UNIT 4 UNITARY SIMILARITY

Structure **Page No.**

4.1 INTRODUCTION

In the previous block, you studied about similarity. Here we discuss a particular kind of similarity, for matrices and linear operators.

To start with, in Sec. 4.2, you will see that a square matrix with complex entries is unitarily similar to an upper triangular matrix. Then, in Sec. 4.3, we discuss an analogue of this result for linear operators on a finite-dimensional inner product space. You have already seen in Unit 1 that if B and B' are bases of a vector space V, then for a linear operator T on V the matrices $[T]_B$ and $[T]_{B'}$ are similar matrices. Here we prove that if V is an inner product space and if B and B' are orthonormal bases of V, then the matrices $[T]_B$ and $[T]_{B'}$ are unitarily similar.

Following this, in Sec. 4.4, we help you recall normal operators and matrices. We also introduce you to more properties of these objects, including unitary diagonalisation.

In the last section, Sec. 4.5, we define the problem of least squares and give a solution of the problem using the idea of normal systems and orthonormal bases.

Objectives

After studying this unit, you should be able to

- define unitary similarity;
- transform a matrix into upper triangular form using unitary similarity;
- decide whether a given linear operator/matrix is unitarily diagonalisable or not;
- explain how a change of orthonormal basis is a unitary similarity;
- prove, and use, some basic properties of a normal linear operator;
- define the least squares problem, and solve it.

4.2 UNITARY SIMILARITY OF MATRICES

In this section, we will see that even if a matrix is not diagonalisable, we can take it 'near' diagonalisability by making it triangular via a special similarity transformation. For this, let us begin with recalling some definitions.

Definition: An $n \times n$ matrix $A = [a_{ij}]$ is called **upper triangular** if all the entries of A below the diagonal are zero, that is, if $a_{ij} = 0$ for $i > j$. For example, $[-3]$,

$\begin{bmatrix} 1 & 2 & 3 \\ 0 & \frac{1}{2} & 4 \\ 0 & 0 & 5 \end{bmatrix}$ and $\begin{bmatrix} 1 & 2 & i \\ 0 & 0 & 4 \\ 0 & 0 & 0 \end{bmatrix}$ are upper triangular, while $\begin{bmatrix} 1 & 2 & 3 \\ 0 & i & 4 \\ 0 & 6 & 5 \end{bmatrix}$ is not. In particular, every diagonal matrix is upper triangular.

Similarly, an $n \times n$ matrix $A = [a_{ij}]$ is called **lower triangular** if $a_{ij} = 0$ for $i < j$.

e.g., $\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 6 & 0 \end{bmatrix}$ and $\begin{bmatrix} i & 0 \\ -1 & 2i+3 \end{bmatrix}$ are lower triangular.

Definition: A set of vectors $\{u_1, \ldots, u_n\}$ in an inner product space $(U, <,>)$ is called **orthogonal** if $< u_i, u_j >= 0$ for $i \neq j$.

The set of vectors is an **orthonormal set** if

i) it is orthogonal; and

ii) $< u_i, u_i >= 1 \; \forall i = 1, \ldots, n$.

Recall that if $F = \mathbf{R}$ or $\mathbf{C}$, the vector space F^n of column vectors of size n is called the **standard inner product space over F** if the inner product is defined by

A* denotes the conjugate transpose of the matrix A.

$$< x, y > = \sum_{i=1}^{n} x_i \bar{y}_i = y^* x, \text{ where } x = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \text{ and } y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}.$$

Also, the norm of x, $\| x \|^2 = \sqrt{< x, x >} = \sqrt{x^* x} \; \forall x \in F^n$.

The Gram-Schmidt orthogonalization process implies that an inner product space has an orthonormal basis. One such orthonormal basis is $\{e_1, \ldots, e_n\}$, where e_i has 1 as the ith entry and 0 elsewhere. This is called **the standard basis of** F^n.

Definition: A matrix $U \in \mathbf{M}_n(\mathbf{C})$ is called **unitary** if $U^*U = I_n = UU^*$, that is, U is an invertible matrix and $U^{-1} = U^*$. U is called **an orthogonal matrix** if $UU^t = I_n = U^tU$. For example, $\begin{bmatrix} \sqrt{2} & i \\ -i & \sqrt{2} \end{bmatrix}$ is orthogonal but not unitary.

Examples of oft used unitary matrices are

$$\begin{bmatrix} \cos t & -\sin t \\ \sin t & \cos t \end{bmatrix} \text{ and } F = \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & \xi & \cdots & \xi^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & \xi^{n-1} & \cdots & \xi^{(n-1)^2} \end{bmatrix},$$ where ξ is a primitive nth root of unity.

Remarks: 1) If $A \in \mathbf{M}_n(\mathbf{R})$, then A is unitary if and only if it is orthogonal.

2) The product of two unitary matrices is also unitary.

3) A matrix $U \in \mathbf{M}_n(\mathbf{C})$ is unitary if and only if its columns form an orthonormal basis of the standard inner product space $\mathbf{C}^n$. [Why is this so? Let us see. Let

$U \in \mathbf{M}_n(\mathbf{C})$. Let its columns be $u_1, \ldots, u_n$. Consider these columns as elements of the standard inner product space $\mathbf{C}^n$. Then, for any two columns u_r and u_s,

$< u_r, u_s > = u_s^* u_r = (Ue_s)^*(Ue_r) = e_s^t(U^*U)e_r$.

So, $< u_r, u_s > = \delta_{rs}$ iff U is unitary,

The Kronecker delta function

$$\delta_{rs} = \begin{cases} 0, & r \neq s \\ 1, & r = s \end{cases}$$

i.e., the n columns of U form an orthonormal set of $\mathbf{C}^n$ iff U is unitary. Also, the columns are linearly independent iff U is non-singular, which it is if it is unitary!]

4) The inverse of a unitary matrix is unitary.

Let's consider an example.

Example 1: Which of the following matrices are unitary, and why?

i) $A = \begin{bmatrix} i & 1 \\ 1 & -i \end{bmatrix}$, ii) $B = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$, iii) $C = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & \sqrt{2} \\ 1 & -1 & 0 \end{bmatrix}$

Solution: i) The matrix A is not unitary since the norm of the first column is $\sqrt{2}$. In a unitary matrix the norm of each column must be 1.

ii) The matrix B is not unitary since Columns 1 and 2 are not orthogonal.

iii) The matrix C is unitary since the norm of each of its columns is 1, and any two columns are orthogonal.

Here are some related exercises now.

E1) Find a 3×3 unitary matrix whose first column is $\frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ i \\ -1 \end{bmatrix}$.

Given a unit vector u in $\mathbf{C}^n$, we can find a unitary matrix whose first column is u.

E2) Characterise unitary matrices that are diagonal, i.e., if $\text{diag}(d_1, \ldots, d_n)$ is unitary, then what properties do d_i satisfy $\forall\, i = 1, \ldots, n$?

E3) What is the determinant of a unitary matrix?

E4) Prove Remarks 1, 2 and 4 above.

E5) Write the statements analogous to Remarks 2, 3, 4 for orthogonal matrices. Also prove these statements.

In Unit 1, you studied about similarity of matrices. Recall that $A, B \in \mathbf{M}_n(\mathbf{C})$ are similar if there is an invertible matrix P such that $P^{-1}AP = B$. In particular, if P is a unitary matrix, then we say that A and B are unitarily similar. Note that if P is a unitary matrix then $P^{-1} = P^*$. So, we have the following definition.

Definition: $A, B \in \mathbf{M}_n(\mathbf{C})$ are **unitarily similar** if there is a unitary matrix P such that $B = P^*AP$.

Let us prove the main result of this section.

This is named after the mathematician Issai Schur (1875-1941).

Theorem 1 (Schur Decomposition): Let $A \in \mathbf{M}_n(\mathbf{C})$. Then A is unitarily similar to an upper triangular matrix. In particular, every square matrix over $\mathbf{C}$ is triangulisable.
Proof: We prove this by induction on n. If $n = 1$, then A is an upper triangular matrix, and so there is nothing to prove. So, for $n = 1$, the statement is trivially true.

If $A \in \mathbf{M}_n(\mathbf{C})$, then it has n eigenvalues in $\mathbf{C}$.

