

## Answer Key and Marking Scheme

**Q1** Alternative approaches would also be accepted and awarded marks

a i) Given  $f(x) = 3x^4 - 20x^3 + 36x^2 + 10$ . So, we have

$$f'(x) = 12x^3 - 60x^2 + 72x = 0$$

$$\Rightarrow 12x(x-2)(x-3) = 0$$

$$\Rightarrow x = 0, 2, 3$$

(1 mark)

$$f''(x) = 36x^2 - 120x + 72$$

$$\Rightarrow f''(0) = 72 > 0 \Rightarrow \text{Pt of minima}$$

$$\Rightarrow f''(2) = -24 < 0 \Rightarrow \text{Pt of maxima}$$

$$\Rightarrow f''(3) = 36 > 0 \Rightarrow \text{Pt of minima}$$

(1.5 marks)

Now  $f(0) = 10 < f(3) = 37$ . Therefore 0 is a pt of global minima. (0.5 marks)

a ii) Clearly 0.5 is closer to 0 ( global minima) and 3.5 is closer to 3 (local minima). So, 0 is a better initial condition. (1 mark)

b Clearly mean is 0 along both dimension. So, the covariance matrix is given by

$$\mathbf{S} = \frac{1}{4} \mathbf{X} \mathbf{X}' = \begin{pmatrix} 2.5 & -5 \\ -5 & 10 \end{pmatrix}$$

$$|\mathbf{S} - \lambda \mathbf{I}| = 0$$

$$\Rightarrow \lambda^2 - 12.5\lambda = 0$$

$$\Rightarrow \lambda = 0, 12.5$$

(0.5 marks)

To find eigenvector corresponding to largest eigenvalue 12.5 consider  $[\mathbf{S} - 12.5\mathbf{I}]\mathbf{x} = \mathbf{0}$

$$\begin{pmatrix} 2.5 - 12.5 & -5 \\ -5 & 10 - 12.5 \end{pmatrix} \mathbf{x} = \mathbf{0}$$

$$\Rightarrow \begin{pmatrix} -10 & -5 \\ -5 & -2.5 \end{pmatrix} \mathbf{x} = \mathbf{0}$$

$$\Rightarrow \begin{pmatrix} -10 & -5 \\ 0 & 0 \end{pmatrix} \mathbf{x} = \mathbf{0} \text{ by R2-0.5 R1}$$

$$\Rightarrow -10x_1 - 5x_2 = 0$$

$$\Rightarrow x_1 = -0.5x_2$$

Therefore  $\begin{pmatrix} -t/2 \\ t \end{pmatrix} \forall t \neq 0$  is an eigenvector corresponding to largest eigenvalue of  $\mathbf{S}$  and hence it is the first principal component and therefore

gives the direction of maximum variance. Thus, the claim of student 3 is correct. (1.5 marks)

- c i) Distance from origin to a point  $(x, y)$  is  $(x^2 + y^2)^{\frac{1}{2}}$ . Therefore the problem is

$$\min(x^2 + y^2)^{\frac{1}{2}}$$

subject to constraints

$$x^2 + y^2 - 5 \leq 0,$$

$$x + 2y - 4 = 0,$$

$$-x \leq 0, -y \leq 0.$$

(1 mark)

- c ii) The Lagrangian function is given by

$$L(x, y, \alpha_1, \alpha_2, \alpha_3, \gamma) = (x^2 + y^2)^{\frac{1}{2}} + \alpha_1(x^2 + y^2 - 5) - \alpha_2x - \alpha_3y + \gamma(x + 2y - 4)$$

(1 mark)

- c iii) Clearly feasibility conditions

$$\begin{aligned} (\frac{4}{5})^2 + (\frac{8}{5})^2 - 5 &= 3.2 \leq 0, \\ (\frac{4}{5}) + 2(\frac{8}{5}) - 4 &= 0, \\ -(\frac{4}{5}) &\leq 0, -(\frac{8}{5}) \leq 0. \end{aligned}$$

are satisfied.(0.5 marks)

$$\text{Now, } \alpha_1((\frac{4}{5})^2 + (\frac{8}{5})^2 - 5) = 0 \Rightarrow \alpha_1 = 0$$

$$\text{Similarly } -\alpha_2(\frac{4}{5}) = 0, -\alpha_3(\frac{8}{5}) = 0 \Rightarrow \alpha_2 = \alpha_3 = 0. \quad (0.5 \text{ marks})$$

