的前进方向并不垂直，由此我们可以推断出该负梯度方法中的步长采用了<u>非精确</u>线搜索准则。

(4) 原点$x^{(0)}=(0,0)^T$到凸集$S=\{x\mid x_1+x_2\ge 4,2x_1+x_2\ge 5\}$的最小距离为<u>$2\sqrt{2}$</u>。

$(-2d_2-\frac{3}{2}d_3,d_2,d_3)$

(5) 写出向量$(2,4,3)^T$关于单位方阵的全部(两个)线性无关的共轭向量<u>$(-2,1,0),(-\frac{3}{2},0,1)$</u>

(6) 使用<u>Wolfe/精确</u>线搜索准则的DFP方法和BFGS方法，可以保证B_{k+1}、H_{k+1}的对称正定性。

(7) 设G是对阵正定矩阵，若非零向量组$\{d_0,d_1,\cdots,d_l\}$满足$d_i^TGd_j=0,\ i\neq j$，则称这个非零向量组是矩阵G的共轭方向。矩阵G的特征向量是共轭方向吗？<u>是</u>。共轭方向一定为特征向量方向吗？<u>不是</u>矩阵G的共轭方向有多少组？<u>无穷组</u>

(8) Broyden族拟牛顿方法是共轭梯度方法吗？<u>是</u>

(9) 带约束的N维优化问题，在任意点上最多有<u>N</u>个约束条件起作用。

$f(x)=x_1^2+x_2^2$

(10) 对一个有三个变量的函数$f(x)$使用FR共轭梯度法进行优化（即求解$\min f(x)$），第1次迭代，搜索方向为$d^{(1)}=(1,-1,2)^T$，沿$d^{(1)}$作精确线搜索，得到点$x^{(2)}$，又设 其

$g_1=(-1,1,-2)$,

$$\frac{\partial f(x^{(2)})}{\partial x_1}=-2,\frac{\partial f(x^{(2)})}{\partial x_2}=-2$$

$g_2=(-2,-2,x)$,

则按照共轭梯度法的规定，从$x^{(2)}$出发的搜索方向为<u>$(2,2,0)$</u>。

$d_2=-g_2+\frac{g_2^Tg_2}{g_1^Tg_1}d_1=(2,2,0)$

$g_2d_1=0\Rightarrow(-2,-2,x)(1,-1,2)=-2+2+2x=0\Rightarrow x=0$

二、证明题(共两题，每题 10 分)

(1) 若存在常数$m>0$，使得$g(x)=f(x)-\frac{m}{2}\|x\|_2^2$为凸函数，则称$f(x)$为强凸函数，证明强凸函数以下两条性质：

$$f(x)=g(x)+\frac{m}{2}x^Tx$$

(a) $\forall x,y\in dom\, f$以及$\theta\in(0,1)$有

$$f(\theta x+(1-\theta)y)\le\theta f(x)+(1-\theta)f(y)-\frac{m}{2}\theta(1-\theta)\|x-y\|_2^2.$$

(b) $\forall x,x'\in domf$，下述不等式成立：

$$\|\nabla f(x)-\nabla f(x')\|_2\ge m\|x-x'\|_2$$

(a) $f(x)=g(x)+\frac{m}{2}x^Tx$

$$f(\theta x+(1-\theta)y)=g(\theta x+(1-\theta)y)+\frac{m}{2}\|\theta x+(1-\theta)y\|^2$$

$$\le\theta g(x)+(1-\theta)g(y)+\frac{m}{2}\|\theta x+(1-\theta)y\|^2$$

$$=\theta\left(g(x)+\frac{m}{2}\|x\|^2\right)+(1-\theta)\left(g(y)+\frac{m}{2}\|y\|^2\right)$$

$$+\frac{m}{2}\|\theta x+(1-\theta)y\|^2-\frac{\theta m}{2}\|x\|^2-\frac{(1-\theta)m}{2}\|y\|^2$$

$$=\theta f(x)+(1-\theta)f(y)+\frac{m}{2}\left(\|\theta x+(1-\theta)y\|^2-\theta\|x\|^2-(1-\theta)\|y\|^2\right)$$

$$=\square+\frac{m}{2}\left(\theta(\theta-1)\|x\|^2+\theta(\theta-1)\|y\|^2+2\theta(1-\theta)x^Ty\right).$$

$$=\square-\frac{m}{2}\left((1-\theta)\theta\|x\|^2+(1-\theta)\theta\|y\|^2-2\theta(1-\theta)x^Ty\right),$$

$$=\square-\frac{m}{2}.(1-\theta)\theta\left(\|x-y\|^2\right),$$

$\square$

(b) $(\nabla f(x)-\nabla f(y))^T(x-y)$

$$= (\nabla g(x)-\nabla g(y)+m(x-y))^T(x-y)$$

$$= (\nabla g(x)-\nabla g(y))^T(x-y)+m\|x-y\|^2$$

$g(x)$ 凸 $\Rightarrow$

$$\boxed{\begin{aligned} g(x) &\ge g(y)+\nabla g(y)^T(x-y) \\ g(y) &\ge g(x)+\nabla g(x)^T(y-x) \end{aligned}}$$

公式记错了!

$$\Rightarrow \quad 0 \ge (\nabla g(y)-g(x))^T(x-y)$$

$$\Rightarrow \quad [\nabla g(x)-\nabla g(y)]^T(x-y) \ge 0$$

$$\therefore \quad (\nabla f(x)-\nabla f(y))^T(x-y) \ge m|x-y|^2$$

而

$$|\nabla f(x)-\nabla f(y)||x-y| \ge |(\nabla f(x)-\nabla f(y))^T(x-y)|$$

$$\ge m|x-y|^2$$

$$\Rightarrow \quad |\nabla f(x)-\nabla f(y)| \ge m|x-y|$$

三、 计算题(共四题，每题 10 分)

(1) 给定初始点$x^{(1)}=(0,0)^T$,使用最速下降法求解如下优化问题(使用精确线搜索,至少算至$x^{(4)}$)

$$\min f(x)=x_1-x_2+2x_1^2+2x_1x_2+x_2^2.$$

(2) 使用共轭梯度法求解如下正定二次函数的优化问题,初始点为$x_0=(0,0)^T$

$$\min f(x)=\frac{3}{2}x_1^2+\frac{1}{2}x_2^2-x_1x_2-2x_1.$$

(1) $g(x)=\begin{pmatrix}1+4x_1+2x_2\\-1+2x_1+2x_2\end{pmatrix}$ $\qquad \underline{x^{(1)}=(0,0)}$ $\qquad x_1=-1\quad x_2=\frac{3}{2}$

$g^{(1)}=(1,-1)^T,\ d^{(1)}=-g^{(1)}=(-1,1)^T$

$\phi(\alpha)=f(x^{(1)}+\alpha d^{(1)})=f(-\alpha,\alpha)=-2\alpha+\cancel{2\alpha^2}-\cancel{2\alpha^2}+\alpha^2=\alpha^2-2\alpha\Rightarrow\alpha=1$

$\therefore x^{(2)}=(0,0)+1\cdot(-1,1)=\underline{(-1,1)}$

$g^{(2)}=(-1,-1)\quad d^{(2)}=(1,1).$

$\phi(\alpha)=f(x^{(2)}+\alpha d^{(2)})=f(-1+\alpha,1+\alpha)=-2+2(\alpha-1)^2+2(\alpha^2-1)+(\alpha+1)^2$

$=5\alpha^2-2\alpha-1$

$\Rightarrow\alpha^*=\frac{1}{5},\quad x^{(3)}=(-1,1)+\frac{1}{5}(1,1)=\underline{(-\frac{4}{5},\frac{6}{5})}$

$\vdots$

(2) $g(x) = \begin{pmatrix} 3x_1 - x_2 - 2 \\ x_2 - x_1 \end{pmatrix}$ $\quad x_1^* = x_2^* = 1$ $\quad \boxed{(1,1)}$ 最优点…

$G(x) = \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix}$

$d_0 = -g_0 = (2,0)$

$x_1 = x_0 + \alpha_0 d_0 = (2\alpha, 0) \quad \phi(\alpha) = 6\alpha^2 - 4\alpha \Rightarrow \alpha = \frac{1}{3}$.

$\therefore \underline{\underline{x_1 = (\frac{2}{3}, 0)}} \quad g_1 = (0, -\frac{2}{3})$

$d_1^T G d_0 = (x.y) \begin{pmatrix} 3 & -1 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} 2 \\ 0 \end{pmatrix} = (3x-y, -x+y) \begin{pmatrix} 2 \\ 0 \end{pmatrix}$

$= 6x - 2y = 0 \Rightarrow y = 3x$

$d_1 = (\frac{2}{9}, \frac{2}{3}), = (1,3)$,

$x_2 = x_1 + \alpha d_1 = (\frac{2}{3} + \alpha, 3\alpha) \quad \phi(\alpha) = \frac{3}{2}(\frac{2}{3}+\alpha)^2 + \frac{1}{2}(3\alpha)^2 - (\frac{2}{3}+\alpha)3\alpha$

$- 2(\frac{2}{3}+\alpha)$

$\phi'(\alpha) = 3(\frac{2}{3}+\alpha) + 9\alpha - 2 - 6\alpha - 2$

$= \cancel{2} + 3\alpha + 9\alpha - \cancel{2} - 6\alpha - 2 = 6\alpha - 2 = 0 \quad \alpha = \frac{1}{3}$.

