Homework 7

3.3.2 a) $A \geq 0 \iff \forall\, w \in \mathbb{R}^n \quad w^T A w \geq 0$. Using $w = e_i$ (the ith unit vector) you see

$$0 \leq e_i^T A e_i = a_{ii} \qquad \therefore \text{ All diagonal elts in } A \text{ are } \geq 0$$

b) Take $A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$. Then A not psd since $\det(A) = -1$ is a principal minor. But diagonal entries are all ≥ 0.

3.3.4 $g : \mathbb{R}^m \to \mathbb{R}$

$$Bx = \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ b_{i1} & \cdots & b_{in} \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{n} b_{ij} x_j \\ \vdots \\ \sum_{j=1}^{n} b_{ij} x_j \\ \vdots \\ \sum_{j=1}^{n} b_{mj} x_j \end{pmatrix}$$

$$f(x) = g(Bx)$$

$$= g\left(\sum_{j=1}^{n} b_{ij} x_j, \sum_{j=1}^{n} b_{2j} x_j, \ldots, \sum_{j=1}^{n} b_{mj} x_j\right) = g(y_1, \ldots, y_m)$$

$$\nabla f(x) = \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \qquad \frac{\partial f}{\partial x_i} = \frac{\partial g}{\partial x_i} = \frac{\partial g}{\partial y_1}\frac{\partial y_1}{\partial x_i} + \frac{\partial g}{\partial y_2}\frac{\partial y_2}{\partial x_i} + \cdots + \frac{\partial g}{\partial y_m}\frac{\partial y_m}{\partial x_i}$$

$$\frac{\partial y_1}{\partial x_i} = \frac{\partial \sum b_{ij} x_j}{\partial x_i} = b_{1i}$$

$$\therefore \frac{\partial f}{\partial x_i} = \left(\frac{\partial g}{\partial y_1}\right) b_{1i} + \left(\frac{\partial g}{\partial y_2}\right) b_{2i} + \cdots + \left(\frac{\partial g}{\partial y_m}\right) b_{mi}$$

$$= (\nabla g)^T \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{mi} \end{pmatrix} =$$

$$\therefore \nabla f^T = (\nabla g)^T B \qquad \therefore \nabla f = B^T \nabla g$$

Similarly do b)

3.3.5 : $f(x) = \frac{1}{2}x^T A x + b^T x$ $A \in \mathbb{R}^{n\times n}_{symm}$ $b \in \mathbb{R}^n$

a) Show $A \succ 0 \Rightarrow f$ coercive

Suppose f is not coercive. Then $\exists$ a sequence $x^{(1)}, x^{(2)}, \dots$ with $\|x^{(k)}\| \to \infty$ but $f(x^{(k)}) \le \alpha$ for some α.

$$\Rightarrow \frac{1}{2}(x^{(k)})^T A x^{(k)} + b^T x^{(k)} \le \alpha \qquad (*)$$

Now consider the sequence $y^{(k)} := \frac{x^{(k)}}{\|x^{(k)}\|}$. Then $\|y^{(k)}\| = \frac{\|x^{(k)}\|}{\|x^{(k)}\|} = 1$

$\therefore y^{(1)}, y^{(2)}, \dots$ all lie on the boundary of the closed ball $\overline{B(0,1)}$. Since this boundary is closed and bounded, by Bolzano-Weierstrass $\exists$ a subsequence in $y^{(1)}, y^{(2)}, \dots$ that converges to a pt w in this boundary. Call this subseq $y^{(1)}, y^{(2)}, \dots$ again. Have $\|w\| = 1$.

Now $(*) \Rightarrow \frac{1}{2}(x^{(k)})^T A x^{(k)} + b^T x^{(k)} \le \alpha$

$$\Rightarrow \frac{1}{2}\frac{(x^{(k)})^T}{\|x^{(k)}\|} A \frac{x^{(k)}}{\|x^{(k)}\|} + \frac{b^T x^{(k)}}{\|x^{(k)}\|^2} \le \frac{\alpha}{\|x^{(k)}\|^2}$$

As $k \to \infty$ $\|x^{(k)}\| \to \infty$ $\therefore$ we get (substituting $y^{(k)} = x^{(k)}/\|x^{(k)}\|$)

$$\lim_{k\to\infty}\left(\frac{1}{2}(y^{(k)})^T A y^{(k)}\right) + 0 \le 0$$

$$\Rightarrow \frac{1}{2} w^T A w \le 0 \Rightarrow w^T A w \le 0$$

$\Rightarrow A$ not pd.

b) Suppose A not pd. Then $\exists\, w \in \mathbb{R}^n : w^T A w \le 0$

Consider $x = tw$ as $t \to \pm\infty$. Then

$f(x) = \frac{1}{2}t^2(w^T A w) + t\, b^T w$ $b^T w$ is a constant, $w^T A w \le 0$

$\therefore$ As $t \to \pm\infty$ $f(x) \to 0$ or $-\infty$. $\therefore f$ not coercive.