Assume that the statement holds for all matrices in $\mathbf{M}_{n-1}(\mathbf{C})$. Let $A \in \mathbf{M}_n(\mathbf{C})$, and let λ be an eigenvalue of A. Let v be a corresponding eigenvector of A, and let $u = v/\|v\|$. Then $\|u\| = 1$. (We call **u** the **normalised form** of v.)

Let U_1 be a unitary matrix whose first column is u (as in E1).
Now, consider the matrix $U_1^* A U_1$. Its first column is
$U_1^* A U_1 e_1 = U_1^* A u = \lambda U_1^* u = \lambda (U_1^* U_1) e_1 = \lambda e_1$.
Thus, $U_1^* A U_1$ is of the form $= \begin{bmatrix} \lambda & ** \\ \mathbf{0}_{n-1} & A_1 \end{bmatrix}$, where $\mathbf{0}_{n-1}$ denotes the column vector of length $n-1$ whose entries are all 0, ** stands for the remaining entries of the first row and $A_1 \in \mathbf{M}_{n-1}(\mathbf{C})$.

Now, by induction, there is an $(n-1) \times (n-1)$ unitary matrix V_2 such that $V_2^* A_1 V_2$ is upper triangular. Let $U_2 = \begin{bmatrix} 1 & \mathbf{0}_{n-1}^t \\ \mathbf{0}_{n-1} & V_2 \end{bmatrix}$. You can verify that U_2 is unitary. Also,

$$U_2^*(U_1^* A U_1) U_2 = U_2^* \begin{bmatrix} \lambda & ** \\ \mathbf{0}_{n-1} & A_1 \end{bmatrix} U_2 = \begin{bmatrix} \lambda & ** \\ \mathbf{0}_{n-1} & V_2^* A_1 V_2 \end{bmatrix},$$

which is an upper triangular matrix.
Hence, if $U = U_1 U_2$, then U is unitary and $U^* A U$ is upper triangular. ■

Now, what if $A \in \mathbf{M}_n(\mathbf{C})$ is such that all its entries are in $\mathbf{R}$ and all its n eigenvalues are real numbers? Then the corresponding eigenvectors will be in $\mathbf{R}^n$. (Why?) So, in the proof of Theorem 1 we can work on the standard inner product space $\mathbf{R}^n$ rather than $\mathbf{C}^n$, and the corresponding unitary matrix U will have real entries. In this case, U will be such that $U^t U = I_n = U U^t$, i.e., U will be orthogonal. Hence, we have the following result.

Theorem 2: If $A \in \mathbf{M}_n(\mathbf{R})$ and all its eigenvalues are real numbers, then there is an orthogonal matrix Q such that $Q^t A Q$ is upper triangular. ■

Note: Since unitary/orthogonal similarity is a particular case of general similarity, all the results (in Unit 1) about similar matrices hold true in these cases too.

Let us consider some examples of unitary similarity now.

Example 2: Let $A = \begin{bmatrix} 0 & 0 & -1 \\ 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix}$. Find a unitary matrix U such that $U^* A U$ is upper triangular.

Solution: The characteristic polynomial of A is $t^3 + t^2 + t + 1$. So, $-1, i$ and $-i$ are the eigenvalues of A. A normalized eigenvector corresponding to the eigenvalue -1

is $u = \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\end{bmatrix}$. A unitary matrix, whose first column is u, is

$$U_1 = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix}. \text{ Then } U_1^* A U_1 = \begin{bmatrix} -1 & 1 & 1/\sqrt{2} \\ 0 & 0 & -1/\sqrt{2} \\ 0 & \sqrt{2} & 0 \end{bmatrix}.$$

Next, we consider the 2×2 matrix $A_1 = \begin{bmatrix} 0 & -1/\sqrt{2} \\ \sqrt{2} & 0 \end{bmatrix}$. This matrix has an

eigenvalue i, and its corresponding normalized eigenvector is $\frac{1}{\sqrt{3}}\begin{bmatrix}1\\-i\sqrt{2}\end{bmatrix}$. A 2×2

unitary matrix whose first column is this vector is $\frac{1}{\sqrt{3}}\begin{bmatrix}1 & i\sqrt{2}\\ -i\sqrt{2} & 1\end{bmatrix}$. Hence, if

$U_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{\sqrt{3}} & i\sqrt{2/3} \\ 0 & -i\sqrt{2/3} & 1/\sqrt{3} \end{bmatrix}$, then U_2 is unitary and

$$U_2^*(U_1^* A U_1)U_2 = \begin{bmatrix} -1 & (1-i)/\sqrt{3} & (1-2i)/\sqrt{6} \\ 0 & i & 1/\sqrt{2} \\ 0 & 0 & -i \end{bmatrix}.$$

Thus, if $U = U_1U_2$, then U^*AU is an upper triangular matrix.

Example 3: Let $A = \begin{bmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{bmatrix}$. Does there exist a real orthogonal matrix Q such

that Q^tAQ is upper triangular? If so, find one such Q.

Solution: The characteristic polynomial of A is $(t-5)(t-2)^2$. Since all the roots of the characteristic polynomial are real numbers, by Theorem 2 there exists an orthogonal matrix Q with real entries such that Q^tAQ is upper triangular.
To find Q, we first consider the eigenvalue 5. A normalized eigenvector

corresponding to 5 is $u = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\1\\1\end{bmatrix}$. Next, an orthogonal matrix with first column u is

$$Q_1 = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \end{bmatrix}, \text{ and } Q_1^t A Q_1 = \begin{bmatrix} 5 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 2 \end{bmatrix}.$$

Hence, $Q = Q_1$. In fact, this Q not only triangulises A, but diagonalises it.

Note: In the example above, our choice of Q_1 turned out to be good. But, if we had taken some other choice of Q_1 we may have needed to work more, as in Example 2.

Here are some related exercises now.

E6) Find a unitary matrix U such that U^*AU is upper triangular, where

$$A = \begin{bmatrix} 2 & 1 & -3 \\ 1 & 2 & -3 \\ -4 & -2 & 6 \end{bmatrix}.$$

E7) Let $u \in \mathbf{C}^n$ have norm 1. Show that if $v \in \mathbf{C}^n$ is orthogonal to u, then v is an eigenvector of the $n \times n$ matrix $A = I_n - uu^*$. Use this to find a unitary matrix U such that U^*AU is a diagonal matrix.

E8) Give an example of a matrix that is **not** diagonalisable, but is triangulisable.

E9) Show that the triangular matrix to which a given matrix is unitarily similar is not unique.

[**Hint**: For instance, verify that for $A = \begin{bmatrix} 1 & 0 & 3 \\ 0 & 1 & 4 \\ 0 & 0 & 2 \end{bmatrix}$, P^*AP is also triangular,

where $P = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$ is a unitary matrix.]

E10) If A is unitarily similar to T and A^* is unitarily similar to B, is there any relationship between T and B? Give reasons for your answer.

So far you have seen that every matrix is unitarily similar to a triangular matrix. You also know, from Block 1, that some matrices are diagonalisable. If you go back to those examples, you will see that the non-singular matrix that diagonalises a given matrix need not be unitary. However, in some cases this matrix is unitary.

Definition: An $n \times n$ matrix with entries from $\mathbf{C}$ is called **unitarily diagonalisable** if it is unitarily similar to a diagonal matrix.

Let us start with a simple observation. If X is an upper triangular matrix, then X^* will be a lower triangular matrix, that is, all its entries above the main diagonal will be zero. Therefore, **an upper triangular matrix is hermitian if and only if it is a diagonal matrix. Further, the entries on the diagonal have to be real numbers.**

Now, suppose that A is a hermitian matrix. Let U be a unitary matrix such that $U^*AU = X$, where X is upper triangular. Then

$X^* = (U^*AU)^* = U^*A^*U = U^*AU = X$,

and so X is hermitian and upper triangular. Therefore, X is a diagonal matrix with real entries. This shows that **a hermitian matrix is unitarily diagonalisable** and its eigenvalues are all real numbers. In particular, a hermitian operator has only real eigenvalues.

Now, some recalling from your undergraduate studies!

Definition: An $n \times n$ matrix A with entries from **C** is called normal if $AA^* = A^*A$.

For instance, hermitian matrices and unitary matrices are normal. Can you find an example of a normal matrix which is neither unitary nor hermitian?

