8-1-1971

On unitary and orthogonal transformations

Rufus Watts
Atlanta University

Follow this and additional works at: http://digitalcommons.auctr.edu/dissertations
Part of the Mathematics Commons

Recommended Citation
ON UNITARY AND ORTHOGONAL TRANSFORMATIONS

A THESIS
SUBMITTED TO THE FACULTY OF ATLANTA UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR
THE DEGREE OF MASTER OF SCIENCE

BY
RUFUS WATTS

DEPARTMENT OF MATHEMATICS

ATLANTA, GEORGIA
AUGUST 1971
ACKNOWLEDGEMENTS

I wish to express my sincere thanks to Dr. N. A. Warsi, of Atlanta University, my thesis advisor, for his assistance and expert direction in the preparation of this thesis.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>i</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td><strong>Chapter</strong></td>
<td></td>
</tr>
<tr>
<td>I. VECTOR SPACES AND MATRICES</td>
<td>3</td>
</tr>
<tr>
<td>Vector Spaces</td>
<td></td>
</tr>
<tr>
<td>Linear Transformations</td>
<td></td>
</tr>
<tr>
<td>Matrices</td>
<td></td>
</tr>
<tr>
<td>Addition and Scalar Multiplication of Matrices</td>
<td></td>
</tr>
<tr>
<td>Some Special Matrices</td>
<td></td>
</tr>
<tr>
<td>Change of Basis and Hermite Normal Form</td>
<td></td>
</tr>
<tr>
<td>Determinants and Similarity Transformations</td>
<td></td>
</tr>
<tr>
<td>Eigenvectors and Eigenvalues</td>
<td></td>
</tr>
<tr>
<td>II. ORTHOGONAL, UNITARY AND NORMAL MATRICES</td>
<td>26</td>
</tr>
<tr>
<td>Orthogonal and Unitary Transformations</td>
<td></td>
</tr>
<tr>
<td>Normal Matrices</td>
<td></td>
</tr>
<tr>
<td>III. DETERMINING A UNITARY (OR ORTHOGONAL) MATRIX OF TRANSITION</td>
<td>35</td>
</tr>
<tr>
<td>BIBLIOGRAPHY</td>
<td>40</td>
</tr>
</tbody>
</table>
INTRODUCTION

Many of the most important applications of mathematics involve what are known as linear methods. The idea of what is meant by a linear method applied to a linear problem or a linear system is so important that it deserves attention in its own right.

Matrix theory, vector analysis, Fourier series, and Laplace transforms are examples of mathematical techniques which are particularly suited for handling linear problems. In order for the linear theory to apply to the linear problem it is necessary that what we have called "inputs and outputs" and "linear systems" be representable within the theory. Therefore we introduce the concept of a vector space. The law of combination which will be defined for vector spaces are intended to make precise the meaning of our vague statement "combining these inputs in various ways". Generally, one vector space will be used for the set of inputs and one vector space will be used for the set of outputs. We also need something to represent the "linear system" and for this purpose we introduce the concept of linear transformation.

The next step is to introduce a practical method
for performing the needed calculations with vectors and linear transformations. We restrict ourselves to the case in which vector spaces are finite dimensional. Here it is appropriate to represent vectors by n-tuples and to represent linear transformations by matrices.

The point is that the concepts of vector spaces and linear transformations are common to all linear methods while linear matrix theory applies only to those linear problems which are finite dimensional. Thus it is of practical value to discuss vector spaces and linear transformations before introducing the formalism of n-tuples and matrices. And, generally, proofs that can be given without recourse to n-tuples and matrices will be shorter, simpler and clearer.

In chapters I and II of this thesis vector spaces, linear transformations and matrices are discussed. Considerable information is given on normal, unitary and orthogonal matrices. Many interesting theorems and their applications are discussed. In chapter III we summarize a complete set of computational steps which will effectively determine a unitary (or orthogonal) matrix of transition for diagonalizing a given normal matrix.
CHAPTER I

VECTOR SPACES AND MATRICES

1.1 VECTOR SPACES

Definition 1.1.1. By a field we mean a non-empty set of elements with two laws of combination, which we call addition and multiplication, satisfying the following conditions:

F1. To every pair of elements \( a, b \in \mathbb{F} \) there is associated a unique element, called their sum, which we denote by \( a + b \).

F2. Addition is associative: \( (a + b) + c = a + (b + c) \).

F3. There exists an element, which we denote by \( 0 \), such that \( a + 0 = a \) for every \( a \in \mathbb{F} \).

F4. For every \( a \in \mathbb{F} \) there exists an element, which we denote by \( -a \), such that \( a + (-a) = 0 \).

F5. Addition is commutative: \( a + b = b + a \).

F6. To every pair of elements \( a, b \in \mathbb{F} \), there is associated a unique element, called their product, which we denote by \( ab \) or \( a \cdot b \).

F7. Multiplication is associative: \( (ab)c = a(bc) \).

F8. There exists an element different from 0, which
we denote by 1, such that $a \cdot 1 = a$ for every $a \in F$.

$F_9$. For every $a \in F$, $a \neq 0$, there exists an element, which we denote by $a^{-1}$, such that $a \cdot a^{-1} = 1$.

$F_{10}$. Multiplication is commutative: $ab = ba$.

$F_{11}$. Multiplication is distributive with respect to addition: $(a+b)c = ac + bc$.

The elements of a field $F$ are called scalars. The rational numbers, real numbers and complex numbers are familiar and important examples of fields. We do not develop the various properties of abstract fields other than the rationals, reals and complex numbers.

**Definition 1.1.2.** A vector space $V$ over $F$ is a set of elements, called vectors, with two laws of combination, called vector addition and scalar multiplication, satisfying the following conditions:

$A_1$. To every pair of vectors $\alpha, \beta \in V$ there is associated a unique vector in $V$ called their sum, which we denote by $\alpha + \beta$.

