## Problem 1

(a)

$$\begin{aligned}\nabla_{\vec{x}} f(\vec{x}) &= A\vec{x}-\vec{b}\\ \nabla_{\vec{x}}^2 f(\vec{x}) &= \nabla(A\vec{x}-\vec{b})\\ &= A^T\\ &= A\end{aligned}$$

令第一个式子为$\vec{0}$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$\vec{v}_1,\vec{v}_2$，取$\vec{w}_1=\vec{v}_1$，设

$$\vec{w}_2 = \vec{v}_2 - \alpha\vec{w}_1$$

要使得$\vec{w}_1,\vec{w}_2$为$A-$共轭，那么

$$\begin{aligned}\langle\vec{w}_2,\vec{w}_1\rangle &= \langle\vec{v}_2-\alpha\vec{w}_1,\vec{w}_1\rangle\\ &= \langle\vec{v}_2,\vec{w}_1\rangle - \alpha\langle\vec{w}_1,\vec{w}_1\rangle\\ &= 0\\ \alpha &= \frac{\langle\vec{v}_2,\vec{w}_1\rangle}{\langle\vec{w}_1,\vec{w}_1\rangle}\end{aligned}$$

不难看出

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

接着，假设我们已经有了$k$个$A-$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$\vec{w}_{k+1}$，假设

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

我们需要的条件为

$$\langle\vec{w}_{k+1},\vec{w}_i\rangle = 0, 1\le i\le k$$

那么$\forall 1\le i\le k$

$$
\begin{aligned}
\langle \vec{w}_{k+1}, \vec{w}_i \rangle &= \langle \vec{v}_{k+1} - \sum_{j=1}^{k} \alpha_j \vec{w}_j, \vec{w}_i \rangle \\
&= \langle \vec{v}_{k+1}, \vec{w}_i \rangle - \sum_{j=1}^{k} \alpha_j \langle \vec{w}_j, \vec{w}_i \rangle \\
&= \langle \vec{v}_{k+1}, \vec{w}_i \rangle - \alpha_i \langle \vec{w}_i, \vec{w}_i \rangle \\
&= 0 \\
\alpha_i &= \frac{\langle \vec{v}_{k+1}, \vec{w}_i \rangle}{\langle \vec{w}_i, \vec{w}_i \rangle}
\end{aligned}
$$

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$k = 2 \ldots n$
    1. 对$i = 1, \ldots, k-1$，计算
    $$
    \alpha_i = \frac{\langle \vec{v}_{k+1}, \vec{w}_i \rangle}{\langle \vec{w}_i, \vec{w}_i \rangle}
    $$
    2.
    $$
    \vec{w}_k = \vec{v}_k - \sum_{i=1}^{k-1} \alpha_i \vec{w}_i
    $$

## Problem 2

(a)拉朗格朗日乘子为

$$
\Lambda(\vec{x}, \vec{\lambda}, \vec{\mu}) = \vec{c}^T \vec{x} - \vec{\lambda}^T (A\vec{x} - \vec{b}) - \vec{\mu}^T \vec{x}
$$

所以KKT条件为

1. stationarity
$$
\nabla_{\vec{x}} \Lambda(\vec{x}, \vec{\lambda}, \vec{\mu}) = \vec{c} - A^T \vec{\lambda} - \vec{\mu} = \vec{0}
$$
2. primal feasibility
$$
\begin{aligned}
A\vec{x} &= \vec{b} \\
\vec{x} &\geq \vec{0}
\end{aligned}
$$
3. complementary slackness
$$
\mu_j x_j = 0
$$
4. dual feasibility

$$\mu_j \geq 0$$

(b)令

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

记

$$\tilde{A} = \left[\begin{array}{c|c} A & \vec{0} \\ \hline \vec{v}^T & 1 \end{array}\right], \tilde{\vec{b}} = \begin{bmatrix} \vec{b} \\ d \end{bmatrix}, \tilde{\vec{c}} = \begin{bmatrix} \vec{c} \\ 0 \end{bmatrix}, \tilde{\vec{x}} = \begin{bmatrix} \vec{x} \\ x_{n+1} \end{bmatrix}$$

那么

$$\begin{aligned} \tilde{\vec{c}}^T\tilde{\vec{x}} &= \vec{c}^T\vec{x} \\ \tilde{A}\tilde{\vec{x}} &= \begin{bmatrix} A\vec{x} \\ \vec{v}^T\vec{x} + x_{n+1} \end{bmatrix} = \begin{bmatrix} \vec{b} \\ d \end{bmatrix} = \tilde{\vec{b}} \\ \tilde{\vec{x}} &= \begin{bmatrix} \vec{x} \\ x_{n+1} \end{bmatrix} \geq 0 \end{aligned}$$

所以新的线性规划问题为

$$\begin{aligned} \text{minimize } & \tilde{\vec{c}}^T\tilde{\vec{x}} \\ \text{such that } & \tilde{A}\tilde{\vec{x}} = \tilde{\vec{b}} \\ & \tilde{\vec{x}} \geq \vec{0} \end{aligned}$$

(c)对偶问题可以化为

$$\begin{aligned} \text{minimize } & -\vec{b}^T\vec{y} \\ \text{such that } & -A^T\vec{y} + \vec{c} \geq \vec{0} \end{aligned}$$

记

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

所以stationarity条件为

$$\begin{aligned} \nabla_{\vec{y}}\Lambda'(\vec{y}, \vec{\lambda}) = -\vec{b} + A\vec{\lambda} = \vec{0} \\ \vec{b} = A\vec{\lambda} \end{aligned}$$