Now, we will state a result which gives us an important property of normal matrices. But, first some related observations.

Remarks: 1) Let $B \in \mathbf{M}_n(\mathbf{C})$. Let b_{ij} be the (i, j) th entry of B. Then you should check that the (k, k) th entry of BB^* is $\sum_{i=1}^{n} |b_{ki}|^2$, and the (k, k) th entry of B^*B is $\sum_{i=1}^{n} |b_{ik}|^2$. So, if $BB^* = B^*B$, then for all $k = 1, \ldots, n$, $\sum_{i=1}^{n} |b_{ki}|^2 = \sum_{i=1}^{n} |b_{ik}|^2$.

2) If D is a diagonal matrix, $DD^* = D^*D$.

Let us now prove the following theorem.

Theorem 3: Let $A \in \mathbf{M}_n(\mathbf{C})$. Then A is unitarily diagonalisable if and only if A is a normal matrix.

Proof: Suppose, first, that A is unitarily diagonalisable. Let U be a unitary matrix such that $U^*AU = D$, where D is a diagonal matrix. Then $A = UDU^*$ and $AA^* = (UDU^*)(UDU^*)^* = UDD^*U = UD^*DU^* = (UDU^*)^*(UDU^*) = A^*A$, since D is a diagonal matrix.

Conversely, let A be a normal matrix. Let U be a unitary matrix such that $U^*AU = X$, where X is an upper triangular matrix. Then $XX^* = (U^*AU)(U^*AU)^* = (U^*AA^*U) = (U^*A^*AU) = X^*X$, and so X is normal. Let x_{ij} denote the (i, j) th entry of X. Equating the (1, 1) th entries of XX^* and X^*X, we have $\sum_{i=1}^{n} |x_{1i}|^2 = \sum_{i=1}^{n} |x_{i1}|^2$ (from Remark 1 above). Also, as X is upper triangular, $x_{ij} = 0$ if $i > j$. So $\sum_{i=1}^{n} |x_{i1}|^2 = |x_{11}|^2$.

So, $\sum_{i=1}^{n} |x_{1i}|^2 = |x_{11}|^2$, i.e., $\sum_{i=2}^{n} |x_{1i}|^2 = 0$, and so $x_{12} = \ldots = x_{1n} = 0$. Thus, all the entries of the first row, except possibly at the (1, 1) th place, are zero.

Next, equating the (2,2) th entries of XX^* and X^*X, and using the fact that X is upper triangular, we have $\sum_{i=1}^{n} |x_{2i}|^2 = |x_{12}|^2 + |x_{22}|^2$. Since $x_{12} = 0$, we see that $x_{23} = \ldots = x_{2n} = 0$. Thus, all the entries of the second row, except possibly at the (2, 2) th place, are zero.
Continuing in this way, we find that X is a diagonal matrix. ■

The following theorem is also useful sometimes to determine if a given matrix is unitarily diagonalisable.

The trace of a square matrix A $= tr(A) = \sum_{i=1}^{n} a_{ii}$.

Theorem 4: $A \in \mathbf{M}_n(\mathbf{C})$ is unitarily diagonalisable if and only if $tr(AA^*) = \sum_{i=1}^{n} |\lambda_i|^2$, where $\lambda_1, \ldots, \lambda_n$ are eigenvalues of A.

Proof: Suppose A is unitarily diagonalisable, with eigenvalues $\lambda_1, \ldots, \lambda_n$. Let U be a unitary matrix such that $U^*AU = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$.

Then $(U^*AU)(U^*AU)^* = U^*AA^*U = \begin{bmatrix} |\lambda_1|^2 & & 0 \\ & \ddots & \\ 0 & & |\lambda_n|^2 \end{bmatrix}$.

So, $tr(AA^*) = tr(U^*AA^*U)$ (Why?)

$= \sum_{i=1}^{n} |\lambda_i|^2$

For the converse, assume U is a unitary matrix such that $U^*AU = X$, an upper triangular matrix. Denote the (i, j) th entry of X by x_{ij}. Since $U^*AA^*U = (U^*AU)(U^*AU)^* = XX^*$, it follows that

$tr(AA^*) = tr(XX^*) = \sum_{j=1}^{n}\sum_{i=1}^{n} |x_{ij}|^2$.

So, $tr(AA^*) = \sum_{i=1}^{n} |x_{ii}|^2 + \sum_{i<j} |x_{ij}|^2$.

Now, the diagonal entries of X are the eigenvalues of A. (Why?) Therefore, $\sum_{i=1}^{n} |x_{ii}|^2 = \sum_{i=1}^{n} |\lambda_i|^2 = tr(AA^*)$ (given to us). Hence $\sum_{i<j} |x_{ij}|^2 = 0$, and so $x_{ij} = 0$ for $i < j$. This proves that X is a diagonal matrix, and so A is unitarily diagonalisable. ■

Let us look at some examples now.

Example 4: Check whether or not $A = \begin{bmatrix} 5 & -6 & -6 \\ -1 & 4 & 2 \\ 3 & -6 & -4 \end{bmatrix}$ is unitarily diagonalisable.

Solution: Since A has only real entries, $A^* = A^t$. To show that A is unitarily diagonalisable we can either apply Theorem 3 or Theorem 4. Theorem 3 requires us to check if $AA^* = A^*A$ or not. Theorem 4 states that for A to be unitarily diagonalisable we need to check whether $tr(AA^*) = \sum_{i=1}^{n} |\lambda_i|^2$. But, for this, we need to find the eigenvalues of A.

To apply Theorem 3, note that the (1, 1) th entry of AA^* is the sum of the squares of the entries of the first row, that is, 97. On the other hand, the (1,1) th entry of A^*A is the sum of the squares of the entries in the first column, which is 35. Hence $AA^* \neq A^*A$, and so A is not unitarily diagonalisable.

To apply Theorem 4, note that the eigenvalues of A are 2, 2, 1. Since A has real entries, $tr(AA^*)$ is the sum of the squares of the entries of A, which is greater than $2^2 + 2^2 + 1^2 = 9$. So, $tr(AA^*) \neq \sum_{i=1}^{n} |\lambda_i|^2$.

You can check that A is actually a diagonalisable matrix, that is, A is similar to a diagonal matrix. Hence, A is diagonalisable but not unitarily diagonalisable.

Example 5: Check if $A = \begin{bmatrix} 1 & 0 & 2 \\ 0 & 3 & 0 \\ -2 & 0 & 1 \end{bmatrix}$ is unitarily diagonalisable. If it is, find a unitary matrix U such that U^*AU is diagonal. If it is not, then find a triangular matrix to which it is unitarily similar.

Solution: The characteristic polynomial of A is $(t-3)\{(t-1)^2+4\}$. The eigenvalues are $3, 1+2i, 1-2i$.

$tr(AA^*) = 1^2+2^2+3^2+(-2)^2+1^2 = 19 = 3^2+|1+2i|^2+|1-2i|^2$. So, by Theorem 4, A is unitarily diagonalisable.

So, now let us find U. You can check that the normalized eigenvectors corresponding to $1+2i, 1-2i$ and 3 are $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ 0 \\ i \end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix} i \\ 0 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$. These are also mutually orthogonal. Therefore, U is $\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & i & 0 \\ 0 & 0 & \sqrt{2} \\ i & 1 & 0 \end{bmatrix}$.

Some related exercises now.

E11) Give an example to show that the sum of two unitarily diagonalisable matrices need not be unitarily diagonalisable. Is the product of two unitarily diagonalisable matrices unitarily diagonalisable? Why?

E12) Find a unitary matrix U such that U^*AU is diagonal, where $A = \begin{bmatrix} 1 & -i \\ i & 1 \end{bmatrix}$.

E13) Are the matrices $\begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$ and $\begin{bmatrix} -1 & 1 \\ 1 & 1 \end{bmatrix}$ unitarily similar?

E14) Show that if A is a unitary matrix with real eigenvalues, then A is hermitian.

E15) Check whether or not the following matrices are unitarily diagonalisable:

(i) $\begin{bmatrix} 0 & 0 & -1 & 1 \\ 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{bmatrix}$ (ii) $\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & -1 & 1 \end{bmatrix}$.