3.3.7 $f(x_1, x_2, x_3) = e^{x_1} + e^{x_2} + e^{x_3} + 2e^{-x_1-x_2-x_3}$

$$\nabla f = \begin{pmatrix} e^{x_1} - 2e^{-x_1-x_2-x_3} \\ e^{x_2} - 2e^{-x_1-x_2-x_3} \\ e^{x_3} - 2e^{-x_1-x_2-x_3} \end{pmatrix}$$

$$Hf = \begin{pmatrix} e^{x_1} + 2e^{-x_1-x_2-x_3} & +2e^{-x_1-x_2-x_3} & 2e^{-x_1-x_2-x_3} \\ 2e^{-x_1-x_2-x_3} & e^{x_2} + 2e^{-x_1-x_2-x_3} & 2e^{-x_1-x_2-x_3} \\ 2e^{-x_1-x_2-x_3} & 2e^{-x_1-x_2-x_3} & e^{x_3} + 2e^{-x_1-x_2-x_3} \end{pmatrix}$$

Critical pts: $\nabla f = 0$ $\quad e^{x_i} = 2e^{-x_1-x_2-x_3}$

$x_i = \ln 2 + (-x_1 - x_2 - x_3)$

$\Rightarrow 2x_1 = \ln 2 - x_2 - x_3$ $\quad \boxed{x_1 = \frac{\ln 2 - x_2 - x_3}{2}}$

$\boxed{x_2 = \frac{\ln 2}{3} - \frac{x_3}{3}}$

$2x_3 = \ln 2 - x_1 - x_2 \quad \Rightarrow \quad 2x_3 = \ln 2 - \left(\frac{\ln 2 - x_2 - x_3}{2}\right) - \left(\frac{\ln 2}{3} - \frac{x_3}{3}\right)$

$2x_3 = \ln 2 - \frac{\ln 2}{2} + \frac{x_2}{2} + \frac{x_3}{2} - \frac{\ln 2}{3} + \frac{x_3}{3} = \frac{\ln 2}{2} - \frac{\ln 2}{3} + \frac{x_2}{2} + \frac{5x_3}{6}$

$\frac{7}{6}x_3 = \frac{\ln 2}{6} + \frac{1}{2}\left(\frac{\ln 2}{3} - \frac{x_3}{3}\right) = \frac{\ln 2}{3} - \frac{x_3}{6}$

$\frac{8}{6}x_3 = \frac{\ln 2}{3} \quad \Rightarrow \quad \boxed{x_3 = \frac{\cancel{6}\,\ln 2}{\cancel{8}_4\,\cancel{3}} = \frac{\ln 2}{4}}$

$\therefore \boxed{x_3 = \frac{\ln 2}{4}}$ $\quad x_2 = \frac{\ln 2}{3} - \frac{\ln 2}{12} = \frac{3}{12}\ln 2 = \boxed{\frac{1}{4}\ln 2 = x_2}$

$x_1 = \frac{\ln 2}{4}$

$\therefore$ Critical pt is $\left(\frac{\ln 2}{4}, \frac{\ln 2}{4}, \frac{\ln 2}{4}\right)$

Plugging this into the Hessian we get: $Hf\left(\frac{\ln 2}{4}, \frac{\ln 2}{4}, \frac{\ln 2}{4}\right)$

$$= \begin{pmatrix} 2^{5/4} & 2^{1/4} & 2^{1/4} \\ 2^{1/4} & 2^{5/4} & 2^{1/4} \\ 2^{1/4} & 2^{1/4} & 2^{5/4} \end{pmatrix}$$

Check if it is pd or psd.
It is pd (check all leading principal minors)
$\therefore$ This is a local min

<u>3.3.10</u> Fix $x^* \in \mathbb{R}^n$, $w \in \mathbb{R}^n$ s.t $\nabla f(x^*)^T w = 0$

a) Suppose $w^T Hf(x^*)^T w < 0$. Consider $x^* + tw$ where $t > 0$

Then $f(x^* + tw) = f(x^*) + \nabla f(x^*)^T tw + \frac{1}{2}(tw)^T Hf(z(t))\, tw$

for some $z(t) \in [x^*, x^* + tw]$

Since $\nabla f(x^*)^T w = 0$ we get

$$f(x^* + tw) = f(x^*) + \frac{1}{2} t^2 w^T Hf(z(t))\, w$$

Now since all 2nd partials are conts, $Hf(z(t)) \to Hf(x^*)$ as $t \to 0$

$\therefore \exists$ some $\bar{t}$ s.t $\forall\ t \le \bar{t}$ $Hf(z(t)) < 0$ since $Hf(x^*) < 0$

$\therefore \forall\ t \in (0, \bar{t}]$ we will get

$f(x^* + tw) < f(x^*)$

b) Suppose $\exists\ \bar{t} > 0$ s.t $f(x^* + tw) < f(x^*)$ $\forall\ t \in (0, \bar{t}]$

Then by same argument as above, we get

$$\frac{1}{2} t^2 w^T Hf(z(t))\, w < 0 \quad \forall\ t \le \bar{t} \Rightarrow w^T Hf(z(t)) w < 0 \quad \forall\ t \le \bar{t}$$

Using continuity of the Hessian we get

$w^T Hf(x^*) w \le 0$ (letting $t \to 0$) ∎