Further Mathematical Methods (Linear Algebra)

Solutions For The 2000 Examination

Question 1

(a) For a non-empty subset W of V to be a subspace of V we require that, for all vectors $\mathbf{x}, \mathbf{y} \in W$ and all scalars $\alpha \in \mathbb{R}$:

i. **Closure under vector addition:** $\mathbf{x} + \mathbf{y} \in W$.

ii. **Closure under scalar multiplication:** $\alpha\mathbf{x} \in W$.

To be an inner product on V, a function $\langle \mathbf{x}, \mathbf{y} \rangle$ which maps vectors $\mathbf{x}, \mathbf{y} \in V$ to $\mathbb{R}$ must be such that:

i. **Positivity:** $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$ and, $\langle \mathbf{x}, \mathbf{x} \rangle = 0$ if and only if $\mathbf{x} = \mathbf{0}$.

ii. **Symmetry:** $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$.

iii. **Linearity:** $\langle \alpha\mathbf{x} + \beta\mathbf{y}, \mathbf{z} \rangle = \alpha\langle \mathbf{x}, \mathbf{z} \rangle + \beta\langle \mathbf{y}, \mathbf{z} \rangle$.

for all vectors $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and all scalars $\alpha, \beta \in \mathbb{R}$.

(b) We consider the vector space which consists of all 2×2 real matrices given by

$$V = \left\{ \begin{bmatrix} a & b \\ c & d \end{bmatrix} \;\middle|\; a, b, c, d \in \mathbb{R} \right\},$$

where the operations of vector addition and scalar multiplication are defined in the normal way. To show that the set $W \subset V$ defined by

$$W = \left\{ \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} \;\middle|\; a, c, d \in \mathbb{R} \right\},$$

is a subspace of V, we start by noting that W is a sub*set* of V since any matrix of the form

$$\begin{bmatrix} a & 0 \\ c & d \end{bmatrix} \text{ can be written as } \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

with $b = 0$. So, to show that W is a sub*space* of V, we take any two vectors in W, say

$$\mathbf{u} = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} \text{ and } \mathbf{v} = \begin{bmatrix} a' & 0 \\ c' & d' \end{bmatrix},$$

and any scalar $\alpha \in \mathbb{R}$ and note that W is closed under:

- **vector addition** since

$$\mathbf{u} + \mathbf{v} = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} + \begin{bmatrix} a' & 0 \\ c' & d' \end{bmatrix} = \begin{bmatrix} a + a' & 0 \\ c + c' & d + d' \end{bmatrix},$$

  and so, $\mathbf{u} + \mathbf{v} \in W$ too since $a + a', b + b', d + d' \in \mathbb{R}$.

- **scalar multiplication** since

$$\alpha\mathbf{u} = \begin{bmatrix} \alpha a & 0 \\ \alpha c & \alpha d \end{bmatrix},$$

  and so, $\alpha\mathbf{u} \in W$ too since $\alpha a, \alpha b, \alpha d \in \mathbb{R}$.

as required. Clearly, the matrix which plays the role of the additive identity (i.e. $\mathbf{0}$) in this vector space is

$$\begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

since

$$\mathbf{u} + \mathbf{0} = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} = \mathbf{0}.$$

(c) We are asked to show that the function defined by

$$\left\langle \begin{bmatrix} a & 0 \\ c & d \end{bmatrix}, \begin{bmatrix} a' & 0 \\ c' & d' \end{bmatrix} \right\rangle = aa' + cc' + dd',$$

is an inner product on W. To do this, we show that this formula satisfies all of the conditions given in part **(a)**. Thus, taking any three vectors

$$\mathbf{u} = \begin{bmatrix} a & 0 \\ c & d \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} a' & 0 \\ c' & d' \end{bmatrix} \quad \text{and} \quad \mathbf{w} = \begin{bmatrix} a'' & 0 \\ c'' & d'' \end{bmatrix},$$

in W and any two scalars α and β in $\mathbb{R}$ we have:

i. $\langle \mathbf{u}, \mathbf{u} \rangle = a^2 + c^2 + d^2$ which is the sum of the squares of three real numbers and as such is real and non-negative. Further, to show that $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$ (where here, $\mathbf{0}$ is the 2×2 zero matrix), we note that:

   - **LTR:** If $\langle \mathbf{u}, \mathbf{u} \rangle = 0$, then $a^2 + c^2 + d^2 = 0$. But, this is the sum of the squares of three real numbers and so it must be the case that $a = c = d = 0$. Thus, $\mathbf{u} = \mathbf{0}$.
   - **RTL:** If $\mathbf{u} = 0$, then $a = c = d = 0$. Thus, $\langle \mathbf{u}, \mathbf{u} \rangle = 0$.

   (as required).

ii. Obviously, $\langle \mathbf{u}, \mathbf{v} \rangle = aa' + cc' + dd' = a'a + c'c + d'd = \langle \mathbf{v}, \mathbf{u} \rangle$.

iii. We note that the vector $\alpha\mathbf{u} + \beta\mathbf{v}$ is given by

$$\alpha\mathbf{u} + \beta\mathbf{v} = \alpha \begin{bmatrix} a & 0 \\ c & d \end{bmatrix} + \beta \begin{bmatrix} a' & 0 \\ c' & d' \end{bmatrix} = \begin{bmatrix} \alpha a + \beta a' & 0 \\ \alpha c + \beta c' & \alpha d + \beta d' \end{bmatrix},$$

   and so we have

$$\begin{aligned} \langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle &= (\alpha a + \beta a')a'' + (\alpha c + \beta c')c'' + (\alpha d + \beta d')d'' \\ &= \alpha(aa'' + cc'' + dd'') + \beta(a'a'' + b'b'' + c'c'') \\ \therefore\ \langle \alpha\mathbf{u} + \beta\mathbf{v}, \mathbf{w} \rangle &= \alpha\langle \mathbf{u}, \mathbf{w} \rangle + \beta\langle \mathbf{v}, \mathbf{w} \rangle, \end{aligned}$$

Consequently, the formula given above does define an inner product on W (as required).

