1. (a) Suppose $x \in N(A^T A)$, i.e. $A^T A x = 0$. Thus $A x \in N(A^T) = R(A)^\perp$. But also $A x \in R(A)$, so since $R(A) \cap R(A)^\perp = \{0\}$, we conclude that $A x = 0$, i.e. $x \in N(A)$. Therefore $N(A^T A) \subseteq N(A)$.

Conversely, if $x \in N(A)$. Then $A x = 0$ and hence $A^T A x = 0$, so $x \in N(A^T A)$. Therefore $N(A) \subseteq N(A^T A)$.

We conclude that $N(A^T A) = N(A)$.

(b) If $A A^T A x = 0$ then as in (a) we have $A^T A x \in N(A) \cap R(A) = N(A) \cap N(A)^\perp = \{0\}$, and hence $A x = 0$. By part (a) above, this implies that $A x = 0$. Thus we have $N(A A^T A) \subseteq N(A)$. Since also $N(A) \subseteq N(A A^T A)$ (again as in part (a)), we conclude that $N(A A^T A) = N(A)$.

(c) The matrices $A A^T A$, $A^T A$, $A$ have $n$ columns and the same nullity, say $k$, by parts (a) and (b). Thus by the Rank-Nullity theorem they all have rank equal to $n - k$.

2. (a) We have

$$\langle p, p \rangle = \int_{-1}^{1} p^2(x) \, dx \geq 0$$

Equality occurs when $\int_{-1}^{1} p^2(x) \, dx = 0$, which implies that $p(x) = 0$ since $p$ is a polynomial (or more generally a continuous function).

For symmetry, $\langle p, q \rangle = \int_{-1}^{1} p(x)q(x) \, dx = \int_{-1}^{1} q(x)p(x) \, dx = \langle q, p \rangle$.

Finally, we have

$$\langle \alpha p + \beta q, r \rangle = \int_{-1}^{1} (\alpha p(x) + \beta q(x)) r(x) \, dx =$$

$$= \alpha \int_{-1}^{1} p(x)r(x) \, dx + \beta \int_{-1}^{1} q(x)r(x) \, dx = \alpha \langle p, r \rangle + \beta \langle q, r \rangle$$

(b) Since $\langle 1, x \rangle = \int_{-1}^{1} x \, dx = \left[ \frac{x^2}{2} \right]_{-1}^{1} = 0$, we see that $1, x$ are orthogonal.

(c) We have

$$||1|| = \sqrt{\langle 1, 1 \rangle} = \sqrt{\int_{-1}^{1} 1 \, dx} = \sqrt{2}$$

$$||1 + x|| = \sqrt{\langle 1 + x, 1 + x \rangle} = \sqrt{\int_{-1}^{1} (1 + x)^2 \, dx} = \sqrt{\frac{8}{3}}$$

$$\langle 1, 1 + x \rangle = \int_{-1}^{1} (1 + x) \, dx = 2$$

Hence

$$\cos \theta = \frac{\langle 1, 1 + x \rangle}{||1|| \cdot ||1 + x||} = \frac{\sqrt{3}}{2} \Rightarrow \theta = \frac{\pi}{6}$$
3. (a) We have
\[ \langle x, x \rangle = 4x_1^2 + 9x_2^2 \geq 0 \]
with equality if and only if \( x_1 = x_2 = 0 \iff x = 0 \).

For symmetry, \( \langle x, y \rangle = 4x_1y_1 + 9x_2y_2 = 4y_1x_1 + 9y_2x_2 = \langle y, x \rangle \).

Finally, we have
\[ \langle \alpha x + \beta y, z \rangle = 4(\alpha x_1 + \beta y_1)z_1 + 9(\alpha x_2 + \beta y_2)z_2 = \]
\[ = \alpha(4x_1z_1 + 9x_2z_2) + \beta(4y_1z_1 + 9y_2z_2) = \]
\[ = \langle x, z \rangle + \beta \langle y, z \rangle \]

(b) The scalar projection is
\[ \alpha = \frac{\langle v_2, v_1 \rangle}{\|v_1\|} = \frac{8}{2} = 4 \]

The vector projection is
\[ p_1 = \frac{1}{\|v_1\|} v_1 = 2v_1 = \left( \begin{array}{c} 2 \\ 0 \end{array} \right) \]

(c) Normalizing \( v_1 \), we get \( u_1 = \frac{1}{\|v_1\|} v_1 = \left( \begin{array}{c} \frac{1}{2} \\ 0 \end{array} \right) \). By (b), we know that the projection of \( v_2 \) onto \( u_1 \) is \( p_1 \). Thus normalizing \( v_2 - p_1 = \left( \begin{array}{c} 0 \\ 2 \end{array} \right) \), we get \( u_2 = \frac{1}{6}(v_2 - p_1) = \left( \begin{array}{c} 0 \\ \frac{1}{3} \end{array} \right) \).

