#### **Problem 1**

(a)

$$egin{aligned} 
abla_{ec{x}}f(ec{x}) &= Aec{x} - ec{b} \ 
abla_{ec{x}}^2f(ec{x}) &= 
abla(Aec{x} - ec{b}) \ &= A^T \ &= A \end{aligned}$$

令第一个式子为0可得

$$\vec{x}' = A^{-1}\vec{b}$$

因为A正定,所以 $\vec{x}'$ 是极小值点。

(b)首先考虑如何将两个向量变成A-共轭,假设两个向量为 $ec{v}_1, ec{v}_2$ ,取 $ec{w}_1 = ec{v}_1$ ,设

$$ec{w}_2 = ec{v}_2 - lpha ec{w}_1$$

要使得 $\vec{w}_1, \vec{w}_2$ 为A-共轭,那么

$$egin{aligned} \langle ec{w}_2, ec{w}_1 
angle &= \langle ec{v}_2 - lpha ec{w}_1, ec{w}_1 
angle \ &= \langle ec{v}_2, ec{w}_1 
angle - lpha \langle ec{w}_1, ec{w}_1 
angle \ &= 0 \ lpha &= rac{\langle ec{v}_2, ec{w}_1 
angle}{\langle ec{w}_1, ec{w}_1 
angle} \end{aligned}$$

不难看出

$$\operatorname{span}\{ec{w}_1,ec{w}_2\}=\operatorname{span}\{ec{v}_1,ec{v}_2\}$$

接着,假设我们已经有了 $k \cap A$ —共轭向量 $\vec{w}_1, \ldots, \vec{w}_k$ ,并且

$$\operatorname{span}\{\vec{w}_1,\ldots,\vec{w}_k\} = \operatorname{span}\{\vec{v}_1,\ldots,\vec{v}_k\}$$

现在讨论如何获得第k+1个A-共轭向量 $ec{w}_{k+1}$ ,假设

$$ec{w}_{k+1} = ec{v}_{k+1} - \sum_{j=1}^k lpha_j ec{w}_j$$

我们需要的条件为

$$\langle ec{w}_{k+1}, ec{w}_i 
angle = 0, 1 \leq i \leq k$$

那么 $\forall 1 \leq i \leq k$ 

$$egin{aligned} \langle ec{w}_{k+1}, ec{w}_i 
angle &= \langle ec{v}_{k+1} - \sum_{j=1}^k lpha_j ec{w}_j, ec{w}_i 
angle \ &= \langle ec{v}_{k+1}, ec{w}_i 
angle - \sum_{j=1}^k lpha_j \langle ec{w}_j, ec{w}_i 
angle \ &= \langle ec{v}_{k+1}, ec{w}_i 
angle - lpha_i \langle ec{w}_i, ec{w}_i 
angle \ &= 0 \ &lpha_i &= rac{\langle ec{v}_{k+1}, ec{w}_i 
angle}{\langle ec{w}_i, ec{w}_i 
angle} \end{aligned}$$

显然有

$$\mathrm{span}\{\vec{w}_1,\ldots,\vec{w}_{k+1}\} = \mathrm{span}\{\vec{v}_1,\ldots,\vec{v}_{k+1}\}$$

上述讨论可以得到如下算法:

1. 令

$$\vec{w}_1 = \vec{v}_1$$

2. 对k = 2...n

1. 对 $i = 1, \ldots, k - 1$ , 计算

$$lpha_i = rac{\langle ec{v}_{k+1}, ec{w}_i 
angle}{\langle ec{w}_i, ec{w}_i 
angle}$$

2.

$$ec{w}_k = ec{v}_k - \sum_{i=1}^{k-1} lpha_i ec{w}_i$$

## **Problem 2**

(a)拉朗格朗日乘子为

$$\Lambda(ec{x},ec{\lambda},ec{\mu}) = ec{c}^Tec{x} - ec{\lambda}^T(Aec{x}-ec{b}) - ec{\mu}^Tec{x}$$