Similarly

$$\begin{aligned} \frac{\partial L}{\partial x} &= \frac{x}{(x^2 + y^2)^{\frac{1}{2}}} + \gamma = 0 \\ \Rightarrow & \frac{(\frac{4}{5})}{((\frac{4}{5})^2 + (\frac{8}{5})^2)^{\frac{1}{2}}} + \gamma = 0 \\ \Rightarrow & \gamma = \frac{-1}{\sqrt{5}} \\ \frac{\partial L}{\partial y} &= \frac{y}{(x^2 + y^2)^{\frac{1}{2}}} + 2\gamma = 0 \\ \Rightarrow & \frac{(\frac{8}{5})}{((\frac{4}{5})^2 + (\frac{8}{5})^2)^{\frac{1}{2}}} + 2\gamma = 0 \\ \Rightarrow & \gamma = \frac{-1}{\sqrt{5}} \end{aligned}$$

Therefore  $(\frac{4}{5}, \frac{8}{5})$  satisfies all KKT conditions and  $(\alpha_1, \alpha_2, \alpha_3, \gamma) = (0, 0, 0, \frac{-1}{\sqrt{5}})$  ( 1 mark)

**Q2** Alternative approaches would also be accepted and awarded marks

- i) The rank is 2 in either case. The second column can be obtained from columns 1 and 3 for example. 1 mark for each answer.
- ii)  $\mathbb{C}$  over  $\mathbb{R}$  is two dimensional whereas  $\mathbb{C}$  over  $\mathbb{C}$  is one dimensional. One mark for each. Other examples could also be accepted.
- iii) Since  $v_1 + v_2 + v_3 = 0$ ,  $v_1 = -v_2 - v_3$ . Span of  $\{v_1, v_2\} = \{\alpha_1 v_1 + \alpha_2 v_2 | \alpha_1, \alpha_2 \in F\}$  and this could be written in terms of  $v_2$  and  $v_3$  and hence the result follows. 1 mark for the first observation and 1 mark for rewriting the linear combination.
- iv) C is wrong as RREF is unique and hence D is correct. (1 mark) E is wrong as one cannot get X from its RREF in a unique way. (1 mark)
- v) Take any invertible matrix P of size n x n and  $B = P\Lambda P^{-1}$  would have the same eigenvalues  $\lambda_1, \lambda_2, \dots, \lambda_n$ , where  $\Lambda$  is the diagonal matrix with  $\lambda_1, \lambda_2, \dots, \lambda_n$  as diagonal values. (1 mark)  
The eigenvectors of  $B$  would be linearly independent and one can use Gram Schmidt orthogonalization process to generate the required orthogonal matrix. (1 mark)

**Q3** Alternative approaches would also be accepted and awarded marks

- 3.1 i) We need to prove 3 properties of distance metric as studied in class.

Property 1:

$d(x, y) \geq 0$  as  $d(x, y)$  involves taking sum of non-negative numbers due to the use of  $|\cdot|$ ,  $\sum_{i=1}^n |x_i - y_i|$  can be zero only when each component of form  $x_i - y_i = 0$ . This only happens when  $x = y$ . Hence  $d(x, y) = 0$  only for  $x = y$ . (0.5 marks)

Property 2:

Due to the fact that  $|x_i - y_i| = |y_i - x_i|$ , it can be safely concluded that  $d(x, y) = d(y, x)$ . (0.5 marks)

Property 3:

Note  $d(x, y) = \sum_{i=1}^n |x_i - y_i| = \sum_{i=1}^n |x_i - z_i + z_i - y_i|$   
But

$$\sum_{i=1}^n |x_i - z_i + z_i - y_i| \leq \sum_{i=1}^n |x_i - z_i| + \sum_{i=1}^n |z_i - y_i|$$

Above we have used triangle inequality of  $\ell_1$  norm.

$$\sum_{i=1}^n |x_i - z_i| + \sum_{i=1}^n |z_i - y_i| = d(x, z) + d(z, y)$$

Hence

$$d(x, y) \leq d(x, z) + d(z, y) \quad (1 \text{ marks})$$

In summary  $d(x,y)$  is a distance metric.

ii) Let  $z_1 \in \mathcal{C}$  which means  $Az_1 = b$ .