(d) To find a matrix A such that

$$[\mathbf{M}]_S = \mathsf{A}[\mathbf{M}]_{S'},$$

where $[\mathbf{M}]_S$ and $[\mathbf{M}]_{S'}$ are the coordinate vectors of $\mathbf{M} \in W$ relative to the bases

$$S = \left\{ \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \right\},$$

and

$$S' = \left\{ \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix}, \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix} \right\},$$

respectively, we use the definition of coordinate vector. That is, we use the fact that the equality

$$a \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} + b \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} + c \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} = a' \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} + b' \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + c' \begin{bmatrix} 0 & 0 \\ -1 & 1 \end{bmatrix},$$

holds if and only if

$$\begin{bmatrix} a \\ b \\ c \end{bmatrix}_S = \mathsf{A} \begin{bmatrix} a' \\ b' \\ c' \end{bmatrix}_{S'},$$

where

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

is the required matrix.

Question 2. The required Leslie matrix is

$$\mathsf{L} = \begin{bmatrix} 1 & 3 & 0 \\ 1/4 & 0 & 0 \\ 0 & 1/2 & 0 \end{bmatrix},$$

and to find its unique real positive eigenvalue we solve the equation

$$\begin{vmatrix} 1-\lambda & 3 & 0 \\ 1/4 & -\lambda & 0 \\ 0 & 1/2 & -\lambda \end{vmatrix} = 0 \implies (1-\lambda)\lambda^2 + \frac{3}{4}\lambda = 0 \implies \lambda\left(\lambda^2 - \lambda - \frac{3}{4}\right) = 0,$$

where we have simplified the determinant using a cofactor expansion on the third column. This can be factorised to give:

$$\lambda\left(\lambda - \frac{3}{2}\right)\left(\lambda + \frac{1}{2}\right) = 0,$$

and so the eigenvalues are $\lambda = 0, \frac{3}{2}, -\frac{1}{2}$. Thus, the unique positive real eigenvalue is $\lambda_1 = \frac{3}{2}$ and this is clearly dominant since it is larger in magnitude than the other two (or, alternatively, since we have two successive fertile classes). To describe the long-term behaviour of this population, we recall from the lectures that

$$\mathbf{x}^{(k)} \simeq c\lambda_1^k \mathbf{v}_1 \text{ and } \mathbf{x}^{(k)} \simeq \lambda_1 \mathbf{x}^{(k-1)},$$

for large k since we have a dominant eigenvalue. To find $\mathbf{v}_1$, an eigenvector corresponding to λ_1, it is probably easiest to recall that:

$$\mathbf{v}_1 = \begin{bmatrix} 1 \\ b_1/\lambda_1 \\ b_1 b_2 \ \lambda_1^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1/6 \\ 1/18 \end{bmatrix} \text{ and so we take } \mathbf{v}_1 = \begin{bmatrix} 18 \\ 3 \\ 1 \end{bmatrix}.$$

Consequently, in the long-term, we have

$$\mathbf{x}^{(k)} \simeq c\left(\frac{3}{2}\right)^k \begin{bmatrix} 18 \\ 3 \\ 1 \end{bmatrix} \text{ and } \mathbf{x}^{(k)} \simeq \frac{3}{2}\mathbf{x}^{(k-1)},$$

where the first result tells us that

> The proportion of the female population in each age-class becomes *constant* in the ratio 18:3:1.

whereas the second result tells us that

> The population in each age class increases by a factor of $\frac{3}{2}$ (i.e. increases by 50%) every time period (i.e. every twenty years in this case).

(b) The required Leslie matrix and the initial population distribution vector are given by:

$$\mathsf{L} = \begin{bmatrix} 1 & 3 \\ 1/4 & 0 \end{bmatrix} \text{ and } \mathbf{x}^{(0)} = \begin{bmatrix} 320 \\ 320 \end{bmatrix}.$$

So, to find an *exact* formula for the population distribution vector at the end of the kth time period we note that

$$\mathbf{x}^{(k)} = \mathsf{L}^k \mathbf{x}^{(0)} = \mathsf{P}\mathsf{D}^k\mathsf{P}^{-1}\mathbf{x}^{(0)},$$

where D is a diagonal matrix and the matrix P is invertible. To find these two matrices we can see that the eigenvalues of L are given by the equation:

$$\begin{vmatrix} 1-\lambda & 3 \\ 1/4 & -l \end{vmatrix} = 0 \implies -\lambda(1-\lambda) - \frac{3}{4} = 0 \implies \lambda^2 - \lambda - \frac{3}{4} = 0 \implies \left(\lambda - \frac{3}{2}\right)\left(\lambda + \frac{1}{2}\right) = 0,$$

i.e. the eigenvalues of L are $\lambda = \frac{3}{2}$ and $\lambda = -\frac{1}{2}$. The corresponding eigenvectors are then given by:

- For $\lambda = \frac{3}{2}$, we have

$$\begin{bmatrix} -1/2 & 3 \\ 1/4 & -3/2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{0} \implies -x + 6y = 0,$$

  i.e. $x = 6y$ for any $y \in \mathbb{R}$, and so a corresponding eigenvector would be $[6, 1]^t$.

- For $\lambda = -\frac{1}{2}$, we have

$$\begin{bmatrix} 3/2 & 3 \\ 1/4 & 1/2 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \mathbf{0} \implies x + 2y = 0,$$

  i.e. $x = -2y$ for any $y \in \mathbb{R}$, and so a corresponding eigenvector would be $[-2, 1]^t$.

