Q1

a) $f(x)=\frac{1}{2}x^T H x.$ $\quad x\in R^n \quad H\in R^{n\times n}$

$$df=\frac{1}{2}[(dx)^T Hx + x^T H\,dx]$$

$$=\frac{1}{2}[(Hx)^T + x^T H]\,dx$$

Since $df=(\nabla f(x))^T dx$, then $\nabla f(x)=\frac{1}{2}(Hx+H^T x)=\frac{1}{2}(H+H^T)x$

$d(\nabla f(x))=\frac{1}{2}(H+H^T)dx$, so $\nabla^2 f(x)=\frac{1}{2}(H+H^T)$

if H is symmetric, $H=H^T$ then $\nabla f(x)=Hx$, $\nabla^2 f(x)=H$

b) $f(x)=b^T A x-\frac{1}{2}x^T A^T A x \quad A\in R^{m\times n} \quad b\in R^m$

$$df=d(b^T A x)-\frac{1}{2}d(x^T A^T A x)$$

$$=(b^T A)dx-\frac{1}{2}[(A^T A x)^T + x^T\cdot(A^T A)]dx$$

then, $\nabla f(x)=A^T b-\frac{1}{2}(A^T A+(A^T A)^T)x = A^T b - A^T A x$

$d(\nabla f(x))=-A^T A\,dx$, so $\nabla^2 f(x)=-A^T A$

c) $f(x)=\|x\|_2=(\sum_{i=1}^n x_i^2)^{\frac{1}{2}}$

Identity matrix

$f(x)=(x^T x)^{\frac{1}{2}}$, $\quad df=\frac{1}{2}(x^T x)^{-\frac{1}{2}}\cdot \underbrace{d(x^T\cdot I x)}_{=x}$

thus, $\nabla f(x)=(x^T x)^{-\frac{1}{2}}\cdot x=\|x\|_2^{-1}\cdot x$

$$d(\nabla f(x))=(x^T x)^{-\frac{1}{2}}dx + x\,d((x^T x)^{-\frac{1}{2}}) \quad ①$$

$$d((x^T x)^{-\frac{1}{2}})=-(x^T x)^{-\frac{3}{2}}\cdot d(x^T x)=-(x^T x)^{-\frac{3}{2}}x^T dx \quad ②$$

So from ① and ② we can have:

$$\nabla^2 f(x)=-(x^T x)^{-\frac{3}{2}}xx^T+(x^T x)^{-\frac{1}{2}}\cdot I$$

Identity matrix

$$=-\|x\|_2^{-3}xx^T+\|x\|_2^{-1}I$$

1 d) $f(x) = \|Ax-b\|_2 = \left((Ax-b)^T (Ax-b)\right)^{\frac{1}{2}}$

$$df(x) = \frac{1}{2}\left((Ax-b)^T(Ax-b)\right)^{-\frac{1}{2}} \cdot d\left((Ax-b)^T(Ax-b)\right)$$

from (a) ~ (c), we can get:

$$df(x) = \frac{1}{2}\|Ax-b\|_2^{-1}(2x^TA^TA - 2b^TA)dx$$

$$= \|Ax-b\|_2^{-1}(x^TA^TA - b^TA)dx$$

so $\nabla f(x) = \|Ax-b\|_2^{-1}(A^TAx - A^Tb)$

$$d(\nabla f(x)) = (A^TAx - A^Tb)\, d\left(\|Ax-b\|_2^{-1}\right) + \|Ax-b\|_2^{-1} d(A^TAx)$$

$$= (A^TAx - A^Tb)(-\|Ax-b\|_2^{-2})\, d\|Ax-b\|_2 + \|Ax-b\|_2^{-1}A^TA\,dx$$

$$= -\|Ax-b\|_2^{-2}(A^TAx - A^Tb)\|Ax-b\|_2^{-1}(x^TA^TA - b^TA)dx + \|Ax-b\|_2^{-1}A^TA\,dx$$

So $\nabla^2 f(x) = -\|Ax-b\|_2^{-3}(A^TAxx^TA^TA - 2A^TAxb^TA + A^Tbb^TA) + \|Ax-b\|_2^{-1}A^TA$

$$= -\|Ax-b\|_2^{-3}A^T(Ax-b)(Ax-b)^TA + \|Ax-b\|_2^{-1}A^TA$$

2)

a.) $f(x+s) = \theta(1) = \theta(0) + \int_0^1 \theta'(t)\,dt = \theta(0) + \theta'(0) + \int_0^1 \theta'(t) - \theta'(0)\,dt$

so, $\theta(1) - \theta(0) - \theta'(0) = \int_0^1 \theta'(t) - \theta'(0)\,dt$

take absolute value:

$$|\theta(1) - \theta(0) - \theta'(0)| = \left|\int_0^1 \theta'(t) - \theta'(0)\,dt\right| \le \int_0^1 |\theta'(t) - \theta'(0)|\,dt \quad ①$$

