ECE 542 HW1

Shubham Miglani
200284655

1 $v \in R^n$

To prove $v^T v = \text{trace}(v v^T)$
where $\text{trace}(A) = \Sigma_i A_{ii}$

$$v = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_{n \times 1} \qquad v^T = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}_{1 \times n}$$

$$v^T v = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}_{1 \times n} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_{n \times 1}$$

$$= v_1^2 + v_2^2 + v_3^2 + \cdots \; v_n^2 \quad - (1)$$

Now

$$v v^T = \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix}_{n \times 1} \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}_{1 \times n}$$

$$= \begin{bmatrix} v_1^2 + v_1 v_2 & \cdots & + v_1 v_n \\ v_2 v_1 + v_2^2 & \cdots & v_2 v_n \\ \vdots & & \vdots \\ v_n v_1 + v_n v_2 & \cdots & v_n^2 \end{bmatrix}_{n \times n}$$

As defined, $\text{trace}(A) = \Sigma_i A_{ii}$

$\text{trace}(vv^T) = \Sigma_i (vv^T)_{ii}$

$= v_1^2 + v_2^2 + \cdots \cdots v_n^2 \quad - (2)$

(Sum of diagonal elements where row number = column number)

As (1) = (2)

So $v^T v = \text{trace}(v v^T)$

Hence proved.

2

$$\begin{bmatrix} A_{11} & 0 \\ A_{21} & A_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \end{bmatrix}$$

$$A_{11} x_1 + 0 x_2 = b_1$$

$$A_{11} x_1 = b_1$$

As A_{11} is invertible

$$x_1 = (A_{11})^{-1} b_1 \quad — (1)$$

$$A_{21} x_1 + A_{22} x_2 = b_2$$

Using value of $x_1 = (A_{11})^{-1} b_1$ from (1)

$$A_{21} (A_{11})^{-1} b_1 + A_{22} x_2 = b_2$$

$$A_{22} x_2 = b_2 - (A_{21})(A_{11})^{-1} b_1$$

As A_{22} is given as invertible

$$x_2 = (A_{22})^{-1} \left[b_2 - (A_{21})(A_{11})^{-1} b_1 \right]$$

and $x_1 = A_{11}^{-1} b_1$

3 a) $f(x) = \frac{1}{2}(x^T P^T P x) + q^T x + r$

Firstly we know that

$P^T P$ is a symmetric matrix.

as $(P^T P)^T = P^T (P^T)^T = P^T P$

as $(P^T P)^T = P^T P$, It is symmetric.

$$\Delta f(x) = \frac{\partial}{\partial x}\left(\frac{1}{2} x^T P^T P x + q^T x + r\right)$$

First $\frac{\partial}{\partial x}\left(\frac{1}{2} x^T P^T P x\right)$

From class derivatⁿ we know that for $f(x) = x^T A x$ where A is symmetric then $\nabla_x f(x) = 2Ax$

Here we can take $P^T P = A$

we know $P^T P$ is symmetric so

$$\nabla_x\left(\frac{1}{2} x^T P^T P x\right) = \frac{\cancel{2} P^T P x}{\cancel{2}}$$

$$= P^T P x$$

$\nabla_x(q^T x) = q$ (As derived in class)

$\nabla_x(r) = 0$

So
$\nabla_x f(x) = P^T P x + q$

b) For minimum

$$\nabla_x f(x) = 0$$

$$P^T P x + q = 0$$

$$P^T P x = -q$$

$$x = -(P^T P)^{-1} q$$

as we are given P is full rank

Thus $x = -(P^T P)^{-1} q$

$$= -P^{-1} P^{T^{-1}} q$$

is the point where $f(x)$ is minimum

Using this value of x in $f(x)$

$$f(x) = \frac{1}{2} x^T P^T P x + q^T x + r$$

$$= \frac{1}{2}\left(-(P^T P)^{-1} q\right)^T P^T P \left(-P^T P\right)^{-1} q$$

$$+ q^T x - (P^T P)^{-1} q + r$$

$$\epsilon \text{ as } (P^T P)(P^T P)^{-1} = I$$

$$\text{So } f(x) = \frac{1}{2}\left((P^T P)^{-1} q\right)^T q$$

$$- q^T (P^T P)^{-1} q + r$$

$$= \frac{q^T (P^T P)^{-1} q}{2} - q^T (P^T P)^{-1} q$$

$$+ r$$

$$= r - \frac{q^T (P^T P)^{-1} q}{2}$$

4 $$f(x) = \ln\left(\sum_{i=1}^{m} \exp\left(a_i^T x + b_i\right)\right)$$

$\nabla f(x)$

Take $\sum_{i=1}^{m} \exp\left(a_i^T x + b_i\right) = v$

$$\nabla f(x) = \nabla_x \ln v(x)$$

$$= \frac{1}{v(x)} \frac{\partial}{\partial x} v(x) \quad \left[\text{Chain Rule}\right)\right]$$

$$= \frac{1}{\sum_{i=1}^{m} \exp\left(a_i^T x + b_i\right)} \frac{\partial}{\partial x} \sum_{i=1}^{m} \exp\left(a_i^T x + b_i\right)$$

$$\frac{\partial}{\partial x} \sum_{i=1}^{m} \exp\left(a_i^T x + b_i\right) =$$

$$\sum_{i=1}^{m} \frac{\partial}{\partial x}\left(\exp\left(a_i^T x + b_i\right)\right)$$

$$= \sum_{i=1}^{m} \exp\left(a_i^T x + b_i\right) \frac{\partial}{\partial x}\left(a_i^T x + b_i\right)$$

$$= \sum_{i=1}^{m} \exp\left(a_i^T x + b_i\right) a_i$$

as we already know from class

$$\frac{\partial}{\partial x} a_i^T x = a_i$$

So $\nabla f(x)$

$$= \frac{\sum_{i=1}^{m} \exp(a_i^T x + b_i) a_i}{\sum_{i=1}^{m} \exp(a_i^T x + b_i)}$$

This can also be written in another way

Take $z = \begin{bmatrix} \exp(a_1^T x + b_1) \\ \exp(a_2^T x + b_2) \\ \vdots \\ \exp(a_m^T x + b_m) \end{bmatrix}$

Then $\nabla f(x) = \dfrac{a^T z}{[1\ 1\ \ldots\ldots]\ [z]}$

or take $[1\ 1\ \ldots\ldots\ 1]_{1 \times m} = O^T$

where $O \in R^m$

$$\nabla f(x) = \frac{a^T z}{O^T z}$$