1) a) $f(x) = a^T x$

Let $x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$ $a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}_{n\times 1}$

$f(x) = a^T x = a_1 x_1 + a_2 x_2 + \cdots + a_n x_n$

$$\nabla f(x) = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix} = a$$

$\therefore \nabla f(x) = a$

1) b) Let $A = [\hat{a}_1 \ \hat{a}_2 \ \hat{a}_3 \ . \ . \ \hat{a}_m]^T$

$$Ax = \begin{bmatrix} \hat{a}_1^T x \\ \hat{a}_2^T x \\ \vdots \\ \hat{a}_m^T x \end{bmatrix}$$

$$\nabla(Ax) = \begin{bmatrix} \nabla(\hat{a}_1^T x) \\ \nabla(\hat{a}_2^T x) \\ \vdots \\ \nabla(\hat{a}_m^T x) \end{bmatrix} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_m \end{bmatrix}^T = A^T$$

1c) $x^T A x = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$= \begin{bmatrix} \sum_{i=1}^{n} a_{i1} x_i & \cdots & \sum_{i=1}^{n} a_{in} x_i \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

$= \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij} x_i$

$x^T A x = \sum_{j=1}^{n} \sum_{i=1}^{n} a_{ij} x_j x_i$

$\frac{\partial x^T A x}{\partial x_p} = \sum_{j=1}^{n} a_{pj} x_j + \sum_{j=1}^{n} a_{jp} x_j$

$\frac{\partial x^T A x}{\partial x} = \begin{bmatrix} \sum_{j=1}^{n} a_{1j} x_j + \sum_{j=1}^{n} a_{j1} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj} x_j + \sum_{j=1}^{n} a_{jn} x_j \end{bmatrix}$

$$= \begin{bmatrix} \sum_{j=1}^{n} a_{1j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{nj} x_j \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^{n} a_{j1} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{jn} x_j \end{bmatrix}$$

$$\downarrow \qquad\qquad \downarrow$$

$$= Ax + A^T x$$

$$\nabla(x^T A x) = (A + A^T) x$$

1)d) $f(x) = \dfrac{e^{a^T x}}{1+e^{a^T x}}$

$f(x) = 1 - \dfrac{1}{1+e^{a^T x}}$

$\nabla f = \left(\dfrac{1}{1+e^{a^T x}}\right)^2 \times e^{a^T x} \times \nabla(a^T x)$ (∵ chain rule)

$= \dfrac{e^{a^T x}}{(1+e^{a^T x})^2} \times a = \dfrac{a e^{a^T x}}{(1+e^{a^T x})^2}$

1)e) $f(x) = \|Ax-b\|_2^2 = (Ax-b)^T(Ax-b)$

$\nabla f = \nabla\left((Ax-b)^T(Ax-b)\right)$

$= 2(Ax-b)\nabla((Ax-b))$

$= 2A^T(Ax-b)$

$= 2A^TAx - 2A^Tb$

1)f) $\nabla f = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & & \vdots \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix}$

$$\frac{\partial f_m}{\partial x_m} = \frac{\left[\sum_{j=1}^{n} e^{x_j}\right]\left[e^{x_m}\right] - e^{x_m} \cdot e^{x_m}}{\left(\sum_{j=1}^{n} e^{x_j}\right)^2}$$

$$\frac{\partial f_m}{\partial x_m} = \frac{\left(\sum_{\substack{j=1 \\ j \neq m}}^{n} e^{x_j}\right) e^{x_m}}{\left(\sum_{j=1}^{n} e^{x_j}\right)^2}$$

$$\frac{\partial f_m}{\partial x_n} = \frac{-e^{x_m} \cdot e^{x_n}}{\left(\sum_{j=1}^{n} e^{x_j}\right)^2} \quad \text{where } m \neq n$$