带回原式可得

$$\begin{aligned} \Lambda'(\vec{y}, \vec{\lambda}) &= -\vec{\lambda}^T A^T\vec{y} + \vec{\lambda}^T A^T\vec{y} - \vec{\lambda}^T\vec{c} \\ &= -\vec{\lambda}^T\vec{c} \\ &= -\vec{c}^T\vec{\lambda} \end{aligned}$$

complementary slackness条件为

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

所以在最优解处

$$\begin{aligned}-\vec{b}^T\vec{y}&=\Lambda'(\vec{y},\vec{\lambda})=-\vec{c}^T\vec{\lambda}\\ \vec{b}^T\vec{y}&=\vec{c}^T\vec{\lambda}\end{aligned}$$

因为驻点唯一，所以如下方程的解唯一：

$$A\vec{\lambda}=\vec{b} \tag{1}$$
$$\vec{\lambda}\geq\vec{0} \tag{2}$$

设上述方程的解为

$$\vec{\lambda}'$$

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

$$A\vec{x}=\vec{b} \tag{5}$$
$$\vec{x}\geq\vec{0} \tag{6}$$

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

$$\nabla_{\vec{x}}g(\vec{x})=2\vec{x}$$

所以随机梯度下降法计算的梯度为

$$\frac{2}{k}\sum_{i=1}^{k}(\vec{x}_i-\vec{x})=2\left(\frac{1}{k}\sum_{i=1}^{k}\vec{x}_i-\vec{x}\right)$$

(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

$$\forall \vec{y}, \text{或者} \exists i, s.t\ f_i(\vec{x}) > f_i(\vec{y}), \text{或者} \forall i, f_i(\vec{x}) \geq f_i(\vec{y})$$

(i)如果第一个条件不成立，即$\forall y, i$，

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

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$f_i \neq f_j$，那么必然存在函数值不相同的点$\vec{x}$，设

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

(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

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

构造拉格朗日乘子：

$$L(\vec{x}, \gamma) = \sum_{i=1}^{k} \gamma_i f_i(\vec{x}) - \sum_{i=1}^{k} \alpha_i \gamma_i - \beta \left( \sum_{i=1}^{k} \gamma_i - 1 \right)$$

关于$\vec{x}, \gamma$求梯度并为0可得

$$\begin{aligned} \nabla_{\vec{x}} L(\vec{x}, \gamma) &= \sum_{i=1}^{k} \gamma_i \nabla_{\vec{x}} f_i(\vec{x}) = \vec{0} \\ \frac{\partial L(\vec{x}, \gamma)}{\partial \gamma_i} &= f_i(\vec{x}) - \alpha_i - \beta = 0, i = 1, \ldots, n \end{aligned}$$

所以

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

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

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

以及

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

注意我们有

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

如果不存在$i, \vec{x}$，使得

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

那么必然有

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

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$\vec{x}$求梯度可得

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

因为$\forall i$

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

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

注意$\forall \vec{x} \neq \vec{x}'$

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

所以必然存在$i$，使得

$$\begin{aligned}(f_i(\vec{x}') - z_i)^2 &< (f_i(\vec{x}) - z_i)^2 \\ f_i(\vec{x}') - z_i &< f_i(\vec{x}) - z_i \\ f_i(\vec{x}') &< f_i(\vec{x})\end{aligned}$$

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)

$$\begin{aligned}\frac{\partial f}{\partial x_1} &= (x_1^2 - x_2) \times 2x_1 + (x_1 - 1) \\ &= 2x_1^3 - 2x_1x_2 + x_1 - 1 \\ \frac{\partial f}{\partial x_2} &= (x_1^2 - x_2) \times (-1) \\ &= x_2 - x_1^2 \\ \frac{\partial^2 f}{\partial x_1^2} &= \frac{\partial}{\partial x_1}(2x_1^3 - 2x_1x_2 + x_1 - 1) \\ &= 6x_1^2 - 2x_2 + 1 \\ \frac{\partial^2 f}{\partial x_2^2} &= \frac{\partial}{\partial x_2}(x_2 - x_1^2) \\ &= 1 \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} &= \frac{\partial}{\partial x_2}(2x_1^3 - 2x_1x_2 + x_1 - 1) \\ &= -2x_1\end{aligned}$$

所以

$$
\begin{aligned}
\nabla f\Big|_{\vec{x}_0=(2,2)} &= \begin{bmatrix} 9 \\ -2 \end{bmatrix} \\
H = \nabla^2 f\Big|_{\vec{x}_0=(2,2)} &= \begin{bmatrix} 21 & -4 \\ -4 & 1 \end{bmatrix} \\
H^{-1}\nabla f &= \begin{bmatrix} \frac{1}{5} \\ -\frac{6}{5} \end{bmatrix}
\end{aligned}
$$

所以

$$
\begin{aligned}
\vec{x}_1 &= \vec{x}_0 - H^{-1}\nabla f \\
&= \begin{bmatrix} 2 \\ 2 \end{bmatrix} - \begin{bmatrix} \frac{1}{5} \\ -\frac{6}{5} \end{bmatrix} \\
&= \begin{bmatrix} \frac{9}{5} \\ \frac{16}{5} \end{bmatrix}
\end{aligned}
$$

(c)显然

$$
f(\vec{x}_1) < f(\vec{x}_0)
$$

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。