$\boxed{x_2 = (1, 1)}$ 共轭梯度法 二次终止性

(4) 用增广拉格朗日函数方法求解如下优化问题:

$$\min x_1 + \frac{1}{3}(x_2+1)^2$$
$$s.t.\, x_1 \ge 0, x_2 \ge 1$$

不等式约束 $C_1(X)=x_1.\quad C_2(X)=x_2-1$

增广 Lagrangian 函数为

$$L = x_1+\frac{1}{3}(x_2+1)^2 - \sum_{i\in E}\lambda_i c_i(x) + \frac{\sigma}{2}\sum_{i\in E}(c_i(x))^2 + \sum_{i\in I}\phi_i$$

$$= x_1+\frac{1}{3}(x_2+1)^2 + \frac{1}{2\sigma}\left(\max(0, \lambda_1-\sigma x_1)^2-\lambda_1^2\right) + \frac{1}{2\sigma}\left(\max(0, \lambda_2-\sigma(x_2-1))^2-\lambda_2^2\right)$$

$$= x_1+\frac{1}{3}(x_2+1)^2 + \frac{1}{2\sigma}\left(\max\{0,\lambda_1-\sigma x_1\}^2+\max\{0,\lambda_2-\sigma(x_2-1)\}^2-\lambda_1^2-\lambda_2^2\right)$$

$$\frac{\partial L}{\partial x_1} = 1+\frac{1}{2\sigma}\left(\frac{\partial \max(0,\lambda_1-\sigma x_1)^2}{\partial x_1}\right) = \begin{cases} 1-\lambda_1+\sigma x_1 & x_1 \le \frac{\lambda_1}{\sigma} \\ 1 & x_1 > \frac{\lambda_1}{\sigma}\end{cases}$$

$$\frac{\partial L}{\partial x_2} = \frac{2}{3}(x_2+1)+\frac{1}{2\sigma}\left(\frac{\partial \max(0, \lambda_2-\sigma(x_2-1))^2}{\partial x_2}\right) = \begin{cases} \left(\frac{2}{3}+\sigma\right)x_2+\frac{2}{3}-\lambda_2-\sigma & x_2-1 \le \frac{\lambda_2}{\sigma} \\ \frac{2}{3}(x_2+1) & x_2-1 > \frac{\lambda_2}{\sigma}\end{cases}$$

当 $x_1 > \frac{\lambda_1}{\sigma}$ 时 $\frac{\partial L}{\partial x_1} \ne 0$，不可能为最优解，只能 $x_1 \le \frac{\lambda_1}{\sigma}$

此时 $\frac{\partial L}{\partial x_1} = 1-\lambda_1+\sigma x_1 = 0 \Rightarrow \boxed{x_1 = \frac{\lambda_1^{(k)}-1}{\sigma}}$

① 若 $x_2-1 > \frac{\lambda_2}{\sigma}$，则 $\frac{\partial L}{\partial x_2} = \frac{2}{3}(x_2+1)=0 \Rightarrow x_2=-1 \quad \lambda_2 < \sigma(x_2-1) < -2\sigma < 0$ ✗

只能 $x_2-1 \le \frac{\lambda_2}{\sigma}$，则有 $\boxed{x_2 = \frac{\sigma+\lambda_2-\frac{2}{3}}{\frac{2}{3}+\sigma} = \frac{3\sigma+3\lambda_2^{(k)}-2}{2+3\sigma}}$

乘子修正 $\lambda_1^{(k+1)} = \max\{0, \lambda_1^{(k)}-\sigma C_1(x_k)\} = \max\{0, \lambda_1^{(k)} - \cancel{\sigma}\cdot\frac{\lambda_1-1}{\cancel{\sigma}}\}$

$= 1.$

$\lambda_2^{(k+1)} = \max\{0, \lambda_2^{(k)} - \sigma C_2(x_k)\} = \max\{0, \lambda_2^{(k)} - \sigma(x_2-1)\}$

$= \max\{0, \lambda_2^{(k)} - \sigma\frac{\cancel{3\sigma}+3\lambda_2-2-\cancel{3\sigma}-2}{3\sigma+2}\}$

$$=\max\left(0, \frac{2(\lambda_2+2\sigma)}{3\sigma+2}\right)$$

$$x = \frac{2(x+2\sigma)}{3\sigma+2} \Rightarrow (3\sigma+2)x = 2x+4\sigma$$

$$3\sigma x = 4\sigma \quad x = \frac{4}{3}$$

$\therefore \lambda_2^* \to \frac{4}{3}$. $\lambda_1^* \to 1$.

那么 $(x_1^*, x_2^*) = (0,1)$

（3） 采用基于 KKT 的方法，求解如下带约束的优化问题

$$\min x_1^2 + x_2^2$$

$$\text{s.t. } x_1 + x_2 = 1, \qquad x_2 \le \alpha,$$

其中$(x_1, x_2) \in R^2$，α为实数。

$$L = x_1^2 + x_2^2 - \lambda_1(x_1 + x_2 - 1) - \lambda_2(\alpha - x_2),$$

$$\begin{cases} \dfrac{\partial L}{\partial x_1} = 2x_1 - \lambda_1 = 0 \\ \dfrac{\partial L}{\partial x_2} = 2x_2 - \lambda_1 + \lambda_2 = 0 \\ \lambda_1(x_1 + x_2 - 1) = 0 \\ \lambda_2(\alpha - x_2) = 0 \end{cases} \Rightarrow \left| \begin{array}{l} x_1 = 1-\alpha. \\ x_2 = \alpha \\ \lambda_1 = 2(1-\alpha) \ge 0 \\ \lambda_2 = 2(1-2\alpha) \ge 0 \quad \boxed{\alpha \le \frac{1}{2}}. \end{array} \right. \quad \left| \begin{array}{l} x_1 = 1/2 \\ x_2 = 1/2 \\ \lambda_1 = 1 \\ \lambda_2 = 0 \quad (这要求 \alpha \ge 1/2) \end{array} \right.$$

$$\nabla^2 L = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} 正定 \quad \therefore d^T \nabla^2 L d \text{ 在 } (1-\alpha, \alpha) > 0,$$

$$\therefore \min \ \underline{(1-\alpha, \alpha)}.$$

当$\alpha \ge \frac{1}{2}$时. 最优解 $(1/2, 1/2)$,

当$\alpha < \frac{1}{2}$时 最优解 $(1-\alpha, \alpha)$.

Textbook Problem

Chap. 2

11. 证明: 若 $\rho < 1/2$, 则正定二次函数精确线搜索的步长满足 Goldstein 准则.

$f(x) = \frac{1}{2}x^T G x + b^T x$

$$\alpha_k = \frac{-d_k^T g_k}{d_k^T G d_k} \qquad 0 < \rho < \frac{1}{2}.$$

Goldstein:

$$f(x_k + \alpha d_k) \leq f(x_k) + \rho g_k^T d_k \alpha$$

$$f(x_k + \alpha d_k) \geq f(x_k) + (1-\rho) g_k^T d_k \alpha$$

对正定二次函数.

$$f(x_k + \alpha_k d_k) - f(x_k) = \alpha_k g_k^T d_k + \frac{1}{2}\alpha_k^2 d_k^T G d_k.$$

$$= \alpha_k g_k^T d_k + \frac{1}{2}\frac{(d_k^T g_k)^2}{(d_k^T G d_k)^2} \cdot \cancel{d_k^T} G d_k.$$

$$= \alpha_k g_k^T d_k + \frac{1}{2}(-\alpha_k) \cdot g_k^T d_k$$

$$= \left(\frac{1}{2}\right) g_k^T d_k \cdot \alpha_k$$

$\because g_k^T d_k < 0,\ 0 < \rho < 1$

$$\therefore (1-\rho) g_k^T d_k \cdot \alpha < f(x_k + \alpha_k d_k) - f(x_k) < \rho g_k^T d_k \cdot \alpha.$$

Chap 3

4. 考虑函数

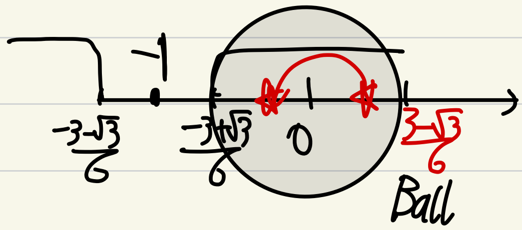
$$f(x)=2x_1^2+x_2^2-2x_1x_2+2x_1^3+x_1^4,$$

确定关于点 $x^*=(0,0)^{\mathrm{T}}$, 使 $G(x)$ 正定的最大开球. 问在此球中如何取初始点 $x^{(0)}$, 其中 $x_1^{(0)}=x_2^{(0)}$, 使基本 Newton 方法收敛.

$$g(x)=\begin{pmatrix}4x_1-2x_2+6x_1^2+4x_1^3\\ 2x_2-2x_1\end{pmatrix},\quad G(x)=\begin{pmatrix}12x_1^2+12x_1+4, & -2\\ -2 & 2\end{pmatrix}$$

$$\begin{cases}12x_1^2+12x_1+4>0\\ 2(12x_1^2+12x_1+4)-4>0\end{cases}\Rightarrow\begin{cases}\text{恒成立}\\ x_1<\frac{-3-\sqrt{3}}{6}\quad x_1>\frac{-3+\sqrt{3}}{6}\end{cases}$$

∴最大开球半径为 $\frac{3-\sqrt{3}}{6}$



f的最优点为 $(0,0)$ $(-1,-1)$

10. 假定 $s_k^{\mathrm{T}}y_k>0$, 并且 H_k 正定. 证明: 对称秩 1 公式属于 Broyden 族公式类, 但其 $\varphi\notin[0,1]$.

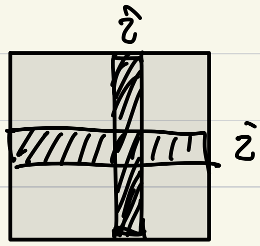
多次计算无果,遂放弃

$$B_{k+1}^{SR1}=\phi_k B_{k+1}^{BFGS}+(1-\phi_k)B_{k+1}^{DFP}$$

$$\phi_k=\frac{s_k^T y_k}{s_k^T y_k-s_k^T B_k s_k}$$

14. 证明: 对于 BFGS 方法, 如果矩阵 H_0 的第 i 行与第 i 列为零, 则所有 H_k 的第 i 行与第 i 列均为零, 并且 $x_i^{(k)} = x_i^{(0)}$.

BFGS: $\overset{BFGS}{H_{k+1}} = \left(I - \frac{s_k y_k^T}{y_k^T s_k}\right) H_k \left(I - \frac{y_k s_k^T}{y_k^T s_k}\right) + \frac{s_k s_k^T}{y_k^T s_k}$



这时看 $H_0 e_i = 0$, $e_i^T H_0 = 0$

$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix}$ 第i行　　$(0 \dots 1 \dots 0)$ 第i列

$x_{k+1} = x_k - \alpha_k H_k g_k \Rightarrow x_k = x_0 - \sum_{i=0}^{k-1} \alpha_i H_i g_i$ 而 $x_i^{(k)} = x_k^T e_i$

要证明 $H_k e_i = 0$　$e_i^T H_k = 0$,　$x_k^T e_i = x_0^T e_i$　(数归. 设k时成立)

① $H_0 e_i = 0$, $e_i^T H_0 = 0$, $x_0^T e_i = x_0^T e_i$, ✓

② 设原命题在k时成立 对k+1时

$$H_{k+1} = H_k - \frac{H_k y_k s_k^T}{y_k^T s_k} - \frac{s_k y_k^T H_k}{y_k^T s_k} + \frac{s_k y_k^T H_k y_k s_k^T}{(y_k^T s_k)^2} + \frac{s_k s_k^T}{y_k^T s_k}$$

$$= H_k + \frac{s_k s_k^T - 2 H_k y_k s_k^T}{y_k^T s_k} + \frac{s_k y_k^T H_k y_k s_k^T}{(y_k^T s_k)^2}$$

$$H_{k+1} e_i = H_k e_i + \frac{s_k s_k^T e_i - 2 s_k y_k^T H_k e_i}{y_k^T s_k} - \frac{s_k y_k^T s_k y_k^T H_k e_i}{(y_k^T s_k)^2}$$

$$= 0 + \frac{(-\alpha_k H_k g_k)(-\alpha_k H_k g_k)^T e_i - 0}{y_k^T s_k} - 0 = \frac{\alpha_k^2 H_k g_k g_k^T H_k e_i}{y_k^T s_k} = 0$$

$$e_i^T H_{k+1} = e_i^T H_k + \frac{\alpha_k^2 (e_i^T H_k) g_k g_k^T H_k - 2 (e_i^T H_k) y_k s_k^T}{y_k^T s_k} - \frac{(e_i H_k) y_k s_k^T y_k s_k^T}{(y_k^T s_k)^2}$$

$$= 0$$

$$x_{k+1}^T e_i = (x_k - \alpha_k H_k g_k)^T e_i = x_k^T e_i - \alpha_k g_k^T (H_k e_i) = x_k^T e_i = x_0 e_i$$