E16) Check whether or not the matrix $A = \begin{bmatrix} 3 & -i & 1-2i \\ i & 1 & i \\ 1-2i & -i & 3 \end{bmatrix}$ is unitarily diagonalisable. If it is, find a unitary matrix U such that U^*AU is diagonal. If A is not unitarily diagonalisable, check whether it is diagonalisable.

E17) Show that a normal upper triangular matrix is a diagonal matrix.

E18) Let $A \in \mathbf{M}_n(\mathbf{C})$. Show that $tr(AA^*) = \sum_{j=1}^{n}\sum_{i=1}^{n}|a_{ij}|^2$, where a_{ij} is the (i, j) th

$tr(AA^*)$ is the sum of squares of the absolute values of the entries of A.

entry of A. Deduce that $A = \mathbf{0}$ if and only if $tr(AA^*) = 0$.

In this section, we have looked at unitarily similar matrices. Let us take a look at the linear transformations analogously.

4.3 CHANGE OF ORTHONORMAL BASIS

In Block 1, you have seen that if B_1 and B_2 are two bases of a finite-dimensional vector space V over F and T is a linear operator on V, then $[T]_{B_1}$ and $[T]_{B_2}$ are similar.

In this section you will see that the change of orthonormal basis for the matrix representation of a linear operator on an inner product space is actually a similarity transform by a unitary matrix.

For this, recall that if $B = \{x_1, \ldots, x_n\}$ is an ordered orthonormal basis of an inner product space V over $F(= \mathbf{R}$ or $\mathbf{C})$, then for any $v \in V$,

$v = < v, x_1 > x_1 + \cdots + < v, x_n > x_n$.

So, $w \in V$ is such that $v \perp w$ if and only if $< v, w > = 0$

iff $\left\langle \sum_{i=1}^{n} < v, x_i > x_i, \sum_{j=1}^{n} < w, x_j > x_j \right\rangle = 0$.

iff $\sum_{i=1}^{n} < v, x_i > \overline{< w, x_i >} = 0$, ...(1)

a fact we are going to use while proving the following theorem.

Theorem 5: Let V be an inner product space over a field F. Let $\mathcal{B}$ and $\mathcal{B}'$ be ordered orthonormal bases of V. Then, for any linear operator T on V, the matrices $[T]_B$ and $[T]_{B'}$ are unitarily similar.

Proof: Let $\mathcal{B} = \{x_1, \ldots, x_n\}$ and $\mathcal{B}' = \{y_1, \ldots, y_n\}$. Then, as we have seen in Unit 1, $[T]_{\mathcal{B}'} = {}_{\mathcal{B}'}[I]_{\mathcal{B}}\ [T]_{\mathcal{B}}\ {}_{\mathcal{B}}[I]_{\mathcal{B}'}$. Thus, to prove the result, we need to check that the matrix ${}_{\mathcal{B}}[I]_{\mathcal{B}'}$ is unitary. For this, first note that the jth column of the matrix ${}_{\mathcal{B}}[I]_{\mathcal{B}'}$ is $[y_j]_{\mathcal{B}}$.

Since $y_j = \sum_{i=1}^{n} < y_j, x_i > x_i$, we see that $[y_j]_{\mathcal{B}} = \begin{bmatrix} < y_1, x_1 > \\ \vdots \\ < y_n, x_n > \end{bmatrix}$.

Next, since the basis elements are orthonormal, $< y_r, y_s > = \begin{cases} 0, & \text{if } r \neq s \\ 1, & \text{if } r = s \end{cases}$.

Hence, by (1), $\sum_{i=1}^{n} < y_r, x_i > \overline{< y_s, x_i >} = \begin{cases} 0, & \text{if } r \neq s \\ 1, & \text{if } r = s \end{cases}$

This shows that the columns of ${}_{\mathcal{B}}[I]_{\mathcal{B}'}$ are orthonormal.

Hence ${}_{\mathcal{B}}[I]_{\mathcal{B}'}$ is a unitary matrix. (See Remark 3 of this unit.) ■

For linear operators on an inner product space, we have the following analogue of Theorem 1.

Theorem 6 (Schur Form): Let T be a linear operator on a finite-dimensional inner product space. Then there is an orthonormal basis with respect to which the matrix of T is upper triangular.

Proof: Let $\mathcal{B} = \{x_1, \ldots, x_n\}$ be an orthonormal basis of V. Then, by Theorem 1, there is a unitary matrix U such that $U^*[T]_{\mathcal{B}}U$ is upper triangular. Let the (i, j)th entry of U be u_{ij}. Let $y_j = \sum_{i=1}^{n} u_{ij}x_i$. Then you can verify that $< y_r, y_s > = \begin{cases} 0, \text{ if } r \neq s \\ 1, \text{ otherwise} \end{cases}$.
Therefore, $\mathcal{B}' = \{y_1, \ldots, y_n\}$ is an orthonormal basis of V. (Check that this set is linearly independent.) Now from Unit 1, Sec. 1.2, it follows that $[T]_{\mathcal{B}'} = U^*[T]_{\mathcal{B}}U$, an upper triangular matrix. ■

Here is an exercise now.

E19) Let T be a linear operator on $\mathbf{R}^3$ whose action on the standard orthonormal basis $\{e_1, e_2, e_3\}$ is given by: $Te_1 = e_1 + 3e_2 + e_3$, $Te_2 = 4e_2 + e_3$ and $Te_3 = 2e_1 + 2e_2 + e_3$. Find an orthonormal basis with respect to which the matrix of T is upper triangular.
(**Hint:** Write the matrix representation of T with respect to the standard basis first, and then find a unitary matrix. Finally write a basis with the help of this unitary matrix, as is done in the proof of Theorem 6.)

So far we have considered unitary matrices and similarity using unitary operators. Let us now consider another kind of operator of which a unitary operator is a particular case.

4.4 NORMAL OPERATORS

You know that a matrix which commutes with its adjoint is a **normal matrix**. Similarly, a linear operator T on an n-dimensional inner product space V, is **normal** if $TT^* = T^*T$, where the linear operator T^*, from V to V, is defined by the relationship $< Tu, v > = < u, T^*v > \forall u, v \in V$.

Remark: If T is unitary or self-adjoint, then it is normal.

Recall that T is triangulisable if V has a basis $\mathcal{B}$ such that the matrix of T with respect to $\mathcal{B}$ is upper triangular. Thus, it follows that the characteristic polynomial, and hence the minimal polynomial, is a product of linear factors.

E20) Is a normal operator triangulisable? Give reasons for your answers.

E21) If T is a linear operator on a finite-dimensional inner product space, then show that $V = \operatorname{Im} T \oplus \operatorname{Ker} T^*$.

E22) Prove that if T is a normal operator, then for any $a \in \mathbf{C}$, the operator $aI + T$ is also normal.

Now we list some important properties of a normal operator.

Theorem 7: Let T be a normal operator on a finite-dimensional inner product space V. Then

(i) $\operatorname{Ker} T = \operatorname{Ker} T^*$;

(ii) $\operatorname{Ker} T^n = \operatorname{Ker} T$ for all $n \in \mathbf{N}$;

(iii) $Tv = \lambda v$ if and only if $T^* v = \bar{\lambda} v$;

(iv) eigenvectors corresponding to distinct eigenvalues are orthogonal.

Proof: (i) For $v \in V$, $\| Tv \|^2 = < Tv, Tv > = < v, T^*Tv > = < v, TT^*v >$

$= \overline{< TT^*v, v >} = \overline{< T^*v, T^*v >} = < T^*v, T^*v > = \| T^*v \|^2$.

Therefore, $Tv = \mathbf{0}$ if and only if $T^*v = \mathbf{0}$. Hence (i) is proved.

(ii) Since $V = \operatorname{Im} T \oplus \operatorname{Ker} T^*$, by (i) we have $V = \operatorname{Im} T \oplus \operatorname{Ker} T$. Now we prove (ii) by induction on n.
The statement is trivially true if $n = 1$. Suppose it is true for $n = m > 1$, i.e., $\operatorname{Ker} T^m = \operatorname{Ker} T$.
Now, let $v \in \operatorname{Ker} T^{m+1}$.
$\Rightarrow T^m(Tv) = \mathbf{0} \Rightarrow Tv \in \operatorname{Ker} T^m = \operatorname{Ker} T$
Also, by definition, $Tv \in \operatorname{Im} T$.
Therefore, $Tv \in \operatorname{Ker} T \cap \operatorname{Im} T = \{\mathbf{0}\}$.
Hence $Tv = \mathbf{0}$, and $v \in \operatorname{Ker} T$.
Thus (ii) is true $\forall n \geq 1$.