Consequently, we can take our matrices P and D to be

$$\mathsf{P} = \begin{bmatrix} 6 & -2 \\ 1 & 1 \end{bmatrix} \quad \text{and} \quad \mathsf{D} = \begin{bmatrix} 3/2 & 0 \\ 0 & -1/2 \end{bmatrix},$$

and so, we have

$$\mathsf{P}^{-1} = \frac{1}{8} \begin{bmatrix} 1 & 2 \\ -1 & 6 \end{bmatrix},$$

as well. Hence, to calculate $\mathbf{x}^{(k)} = \mathsf{P}\mathsf{D}^k\mathsf{P}^{-1}\mathbf{x}^{(0)}$, we note that

$$\mathsf{D}^k\mathsf{P}^{-1}\mathbf{x}^{(0)} = \frac{1}{8} \begin{bmatrix} (3/2)^k & 0 \\ 0 & (-1/2)^k \end{bmatrix} \begin{bmatrix} 1 & 2 \\ -1 & 6 \end{bmatrix} \begin{bmatrix} 320 \\ 320 \end{bmatrix} = \begin{bmatrix} (3/2)^k & 2(3/2)^k \\ -(-1/2)^k & (-1/2)^k \end{bmatrix} \begin{bmatrix} 40 \\ 40 \end{bmatrix} = 40 \begin{bmatrix} 3(3/2)^k \\ 5(-1/2)^k \end{bmatrix},$$

and so,

$$\mathbf{x}^{(k)} = \mathsf{P}\mathsf{D}^k\mathsf{P}^{-1}\mathbf{x}^{(0)} = 40 \begin{bmatrix} 6 & -2 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3(3/2)^k \\ 5(-1/2)^k \end{bmatrix} = 40 \begin{bmatrix} 18(3/2)^k - 10(-1/2)^k \\ 3(3/2)^k + 5(-1/2)^k \end{bmatrix},$$

is the required exact formula for $\mathbf{x}^{(k)}$.

The above expression for $\mathbf{x}^{(k)}$ gives the population distribution vector at time $t_k = 3000 + 20k$ years, so for the population distribution vector in the year 3040, we want to evaluate this for $k = 2$. Thus,

$$\mathbf{x}^{(2)} = 40 \begin{bmatrix} 18(3/2)^2 - 10(-1/2)^2 \\ 3(3/2)^2 + 5(-1/2)^2 \end{bmatrix} = 10 \begin{bmatrix} 18 \times 9 - 10 \\ 3 \times 9 + 5 \end{bmatrix} = \begin{bmatrix} 1520 \\ 320 \end{bmatrix},$$

is the required population distribution vector.

(c) As the females are now infertile, the Leslie matrix for this population is

$$\mathsf{L} = \begin{bmatrix} 0 & 0 \\ 1/4 & 0 \end{bmatrix}.$$

Thus, we can see that

$$\mathbf{x}^{(3)} = \mathsf{L}\mathbf{x}^{(2)} = \begin{bmatrix} 0 & 0 \\ 1/4 & 0 \end{bmatrix} \begin{bmatrix} 1520 \\ 320 \end{bmatrix} = \begin{bmatrix} 0 \\ 1520/4 \end{bmatrix},$$

and so,

$$\mathbf{x}^{(4)} = \mathsf{L}\mathbf{x}^{(3)} = \begin{bmatrix} 0 & 0 \\ 1/4 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 1520/4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

i.e. the female population dies out after 40 years, i.e. in the year 3080.

Or, alternatively, noting that the maximum lifespan of a female in this population is forty years, it should be clear that (since no new females are being born after 3040) they will all be dead within forty years, i.e. by the year 3080.

Question 3

(a) We are given that Y and Z are subspaces of a vector space X and we are asked to show that the following two statements are equivalent:

I. $X = Y + Z$ and every $\mathbf{x} \in X$ can be written uniquely as $\mathbf{x} = \mathbf{y} + \mathbf{z}$ with $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$.

II. $X = Y + Z$ and $Y \cap Z = \{\mathbf{0}\}$.

That is, we need to show that I is true *if and only if* II is true. So, as this is an 'if and only if' statement we have to show that it holds 'both ways':

If I, then II: We are given that $X = Y + Z$ and every $\mathbf{x} \in X$ can be written uniquely as $\mathbf{x} = \mathbf{y} + \mathbf{z}$ with $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$ and so noting that:

- We are allowed to assume that $X = Y + Z$.
- Taking any vector $\mathbf{u} \in Y \cap Z$ we have $\mathbf{u} \in Y$ and $\mathbf{u} \in Z$. Now, as $\mathbf{u} \in X$ and $\mathbf{0}$ is in both Y and Z (since they are subspaces of X), we can write

$$\mathbf{u} = \mathbf{0} + \mathbf{u} = \mathbf{u} + \mathbf{0},$$

  and so, by uniqueness, we have $\mathbf{u} = \mathbf{0}$. Thus, $Y \cap Z = \{\mathbf{0}\}$.

we have $X = Y + Z$ and $Y \cap Z = \{\mathbf{0}\}$ (as required).

If II, then I: We are given that $X = Y + Z$ and $Y \cap Z = \{\mathbf{0}\}$ and so noting that:

- We are allowed to assume that $X = Y + Z$.
- Consider any vectors $y, y' \in Y$ and any vectors $z, z' \in Z$ which are such that

$$\mathbf{x} = \mathbf{y} + \mathbf{z} = \mathbf{y}' + \mathbf{z}'.$$

  On re-arranging, this gives

$$\mathbf{y} - \mathbf{y}' = \mathbf{z}' - \mathbf{z},$$

  where $\mathbf{y} - \mathbf{y}' \in Y$ and $\mathbf{z}' - \mathbf{z} \in Z$ as Y and Z are subspaces of X. So, as $\mathbf{y} - \mathbf{y}' = \mathbf{z}' - \mathbf{z}$, the vector given by these differences must be in $Y \cap Z = \{\mathbf{0}\}$, i.e. $\mathbf{y} - \mathbf{y}' = \mathbf{z}' - \mathbf{z} = \mathbf{0}$. Thus, we have $\mathbf{y} = \mathbf{y}'$ and $\mathbf{z} = \mathbf{z}'$, and so there is only one way of writing $\mathbf{x} \in X$ in terms of $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$.

we have $X = Y + Z$ and every $\mathbf{x} \in X$ can be written uniquely as $\mathbf{x} = \mathbf{y} + \mathbf{z}$ with $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$ (as required).

