COMP/ENGN8535 Homework Assignment #1 Solutions

1. (20 points) The *Frobenius norm* of a matrix $H \in \mathbb{R}^{n \times m}$ is defined as

$$\|H\|_F = \left(\sum_{i=1}^{n} \sum_{j=1}^{m} H_{ij}^2 \right)^{1/2}.$$

Prove the following:

(a) (5 points) $\|H\|_F$ is equal to the square root of the matrix trace of HH^T.

$$\|H\|_F = \sqrt{\mathbf{tr}(HH^T)}$$

Solution:

$$\mathbf{tr}(HH^T) = \sum_i [HH^T]_{ii} = \sum_i H_i H_i^T \tag{1}$$

$$= \sum_i \sum_j H_{ij}^2 = \|H\|_F^2. \tag{2}$$

(b) (5 points) Any orthogonal matrix $P \in \mathbb{R}^{n \times n}$ or $Q \in \mathbb{R}^{m \times m}$ preserves the Frobenius norm, that is,

$$\|H\|_F = \|PH\|_F = \|HQ\|_F$$

Solution:

$$\mathbf{tr}(HH^T) = \mathbf{tr}(HIH^T) = \mathbf{tr}(HQQ^TH^T) = \mathbf{tr}\big(HQ(HQ)^T\big). \tag{3}$$

Recall, that

$$\|H\|_F = \sqrt{\mathbf{tr}(HH^T)}. \tag{4}$$

By combining equation 3 and equation 4

$$\|H\|_F = \|HQ\|_F. \tag{5}$$

Therefore

$$\|H\|_F = \|H^T\|_F = \|H^TP^T\|_F = \|PH\|_F. \tag{6}$$

(c) (10 points) Prove that $\|H\|_F$ is equal to the root sum square of H's singular values, that is,

$$\|H\|_F = \sqrt{\sum_i \sigma_i^2}$$

Solution:

$$H = U\Lambda V^T \tag{7}$$

$$\sum_i \sigma_i^2 = \mathbf{tr}\left(\Lambda\Lambda^T\right) \tag{8}$$

$$\begin{aligned}
\|H\|_F^2 = \mathbf{tr}\left(HH^T\right) &= \mathbf{tr}\left(U\Lambda V^T V\Lambda^T U^T\right) \\
&= \mathbf{tr}\left(U\Lambda\Lambda^T U^T\right) \\
&= \mathbf{tr}\left(\Lambda\Lambda^T U^T U\right) \\
&= \mathbf{tr}\left(\Lambda\Lambda^T\right) \\
&= \sum_i \sigma_i^2
\end{aligned} \tag{9}$$

2. (20 points) (a) (10 points) Work out the singular value decomposition (SVD) for the following matrix by hand. List the **key steps** of your calculation.

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Solution: We may determine singular values of A by solving

$$\mathbf{det}(A^T A - \lambda I) = 0. \tag{10}$$

The roots are $\lambda_1 = 3$, $\lambda_2 = 1$. Therefore singular values (assumed to be non-negative) are

$$S = \mathbf{diag}(\sqrt{3}, 1) \tag{11}$$

Hence A's right singular vectors v are given by

$$A^T A v = \lambda v, \tag{12}$$

which yields $v_1 = [t_1, t_1]^T$ and $v_2 = [-t_2, t_2]^T$ where $t_1, t_2 \in \mathbb{R}$ are scaling factors such that $\|v_1\|^2 = \|v_2\|^2 = 1$.

Without loss of generality, we shall assume $t_1, t_2 \geq 0$, hence

$$V = [v_1, v_2] = \begin{bmatrix} \sqrt{1/2} & -\sqrt{1/2} \\ \sqrt{1/2} & \sqrt{1/2} \end{bmatrix} \tag{13}$$

Left singular vectors can be solved with

$$U = AVS^{-1} = \begin{bmatrix} \sqrt{2/3} & 0 \\ \sqrt{1/6} & \sqrt{1/2} \\ \sqrt{1/6} & -\sqrt{1/2} \end{bmatrix} \tag{14}$$

Hence the SVD is given by

$$A = \begin{bmatrix} \sqrt{2/3} & 0 \\ \sqrt{1/6} & \sqrt{1/2} \\ \sqrt{1/6} & -\sqrt{1/2} \end{bmatrix} \mathbf{diag}(\sqrt{3}, 1) \begin{bmatrix} \sqrt{1/2} & -\sqrt{1/2} \\ \sqrt{1/2} & \sqrt{1/2} \end{bmatrix}^T. \tag{15}$$

Because $\mathbf{rank}(A) = 2$, the above SVD only reveals two singular vectors — this is sometimes referred to as **economic/compact** SVD. The **full** SVD of A can be trivially derived from equation 15, by appending zero row(s) and left singular vector(s) to S and V, respectively.