(iii) Firstly, since T is normal, $(T - \lambda I)(T - \lambda I)^* = (T - \lambda I)^*(T - \lambda I)$, that is, $(T - \lambda I)$ is a normal operator. Thus, by (i) we have
$\operatorname{Ker}(T - \lambda I) = \operatorname{Ker}(T - \lambda I)^* = \operatorname{Ker}(T^* - \bar{\lambda} I)$
Therefore, $(T - \lambda I)v = \mathbf{0}$ if and only if $(T^* - \bar{\lambda} I)v = \mathbf{0}$; hence (iii).

(iv) Let $Tv = \lambda v$ and $Tw = \mu w$, $\lambda \neq \mu$. Then, we see that
$\lambda < v, w > = < Tv, w > = < v, \bar{\mu} w >$, using (iii).
$= \mu < v, w >$.
So, $(\lambda - \mu) < v, w > = 0$, and hence $< v, w > = 0$, as $\lambda \neq \mu$. ■

Why don't you try some exercises now?

E23) Show that a normal operator is self-adjoint if and only if all its eigenvalues are real.

E24) Show that a normal operator is unitary if and only if all its eigenvalues have absolute value 1.

And now we come to a related application, which is widely used.

4.5 THE METHOD OF LEAST SQUARES

Consider a system of linear equations, $Ax = b$, where A is an $n \times m$ matrix with entries from the field of real numbers and $b \in \mathbf{R}^n$. Recall that the given system has a solution if and only if $b \in \operatorname{Im} A$. Now the problem is what happens if $b \notin \operatorname{Im} A$, that is, if we have an inconsistent system. In this case the best we can do is to find a vector $u \in \mathbf{R}^m$ such that Au is the closest to b. What is the measure of closeness that we can pick? Let us see.

We have the underlying vector space $\mathbf{R}^m$ with the standard inner product, where $\| x-y \| = \sqrt{(x-y)^*(x-y)} = \sqrt{(x_1-y_1)^2+\cdots+(x_m-y_m)^2}$ is the **distance between the vectors** x **and** y. In this sense, if we want a vector $u \in \mathbf{R}^m$ such that Au is as close to b as possible, then we want u to be such that the distance $\| Au-b \|$ is the least possible. If we can find such a vector u, then this is called a least squares solution of the system. Formally, we have the following definition.

$Ax = y$ is consistent iff $\exists\, u \in \mathbf{R}^m$ such that $\|Au - b\| = 0$.

Definition: $u \in \mathbf{R}^m$ is called a **least squares solution** of the system $Ax = b$, if $\| Au-b \| \le \| Ax-b \|$ for all $x \in \mathbf{R}^m$.

The method of least squares grew out of the fields of astronomy and geodesy as scientists and mathematicians tried to help the sailors to navigate the oceans. In this method, model numerical data obtained from observations is approximated by a curve that gives an optimal fit of the data. The best fit is obtained by taking the sum of squared residuals with least value, a residual being the difference between an observed value (the dots in Fig. 1) and the value given by the curve (the curve in Fig. 1). The method was first described by the mathematician Carl Friedrich Gauss around 1794. Least squares corresponds to the maximum likelihood criterion if the experimental errors have a normal distribution (see MMT-008).

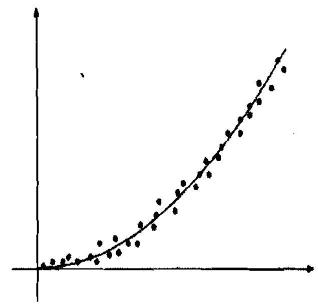

Fig. 1: The result of fitting a set of data points with a quadratic function

The first question is the existence of a vector that provides a least squares solution. To look into this, we need to recall some facts.

Fig. 2: The German mathematician Carl Friedrich Gauss (1777-1855)

Firstly, let W be a subspace of an inner product space V. Then $W^\perp = \{u \in V \mid <u, w> = 0 \text{ for all } w \in W\}$ is a subspace of V, and $\mathbf{V} = \mathbf{W} \oplus \mathbf{W}^\perp$. So, for each $v \in V$ there are unique $w \in W$ and $u \in W^\perp$ such that $v = w + u$. This defines a map $Pr_W : V \to V : Pr_W(v) = w$. This map is called the **orthogonal projection** of V onto W. You should verify that this map is a linear operator on V such that $\operatorname{Im} Pr_W = W$ and $\operatorname{Ker} Pr_W = W^\perp$.

Next, recall that if u and v are orthogonal vectors in an inner product space, then $\| u+v \|^2 = \| u \|^2 + \| v \|^2$.

Now, we can see if a least squares solution exists.

Theorem 8: Consider the linear system $Ax = y$. Let $W = \operatorname{Im} A$ and $b = Pr_W(y)$. If x_0 is a solution of the system $Ax = b$, then x_0 is a least squares solution of $Ax = y$, that is, $\| Ax_0 - y \| \le \| Ax - y \|$ for all $x \in \mathbf{R}^m$.

Proof: Note that since $b = Pr_W(y) \in \operatorname{Im} A$, the system $Ax = b$ is consistent. Next, $Pr_W(y-b) = Pr_W(y) - Pr_W(b) = b - b = \mathbf{0}$, so that $y - b \in \operatorname{Ker} Pr_W = W^\perp$.
Since $Ax - b \in W = \operatorname{Im} A$, and $b - y \in W^\perp$, $< Ax-b,\ b-y > = 0$.
Therefore, $\| Ax-y \|^2 = \| (Ax-b)+(b-y) \|^2$
$$= \| Ax-b \|^2 + \| b-y \|^2.$$
Thus, if x_0 is a solution of $Ax = b$, then
$\| Ax_0 - y \|^2 = \| Ax_0 - b \|^2 + \| b-y \|^2 = \| b-y \|^2$.
Therefore, for all $x \in \mathbf{R}^m$,
$\| Ax-y \|^2 = \| Ax-b \|^2 + \| b-y \|^2 = \| Ax-b \|^2 + \| Ax_0 - y \|^2 \ge \| Ax_0 - y \|^2$.
This shows that x_0 is a least squares solution of $Ax = y$. ■

Our next result shows that a least squares solution is actually a solution of a linear system called the normal system. Before we prove it, note that

$v \in \operatorname{Im}(A)^{\perp}$

$\Leftrightarrow \langle Au, v \rangle = \mathbf{0} \,\forall u \in \mathbf{R}^m$

$\Leftrightarrow v^t Au = \mathbf{0} \;\forall\; u \in \mathbf{R}^m$

$\Leftrightarrow u^t(A^t v) = \mathbf{0} \;\forall u \in \mathbf{R}^m.$

$\Leftrightarrow A^t v = \mathbf{0}$. (Why?)

$\mathbf{R}^m = \operatorname{Im} A \oplus \operatorname{Ker} A^t$

Therefore, $\mathbf{Im(A)^{\perp} = Ker(A^t)}$.

This fact will be used in proving the following theorem.

Theorem 9: $x_0 \in \mathbf{R}^m$ is a least squares solution of $Ax = y$ if and only if x_0 is a solution of the system $A^t Ax = A^t y$.

Proof: First, assume that x_0 is a least squares solution. Then $Ax_0 = b$, where $b = Pr_W(y)$, where $W = \operatorname{Im} A$.

Write $y = b + b'$, where $b' \in W^{\perp} = \operatorname{Ker}(A^t)$.

Then $Ax_0 - y = Ax_0 - b - b' = -b' \in \operatorname{Ker}(A^t)$.

Therefore, $A^t(Ax_0 - y) = \mathbf{0}$, and so $A^t Ax_0 = A^t y$.

Hence x_0 is a solution of $A^t Ax = A^t y$.

Conversely, if x_0 is a solution of $A^t Ax = A^t y$, then $A^t(Ax_0 - y) = \mathbf{0}$, and so $Ax_0 - y \in \operatorname{Ker}(A^t)$.