$A_2$. Addition is associative: $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$.

$A_3$. There exists a vector, which we denote by $0$, such that $\alpha + 0 = \alpha$ for all $\alpha \in V$.

$A_4$. For each $\alpha \in V$ there exists an element, which we denote by $-\alpha$ such that $\alpha + (-\alpha) = 0$.

$A_5$. Addition is commutative: $\alpha + \beta = \beta + \alpha$.

$B_1$. To every scalar $a \in F$ and vector $\alpha \in V$, there is
associated a unique vector, called the product of \( a \) and \( \mathbf{d} \), which we denote by \( a \mathbf{d} \).

B2. Scalar multiplication is associative: \( a(b \mathbf{d}) = (ab) \mathbf{d} \).

B3. Scalar multiplication is distributive with respect to vector addition: \( a(\mathbf{d} + \mathbf{e}) = a\mathbf{d} + a\mathbf{e} \).

B4. Scalar multiplication is distributive with respect to scalar addition: \( (a + b)\mathbf{d} = a\mathbf{d} + b\mathbf{d} \).

B5. \( 1 \mathbf{d} = \mathbf{d} \) (where \( 1 \in \mathbb{F} \)).

Definition 1.1.3. A set of vectors are said to be linearly dependent if there exists a non-trivial linear relation among them. Otherwise the set is said to be linearly independent.

It should be noted that any set of vectors that includes the zero vector is linearly dependent. A set consisting of exactly one non-zero vector is linearly independent. For if \( a \mathbf{d} = 0 \) with \( a \neq 0 \), then \( \mathbf{d} = a^{-1} \mathbf{d} = a^{-1}(a \mathbf{d}) = a^{-1} \cdot 0 = 0 \). Notice also that the empty set is linearly independent.

Definition 1.1.4. A linearly independent set spanning a vector space \( V \) is called a basis of \( V \).

If \( A = \{\mathbf{d}_1, \mathbf{d}_2, \ldots \} \) is a basis of \( V_1 \) then by definition an \( \mathbf{d} \in V_1 \) can be written in the form \( \mathbf{d} = \sum a_i \mathbf{d}_i \), where \( a_i \in \mathbb{F} \) and \( \mathbf{d}_i \in V \). The interesting thing about a basis as distinct from other spanning sets is that the coefficients are uniquely determined by .
A vector space with a finite basis is called finite dimensional and the number of elements in a basis is called the dimension of the space. There are interesting vector spaces with infinite bases. However, it is not our intention to deal with infinite dimensional vector spaces as such. Whenever we speak of the dimension of a vector space without specifying whether it is finite or infinite dimensional we mean that the dimension is finite.

Definition 1.1.6. A subspace W of a vector space V is a non-empty subset of V which is itself a vector space with respect to the operations of addition and scalar multiplication defined in V.

1.2 LINEAR TRANSFORMATIONS

Let U and V be vector spaces over the same field of scalars F.

Definition 1.2.1. A linear transformation $\sigma$ of U into V is a single-valued mapping of U into V which associates to each $\lambda \in U$ a unique element $\sigma(\lambda) \in V$ such that for all $\lambda, \beta \in U$ and all $a, b \in F$, we have

$$\sigma(a\lambda + b\beta) = a\sigma(\lambda) + b\sigma(\beta).$$

The use of the word "into" in the definition of a linear transformation is intended to mean that it is not necessary that every $\lambda \in V$ be the image of some $\lambda \in U$.

Example: Consider the complex plane. The transformation defined by $\sigma(x, y) = (6x, 6y)$ for some fixed
scalar is a linear transformation which maps each point p to a point q collinear with p.

1.3 MATRICES

Definition 1.3.1. A matrix over a field $F$ is a rectangular array of scalars. The array will be written in the form

\[
\begin{bmatrix}
11 & 12 & \ldots & 1n \\
21 & 22 & \ldots & 2n \\
\vdots & \vdots & \ddots & \vdots \\
m1 & m2 & \ldots & mn
\end{bmatrix}
\]

whenever we wish to display all elements in the array or show the form of the array. A matrix with $m$ rows and $n$ columns is called an $m \times n$ matrix. An $n \times n$ matrix is said to be of order $n$.

We often abbreviate a matrix written in the form above to $[a_{ij}]$ where the first index denotes the number of the row and the second index denotes the number of the column. The particular letter appearing in each index is immaterial; it is the position that is important. With this convention $a_{ij}$ is a scalar and $[a_{ij}]$ is a matrix. The $a_{ij}$ appearing in the array $[a_{ij}]$ is called the components or element of $[a_{ij}]$. Two matrices are equal if and only if they have the same components. The
main diagonal of the matrix \([a_{ij}]\) is the set of elements \([a_{11}, \ldots , a_{tt}]\) where \(t = \min\{m,n\}\). A diagonal matrix \(A\) is a square matrix in which the elements not in the main diagonal are zero. A matrix with every element zero is called the zero matrix.

Matrices can be used to represent a variety of different mathematical concepts. The way matrices are manipulated depends upon the objects which they represent. In our discussion we investigate the properties of matrices as representations of linear transformations. Not only do matrices provide us with a convenient means of doing whatever computations which is necessary with linear transformations, but the theory of vector spaces and linear transformations also prove to be a powerful tool in developing the properties of matrices.

1.4 ADDITION AND SCALAR MULTIPLICATION OF MATRICES

Definition 1.4.1. By the sum, \(A+B\), of two \(m \times n\) matrices \(A = [a_{ij}]\) and \(B = [b_{ij}]\) is meant the \(m \times n\) matrix \(C = [c_{ij}]\) where \(c_{ij} = a_{ij} + b_{ij}\).