③ 对 $\forall k \in \mathbb{N}^*$, 原命题成立

15. 利用对称矩阵迹的性质，证明：

$$\text{trace}(B_{k+1}^{\text{BFGS}}) = \text{trace}(B_k) - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k} + \frac{\|y_k\|^2}{y_k^T s_k}.$$

$$a=(a_1 \dots a_n)^T \quad b=(b_1 \cdots b_n)^T$$

$$B_{k+1}^{BFGS} = B_k + \frac{y_k y_k^T}{y_k^T s_k} - \frac{B_k s_k s_k^T B_k}{s_k^T B_k s_k}.$$

对称矩阵且 $ab^T = \begin{pmatrix} a_1 b_1 & \cdots & a_1 b_n \\ \vdots & & \vdots \\ a_n b_1 & \cdots & a_n b_n \end{pmatrix}$

$$\Rightarrow tr(ab^T) = \sum a_i b_i = a^T b$$

$$tr(B_{k+1}^{BFGS}) = tr(B_k) + \frac{y_k^T y_k}{y_k^T s_k} - \frac{(B_k s_k)^T (B_k s_k)}{s_k^T B_k s_k}$$

$$= tr(B_k) + \frac{\|y_k\|^2}{y_k^T s_k} - \frac{\|B_k s_k\|^2}{s_k^T B_k s_k}$$

16. 对 Broyden 族公式中的矩阵 H_{k+1}^{φ}，考虑下列问题：

(1) 求出使 H_{k+1}^{φ} 奇异的 φ，记为 $\bar{\varphi}$. 若 H_k 正定，用 Cauchy-Schwarz 不等式证明 $\bar{\varphi} < 0$.

(2) 由 H_{k+1}^{φ} 求出 B_{k+1}^{φ}：

$$B_{k+1}^{\varphi} = B_{k+1}^{\text{DFP}} + (\theta - 1) w w^T = B_{k+1}^{\text{BFGS}} + \theta w w^T,$$

其中

$$w = (s_k^T B_k s_k)^{\frac{1}{2}} \left(\frac{y_k}{s_k^T y_k} - \frac{B_k s_k}{s_k^T B_k s_k} \right),$$

$$\theta = (\varphi - 1)/(\varphi - 1 - \varphi\mu),$$

$$\mu = (y_k^T H_k y_k)(s_k^T B_k s_k)/(s_k^T y_k)^2.$$

$$H_{k+1}^{\varphi} = \varphi H_k^{BFGS} + (1-\varphi) H_k^{DFP} = H_k^{DFP} + \varphi (H_k^{BFGS} - H_k^{DFP})$$

$$= H_k^{DFP} + \varphi v_k v_k^T$$

(1)

根据 Sherman-Morrison-Woodbury 公式

$$H_{k+1}^{\varphi} = H_{k+1}^{DFP} + \varphi_k v_k v_k^T \text{ 可逆} \Leftrightarrow 1 + \varphi v_k^T H_{k+1}^{DFP} v_k \neq 0$$

∴若 H_{k+1}^{φ} 奇异，则 $\bar{\varphi} = -\frac{1}{v_k^T H_{k+1}^{DFP} v_k}$

H_k 正定时 $H_{k+1}^{DFP} = H_k + \frac{s_k s_k^T}{y_k^T s_k} - \frac{H_k y_k y_k^T H_k}{y_k^T H_k y_k}$

$$\forall x \in \mathbb{R},\ x^T H_{k+1}^{DFP} x = x^T H_k x + \frac{(x^T s_k)^2}{y_k^T s_k} - \frac{(x^T H_k y_k)^2}{y_k^T H_k y_k}$$

$$= \frac{(x^T H_k x)(y_k^T H_k y_k) - (x^T H_k y_k)^2}{y_k^T H_k y_k} + \frac{(x^T s_k)^2}{y_k^T s_k} \geq \frac{(x^T s_k)^2}{y_k^T s_k} \geq 0$$

(Cauchy)

∴ H_{k+1} 正定 从而 $\bar{\varphi} = -\frac{1}{v_k^T H_{k+1}^{DFP} v_k} < 0$

(2) 根据 Shermann-Morrison-Woodbury 公式.

$$B_{k+1}^{\varphi}=\left(H_{k+1}^{\varphi}\right)^{-1}=\left(H_{k+1}^{DFP}+\varphi v_k v_k^T\right)^{-1}$$

其中 $v_k=(y_k^T H_k y_k)^{1/2}\left(\frac{s_k}{s_k^T y_k}-\frac{H_k y_k}{y_k^T H_k y_k}\right)$

$$=\left(H_{k+1}^{DFP}\right)^{-1}-\frac{1}{\sigma}\left(H_{k+1}^{DFP}\right)^{-1}\varphi v_k v_k^T\left(H_{k+1}^{DFP}\right)^{-1}$$，其中 $\sigma=1+\varphi v_k^T B_{k+1}^{DFP} v_k$

$$=B_{k+1}^{DFP}-\frac{1}{1+\varphi v_k^T B_{k+1}^{DFP} v_k}\left(B_{k+1}^{DFP}\cdot\varphi v_k v_k^T\cdot B_{k+1}^{DFP}\right)$$

$$=?$$

桧，略

17. 考虑线性变换 $\hat{x}=Wx+u$, 其中 W 非奇异. 对于一种方法, 若从 $\hat{x}_k=Wx_k+u$ 可得 $\hat{x}_{k+1}=Wx_{k+1}+u$, 则称该方法在此线性变换下是不变的. 讨论负梯度方法, 带固定步长的 Newton 方法, DFP 方法与 BFGS 方法是否具有不变性.

$x=W^{-1}(\hat{x}-u)$　　$f(\hat{x})=f(W^{-1}(\hat{x}-u))$　　$g(\hat{x})=(W^{-1})^T g(W^{-1}(\hat{x}-u))=(W^{-1})^T g(x)$

$$G(\hat{x})=(W^{-1})^T G(W^{-1}(x-u))(W^{-1})=(W^{-1})^T G(x)(W^{-1})$$

即　$g(\hat{x})=W^{-1}g(x)$,　$G(\hat{x})=(W^{-1})G(x)(W^{-1})^T$

负梯度法: $\hat{x}_{k+1}=\hat{x}_k-\alpha g(\hat{x}_k)=Wx_k+u-\alpha\cdot(W^{-1})^T g(x_k)\neq W(x_k-\alpha g_k)+u$

∴ 负梯度法不具不变性

基本Newton法: $\hat{x}_{k+1}=\hat{x}_k-G^{-1}(\hat{x}_k)g(\hat{x}_k)=\hat{x}_k-WG(x_k)\cancel{W^T}\cancel{(W^{-1})^T}g(x_k)$

$=\hat{x}_k-WG(x_k)g(x_k)$

由于 $\hat{x}_k=Wx_k+u$, 因此 $\hat{x}_{k+1}=Wx_k+u-WG_kg_k$

$=W(x_k-G_kg_k)+u=Wx_{k+1}+u$

∴ 基本 Newton 法具不变性

BFGS/DFP: $\hat{x}_{k+1}=\hat{x}_k-\hat{H}_k\hat{g}_k=\hat{x}_k-WH_k\cancel{W^T}\cancel{(W^{-1})^T}g_k$

$=W(x_k-H_kg_k)+u=Wx_{k+1}+u$

∴ BFGS/DFP 具不变性.

Chap 4

定理 3.2 设 $G \in \mathbb{R}^{n\times n}$ 对称正定, 则对任给的 $u, v \in \mathbb{R}^n$, G 度量意义下的 Cauchy-Schwarz 不等式

$$|u^{\mathrm{T}}Gv| \leqslant \|u\|_G\|v\|_G$$

成立, 当且仅当 u, v 共线时等式成立.

该定理可以由通常度量意义下的 Cauchy-Schwarz 不等式导出, 定理的证明作为第四章的习题.

$$(u^TGu)^2(v^TGv)^2 \geq (u^TGv)^2 \Rightarrow \|u\|_G^2\|v_G\|^2 \geq |u^TGv|^2$$
$$\Rightarrow \|u^TGv\| \leq \|u\|_G\|v_G\|$$

定理 4.2 共轭向量组中的向量一定线性无关.

$d_0 \ldots d_{n-1}$ 中若线性相关, 不妨设 $d_i = \beta_0 d_0 + \cdots + \beta_{i-1}d_{i-1} + \beta_{i+1}d_{i+1} + \cdots + \beta_{n-1}d_{n-1}$. 则

$(i \neq j)$

$d_i^T G d_j = \beta_j d_j^T G d_j > 0$. 与 $d_i^T G d_j = 0$ 矛盾!

定理 4.4 (平方根矩阵的存在唯一性) 若 G 是 $n \times n$ 对称正定矩阵, 则 G 的平方根矩阵存在唯一.

所以 $\lambda_i > 0$

$G = Q\Lambda Q^T = (Q\Lambda_1 Q^T)(Q\Lambda_1 Q^T)$, 其中 $\Lambda_1 = \Lambda^{1/2}$

$\therefore$ 取 $\sqrt{G} = Q\Lambda_1 Q^T$ 即可

4. 设 G 为三对角阵, 其对角元均为 2, 次对角元均为 -1. 证明: 向量

$$d_i = (1, 2, \cdots, \underset{i}{i+1}, 0, \cdots, 0)^{\mathrm{T}} \in \mathbb{R}^n, \quad i = 0, \cdots, n-1$$

为 n 个 G 共轭方向.

$(1, 0, \ldots 0) = d_0$
$(1, 2, \ldots 0) = d_1$
$\vdots$
$(1, 2, \ldots, n) = d_{n-1}$.

$$G = \begin{pmatrix} 2 & -1 & & & & \\ -1 & 2 & -1 & & & \\ & -1 & 2 & -1 & & \\ & & -1 & 2 & -1 & \\ & & & \ddots & 2 & -1 \\ & & & & -1 & 2 \end{pmatrix}$$

$\forall i \neq j$ $\quad d_{n-1}^T G = (1, 2 \ldots n)\begin{pmatrix} 2 & -1 & & & \\ -1 & 2 & -1 & & \\ & -1 & 2 & -1 & \\ & & \ddots & 2 & -1 \\ & & & -1 & 2 \end{pmatrix} = (0, 0, \ldots, n+1)$

$d_{n-2}^T G = (1, 2 \ldots n-1, 0)G = (0, 0, \ldots, n, -(n-1))$

$d_{n-3}^T G = (1, 2, \ldots n-2, 0, 0)G = (0, 0, \ldots, n-1, -(n-2), 0)$

$\vdots$

$d_i^T G = (0, 0, \ldots, i+2, -(i+1), 0 \ldots 0)$, $\quad i \leq n-2$

$\forall i \neq j$ 不妨设 $i>j$

① $j<i\le n-2$. $d_i^T G d_j = (0,0,\cdots,i+2,-(i+1),0\cdots 0)\begin{pmatrix}\vdots\\ j+1\\ \vdots\\ 0\end{pmatrix}=0$.