所以KKT条件为

1. stationarity

$$abla_{ec{x}} \Lambda(ec{x},ec{\lambda},ec{\mu}) = ec{c} - A^T ec{\lambda} - ec{\mu} = ec{0}$$

2. primal feasibility

$$Aec{x} = ec{b} \ ec{x} > ec{0}$$

3. complementary slackness

$$\mu_j x_j = 0$$

4. dual feasibility

$$\mu_i \geq 0$$

(b)令

$$x_{n+1}=d-ec{v}^Tec{x}=d-\sum_{i=1}^n v_i x_i \geq 0$$

记

$$ilde{A} = \left[egin{array}{c|c} A & ec{0} \ \hline ec{v}^T & 1 \end{array}
ight], ilde{ ilde{b}} = \left[egin{array}{c|c} ec{b} \ d \end{array}
ight], ilde{ ilde{c}} = \left[egin{array}{c|c} ec{c} \ 0 \end{array}
ight], ilde{ ilde{x}} = \left[egin{array}{c|c} ec{x} \ x_{n+1} \end{array}
ight]$$

那么

$$egin{aligned} & ilde{ec{c}}^T ilde{ec{x}} = ec{c}^T ec{x} \ & ilde{A} ilde{ec{x}} = egin{bmatrix} A ec{x} \ ec{v}^T ec{x} + x_{n+1} \end{bmatrix} = egin{bmatrix} ec{b} \ d \end{bmatrix} = ilde{ec{b}} \ & ilde{ec{x}} = egin{bmatrix} ec{x} \ x_{n+1} \end{bmatrix} \geq 0 \end{aligned}$$

所以新的线性规划问题为

$$egin{aligned} ext{minimize} & ilde{ar{c}}^T ilde{ar{x}} \ ext{such that} & ilde{A} ilde{ar{x}} = ilde{ar{b}} \ & ilde{ar{x}} \geq ec{0} \end{aligned}$$

(c)对偶问题可以化为

$$egin{aligned} ext{minimize} & - ec{b}^T ec{y} \ ext{such that} & - A^T ec{y} + ec{c} \geq ec{0} \end{aligned}$$

记

$$\Lambda'(ec{y},ec{\lambda}) = -ec{b}^Tec{y} - ec{\lambda}^T(-A^Tec{y} + ec{c})$$

所以stationarity条件为

$$egin{aligned} 
abla_{ec{y}} \Lambda'(ec{y},ec{\lambda}) &= -ec{b} + Aec{\lambda} = ec{0} \ ec{b} &= Aec{\lambda} \end{aligned}$$

带回原式可得

$$egin{aligned} \Lambda'(ec{y},ec{\lambda}) &= -ec{\lambda}^TA^Tec{y} + ec{\lambda}^TA^Tec{y} - ec{\lambda}^Tec{c} \ &= -ec{\lambda}^Tec{c} \ &= -ec{c}^Tec{\lambda} \end{aligned}$$

complementary slackness条件为

$${ec \lambda}^T (-A^T {ec y} + {ec c}) = 0$$

所以在最优解处

$$-ec{b}^Tec{y} = \Lambda'(ec{y},ec{\lambda}) = -ec{c}^Tec{\lambda} \ ec{b}^Tec{y} = ec{c}^Tec{\lambda}$$

因为驻点唯一, 所以如下方程的解唯一:

$$A\vec{\lambda} = \vec{b} \tag{1}$$

 $\vec{\lambda} > \vec{0}$ (2)

设上述方程的解为

 $ec{\lambda}'$ 

所以最优解满足

$$\vec{b}^T \vec{y} = \vec{c}^T \vec{\lambda}' \tag{3}$$

因为原始问题的驻点唯一,所以原始问题的解唯一,设为 $\vec{x}'$ ,那么最优解为

$$\vec{c}^T \vec{x}' \tag{4}$$

其中 $\vec{x}'$ 满足

$$A\vec{x} = \vec{b} \tag{5}$$

$$\vec{x} \geq \vec{0}$$
 (6)