Example:
\[
\begin{bmatrix}
1 & -1 & 2 \\
3 & 0 & 1
\end{bmatrix} + \begin{bmatrix}
2 & 2 & 2 \\
1 & 0 & -1
\end{bmatrix} = \begin{bmatrix}
3 & 1 & 4 \\
4 & 0 & 0
\end{bmatrix}.
\]

Definition 1.4.2. By the scalar product, \(KA\), of an \(m \times n\) matrix \(A = [a_{ij}]\) and a scalar \(K\) is meant the \(m \times n\) matrix \(C = [c_{ij}]\), where \(c_{ij} = K a_{ij}\).
Example:

\[
2 \begin{bmatrix} 1 & -1 & 3 \\ 3 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 2 & -2 & 4 \\ 6 & 0 & 2 \end{bmatrix}.
\]

Theorem 1.4.1. If \( A, B \) and \( C \) are any \( m \times n \) matrices, then

1. \( 0 \cdot A = 0 \). (The "0" on the left is a scalar, and the "0" on the right is a \( m \times n \) matrix.

2. \( 1 \cdot A = A \).

3. \( A(B + C) = AB + AC \).

4. \( (A + B)C = AC + BC \).

5. \( A(BC) = (AB)C \).

Of course, in each of the above statements we must assume the operations proposed are well defined. For example, in 3, \( B \) and \( C \) must be the same size and \( A \) must have the same number of columns as \( B \) and \( C \) have rows. Also, the set of column matrices form a vector space.

Definition 1.4.3. Let \( X \) be a column (row) vector with elements \( \{x_1, \ldots, x_n\} \) and let \( Y \) be a column (row) vector of the same order with elements \( \{y_1, \ldots, y_n\} \). By the inner product \( X \cdot Y \) of \( X \) and \( Y \) is meant the scalar \( \{x_1y_1 + x_2y_2 + \ldots + x_ny_n\} \).

Example:

If \( X = \{2, -1, 3\} \) and \( Y = \{2, 0, -2\} \), then \( X \cdot Y \)

\[
= 2(2) + (-1)(0) + 3(-2)
= -2.
\]

Definition 1.4.4. Let \( A = [a_{ij}] \) be an \( m \times n \) matrix
and \( B = [b_{ij}] \) be an \( n \times p \) matrix. Since the order of any column \( A_i \) of \( A \) is equal to the order of any row \( B_j \) of \( B \), it is possible to form the inner product \( A_i \cdot B_j \).
By the Cayley product \( AB \) of two matrices \( A \) and \( B \), taken in this order, is meant the \( m \times p \) matrix \( C = [c_{ij}] \), where
\[
c_{ij} = A_i \cdot B_j.
\]
Alternatively, we may say by the Cayley product \( AB \)
\[
c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj}.
\]
Examples:

1. \[
\begin{bmatrix}
1 & 2 & 4 \\
-1 & 3 & -4
\end{bmatrix}
= [x_1 - x_2, 2x_1 + 3x_2, 4x_1 - 2x_2].
\]

2. \[
\begin{bmatrix}
1 & -1 & 2 \\
3 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 2 & 0 \\
0 & -1 & 1 \\
1 & 2 & -1
\end{bmatrix}
= \begin{bmatrix}
3 & 7 & -3 \\
4 & 8 & -1
\end{bmatrix}.
\]

Multiplication of matrices is not always commutative, not even for square matrices of the same order. If \( A \) is an \( m \times n \) matrix and \( B \) is an \( n \times p \) matrix, \( p \neq m \), then \( AB \) is defined but \( BA \) is not. If \( p = m \neq n \), then \( AB \) and \( BA \) are both defined but \( AB \neq BA \), since \( AB \) is a square matrix of order \( m \) and \( BA \) is a square matrix of order \( n \). Even if \( p = m = n \), we do not necessarily have \( AB = BA \), as the following examples testify:

1. \[
\begin{bmatrix}
0 & 1 \\
1 & 0
\end{bmatrix}
\begin{bmatrix}
0 & 0 \\
0 & 0
\end{bmatrix}
= \begin{bmatrix}
1 & 0 \\
0 & 0
\end{bmatrix}.
\]
If $A$ and $B$ are matrices such that $AB = 0$, it does not follow that either $A = 0$ or $B = 0$. We have, for example

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

Definition 1.4.5. The rank of a matrix $A$, denoted $\rho(A)$, is the maximum number of independent rows or columns.

1.5 SOME SPECIAL MATRICES

There are many special matrices that play an important role in various parts of matrix theory and in the development of this thesis. We consider some of these special matrices in this section.

Definition 1.5.1. A square matrix $[a_{ij}]^n$ such that $a_{ij} = 0$ if $i \neq j$ and $a_{ij} = 1$ if $i = j$ is called an identity matrix and will be denoted by $I(n)$.

Example:

$$I(3) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

On the other hand, one must not jump to the conclusion that if $BA = A$, $B$ is an identity matrix. We have, for example

$$\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} 0 & 0 \\ 2 & 3 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 2 & 3 \end{bmatrix}$$
Definition 1.5.2. If A is an m x n matrix, then the n x m matrix whose successive rows are successive columns of A is called the transpose of A and will be denoted by $A^T$. In other words, the $(i,j)$th element of $A^T$ is the $(j,i)$th element of A.

Example:

If $A = \begin{bmatrix} 1 & -2 & 0 \\ 1 & 1 & 2 \end{bmatrix}$, then $A^T = \begin{bmatrix} 1 & -2 & 1 \\ -2 & 1 & 2 \end{bmatrix}$.

Definition 1.5.3. A matrix A is said to be symmetric if $A = A^T$ and is skew-symmetric if $A = -A^T$.