② $j\le n-2<i$ $d_i^T G d_j = (0,\cdots,n+1)\begin{pmatrix}\vdots\\ j+1\\ \vdots\\ 0\end{pmatrix}=0$,

$\therefore \{d_0,\dots d_{n-1}\}$ 为 n 个共轭方向

5. 设 G 为具有不同特征值的对称正定矩阵. 证明: G 的特征向量是 G 共轭的.

$G=Q\Lambda Q^T$, $Q=(g_1,\dots,g_n)$ $\Lambda=\mathrm{diag}(\lambda_1,\dots,\lambda_n)$ $(\lambda_i\neq\lambda_j)$

$g_i^T G g_j = g_i^T\cdot\lambda g_j = \lambda g_i^T g_j = 0$ $(i\neq j)$ $\therefore$ G共轭

6. 将共轭梯度方法用于正定二次函数. 证明: 若在点 x_m 处迭代终止, 则序列

$$g_0,\ Gg_0,\ G^2g_0,\ \cdots$$

(应用性质)

中线性无关的向量个数为 m.

共轭梯度法 $\Rightarrow$ 正定二次函数 $\Rightarrow$ 线性共轭梯度法

由线性共轭梯度法性质. $\mathrm{span}\{g_0, Gg_0, G^2g_0\}=\mathrm{span}\{d_0, d_1, d_2\dots\}$
由于 x_m处迭代终止 因此 共有 m个共轭梯度方向 $\Rightarrow \mathrm{span}\{d_0,d_1,d_2,\dots\}$的基有 m个 $\Rightarrow \mathrm{span}\{g_0, Gg_0, G^2g_0,\dots\}$ 只有 m个线性无关向量

7. 设 G 是 $n\times n$ 正定对称矩阵. 对 $\mathbb{R}^n$ 中任意一组线性无关向量 $\{p_0,\cdots,p_{n-1}\}$, Gram-Schmidt 过程产生一组向量

$$d_0=p_0, \tag{4.18}$$

$$d_k=p_k-\sum_{i=0}^{k-1}\frac{p_k^TGd_i}{d_i^TGd_i}d_i,\quad k=1,\cdots,n-1. \tag{4.19}$$

证明: 向量 $d_0,d_1,\cdots,d_{n-1}$ 是 G 共轭的.

$n=2$时. $d_0^TGd_1 = p_0^TG\left(p_1-\frac{p_1^TGd_0}{d_0^TGd_0}\cdot d_0\right)=p_0^TGp_1-\frac{p_1^TGp_0}{\cancel{p_0^TGp_0}}\cdot \cancel{p_0^TGp_0}=0$

设$n=k$时, 成立 $d_i^TGd_j=0$ $(i\neq j,\ i,j=0,1,\dots k-1)$, 则

$n=k+1$时 $(\underbrace{d_0,\dots d_k}_{G共轭},\underbrace{d_{k+1}}_{?})$ 只需证明 $d_{k+1}^TGd_i=0$, $(i=0,\dots,k)$

$d_{k+1}^TGd_0=\left(p_{k+1}-\sum_{i=0}^{k}\frac{p_{k+1}^TGd_i}{d_i^TGd_i}d_i\right)^TGd_0$

$$= p_{k+1}^T G d_0 - \sum_{i=0}^{k} \left(\frac{p_{k+1}^T G d_i}{d_i^T G d_i} \cdot d_i^T G d_0 \right)$$ $i\neq 0$ 时 $d_i^T G d_0 = 0$

$$= p_{k+1}^T G d_0 - (p_{k+1}^T G d_0 + 0) = 0$$

$$d_{k+1}^T G d_j = \left(p_{k+1}^T - \sum_{i=0}^{k} \frac{p_{k+1}^T G d_i}{d_i^T G d_i} \cdot d_i^T\right) G d_j$$

$(j=1 \ldots k)$

$$= p_{k+1}^T G d_j - \sum_{i=0}^{k} \frac{p_{k+1}^T G d_i}{d_i^T G d_i} \cdot d_i^T G d_j$$

$$= p_{k+1}^T G d_j - \left(\frac{p_{k+1}^T G d_j}{\cancel{d_j^T G d_j}} \cdot \cancel{d_j^T G d_j} + 0\right) = 0$$

$\therefore$ 对 $\forall n \in \mathbb{N}^*$，Gram-Schmidt 过程产生的向量组都 G 共轭.

9. 证明：当采用强 Wolfe 线搜索并且 $\sigma < 1$ 时，用共轭下降公式得到的方向为下降方向.

$$d_k = -g_k + \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}} d_{k-1}$$

$$|g_{k+1}^T d_k| < -\sigma g_k^T d_k.$$

可以证明，当采用强 Wolfe 准则时

$$-\frac{1}{1-\sigma} < \frac{g_k^T d_k}{\|g_k\|^2} < \frac{2\sigma-1}{1-\sigma}$$ (数归) (见 FR 方法)

那么 $\sigma > 1$ 时，$\frac{2\sigma-1}{1-\sigma} < 0 \Rightarrow g_k^T d_k < 0$

引理 4.8 (Zoutendijk 条件) 设 $f(x)$ 有下界，$g(x)$ 满足 Lipschitz 条件，使用 Wolfe 线搜索准则或精确线搜索准则的，具有 $x_{k+1} = x_k + \alpha_k d_k$ 迭代格式的一般下降方法满足 Zoutendijk 条件：

$$\sum_{k\geqslant 0} \frac{(g_k^T d_k)^2}{\|d_k\|^2} < \infty.$$

引理的证明留为作业.

$\theta_k = \langle -g_k, d_k \rangle$

要证明

$$\sum_{k\geq 0} \frac{\|g_k\|^2 \|d_k\|^2 \cos^2\langle -g_k, d_k\rangle}{\|d_k\|^2} = \sum_{k\geq 0} \|g_k\|^2 \cos^2\theta_k < \infty,$$

Wolfe 准则 $\Rightarrow$ $g_{k+1}^T d_k > \sigma g_k^T d_k \Rightarrow (g_{k+1} - g_k)^T d_k > (\sigma-1) g_k^T d_k$

Lipschitz $\Rightarrow$ $\|g_{k+1} - g_k\| \leq L \|x_{k+1} - x_k\| = L \cdot \|\alpha_k d_k\| = L\alpha_k \|d_k\|$

$$\Rightarrow (\sigma-1) g_k^T d_k < \|g_{k+1} - g_k\| \cdot \|d_k\| \leq L\alpha_k \|d_k\|^2$$

$$\Rightarrow \alpha_k > \frac{(\sigma-1) g_k^T d_k}{L\|d_k\|^2} = \frac{\sigma-1}{L\|d_k\|^2} (g_k^T d_k).$$

代入 $f_{k+1} < f_k + \rho g_k^T d_k \cdot \alpha_k < f_k + \rho \frac{\sigma-1}{L} \|g_k\|^2 \cos^2\theta_k$

$$f_{k+1}-f_k < \rho\cdot\frac{\sigma-1}{L}(g_k^T d_k)^2 \Rightarrow \sum_{k=0}^{K}(f_{k+1}-f_k) < \sum_{k=0}^{K}\rho\cdot\frac{\sigma-1}{L}\|g_k\|^2\cos\theta_k^2$$

$$\Rightarrow f_K-f_0 < \rho\cdot\frac{\sigma-1}{L}\sum_{k\geq 0}\|g_k\|^2\cos\theta_k^2$$

$$\sum_{k\geq 0}\|g_k\|^2\cos\theta_k^2 < (f_0-f_k)\cdot\frac{1}{\rho(1-\sigma)} < \infty$$ □.

11. 考虑用 FR 方法解正定二次函数的极小化问题. 记

$$R_k = \left[\frac{-g_1}{\|g_1\|}, \frac{-g_2}{\|g_2\|}, \cdots, \frac{-g_k}{\|g_k\|}\right], \quad n\times k,$$

$$S_k = \left[\frac{d_1}{\|g_1\|}, \frac{d_2}{\|g_2\|}, \cdots, \frac{d_k}{\|g_k\|}\right], \quad n\times k,$$

$$B_k = \begin{bmatrix} 1 & & & & \\ -\sqrt{\beta_1} & 1 & & & \\ & -\sqrt{\beta_2} & 1 & & \\ & & \ddots & \ddots & \\ & & & -\sqrt{\beta_{k-1}} & 1 \end{bmatrix},$$

$$D_k = \begin{bmatrix} \alpha_1^{-1} & & & \\ & \alpha_2^{-1} & & \\ & & \ddots & \\ & & & \alpha_k^{-1} \end{bmatrix},$$

其中 α_i $(i=1,\cdots,k)$ 是精确线搜索的步长.

(1) 证明: $GS_kD_k^{-1} = R_kB_k + g_{k+1}e_k^T/\|g_k\|, S_kB_k^T = R_k$, 其中 G 是正定二次函数的 Hesse 矩阵, $e_k = [0,\cdots,0,1,0,\cdots,0]^T$.