To obtain the full SVD, we first compute the third left singular vector u_3. The only requirements are that it's orthogonal to existing columns and normalised, hence

$$Uu_3 = \mathbf{0} \quad \textbf{s.t. } u_3^T u_3 = 1, \tag{16}$$

which have two solutions $u_3 = \pm[\sqrt{1/3}, -\sqrt{1/3}, -\sqrt{1/3}]^T$. The sign does not matter because the last row of below S matrix is 0. Therefore the full SVD is

$$A = \begin{bmatrix} \sqrt{2/3} & 0 & \sqrt{1/3} \\ \sqrt{1/6} & \sqrt{1/2} & -\sqrt{1/3} \\ \sqrt{1/6} & -\sqrt{1/2} & -\sqrt{1/3} \end{bmatrix} \begin{bmatrix} \mathbf{diag}(\sqrt{3}, 1) \\ \mathbf{0}^T \end{bmatrix} \begin{bmatrix} \sqrt{1/2} & -\sqrt{1/2} \\ \sqrt{1/2} & \sqrt{1/2} \end{bmatrix}^T. \quad (17)$$

Full marks for any form of the SVD (compact/efficient or full)

(b) (10 points) Consider the matrix

$$A = \begin{pmatrix} 0 & 6 & 8 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \end{pmatrix}$$

and vectors

$$v = \begin{pmatrix} 16 \\ 4 \\ 1 \end{pmatrix}$$

$$w = \begin{pmatrix} 2 \\ 2 \\ 2 \end{pmatrix}$$

Which of v, w is/are an eigenvector(s) of A and what are their eigenvalues? You should solve this without much calculation.

Solution: For a given matrix A, we can get an equation as:

$$\lambda \mathbf{x} = A\mathbf{x} \tag{18}$$

- A nonzero vector $\mathbf{x}$ is an eigenvector if there is a scalar λ such that $A\mathbf{x} = \lambda\mathbf{x}$
- The scalar value λ is called the eigenvalue.

Firstly we shall assume v is an eigenvector, then we have:

$$Av = \lambda v \tag{19}$$

when $\lambda = 2$, the equation holds, so v is an eigenvector with corresponding eigenvalue 2. Similarly, if we assume w is an eigenvector, the equation doesn't hold, thus it is not an eigenvector. Full marks also if a full first-principles derivation was performed to reach the correct answer.

3. (20 points) A symmetric real matrix A is *positive definite* (PD) if and only if $x^T Ax > 0$ for every vector $x \neq \mathbf{0}$.

Using this definition to determine whether or not a matrix is positive definite is however difficult. An alternative strategy is to use Sylvester's criterion, which states that a symmetric matrix is PD if and only if all its upper-left sub-matrices have a positive determinant.

(a) (10 points) Show whether or not the following two matrices are PD. $A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$, and $B = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{pmatrix}$.

Solution: A is not PD. B is PD. Full marks for correct answers with any reasonable derivation or justification - does not have to involve Sylvester's criterion.

(b) (10 points) Find the range of values for (real number) b such that the following real matrix is a PD matrix. List the **key steps** how you reach your answer.

$$\begin{pmatrix} 3 & -1 & b \\ -1 & 2 & -1 \\ b & -1 & 1 \end{pmatrix}$$

Solution: Sylvester's criterion: a Hermitian matrix is positive-definite if and only if all of its upper left sub-matrices have positive determinants:

- the upper left 1-by-1 corner

$$\mathbf{det}\left((3)\right) = 3 > 0 \tag{20}$$

- the upper left 2-by-2 corner

$$\mathbf{det}\left(\begin{pmatrix} 3 & -1 \\ -1 & 2 \end{pmatrix}\right) = 5 > 0 \tag{21}$$

- itself

$$\mathbf{det}\left(\begin{pmatrix} 3 & -1 & b \\ -1 & 2 & -1 \\ b & -1 & 1 \end{pmatrix}\right) = -2b^2 + 2b + 2 > 0 \Rightarrow \frac{1-\sqrt{5}}{2} < b < \frac{\sqrt{5}+1}{2} \tag{22}$$

4. (10 points) (a) (5 points) Let X and Y be real-valued random variables. Suppose X and Y are statistically independent, *i.e.* $P(X, Y) = P(X)P(Y)$.

   Show that $P(X|Y) = P(X)$, in other words, the probability of X given Y is the same as probability of X. This is equivalent to say, knowing the value of Y gives no information about the probability of X.

   Solution:

$$P(X|Y) = \frac{P(X, Y)}{P(Y)} = \frac{P(X)P(Y)}{P(Y)} = P(X) \tag{23}$$

(b) (5 points) Let A and B be real-valued random variables. Suppose $0 < P(B) < 1$, show that if $P(A|B) = P(A)$, then A and B are statistically independent.