Examples:

\[
\begin{bmatrix} 1 & 2 & 3 \\ 2 & 0 & -4 \\ 3 & -4 & -2 \end{bmatrix}
\] is symmetric

\[
\begin{bmatrix} 0 & 2 & 3 \\ -2 & 0 & -4 \\ -3 & 4 & 0 \end{bmatrix}
\] is skew-symmetric.

Theorem 1.5.1. If A is a matrix, then $AA^T$ is symmetric.

Proof: $(AA^T)^T = (A^T)^TA^T = AA^T$.

Theorem 1.5.2. If A is a square matrix, then $A + A^T$ is symmetric and $A - A^T$ is skew-symmetric.

Proof: $(A + A^T)^T = A^T + (A^T)^T = A^T + A = A + A^T$


We shall consider two important unary operations that can be performed on matrices with complex elements.
We first recall that if \( u = a + ib \) is a complex number, then \( \bar{u} = a - ib \) is known as the complex conjugate of \( u \).

**Definition 1.5.4.** If \( A = [a_{ij}] \) is a matrix with complex numbers as elements, then the matrix \( \bar{A} = [\bar{a}_{ij}] \), where \( \bar{a}_{ij} \) is the complex conjugate of \( a_{ij} \), is called the conjugate. The matrix \( A^* = (\bar{A})^T \) is called the conjugate transpose of \( A \).

**Example:**

If \( A = \begin{bmatrix} 1-i & i & 3 \\ 0 & 1 & 2+i \\ -3i & -4 & 10-2i \end{bmatrix} \), then \( \bar{A} = \begin{bmatrix} 1+i & -i & 3 \\ 0 & 1 & 2-i \\ 3i & -4 & 10+2i \end{bmatrix} \)

and \( A^* = \begin{bmatrix} 1+i & 0 & 3i \\ -i & 1 & -4 \\ 3 & 2-i & 10+2i \end{bmatrix} \).

**Definition 1.5.5.** A matrix \( A \) with complex elements is said to be Hermitian if \( A = A^* \) and skew-Hermitian if \( A = -A^* \). For example, if \( a, b, c, d, e, f, g, h, \) and \( k \) are real numbers and if \( i = -1 \), then

\[
\begin{bmatrix}
a & b+ic & e+if \\
b-ic & d & h+ik \\
e-if & h+if & g
\end{bmatrix}
\]

is Hermitian and

\[
\begin{bmatrix}
ia & b+ic & e+if \\
-b+ic & id & h+ik \\
-e+if & -h+ik & ig
\end{bmatrix}
\]

is skew-Hermitian.
Definition 1.5.6. A matrix \( A \) has an inverse if there exists a matrix \( B \) such that \( AB = BA = I \). Customarily, \( B \) is denoted by \( A^{-1} \).

A matrix that has an inverse is said to be non-singular or invertible; otherwise it is said to be singular. Only a square matrix can have an inverse.

1.6 CHANGE OF BASIS AND HERMITE NORMAL FORM

Let \( A = \{a_1, \ldots, a_n\} \) and \( A' = \{a'_1, \ldots, a'_n\} \) be bases of the vector space \( U \). In a typical "change of basis" situation the representations of various vectors and linear transformations are known in terms of the basis \( A \), and we wish to determine their representation in terms of the basis \( A' \). We, in this connection, refer to \( A \) as the "old" basis and to \( A' \) as the "new" basis. Each \( a_j \) is expressible as a linear combination of the elements of \( A \), that is,

\[
d_j' = \sum_{i=1}^{n} P_{ij} a_i.
\]

The associated matrix \( P = [P_{ij}] \) is called the matrix of transition from the basis \( A \) to \( A' \).

Let there be an \( m \times n \) matrix \( A \) of rank \( \rho \), and let there exist a non-singular \( m \times m \) matrix \( Q \) such that \( A' = Q^{-1}A \) having the following form:

1. There is at least one non-zero element in each of the first \( \rho \) rows of \( A \), and the elements in all remaining rows are zero.

2. The first non-zero element appearing in row i
(1) $i \leq \ell$ is a 1 appearing in column $K_i$, where $K_1 < K_2 \ldots < K_{\ell}$.

(3) In column $K_i$ the only non-zero element is the 1 in row $i$. Then $A$ is said to be in Hermite normal or row-echelon form.

Example:

\[
\begin{bmatrix}
0 & 1 & 0 & 0 & 1 \\
0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0
\end{bmatrix}
\]

1.7 DETERMINANTS, EIGENVALUES AND SIMILARITY TRANSFORMATIONS

We have seen that if $A$ represents a linear transformation $\mathcal{G}$ of $V$ into itself with respect to a basis $A$ and $P$ is the matrix of transition from $A$ to a new basis $A'$, then $P^{-1}AP$ is the matrix representing $\mathcal{G}$ with respect to $A'$. In this case $A$ and $A'$ are said to be similar and the mapping of $A$ onto $A' = P^{-1}AP$ is called a similarity transformation (on the set of matrices, not on $V$).

Given $\mathcal{G}$ we seek a basis for which the matrix representing $\mathcal{G}$ is particularly simple. In practice $\mathcal{G}$ is given only implicitly by giving a matrix $A$ representing $\mathcal{G}$. The problem, then, is to determine the matrix of transition $P$ so that $P^{-1}AP$ has the desired form. The matrix representing $\mathcal{G}$ has its simplest form whenever $\mathcal{G}$ maps each basis vector onto a multiple of itself, that is, whenever for each basis vector $\mathcal{v}$ there exists a scalar $\lambda$ such that
\( \mathcal{G}(\lambda) = \lambda \cdot I \). It is not always possible to find such a basis, but there are some rather general conditions under which it is possible.

We introduce some topics from the theory of determinants solely for the purpose of finding the eigenvalues of a linear transformation. Whenever a basis of eigenvectors exists the use of determinants will provide a method for finding the eigenvalues and, knowing the eigenvalues, use of the Hermite normal form will enable us to find the eigenvectors.