(2) 证明: $R_k^TGR_k = T_k$, 其中 T_k 是三对角阵; 若迭代进行 n 步, T_n 与 G 有相同的特征值.

$$\beta_k = \frac{g_{k+1}^T g_{k+1}}{g_k^T g_k} = \frac{\|g_{k+1}\|^2}{\|g_k\|^2}$$

(1) ① $$\underbrace{\left(\frac{d_1}{\|g_1\|},\cdots,\frac{d_k}{\|g_k\|}\right)}_{1\times k}\underbrace{\begin{pmatrix}1 & -\sqrt{\beta_1} & & \\ & 1 & -\sqrt{\beta_2} & \\ & & \ddots & -\sqrt{\beta_{k-1}} \\ & & & 1\end{pmatrix}}_{k\times k} = \left(\frac{d_1}{\|g_1\|}, -\sqrt{\beta_1}\frac{d_1}{\|g_1\|}+\frac{d_2}{\|g_2\|}, \cdots, -\sqrt{\beta_{k-1}}\frac{d_{k-1}}{\|g_{k-1}\|}+\frac{d_k}{\|g_k\|}\right)$$

$$-\sqrt{\beta_{k-1}}\frac{d_{k-1}}{\|g_{k-1}\|}+\frac{d_k}{\|g_k\|} = -\frac{\|g_k\|}{\|g_{k-1}\|}\frac{d_{k-1}}{\|g_{k-1}\|}+\frac{1}{\|g_k\|}\left(-g_k+\frac{\|g_k\|^2}{\|g_{k-1}\|^2}d_{k-1}\right)$$

$$= \cancel{-\frac{\|g_k\|}{\|g_{k-1}\|^2}d_{k-1}}+\left(\frac{-g_k}{\|g_k\|}\right)+\cancel{\frac{\|g_k\|}{\|g_{k-1}\|^2}d_{k-1}} = \frac{-g_k}{\|g_k\|}$$

② 证 $S_kB_k^T = R_k$

(ㄱ) RHS :

$$R_kB_k = \left(\frac{-g_1}{\|g_1\|} + \sqrt{\beta_1}\frac{g_2}{\|g_2\|}, \ldots, \frac{-g_{k-1}}{\|g_{k-1}\|} + \sqrt{\beta_{k-1}}\frac{g_k}{\|g_k\|}, \frac{-g_k}{\|g_k\|}\right)$$

$$\frac{-g_k}{\|g_k\|} + \sqrt{\beta_k}\cdot\frac{g_{k+1}}{\|g_{k+1}\|} = \frac{-g_k}{\|g_k\|} + \frac{\cancel{\|g_{k+1}\|}}{\|g_k\|}\frac{g_{k+1}}{\cancel{\|g_{k+1}\|}} = \frac{1}{\|g_k\|}(g_{k+1} - g_k)$$

$$= \left(\frac{1}{\|g_1\|}(g_2 - g_1), \ldots, \frac{1}{\|g_{k-1}\|}(g_k - g_{k-1}), \frac{1}{\|g_k\|}(-g_k)\right)$$

$$\frac{g_{k+1}e_k^T}{\|g_k\|} + R_kB_k = \left(\frac{1}{\|g_1\|}(g_2 - g_1), \ldots, \frac{1}{\|g_{k-1}\|}(g_k - g_{k-1}), \frac{1}{\|g_k\|}(g_{k+1} - g_k)\right)$$

LHS :

$$GS_kD_k^{-1} = G\left(\frac{\alpha_1 d_1}{\|g_1\|}, \ldots, \frac{\alpha_k d_k}{\|g_k\|}\right)$$

$$= G\left(\frac{s_1}{\|g_1\|}, \ldots, \frac{s_k}{\|g_k\|}\right) = \left(\frac{g_2 - g_1}{\|g_1\|}, \ldots, \frac{g_{k+1} - g_k}{\|g_k\|}\right)$$

$\therefore$ RHS = LHS

(2)
$$\underbrace{\begin{pmatrix} \frac{-g_1^T}{\|g_1\|} \\ \vdots \\ \frac{-g_k^T}{\|g_k\|} \end{pmatrix}}_{k\times n} \underbrace{G}_{n\times n} \underbrace{\left(\frac{-g_1}{\|g_1\|}, \ldots, \frac{-g_k}{\|g_k\|}\right)}_{n\times k} = \underbrace{\begin{pmatrix} \frac{-g_1^TG}{\|g_1\|} \\ \vdots \\ \frac{-g_k^TG}{\|g_k\|} \end{pmatrix}}_{k\times n} \underbrace{\left(\frac{-g_1}{\|g_1\|}, \ldots, \frac{-g_k}{\|g_k\|}\right)}_{n\times k}$$

$$= \begin{pmatrix} \frac{g_1^TGg_1}{\|g_1\|^2}, & \frac{g_1^TGg_2}{\|g_2\|\|g_1\|}, & \ldots, & \frac{g_1^TGg_k}{\|g_1\|\|g_k\|} \\ \vdots & & & \\ \frac{g_k^TGg_1}{\|g_k\|\|g_1\|}, & \frac{g_k^TGg_2}{\|g_k\|\|g_2\|}, & \ldots, & \frac{g_k^TGg_k}{\|g_k\|^2} \end{pmatrix}_{n\times n}$$

$$g_i^T G g_i = \frac{1}{\alpha_i} g_i^T (g_{i+1} - g_i) = -\frac{1}{\alpha_i} g_i^T g_i < 0$$

$$g_{i+1}^T G g_i = \frac{1}{\alpha_i} g_{i+1}^T G (x_{i+1} - x_i) = \frac{1}{\alpha_i} g_{i+1}^T (g_{i+1} - g_i) = \frac{1}{\alpha_i} g_{i+1}^T g_{i+1} > 0$$

$$g_{i+k}^T G g_i = \frac{1}{\alpha_i} g_{i+k}^T (g_{i+1} - g_i) = 0 \quad (k \geq 2)$$

$$\therefore R_k^T G R_k = \begin{pmatrix} \square & \circ & & \\ \circ & \square & \ddots & \\ & \ddots & \ddots & \circ \\ & & \circ & \square \end{pmatrix}$$

其中 $\square < 0$, $\circ > 0$, 为三对角矩阵

$$T_n = \underset{k\times n}{R_n^T} \underset{n\times n}{G} \underset{n\times k}{R_n}$$

注意到 $R_n = \left(\frac{-g_1}{\|g_1\|}, \cdots \frac{-g_n}{\|g_n\|} \right) \in \mathbb{R}^{n\times n}$

为正交矩阵, 因此 $R_n^T = R_n^{-1}$

$T_n = R_n^{-1} G R_n \Rightarrow T_n \sim G$ (相似矩阵)

$\Rightarrow T_n$ 与 G 有着相同的特征值

12. Miele 与 Cantrell[54] 在 1969 年给出如下算法:

算法 4.2 (MC 共轭梯度方法)

步 1　给出 x_0, $\varepsilon > 0$, $k := 0$;

步 2　进行一步最速下降方法迭代, 得 $x_1 = x_0 + \alpha_0 d_0$, $s_0 = x_1 - x_0$. $k := 1$;

步 3　若终止条件满足, 则迭代停止;

$s_{k-1} = x_k - x_{k-1}$

步 4　求 $(\alpha_k, \beta_k) = \arg\min_{\alpha,\beta} f(x_k + \alpha d_k + \beta s_{k-1})$, 其中 $d_k = -g_k$;

步 5　$x_{k+1} := x_k + \alpha_k d_k + \beta_k s_{k-1}$, $s_k = x_{k+1} - x_k$, $k := k+1$, 转步 3.

$(1+\beta) x_k - x_{k-1} + \alpha d_k$

对该算法, 考虑下面几个问题:

(1) 证明:

$$g_{k+1}^T d_k = 0, \qquad k \geqslant 0, \tag{4.20a}$$

$$g_{k+1}^T s_{k-1} = 0, \qquad k \geqslant 1,$$

$$g_{k+1}^T s_k = 0, \qquad k \geqslant 0.$$

(2) 若 $f(x) = \frac{1}{2} x^T G x + b^T x$, 证明:

$$\alpha_k = -\frac{(g_k^T d_k)(s_{k-1}^T G s_{k-1})}{\Delta_k},$$

$$\beta_k = \frac{(g_k^T d_k)(d_k^T G s_{k-1})}{\Delta_k},$$

$$f_{k+1} - f_k = \frac{1}{2} \alpha_k g_k^T d_k,$$

其中 $\Delta_k = (d_k^T G d_k)(s_{k-1}^T G s_{k-1}) - (d_k^T G s_{k-1})^2$.

(3) 证明: 当目标函数是正定二次函数时, 该方法与 FR 方法等价.

(3) $x_{k+1}^{MC} = x_k^{MC} - \alpha_k g_k + \beta_k^{MC}(x_k - x_{k-1})$,

$$x_{k+1}^{FR} = x_k^{FR} + \alpha_k^{FR}(-g_k + \beta_{k-1}^{FR} d_{k-1}^{FR})$$
$$= x_k^{FR} - \alpha_k^{FR} g_k + \alpha_k^{FR}\beta_{k-1}^{FR}\left(\frac{x_k - x_{k-1}}{\alpha_{k-1}^{FR}}\right)$$
$$= x_k^{FR} - \alpha_k^{FR} g_k + \frac{\alpha_k^{FR}}{\alpha_{k-1}^{FR}}\cdot\beta_{k-1}^{FR}(x_k^{FR} - x_{k-1}^{FR})$$

(数归) 假设已有 $x_k^{FR} = x_k^{MC}$, $x_{k-1}^{FR} = x_{k-1}^{MC}$, (要证明 MC ⇔ FR)
只需证 $\alpha_k^{MC} = \alpha_k^{FR}$, $\frac{\alpha_k^{FR}}{\alpha_{k-1}^{FR}}\beta_{k-1}^{FR} = \beta_k^{MC}$

$$\cancel{\alpha_k^{FR} = \frac{(d_k^{FR})^T g_k}{(d_k^{FR})^T G d_k^{FR}} = \frac{-\left(-g_k + \frac{g_k^T g_k}{g_{k-1}^T g_{k-1}}\cdot d_{k-1}\right)^T g_k}{(-g_k+\beta_{k-1}d_{k-1})^T G(-g_k+\beta_{k-1}d_{k-1})}}$$ (不会)
$$\cancel{= g_k^T g_k - \beta_{k-1} d_{k-1}^T g_k.}$$

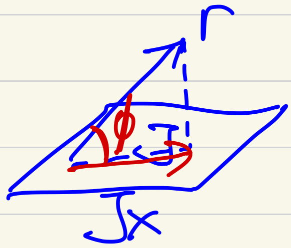
不会,略

Chap5

5. 设 $r \in \mathbb{R}^m$, $J \in \mathbb{R}^{m\times n}$ 列满秩. 证明 r 到由 J 的列张成的子空间的 Euclidean 投影为 $J(J^{\mathrm{T}}J)^{-1}J^{\mathrm{T}}r$, r 与该空间的夹角 ϕ 的余弦为

$$\cos\phi = \frac{r^{\mathrm{T}}J(J^{\mathrm{T}}J)^{-1}J^{\mathrm{T}}r}{\|J(J^{\mathrm{T}}J)^{-1}J^{\mathrm{T}}r\|\|r\|}.$$

问：该结果可用于何处?



$$\min \frac{1}{2}\|Jx-r\|^2 \Rightarrow f(x)=\frac{1}{2}(Jx-r)^T(Jx-r)=\frac{1}{2}x^T(J^TJ)x - x^TJ^Tr + \frac{1}{2}r^Tr$$

$$\nabla f(x) = (J^TJ)x - J^Tr = 0 \Rightarrow x = (J^TJ)^{-1}J^Tr$$

∴ 投影为 $Jx = J(J^TJ)^{-1}(J^Tr)$

$$\cos\phi = \frac{(Jx)^Tr}{\|Jx\|\,\|r\|} = \frac{r^TJ(J^TJ)^{-1}J^Tr}{\|J(J^TJ)^{-1}(J^Tr)\|\,\|r\|}$$

