Assignment 2

1. Let \( y = a + bx \) be the equation of the least squares line of best fit for the following points \((x_i, y_i)\):

\[(0, 1), (1, 2), (2, 2), (3, 5), (4, 5).\]

Calculate \(a\) and \(b\).

Solution.

If all the points were on the line we would have \(y_i = a + bx_i\) for each \(i\), and so

\[
\begin{pmatrix}
1 & 0 \\
1 & 1 \\
1 & 2 \\
1 & 3 \\
1 & 4
\end{pmatrix}
\begin{pmatrix}
a \\
b
\end{pmatrix}
=
\begin{pmatrix}
1 \\
2 \\
2 \\
5 \\
5
\end{pmatrix}.
\]

If this were true, the 5-component column vector on the right-hand side would be in the column space of the \(5 \times 2\) matrix \(A\) on the left-hand side. In fact, it is not possible to find \(a\) and \(b\) to solve the equations exactly; instead \(a\) and \(b\) must be chosen so that

\[
a \begin{pmatrix}1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 4\end{pmatrix} + \frac{b}{5} \begin{pmatrix}1 \\ 2 \\ 3 \\ 4 \\ 5\end{pmatrix}
\]

is as close as possible to \(t(1, 2, 2, 5, 5)\). Thus \(a^t(1, 1, 1, 1) + b^t(2, 3, 4, 5)\) must be the projection of \(t(1, 2, 2, 5, 5)\) onto the column space of \(A\). According to the theory described in the lectures, to find \(a\) and \(b\) we must solve the system of linear equations

\[
^tA A \begin{pmatrix}a \\ b\end{pmatrix} = ^tA \begin{pmatrix}1 \\ 2 \\ 2 \\ 5 \\ 5\end{pmatrix}.
\]

We find that

\[
^tA A = \begin{pmatrix}1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix}1 \\ 0 \\ 1 \\ 2 \end{pmatrix} = \begin{pmatrix}5 & 10 \\ 10 & 30\end{pmatrix}
\]

and

\[
^tA \begin{pmatrix}1 \\ 2 \\ 2 \\ 5 \\ 5\end{pmatrix} = \begin{pmatrix}1 \\ 1 \end{pmatrix} \begin{pmatrix}1 \\ 0 \end{pmatrix} = \begin{pmatrix}15 \\ 41\end{pmatrix},
\]

and so it follows that

\[
\begin{pmatrix}a \\ b\end{pmatrix} = \begin{pmatrix}5 \\ 10 \\ 30 \\ 41\end{pmatrix}^{-1} \begin{pmatrix}15 \\ 41\end{pmatrix} = \frac{1}{50} \begin{pmatrix}30 & -10 \\ -10 & 5\end{pmatrix} \begin{pmatrix}15 \\ 41\end{pmatrix}
\]

\[
= \frac{1}{10} \begin{pmatrix}8 \\ 11\end{pmatrix} = \begin{pmatrix}0.8 \\ 1.1\end{pmatrix}
\]

Thus the line of best fit is \(y = 0.8 + 1.1x\).

The points and the line of best fit are shown in the diagram.

2. Let \(A\) be a square matrix which satisfies \(A^2 - 3A + 2I = 0\). Prove that if \(\lambda\) is an eigenvalue of \(A\) then \(\lambda\) must be 1 or 2.

Solution.

Let \(v\) be an eigenvector corresponding to the eigenvalue \(\lambda\). By definition \(v\) is nonzero, and \(Av = \lambda v\). Multiplying this equation by \(A\) gives \(A^2v = \lambda Av\), and since \(Av = \lambda v\) we deduce that \(A^2v = \lambda^2 v\). Now

\[(A^2 - 3A + 2I)v = A^2v - 3Av + 2v = (\lambda^2 + 3\lambda + 2)v,
\]

and since \(A^2 - 3A + 2I = 0\) it follows that \((\lambda^2 + 3\lambda + 2)v = 0\). Since \(v \neq 0\) this gives \(\lambda^2 + 3\lambda + 2 = 0\), whence \(\lambda\) is 1 or 2.
3. If \( f(x) = a_0 + a_1 x + a_2 x^2 + \cdots + a_d x^d \) is any polynomial, where the coefficients 
\( a_i \) are elements of a field \( F \), and if \( A \) is any square matrix over \( F \), we define \( f(A) \) to be the matrix 
\( a_0 I + a_1 A + a_2 A^2 + \cdots + a_d A^d \).

(i) Suppose that \( D \) is an \( n \times n \) diagonal matrix and \( f(x) = \det(D - xI) \) its characteristic 
polynomial. Show that \( f(D) \) is the zero matrix.

(ii) Use Part (i) to show that if \( A \) is any diagonalizable matrix and \( f(x) \) its characteristic 
polynomial then \( f(A) = 0 \).