Let $S = \left(\sum_{j=1}^{n} e^{x_j}\right)^2$

$$\therefore \nabla f = \begin{bmatrix} \dfrac{e^{x_1}\cdot\left(\sum_{j\neq 1} e^{x_j}\right)}{s} & \dfrac{-e^{x_1+x_2}}{s} & \cdots & \dfrac{-e^{x_1+x_n}}{s} \\ \dfrac{-e^{x_2+x_1}}{s} & \dfrac{e^{x_2}\cdot\left(\sum_{j\neq 2} e^{x_j}\right)}{s} & \cdots & \dfrac{-e^{x_2+x_n}}{s} \\ \vdots & & & \\ \dfrac{-e^{x_n+x_1}}{s} & & \cdots & \dfrac{e^{x_n}\cdot\left(\sum_{j\neq n} e^{x_j}\right)}{s} \end{bmatrix}$$

$$\therefore \nabla f = \frac{1}{s}\begin{bmatrix} e^{x_1}\cdot\sum_{j\neq 1} e^{x_j} & -e^{x_1+x_2} & \cdots & -e^{x_1+x_n} \\ -e^{x_2+x_1} & e^{x_2}\cdot\sum_{j\neq 2} e^{x_j} & \cdots & -e^{x_2+x_n} \\ \vdots & & & \\ -e^{x_n+x_1} & -e^{x_n+x_2} & \cdots & e^{x_n}\cdot\sum_{j\neq n} e^{x_j} \end{bmatrix}$$

where $s = \left(\sum_{j=1}^{n} e^{x_j}\right)^2$

1) g)

$$f(x) = \sum_{i=1}^{n} x_i \ln \frac{x_i}{\alpha_i}$$

$$= \sum_{i=1}^{n} x_i \ln x_i - \sum_{i=1}^{n} x_i \ln \alpha_i$$

$$\frac{\partial f}{\partial x_i} = 1 + \ln x_i - \ln \alpha_i = 1 + \ln \frac{x_i}{\alpha_i}$$

$$\therefore \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 1 + \ln \frac{x_1}{\alpha_1} \\ 1 + \ln \frac{x_2}{\alpha_2} \\ \vdots \\ 1 + \ln \frac{x_n}{\alpha_n} \end{bmatrix}$$

1) a) $C = \{ x \in \mathbb{R}^n \mid x^T A x + b^T x + c \leq 0 \}$

Let $f(x) = x^T A x + b^T x + c$

$C = \{ x \in \mathbb{R}^n \mid f(x) \leq 0 \}$

To show C is convex it is sufficient to show f is convex if $A \succeq 0$

$A \succeq 0$

Let $x, y \in C$ then

$x^T A x + b^T x + c \leq 0$, $y^T A y + b^T y + c \leq 0$

Let $z = tx + (1-t)y$ where $t \in (0,1)$

If we show $z \in C$ then we are done.

Assume $f(z) > t f(x) + (1-t) f(y)$

$$\Rightarrow f(tx + (1-t)y) > t f(x) + (1-t) f(y)$$

$$\Rightarrow t^2 x^T A x + (1-t)^2 y^T A y + t(1-t) x^T A y + t(1-t) y^T A x$$
$$+ t b^T x + (1-t) b^T y + c > t x^T A x + t b^T x$$
$$+ c + (1-t) y^T A y$$
$$+ (1-t) b^T y + (1-t) c$$

$$\Rightarrow t(1-t) x^T A x + t(1-t) y^T A y - t(1-t) x^T A y$$
$$- t(1-t) y^T A x < 0$$

$$\Rightarrow x^T A x + y^T A y - x^T A y - y^T A x < 0$$

$$\Rightarrow (x-y)^T A (x-y) < 0$$

$\Rightarrow \in$ contradiction since $A \succeq 0$

$\therefore$ $f(z) \leq t f(x) + (1-t) f(y)$

$\therefore$ f is convex ($\because$ by defn of convex function)