6. 对最小二乘问题, 假定存在 x^*, 使得 $J(x^*)^{\mathrm{T}}r(x^*)=0$; 对充分接近 x^* 的 x, Jacobi 矩阵 $J(x)$ Lipschitz 连续, 并且 $\|J(x)\| \leqslant \alpha$. 证明:

$$[J(x)-J(x^*)]^{\mathrm{T}}r(x^*) = S(x^*)(x-x^*) + O(\|x-x^*\|^2).$$

对 $r_i(x)$ 关于 x^* 展开再代入 $\sum r_i(x^*)\nabla r_i(x)$

$$\|J(x)-J(x^*)\| \le L\|x-x^*\|$$

$$\nabla r_i(x) = \nabla r_i(x^*) + \nabla^2 r_i(x^*)(x-x^*) + o(\|x-x^*\|)$$

$$\begin{aligned}
g(x^*) = J(x^*)^Tr(x^*) &= \sum r_i(x^*)\nabla r_i(x^*) \\
&= \sum r_i(x^*)\left[\nabla r_i(x) - \nabla^2 r_i(x^*)(x-x^*) + o(\|x-x^*\|)\right] \\
&= \sum\left[r_i(x^*)\nabla r_i(x) - r_i(x^*)\nabla^2 r_i(x^*)\cdot(x-x^*)\right] \\
&= J(x)^Tr(x^*) - S(x^*)(x-x^*)
\end{aligned}$$

$$[J(x)-J(x^*)]^Tr(x^*) = S(x^*)(x-x^*)$$

8. 在最小二乘方法中, 设 d_k^{N}, d_k^{GN} 分别为 x_k 处的 Newton 方向

$$d_k^{\mathrm{N}} = -(J_k^{\mathrm{T}} J_k + S_k)^{-1} J_k^{\mathrm{T}} r_k$$

和 Gauss-Newton 方向

$$d_k^{\mathrm{GN}} = -(J_k^{\mathrm{T}} J_k)^{-1} J_k^{\mathrm{T}} r_k.$$

证明:

$$d_k^{\mathrm{GN}} - d_k^{\mathrm{N}} = (J_k^{\mathrm{T}} J_k)^{-1} S_k d_k^{\mathrm{N}}.$$

由此证明: 对最小二乘问题的最优解 x^*, 若 $\nabla^2 f(x^*)$ 非奇异, $\nabla^2 f(x)$ 在 x^* 的邻域中 Lipschitz 连续, x_k 充分接近 x^*, $x_{k+1}^{\mathrm{GN}} = x_k + d_k^{\mathrm{GN}}$, 则

$$\|x_{k+1}^{\mathrm{GN}} - x^*\| \leqslant \|(J_k^{\mathrm{T}} J_k)^{-1}\| \|S_k\| \|x_k - x^*\| + O(\|x_k - x^*\|^2).$$

当 $n = 1$ 时, 可证

$$(x_{k+1}^{\mathrm{GN}} - x^*) - S_k (J_k^{\mathrm{T}} J_k)^{-1} (x_k - x^*) = O(|x_k - x^*|^2).$$

$$\|G(x) - G(x^*)\| \le L\|x - x^*\|$$

$$(J_k^T J_k + S_k) d_k^N = J_k^T r_k \Rightarrow (J_k^T J_k) d_k^N = J_k^T r_k - S_k d_k^N$$

$$(J_k^T J_k) d_k^{GN} = J_k^T r_k$$

$$\Rightarrow d_k^{GN} - d_k^N = (J_k^T J_k)^{-1} (\cancel{J_k^T r_k} - \cancel{J_k^T r_k} + S_k d_k^N) = (J_k^T J_k)^{-1} S_k d_k^N$$

$$\left.\begin{array}{l} x_{k+1}^{GN} = x_k + d_k^{GN} \\ x_{k+1}^N = x_k + d_k^N \end{array}\right| \Rightarrow x_{k+1}^{GN} - x_{k+1}^N = d_k^{GN} - d_k^N$$

$$\boxed{x_{k+1}^{GN} - x^* = (x_{k+1}^N - x^*) + (d_k^{GN} - d_k^N)} \quad *$$

$$\|x_{k+1}^{GN} - x^*\| \le \|x_{k+1}^N - x^*\| + \|(J_k^T J_k)^{-1} S_k d_k^N\|$$

$$\le \boxed{O(\|x_k - x^*\|^2)} + \|(J_k^T J_k)^{-1}\| \|S_k\| \|d_k^N\|$$

Newton法 二次收敛性

$$= \|(J_k^T J_k)^{-1}\| \|S_k\| \|x_{k+1}^N - x_k\| + O(\|x_k - x^*\|^2)$$

$$\le \|(J_k^T J_k)^{-1}\| \cdot \|S_k\| \cdot \|x_k - x^*\| + O(\|x_k - x^*\|^2)$$

定理 5.5(修正 Newton 方程与信赖域问题的关系) d_k 为信赖域子问题

$$\min_d q_k(d) = \frac{1}{2}d_k^T G_k d_k + d_k^T g_k \tag{5.20a}$$

$$\text{s.t. } \|d\|^2 \leqslant \Delta_k^2,\ \Delta_k > 0 \tag{5.20b}$$

的全局极小解的充分必要条件是, 对满足 (5.20b) 的 d_k, 存在 $\nu_k \geqslant 0$, 使得

$$(G_k + \nu_k I)d_k = -g_k, \tag{5.21a}$$

$$\nu_k(\Delta_k^2 - \|d_k\|^2) = 0, \tag{5.21b}$$

$$G_k + \nu_k I \text{ 半正定}. \tag{5.21c}$$

定理的证明留为作业.

必要性 $\Rightarrow$ 作 Lagrangian 函数 $L = \frac{1}{2}d^T G_k d + g_k^T d - \frac{\nu_k}{2}(\Delta_k^2 - \|d\|^2)$

$$\begin{cases} \frac{\partial L}{\partial d} = G_k d + g_k + \nu_k d = (G_k + \nu_k I)d + g_k = 0 \\ \text{互补条件 } \nu_k(\Delta_k^2 - \|d\|^2) = 0 \\ \nu_k \geq 0 \end{cases}$$

充分性: 作函数 $\overbrace{\frac{1}{2}d^T(G_k+\nu_k I)d_k + g_k^T d}^{\tilde{q}(d)} = \frac{1}{2}d^T G_k d + g_k^T d + \frac{1}{2}\nu_k d^T d$.

易发现 $d = -(G_k+\nu_k I)^{-1} g_k$ 为上述函数的全局极小点.

又 $\tilde{q}(d) = q(d) + \frac{1}{2}\nu_k\|d\|^2 \Rightarrow q(d) = \tilde{q}(d) - \frac{1}{2}\nu_k\|d\|^2$.

故 $q(d) - q(d_k) = \tilde{q}(d) - \tilde{q}(d_k) + \frac{1}{2}\nu_k(\|d_k\|^2 - \|d\|^2)$

$\geq \frac{1}{2}\nu_k(\|d_k\|^2 - \|d\|^2)$

① 若 $\nu_k = 0$, 则 $q(d) \geq q(d_k)$

② 若 $\nu_k \neq 0$, 则只能 $\Delta_k = \|d_k\| \Rightarrow q(d) - q(d_k) \geq \frac{1}{2}\nu_k(\Delta_k^2 - \|d\|^2)$.

无论如何都有 $q(d) - q(d_k) \geq \frac{1}{2}\nu_k(\Delta_k^2 - \|d\|^2)$

当 $\Delta_k \geq \|d\|$ 时, $q(d) \geq q(d_k)$. d_k 为 q 的全局最小点.

即 d_k 为 $\min_d q(d)$. s.t $\Delta_k^2 \geq \|d\|^2$ 的解.

Chap 6

(LJC@)

4. 证明：若 $\{a_i(x^*), i \in \mathscr{A}^*\}$ 线性无关，则 $\mathcal{F}^* = \mathscr{F}^*$.

5. 若在点 x^* 处 KKT 条件满足，$\{a_i(x^*), i \in \mathscr{A}^*\}$ 线性无关，证明：x^* 对应的 Lagrange 乘子 λ^* 唯一.

$$g(x) = \sum_{i\in E\cup I} \lambda_i a_i(x) \Rightarrow 0 = g(x^*) = \sum_{i\in E\cup I} \lambda_i a_i(x^*)$$

假设λ^*不唯一，则 $\exists \lambda^{(1)} \neq \lambda^{(2)}$，使得

$$\sum_i \lambda_i^{(1)} a_i(x^*) = \sum_i \lambda_i^{(2)} a_i(x^*)$$

$$\Rightarrow \sum_i (\lambda_i^{(1)} - \lambda_i^{(2)}) a_i(x^*) = 0$$

由于 $a_i(x^*)$线性无关，因此 $\lambda_i^{(1)} = \lambda_i^{(2)}$ $\forall i \in E\cup I$，
从而 $\lambda^{(1)} = \lambda^{(2)}$. 矛盾！假设不成立 原命题得证.

6. 对等式约束最优化问题，若从约束 $c(x) = 0$ 中能得到 $x_1 = \varphi(x_2)$，其中 $c(x) \in \mathbb{R}^m$，$x_1 \in \mathbb{R}^m$，$x_2 \in \mathbb{R}^{n-m}$，则等式约束最优化问题 $\min f(x_1, x_2)$ 可以化为无约束最优化问题

$$\min \psi(x_2) = f(\varphi(x_2), x_2).$$

(1) 设 $g = \begin{bmatrix} g_1 \\ g_2 \end{bmatrix}$，$A = \begin{bmatrix} A_1 \\ A_2 \end{bmatrix}$，求出 $\nabla\varphi$ 和 $\nabla\psi$;

(2) 设 A^* 列满秩，证明：Lagrange 乘子 λ^* 可唯一地表示为 $\lambda^* = A^{*\dagger} g^*$，其中 $A^{*\dagger}$ 是 A^* 的广义逆，或是 $A_1^* \lambda^* = g_1^*$ 的解.

(隐函数求导)

(1) $C(x_1, x_2) = 0 \Rightarrow \dfrac{\partial c}{\partial x_1} \cdot \dfrac{\partial x_1}{\partial x_2} + \dfrac{\partial c}{\partial x_2} = 0$

$$\Rightarrow A_1 \nabla\phi + A_2 = 0 \Rightarrow \nabla\phi = -(A_1)^{-1} A_2$$

$$\nabla\psi = \frac{\partial f}{\partial x_1} \cdot \frac{\partial x_1}{\partial x_2} + \frac{\partial f}{\partial x_2} = g_1 \cdot \nabla\phi + g_2$$

$$= g_1(-(A_1)^{-1} A_2) + g_2 = -g_1 (A_1)^{-1} A_2 + g_2.$$

(2) $L = f(x_1, x_2) - \lambda c(x_1, x_2)$

$$\nabla L = \begin{pmatrix} g_1 - \lambda \frac{\partial c}{\partial x_1} \\ g_2 - \lambda \frac{\partial c}{\partial x_2} \end{pmatrix} = 0 \Rightarrow \begin{cases} g_1^* = \lambda A_1^* \\ g_2^* = \lambda A_2^* \end{cases} \Rightarrow g^* = \lambda A^* \Rightarrow \lambda = (A^*)^{-1} g^*$$

7. 对不等式约束最优化问题，在什么条件下，KKT 条件是充分条件、必要条件、充分必要条件？请举例说明.