此即为对偶问题的解。注意(3)(4)和(5)(6)的形式一致,所以(3)(4)相等,即原问题和对偶问题的最优值相同。

## **Problem 3**

(a)(i)取

$$g(\vec{x}) = \|\vec{x}\|^2 = \vec{x}^T \vec{x}$$

(ii)因为

$$abla_{ec{x}}g(ec{x})=2ec{x}$$

所以随机梯度下降法计算的梯度为

$$rac{2}{k}\sum_{i=1}^k (ec{x}_i-ec{x}) = 2\left(rac{1}{k}\sum_{i=1}^k ec{x}_i-ec{x}
ight)$$

(b)(i)不一定,如果 f无下界,那么无法收敛到局部最小值。

(ii)利用单调有界数列必收敛即可。

## **Problem 4**

 $\vec{x}$ 是Pareto optimal的含义为

$$orall ec{y},$$
或者  $\exists i, s.\ t\ f_i(ec{x}) > f_i(ec{y})$ ,或者  $orall i, f_i(ec{x}) \geq f_i(ec{y})$ 

(i)如果第一个条件不成立,即∀y,i,

$$f_i(ec{x}) \leq f_i(ec{y})$$

如果 $f_i$ 全部相同,那么结论显然;如果 $f_i$ 不全相同,那么不妨设 $f_i \neq f_i$ ,那么必然存在函数值不相同的点 $\vec{x}$ ,设

$$f_i(\vec{x}) > f_i(\vec{x})$$

(b)首先由 $f_i$ 的凸性以及 $\vec{\gamma} \geq \vec{0}$ ,我们可得g也是凸函数,所以存在最小值。接着将 $\gamma_i$ 视为变量,所以得到如下优化问题

$$egin{aligned} ext{minimize} & \sum_{i=1}^k \gamma_i f_i(ec{x}) \ ext{such that} & \gamma_i \geq 0, i=1,\ldots,n \ & \sum_{i=1}^n \gamma_i = 1 \end{aligned}$$

构造拉格朗日乘子:

$$L(ec{x},\gamma) = \sum_{i=1}^k \gamma_i f_i(ec{x}) - \sum_{i=1}^k lpha_i \gamma_i - eta \left(\sum_{i=1}^k \gamma_i - 1
ight)$$

关于 $\vec{x}$ , $\gamma$ 求梯度并为0可得

$$egin{aligned} 
abla_{ec{x}} L(ec{x}, \gamma) &= \sum_{i=1}^k \gamma_i 
abla_{ec{x}} f_i(ec{x}) = ec{0} \ rac{\partial L(ec{x}, \gamma)}{\partial \gamma_i} &= f_i(ec{x}) - lpha_i - eta = 0, i = 1, \ldots, n \end{aligned}$$

所以

$$f_i(\vec{x}') = \alpha_i + \beta, i = 1, \dots, n$$

对偶互补条件为

$$\alpha_i \gamma_i = 0, i = 1, \ldots, n$$

因为

$$\gamma_i > 0$$

所以

$$\alpha_i = 0$$

因此

$$f_i(ec{x}') = eta, i = 1, \ldots, n$$

以及

$$g(ec{x}') = \sum_{i=1}^k \gamma_i f_i(ec{x}') = \sum_{i=1}^k \gamma_i eta = eta$$

注意我们有

$$g(ec{x}) = \sum_{i=1}^k \gamma_i f_i(ec{x}) \geq eta$$

如果不存在 $i, \vec{x}$ , 使得

$$f_i(\vec{x}) > \beta$$

那么必然有

$$g(ec{x}) = \sum_{i=1}^k \gamma_i f_i(ec{x}) \leq \sum_{i=1}^k \gamma_i eta = eta$$

