Further Mathematical Methods (Linear Algebra)

Solutions For The 1999 Examination

Question 1

(a) The *orthogonal complement* of a set of vectors S, denoted by $S^\perp$, is the set of all vectors that are orthogonal to the vectors in S, i.e.

$$S^\perp = \{\mathbf{v} : \langle \mathbf{v}, \mathbf{s} \rangle = 0 \ \ \forall \mathbf{s} \in S\}.$$

To prove that $N(\mathsf{A}^t) = R(\mathsf{A})^\perp$ for any real $m \times n$ matrix A, we use a 'double inclusion proof,' that is,

- Suppose that $\mathbf{z} \in R(\mathsf{A})^\perp$, i.e. $\langle \mathbf{z}, \mathbf{y} \rangle = 0 \ \forall \mathbf{y} \in R(\mathsf{A})$, or using our convention [which allows us to express inner products in terms of matrix products], $\mathbf{z}^t\mathbf{y} = \mathbf{0} \ \forall \mathbf{y} \in R(\mathsf{A})$. Now, since $\mathsf{A}\mathbf{x} \in R(\mathsf{A}) \ \forall \mathbf{x} \in \mathbb{R}^n$, we have

  $$\mathbf{z}^t\mathsf{A}\mathbf{x} = \mathbf{0} \implies (\mathsf{A}^t\mathbf{z})^t\mathbf{x} = \mathbf{0} \implies \langle \mathsf{A}^t\mathbf{z}, \mathbf{x} \rangle = 0.$$

  But, this is true $\forall \mathbf{x} \in \mathbb{R}^n$, and so we have $\mathsf{A}^t\mathbf{z} = \mathbf{0}$. Thus, $\mathbf{z} \in N(\mathsf{A}^t)$ and $R(\mathsf{A}^t) \subseteq N(\mathsf{A}^t)$.

- Suppose that $\mathbf{z} \in N(\mathsf{A}^t)$, i.e. $\mathsf{A}^t\mathbf{z} = \mathbf{0}$. Now, for any $\mathbf{y} \in R(\mathsf{A})$, $\exists \mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathsf{A}\mathbf{x}$, thus

  $$\mathsf{A}^t\mathbf{z} = \mathbf{0} \implies \mathbf{x}^t\mathsf{A}^t\mathbf{z} = \mathbf{0} \implies (\mathsf{A}\mathbf{x})^t\mathbf{z} = \mathbf{0} \implies \langle \mathbf{y}, \mathbf{z} \rangle = 0,$$

  where, once again, we have used our convention. So, $\mathbf{z}$ is orthogonal to any $\mathbf{y} \in R(\mathsf{A})$, i.e. $\mathbf{z} \in R(\mathsf{A})^\perp$. Thus, $N(\mathsf{A}^t) \subseteq R(\mathsf{A})^\perp$.

Consequently, $N(\mathsf{A}^t) = R(\mathsf{A})^\perp$ (as required).

(b) We are asked to find an orthonormal basis for the subspace S of $\mathbb{R}^4$ spanned by the vectors $\mathbf{v}_1 = [1, 0, 2, 0]^t$, $\mathbf{v}_2 = [0, 2, 1, 0]^t$ and $\mathbf{v}_3 = [0, 1, 1, 0]^t$. To do this, we use the Gram-Schmidt procedure:

- Taking $\mathbf{v}_1 = [1, 0, 2, 0]^t$, we note that $\|\mathbf{v}_1\|^2 = 1 + 0 + 4 + 0 = 5$ and set $\mathbf{e}_1 = \frac{1}{\sqrt{5}}[1, 0, 2, 0]^t$.

- Taking $\mathbf{v}_2 = [0, 2, 1, 0]^t$, we need to construct a vector $\mathbf{u}_2$ where

  $$\mathbf{u}_2 = \mathbf{v}_2 - \langle \mathbf{v}_2, \mathbf{e}_1 \rangle \mathbf{e}_1,$$

  which as

  $$\langle \mathbf{v}_2, \mathbf{e}_1 \rangle = \frac{0 + 0 + 2 + 0}{\sqrt{5}} = \frac{2}{\sqrt{5}},$$

  gives us

  $$\mathbf{u}_2 = \begin{bmatrix} 0 \\ 2 \\ 1 \\ 0 \end{bmatrix} - \frac{2}{5}\begin{bmatrix} 1 \\ 0 \\ 2 \\ 0 \end{bmatrix} = \frac{1}{5}\begin{bmatrix} -2 \\ 10 \\ 1 \\ 0 \end{bmatrix}.$$

  But, we are only interested in the direction of $\mathbf{u}_2$, and so we set $\mathbf{u}_2 = [-2, 10, 1, 0]^t$. Thus, noting that $\|\mathbf{u}_2\|^2 = 4 + 100 + 1 + 0 = 105$ we set $\mathbf{e}_2 = \frac{1}{\sqrt{105}}[-2, 10, 1, 0]^t$.

- Taking $\mathbf{v}_3 = [0, 1, 1, 0]^t$, we need to construct a vector $\mathbf{u}_3$ where

  $$\mathbf{u}_3 = \mathbf{v}_3 - \langle \mathbf{v}_3, \mathbf{e}_1 \rangle \mathbf{e}_1 - \langle \mathbf{v}_3, \mathbf{e}_2 \rangle \mathbf{e}_2,$$

  which as

  $$\langle \mathbf{v}_3, \mathbf{e}_1 \rangle = \frac{0 + 0 + 2 + 0}{\sqrt{5}} = \frac{2}{\sqrt{5}},$$

  and

  $$\langle \mathbf{v}_3, \mathbf{e}_2 \rangle = \frac{0 + 10 + 1 + 0}{\sqrt{105}} = \frac{11}{\sqrt{105}},$$

gives us

$$\mathbf{u}_3 = \begin{bmatrix}0\\1\\1\\0\end{bmatrix} - \frac{2}{5}\begin{bmatrix}1\\0\\2\\0\end{bmatrix} - \frac{11}{105}\begin{bmatrix}-2\\10\\1\\0\end{bmatrix} = \frac{1}{21}\begin{bmatrix}-4\\-1\\2\\0\end{bmatrix}.$$

But, we are only interested in the direction of $\mathbf{u}_3$, and so we set $\mathbf{u}_3 = [-4, -1, 2, 0]^t$. Thus, noting that $\|\mathbf{u}_3\|^2 = 16 + 1 + 4 + 0 = 21$ we set $\mathbf{e}_3 = \frac{1}{\sqrt{21}}[-4, -1, 2, 0]^t$.