KKT条件是最优解的：
- 充分条件 ✗，要考虑二阶充分条件.
- 必要条件：约束 $\{a_i^*(x), i\in\mathcal{A}(x^*)\}$ 线性无关 (LICQ条件).
- 充要条件：凸优化 (换言之，G正定).

8. **MF (Mangasarian-Fromovitz)** 约束规范是这样定义的：若在 x^* 处，存在 $d\in\mathbb{R}^n$，使得

$$
\begin{aligned}
{a^*}^{\mathrm{T}} d &= 0, \quad i\in\mathcal{E},\\
{a^*}^{\mathrm{T}} d &> 0, \quad i\in\mathcal{I}^*,
\end{aligned}
$$

且 $\{a_i^*, i\in\mathcal{E}\}$ 线性无关，则称在 x^* 处 MF 约束规范满足. 证明：对约束 $x_1^3\geqslant x_2$, $x_2\geqslant 0$, 在 $x^*=(0,0)^{\mathrm{T}}$ 处，MF 约束规范不满足，线性无关约束规范 ($\{a_i^*, i\in\mathscr{A}^*\}$ 线性无关) 也不满足.

(1) $C_1(x)=x_1^3-x_2\geq 0$　(0,0)处起作用约束　$a_1(x)=\begin{pmatrix}3x_1^2\\-1\end{pmatrix}$

$C_2(x)=x_2\geq 0$　(0,0)处起作用约束　$a_2(x)=\begin{pmatrix}0\\1\end{pmatrix}$

$(a_1^*)^T d=(0,-1)^T\begin{pmatrix}d_1\\d_2\end{pmatrix}>0 \Rightarrow -d_2>0\Rightarrow d_2<0$

$(a_2^*)^T d=(0,1)^T\begin{pmatrix}d_1\\d_2\end{pmatrix}>0 \Rightarrow d_2>0$

$\Rightarrow$ 这样的 $d=\begin{pmatrix}d_1\\d_2\end{pmatrix}$ 不存在！

$\therefore$ x^* 处MF约束不满足

(2) $C_1(0,0)=0$, $C_2(0,0)=0 \Rightarrow \{1,2\}\subseteq\mathcal{A}^*$.

但 $a_1^*=\begin{pmatrix}0\\-1\end{pmatrix}$ 与 $a_2^*=\begin{pmatrix}0\\1\end{pmatrix}$ 线性相关. 故LICQ不满足

10. 求下列问题的 KKT 点, 判断这些 KKT 点是否是最优解:

(1) $\min\ (x_1-1)^2+(x_2-2)^2,$

s.t. $(x_1-1)^2-5x_2=0;$

(2) $\min\ (x_1+x_2)^2+2x_1+x_2^2,$

s.t. $x_1+3x_2\leqslant 4,$

$2x_1+x_2\leqslant 3,$

$x_1\geqslant 0,$

$x_2\geqslant 0.$

11. 求出最小化

$$f(x)=x_1^2+4x_2^2+16x_3^2$$

在约束 $c(x)=0$ 下的所有 KKT 点, 其中 $c(x)$ 分别为

(1) $c(x)=x_1-1;$

(2) $c(x)=x_1x_2-1=0;$

(3) $c(x)=x_1x_2x_3-1=0.$

判断这些 KKT 点是否是最优解.

(1)

13. 考虑问题

$$\max f(x)=\sum_{i=1}^{n} f_i(x_i),$$

$$\text{s.t. } x_i \geqslant 0,\ i=1,\cdots,n,$$

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

其中 f_i 可微. 设 x^* 是问题的最优解. 证明：存在 μ^*, 使得

$$f_i'(x_i^*)=\mu^*,\quad x_i^*>0,$$

$$f_i'(x_i^*)\geqslant \mu^*,\quad x_i^*=0.$$

$$L=\sum_{i=1}^{n} f_i(x_i)-\sum_{i=1}^{n}(\lambda_i x_i)-\lambda\left(\sum_{i=1}^{n} x_i-1\right)$$

$$\Rightarrow \begin{cases} \frac{\partial L}{\partial x_i}=f_i'(x_i)-\lambda_i-\lambda=0 \quad (i=1,\ldots n). \\ \lambda_i x_i=0 \\ \lambda(\sum x_i-1)=0 \end{cases}$$

对于 $x_i^*>0$ 的 i, 有 $\lambda_i=0$, 从而 $f_i'(x_i^*)=\lambda^*>0$

对于 $x_i^*=0$ 的 i, $f_i'(x_i)=\lambda_i+\lambda\geq\lambda^*$

$\therefore$ 只需取 $\mu^*=\lambda^*$ 即可

14. 对凸规划问题, **Slater 约束规范**为：存在 $\tilde{x}\in D$, 使得 $c_i(\tilde{x})>0, i\in I$. 证明：若凸规划问题满足 Slater 约束规范, 则可行点 x^* 为最优解的充分必要条件是 x^* 为 KKT 点.

x^* 为最优解 $\Leftrightarrow$? x^* 为KKT点.

超纲了

Chap 7

1. 对问题

$$\min -x_1x_2x_3,$$
$$\text{s.t. } 72 - x_1 - 2x_2 - 2x_3 = 0,$$

考虑外点罚函数方法. 求出 $x(\sigma)$ 的显式表达式. 当 $\sigma \to \infty$ 时, 求出问题的最优解和相应的 Lagrange 乘子. 给出 σ 的取值范围, 使矩阵 $\nabla_x^2 P_E(x(\sigma), \sigma)$ 正定.

$$P_E(x,\sigma) = -x_1x_2x_3 + \frac{\sigma}{2}(72 - x_1 - 2x_2 - 2x_3)^2$$

$$\begin{cases} \frac{\partial P}{\partial x_1} = -x_2x_3 - \sigma(72 - x_1 - 2x_2 - 2x_3) = 0 \\ \frac{\partial P}{\partial x_2} = -x_1x_3 - 2\sigma(72 - x_1 - 2x_2 - 2x_3) = 0 \\ \frac{\partial P}{\partial x_3} = -x_1x_2 - 2\sigma(72 - x_1 - 2x_2 - 2x_3) = 0 \end{cases} \Rightarrow \begin{cases} x_1 = 2(3\sigma - 3\sqrt{\sigma^2 - 8\sigma}) \\ x_2 = 3\sigma - 3\sqrt{\sigma^2 - 8\sigma} \\ x_3 = 3\sigma - 3\sqrt{\sigma^2 - 8\sigma} \end{cases}$$

$$\therefore x^* = (3\sigma - 3\sqrt{\sigma^2 - 8\sigma})(2,1,1)^T \to 12(2,1,1)^T$$

$$\frac{9\sigma^2 - (9\sigma^2 - 72\sigma)}{3\sigma + 3\sqrt{\sigma^2 - 8\sigma}} = \frac{72\sigma}{3\sigma + 3\sqrt{\sigma^2 - 8\sigma}} \to 12$$

$$\nabla^2 P = \begin{pmatrix} \sigma & -x_3 + 2\sigma & -x_2 + 2\sigma \\ -x_3 + 2\sigma & 4\sigma & -x_1 + 4\sigma \\ -x_2 + 2\sigma & -x_1 + 4\sigma & 4\sigma \end{pmatrix} \text{ 正定 } \Rightarrow \sigma > 0$$
$$\Rightarrow 4\sigma^2 - (2\sigma - x_3)^2 > 0 \Rightarrow \sigma > 8$$

$$\lambda = \lim_{k\to\infty} \sigma_k c(x_k) = \lim_{k\to\infty} \sigma \cdot (72 - 6x_2)$$
$$= \lim_{k\to\infty} \sigma(72 - 6(3\sigma - 3\sqrt{\sigma^2 - 8\sigma}))$$
$$= 72\sigma - 6\sigma(3\sigma - 3\sqrt{\sigma^2 - 8\sigma})$$
$$\to 144$$

2. 对问题

$$\min x_2^2 - 3x_1,$$
$$\text{s.t. } x_1 + x_2 = 1,$$
$$x_1 - x_2 = 0,$$

应用外点罚函数方法. 当 $\sigma \to \infty$ 时, 求出问题的最优解和相应的 Lagrange 乘子.

$$P_E(x,\sigma) = x_2^2 - 3x_1 + \frac{\sigma}{2}(x_1 + x_2 - 1)^2 + \frac{\sigma}{2}(x_1 - x_2)^2$$

$$\begin{cases} \frac{\partial P}{\partial x_1} = -3 + \sigma(x_1 + x_2 - 1) + \sigma(x_1 - x_2) = 0 \\ \frac{\partial P}{\partial x_2} = 2x_2 + \sigma(x_1 + x_2 - 1) - \sigma(x_1 - x_2) = 0. \end{cases} \Rightarrow \begin{cases} x_1 = \frac{\sigma + 3}{2\sigma} \\ x_2 = \frac{\sigma}{2(\sigma+1)} \end{cases}$$

$$x^* \to (1/2, 1/2)$$

$$\lambda_1 = \lim_{k\to\infty} -\sigma_k c_1(x_k) = -\sigma\left(\frac{\sigma+3}{2\sigma} + \frac{\sigma}{2(\sigma+1)} - 1\right) \to -1$$
$$\lambda_2 = \lim_{k\to\infty} -\sigma_k c_2(x_k) = -\sigma\left(\frac{\sigma+3}{2\sigma} - \frac{\sigma}{2(\sigma+1)}\right) \to -2$$

3. 考虑问题

$$\min x,\quad x\in\mathbb{R},$$
$$\text{s.t. } x^2\geqslant 0,$$
$$x+1\geqslant 0.$$

写出该问题的对数障碍函数 $B_L(x,\mu)$，并求出其局部极小点. 对任意 $\{\mu_k\}$，$\mu_k\to 0$，求出相应的局部极小点序列 $\{x(\mu_k)\}$ 的极限点.

$$B_L(x\cdot\mu)=x-\mu\ln x^2-\mu\ln(x+1)=x-2\mu\ln x-\mu\ln(x+1)$$

$$\frac{\partial B}{\partial x}=1-\frac{2\mu}{x}-\frac{\mu}{x+1}=0\Rightarrow x=\frac{3\mu-1\pm\sqrt{9\mu^2+2\mu+1}}{2}$$

$\mu\to 0$时 $x^*=\frac{-1\pm\sqrt{1}}{2}=0$ 或 -1

4. 对问题

$$\min 2x_1+3x_2,$$
$$\text{s.t. } 1-2x_1^2-x_2^2\geqslant 0,$$

考虑对数障碍函数方法. 当 $\mu\to 0$ 时，求出问题的最优解和相应的 Lagrange 乘子.