(b) Given that $X = Y \oplus Z$ (i.e. the subspaces X, Y and Z satisfy I or II as given above), we define the *projection* P from X onto Y parallel to Z to be a mapping $\mathsf{P} : X \to Y$ such that for any $\mathbf{x} \in X$ we have $\mathsf{P}\mathbf{x} = \mathbf{y}$ where $\mathbf{x} = \mathbf{y} + \mathbf{z}$ with $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$.

To show that P is idempotent, we note that:

- For any $\mathbf{y} \in Y \subseteq X$, we can write $\mathbf{y} = \mathbf{y} + \mathbf{0}$ and so $\mathsf{P}\mathbf{y} = \mathbf{y}$.
- So, for any $\mathbf{x} \in X$, we can write $\mathbf{x} = \mathbf{y} + \mathbf{z}$ with $\mathbf{y} \in Y$ and $\mathbf{z} \in Z$, and so we have

$$\mathsf{P}\mathbf{x} = \mathbf{y} \implies \mathsf{P}^2\mathbf{x} = \mathsf{P}\mathbf{y} \implies \mathsf{P}^2\mathbf{x} = \mathbf{y},$$

  since $\mathsf{P}\mathbf{y} = \mathbf{y}$. But, $\mathsf{P}\mathbf{x} = \mathbf{y}$, and so this gives us $\mathsf{P}^2\mathbf{x} = \mathsf{P}\mathbf{x}$ for any $\mathbf{x} \in X$. Thus, $\mathsf{P}^2 = \mathsf{P}$ and so P is idempotent.

(as required).

(c) We are given that A be a real matrix where $R(\mathsf{A})$ and $R(\mathsf{A})^\perp$ denote the range of A and the orthogonal complement of $R(\mathsf{A})$ respectively. Now, we let $\mathsf{P}\mathbf{s}$ be a projection of any vector $\mathbf{s}$ onto $R(\mathsf{A})$ parallel to $R(\mathsf{A})^\perp$, and this is clearly going to be the orthogonal projection of $\mathbf{s}$ onto $R(\mathsf{A})$. So, to show that this vector, i.e. $\mathsf{P}\mathbf{s}$, is the vector in $R(\mathsf{A})$ closest to $\mathbf{s}$, we note that:

- $\mathsf{P}\mathbf{s}$ is clearly in $R(\mathsf{A})$ since P projects vectors onto $R(\mathsf{A})$.
- Taking $\mathbf{r}$ to be any vector in $R(\mathsf{A})$, we can construct the vector $\mathbf{u} = \mathsf{P}\mathbf{s} - \mathbf{r}$ and this is in $R(\mathsf{A})$ too (since $R(\mathsf{A})$ is a vector space). Now, the 'distance' between the vectors $\mathbf{s}$ and $\mathbf{r}$ is given by

$$\|\mathbf{s} - \mathbf{r}\|^2 = \|\mathbf{s} - \mathsf{P}\mathbf{s} + \mathbf{u}\|^2 = \|(\mathsf{I} - \mathsf{P})\mathbf{s} + \mathbf{u}\|^2,$$

So, as $\mathbf{u} \in R(\mathsf{A})$ and $(\mathsf{I} - \mathsf{P})\mathbf{s} \in R(\mathsf{A})^{\perp}$ we can apply the Generalised Theorem of Pythagoras, to get

$$\|\mathbf{s} - \mathbf{r}\|^2 = \|(\mathsf{I} - \mathsf{P})\mathbf{s}\|^2 + \|\mathbf{u}\|^2,$$

and this quantity is minimised when $\mathbf{u} = \mathbf{0}$. Thus, the vector in $R(\mathsf{A})$ closest to $\mathbf{s}$ is given by $\mathbf{u} = \mathbf{0}$, i.e. $\mathbf{r} = \mathsf{P}\mathbf{s}$.

(as required).

(d) Hence, or otherwise, we are asked to find the least squares estimate of the parameters m and c when x and y are related by the expression $y = mx + c$ and we are given the data:

x	1	0	-1
y	3	-1	1

To start with, we note that the relationship that the parameters that m and c should satisfy gives us a set of three [inconsistent] equations, i.e.

$$\begin{aligned} m + c &= 3 \\ c &= -1 \\ -m + c &= 1 \end{aligned}$$

and we can write these in the form $\mathsf{A}\mathbf{x} = \mathbf{b}$ using

$$\mathsf{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} m \\ c \end{bmatrix} \quad \text{and} \quad \mathbf{b} = \begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix}.$$

Now, the 'hence or otherwise' indicates that there are [at least] two methods and so, let's look at them both:

'Hence' The least squares estimate of these parameters are the values of m and c which minimise the quantity $\|\mathsf{A}\mathbf{x} - \mathbf{b}\|$. But, since $\mathsf{A}\mathbf{x}$ is just a vector in $R(\mathsf{A})$, we have

$$R(\mathsf{A}) = \text{Lin}\left\{ \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} \right\},$$

and so, the vector we seek will be of the form

$$m \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + c \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}.$$

Indeed, we seek the values of m and c which make this vector closest to the vector $[3, -1, 1]^t$. But, by inspection we can see that

$$R(\mathsf{A})^{\perp} = \text{Lin}\left\{ \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix} \right\},$$

(since the vector $[1, -2, 1]^t$ is orthogonal to both $[1, 0, -1]^t$ and $[1, 1, 1]^t$) and so, by part **(c)**, as

$$\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}}_{\in R(\mathsf{A})} + \underbrace{\begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}}_{\in R(\mathsf{A})^{\perp}},$$

we can see that orthogonally projecting the vector $[3, -1, 1]^t$ onto $R(\mathsf{A})$ using P yields

$$\mathsf{P}\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix},$$

i.e. $\mathbf{x}^* = [m^*, c^*]^t = [1, 1]^t$.