Solution:

$$P(A|B) = \frac{P(A,B)}{P(B)} = P(A) \tag{24}$$

Multiplying centre and right equations by $P(B) > 0$ gives

$$P(A,B) = P(A)P(B).$$

5. (20 points) Tom goes to the hospital to do the yearly body checkup, and the doctor has a bad news and a good news. The bad news is that Tom tested positive for a serious disease, and the test is 99% accurate (*i.e.*, the probability of testing positive given that someone has the disease is 0.99, as is the probability of testing negative given that they don't have the disease). The good news is that, according to the latest research, this is a rare disease in Australia, striking only one in 100,000 people. What are the chances that Tom actually has the disease? (Show your calculations as well as giving the final result.)

   Solution: Denote:

   - D: if actually have disease
   - T: if test positive
   - Test is 99% accurate

     $$P(T=1|D=1) = P(T=0|D=0) = 0.99$$

   - Rare disease

     $$P(D=1) = 0.00001\ ,\ P(D=0) = 0.99999$$

   - What is the chance of disease if tested positive?

     $$P(D=1|T=1)$$

$$P(D=1|T=1) = \frac{P(D=1,T=1)}{P(T=1)} = \frac{P(D=1,T=1)}{P(D=1,T=1)+P(D=0,T=1)} \tag{25}$$

$$P(D=1,T=1) = P(T=1|D=1)P(D=1) = 0.0000099 \tag{26}$$

$$P(D=0,T=1) = \Big(1 - P(T=0|D=0)\Big)P(D=0) = 0.0099999 \tag{27}$$

$$P(D=1|T=1) = \frac{99}{99+99999} \approx 0.000989 \tag{28}$$

6. (10 points) (a) (3 points) Given $z = x(y+4)$ subject to $x + y = 8$. Use the Lagrange multiplier method to find the local optimum value of z, as well as the optimal x and y. Is the above solution a minimum or a maximum?

Solution: By using Lagrange multiplier method, we have:

$$\begin{aligned} L(x,y,\lambda) &= x(y+4) + \lambda(x+y-8) \\ &= xy + 4x + x\lambda + y\lambda - 8\lambda \end{aligned} \tag{29}$$

$$\begin{aligned} \frac{\partial L(x,y,z,\lambda)}{\partial x} &= y + 4 + \lambda = 0 \\ \frac{\partial L(x,y,z,\lambda)}{\partial y} &= x + \lambda = 0 \\ \frac{\partial L(x,y,z,\lambda)}{\partial z} &= x + y - 8 = 0 \end{aligned} \tag{30}$$

$x = 6, y = 2, z = 36, \text{and } \lambda = -6.$

The stationary point is $(x,y) = (6,2)$, with a local optimal value as $z = 36$.

$$\begin{aligned} z &= x(12-x) \\ &= -x^2 + 12x \end{aligned} \tag{31}$$

- quadratic function graph (high school knowledge)
- $\frac{dz}{dx} = -2x + 12$, $\frac{d^2z}{dx^2} = -2 < 0$

It implies z is a maximum.

(b) (7 points) Use the Lagrange Multiplier method to prove that a triangle with maximum area that has a given perimeter of 2 must be equilateral. Can the solution be found using the Lagrange method ?

Solution:

Yes, it can. Assuming a, b, c are three edges. According to Heron's formula, we have:

$$\varphi(a,b,c) : 2p = a + b + c = 2 \tag{32}$$

$$S(a,b,c) = \sqrt{p(p-a)(p-b)(p-c)} = \sqrt{(1-a)(1-b)(1-c)} \tag{33}$$

Then according to Lagrange multiplier method, we have:

$$\begin{aligned} F(a,b,c,\lambda) &= S(a,b,c) + \lambda\varphi(a,b,c) \\ \frac{\partial F(a,b,c,\lambda)}{\partial a} &= \frac{1}{2S(a,b,c)}(-bc+b+c-1) + \lambda = 0 \\ \frac{\partial F(a,b,c,\lambda)}{\partial b} &= \frac{1}{2S(a,b,c)}(-ac+a+c-1) + \lambda = 0 \\ \frac{\partial F(a,b,c,\lambda)}{\partial c} &= \frac{1}{2S(a,b,c)}(-ab+a+b-1) + \lambda = 0 \\ \frac{\partial F(a,b,c,\lambda)}{\partial \lambda} &= a+b+c-2p = 0 \end{aligned} \tag{34}$$

Then according to Lagrange multiplier method, we have:

$$\begin{aligned} -bc+b+c-1 &= -2\lambda S(a,b,c) \\ -ac+a+c-1 &= -2\lambda S(a,b,c) \\ -ab+a+b-1 &= -2\lambda S(a,b,c) \\ a+b+c &= 2 \\ a<1; \quad b<1; &\quad c<1 \end{aligned} \tag{35}$$

Then you can get $a = b = c$.

END.