$\therefore$ C is convex if $A \succeq 0$

b) Similarly let $B = A + \lambda_0 aa^T \geq 0$

where $a^T x + \beta = 0$ let x, y, z be intersection of C & hyperplane

Assume $f(z) > t f(x) + (1-t) f(y)$

$$(x-y)^T A (x-y) < 0$$

$$(x-y)^T (B - \lambda_0 aa^T)(x-y) < 0$$

$$(x-y)^T B (x-y) - \lambda_0 (x-y)^T a a^T (x-y) < 0$$

$$(x-y)^T B(x-y) - \lambda_0 x^T a a^T x + \lambda_0 x^T a a^T y + \lambda_0 y^T a a^T x - \lambda_0 y^T a a^T y < 0$$

$$(x-y)^T B(x-y) - \lambda_0 \cancel{(-\beta)^T (-\beta)} + \lambda_0 (-\beta)^T (-\beta) + \lambda_0 (-\beta)^T (-\beta) - \lambda_0 \cancel{(-\beta)^T (-\beta)^T} < 0$$

$$(x-y)^T B (x-y) < 0$$

$\Rightarrow \Leftarrow$ since $B \geq 0$

$\therefore$ If $A + \lambda_0 aa^T \geq 0$ then f is convex similarly Assume $g(z) > t\,g(x) + (1-t)\,g(y)$

$$\alpha^T z + \beta > t(\alpha^T x + \beta) + (1-t)(\alpha^T y + \beta)$$

$$\alpha^T(tx + (1-t)y) + \beta > t\alpha^T x + t\beta + (1-t)\alpha^T y + (1-t)\beta$$

$$t\alpha^T x + (1-t)\alpha^T y + \beta > t\alpha^T x + (1-t)\alpha^T y + \beta$$

$$0 > 0$$

$$\Rightarrow \in$$

$\therefore\ g(z) \leq t\,g(x) + (1-t)\,g(y)$

$\therefore$ g is convex.

$\therefore$ Intersection of f & g is convex if

$A + X_0 \alpha\alpha^T \geq 0$.

2) Show that $x = (1, \frac{1}{2}, -1)$ is optimal solution for minimization problem.

$$\min_{x} \ \frac{1}{2} x^T A x + b^T x + c$$

$$\text{s.t} \quad -1 \le x_i \le 1 \quad i = 1, 2, 3$$

where $A = \begin{bmatrix} 13 & 12 & -2 \\ 12 & 17 & 6 \\ -2 & 6 & 12 \end{bmatrix}$ $\quad b = \begin{bmatrix} -22 \\ -14.5 \\ 13 \end{bmatrix}$

Soln $\quad L(x, \lambda, \mu) = \frac{1}{2} x^T A x + b^T x + c + \left[x - \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right]^T \lambda - \left[x + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}\right]^T \mu$

$\lambda, \mu \ge 0$

$$\nabla L(x^*, \lambda^*, \mu^*) = 0$$

$$A x^* + b + \lambda^* - \mu^* = 0$$

Take $\lambda^* = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ $\quad \mu^* = \begin{bmatrix} 0 \\ 0 \\ 2 \end{bmatrix}$

$$A x^* + \begin{bmatrix} -21 \\ -14.5 \\ 11 \end{bmatrix} = 0 \Rightarrow A x^* = \begin{bmatrix} 21 \\ 14.5 \\ -11 \end{bmatrix}$$

$$x^* = A^{-1}\begin{bmatrix} 21 \\ 14.5 \\ -11 \end{bmatrix}$$

$$x^* = \begin{bmatrix} \frac{42}{25} & -\frac{39}{25} & \frac{53}{50} \\ \frac{-39}{25} & \frac{38}{25} & \frac{-51}{50} \\ \frac{53}{50} & \frac{-51}{50} & \frac{77}{100} \end{bmatrix}\begin{bmatrix} 21 \\ 14.5 \\ -11 \end{bmatrix}$$

$$x^* = \begin{bmatrix} 1 \\ 1/2 \\ -1 \end{bmatrix}$$

Hence proved.