Let  $z_2 \in \mathcal{C}$  which means  $Az_2 = b$ .

We need to consider a general  $z_3 = \lambda z_1 + (1 - \lambda)z_2$  where  $\lambda \in [0, 1]$ .

$$Az_3 = A(\lambda z_1 + (1 - \lambda)z_2) \quad (0.25 \text{ marks})$$

This means that

$$Az_3 = (\lambda)Az_1 + (1 - \lambda)Az_2 = \lambda.b + (1 - \lambda).b = b \quad (0.5 \text{ marks})$$

In above step we used the initial assumptions regarding  $z_1$  and  $z_2$ . Hence we can conclude that  $Az_3 = b$  or equivalently  $z_3 \in \mathcal{C}$  for a general  $z_3$  as defined before. Hence  $\mathcal{C}$  is a convex set. (0.25 marks)

3.2 It may be observed that here  $S = XX^T$  (similar to covariance matrix in PCA), where  $X$  is a matrix which has  $x_1, x_2, \dots, x_{20}$  as columns. Since the provided  $eig(\cdot)$  can only deal with matrices of maximum size  $30 \times 30$ , we cannot use it to directly find eigenvalues and eigenvectors of  $S = XX^T$ . Note: Here  $X^T X \in \mathbb{R}^{20 \times 20}$ . So its possible to use given function  $eig(\cdot)$  to find eigenvalues and eigenvectors of  $X^T X$  as it can handle matrices of size upto  $30 \times 30$ . (0.5 marks)

a) Recall from PCA discussion that

$$\begin{aligned} XX^T b &= \lambda b \\ X^T XX^T b &= \lambda X^T b \\ X^T X c &= \lambda c \text{ where } X^T b = c \\ XX^T X c &= \lambda X c \\ XX^T d &= \lambda d \text{ where } X c = d \end{aligned}$$

Hence, if  $(c, \lambda)$  is an eigenpair of  $X^T X$ , then  $Xc$  is eigenvector and  $\lambda$  eigenvalue for  $XX^T$ . (0.5 marks)

Based on above discussion, we can use  $X^T X$  matrix to find eigenvalues of  $XX^T$ . This means that if  $\lambda$  is eigenvalue of  $XX^T$  then its also an eigenvalue of  $X^T X$ . Other eigenvalues of  $XX^T$  are zero. (1 marks)

b) Similarly, from previous derivation in (a) to find the eigenvectors of  $XX^T$  we have to first find the eigenvectors of  $X^T X$ . Let it be  $c$ , then eigenvector of  $XX^T$  can be obtained as  $Xc$  where eigenvalue is given by  $\lambda$ . (1 marks)

3.3 (a)

$$g(z) = \frac{1}{2} \|Az - b\|_2^2 + \frac{1}{2} \|z\|_2^2 + \frac{1}{2} \|b\|_2^2$$

It can also be written as  $g(z) = \frac{1}{2} z^T A^T A z - z^T A^T b + \frac{1}{2} b^T b + \frac{1}{2} z^T z + b^T b$

Its gradient is  $\nabla g(z) = -A^T A z + A^T b - z$  (2 marks)

Hence  $d_k = -\nabla g(z) = A^T A z + z - A^T b$

Now, we need following:  $\alpha = \underset{\alpha}{\operatorname{argmin}} g(z_k + \alpha d_k)$

$$g(z_k + \alpha d_k) = \frac{1}{2} (z_k + \alpha d_k)^T A^T A (z_k + \alpha d_k) - (z_k + \alpha d_k)^T A^T b + \frac{1}{2} b^T b + \frac{1}{2} (z_k + \alpha d_k)^T (z_k + \alpha d_k) + b^T b$$

Now taking derivative with respect to  $\alpha$  we get

$$\frac{\partial g(z_k + \alpha d_k)}{\partial \alpha} = z_k^T A^T A d_k + \alpha d_k^T A^T A d_k + \alpha d_k^T d_k - d_k^T A^T b + z_k^T d_k = 0 \quad (1 \text{ marks})$$