4. Clearly \( \|x\|_\infty = \max_{1 \leq i \leq n} |x_i| \geq 0 \) with equality if and only if \( |x_i| = 0 \Rightarrow x_i = 0 \) for all \( i \), i.e. \( x = 0 \).

Now
\[ \|\alpha x\|_\infty = \max_{1 \leq i \leq n} |\alpha x_i| = \max_{1 \leq i \leq n} |\alpha||x_i| = |\alpha| \max_{1 \leq i \leq n} |x_i| = |\alpha||x||_\infty \]

Finally, for the triangle inequality
\[ \|x + y\| = \max_{1 \leq i \leq n} |x_i + y_i| \leq \max_{1 \leq i \leq n} (|x_i| + |y_i|) \leq \max_{1 \leq i \leq n} |x_i| + \max_{1 \leq i \leq n} |y_i| = \|x\| + \|y\| \]

since \( |x_i + y_i| \leq |x_i| + |y_i| \) for all \( i \).

5. (a) For \( x = \left( \begin{array}{c} 0 \\ 1 \end{array} \right) \) we have \( \|x\| = -3 < 0 \) so this is not a norm.

(b) Since \( (1 + x^2)|f(x)| \geq 0 \) for \( 0 \leq x \leq 1 \), we have
\[ ||f|| = \int_0^1 (1 + x^2)|f(x)| \, dx \geq 0 \]

with equality if and only if \( (1 + x^2)|f(x)| = 0 \iff f(x) = 0 \) for all \( x \in [0, 1] \), since \( f \) is continuous.

Now
\[ ||\alpha f|| = \int_0^1 (1 + x^2)|\alpha f(x)| \, dx = \int_0^1 (1 + x^2)|\alpha||f(x)| \, dx = \]
\[ = \alpha \int_0^1 (1 + x^2)|f(x)| \, dx = |\alpha||f|| \]
Finally, for the triangle inequality, since \( |f(x) + g(x)| \leq |f(x)| + |g(x)| \), we have
\[
\|f + g\| = \int_0^1 (1 + x^2)|f(x) + g(x)| \, dx 
\leq \int_0^1 (1 + x^2) \left( |f(x)| + |g(x)| \right) \, dx = \\
= \int_0^1 (1 + x^2)|f(x)| \, dx + \int_0^1 (1 + x^2)|g(x)| \, dx = \|f\| + \|g\|
\]

(c) Clearly \( \|p\| = |a_0| + 2|a_1| + 3|a_2| \geq 0 \) with equality if and only if \( a_0 = a_1 = a_2 = 0 \), i.e. \( p(x) = 0 \).

Now, since \( \alpha p(x) = \alpha a_2 x^2 + \alpha a_1 x + \alpha a_0 \),
\[
\|\alpha p\| = |\alpha a_0| + 2|\alpha a_1| + 3|\alpha a_2| = |\alpha| \left( |a_0| + 2|a_1| + 3|a_2| \right) = |\alpha|\|p\|
\]

Finally, for the triangle inequality for two polynomials \( p(x) = a_2 x^2 + a_1 x + a_0 \) and \( q(x) = b_2 x^2 + b_1 x + b_0 \) we have
\[
\|p + q\| = |a_0 + b_0| + 2|a_1 + b_1| + 3|a_2 + b_2| \leq \\
\leq |a_0| + |b_0| + 2(|a_1 + b_1|) + 3(|a_2 + b_2|) = \\
= |a_0| + 2|a_1| + 3|a_2| + |b_0| + 2|b_1| + 3|b_2| = \\
= \|p\| + \|q\|
\]

6. (a) We have
\[
(A^2 + A + I)x = A^2 x + Ax + x = A(Ax) + \lambda x + x = \\
= \lambda Ax + (\lambda + 1)x = (\lambda^2 + \lambda + 1)x
\]
so \( \lambda^2 + \lambda + 1 \) is an eigenvalue of \( A^2 + A + I \).

(b) We have
\[
p_A(\lambda) = \det(A - \lambda I) = \det(A - \lambda I)^T = \det(A^T - \lambda I) = p_{A^T}(\lambda)
\]
Thus \( A \) and \( A^T \) have the same characteristic polynomial and the same eigenvalues.

(c) Take \( A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \). Then \( x = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \) is an eigenvector of \( A \) for its single eigenvalue \( \lambda = 1 \).