Thus, *an* orthonormal basis for S is

$$\left\{ \frac{1}{\sqrt{5}}\begin{bmatrix}1\\0\\2\\0\end{bmatrix}, \frac{1}{\sqrt{105}}\begin{bmatrix}-2\\10\\1\\0\end{bmatrix}, \frac{1}{\sqrt{21}}\begin{bmatrix}-4\\-1\\2\\0\end{bmatrix} \right\}.$$

We are then asked to find the orthogonal complement of S. It should be obvious that the vector $[0, 0, 0, 1]^t$ is orthogonal to each of the vectors in the set $\{\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3\}$ (or indeed, $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$), and so this vector lies in $S^\perp$. Thus, as there can be no more than four orthogonal vectors in any subset of $\mathbb{R}^4$, we have

$$S^\perp = \text{Lin}\left\{ \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \right\}.$$

(c) The *direct sum* of two subspaces X and Y of a vector space, denoted by $X \oplus Y$, is the sum of X and Y, i.e.

$$X + Y = \{\mathbf{x} + \mathbf{y} : \mathbf{x} \in X \text{ and } \mathbf{y} \in Y\},$$

where $\forall \mathbf{z} \in X + Y$, $\mathbf{z}$ can be written in the form $\mathbf{x} + \mathbf{y}$ (where $\mathbf{x} \in X$ and $\mathbf{y} \in Y$) in just one way.

To show that: If S is a subspace of $\mathbb{R}^n$, then $S \oplus S^\perp = \mathbb{R}^n$, we take $\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ to be an orthonormal basis of S, and extend this to an orthonormal basis of $\mathbb{R}^n$, i.e. $\mathbb{R}^n = \text{Lin}\{\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$. Now, we know that for any set $S \subseteq \mathbb{R}^n$, $S^\perp$ is a subspace of $\mathbb{R}^n$ and it should be clear that

$$S^\perp = \text{Lin}\{\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$$

Thus, $\mathbb{R}^n = S \oplus S^\perp$ since $\mathbb{R}^n = S + S^\perp$ and any $\mathbf{x} \in \mathbb{R}^n$ can be written in the form $\mathbf{x} = \mathbf{y} + \mathbf{z}$ (where $\mathbf{y} \in S = \text{Lin}\{\mathbf{e}_1, \ldots, \mathbf{e}_k\}$ and $\mathbf{z} \in S^\perp = \text{Lin}\{\mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$) in just one way as $\{\mathbf{e}_1, \ldots, \mathbf{e}_k, \mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$ is a basis of $\mathbb{R}^n$.

In (b), we had

$$S = \text{Lin}\left\{ \frac{1}{\sqrt{5}}\begin{bmatrix}1\\0\\2\\0\end{bmatrix}, \frac{1}{\sqrt{105}}\begin{bmatrix}-2\\10\\1\\0\end{bmatrix}, \frac{1}{\sqrt{21}}\begin{bmatrix}-4\\-1\\2\\0\end{bmatrix} \right\} \text{ and } S^\perp = \text{Lin}\left\{ \begin{bmatrix}0\\0\\0\\1\end{bmatrix} \right\},$$

and as the orthonormal basis for S and the basis for $S^\perp$ form an orthonormal basis for $\mathbb{R}^4$, it is clear that $S \oplus S^\perp = \mathbb{R}^4$.

(d) To prove that the null-space of a real $m \times n$ matrix A is a subspace of $\mathbb{R}^n$, we note that the null-space is defined to be the set of vectors

$$N(\mathsf{A}) = \{\mathbf{x} \in \mathbb{R}^n : \mathsf{A}\mathbf{x} = \mathbf{0}\},$$

and so it is clearly a sub*set* of $\mathbb{R}^n$. To show that it is a sub*space*, we need to establish that it is also closed under vector addition and scalar multiplication:

- For *any* two vectors $\mathbf{x}, \mathbf{y} \in N(\mathsf{A})$, i.e. $\mathbf{x}$ and $\mathbf{y}$ are such that $\mathsf{A}\mathbf{x} = \mathbf{0}$ and $\mathsf{A}\mathbf{y} = \mathbf{0}$, we ask: Is $\mathbf{x} + \mathbf{y} \in N(\mathsf{A})$? The answer is yes since:

$$\mathsf{A}(\mathbf{x} + \mathbf{y}) = \mathsf{A}\mathbf{x} + \mathsf{A}\mathbf{y} = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

  as $\mathbf{x}, \mathbf{y} \in N(\mathsf{A})$. Thus, $N(\mathsf{A})$ is closed under vector addition.

- For *any* vector $\mathbf{x} \in N(\mathsf{A})$, i.e. $\mathbf{x}$ is such that $\mathsf{A}\mathbf{x} = \mathbf{0}$, and *any* scalar $\alpha \in \mathbb{R}$, we ask: Is $\alpha\mathbf{x} \in N(\mathsf{A})$? The answer is yes since:

$$\mathsf{A}(\alpha\mathbf{x}) = \alpha(\mathsf{A}\mathbf{x}) = \alpha\mathbf{0} = \mathbf{0},$$

  as $\mathbf{x} \in N(\mathsf{A})$. Thus, $N(\mathsf{A})$ is closed under scalar multiplication.

Thus, $N(\mathsf{A})$ is a subspace of $\mathbb{R}^n$ (as required).

We are given the matrix

$$\mathsf{A} = \begin{bmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 2 & 0 \\ 0 & 2 & 1 & 0 \end{bmatrix},$$

and we want to find its null-space. To do this, we note that the range of A^t, i.e. the row-space of the matrix A, is just the linear span of the vectors considered in the first part of (b) and so, $R(\mathsf{A}^t) = S$. Then, using (a) we can see that $N(\mathsf{A}) = N(\mathsf{A}^{tt}) = R(\mathsf{A}^t)^\perp$ and so, $N(\mathsf{A}) = S^\perp$. Consequently, from the last part of (b), we have

$$N(\mathsf{A}) = S^\perp = \text{Lin}\left\{ \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

Question 2.

(a) An eigenvalue λ_1 is called *dominant* if for any other eigenvalue λ, $|\lambda| < \lambda_1$.