(iii) In fact it is true for all square matrices \( A \) that if \( f(x) \) is the characteristic 
polynomial then \( f(A) = 0 \). Check this by direct calculation in the following cases:

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad \begin{pmatrix} \lambda & 0 \\ 0 & \lambda \end{pmatrix}.
\]

Solution.

(i) Suppose that the diagonal entries of \( D \) are \( d_1, d_2, \ldots, d_n \). Then

\[
f(x) = \det(D - xI) = \det \begin{pmatrix} d_1 - x & 0 & \cdots & 0 \\ 0 & d_2 - x & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n - x \end{pmatrix}
\]

and so we see that \( f(d_1) = f(d_2) = \cdots = f(d_n) = 0 \). Now for every integer \( k \geq 0 \) the matrix \( D^k \) is diagonal, and its diagonal entries are \( d_1^k, d_2^k, \ldots, d_n^k \). So for any polynomial \( p(x) = \alpha_0 + \alpha_1 x + \cdots + \alpha_r x^r \) we find that

\[
p(D) = \alpha_0 I + \alpha_1 D + \cdots + \alpha_r D^r
\]

\[
= (\alpha_0 \begin{pmatrix} a_0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_0 \end{pmatrix}) + (\alpha_1 \begin{pmatrix} a_1 d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_1 d_1 \end{pmatrix}) + \cdots
\]

\[
+ (\alpha_r \begin{pmatrix} \alpha_r d_1^r & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_r d_1^r \end{pmatrix})
\]

\[
= \begin{pmatrix} p(d_1) & 0 & \cdots & 0 \\ 0 & p(d_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & p(d_n) \end{pmatrix}.
\]

In particular, \( f(D) \) is a diagonal matrix, and its diagonal entries are, 
respectively, \( f(d_1), f(d_2), \ldots, f(d_n) \), which are all 0. So \( f(D) \) is the 
zero matrix.

Alternatively, from the formula for \( f(x) \) above we can see that

\[
f(D) = (d_1 I - X)(d_2 I - D) \cdots (d_n I - D)
\]

\[
= \begin{pmatrix} d_2 - d_1 & 0 & \cdots & 0 \\ 0 & d_1 - d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_1 - d_n \end{pmatrix}
\]

\[
= \begin{pmatrix} d_n - d_1 & 0 & \cdots & 0 \\ 0 & d_n - d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}.
\]

In the \( i \)-th factor the \( i \)-th diagonal entry is zero, and so when we compute 
the product we find that all the diagonal entries are zero.

(ii) If \( A \) is diagonalizable then there exists an invertible matrix \( P \) such 
that \( P^{-1} A P = D \), where the diagonal entries of \( D \) are the eigenvalues of \( A \). Thus if \( f(x) \) is the characteristic polynomial of \( A \) then

\[
f(x) = (x - d_1)(x - d_2) \cdots (x - d_n),
\]

where \( d_1, d_2, \ldots, d_n \) are the diagonal entries of \( D \). In particular, \( f(x) \) is also the characteristic polynomial of \( D \). (It is true in general—this was proved in lectures—that similar matrices 
have the same characteristic polynomial.)

Write \( f(x) = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \cdots + \alpha_n x^n \). Since \( A = P D P^{-1} \) it follows 
that for all positive integers \( r \),

\[
A^r = (P D P^{-1})^r = (P D P^{-1})(P D P^{-1}) \cdots (P D P^{-1}) = P D P^{-1}
\]

since the \( P \)'s in the middle of the expression cancel out. So

\[
f(A) = \alpha_0 I + \alpha_1 A + \alpha_2 A^2 + \cdots + \alpha_n A^n
\]

\[
= \alpha_0 I + \alpha_1 P D P^{-1} + \alpha_2 P D^2 P^{-1} + \cdots + \alpha_n P D^n P^{-1}
\]

\[
= P(\alpha_0 I + \alpha_1 D + \alpha_2 D^2 + \cdots + \alpha_n D^n) P^{-1} = P f(D) P^{-1} = 0
\]

since \( f(D) = 0 \) by Part prt (i).

(iii) In the second case the characteristic polynomial of \( A \) is \( f(x) = (x - \lambda)^3 \), 
and so the task is to show that \( (\lambda - A)^3 \) is the zero matrix. Now

\[
\lambda I - A = \begin{pmatrix} 0 & 0 & -1 \\ 0 & 0 & -1 \\ 0 & 0 & 0 \end{pmatrix},
\]

and it is trivial to check that the cube of this is 
zero.

In the other case \( f(x) = x^2 - (a + d)x + (a d - bc) \), and

\[
f(A) = \begin{pmatrix} a^2 + bc & ab + bd \\ ca + dc & bc + d^2 \end{pmatrix} - (a + d) \begin{pmatrix} a & b \\ c & d \end{pmatrix} + \begin{pmatrix} ad - bc & 0 \\ 0 & ad - bc \end{pmatrix}
\]

which is indeed zero.