However, \( A^T x = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \) which is not a multiple of \( x \). So \( x \) is not an eigenvector of \( A^T \).

7. (a) By Parseval’s formula, we have
\[
\|x\|^2 = \langle x, u_1 \rangle^2 + \langle x, u_2 \rangle^2 + \langle x, u_3 \rangle^2 \Rightarrow \langle x, u_3 \rangle^2 = 25 - 9 - 16 = 0
\]
Thus \( \langle x, u_3 \rangle = 0 \).

(b) By the Cauchy-Schwarz inequality, we have
\[
\|x\| \cdot \|y\| \geq \|\langle x, y \rangle\| \Rightarrow \|y\| \geq \frac{|\langle x, y \rangle|}{\|x\|} = 1
\]

8. We are looking for a linear function \( l(x) = c_0 + c_1 x \) to fit the data. Thus we have the least squares system \( A c = b \) where
\[
A = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 1 \\ 3 \\ 3 \end{pmatrix}
\]
We have
\[
A^T A = \begin{pmatrix} 3 & 3 \\ 3 & 5 \end{pmatrix}, \quad A^T b = \begin{pmatrix} 7 \\ 9 \end{pmatrix}
\]
so the normal equations \( A^T A c = A^T b \) have solution
\[
c = (A^T A)^{-1} A^T b = \begin{pmatrix} 4 \\ 1 \end{pmatrix}
\]
The best linear least squares fit is thus given by the line \( l(x) = \frac{4}{3} + x \).

9. (a) \( x = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \in S^\perp \) if and only if
\[
\begin{pmatrix} 2 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix} x = 0 \iff 2x_1 = 0 \iff x_1 = 0
\]
\[
\begin{pmatrix} 0 & 1 & 1 \end{pmatrix} x = 0 \iff x_1 + x_2 + x_3 = 0 \iff x_2 = -x_3
\]
We conclude that
\[
S^\perp = \text{span} \left\{ \begin{pmatrix} 0 \\ -1 \\ 1 \end{pmatrix} \right\}
\]

(b) Let
\[
x_1 = \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix}, \quad x_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}
\]
Applying the Gram-Schmidt process to the basis \( \{x_1, x_2\} \), we have
\[
u_1 = \frac{1}{\|x_1\|} x_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}
\]
The projection of \( x_2 \) onto \( u_1 \) is
\[
p_1 = (x_2^T u_1) u_1 = u_1
\]
Since \( x_2 - p_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} \) we have
\[
u_2 = \frac{1}{\|x_2 - u_1\|} (x_2 - u_1) = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}
\]
\( \{u_1, u_2\} \) is an orthonormal basis of \( S \).

(c) The projection of \( e_3 \) onto \( S \) is
\[
p = (e_3^T u_1) u_1 + (e_3^T u_2) u_2 = \frac{1}{2} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}
\]

10. (a) Let \( x_1, x_2 \) be the eigenvectors corresponding to \( \lambda_1 = 1, \lambda_2 = 2 \) respectively. Then
\[
A^2 x_1 = A(A x_1) = A x_1 = x_1
\]
\[
A^2 x_2 = A(A x_2) = A(2 x_2) = 2 A x_2 = 4 x_2
\]
and thus \( \mu_1 = 1, \mu_2 = 4 \) are the eigenvalues of \( A^2 \). We then have
\[
\text{tr}(A^2) = \mu_1 + \mu_2 = 5
\]
\[
\det(A^2) = \mu_1 \mu_2 = 4
\]
(b) Since its leading coefficient is \((-1)^2 = 1\) and its roots are 1, 4, the characteristic polynomial is 
\[ p_{A^2}(\lambda) = (\lambda - 1)(\lambda - 4) = \lambda^2 - 5\lambda + 4 \]

11. (a) We have
\[ z \in N(A) \iff Az = 0 \iff xx^T z + yy^T z = 0 \iff (x^T z)x + (y^T z)y = 0 \]
Since \(x, y\) are linearly independent, \((x^T z)x + (y^T z)y = 0\) if and only if \(x^T z = y^T z = 0\). Equivalently, this means that \(z\) is orthogonal to both \(x, y\) and so \(z \in S^\perp\). We conclude that \(N(A) = S^\perp\).

(b) Since \(S\) has dimension 2, \(S^\perp = N(A)\) has dimension \(n - 2\). Thus, by the Rank-Nullity theorem, the rank of \(A\) is \(n - (n - 2) = 2\).

(c) Since \(\{x, y\}\) is orthonormal, we have
\[ Ax = xx^T x + yy^T x = x \]
\[ Ay = xx^T y + yy^T y = y \]
so \(x, y\) are eigenvectors of \(A\) with eigenvalue 1.