所以 $\forall i, \vec{x}$ ,

$$f_i(\vec{x}) = \beta$$

此时 $\vec{x}'$ 是Pareto optimal。

如果存在 $i, \vec{x}$ , 使得

$$f_i(\vec{x}) > \beta = f_i(\vec{x}')$$

那么 $\vec{x}'$ 同样是Pareto optimal。

(c)感觉原题有误, $\vec{x}'$ 应该是Pareto dominate。

关于求求梯度可得

$$abla h(ec{x}) = 2 \sum_{i=1}^k (f_i(ec{x}) - z_i) 
abla f_i(ec{x})$$

$$abla^2 h(ec{x}) = 2 \sum_{i=1}^k (f_i(ec{x}) - z_i) 
abla^2 f_i(ec{x})$$

因为 $\forall i$ 

$$f_i(ec{x}) - z_i \geq 0$$

以及 $abla^2 f_i(ec{x}) \geq$  (半正定),所以 $abla^2 h(ec{x})$ 半正定,因此最小值点唯一。

注意 $\forall \vec{x} \neq \vec{x}'$ 

$$\sum_{i=1}^k (f_i(ec{x}') - z_i)^2 < \sum_{i=1}^k (f_i(ec{x}) - z_i)^2$$

所以必然存在i, 使得

$$egin{split} (f_i(ec{x}') - z_i)^2 &< (f_i(ec{x}) - z_i)^2 \ f_i(ec{x}') - z_i &< f_i(ec{x}) - z_i \ f_i(ec{x}') &< f_i(ec{x}) \end{split}$$

因此 $\vec{x}'$ 是Pareto dominate。

# **Problem 5**

(a)当满足如下条件时f取最小值

$$x_1=1, x_2=x_1^2=1$$

(b)

$$egin{aligned} rac{\partial f}{\partial x_1} &= (x_1^2 - x_2) imes 2x_1 + (x_1 - 1) \ &= 2x_1^3 - 2x_1x_2 + x_1 - 1 \ rac{\partial f}{\partial x_2} &= (x_1^2 - x_2) imes (-1) \ &= x_2 - x_1^2 \ rac{\partial^2 f}{\partial x_1^2} &= rac{\partial}{\partial x_1} (2x_1^3 - 2x_1x_2 + x_1 - 1) \ &= 6x_1^2 - 2x_2 + 1 \ rac{\partial^2 f}{\partial x_2^2} &= rac{\partial}{\partial x_2} (x_2 - x_1^2) \ &= 1 \ rac{\partial^2 f}{\partial x_2 \partial x_1} &= rac{\partial}{\partial x_2} (2x_1^3 - 2x_1x_2 + x_1 - 1) \ &= -2x_1 \end{aligned}$$

所以

$$egin{aligned} 
abla f \Big|_{ec{x}_0=(2,2)} &= egin{bmatrix} 9 \ -2 \end{bmatrix} \ H &= 
abla^2 f \Big|_{ec{x}_0=(2,2)} &= egin{bmatrix} 21 & -4 \ -4 & 1 \end{bmatrix} \ H^{-1} 
abla f &= egin{bmatrix} rac{1}{5} \ -rac{6}{5} \end{bmatrix} \end{aligned}$$

所以

$$egin{aligned} ec{x}_1 &= ec{x}_0 - H^{-1} 
abla f \ &= egin{bmatrix} 2 \ 2 \end{bmatrix} - egin{bmatrix} rac{1}{5} \ -rac{6}{5} \end{bmatrix} \ &= egin{bmatrix} rac{9}{5} \ rac{16}{5} \end{bmatrix} \end{aligned}$$

(c)显然

$$f(ec{x}_1) < f(ec{x}_0)$$

从这个角度来说这步更新是好的。

(d)注意最优点为

$$[1, 1]^T$$

这和 $\vec{x}_0$ 的方向一致,但是和 $\vec{x}_1$ 的方向不一致,所以从这个角度来说不是一个好的更新。