Hence

$$\alpha = \frac{d_k^T A^T b - z_k^T A^T A d_k - z_k^T d_k}{d_k^T A^T A d_k + d_k^T d_k}$$

where  $d_k = A^T A z + z - A^T b$  (1 marks)

**Q4** Alternative approaches would also be accepted and awarded marks

- (1) Using gradient descent we can write  $x_{n+1} = x_n - \lambda \frac{\partial f}{\partial x}$  or  $x_{n+1} = x_n - \lambda(2ax_n + b)$ , or  $x_{n+1} = x_n(1 - 2a\lambda) - \lambda b$ . Then we can write  $x_{n+2} = x_n(1 - 2a\lambda)^2 - \lambda b(1 - 2a\lambda) - \lambda b$ . Continuing on this way we can write  $x_{n+k} = x_n(1 - 2a\lambda)^k - \lambda b(1 - 2a\lambda)^{k-1} - \lambda b(1 - 2a\lambda)^{k-2} - \dots - \lambda b$ . This is in the form  $x_{n+k} = x_n P^k + Q$  as required. Now if we let  $0 < |1 - 2a\lambda| < 1$  or  $-1 < (1 - 2a\lambda) < 1$  and let  $k \rightarrow \infty$  we see that  $P^k \rightarrow 0$  and we are left with  $x_\infty = -\lambda b - \lambda b(1 - 2a\lambda) - \lambda b(1 - 2a\lambda)^2 - \dots$ . This is an infinite series which converges if  $-1 < (1 - 2a\lambda) < 1$  and sums up to  $\frac{-\lambda b}{1 - (1 - 2a\lambda)} = \frac{-b}{2a}$  which is the local (and global) minimum of the given quadratic. Thus  $0 < \lambda < \frac{1}{a}$  is the condition needed to ensure convergence of gradient descent.

Marking Scheme: 2 Marks  $\rightarrow$  setting up the expression for  $x_{n+k}$  in the form  $x_n P^k + Q$ . 3 Marks  $\rightarrow$  recognizing that the infinite series converges for  $\lambda$  in a certain range and completing the argument.

- (2) The given kernel function can be written as  $(\mathbf{x}^T \mathbf{z})^2 + 3(\mathbf{x}^T \mathbf{z} + 2)^2 = 4(\mathbf{x}^T \mathbf{z})^2 + 12\mathbf{x}^T \mathbf{z} + 12$ . This can be seen as an inner product  $\phi$  of the following form  $\phi = [\phi_1(\mathbf{x}), \phi_2(\mathbf{x}), \phi_3(\mathbf{x})]^T$  where  $\phi_1$  is a  $n^2 \times 1$  mapping representing the term  $4(\mathbf{x}^T \mathbf{z})^2$  and can be derived as follows:  $4(\mathbf{x}^T \mathbf{z})^2 = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} 4x_i z_i x_j z_j = \sum_{i=1}^{i=n} \sum_{j=1}^{j=n} 2x_i x_j * 2z_i z_j$ . This leads us to the mapping  $\phi_1(\mathbf{x}) = [2x_1 x_1, 2x_1 x_2, \dots, 2x_1 x_n, 2x_2 x_1, 2x_2 x_2, \dots, 2x_2 x_n, \dots, 2x_n x_1, 2x_n x_2, \dots, 2x_n x_n]^T$  which is a  $n^2 \times 1$  mapping.  $\phi_2(\mathbf{x})$  is the

$n \times 1$  mapping  $[\sqrt{12}x_1, \sqrt{12}x_2, \dots, \sqrt{12}x_n]$  representing the term  $12\mathbf{x}^T \mathbf{z}$  and  $\phi_3(\mathbf{x}) = \sqrt{12}$  representing the constant term in the kernel function. Now  $\phi^T(\mathbf{x})\phi(\mathbf{x})$  can be seen to equal the given Kernel function and  $\phi$  is of dimension  $n^2 + n + 1$ .

Marking Scheme: 2 Marks  $\rightarrow$  simplifying the given expression only in terms of  $4(\mathbf{x}^T \mathbf{z})^2 + 12\mathbf{x}^T \mathbf{z} + 12$ . 3 Marks  $\rightarrow$  splitting this sum into different mappings and stitching them together to find the final mapping.