'Otherwise' We know from lectures that a least squares solution to the matrix equation $\mathsf{A}\mathbf{x} = \mathbf{b}$ is given by $\mathbf{x}^* = (\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathbf{b}$ and so, since

$$\mathsf{A}^t\mathsf{A} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 0 & 1 \\ -1 & 1 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix} \implies (\mathsf{A}^t\mathsf{A})^{-1} = \frac{1}{6}\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix},$$

and

$$\mathsf{A}^t\mathbf{b} = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 1 & 1 \end{bmatrix}\begin{bmatrix} 3 \\ -1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

we have

$$\begin{bmatrix} m^* \\ c^* \end{bmatrix} = \mathbf{x}^* = (\mathsf{A}^t\mathsf{A})^{-1}\mathsf{A}^t\mathbf{b} = \frac{1}{6}\begin{bmatrix} 3 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

as the least squares estimate for m and c.

Question 4.

(a) A complex matrix A is called *anti-Hermitian* if $\mathsf{A}^\dagger = -\mathsf{A}$ (where $\mathsf{A}^\dagger$ denotes the complex conjugate transpose of A). We are asked to show that:

- The non-zero eigenvalues of an anti-Hermitian matrix are all purely imaginary.

  Let λ be a non-zero eigenvalue of the anti-Hermitian matrix A with corresponding eigenvector $\mathbf{x}$, i.e. $\mathsf{A}\mathbf{x} = \lambda\mathbf{x}$. Multiplying through by $\mathbf{x}^\dagger$, we get

  $$\mathbf{x}^\dagger \mathsf{A}\mathbf{x} = \lambda \mathbf{x}^\dagger \mathbf{x}, \qquad \text{(i)}$$

  and taking the complex conjugate transpose of this yields,

  $$\mathbf{x}^\dagger \mathsf{A}^\dagger \mathbf{x} = \lambda^* \mathbf{x}^\dagger \mathbf{x} \implies -\mathbf{x}^\dagger \mathsf{A}\mathbf{x} = \lambda^* \mathbf{x}^\dagger \mathbf{x}, \qquad \text{(ii)}$$

  since A is anti-Hermitian. Now, adding (i) and (ii) together we get,

  $$(\lambda + \lambda^*)\mathbf{x}^\dagger \mathbf{x} = \mathbf{0},$$

  and so as $\mathbf{x}^\dagger \mathbf{x} = \|\mathbf{x}\|^2 \neq 0$ (since $\mathbf{x}$ is an eigenvector) we have $\lambda = -\lambda^*$. Thus, since $\lambda \neq 0$, λ is purely imaginary (as required).

- The eigenvectors of an anti-Hermitian matrix corresponding to distinct eigenvalues are orthogonal.

  Let A be an anti-Hermitian matrix and let $\mathbf{x}$ and $\mathbf{y}$ be eigenvectors corresponding to the distinct eigenvalues λ and μ respectively of A. (Without loss of generality, we assume that $\mu \neq 0$.) Thus, we have

  $$\mathsf{A}\mathbf{x} = \lambda\mathbf{x} \text{ and } \mathsf{A}\mathbf{y} = \mu\mathbf{y},$$

  which on multiplying both sides of these expressions by $\mathbf{y}^\dagger$ and $\mathbf{x}^\dagger$ respectively yields

  $$\mathbf{y}^\dagger \mathsf{A}\mathbf{x} = \lambda \mathbf{y}^\dagger \mathbf{x} \text{ and } \mathbf{x}^\dagger \mathsf{A}\mathbf{y} = \mu \mathbf{x}^\dagger \mathbf{y}.$$

  Now, taking the complex conjugate transpose of the latter expression gives

  $$\mathbf{y}^\dagger \mathsf{A}^\dagger \mathbf{x} = \mu^* \mathbf{y}^\dagger \mathbf{x} \implies -\mathbf{y}^\dagger \mathsf{A}\mathbf{x} = \mu^* \mathbf{y}^\dagger \mathbf{x},$$

  since A is anti-Hermitian. So, equating our two expressions for $\mathbf{y}^\dagger \mathsf{A}\mathbf{x}$ we get

  $$(\lambda + \mu^*)\mathbf{y}^\dagger \mathbf{x} = \mathbf{0} \implies (\lambda - \mu)\mathbf{y}^\dagger \mathbf{x} = \mathbf{0},$$

  since μ is non-zero, and hence purely imaginary. Thus, we have

  $$\mathbf{y}^\dagger \mathbf{x} = \mathbf{0} \implies \langle \mathbf{y}, \mathbf{x} \rangle^* = 0 \implies \langle \mathbf{x}, \mathbf{y} \rangle = 0,$$

  and so the eigenvectors $\mathbf{x}$ and $\mathbf{y}$ are orthogonal (as required).

(b) A complex matrix A is normal if $\mathsf{A}^\dagger \mathsf{A} = \mathsf{A}\mathsf{A}^\dagger$. Clearly, anti-Hermitian matrices are normal since

$$\mathsf{A}^\dagger \mathsf{A} = -\mathsf{A}^2 = \mathsf{A}\mathsf{A}^\dagger.$$

(c) To find the spectral decomposition of the matrix

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

where $i^2 = -1$, we start by noting that this matrix is anti-Hermitian and hence it is normal by part **(b)**. So, to find the eigenvalues of A, we solve the determinant equation given by

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

Factorising this gives $(\lambda - i)(\lambda - 2i)(\lambda + i) = 0$ and so, the eigenvalues are $\lambda = \pm i, 2i$. We now need to find an orthonormal set of eigenvectors corresponding to these eigenvalues, i.e.

- For $\lambda = i$, a corresponding eigenvector $[x, y, z]^t$ is given by

$$\begin{bmatrix} 0 & 1 & 0 \\ -1 & -i & -1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies \begin{aligned} y &= 0 \\ x + iy + z &= 0 \\ y &= 0 \end{aligned} \implies \begin{aligned} y &= 0 \\ x + z &= 0 \end{aligned},$$

i.e. $x = -z$ for $z \in \mathbb{R}$ and $y = 0$. Thus, a corresponding eigenvector is $[-1, 0, 1]^t$.