Write $y = b + b'$, with $b \in W = \operatorname{Im}(A)$ and $b' \in \operatorname{Ker}(A^t)$. Then

$Ax_0 - b = (Ax_0 - y) + b' \in \operatorname{Ker}(A^t) = W^{\perp}$. Also $Ax_0 - b \in \operatorname{Im}(A) = W$.

$\therefore \; Ax_0 - b \in W \cap W^{\perp} = \{\mathbf{0}\}$.

$\therefore \; Ax_0 = b$

So, by Theorem 8, x_0 is a least squares solution of $Ax = y$. ■

Thus, this theorem gives the solution of the problem of least squares. But, this gives a solution only if $A^t Ax = A^t y$ has a solution. Does it? In E 25 we have asked you to check this.

E25) Show that $\operatorname{rank}(A) = \operatorname{rank}(A^t A)$.

E26) Let $A \in \mathbf{M}_{m \times n}(F)$. Show that $A^t Ax = A^t y$ is consistent, for any $y \in \mathbf{R}^m$.

It all now depends on the matrix $A^t A$. If A is invertible, then by E 26, $A^t A$ is invertible, and the least squares solution in this case is given by $(A^t A)^{-1} A^t y$. If A is not invertible, then the corresponding normal system $(A^t A)x = A^t y$ will have more than one solution. In this case we may take the least squares solution x_0 such that $\| x_0 \| = \min\{\| u \| \mid u \in \mathbf{R}^m, A^t Au = A^t y\}$.

Let us look at a few examples.

Example 6: Check whether $Ax = y$, where $A = \begin{bmatrix} 3 & -1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix}$ and $y = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$, is consistent.

If it is, solve it. If it is inconsistent, find a least squares solution of the system.

Solution: The system $Ax = y$ is inconsistent. (You should verify that this is so!) To find a least squares solution, we solve the normal system $A^tAx = A^ty$. Here

$$A^tA = \begin{bmatrix} 3 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 3 & -1 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 11 & -2 \\ -2 & 2 \end{bmatrix}.$$

Thus, A^tA is invertible, and $(A^tA)^{-1} = \frac{1}{18}\begin{bmatrix} 2 & 2 \\ 2 & 11 \end{bmatrix}$.

So the least squares solution is given by

$$x_0 = (A^tA)^{-1}A^ty = \frac{1}{18}\begin{bmatrix} 2 & 2 \\ 2 & 11 \end{bmatrix}\begin{bmatrix} 3 & 1 & 1 \\ -1 & 1 & 0 \end{bmatrix}\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 3/2 \end{bmatrix}.$$

Example 7: Find a least squares solution of the system $Ax = y$, where

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 1 & -1 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 2 \end{bmatrix} \text{ and } y = \begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

Solution: The given system is inconsistent. (Why?)

Here $A^tA = \begin{bmatrix} 3 & 2 & 5 \\ 2 & 6 & 8 \\ 5 & 8 & 13 \end{bmatrix}$, which is a matrix of rank 2. The corresponding normal

system is $A^tAx = A^ty$, that is, $\begin{bmatrix} 3 & 2 & 5 \\ 2 & 6 & 8 \\ 5 & 8 & 13 \end{bmatrix}\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 2 & -1 & 0 & 1 \\ 3 & 0 & 0 & 2 \end{bmatrix}\begin{bmatrix} 1 \\ 0 \\ 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}$.

You can apply Gaussian elimination to obtain the general solution of this system as

$$\left\{ \frac{1}{7}\begin{bmatrix} 1-7t \\ 2-7t \\ 7t \end{bmatrix} \middle| t \in \mathbf{R} \right\}.$$

All these are least squares solutions. In this case we find t such that $\| u \| = \frac{1}{7}\sqrt{(1-7t)^2 + (2-7t)^2 + 49t^2}$ is the least, i.e., for which the value of the quadratic polynomial $147t^2 - 42t + 5$ is the least. Since this represents a parabola, its least value is for $t = \frac{42}{2 \times 147} = \frac{1}{7}$.

(Another way of finding the minimum is to find f' and f'', where $f(t) = 147t^2 - 42t + 5$. Then apply calculus to obtain the result.)

Try some similar exercises now.

E27) Do two different least squares solutions of the following system exist ? If yes, find them. If not, give reasons for saying so.
$x-y+2z=1, -x+2z=1, x+4z=1$.

E28) Find a least squares solution of the linear system $Ax = y$,

$$\text{where } A=\begin{bmatrix}1 & 0\\ -1 & 1\\ 0 & -1\\ 2 & 1\end{bmatrix} \text{ and } y=\begin{bmatrix}1\\ -1\\ 2\\ 3\end{bmatrix}.$$

E29) Find a least squares solution u, of minimum norm, of the linear system

$$Ax = y, \text{ where } A=\begin{bmatrix}1 & 2 & 0 & -1\\ 1 & -1 & 3 & 2\\ 1 & -1 & 3 & 2\\ -1 & 1 & -3 & 1\end{bmatrix} \text{ and } y=\begin{bmatrix}1\\ -1\\ 2\\ 1\end{bmatrix}.$$

Let us now consider some situations which illustrate the utility of this type of solution. The least squares solution basically comes into play when a linear system is over determined, i.e., 'too many' variables are involved. This solution is a projection onto a subspace, which, of course, contains fewer basis elements. This method for obtaining a best possible solution is used widely in many areas, particularly when there are systems involving several thousands of variables, as in the models used for obtaining clearer satellite pictures, in many regression models, in models depicting financial situations, in models studying biology, etc. However, it has its limitations, particularly related to the size of error one gets because it is an approximate solution. We will not discuss this here, but any book that discusses the least squares problem will give you some idea about this.

Let us look at an example of the use of the method of least squares in real-life situations.

Example 8: The owner of a rapidly expanding business finds that for the first five months of the year her sales are (in lakhs) Rs. 4, Rs. 4.4, Rs. 5.2, Rs. 6.4 and Rs. 8, respectively. She plots these figures on a graph and conjectures that for the rest of the year her sales curve can be approximated by a quadratic polynomial. Find the least squares quadratic polynomial fit to the sales curve, and use it to project the sales for December of that year.

Solution: Suppose the polynomial is $a+bx+cx^2$. This is to be fitted to the data (1, 4), (2, 4.4), (3, 5.2), (4, 6.4), (5, 8). Here

$$A=\begin{bmatrix}1 & x_1 & x_1^2\\ 1 & x_2 & x_2^2\\ 1 & x_3 & x_3^2\\ 1 & x_4 & x_4^2\\ 1 & x_5 & x_5^2\end{bmatrix}=\begin{bmatrix}1 & 1 & 1\\ 1 & 2 & 4\\ 1 & 3 & 9\\ 1 & 4 & 16\\ 1 & 5 & 25\end{bmatrix} \text{ and } y=\begin{bmatrix}4\\ 4.4\\ 5.2\\ 6.4\\ 8\end{bmatrix}$$

$$\text{So, } v=\begin{bmatrix}a\\ b\\ c\end{bmatrix}=(A^tA)^{-1}A^ty=\begin{bmatrix}4\\ -0.2\\ 0.2\end{bmatrix}$$

So the polynomial is $4-0.2x+0.2x^2$.
For $x=12$, its value is 30.4.

So her sales for December will be worth Rs. 30.4 lakhs.

Try an exercise now.

E30) Find the quadratic polynomial which best fits the points (2, 0), (3, −10), (5, −48) and (6, −76).

With this we come to the end of this unit. Let us take a brief look at what we have covered in it.

4.6 SUMMARY

In this unit, we have covered the following points:

1. Any $A \in \mathbf{M}_n(C)$ is unitarily similar to an upper triangular matrix, i.e. A is triangulisable.

2. A is normal if and only if it is unitarily diagonalisable, i.e., it is unitarily similar to a diagonal matrix.

3. $A \in \mathbf{M}_n(C)$ is unitary diagonalisable (and hence normal) iff $\operatorname{tr}(AA^*) = \sum_{i=1}^{n} |\lambda_i|$, where $\lambda_1, \ldots, \lambda_n$ are the eigenvalues of A.

4. The change of orthonormal bases of the matrix of a linear operator on a finite-dimensional inner product space is a similarity transform by a unitary matrix.