(b) The Leslie matrix for this species is:

$$\mathsf{L} = \begin{bmatrix} 0 & 7 & 6 \\ \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{3} & 0 \end{bmatrix}.$$

To find the eigenvalues of this matrix, we have to solve the determinant equation:

$$|\mathsf{L} - \lambda \mathsf{I}| = 0 \implies \begin{vmatrix} -\lambda & 7 & 6 \\ \frac{1}{2} & -\lambda & 0 \\ 0 & \frac{1}{3} & -\lambda \end{vmatrix} = 0 \implies \lambda^3 - \tfrac{7}{2}\lambda - 1 = 0.$$

Using trial-and-error we find that $\lambda_1 = 2$ is a solution, and hence an eigenvalue of L. To find the others, we divide the polynomial $\lambda^3 - \frac{7}{2}\lambda - 1$ by $\lambda - 2$ to get the quadratic

$$\lambda^2 + 2\lambda + \frac{1}{2} = 0,$$

and solving this, we find that the remaining two eigenvalues are $\lambda_{2,3} = -1 \pm \frac{1}{\sqrt{2}}$. (Notice that both of these are negative!) To establish that the unique positive real eigenvalue, i.e. $\lambda = 2$, is dominant we note that

$$|\lambda_2| = |-1 + \tfrac{1}{\sqrt{2}}| = |-0.29| = 0.29 < 2 = \lambda_1,$$

and,

$$|\lambda_3| = |-1 - \tfrac{1}{\sqrt{2}}| = |-1.71| = 1.71 < 2 = \lambda_1,$$

(as required).

To find an eigenvector, $\mathbf{v}_1$, corresponding to $\lambda = 2$, you can either recall that

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ b_1/\lambda_1 \\ b_1 b_2/\lambda_1^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1/4 \\ 1/24 \end{bmatrix},$$

from the lectures, or you can find an eigenvector in the normal way by solving the matrix equation $(\mathsf{L} - 2\mathsf{I})\mathbf{v}_1 = \mathbf{0}$.

As the unique real positive eigenvalue is dominant, the long-term behaviour of this population is described by two relations:

- $\mathbf{x}^{(k)} \simeq c2^k\mathbf{v}_1$ and so the proportion of the population in each age class is, in the long run, constant and given by the ratio $1 : \frac{1}{4} : \frac{1}{24}$.
- $\mathbf{x}^{(k)} \simeq 2\mathbf{x}^{(k-1)}$ and so the population in each age class grows by a factor of 2 (i.e. increases by 100%) every time period, which in this case is every 10 years.

(c) The Leslie matrix for this species is:

$$\mathsf{L} = \begin{bmatrix} 0 & 0 & 2 \\ \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \end{bmatrix}.$$

To find the eigenvalues of this matrix, we have to solve the determinant equation:

$$|\mathsf{L} - \lambda \mathsf{I}| = 0 \implies \begin{vmatrix} -\lambda & 0 & 2 \\ \frac{1}{6} & -\lambda & 0 \\ 0 & \frac{1}{9} & -\lambda \end{vmatrix} = 0 \implies \lambda^3 - \tfrac{1}{27} = 0,$$

and so we find that the solutions are $\lambda = \frac{1}{3}e^{2\pi n i/3}$ where we only need to consider the cases where $n = 1, 2, 3$. Thus, the eigenvalues are:

- $\lambda_1 = \frac{1}{3}$ if $n = 0$ (this is the unique positive real eigenvalue),
- $\lambda_2 = \frac{1}{3}\left(\cos\frac{2\pi}{3} + i\sin\frac{2\pi}{3}\right) = \frac{1}{6}(-1 + i\sqrt{3})$ if $n = 1$, and
- $\lambda_3 = \frac{1}{3}\left(\cos\frac{4\pi}{3} + i\sin\frac{4\pi}{3}\right) = \frac{1}{6}(-1 - i\sqrt{3})$ if $n = 2$.

To establish that the unique dominant eigenvalue (i.e. λ_1) is *not* dominant, we note that $|\lambda_{2,3}| \not< \lambda_1$ as

$$|\lambda_{2,3}| = \left|\tfrac{1}{6}(-1 \pm i\sqrt{3})\right| = \tfrac{1}{6}\left|-1 \pm i\sqrt{3}\right| = \tfrac{1}{6}\sqrt{1+3} = \tfrac{1}{3} = \lambda_1,$$

(as required).

To calculate L^3, we note that

$$\mathsf{L}^2 = \begin{bmatrix} 0 & 0 & 2 \\ \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 2 \\ \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \end{bmatrix} = \begin{bmatrix} 0 & \frac{2}{9} & 0 \\ 0 & 0 & \frac{1}{3} \\ \frac{1}{54} & 0 & 0 \end{bmatrix},$$

and so,

$$\mathsf{L}^3 = \begin{bmatrix} 0 & \frac{2}{9} & 0 \\ 0 & 0 & \frac{1}{3} \\ \frac{1}{54} & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 2 \\ \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{9} & 0 \end{bmatrix} = \frac{1}{27}\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

i.e. $\mathsf{L}^3 = \frac{1}{27}\mathsf{I}$. Now, noting that the population of this species is described by the equation

$$\mathbf{x}^{(k)} = \mathsf{L}\mathbf{x}^{(k-1)},$$

where $\mathbf{x}^{(k)}$ is the population distribution vector for the species in the kth time period and that a time interval of sixty years is the same as three time periods (as each individual time period lasts for twenty years), the relevant population distribution vectors for this question are:

$$\mathbf{x}^{(3)} = \mathsf{L}^3\mathbf{x}^{(0)} = \tfrac{1}{27}\mathbf{x}^{(0)}, \;\; \mathbf{x}^{(6)} = \tfrac{1}{27}\mathsf{L}^3\mathbf{x}^{(3)} = \tfrac{1}{27^2}\mathbf{x}^{(0)}, \;\; \mathbf{x}^{(9)} = \tfrac{1}{27^2}\mathsf{L}^3\mathbf{x}^{(6)} = \tfrac{1}{27^3}\mathbf{x}^{(0)}, \ldots$$

Thus, at the end of the jth sixty year interval (i.e. the $3j$th time period), the population distribution vector will be given by

$$\mathbf{x}^{(3j)} = \frac{\mathbf{x}^{(0)}}{27^j},$$

and so we can see that:

- The population in each age class decreases by a factor of $\frac{1}{27}$ every sixty years.
- Every sixty years the proportion of the population in each age class is the same as it was 'at the beginning' (i.e. as it was when $k = 0$).