- For $\lambda = -i$, a corresponding eigenvector $[x, y, z]^t$ is given by

$$\begin{bmatrix} 2i & 1 & 0 \\ -1 & i & -1 \\ 0 & 1 & 2i \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies \begin{aligned} 2ix + y &= 0 \\ x - iy + z &= 0 \\ y + 2iz &= 0 \end{aligned} \implies \begin{aligned} 2ix + y &= 0 \\ 2ix + 2y + 2iz &= 0 \\ y + 2iz &= 0 \end{aligned},$$

i.e. $y = -2ix$ and $z = x$ for $x \in \mathbb{R}$. Thus, a corresponding eigenvector is $[1, -2i, 1]^t$.

- For $\lambda = 2i$, a corresponding eigenvector $[x, y, z]^t$ is given by

$$\begin{bmatrix} -i & 1 & 0 \\ -1 & -2i & -1 \\ 0 & 1 & -i \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies \begin{aligned} ix - y &= 0 \\ x + 2iy + z &= 0 \\ y - iz &= 0 \end{aligned} \implies \begin{aligned} ix - y &= 0 \\ ix - 2y + iz &= 0 \\ y - iz &= 0 \end{aligned},$$

i.e. $y = ix$ and $z = x$ for $x \in \mathbb{R}$. Thus, a corresponding eigenvector is $[1, i, 1]^t$.

and so an orthonormal set of eigenvectors is

$$\left\{ \frac{1}{\sqrt{2}} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}, \frac{1}{\sqrt{6}} \begin{bmatrix} 1 \\ -2i \\ 1 \end{bmatrix}, \frac{1}{\sqrt{3}} \begin{bmatrix} 1 \\ i \\ 1 \end{bmatrix} \right\},$$

(since the eigenvectors are already mutually orthogonal). Thus,

$$\mathsf{A} = \frac{i}{2} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} - \frac{i}{6} \begin{bmatrix} 1 \\ -2i \\ 1 \end{bmatrix} \begin{bmatrix} 1 & 2i & 1 \end{bmatrix} + \frac{2i}{3} \begin{bmatrix} 1 \\ i \\ 1 \end{bmatrix} \begin{bmatrix} 1 & -i & 1 \end{bmatrix}$$

$$= \frac{i}{2} \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} - \frac{i}{6} \begin{bmatrix} 1 & 2i & 1 \\ -2i & 4 & -2i \\ 1 & 2i & 1 \end{bmatrix} + \frac{2i}{3} \begin{bmatrix} 1 & -i & 1 \\ i & 1 & i \\ 1 & -i & 1 \end{bmatrix}.$$

is the spectral decomposition of A.

Question 5.

(a) To find the general solution to the system of coupled linear differential equations given by

$$\begin{aligned}\dot{y}_1 &= -y_1 - 2y_2\\ \dot{y}_2 &= -2y_1 + 2y_3\\ \dot{y}_3 &= 2y_2 + y_3\end{aligned}$$

we re-write them as $\dot{\mathbf{y}} = \mathsf{A}\mathbf{y}$ where $\mathbf{y} = [y_1, y_2, y_3]^t$ and

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

and diagonalise this matrix. So, to find the eigenvalues of A, we solve the determinant equation

$$\begin{vmatrix} -1-\lambda & -2 & 0 \\ -2 & -\lambda & 2 \\ 0 & 2 & 1-\lambda \end{vmatrix} = 0 \implies -(1+\lambda)[-\lambda(1-\lambda) - 4] - 4(1-\lambda) = 0,$$

which on simplifying yields

$$\lambda(1-\lambda^2) + 4(1+\lambda) - 4(1-\lambda) = 0 \implies \lambda(\lambda^2 - 9) = 0,$$

and so the eigenvalues are $\lambda = 0, \pm 3$. Then, to find the eigenvectors, we note that:

- For $\lambda = 0$, a corresponding eigenvector $[x, y, z]^t$ is given by

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

  i.e. $x = z$ and $y = -z/2$ for $z \in \mathbb{R}$. Thus, a corresponding eigenvector is $[1, -1/2, 1]^t$ or $[2, -1, 2]^t$.

- For $\lambda = 3$, a corresponding eigenvector $[x, y, z]^t$ is given by

$$\begin{bmatrix} -4 & -2 & 0 \\ -2 & -3 & 2 \\ 0 & 2 & -2 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies \begin{aligned} 2x + y &= 0 \\ 2x + 3y - 2z &= 0 \\ 2y - 2z &= 0 \end{aligned} \implies \begin{aligned} 2x + y &= 0 \\ y - z &= 0 \end{aligned},$$

  i.e. $x = -z/2$ and $y = z$ for $z \in \mathbb{R}$. Thus, a corresponding eigenvector is $[-1/2, 1, 1]^t$ or $[-1, 2, 2]^t$.

- For $\lambda = -3$, a corresponding eigenvector $[x, y, z]^t$ is given by

$$\begin{bmatrix} 2 & -2 & 0 \\ -2 & 3 & 2 \\ 0 & 2 & 4 \end{bmatrix}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \mathbf{0} \implies \begin{aligned} x - y &= 0 \\ -2x + 3y + 2z &= 0 \\ y + 2z &= 0 \end{aligned} \implies \begin{aligned} x - y &= 0 \\ y + 2z &= 0 \end{aligned},$$

  i.e. $x = y$ and $z = -y/2$ for $y \in \mathbb{R}$. Thus, a corresponding eigenvector is $[1, 1, -1/2]^t$ or $[2, 2, -1]^t$

Thus, we have $\mathsf{P}^{-1}\mathsf{A}\mathsf{P} = \mathsf{D}$ where

$$\mathsf{P} = \begin{bmatrix} 2 & -1 & -2 \\ -1 & 2 & -2 \\ 2 & 2 & 1 \end{bmatrix} \quad \text{and} \quad \mathsf{D} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{bmatrix}.$$