To define determinants and handle them we have to know something about permutations. Accordingly, we introduce permutations in a form most suitable for our purpose.

Definition 1.7.1. A permutation \( \pi \) of a set \( S \) is a one-to-one mapping of \( S \) onto itself.

We are dealing with permutations of finite sets and we take \( S \) to be set of the first \( N \) integers; \( S = \{1, 2, \ldots, N\} \). Let \( \pi(i) \) denote the element which \( \pi \) associates with \( i \). Whenever we wish to specify a particular permutation we describe it by writing the elements of \( S \) in two rows; the first row containing the elements of \( S \) in any order and the second row containing the element \( \pi(i) \) directly below the element \( i \). Thus for \( S = \{1, 2, 3, 4\} \), the permutation \( \pi \) for which \( \pi(1) = 2, \pi(2) = 4, \pi(3) = 3, \) and
$\Pi(4) = 1$, can conveniently be described by

$$\Pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 4 & 3 & 1 \end{pmatrix} \text{ or } \begin{pmatrix} 2 & 4 & 1 & 3 \\ 4 & 1 & 2 & 3 \end{pmatrix} \text{ or } \begin{pmatrix} 4 & 1 & 3 & 2 \\ 1 & 2 & 3 & 4 \end{pmatrix}.$$

Two permutations acting on the same set of elements can be combined as a function. Thus if $\Pi$ and $\sigma$ are two permutations, $\sigma \Pi$ will denote that permutation mapping $i$ onto $\sigma[\Pi(i)]$; $(\sigma \Pi) i = \sigma[\Pi(i)]$. As an example, let $\Pi$ denote the permutation described above and let

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 1 & 3 & 4 & 2 \end{pmatrix}.$$

Then

$$\sigma \Pi = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 2 & 4 & 1 \end{pmatrix}.$$

Notice particularly that $\sigma \Pi \neq \Pi \sigma$.

The permutation that leaves all elements of $S$ fixed is called the identity permutation and will be denoted by $\epsilon$. For a given $\Pi$ the unique permutation $\Pi^{-1}$ such that $\Pi^{-1}\Pi = \epsilon$ is called the inverse of $\Pi$.

For a permutation $\Pi$, let $\text{sgn } \Pi$ denote the $(-1)^k(\Pi)$. "sgn" is an abbreviation for signum and we use the term "sgn $\Pi$" to mean the sign of $\Pi$. If $\text{sgn } \Pi$ is 1 we say that is even; if $\text{sgn } \Pi$ is -1 we say that $\Pi$ is odd.

Let $A = [a_{ij}]$ be a square $n \times n$ matrix. We wish to associate with this matrix a scalar that will in some sense measure the "size" of $A$ and tell us whether or not $A$ is non-singular.

Definition 1.7.2. The determinant of the matrix $A = [a_{ij}]$ is defined to be the scalar $|A| = |a_{ij}|$ computed
according to the rule

$$|A| = |a_{ij}| = \sum_{\pi} (\text{sgn} \pi) a_{1\pi(1)} a_{2\pi(2)} \cdots a_{n\pi(n)},$$

where the sum is taken over all permutations of the elements of $S = \{1, \ldots, n\}$. Each term of the sum is a product of $n$ elements, each taken from a different row of $A$ and from a different column of $A$ and $\text{sgn} \pi$. The number $n$ is called the order of the determinant.

As a direct application of this definition we see that

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}.$$

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31} - a_{11}a_{23}a_{32}.$$

**Theorem 1.7.1.** If $A'$ is the matrix obtained from $A$ by multiplying a row (or column) of $A$ by a scalar $c$, then $|A'| = c|A|$.

**Proof:** Each term of the expansion of $|A|$ contains just one element from each row of $A$. Thus multiplying a row of $A$ by $c$ introduces the factor $c$ into each term of $|A|$. Thus $|A'| = c|A|$.

**Theorem 1.7.2.** If $A'$ is the matrix obtained from $A$ by interchanging any two rows (or columns) of $A$, then $|A'| = -|A|$.

**Proof:** Interchanging two rows of $A$ has the effect...
of interchanging two row indices of the elements appearing in $A$. If $\sigma$ is the permutation interchanging these two indices this operation has the effect of replacing each permutation $\pi$ by $\pi \sigma$. Since $\sigma$ is an odd permutation, this has the effect of changing the sign of every term in the expansion of $|A|$. Therefore, $|A'| = -|A|$. 

Theorem 1.7.3. If $A$ has two equal rows, $|A| = 0$.

Proof: The matrix obtained from $A$ by interchanging the two equal rows is identical to $A$, and yet, by theorem 1.7.2, this operation must change the sign of the determinant. Since the only number equal to its negative is zero, $|A| = 0$.

Definition 1.7.3. For a given pair $i,j$, consider in the expansion for $|A|$ those terms which have $a_{ij}$ as a factor. $|A|$ is of the form $|A| = a_{ij}A_{ij} +$ (terms which do not contain $a_{ij}$ as a factor). The scalar $A_{ij}$ is called the cofactor of $a_{ij}$.

Definition 1.7.4. If $A = [a_{ij}]$ is any matrix and $A_{ij}$ is the cofactor of $a_{ij}$, the matrix $[A_{ij}]^T$ is called the adjoint of $A$.

Now we should consider matrices for which the elements are polynomials. If $F$ is the field of scalars for the set of polynomials in the indeterminate $X$, let $K$ be the set of all rational functions in $X$, that is, the set of all permissible quotients of polynomials in $X$. A
matrix with polynomial components is a special case of matrices with elements in $K$.