Question 3

(a) To show that the $m \times n$ matrix

$$\mathsf{P} = \mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t,$$

is an orthogonal projection of $\mathbb{R}^m$ onto $R(\mathsf{A})$ we have to show three things:

- P is idempotent as $\mathsf{P}^2 = \mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t = \mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t = \mathsf{P}$ (as required).
- P is symmetric as $\mathsf{P}^t = [\mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t]^t = \mathsf{A}^{tt}[(\mathsf{A}^t\mathsf{A})^{-1}]^t\mathsf{A}^t = \mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t = \mathsf{P}$ (as required).
- $R(\mathsf{P}) = R(\mathsf{A})$ as (using a 'double inclusion proof'):
  - If $\mathbf{y} \in R(\mathsf{A}) \subseteq \mathbb{R}^m$, $\exists \mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathsf{A}\mathbf{x}$ and so
  $$\mathsf{P}\mathbf{y} = \mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathbf{y} = \mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathsf{A}\mathbf{x} = \mathsf{A}\mathbf{x} = \mathbf{y},$$
  i.e. $\mathbf{y} \in R(\mathsf{P})$. Thus, $R(\mathsf{A}) \subseteq R(\mathsf{P})$.
  - If $\mathbf{y} \in R(\mathsf{P}) \subseteq \mathbb{R}^m$, $\exists \mathbf{x} \in \mathbb{R}^n$ such that $\mathbf{y} = \mathsf{P}\mathbf{x}$ and so
  $$\mathbf{y} = \mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathbf{x} = \mathsf{A}\mathbf{z},$$
  where $\mathbf{z} = (\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathbf{x}$. Thus, $\mathbf{y} \in R(\mathsf{A})$ and so, $R(\mathsf{P}) \subseteq R(\mathsf{A})$.

  Consequently, $R(\mathsf{P}) = R(\mathsf{A})$ (as required).

A *least squares solution* to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is an $\mathbf{x}$, say $\mathbf{x}^*$, such that $\|\mathsf{A}\mathbf{x} - \mathbf{b}\|$ is minimised over all $\mathbf{x}$. Thus, $\mathsf{A}\mathbf{x}^*$ must be the orthogonal projection of $\mathbf{b}$ onto $R(\mathsf{A})$, i.e.

$$\mathsf{A}\mathbf{x}^* = \mathsf{A}(\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathbf{b},$$

and so, $\mathbf{x}^* = (\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathbf{b}$ is a least squares solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$.

(b) The quantities x and y are related by a rule of the form $y = mx + c$ for some constants m and c. Using this relationship and the given data, we construct a matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$, say

$$\begin{bmatrix} 2 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{bmatrix} \begin{bmatrix} m \\ c \end{bmatrix} = \begin{bmatrix} 13 \\ 17 \\ 22 \\ 25 \end{bmatrix}.$$

To estimate m and c using the least squares method, we have to find $\mathbf{x}^* = (\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathbf{b}$. Firstly, we find

$$\mathsf{A}^t\mathsf{A} = \begin{bmatrix} 2 & 4 & 5 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{bmatrix} = \begin{bmatrix} 81 & 17 \\ 17 & 4 \end{bmatrix},$$

and so, we can see that

$$(\mathsf{A}^t\mathsf{A})^{-1} = \frac{1}{35}\begin{bmatrix} 4 & -17 \\ -17 & 81 \end{bmatrix}.$$

Secondly, we find

$$\mathsf{A}^t\mathbf{b} = \begin{bmatrix} 2 & 4 & 5 & 6 \\ 1 & 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 4 & 1 \\ 5 & 1 \\ 6 & 1 \end{bmatrix} = \begin{bmatrix} 354 \\ 77 \end{bmatrix}.$$

And so, thirdly, we get

$$\mathbf{x}^* = \begin{bmatrix} m^* \\ c^* \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 4 & -17 \\ -17 & 81 \end{bmatrix} \begin{bmatrix} 354 \\ 77 \end{bmatrix} = \frac{1}{35}\begin{bmatrix} 107 \\ 219 \end{bmatrix},$$

which means that our least squares estimates are $m^* = 3\frac{2}{35}$ and $c^* = 6\frac{9}{35}$.

(c) The Fourier series of order n of a function $f(x)$ defined on the interval $[-\pi, \pi]$ is a projection onto the subspace of the vector space $\mathbb{F}^{[-\pi,\pi]}$ spanned by the orthonormal set of vectors

$$G_n = \left\{ \frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \ldots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \ldots, \frac{\sin nx}{\sqrt{\pi}} \right\},$$

which minimises the quantity

$$\int_{-\pi}^{\pi} [f(x) - g(x)]^2 dx,$$

where $g(x) \in \text{Lin}G_n$. So, if we define the inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(x) g(x) dx,$$

on $\mathbb{F}^{[-\pi,\pi]}$, the Fourier series of such a function is the orthogonal projection of $f(x)$ onto $\text{Lin}G_n$, and this can be written as:

$$\begin{aligned} &\tfrac{1}{2\pi}\langle f(x), 1\rangle\, 1 + \tfrac{1}{\pi}\langle f(x), \cos x\rangle \cos x + \cdots + \tfrac{1}{\pi}\langle f(x), \cos nx\rangle \cos nx \\ &\qquad + \tfrac{1}{\pi}\langle f(x), \sin x\rangle \sin x + \cdots + \tfrac{1}{\pi}\langle f(x), \sin nx\rangle \sin nx. \end{aligned}$$

Now, given that $f(x)$ is an *odd* function we have

$$\langle f(x), 1\rangle = \int_{-\pi}^{\pi} f(x) dx = 0,$$

and for each k (where $1 \leq k \leq n$) we have:

$$\langle f(x), \cos kx\rangle = \int_{-\pi}^{\pi} f(x) \cos kx \, dx = 0.$$

Thus, defining b_k as in the question, i.e.

$$b_k = \tfrac{1}{\pi}\langle f(x), \sin kx\rangle = \frac{1}{\pi}\int_{-\pi}^{\pi} f(x) \sin kx \, dx,$$

we can see that the expression for the orthogonal projection given above reduces to:

$$\sum_{k=1}^{n} b_k \sin kx,$$

(as required).

(d) To find the Fourier series of the function $f(x) = x$ defined on the interval $[-\pi, \pi]$, we note that this function is odd and so we can use the result from the previous part. So, for k such that $1 \leq k \leq n$, we use integration by parts to evaluate:

$$b_k = \frac{1}{\pi}\int_{-\pi}^{\pi} x \sin kx \, dx = \frac{1}{\pi}\left\{ \left[x \frac{-\cos kx}{k} \right]_{-\pi}^{\pi} + \int_{-\pi}^{\pi} \frac{\cos kx}{k} dx \right\},$$

and this gives,

$$b_k = -\frac{2\pi \cos k\pi}{\pi k} + 0 = -\frac{2(-1)^k}{k},$$

where we have used the fact that $\cos k\pi = (-1)^k$. Thus the Fourier series of order n for the function $f(x) = x$ defined in the interval $[-\pi, \pi]$ is

$$-2\sum_{k=1}^{n} \frac{(-1)^k}{k} \sin kx.$$

Question 4

A complex matrix A is:

- *unitary* iff $\mathsf{A}\mathsf{A}^\dagger = \mathsf{I}$.
- *normal* iff $\mathsf{A}\mathsf{A}^\dagger = \mathsf{A}^\dagger\mathsf{A}$.
- *unitarily diagonalisable* iff there is a unitary matrix P such that the matrix $\mathsf{P}^\dagger\mathsf{A}\mathsf{P}$ is diagonal.