5. For a given linear operator on an inner product space, there is an orthonormal basis with respect to which the matrix of the linear operator is upper triangular.

6. If T is a normal operator on a finite-dimensional inner product space, then
   (i) λ is an eigenvalue of T iff $\bar{\lambda}$ is an eigenvalue of T^*, and
   (ii) eigenvectors corresponding to distinct eigenvalues of T are orthogonal.

7. If a linear system $Ax = y$ is inconsistent, it has a least squares solution, which is a solution of the normal system $A^t Ax = A^t y$.

4.7 SOLUTIONS/ANSWERS

E1) Let $u_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ i \\ -1 \end{bmatrix}$. Pick v_2 and v_3 in $\mathbf{C}^3$ such that $\{u_1, v_2, v_3\}$ form a basis of $\mathbf{C}^3$, say, $v_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$, $v_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}$.

Now apply the Gram-Schmidt process to obtain an orthogonal basis $\{w_1, w_2, w_3\}$ from this basis. We get $w_1 = u_1$,

$$w_2 = v_2 - \frac{< v_2, w_1 >}{< w_1, w_1 >} w_1, \; w_3 = v_3 - \frac{< v_3 - w_1 >}{< w_1, w_1 >} w_1 - \frac{< v_3, w_2 >}{< w_2, w_2 >} w_2$$

So, $w_1 = \frac{1}{\sqrt{3}}\begin{bmatrix} 1 \\ i \\ -1 \end{bmatrix}, w_2 = \frac{1}{3}\begin{bmatrix} 2 \\ -i \\ 1 \end{bmatrix}, w_3 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$

Then the unitary matrix is

$$\begin{bmatrix} 1/\sqrt{3} & 2/\sqrt{6} & 0 \\ i/\sqrt{3} & -i/\sqrt{6} & 1 \\ -1/\sqrt{3} & \frac{1}{\sqrt{6}} & 0 \end{bmatrix}.$$

E2) Let $D = \text{diag}(d_1, \ldots, d_n)$.
D is unitary iff $DD^* = I$ iff $d_i \bar{d_i} = 1 \forall i$ iff $\| d_i \| = \pm 1 \forall i$.

E3) U is unitary $\Rightarrow UU^* = I \Rightarrow (\det U)(\overline{\det U}) = 1 \Rightarrow \det U = \pm 1$.

E4) i) For $A \in \mathbf{M}_n(\mathbf{R})$, $A^* = A^t$. Hence the remark.

ii) Let A and B be unitary $n \times n$ matrices.
Then $(AB)(AB)^* = ABB^*A^* = I$.
Similarly, $(AB)^*(AB) = I$.

iii) $U^{-1} = U^*$.
Now $(U^*)(U^*)^* = U^*U = I$
Similarly, $(U^*)^*(U^*) = I$
Hence U^{-1} is also unitary.

E5) The statements are:

i) The product of two orthogonal matrices is orthogonal.

ii) A matrix $A \in \mathbf{M}_n(\mathbf{R})$ is orthogonal iff its columns form an orthonormal basis of the standard inner product space $\mathbf{R}^n$.

iii) The inverse of an orthogonal matrix is orthogonal.

The proofs are on the same lines as the proofs of the 'unitary' cases.

E6) The characteristic equation of A is $t^3 - 10t^2 + 9t = 0$.
So, 9, 1, 0 are its eigenvalues. So, we appeal to Theorem 2.

A unit eigenvector corresponding to $\lambda = 9$ is $\frac{1}{\sqrt{6}}\begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix}$.

Take the unitary matrix $Q_1 = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{3} & 1/\sqrt{2} \\ 1/\sqrt{6} & 1/\sqrt{3} & -1/\sqrt{2} \\ -2/\sqrt{6} & 1/\sqrt{3} & 0 \end{bmatrix}$.

Then $Q_1^t AQ_1 = \begin{bmatrix} 9 & 0 & 2/\sqrt{3} \\ 0 & 0 & -2/\sqrt{6} \\ 0 & 0 & 1 \end{bmatrix}$, which is in triangular form.

E7) $Av = (I - uu^*)v = v$, since $u^*v = 0$.
Also $Au = u - u(u^*u) = \mathbf{0}$.
Now, let $W = u^{\perp} = \{x \in \mathbf{C}^n \mid \langle u, x\rangle = 0\}$. This is a vector space of dimension $n-1$. Let $\{w_2, \ldots, w_n\}$ be a basis.
Then, since each $w_i \perp u$, it is an eigenvector of A.
So, $\{u, w_2, \ldots, w_n\}$ is a l.i. set of eigenvectors of A.
Applying the Gram-Schmidt process we get an orthonormal set of eigenvectors of A, $\{u, u_2, \ldots, u_n\}$.
Then $U = [u, u_2, \ldots, u_n]$ will diagonalise A.

E8) Every element of $\mathbf{M}_n(\mathbf{C})$ is triangulisable. But, e.g. $\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ does not have a set of 2 linearly independent eigenvectors, and hence cannot be diagonalised.

E9) A is triangular, and so is $P^*AP = \begin{bmatrix} 1 & 0 & 4 \\ 0 & 1 & -3 \\ 0 & 0 & 2 \end{bmatrix}$.

Hence the statement is correct.

E10) Let $A = U^*TU$ and $A^* = P^*BP$, where U and P are unitary.
Then $T^* = UP^*BPU^*$
i.e., T^* is unitarily similar to B.

E11) Take, e.g., $A = \begin{bmatrix} 2 & i \\ -i & 0 \end{bmatrix}$ and $B = \begin{bmatrix} i & 0 \\ 0 & 2i \end{bmatrix}$. Since A is Hermitian and B is normal, they are unitarily diagonalisable. But $A + B = \begin{bmatrix} 2+i & i \\ -i & 2i \end{bmatrix}$ is not normal, hence not unitarily diagonalisable.

Similarly, AB is not normal, and hence not unitarily diagonalisable.

E12) The eigenvalues of A are 2, 0. The corresponding normalised eigenvectors are $u_1 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ i \end{bmatrix}$ and $\frac{1}{\sqrt{2}}\begin{bmatrix} i \\ 1 \end{bmatrix}$.
$\therefore U = [u_1 \quad u_2]$ gives $U^*AU = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$.

E13) No. Had they been similar, their eigenvalues would have been the same.

E14) Since A is unitary, it is unitarily diagonalisable.
$\therefore \exists U$ s.t. $A = U^*DU$, where $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, with $\lambda_i \in \mathbf{R}\, \forall i$ (given).

$\therefore D^* = D.$

$\therefore A^* = A.$

E15) Since (i) is orthogonal and (ii) is hermitian, they are unitarily diagonalisable.

E16) Since the (2, 1)th elements of AA^* and A^*A are not the same, $AA^* \neq A^*A$. So A is not unitarily diagonalisable.
The characteristic polynomial of A is
$(x-2-2i)\,[(x^2-(5+2i)\,x+2(1-i)]$, which has 3 distinct roots. Therefore, A is diagonalisable.

E17) Let A be normal and $a_{ij} = 0$ for $i > j$, where $A = [a_{ij}]$. So, $A^* = [b_{ij}]$, where $b_{ij} = \overline{a_{ji}}$.
The (n, n)th elements of AA^* and A^*A must be equal. So, we see that
$|a_{1n}|^2 + |a_{2n}|^2 + \ldots + |a_{n-1\,n}|^2 = 0.$
So, $a_{1n} = 0 = a_{2n} = \ldots = a_{n-1\,n}$
In the same way, you can show that $a_{ij} = 0 \,\forall\, i < j$, so that A is a diagonal matrix.