From this point of view a polynomial with matrix coefficients can be expressed as a single matrix with polynomials as components. For example

$$\begin{bmatrix} 1 & 0 \\ -1 & 2 \end{bmatrix} x^2 + \begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix} x + \begin{bmatrix} 2 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} x^2+2 & 2x-1 \\ -x^2-2x+1 & 2x^2+1 \end{bmatrix}.$$  

Conversely, a matrix with polynomial elements in an indeterminate $X$ can be expanded into polynomials with matrix coefficients. Since polynomials with matrix coefficients and matrices with polynomial components can be converted into one another, we refer to both types of expressions as polynomial matrices.

Definition 1.7.5. If $A$ is any square matrix, the polynomial matrix $A-xI = C$ is called the characteristic matrix of $A$. $C$ has the form

$$\begin{bmatrix} a_{11}-x & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22}-x & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nn}-x \end{bmatrix}.$$  

The determinant of $C$ is a polynomial $|C| = f(x) = k_n x^n + k_{n-1} x^{n-1} + \cdots k_0$ of degree $n$; it is called the characteristic polynomial of $A$. The equation $f(x) = 0$
is called the characteristic equation of \( A \). We should observe that the coefficient of \( x^n \) in the characteristic polynomial is \((-1)^n\); the coefficient of \( x^{n-1} \) is \((-1)^{n-1} \times \sum_{i=1}^{n} a_{ij}\) and the constant term \( k_0 = |A| \).

1.8 EIGENVALUES AND EIGENVECTORS

Let \( \mathcal{G} \) be a linear transformation of \( V \) into itself. It is often useful to find subspaces of \( V \) in which \( \mathcal{G} \) also acts as a linear transformation. If \( W \) is such a subspace, this means that \( \mathcal{G}(W) \subseteq W \). A subspace with this property is called an invariant subspace of \( V \) under \( \mathcal{G} \). Generally, the problem of determining the properties of \( \mathcal{G} \) on \( V \) can be reduced to the problem of determining the properties of \( \mathcal{G} \) on the invariant subspaces.

In general, a problem of finding those scalars \( \lambda \) and associated vectors \( \xi \) for which \( \mathcal{G}(\xi) = \lambda \xi \), is called an eigenvalue problem.

Definition 1.8.1. A non-zero vector \( \xi \) is called an eigenvector of \( \mathcal{G} \) if there exists a scalar \( \lambda \) such that \( \mathcal{G}(\xi) = \lambda \xi \).

Definition 1.8.2. A scalar \( \lambda \) is called an eigenvalue of \( \mathcal{G} \) if there exists a non-zero vector \( \xi \) such that \( \mathcal{G}(\xi) = \lambda \xi \).

Notice that the equation \( \mathcal{G}(\xi) = \lambda \xi \) is an equation in two variables, one of which is a vector and the other a scalar. The solution \( \xi = 0 \) and \( \lambda \) any scalar is a
solution we choose to ignore since it will not lead to an invariant subspace of positive dimension. Eigenvectors are sometimes referred to as characteristic vectors and similarly eigenvalues as characteristic values.

Theorem 1.8.1. Similar matrices have the same eigenvectors and eigenvalues.

Proof: This follows directly from the definition since eigenvalues and eigenvectors are associated with the underlying linear transformation.

Theorem 1.8.2. Similar matrices have the same characteristic polynomial.

Proof: Let A and A' = P^{-1}AP be similar. Then

$$|A-xI| = |P^{-1}AP-xI| = |P^{-1}(A-xI)P| = |P^{-1}| |A-xI| |P|$$

$$= |A-xI| = f(x).$$

We shall call the characteristic polynomial of any matrix representing \( \mathcal{C} \) the characteristic polynomial of \( \mathcal{C} \). This theorem shows that the characteristic polynomial of a linear transformation is uniquely defined.

Let \( S(\lambda) \) be the set of all eigenvectors of \( \mathcal{C} \) corresponding to \( \lambda \), together with 0.

Theorem 1.8.3. \( S(\lambda) \) is a subspace of \( V \).

Proof: If \( \alpha \) and \( \beta \in S(\lambda) \), then

$$\mathcal{C}(a\alpha + b\beta) = a\mathcal{C}(\alpha) + b\mathcal{C}(\beta)$$

$$= a\lambda \alpha + b\lambda \beta$$

$$= \lambda(a\alpha + b\beta).$$

Hence \( a\alpha + b\beta \in S(\lambda) \) and \( S(\lambda) \) is a subspace.
We shall call $S(\lambda)$ the eigenspace of $\xi$ corresponding to $\lambda$, and any subspace of $S(\lambda)$ will be called an eigenspace of $\xi$. $\lambda$ is a solution of the characteristic equation $f(x) = 0$. Hence $(x-\lambda)$ is a factor of $f(x)$. If $(x-\lambda)^k$ is a factor of $f(x)$ while $(x-\lambda)^{k+1}$ is not, $\lambda$ is a root of $f(x) = 0$ of multiplicity $k$. We refer to this multiplicity as the algebraic multiplicity of $\lambda$. The dimension of $S(\lambda)$ is called the geometric multiplicity of $\lambda$.

Theorem 1.8.4. If the eigenvalues $\lambda_1, \ldots, \lambda_s$ are all different and $\{\xi_1, \ldots, \xi_s\}$ is a set of eigenvectors, $\lambda_i$ corresponding to $\xi_i$, the set $\{\xi_1, \ldots, \xi_s\}$ is linearly independent.