We are then asked to prove that: If P is a unitary matrix, then all of the eigenvalues of P have modulus one. To do this, we let λ be any eigenvalue of the matrix P and we let $\mathbf{x}$ be an eigenvector corresponding to λ, i.e. $\mathsf{P}\mathbf{x} = \lambda\mathbf{x}$. Now, we note that

$$\mathbf{x}^\dagger\mathsf{P}^\dagger\mathsf{P}\mathbf{x} = (\mathsf{P}\mathbf{x})^\dagger\mathsf{P}\mathbf{x} = (\lambda\mathbf{x})^\dagger\lambda\mathbf{x} = \lambda^*\lambda\mathbf{x}^\dagger\mathbf{x} = |\lambda|^2\|\mathbf{x}\|^2,$$

and as P is unitary, we have $\mathsf{P}^\dagger\mathsf{P} = \mathsf{I}$, and so

$$\mathbf{x}^\dagger\mathsf{P}^\dagger\mathsf{P}\mathbf{x} = \mathbf{x}^\dagger\mathsf{I}\mathbf{x} = \mathbf{x}^\dagger\mathbf{x} = \|\mathbf{x}\|^2.$$

Then, equating these two expressions, we get

$$|\lambda|^2\|\mathbf{x}\|^2 = \|\mathbf{x}\|^2 \implies (|\lambda|^2 - 1)\|\mathbf{x}\|^2 = 0 \implies |\lambda| = 1 \text{ (since } \|\mathbf{x}\| \neq 0),$$

(as required).

Further, to show that the column vectors of P form an orthonormal set, we denote these vectors by $\mathbf{v}_i$ (for $1 \leq i \leq n$ say) and then note that

$$\mathsf{P}^\dagger\mathsf{P} = \begin{bmatrix} — & \mathbf{v}_1^\dagger & — \\ — & \mathbf{v}_2^\dagger & — \\ & \vdots & \\ — & \mathbf{v}_n^\dagger & — \end{bmatrix} \begin{bmatrix} | & | & & | \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_n \\ | & | & & | \end{bmatrix} = \begin{bmatrix} \langle\mathbf{v}_1,\mathbf{v}_1\rangle^* & \langle\mathbf{v}_1,\mathbf{v}_2\rangle^* & \cdots & \langle\mathbf{v}_1,\mathbf{v}_n\rangle^* \\ \langle\mathbf{v}_2,\mathbf{v}_1\rangle^* & \langle\mathbf{v}_2,\mathbf{v}_2\rangle^* & \cdots & \langle\mathbf{v}_2,\mathbf{v}_n\rangle^* \\ \vdots & \vdots & \ddots & \vdots \\ \langle\mathbf{v}_n,\mathbf{v}_1\rangle^* & \langle\mathbf{v}_n,\mathbf{v}_2\rangle^* & \cdots & \langle\mathbf{v}_n,\mathbf{v}_n\rangle^* \end{bmatrix},$$

where we have used the Complex Euclidean inner product and our convention [which allows us to express inner products in terms of matrix products (i.e. we have used the fact that $\mathbf{x}^\dagger\mathbf{y} = x_1^*y_1 + \cdots x_n^*y_n = (x_1y_1^* + \cdots x_ny_n^*)^* = \langle\mathbf{x},\mathbf{y}\rangle^*$). Consequently, as P is unitary, $\mathsf{P}^\dagger\mathsf{P} = \mathsf{I}$ and so equating these two expressions and matching up corresponding entries we can see that

$$\langle\mathbf{v}_i,\mathbf{v}_j\rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

that is, the set $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ of column vectors of P is orthonormal (as required).

Let us suppose that A is a normal matrix, and as such, it is unitarily diagonalisable, i.e. there is a unitary matrix P such that $\mathsf{P}^\dagger\mathsf{A}\mathsf{P}$ is diagonal. Further, the column vectors of P are given by the orthonormal set of vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ and $\mathbf{x}_i$ is an eigenvector corresponding to the eigenvalue λ_i of A (for $i = 1, 2, \ldots, n$). Now, as P is unitary, we can write $\mathsf{P}\mathsf{P}^\dagger = \mathsf{I}$, and so

$$\mathsf{I} = \begin{bmatrix} | & | & & | \\ \mathbf{x}_1 & \mathbf{x}_2 & \cdots & \mathbf{x}_n \\ | & | & & | \end{bmatrix} \begin{bmatrix} — & \mathbf{x}_1^\dagger & — \\ — & \mathbf{x}_2^\dagger & — \\ & \vdots & \\ — & \mathbf{x}_n^\dagger & — \end{bmatrix},$$

where evaluating the matrix product gives

$$\mathsf{I} = \mathbf{x}_1\mathbf{x}_1^\dagger + \mathbf{x}_2\mathbf{x}_2^\dagger + \cdots + \mathbf{x}_n\mathbf{x}_n^\dagger.$$

Then, multiplying both sides by A and noting that $\mathsf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$ we get

$$\mathsf{A} = \lambda_1\mathbf{x}_1\mathbf{x}_1^\dagger + \lambda_2\mathbf{x}_2\mathbf{x}_2^\dagger + \cdots + \lambda_n\mathbf{x}_n\mathbf{x}_n^\dagger,$$

which means that setting $\mathsf{E}_i = \mathbf{x}_i\mathbf{x}_i^\dagger$ (for $i = 1, 2, \ldots, n$) as suggested in the question will yield the *spectral decomposition* of A, i.e.

$$\mathsf{A} = \lambda_1\mathsf{E}_1 + \lambda_2\mathsf{E}_2 + \cdots + \lambda_n\mathsf{E}_n,$$

(as required). Indeed each E_i is an $n \times n$ matrix and we can see that

$$\mathsf{E}_i\mathsf{E}_j = \mathbf{x}_i\mathbf{x}_i^\dagger\mathbf{x}_j\mathbf{x}_j^\dagger = \langle\mathbf{x}_i, \mathbf{x}_j\rangle^*\mathbf{x}_i\mathbf{x}_j^\dagger = \begin{cases} \mathbf{x}_i\mathbf{x}_i^\dagger & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

since the set of eigenvectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ is orthonormal, thus,

$$\mathsf{E}_i\mathsf{E}_j = \begin{cases} \mathsf{E}_i & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases},$$

(as required).