E18) As in E17, you can see that the sum of the diagonal elements of AA^* is
$\sum_{i,j=1}^{n} |a_{ij}|^2$.
Now, $A = \mathbf{0} \Rightarrow AA^* = \mathbf{0} \Rightarrow tr(AA)^* = 0.$
Conversely, $tr(AA^*) = 0 \Rightarrow \sum_{i,j=1}^{n} |a_{ij}|^2 = 0 \Rightarrow a_{ij} = 0 \,\forall\, i, j.$

E19) Let $B = \{e_1, e_2, e_3\}$. Then $[T]_B = \begin{bmatrix} 1 & 0 & 2 \\ 3 & 4 & 2 \\ 1 & 1 & 1 \end{bmatrix}$.
The eigenvalues of $[T]_B$ are 5, 1, 0.
Corresponding eigenvectors are $u_1 = \frac{1}{3\sqrt{6}}\begin{bmatrix} 1 \\ 7 \\ 2 \end{bmatrix}$, $u_2 = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}$, $u_3 = \frac{1}{\sqrt{6}}\begin{bmatrix} -2 \\ 1 \\ 1 \end{bmatrix}$
Then obtaining an orthogonal basis from $\{u_1, u_2, u_3\}$, we get an orthonormal matrix $V = \{v_1, v_2, v_3\}$, where $v_1 = u_1$, $v_2 = \frac{1}{3\sqrt{3}}\begin{bmatrix} 5 \\ -1 \\ 1 \end{bmatrix}$, $v_3 = \frac{1}{\sqrt{18}}\begin{bmatrix} -1 \\ -1 \\ 4 \end{bmatrix}$.
Then $V^t [T]_B V = \text{diag}(5, 1, 0)$
Take $B' = \{y_1, y_2, y_3\}$, where
$y_1 = v_{11}\,e_1 + v_{21}\,e_2 + v_{31}\,e_3 = v_1$
$y_2 = v_{12}\,e_1 + v_{22}\,e_2 + v_{32}\,e_3 = v_2$
$y_3 = v_{13}\,e_1 + v_{23}\,e_2 + v_{33}\,e_3 = v_3$
Then $[T]_{B'} = V^t [T]_B V$ will be upper triangular.

E20) As we've seen in Theorem 6, it is triangulisable.

E21) We know that $V = \text{Im}\,T \oplus (\text{Im}\,T)^{\perp}$. Now,
$u \in (\text{Im}T)^{\perp}$ if and only if $< Tv, u >= 0$ for all $v \in V$, that is, $< v, T^*u >= 0$ for all $v \in V$, and so $T^*u = \mathbf{0}$. Therefore, $(\text{Im}T)^{\perp} = \text{Ker}\,T^*$, and hence,
$V = \text{Im}T \oplus \text{Ker}\,T^*$.

E22) $(a\,I+T)(a\,I+T)^* = (a\,I+T)(\bar{a}\,I+T^*) = |a|^2\,I + \bar{a}\,T + a\,T^* + TT^*$
$= (a\,I+T)^*(a\,I+T)$.
$\therefore a\,I+T$ is normal $\forall a \in \mathbf{C}$.

E23) First let $T = T^*$. Let λ be an eigenvalue of T and
$Tv = \lambda v \Rightarrow T^*v = \bar{\lambda} v \Rightarrow Tv = \bar{\lambda} v \Rightarrow (\lambda - \bar{\lambda})v = \mathbf{0}$, but $v \neq \mathbf{0}$. So $\lambda = \bar{\lambda}$, i.e., λ is real.
Conversely, if T is self-adjoint, we have already seen that all its eigenvalues must be real.

E24) Let T be unitary, and λ be an eigenvalue of T with $Tv = \lambda v$, $v \neq \mathbf{0}$.
Then $< \lambda v, \lambda v > = < Tv, Tv > = < v, v >$
$\Rightarrow \left(|\lambda|^2 - 1\right) \| v \|^2 = 0 \Rightarrow |\lambda| = 1$.
Conversely, let $\{\lambda_1, \ldots, \lambda_n\}$ be the eigenvalues of T.
Since T is normal, it is unitarily diagonalisable, i.e. $\exists$ an orthonormal basis $B = \{v_1, \ldots, v_n\}$ of V consisting of eigenvectors of T.
Then, for any i, $\| Tv_i \| = \| v_i \|$ since $|\lambda_i| = 1$.
Therefore, T is unitary.

E25) $\text{Ker}\,A = \text{Ker}(A^tA)\left[\because Ax = \mathbf{0} \Rightarrow A^tAx = 0.\right.$
Also, $\left. A^tAx = \mathbf{0} \Rightarrow x^tA^tAx = \mathbf{0} \Rightarrow (Ax)^tAx = \mathbf{0} \Rightarrow Ax = \mathbf{0}.\right]$
So, by the rank-nullity theorem, rank (A) = rank (A^tA).

E26) Note that by E25, rank (A^t) = rank (A) = rank (A^tA), and the column space of A^tA is clearly contained in the column space of A^t. Since the ranks of A^tA and A^t are equal, their dimensions are equal. Hence, the column spaces of A^t and A^tA are equal. Now A^ty, which belongs to the column space of A^t must belong to the column space of A^tA, and hence $A^tAx = A^ty$ is consistent.

E27) This is of the form $Ax = b$, where $A = \begin{bmatrix} 1 & -1 & 2 \\ -1 & 0 & 2 \\ 1 & 0 & 4 \end{bmatrix}$ and $b = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.
Since A is invertible, the system is consistent. The **unique** solution (and hence least squares solution) is $x = A^{-1}b = \left(\frac{-1}{6}\right)\begin{bmatrix} 0 & 4 & -2 \\ 6 & 2 & -4 \\ 0 & -1 & -1 \end{bmatrix}\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \left(\frac{-1}{6}\right)\begin{bmatrix} 2 \\ 4 \\ -2 \end{bmatrix}$.

E28) $A^tA = \begin{bmatrix} 4 & 1 \\ 1 & 3 \end{bmatrix}$, which is invertible.

Hence the unique solution of $A^tAx = A^ty$ is

$$(A^tA)^{-1}A^ty = \frac{1}{11}\begin{bmatrix} 3 & -1 \\ -1 & 4 \end{bmatrix}\begin{bmatrix} 1 & -1 & 0 & 2 \\ 0 & 1 & -1 & 1 \end{bmatrix}\begin{bmatrix} 1 \\ -1 \\ 2 \\ 3 \end{bmatrix}$$

$$= \frac{1}{11}\begin{bmatrix} 24 \\ -8 \end{bmatrix}$$

E29) Rank $A = 3$, Rank $[A \quad y] = 4$.

Therefore the system is not consistent.

$$A^tA = \begin{bmatrix} 4 & -1 & 9 & 2 \\ -1 & 7 & -9 & -5 \\ 9 & -9 & 27 & 9 \\ 2 & -5 & 9 & 10 \end{bmatrix} \text{ and } A^ty = \begin{bmatrix} 1 \\ 2 \\ 0 \\ 2 \end{bmatrix}$$

Applying Gaussian elimination to $A^tAx = A^ty$, we get the least squares solution set as $\left\{\left(\frac{1}{6} - 2t, \frac{2}{3} + t, t, \frac{1}{2}\right)^t \middle| t \in \mathbf{R}\right\}$.

To obtain the element of this of minimum norm, we consider

$$\left(\frac{1}{6} - 2t\right)^2 + \left(\frac{2}{3} + t\right)^2 + t^2 + \frac{1}{4} = 6t^2 + \frac{2t}{3} + \frac{13}{18}.$$

This is minimum for $t = \frac{-1}{18}$

Hence the least squares solution with minimum norm is $\left(\frac{5}{18}, \frac{11}{18}, \frac{-1}{18}, \frac{1}{2}\right)^t$.

E30) Let the polynomial be $a + bx + cx^2$. Then

$$A = \begin{bmatrix} 1 & 2 & 4 \\ 1 & 3 & 9 \\ 1 & 5 & 25 \\ 1 & 6 & 36 \end{bmatrix} \text{ and } y = \begin{bmatrix} 0 \\ -10 \\ -48 \\ -76 \end{bmatrix}.$$

$$A^tA = \begin{bmatrix} 1 & 16 & 74 \\ 16 & 74 & 376 \\ 74 & 376 & 2018 \end{bmatrix}$$

$$(A^tA)^{-1} = \frac{1}{653}\begin{bmatrix} -221 & 124 & -15 \\ 124 & \frac{1729}{18} & \frac{-202}{9} \\ -15 & \frac{-202}{9} & \frac{91}{18} \end{bmatrix}$$

$$\text{So } \begin{bmatrix} a \\ b \\ c \end{bmatrix} = (A^tA)^{-1}A^ty = \frac{1}{653}\begin{bmatrix} -20 \\ 4009 \\ -2049 \end{bmatrix}$$

So, the required polynomial is $\frac{1}{653}(-20 + 4009x - 2049x^2)$.