Since we are interested here mainly in numerical procedures, we start with matrices representing the linear transformations and obtain their eigenvalues and the representations of the eigenvectors. For example, let

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

The first step is to obtain the characteristic matrix

$$C(x) = \begin{bmatrix} -1-x & 2 & 2 \\ 2 & 2-x & 2 \\ -3 & -6 & -6-x \end{bmatrix},$$
and then the characteristic polynomial

$$|C(x)| = -(x+2)(x+3)x.$$ 

Thus the eigenvalues of $A$ are $\lambda_1 = -2, \lambda_2 = -3$ and $\lambda_3 = 0$. The next steps are to substitute, successively, the eigenvalues for $x$ in the characteristic matrix. Thus we have

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

The Hermite normal form obtained from $C(-2)$ is

$$\begin{bmatrix} 1 & 2 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$ 

The components of the eigenvector corresponding to $\lambda_1 = -2$ are found by solving the equations

$$x_1 + 2x_2 = 0.$$ 

$$x_3 = 0.$$ 

Thus $(2, -1, 0)$ is the representation of an eigenvector corresponding to $\lambda_1$; for simplicity we shall write $\xi_1 = (2, -1, 0)$, identifying the vector with its representation.

In a similar fashion we obtain

$$C(-3) = \begin{bmatrix} 2 & 2 & 2 \\ 2 & 5 & 2 \\ -3 & -6 & -3 \end{bmatrix}.$$ 

From $C(-3)$ we obtain the Hermite normal form.
\[
\begin{bmatrix}
1 & 0 & 1 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{bmatrix}
\]

and hence the eigenvector \( \xi_2 = (1, 0, -1) \).

Similarly, from
\[
C(0) = \begin{bmatrix}
-1 & 2 & 2 \\
2 & 2 & 2 \\
-3 & -6 & -6
\end{bmatrix}
\]

we obtain the eigenvector \( \xi_3 = (0, 1, -1) \).

By theorem 1.8.4, the three eigenvectors obtained for the three different eigenvalues are linearly independent.
2.1 ORTHOGONAL AND UNITARY TRANSFORMATIONS

Definition 2.1.1. A linear transformation of \( V \) into itself is called an isometry if and only if it preserves length, that is, if and only if \( \| \delta(\omega) \| = \| \omega \| \) for all \( \omega \in V \). An isometry in a vector space over the real numbers is called an orthogonal transformation. An isometry in a vector space over the complex numbers is called a unitary transformation. We try to save duplication and repetition by treating the real and complex cases together whenever possible.

Consider a basis \( A = \{ \omega_1, \ldots, \omega_n \} \) for a vector space for which the representing matrix is the unit matrix. Every set of vectors which has this property is called an orthonormal set. The word "orthonormal" is a combination of the words "orthogonal" and "normal". A vector is normal if it is of unit length and two vectors are orthogonal if \( (\omega, \beta) = (\beta, \omega) = 0 \); \( (\omega, \beta) \) is the scalar product of \( \omega \) and \( \beta \). Thus the vectors of an orthonormal
set are mutually orthogonal and normalized. The basis of \( A \) under consideration is an orthonormal basis.

**Theorem 2.1.1.** An orthonormal set is linearly independent.

**Proof:** Suppose \( \{\xi_1, \xi_2, \ldots, \xi_n\} \) is an orthonormal set and that \( \xi_i x_i \xi_i = 0 \). Then \( 0 = (\xi_j, 0) = (\xi_j, \xi_i x_i \xi_i) = \xi_i x_i (\xi_j, \xi_i) = x_j \). Thus the set is linearly independent.

**Theorem 2.1.2.** If \( A = \{a_1, \ldots, a_s\} \) is any linearly independent set whatever in \( V \), there exists an orthonormal set \( X = \{\xi_1, \ldots, \xi_s\} \) such that \( \xi_k = \sum a_{ik} a_i \).

**Proof:** (The Gram-Schmidt orthonormalization process). Recall that the length of a vector \( \mathbf{d} \) is \( ||\mathbf{d}|| \) where \( ||\mathbf{d}|| \) is a nonnegative real number. Since \( a_1 \) is an element of a linearly independent set \( a_1 \neq 0 \), and therefore \( ||a_1|| > 0 \). Let \( \xi_1 = \frac{1}{||a_1||} a_1 \). Clearly, \( ||\xi_1|| = 1 \).

Suppose, then, \( \{\xi_1, \ldots, \xi_r\} \) has been found so that it is an orthonormal set and such that each \( \xi_k \) is a linear combination of \( \{a_1, \ldots, a_r\} \). Let

\[
d'_{r+1} = a_{r+1} - (\xi_1, a_{r+1}) \xi_1 - \cdots - (\xi_r, a_{r+1}) \xi_r.
\]

Then for any \( \xi_i, 1 \leq i \leq r \), we have

\[
(\xi_i, d'_{r+1}) = (\xi_i, a_{r+1}) - (\xi_i, a_{r+1}) = 0.
\]

Furthermore, since each \( \xi_k \) is a linear combination of the \( \{a_1, \ldots, a_k\} \), \( d'_{r+1} \) is a linear combination of \( \{a_1, \ldots, a_{r+1}\} \). Also, \( d'_{r+1} \) is not zero since \( \{a_1, \ldots, a_{r+1}\} \).
is a linearly set and the coefficient of \( l_{r+1} \) in the representation of \( l'_r+1 \) is 1. Thus we can define

\[
\xi_{r+1} = \frac{1}{\| l'_r+1 \|}.
\]

Clearly, \( \{\xi_1, \ldots, \xi_{r+1}\} \) is an orthonormal set with the desired properties. We can continue in this fashion until we exhaust the elements of \( A \). The set \( X = \{\xi_1, \ldots, \xi_s\} \) has the required properties.

**Definition 2.1.2.** Any matrix is said to be unitary if \( U^*U = I \). Obviously \( U \) is non-singular, its inverse \( U^{-1} \) being \( U^* \). Also, \( UU^* = I \). If \( U \) is real, then it will be unitary if \( U^TU = I \), for in this case \( U^* = U^T \).

**Definition 2.1.3.** A matrix \( U \) is said to be orthogonal if it is real and \( U^TU = I \). Clearly every orthogonal matrix is unitary.