$$B_L(x\cdot\mu)=2x_1+3x_2-\mu\ln(1-2x_1^2-x_2^2)$$

$$\frac{\partial B}{\partial x_1}=2-\frac{\mu}{1-2x_1^2-x_2^2}\cdot(-4x_1)=2+\frac{4\mu x_1}{1-2x_1^2-x_2^2}=0$$

$$\frac{\partial B}{\partial x_2}=3-\frac{\mu}{1-2x_1^2-x_2^2}\cdot(-2x_2)=3+\frac{2\mu x_2}{1-2x_1^2-x_2^2}=0$$

$$\Rightarrow\begin{cases}x_1=\dfrac{\mu-\sqrt{\mu^2+11}}{11}\\[2mm] x_2=\dfrac{3(\mu-\sqrt{\mu^2+11})}{11}\end{cases}$$

当 $\mu\to 0$ 时 $(x_1,x_2)\to\left(-\frac{\sqrt{11}}{11},-\frac{3\sqrt{11}}{11}\right)$

$$\lambda^*=\frac{\mu}{1-2\left(\frac{\mu-\sqrt{\mu^2+11}}{11}\right)^2-9\left(\frac{\mu-\sqrt{\mu^2+11}}{11}\right)^2}=\frac{\mu}{1-\frac{(\mu-\sqrt{\mu^2+11})^2}{11}}$$

$$=\frac{11\mu}{11-(\mu-\sqrt{\mu^2+11})^2}=\frac{11\mu}{11-(\mu^2+\mu^2+11-2\mu\sqrt{\mu^2+11})}=\frac{11\mu}{2\mu\sqrt{\mu^2+11}-2\mu^2}=\frac{11}{2\sqrt{\mu^2+11}-2\mu}$$

$$\to\frac{11}{2\sqrt{11}}=\frac{\sqrt{11}}{2}.$$

7. 对倒数障碍函数 $B_I(x,\mu)$, 证明：在点 $x^{(k)}$ 处, Lagrange 乘子估计为 $\lambda_i^{(k)} = \frac{\mu_k}{(c_i^{(k)})^2}$, $i \in \mathcal{I}$. 由此证明对 $x^{(k)} \to x^*, \lambda^{(k)} \to \lambda^*$, 若 $i \notin \mathcal{I}^*$, 则 $\lambda_i^{(k)} \to 0$, 且 x^*, λ^* 为 KKT 对.

$$B_I(x_k, \mu) = f(x_k) - \mu_k \sum_{i \in \mathcal{I}} \ln c_i(x)$$

$$\nabla B_I(x_k, \mu) = g_k - \sum_{i \in \mathcal{I}} \mu_k \frac{1}{c_i(x)_k} \cdot \nabla c_i(x)_k = 0$$

$$\Rightarrow g^* = \sum_{i \in \mathcal{I}} \mu_k \frac{1}{c_i(x^*)} \nabla c_i(x^*)$$

9. 对问题

$$\min \frac{1}{1+x^2}, \quad x \in \mathbb{R},$$
$$\text{s.t. } x \geqslant 1,$$

考虑障碍函数方法. 证明：对任何 $\mu > 0$, $B_I(x,\mu)$ 和 $B_L(x,\mu)$ 均无下界.

$B_I(x,\mu) = \frac{1}{1+x^2} + \mu \cdot \frac{1}{(x-1)}$, 取 $x \to 1^-$ 则 B_I 无下界

$B_L(x,\mu) = \frac{1}{x^2+1} - \mu \ln(x-1)$, 取 $x \to 1^+$ 则 B_L 无下界

11. 对问题

$$\min \frac{1}{2} x^T G x + \alpha b^T x,$$
$$\text{s.t. } b^T x = 0,$$

~~正定?~~

其中 $G \in \mathbb{R}^{n\times n}$ 对称非奇异, $\alpha \in \mathbb{R}$, 且对任意满足 $b^T x = 0$ 的 $x \neq 0$, $x^T G x > 0$. 应用乘子罚函数方法, 取 $\lambda_0 = 0$. 证明：当

$$|1 + \sigma b^T G^{-1} b| > 1$$

时, 乘子罚函数方法产生的 $\{x_k\}$ 收敛于最优解 $x^* = 0$.

$$L(x,\lambda,\sigma) = \frac{1}{2} x^T G x + \alpha b^T x - \lambda (b^T x) + \frac{\sigma}{2} (b^T x)^2$$

$$\begin{cases} \frac{\partial L}{\partial x_k} = Gx_k + (\alpha - \lambda) b + \sigma (b^T x)_k \cdot b = Gx_k + (\alpha - \lambda + \sigma b^T x)_k b = 0 \\ \lambda (b^T x) = 0 \\ \lambda \geq 0 \end{cases}$$

$$\Downarrow$$

$$Gx_k + (\alpha - \lambda) b + (\sigma b b^T) x_k = 0$$

$$(G + \sigma b b^T) x_k = (\lambda - \alpha) b$$

依 Sherman-Morrison-Woodbury 公式：

$(G+\sigma bb^T)$ 可逆 $\Leftrightarrow 1+\sigma b^T G^{-1} b \neq 0$

当 $|1+\sigma b^T G^{-1} b|>1$ 时 $\sigma b^T G^{-1} b+1>1$ 或 <-1 ✓

$\therefore x_k = (G+\sigma bb^T)^{-1}(\lambda-\alpha)b$

$\lambda_{k+1} = \lambda_k - \sigma C_i(x_k)$
$= \lambda_k - \sigma\left(b^T(G+\sigma bb^T)^{-1}(\lambda_k-\alpha)b\right)$

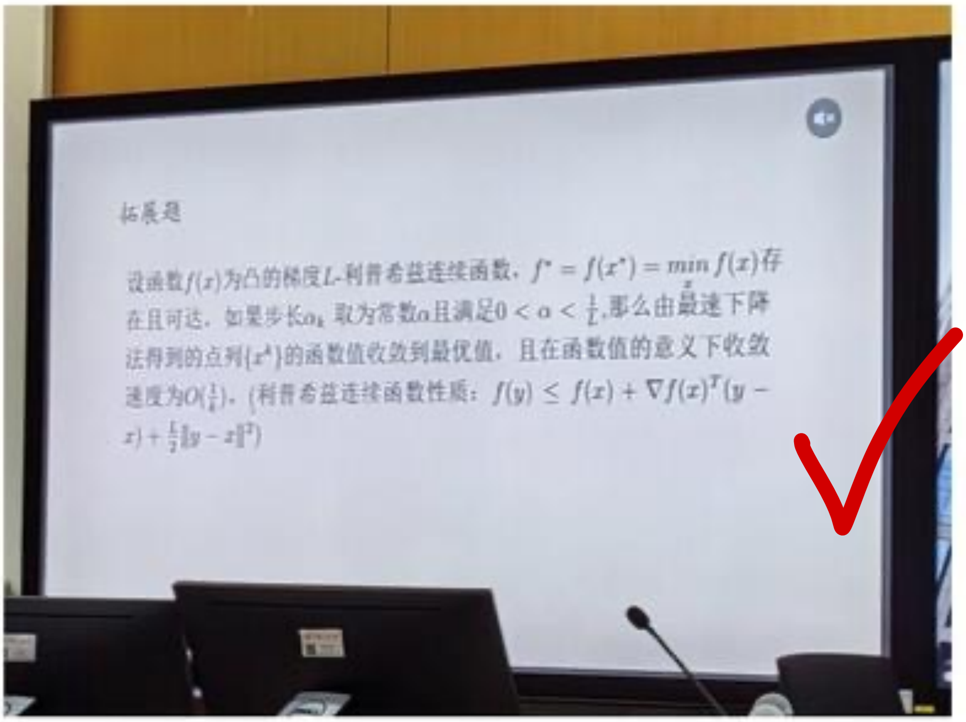
不动点：$\lambda = \lambda - \sigma(\lambda-\alpha)\left[b^T(G+\sigma bb^T)^{-1}b\right] \Rightarrow \lambda=\alpha$

$\therefore \lambda \to \lambda^* = \alpha \Rightarrow x^* = (G+\sigma bb^T)^{-1}(\lambda^*-\alpha)b = 0$

2024 Exam Sun YiFan

1. (15 分) f(x)是定义在凸集 D 上的凸函数，证明：Fα={x | f(x)≤α}是凸集

2. (15 分) 习题课讲过的附加题



3. (20 分) 精确线搜索的收敛性的证明

4. (20 分) f(x)=||x||^β，使用基本牛顿方法极小化 f(x)，x0≠0。

已知 Sherman-Morrison 公式：

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

证明：

(1) k > 1 且 k≠2 时，{xk}的收敛速度为 Q-线性的

(2) 0 < k < 1 时，牛顿方法无法收敛

12. 设在水平集 $\{x|f(x)\leqslant f(x_0)\}$ 上, $f(x)$ 有下界, $g(x)$ 一致连续; 在算法 2.1 中, 方向 d_k 与 $-g_k$ 之间的夹角 θ_k 一致有界, 即对某一 $\mu>0$, 成立

$$0\leqslant\theta_k\leqslant\frac{\pi}{2}-\mu.$$

证明: 若精确线搜索准则对任给 k 都成立, 则或者存在 N, 使 $g_N=0$, 或者 $g_k\to 0, k\to\infty$.

$t_k\in[0,1]$

$$f(x_k+\alpha_k d_k)=f(x_k)+\alpha_k\, g(x_k+t_k\alpha_k d_k)^T d_k$$

$$=f(x_k)+\alpha_k\left[g(x_k+t_k\alpha_k d_k)-g(x_k)\right]^T d_k+\alpha_k g_k^T d_k.$$

$$\leq f(x_k)+\alpha_k\,|g(x_k+t_k\alpha_k d_k)-g(x_k)|\,|d_k|+\alpha_k g_k^T d_k$$

$$=f(x_k)+\alpha_k\; o(\|t_k\alpha_k d_k\|)+\alpha_k g_k^T d_k$$

$=$

$$\cos\theta_k\geq\cos\left(\frac{\pi}{2}-\mu\right)=\sin\mu$$

By Zoutendijk 条件.

$$\sum_{k\geq 0}\|g_k\|^2=\sum_{k\geq 0}\frac{\|g_k\|^2\cdot\cos^2\theta_k}{\cos^2\theta_k}$$

$$\leq\frac{1}{\sin^2\mu}\underbrace{\sum_{k\geq 0}\|g_k\|^2\cdot\cos^2\theta_k<\infty}.$$

$\therefore$ 收敛 $\|g_k\|\to 0$, 即 $g_k\to 0$.

5. (15 分) min (x1-1)^2 + x2，2-x1-x2≥0，x2-x1-1=0。求解 KKT 点

6. (15 分) 最优化问题 min f1(x)+f2(Ax)，A 为 n*n 的矩阵，给出 ADMM 方法的迭代公式