We are given the matrix

$$\mathsf{A} = \begin{bmatrix} 1 & i & 0 \\ i & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

and we are asked to find its spectral decomposition. To do this, we need to find its eigenvalues, which we do by solving the determinant equation $|\mathsf{A} - \lambda\mathsf{I}| = 0$, i.e.

$$\mathsf{A} = \begin{vmatrix} 1-\lambda & i & 0 \\ i & 1-\lambda & 0 \\ 0 & 0 & 1-\lambda \end{vmatrix} \implies (1-\lambda)[(1-\lambda)^2 + 1] = 0,$$

where we have evaluated the determinant by performing a co-factor expansion along its third column. Simplifying the resulting polynomial equation we get

$$(1-\lambda)(\lambda^2 - 2\lambda + 2) = 0 \implies \lambda = 1,\ 1 \pm i,$$

are the eigenvalues. To find the eigenvectors, we solve the matrix equation $(\mathsf{A} - \lambda\mathsf{I})\mathbf{x} = \mathbf{0}$ for each eigenvalue:

- For $\lambda = 1$, we have

$$\begin{bmatrix} 0 & i & 0 \\ i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies x = 0 \text{ and } y = 0 \implies \mathbf{x} = s\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad \text{for } s \in \mathbb{R}.$$

  So, setting $s = 1$ (as we want the eigenvectors to form an orthonormal set), we take $[0, 0, 1]^t$ to be the eigenvector corresponding to $\lambda = 1$.

- For $\lambda = 1 \pm i$, we have

$$\begin{bmatrix} \mp i & i & 0 \\ i & \mp i & 0 \\ 0 & 0 & \mp i \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies \mp x + y = 0 \text{ and } z = 0 \implies \mathbf{x} = s\begin{bmatrix} \pm 1 \\ 1 \\ 0 \end{bmatrix} \quad \text{for } s \in \mathbb{R}.$$

  So, setting $s = 1/\sqrt{2}$ (as we want the eigenvectors to form an orthonormal set), we take $\frac{1}{\sqrt{2}}[\pm 1, 1, 0]^t$ to be the eigenvectors corresponding to $\lambda = 1 \pm i$.

Consequently, the spectral decomposition for A will be

$$\mathsf{A} = 1\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\begin{bmatrix} 0 & 0 & 1 \end{bmatrix} + (1+i)\begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix} + (1-i)\begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{bmatrix}\begin{bmatrix} -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} & 0 \end{bmatrix},$$

or, using the representation given in the question,

$$\mathsf{A} = 1 \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} + \frac{1+i}{2} \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} + \frac{1-i}{2} \begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

(Notice that this result can easily be checked by adding up the three matrices on the right-hand-side and verifying that their sum does indeed give A.)

Question 5

(a) We write the system of coupled differential equations given in the question in the form $\mathbf{y}' = \mathsf{A}\mathbf{y}$ where

$$\mathsf{A} = \begin{bmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{bmatrix} \text{ and } \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \end{bmatrix}.$$

To solve this system, we need to diagonalise the matrix A, and we do this by first finding the eigenvalues of A, i.e. we solve the determinant equation $|\mathsf{A} - \lambda \mathsf{I}| = 0$:

$$\begin{vmatrix} 2-\lambda & 1 & 1 \\ 1 & 2-\lambda & 1 \\ 1 & 1 & 2-\lambda \end{vmatrix} = 0 \implies (2-\lambda)[(2-\lambda)^2 - 1] - [(2-\lambda) - 1] + [1 - (2-\lambda)] = 0,$$

and the easiest way to solve this equation is to perform the following simplification:

$$\begin{aligned} & (2-\lambda)(\lambda^2 - 4\lambda + 3) - 2 + 2\lambda = 0 \\ \implies & (2-\lambda)(\lambda - 1)(\lambda - 3) + 2(\lambda - 1) = 0 \\ \implies & (\lambda - 1)[(2-\lambda)(\lambda - 3) + 2] = 0 \\ \implies & (\lambda - 1)[-\lambda^2 + 5\lambda - 4] = 0 \\ \implies & (\lambda - 1)(\lambda - 1)(\lambda - 4) = 0 \end{aligned}$$

i.e. the eigenvalues are $\lambda = 1, 1, 4$. Secondly, we need to find the eigenvectors corresponding to these eigenvalues, i.e. we need to solve the matrix equation $(\mathsf{A} - \lambda \mathsf{I})\mathbf{x} = \mathbf{0}$ for each eigenvalue:

- For $\lambda = 1$, we have

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies x + y + z = 0 \implies \mathbf{x} = \begin{bmatrix} -s-t \\ s \\ t \end{bmatrix} = s \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \text{ for } s, t \in \mathbb{R}.$$

  So, setting $[s, t]^t = [1, 0]^t$ and $[s, t]^t = [0, 1]^t$ (as we want two linearly independent eigenvectors corresponding to this multiplicitous eigenvalue), we take $[-1, 1, 0]^t$ and $[-1, 0, 1]^t$ to be the eigenvectors corresponding to $\lambda = 1$.

- For $\lambda = 4$, we have

$$\begin{bmatrix} -2 & 1 & 1 \\ 1 & -2 & 1 \\ 1 & 1 & -2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies \left.\begin{aligned} -2x + y + z = 0 \\ x - 2y + z = 0 \\ x + y - 2z = 0 \end{aligned}\right\} \implies x = y = z \implies \mathbf{x} = s \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

  for $s \in \mathbb{R}$. So, setting $s = 1$, we take $[1, 1, 1]^t$ to be the eigenvector corresponding to $\lambda = 4$.

Thirdly, we construct the matrices P and D, i.e.

$$\mathsf{P} = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \text{ and } \mathsf{D} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{bmatrix}.$$

We now note that as $\mathsf{P}^{-1}\mathsf{A}\mathsf{P} = \mathsf{D}$, we can write $\mathsf{A} = \mathsf{P}\mathsf{D}\mathsf{P}^{-1}$, and so $\mathbf{y}' = \mathsf{A}\mathbf{y}$ can be written as $\mathbf{y}' = \mathsf{P}\mathsf{D}\mathsf{P}^{-1}\mathbf{y}$. Thus, making the substitution $\mathbf{z} = [z_1, z_2, z_3]^t = \mathsf{P}^{-1}\mathbf{y}$, we obtain a system of *un*coupled differential equations, namely $\mathbf{z}' = \mathsf{D}\mathbf{z}$. Expanding this out we get

$$\left.\begin{aligned} z_1' &= z_1 \\ z_2' &= z_2 \\ z_3' &= 4z_3 \end{aligned}\right\} \implies \left.\begin{aligned} z_1 &= Ae^t \\ z_2 &= Be^t \\ z_3 &= Ce^{4t} \end{aligned}\right\} \implies \mathbf{z} = \begin{bmatrix} Ae^t \\ Be^t \\ Ce^{4t} \end{bmatrix},$$

where A, B and C are constants. Consequently, using $\mathbf{y} = \mathsf{P}\mathbf{z}$, yields

$$\mathbf{y}(t) = \begin{bmatrix} -1 & -1 & 1 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} Ae^t \\ Be^t \\ Ce^{4t} \end{bmatrix} = \begin{bmatrix} -Ae^t - Be^t + Ce^{4t} \\ Ae^t + Ce^{4t} \\ Be^t + Ce^{4t} \end{bmatrix},$$

which is the *general* solution to the system of coupled differential equations in the question. To find the *particular* solution corresponding to the initial conditions $y_1(0) = -1$, $y_2(0) = 1$ and $y_3(0) = 0$, we note that

$$\mathbf{y}(0) = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -A - B + C \\ A + C \\ B + C \end{bmatrix},$$

which gives us three simultaneous equations for A, B and C. Adding these three equations together yields $C = 0$, and this in turn gives $B = 0$ and $A = 1$. Thus, substituting these values into our expression for $\mathbf{y}(t)$, we get the sought after solution, i.e.