So, given our set of coupled linear differential equations, we can use this to write

$$\dot{\mathbf{y}} = \mathsf{A}\mathbf{y} \implies \dot{\mathbf{y}} = \mathsf{P}\mathsf{D}\mathsf{P}^{-1}\mathbf{y} \implies \dot{\mathbf{z}} = \mathsf{D}\mathbf{z},$$

where $\mathbf{z} = \mathsf{P}^{-1}\mathbf{y}$. Thus, we proceed by solving the uncoupled linear differential equations given by

$$\begin{bmatrix} \dot{z}_1 \\ \dot{z}_2 \\ \dot{z}_3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \implies \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} \begin{bmatrix} A \\ Be^{3t} \\ Ce^{-3t} \end{bmatrix},$$

for arbitrary constants A, B and C. Consequently, using the fact that $\mathbf{y} = \mathsf{P}\mathbf{z}$, we can see that

$$\mathbf{y}(t) = \begin{bmatrix} 2 & -1 & -2 \\ -1 & 2 & -2 \\ 2 & 2 & 1 \end{bmatrix} \begin{bmatrix} A \\ Be^{3t} \\ Ce^{-3t} \end{bmatrix} = \begin{bmatrix} 2A - Be^{3t} - 2Ce^{-3t} \\ -A + 2Be^{3t} - 2Ce^{-3t} \\ 2A + 2Be^{3t} + Ce^{-3t} \end{bmatrix},$$

is the required general solution to $\dot{\mathbf{y}} = \mathsf{A}\mathbf{y}$.

The initial conditions for this system are restricted so that they satisfy the equation

$$y_1(0) - 2y_2(0) - 2y_3(0) = 0.$$

and so noting that our general solution gives

$$\mathbf{y}(t) = \begin{bmatrix} y_1(0) \\ y_2(0) \\ y_3(0) \end{bmatrix} = \begin{bmatrix} 2A - B - 2C \\ -A + 2B - 2C \\ 2A + 2B + C \end{bmatrix},$$

we substitute these expressions for $y_1(0)$, $y_2(0)$ and $y_3(0)$ into the restriction to get

$$(2A - B - 2C) - 2(-A + 2B - 2C) - 2(2A + 2B + C) = 0 \implies B = 0.$$

Thus, substituting this into our general solution, we find that

$$\mathbf{y}(t) = \begin{bmatrix} 2A - 2Ce^{-3t} \\ -A - 2Ce^{-3t} \\ 2A + Ce^{-3t} \end{bmatrix}.$$

is the general solution in the presence of this restriction.

If the initial conditions are further restricted by stipulating that

$$\mathbf{y}(0) = \begin{bmatrix} -2 - 2\gamma \\ 1 - 2\gamma \\ -2 + \gamma \end{bmatrix},$$

for some constant γ, we can see that at $t = 0$, we have

$$\begin{bmatrix} -2 - 2\gamma \\ 1 - 2\gamma \\ -2 + \gamma \end{bmatrix} = \begin{bmatrix} 2A - 2C \\ -A - 2C \\ 2A + C \end{bmatrix} \implies (A+1) \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix} + (C - \gamma) \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix} = \mathbf{0}.$$

But, the vectors $[2, -1, 2]^t$ and $[-2, -2, 1]^t$ are clearly linearly independent and so we must have $A = -1$ and $C = \gamma$. Thus, in the presence of this new restriction, the general solution becomes

$$\mathbf{y}(t) = \begin{bmatrix} -2 - 2\gamma e^{-3t} \\ 1 - 2\gamma e^{-3t} \\ -2 + \gamma e^{-3t} \end{bmatrix} = \begin{bmatrix} -2 \\ 1 \\ -2 \end{bmatrix} + \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix} e^{-3t},$$

and so, in the long-term, we can see that $\mathbf{y}(t) \to [-2, 1, -2]^t$.

(b) The steady states of the coupled non-linear differential equations

$$\dot{y}_1 = y_1(y_1 - 3y_2 - 4)$$
$$\dot{y}_2 = y_2(y_2 - 4y_1 - 6)$$

are given by the solutions of the simultaneous equations

$$\begin{aligned} y_1(y_1 - 3y_2 - 4) &= 0 \\ y_2(y_2 - 4y_1 - 6) &= 0 \end{aligned}$$

i.e. by $(y_1, y_2) = (0,0), (0,6), (4,0)$ and $(-2,-2)$.
To assess the stability of the steady state given by $(y_1, y_2) = (-2,-2)$, we evaluate the Jacobian for this system at $(-2,-2)$, i.e.

$$\mathsf{DF}[(y_1, y_2)] = \begin{bmatrix} 2y_1 - 3y_2 - 4 & -3y_1 \\ -4y_2 & 2y_2 - 4y_1 - 6 \end{bmatrix} \implies \mathsf{DF}[(-2,-2)] = \begin{bmatrix} -2 & 6 \\ 8 & -2 \end{bmatrix},$$

and find the eigenvalues of this matrix. So, solving

$$\begin{vmatrix} -2-\lambda & 6 \\ 8 & -2-\lambda \end{vmatrix} = 0 \implies (2+\lambda)^2 - 6 \times 8 = 0 \implies 2 + \lambda = \pm 4\sqrt{3},$$

we find that the eigenvalues are $\lambda = -2 \pm 4\sqrt{3}$. According to a result given in the lectures, since these are not both both real and negative (note that $4\sqrt{3} > 2$), the steady state $(y_1, y_2) = (-2,-2)$ is not asymptotically stable.

Question 6

(a) To test the set of functions $\{1, x, x^2\}$ for linear independence we calculate the Wronskian as instructed, i.e.

$$W(x) = \begin{vmatrix} f_1 & f_2 & f_3 \\ f_1' & f_2' & f_3' \\ f_1'' & f_2'' & f_3'' \end{vmatrix} = \begin{vmatrix} 1 & x & x^2 \\ 0 & 1 & 2x \\ 0 & 0 & 2 \end{vmatrix} = 2,$$

and as $W(x) \neq 0$ for all $x \in \mathbb{R}$, this set of functions is linearly independent (as required).