3) $\min \quad \dfrac{\|Ax-b\|_2^2}{c^T x + d}$

st

$c^T x + d > 0$

$$L(x,\lambda) = \frac{\|Ax-b\|_2^2}{c^T x + d} + \lambda(-c^T x - d)$$

$$L(x,\lambda) = \frac{\|Ax-b\|_2^2}{c^T x + d} - \lambda(c^T x + d)$$

KKT conditions

$\nabla L(x^*, \tilde{\lambda}) = 0$

$\tilde{\lambda}(c^T x^* + d) = 0$

Since $c^T x^* + d > 0$

Hence $\tilde{\lambda} = 0$

$\therefore \nabla L(x^*, \tilde{\lambda}) = 0$

$$\Rightarrow \frac{2A^T(Ax^* - b)}{c^T x^* + d} - \frac{\|Ax^* - b\|_2^2}{(c^T x^* + d)^2} c = 0$$

$$\Rightarrow \frac{2A^T A x^*}{c^T x^* + d} = \frac{2A^T b}{c^T x^* + d} + \frac{\|Ax^* - b\|_2^2}{(c^T x^* + d)^2} c$$

$$A^T A x^* = A^T b + \frac{\|Ax^* - b\|_2^2}{2(c^T x + d)} c$$

$$x^* = (A^T A)^{-1} A^T b + \frac{\|Ax^* - b\|_2^2}{2(c^T x^* + d)} (A^T A)^{-1} c$$

Let $t = \frac{\|Ax^* - b\|_2^2}{2(c^T x^* + d)}$ where $t \in \mathbb{R}$ since

$(c^T x^* + d) \in \mathbb{R}$

$\|Ax^* - b\|_2^2 \in \mathbb{R}$

$$\boxed{\begin{aligned} &\therefore x^* = (A^T A)^{-1} A^T b + t (A^T A)^{-1} c \\ &\text{Let } x_1 = (A^T A)^{-1} A^T b, \; x_2 = (A^T A)^{-1} c \\ &\therefore x^* = x_1 + t x_2 \end{aligned}}$$

Hence Proved

ii) we know from second order approximation theorem.

$$f(x+t\nabla x) = f(x) + t\,\nabla f(x)^T \nabla x + \frac{t^2}{2}(\nabla x)^T(\nabla^2 f(x))\nabla x$$

$\forall\ t \geq 0$

$$\therefore f(x+t\nabla x) \leq f(x) + t\,\nabla f(x)^T \nabla x + \frac{Mt^2}{2}(\nabla x)^T(\nabla x)$$

Since $(\nabla^2 f(x) \preceq MI)$

we already know stopping criterion

$$f(x+t\nabla x) \leq f(x) + \alpha t\,\nabla f(x)^T \nabla x$$

WLOG Assume

$$f(x) + t\,\nabla f(x)^T \nabla x + \frac{Mt^2}{2}(\nabla x)^T(\nabla x) \leq f(x) + \alpha t\,\nabla f(x)^T \nabla x$$

$$t(1-\alpha)\,\nabla f(x)^T \nabla x + \frac{Mt^2}{2}(\nabla x)^T(\nabla x) \leq 0$$

$$(1-\alpha)\,\nabla f(x)^T \nabla x + \frac{Mt}{2}(\nabla x)^T(\nabla x) \leq 0$$

$$t \leq \frac{-2(1-\alpha)(\nabla f(x))^T \nabla x}{M(\nabla x)^T \nabla x}$$

$$t \leq \frac{-2(1-\alpha)(\nabla f(x)\cdot\nabla x)}{M\|\nabla x\|_2^2}$$

$$\frac{-(\nabla f(x)\cdot\nabla x)}{M\|\nabla x\|_2^2} \leq \frac{-2(1-\alpha)(\nabla f(x)\cdot\nabla x)}{M\|\nabla x\|_2^2}$$

for some $\alpha \in (0,1)$

$\therefore$ Backtracking stopping condition holds for

$$0 < t \leq \frac{-(\nabla f(x) \cdot \nabla x)}{M \|\nabla x\|_2^2}$$

Assume $\dfrac{-2(1-\alpha)(\nabla f(x) \cdot x)}{M \|\nabla x\|_2^2} \leq 1$

$$\Rightarrow \quad \beta^k t \leq \frac{-2(1-\alpha)(\nabla f(x) \cdot \nabla x)}{M \|\nabla x\|_2^2}$$

$$\Rightarrow \quad k \leq \frac{\log\left(\dfrac{-2(1-\alpha)(\nabla f(x) \cdot \nabla x)}{t M \|\nabla x\|_2^2}\right)}{\log(\beta)}$$