$$\mathbf{y}(t) = \begin{bmatrix} -e^t \\ e^t \\ 0 \end{bmatrix}.$$

(b) To find the steady states of the system of non-linear differential equations given in the question we set the right-hand-sides equal to zero and solve the resulting simultaneous equations, i.e.

$$\left.\begin{aligned} y_1(3 - 2y_1 - y_2) = 0 \\ y_2(2 - y_1 - y_2) = 0 \end{aligned}\right\} \implies \left.\begin{array}{llll} y_1 = 0 & (i) & \text{or} & 3 - 2y_1 - y_2 = 0 \quad (ii) \\ y_2 = 0 & (iii) & \text{or} & 2 - y_1 - y_2 = 0 \quad (iv) \end{array}\right\}$$

So taking these in turn we can find the steady states $\mathbf{y}^* = [y_1^*, y_2^*]^t$, namely:

- (i) and (iii) give $[0, 0]^t$.
- (i) and (iv) give $[0, 2]^t$.
- (ii) and (iii) give $[\frac{3}{2}, 0]^t$.
- (ii) and (iv) give $[1, 1]^t$.

Now, the steady state where neither y_1^* nor y_2^* equals zero is clearly $[1, 1]^t$. So, to show that this is asymptotically stable, we calculate the Jacobian of the right-hand-sides, i.e.

$$DF(\mathbf{y}) = \begin{bmatrix} 3 - 4y_1 - y_2 & -y_1 \\ -y_2 & 2 - y_1 - 2y_2 \end{bmatrix},$$

and evaluate it at $\mathbf{y}^* = [1, 1]^t$ to get the matrix

$$\mathsf{DF}([1, 1]^t) = \begin{bmatrix} -2 & -1 \\ -1 & -1 \end{bmatrix}.$$

We now calculate the eigenvalues of this matrix by solving the determinant equation given by

$$\begin{vmatrix} -2 - \lambda & -1 \\ -1 & -1 - \lambda \end{vmatrix} = 0 \implies (2 + \lambda)(1 + \lambda) - 1 = 0 \implies \lambda^2 + 3\lambda + 1 = 0,$$

which gives us $\lambda = \frac{1}{2}(-3 \pm \sqrt{5}) = -0.38, -2.62$. As these eigenvalues are real and negative, the steady state $\mathbf{y}^* = [1, 1]^t$ is asymptotically stable (as required).

Question 6

(a) We are asked to prove that $\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|$ for all $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$. (Notice that this is just the Triangle Inequality!) To do this we note that:

$$\begin{aligned} \|\mathbf{x}+\mathbf{y}\|^2 &= \langle \mathbf{x}+\mathbf{y}, \mathbf{x}+\mathbf{y} \rangle \\ &= \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \\ \implies \quad \|\mathbf{x}+\mathbf{y}\|^2 &= \|\mathbf{x}\|^2 + 2\langle \mathbf{x}, \mathbf{y} \rangle + \|\mathbf{y}\|^2 \end{aligned}$$

where we have used the symmetry and linearity properties of real inner products. However, the Cauchy-Schwarz inequality tells us that $|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\|\|\mathbf{y}\|$, and so

$$\|\mathbf{x}+\mathbf{y}\|^2 \leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2.$$

But, factorising the right-hand-side then gives:

$$\|\mathbf{x}+\mathbf{y}\|^2 \leq (\|\mathbf{x}\| + \|\mathbf{y}\|)^2,$$

and hence,

$$\|\mathbf{x}+\mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\|,$$

(as required)

(b) A subset C of $\mathbb{R}^n$ is *convex* iff for any $\mathbf{x}, \mathbf{y} \in C$, $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C$ too for all α such that $0 \leq \alpha \leq 1$.

To prove that: If C_1 and C_2 are convex sets, then $C_1 \cap C_2$ is a convex set too, we use the following argument:

> Suppose that C_1 and C_2 are convex sets, that is for α such that $0 \leq \alpha \leq 1$:
>
> - For any $\mathbf{x}_1, \mathbf{y}_1 \in C_1$, $\alpha\mathbf{x}_1 + (1-\alpha)\mathbf{y}_1 \in C_1$.
> - For any $\mathbf{x}_2, \mathbf{y}_2 \in C_2$, $\alpha\mathbf{x}_2 + (1-\alpha)\mathbf{y}_2 \in C_2$.
>
> Now, to show that the set $C_1 \cap C_2$ is convex, we need to establish that for any $\mathbf{x}, \mathbf{y} \in C_1 \cap C_2$ and α such that $0 \leq \alpha \leq 1$, the vector $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C_1 \cap C_2$ too. To do this we note that as $\mathbf{x}, \mathbf{y} \in C_1 \cap C_2$, it must be the case that $\mathbf{x}, \mathbf{y} \in C_1$ and $\mathbf{x}, \mathbf{y} \in C_2$. But, as C_1 and C_2 are convex sets, this means that $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C_1$ and $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C_2$, i.e. $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C_1 \cap C_2$ too (as required).

To show that the set $C_1 = \{\mathbf{x} \in \mathbb{R}^2 | \, \|\mathbf{x}\| \leq 3\}$ is convex we need to establish that for any $\mathbf{x}, \mathbf{y} \in C_1$ and α such that $0 \leq \alpha \leq 1$, the vector $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C_1$. To do this, we note that:

- As $\mathbf{x} \in C_1$, $\|\mathbf{x}\| \leq 3$.
- As $\mathbf{y} \in C_1$, $\|\mathbf{y}\| \leq 3$.

and to show that $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C_1$ we need to show that $\|\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\| \leq 3$. To do this, we note that:

$$\begin{aligned} \|\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\| &\leq \|\alpha\mathbf{x}\| + \|(1-\alpha)\mathbf{y}\| && \text{:Using (a)} \\ &= \alpha\|\mathbf{x}\| + (1-\alpha)\|\mathbf{y}\| && \text{:As } \alpha, 1-\alpha \geq 0 \\ &\leq 3\alpha + 3(1-\alpha) && \text{:As } \mathbf{x}, \mathbf{y} \in C_1 \\ &= 3 \end{aligned}$$

and so $\|\alpha\mathbf{x} + (1-\alpha)\mathbf{y}\| \leq 3$ (as required).