(b) We consider the inner product space formed by the vector space $\mathbb{P}_3^{[-2,2]}$ and the inner product

$$\langle f(x), g(x)\rangle = f(-2)g(-2) + f(-1)g(-1) + f(0)g(0) + f(1)g(1) + f(2)g(2).$$

To find an orthonormal basis of the space $\mathrm{Lin}\{1, x, x^2\}$, we use the Gram-Schmidt procedure:

- We start with the vector 1, and note that $\|1\|^2 = \langle 1, 1\rangle = 1 + 1 + 1 + 1 + 1 = 5$. Consequently, we set $\mathbf{e}_1 = 1/\sqrt{5}$.

- We need to find a vector $\mathbf{u}_2$ where

$$\mathbf{u}_2 = x - \left\langle x, \frac{1}{\sqrt{5}} \right\rangle \frac{1}{\sqrt{5}} = x - \frac{\langle x, 1\rangle}{5},$$

  But, as

$$\langle x, 1\rangle = (-2)(1) + (-1)(1) + (0)(1) + (1)(1) + (2)(1) = 0,$$

  we have $\mathbf{u}_2 = x$. Now, as

$$\|x\|^2 = \langle x, x\rangle = (-2)(-2) + (-1)(-1) + (0)(0) + (1)(1) + (2)(2) = 10,$$

  we set $\mathbf{e}_2 = x/\sqrt{10}$.

- Lastly, we need to find a vector $\mathbf{u}_3$ where

$$\mathbf{u}_3 = x^2 - \left\langle x^2, \frac{x}{\sqrt{10}} \right\rangle \frac{1}{\sqrt{10}} - \left\langle x^2, \frac{1}{\sqrt{5}} \right\rangle \frac{1}{\sqrt{5}} = x^2 - \frac{\langle x^2, x\rangle}{10} - \frac{\langle x^2, 1\rangle}{5},$$

  But, as

$$\langle x^2, x\rangle = (4)(-2) + (1)(-1) + (0)(0) + (1)(1) + (4)(2) = 0,$$

  and,

$$\langle x^2, 1\rangle = (4)(1) + (1)(1) + (0)(1) + (1)(1) + (4)(1) = 10,$$

  we have $\mathbf{u}_3 = x^2 - 2$. Now, as

$$\|x^2 - 2\|^2 = \langle x^2 - 2, x^2 - 2\rangle = (2)(2) + (-1)(-1) + (-2)(-2) + (-1)(-1) + (2)(2) = 14,$$

  we set $\mathbf{e}_3 = (x^2 - 2)/\sqrt{14}$.

Consequently, the set

$$\left\{ \frac{1}{\sqrt{5}}, \frac{x}{\sqrt{10}}, \frac{x^2 - 2}{\sqrt{14}} \right\},$$

is an orthonormal basis for the space $\mathrm{Lin}\{1, x, x^2\}$.

(c) We are given that $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k\}$ is an orthonormal basis of a subspace S of an inner product space V. Extending this to an orthonormal basis $\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_k, \mathbf{e}_{k+1}, \ldots, \mathbf{e}_n\}$ of V, we note that for any vector $\mathbf{x} \in V$,

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{e}_i.$$

Now, for any j (where $1 \leq j \leq n$) we have

$$\langle \mathbf{x}, \mathbf{e}_j \rangle = \left\langle \sum_{i=1}^{n} \alpha_i \mathbf{e}_i, \mathbf{e}_j \right\rangle = \sum_{i=1}^{n} \alpha_i \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \alpha_j,$$

since we are using an orthonormal basis. Thus, we can write

$$\mathbf{x} = \sum_{i=1}^{n} \langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{e}_i = \underbrace{\sum_{i=1}^{k} \langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{e}_i}_{\text{in } S} + \underbrace{\sum_{i=k+1}^{n} \langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{e}_i}_{\text{in } S^\perp}.$$

and so, the orthogonal projection of V onto S [parallel to $S^\perp$] is given by

$$\mathsf{P}\mathbf{x} = \sum_{i=1}^{k} \langle \mathbf{x}, \mathbf{e}_i \rangle \mathbf{e}_i,$$

for any $\mathbf{x} \in V$ (as required). Further, since $\|\mathbf{x} - \mathbf{y}\|$ measures the 'distance' between $\mathbf{x}$ and a vector $\mathbf{y} \in S$, this quantity is minimised since orthogonal projections give the vector $\mathsf{P}\mathbf{x} \in S$ which is 'closest' to $\mathbf{x}$.

(d) Using the results in **(c)** it should be clear that a least squares approximation to x^3 in $\text{Lin}\{1, x, x^2\}$ will be given by $\mathsf{P}x^3$. So, using the inner product in **(b)** and the orthonormal basis for $\text{Lin}\{1, x, x^2\}$ which we found there, we have:

$$\begin{aligned} \mathsf{P}x^3 &= \langle x^3, \mathbf{e}_1 \rangle \mathbf{e}_1 + \langle x^3, \mathbf{e}_2 \rangle \mathbf{e}_2 + \langle x^3, \mathbf{e}_3 \rangle \mathbf{e}_3 \\ &= \frac{\langle x^3, 1 \rangle}{5} + \frac{\langle x^3, x \rangle}{10} x + \frac{\langle x^3, x^2 - 2 \rangle}{14} (x^2 - 2) \\ \therefore\ \mathsf{P}x^3 &= \frac{34}{10} x. \end{aligned}$$

since,

- $\langle x^3, 1 \rangle = (-8)(1) + (-1)(1) + (0)(1) + (1)(1) + (8)(1) = 0,$
- $\langle x^3, x \rangle = (-8)(-2) + (-1)(-1) + (0)(0) + (1)(1) + (8)(2) = 34,$
- $\langle x^3, x^2 - 2 \rangle = (-8)(2) + (-1)(-1) + (0)(0) + (1)(-1) + (8)(2) = 0.$

Thus, our least squares approximation to x^3 is $\frac{17}{5}x$.