5) Given $f(x) = \frac{1}{2}x_1^2 + \gamma x_2^2$

we know $x_{k+1} = x_k - t_k \nabla f(x_k)$

$$\nabla f(x_k) = \begin{bmatrix} x_{k_1} \\ 2\gamma x_{k_2} \end{bmatrix}$$

$$\Rightarrow x_{k+1} = x_k - t \nabla f(x_k)$$

$$\Rightarrow \begin{bmatrix} x_{(k+1)_1} \\ x_{(k+1)_2} \end{bmatrix} = \begin{bmatrix} x_{k_1} - t x_{k_1} \\ x_{k_2} - 2\gamma t x_{k_2} \end{bmatrix} = \begin{bmatrix} (1-t) x_{k_1} \\ (1-2\gamma t) x_{k_2} \end{bmatrix}$$

$$\begin{bmatrix} x_{1,1} \\ x_{1,2} \end{bmatrix} = \begin{bmatrix} (1-t)\gamma \\ (1-2\gamma t) \end{bmatrix}$$

$$\begin{bmatrix} x_{2,1} \\ x_{2,2} \end{bmatrix} = \begin{bmatrix} (1-t)^2 \gamma \\ (1-2\gamma t)^2 \end{bmatrix}$$

$$\therefore \begin{bmatrix} x_{k,1} \\ x_{k,2} \end{bmatrix} = \begin{bmatrix} (1-t)^k \gamma \\ (1-2\gamma t)^k \end{bmatrix}$$

Now

$$f(x_k - t\nabla f(x_k)) = \frac{1}{2}(1-t)^2 x_{k_1}^2 + \gamma(1-2\gamma t)^2 x_{k_2}^2$$

$$= \frac{1}{2}(1-t)^2 (1-t)^{2k} \gamma^2 + \gamma(1-2\gamma t)^2 (1-2\gamma t)^{2k}$$

$$f(x_k - t\nabla f(x_k)) = \frac{\gamma^2}{2}(1-t)^{2k+2} + \gamma(1-2\gamma t)^{2k+2}$$

$$\frac{df}{dt} = 0$$

$$-\gamma^2(k+1)(1-t)^{2k+1} - 4\gamma^2(k+1)(1-2\gamma t)^{2k+1} = 0$$

$$\left(\frac{t-1}{1-2\gamma t}\right) = 4^{\frac{1}{2k+1}}$$

$$\lim_{k\to\infty} \frac{t-1}{1-2\gamma t} = \lim_{k\to\infty} 4^{\frac{1}{2k+1}}$$

$$\therefore \frac{t-1}{1-\gamma 2t} = 1$$

$$t-1 = 1-2\gamma t \Rightarrow t = \frac{2}{1+2\gamma}$$

$$\therefore f(x_k) = \frac{1}{2} x_{k_1}^2 + \gamma x_{k_2}^2$$

$$= \frac{1}{2}\gamma^2(1-t)^{2k} + \gamma(1-2\gamma t)^{2k}$$

$$= \frac{1}{2}\gamma^2\left(1 - \frac{2}{1+2\gamma}\right)^{2k} + \gamma\left(1 - \frac{4\gamma}{1+2\gamma}\right)^{2k}$$

$$= \frac{\gamma^2}{2}\left(\frac{2\gamma-1}{2\gamma+1}\right)^{2k} + \gamma\left(\frac{2\gamma-1}{2\gamma+1}\right)^{2k}$$

$$\therefore f(x_k) = \left(\frac{2\gamma-1}{2\gamma+1}\right)^{2k}\left[\frac{\gamma^2}{2} + \gamma\right]$$