To show that the set $C_2 = \{\mathbf{x} = [x_1, x_2]^t \in \mathbb{R}^2 | \, x_1 + x_2 \leq 0\}$ is convex we need to establish that for any $\mathbf{x}, \mathbf{y} \in C_2$ and α such that $0 \leq \alpha \leq 1$, the vector $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} \in C_2$. To do this, we note that:

- As $\mathbf{x} = [x_1, x_2]^t \in C_2$, $x_1 + x_2 \leq 0$.
- As $\mathbf{y} = [y_1, y_2]^t \in C_2$, $y_1 + y_2 \leq 0$.

and to show that $\alpha\mathbf{x} + (1-\alpha)\mathbf{y} = [\alpha x_1 + (1-\alpha)y_1, \alpha x_2 + (1-\alpha)y_2]^t \in C_2$ we need to show that $\alpha x_1 + (1-\alpha)y_1 + \alpha x_2 + (1-\alpha)y_2 \leq 0$. To do this, we note that:

$$\alpha x_1 + (1-\alpha)y_1 + \alpha x_2 + (1-\alpha)y_2 = \alpha(x_1 + x_2) + (1-\alpha)(y_1 + y_2) \leq 0$$

since $\alpha, 1-\alpha \geq 0$ and $\mathbf{x}, \mathbf{y} \in C_2$ (as required).

An *extreme point* of a convex set C is a point $\mathbf{x} \in C$ which can **not** be expressed in the form $\mathbf{x} = \alpha\mathbf{y} + (1-\alpha)\mathbf{z}$ — where $0 \leq \alpha \leq 1$ and $\mathbf{y}, \mathbf{z} \in C$ — **unless** $\alpha = 0$ or $\alpha = 1$.

To find the extreme points of the set $C_1 \cap C_2$, we start by visualising what this set looks like — see Figure 1. It should then be clear that the extreme points of $C_1 \cap C_2$ are all points on the arc ACB

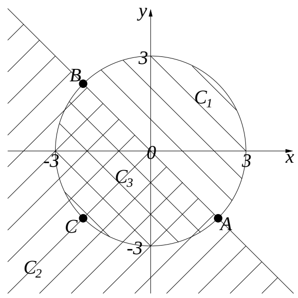

Figure 1: *In this diagram, the set C_1 is the disc of radius 3 centred on the origin and the set C_2 is the half-plane. The set $C_3 = C_1 \cap C_2$ is the region where they intersect. The extreme points of this set are given by all points on the arc ACB.*

which is given by the equation $\|\mathbf{x}\| = 3$ (i.e. for $\mathbf{x} = [x_1, x_2]^t$, this is the curve where $x_1^2 + x_2^2 = 9$) and the points A and B are $(3/\sqrt{2}, -3/\sqrt{2})$ and $(-3/\sqrt{2}, 3/\sqrt{2})$ respectively. (Notice that you find the points A and B by solving the equations $x_1^2 + x_2^2 = 9$ and $x_1 + x_2 = 0$ simultaneously.)

(c) Writing the given linear programming problem in matrix notation, we have:

$$\text{MINIMISE } \mathbf{c}^t\mathbf{x} \text{ subject to } \mathsf{A}\mathbf{x} \geq \mathbf{b} \text{ and } \mathbf{x} \geq \mathbf{0},$$

where

$$\mathbf{c} = \begin{bmatrix} 12 \\ 8 \\ 14 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x \\ y \\ z \end{bmatrix}, \quad \mathsf{A} = \begin{bmatrix} 2 & 2 & 4 \\ 4 & 2 & 2 \end{bmatrix} \text{ and } \mathbf{b} = \begin{bmatrix} 6 \\ 4 \end{bmatrix}.$$

Its dual is then given by:

$$\text{MAXIMISE } \mathbf{b}^t\mathbf{y} \text{ subject to } \mathsf{A}^t\mathbf{y} \leq \mathbf{c} \text{ and } \mathbf{y} \geq \mathbf{0},$$

where $\mathbf{y} = [y_1, y_2]^t$, or writing this out in full, we get:

$$\text{MAXIMISE } 6y_1 + 4y_2 \text{ subject to } \begin{bmatrix} 2 & 4 \\ 2 & 2 \\ 4 & 2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \leq \begin{bmatrix} 12 \\ 8 \\ 14 \end{bmatrix} \text{ and } \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \geq \mathbf{0}.$$

To solve the dual, we start by drawing the feasible region and finding the extreme points — see Figure 2. We now calculate the coordinates of each of these extreme points by finding the points of intersection of the relevant lines. Once this is done, we can find the value of the objective function, i.e. $6y_1 + 4y_2$, at each of these points to find out where it is maximised. Doing this, we find that:

Point	$6y_1 + 4y_2$
A $(0, 3)$	12
B $(2, 2)$	20
C $(3, 1)$	22
D $(3.5, 0)$	21
0 $(0, 0)$	0

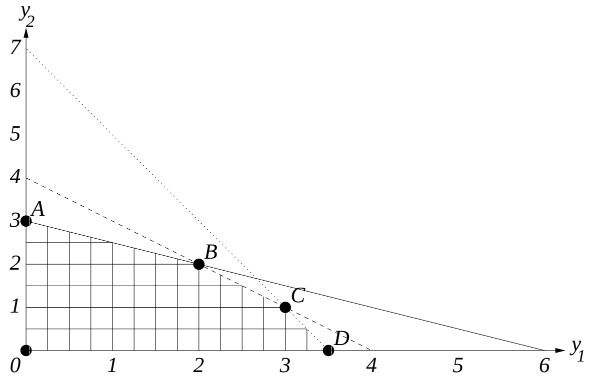

Figure 2: *The feasible region for the dual problem is the cross-hatched region bounded by the y_1 and y_2-axes and the lines $y_1 + 2y_2 = 6$ (i.e. —), $y_1 + y_2 = 4$ (i.e. $- - -$) and $2y_1 + y_2 = 7$ (i.e. $\cdots$). The extreme points are the five 'corners' labelled 0, A, B, C and D.*

and so, clearly, the objective function of the dual problem is maximised at C where it takes the value 22. Consequently, by the Duality Theorem, the solution to the original linear programming problem (i.e. the minimum value of $12x + 8y + 14z$ subject to the given constraints) is also 22.