$\theta'(t) = \nabla f(x+ts)^T s \qquad \theta'(0) = \nabla f(x)^T s$, so $\int_0^1 |\nabla f(x+ts)^T s - \nabla f(x)^T s|\,dt$

$$\int_0^1 |\nabla f(x+ts)^T s - \nabla f(x)^T s|\,dt \le \int_0^1 \|s\|_2 \cdot \|\nabla f(x+ts)^T - \nabla f(x)^T\|_2\,dt \quad ②$$

Since for L-lipschitz continuous gradient: $\|\nabla f(x)-\nabla f(y)\| \le L\|x-y\|$

So from ② we can have:

$$\int_0^1 \left|\nabla f(x+ts)^T s - \nabla f(x)^T s\right| dt \le \int_0^1 L\|s\|_2^2 \, t \, dt = \frac{1}{2} L\|s\|_2^2$$

from ① and ② we can have: $\left|f(x+s)-f(x)-\nabla f(x)^T s\right| \le \frac{1}{2} L\|s\|_2^2$

b) Prove: $\theta(1) = \theta(0) + \theta'(0) + \int_0^1 \int_0^t \theta''(\alpha) \, d\alpha \, dt$

right hand side $= \theta(0)+\theta'(0)+\int_0^1\int_0^t \theta''(\alpha)\, d\alpha \, dt = \theta(0)+\theta'(0)+\int_0^1 \left[\theta'(t)\Big|_0^\alpha\right] d\alpha$

$= \theta(0)+\theta'(0)+\int_0^1 \left[\theta'(\alpha)-\theta'(0)\right] d\alpha = \theta(0)+\theta'(0)+\theta(1)-\theta(0)-\theta'(0) = \theta(1)$

left hand side

$$f(x+s) = \theta(1) = \theta(0)+\theta'(0)+\int_0^1\int_0^t \theta''(\alpha)\, dt \, d\alpha$$

since $\int_0^1\int_0^\alpha \theta''(0)\, dt \, d\alpha = -\frac{1}{2}\theta''(0)$

$$\theta(1) = \theta(0)+\theta'(0)+\frac{1}{2}\theta''(0)+\int_0^1\int_0^\alpha \theta''(t)-\theta''(0)\, dt\, d\alpha$$

so $\theta(1)-\theta(0)-\theta'(0)-\frac{1}{2}\theta''(0) = \int_0^1\int_0^\alpha \theta''(t)-\theta''(0)\, dt\, d\alpha$

take absolute value on both sides:

$$\left|\theta(1)-\theta(0)-\theta'(0)-\frac{1}{2}\theta''(0)\right| = \int_0^1\int_0^\alpha \theta''(t)-\theta''(0)\, dt\, d\alpha \le \int_0^1\int_0^\alpha \left|\theta''(t)-\theta''(0)\right| dt\, d\alpha \quad ①$$

$\theta''(t) = s^T \nabla^2 f(x+ts)\, s \qquad \theta''(0) = s^T \nabla^2 f(x)\, s$

$$\int_0^1\int_0^\alpha \left|\theta''(t)-\theta''(0)\right| dt\, d\alpha \le \int_0^1\int_0^\alpha \|s\|_2^2 \cdot \|\nabla^2 f(x+ts)-\nabla^2 f(x)\|_2 \, dt\, d\alpha \quad ②$$

With Lipschitz-continuous property, we have:

$$\int_0^1\int_0^\alpha \|S\|_2^2 \cdot \|\nabla^2 f(x+ts) - \nabla^2 f(x)\|_2 \, dt \, d\alpha \leq \int_0^1\int_0^\alpha \|S\|_2^2 Q \|ts\|_2 \, dt \, d\alpha \quad ③$$

$$\int_0^1\int_0^\alpha \|S\|_2^2 Q \|ts\|_2 \, dt \, d\alpha = \|S\|_2^3 Q \int_0^1\int_0^\alpha |t| \, dt \, d\alpha = \frac{1}{6} Q \cdot \|S\|_2^3 \quad ④$$

with ①②③④, we can have $\left| f(x+s) - f(x) - \nabla f(x)^T s - \frac{1}{2} s^T \nabla^2 f(x) s \right| \leq \frac{1}{6} Q \|S\|_2^3$

3) Since x^* is a global minimizer, then $\nabla f(x^*) = 0$

Since f is a convex function, for $\forall x, y \in R^n$

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

if x^* is not unique then $\exists z \in R^n$ that $f(z) = f(x^*)$

then $f(x^* + \alpha(z - x^*)) = f((1-\alpha)x^* + \alpha z) \leq \underbrace{(1-\alpha) f(x^*) + \alpha f(z)}_{= f(x^*)}$, $\alpha \in [0,1]$

Since $f(x^*)$ is global minimizer, then $f((1-\alpha)x^* + \alpha z) = f(x^*)$

which means for $\forall s \in [x^*, z]$ $f(s) = f(x^*)$

then $\nabla^2 f(x^*) = 0$ which contradicts to $\nabla^2 f(x^*) > 0$

So the global minimizer of f is unique

4) let $x = (H_x, M_x, F_x)$ $y: (H_y, M_y, F_y)$, $x, y \in P$

a) $\lambda \in [0,1]$

we have:

$H_x + M_x + F_x \le 100$, $H_x, M_x \ge 15$ $F_x \ge M_x$ $50 \le M_x + F_x \le 80$ $H_x + M_x + F_x \ge 90$

$H_y + M_y + F_y \le 100$, $H_y, M_y \ge 15$ $F_y \ge M_y$ $50 \le M_y + F_y \le 80$ $H_y + M_y + F_y \ge 90$

then:

$\lambda(H_x + M_x + F_x) + (1-y)(H_y + M_y + F_y) \le \lambda \cdot 100 + (1-\lambda)100 = 100$

$\lambda H_x + (1-\lambda) H_y \ge (\lambda + (1-\lambda)) \cdot 15 = 15$

$\lambda M_x + (1-\lambda) M_y \ge (\lambda + (1-\lambda)) \cdot 15 = 15$

$\lambda F_x + (1-\lambda) F_y \ge \lambda M_x + (1-\lambda) M_y$

$50 = (\lambda + 1-\lambda) 50 \le \lambda(M_x + F_x) + (1-\lambda)(M_y + F_y) \le (\lambda + 1-\lambda) 80 = 80$

$\lambda(H_x + M_x + F_x) + (1-\lambda)(H_y + M_y + F_y) \ge (\lambda + 1-\lambda) \cdot 90 = 90$

so, $\lambda x + (1-\lambda) y$ satisfies all the constraints, then $\lambda x + (1-\lambda) y \in P$

b)

Since for any $x \in P$, exist $\lambda_p \ge 0$ that $\sum_{p \in S} \lambda_p P = x$ and $\sum_{p \in S} \lambda_p = 1$

reward function: $C_H \cdot H_x + C_M \cdot M_x + C_F \cdot F_x + C_p \cdot (100 - H_x - M_x - F_x)$

$= r(\lambda_p)$, λ_p affects H_x, M_x, F_x; C_H, C_M, C_F are determined for each student

let $Z(x) = (H_x, M_x, F_x, 100 - H_x - M_x - F_x)$ $Z(x) \in \mathbb{R}^4$

$Z_i = Z(s_i)$ $\lambda = \{\lambda_p \mid \lambda_p \ge 0 \ p \in S\}$

then, reward function $r(\lambda) = \sum_{p \in S} \lambda_p \cdot (Z_p^T \cdot c)$

let $c = (C_H, C_M, C_F, C_p)$ $c \in \mathbb{R}^4$

let λ_i, λ_j be the two components of λ that is not 0, $\lambda = \{\lambda_i, \lambda_j \dots\}$

λ_i, λ_j are coefficients of S_i, S_j, for simplicity, put them to the front let λ' be $\{\overset{i\text{-th}}{\lambda_i+\lambda_j}, \overset{j\text{-th}}{0} \cdots\cdots\}$, only the first 2 elements are different to λ

let λ'' be $\{0, \lambda_i+\lambda_j \cdots\cdots\}$, only the first 2 elements are different to λ

then $r(\lambda') - r(\lambda) = \lambda_j \cdot (Z_i^T \cdot C) - \lambda_j \cdot (Z_j^T \cdot C) = \lambda_j \cdot (Z_i - Z_j)^T \cdot C$

$r(\lambda'') - r(\lambda) = \lambda_i (Z_j^T \cdot c) - \lambda_i (Z_i^T \cdot c) = \lambda_i (Z_j - Z_i)^T C$

then there must be one of $r(\lambda')$ or $r(\lambda'') \geq r(\lambda)$

let $\lambda = \arg\max\{r(\lambda'), r(\lambda'')\}$, then λ has one more zero element

we can repeat this process until λ has only one non-zero element

which will be one of the corners in S maximizes the score

c) screeshot:

```
for student  1  the maximum course score is 92.5  on corner :( 50.0   25.0   25.0 )

for student  2  the maximum course score is 85.0  on corner :( 15.0   40.0   40.0 )
for student  2  the maximum course score is 85.0  on corner :( 20.0   40.0   40.0 )
for student  2  the maximum course score is 85.0  on corner :( 50.0   25.0   25.0 )
for student  2  the maximum course score is 85.0  on corner :( 40.0   25.0   25.0 )
for student  2  the maximum course score is 85.0  on corner :( 15.0   37.5   37.5 )
for student  2  the maximum course score is 85.0  on corner :( 15.0   15.0   65.0 )
for student  2  the maximum course score is 85.0  on corner :( 20.0   15.0   65.0 )
for student  2  the maximum course score is 85.0  on corner :( 50.0   15.0   35.0 )
for student  2  the maximum course score is 85.0  on corner :( 40.0   15.0   35.0 )
for student  2  the maximum course score is 85.0  on corner :( 15.0   15.0   60.0 )

for student  3  the maximum course score is 86.5  on corner :( 15.0   15.0   60.0 )
```