**Theorem 2.1.3.** The product of two unitary matrices of the same order is unitary.

**Proof:** If \( A \) and \( B \) are two unitary matrices of the same order such that

\[
A^*A = AA^* = I
\]

\[
B^*B = BB^* = I.
\]

Then we have \( (AB)^*AB = B^*A^*AB = B^*IB = I \) and according \( AB \) is unitary.

**Theorem 2.1.4.** The inverse of a unitary matrix is unitary.

**Proof:** Let \( A \) be a unitary matrix such that \( AA^* = I \). Taking inverses \( (A^*)^{-1}A^{-1} = I \) so that \( A^{-1} \) is unitary.
Theorem 2.1.5. The set of all n-rowed orthogonal matrices is a group with respect to multiplication.

Proof: Let A and B be any two orthogonal matrices of the same order such that $A^TA = I$, $B^TB = I$. We have $(AB)(AB)^T = (AB)(B^TA^T) = A(B^TB)A^T = AIA^T = I$ so that $AB$ is also an orthogonal matrix. Taking inverses we obtain $AA^T = I$. $(AA^T)^{-1} = I^{-1}$, ie. $(A^T)^{-1}(A^{-1}) = I$ so that $A^{-1}$ is also orthogonal.

2.2 NORMAL MATRICES

It is possible to give a necessary and sufficient condition that a matrix be unitary similar to a diagonal matrix. The real value in establishing this condition is that several important types of matrices do satisfy this condition. This leads to the following definition.

Definition 2.2.1. If all elements of a matrix below the main diagonal are zero the matrix is said to be in superdiagonal form, that is $a_{ij} = 0$ for $i > j$.

Definition 2.2.2. The matrices $A$ and $A'$ are unitary (orthogonal) similar if and only if there exists a unitary (orthogonal) matrix $P$ such that $A' = P^{-1}AP = P*AP$ or $A' = P^{-1}AP = P^TAP$.

Theorem 2.2.1. A matrix $A$ in superdiagonal form is a diagonal matrix if and only if $A*A = AA*$.

Proof: Let $A = [a_{ij}]$ where $a_{ij} = 0$ if $i > j$. Suppose $A*A = AA*$. This means, in particular, that
But since $a_{ij} = 0$ for $i > j$ this reduces to
\[ \sum_{j=1}^{n} |a_{ji}|^2 = \sum_{k=1}^{n} |a_{ik}|^2. \]

Now if $A$ were not a diagonal matrix there would be a first index $i$ for which there exists an index $k > i$ such that $a_{ik} \neq 0$. For this choice of the index $i$ the sum on the left in (2.2) reduces to one while the sum on the right contains at least two non-zero terms. Thus
\[ \sum_{j=1}^{i} |a_{ji}|^2 = |a_{ii}|^2 = \sum_{k=1}^{n} |a_{ik}|^2, \]
which is a contradiction. Thus $A$ must be a diagonal matrix. Conversely, if $A$ is a diagonal matrix then clearly $A^*A = AA^*$.

Definition 2.2.3. A matrix $A$ for which $A^*A = AA^*$ is called a normal matrix.

Definition 2.2.4. A matrix $A$ is similar to a diagonal matrix if and only if there exists $n$ linearly independent eigenvectors of $A$.

Theorem 2.2.2. A matrix $A$ is unitary similar to a diagonal matrix if and only if it is normal.

Proof: If $A$ is normal, then any matrix unitary similar to $A$ is also normal. Namely, if $U$ is unitary,
\[ (U^*AU)^*(U^*AU) = U^*A^*UU^*AU = U^*A^*U = U^*AA^*U \]
Thus if $A$ is normal, the superdiagonal form to which it
is unitary similar is also normal and, hence, diagonal.
Conversely, if $A$ is unitary similar to a diagonal matrix
it is unitary similar to a normal matrix, and it is there-
fore normal itself.

Definition 2.2.5. A matrix $H$ is Hermitian if and
only if $H = H^*$. 

Theorem 2.2.3. Unitary matrices and Hermitian ma-
trices are normal.

Proof: If $U$ is unitary, then
$$U^*U = U^{-1}U = UU^{-1}.$$ 
If $H$ is Hermitian, then
$$H^*H = HH = HH^*.$$ 

Theorem 2.2.4. If there exists an orthonormal basis
consisting of characteristic vectors of a linear trans-
formation $\sigma$, $\sigma^* \sigma = \sigma \sigma^*$. 

Proof: Let $X = \{\xi_1, \ldots, \xi_n\}$ be an orthonormal basis
consisting of characteristic vectors of $\sigma$. Let $\lambda_i$ be
the characteristic value corresponding to $\xi_i$. Then
$$\langle \sigma^*(\xi_i), \xi_j \rangle = \langle \xi_i, \sigma(\xi_j) \rangle = \langle \xi_i, \lambda_j \xi_j \rangle = \lambda_j \xi_i \xi_j = \lambda_i \langle \xi_i, \xi_j \rangle = (\bar{\lambda}_i \xi_i, \xi_j).$$ For a fixed $\xi_i$ this equation holds for all
$\xi_j$ and hence $\langle \sigma^*(\xi_i), \lambda \rangle = (\bar{\lambda}_i \xi_i, \lambda)$ for all $\lambda \in \mathbb{V}$. This
means $\sigma^*(\xi_i) = \bar{\lambda}_i \xi_i$ and $\xi_i$ is an characteristic vector
of $\mathcal{C}^*$ with characteristic value $\lambda_i$. Then $\sigma \sigma^*(\xi_i) = 
abla(\lambda_i \xi_i) = \lambda_i \lambda_i \xi_i = \sigma^* \sigma(\xi_i)$. Since $\sigma \sigma^* = \sigma^* \sigma$ on a basis of $V$, $\sigma \sigma^* = \sigma^* \sigma$ on all of $V$.

Definition 2.2.6. A linear transformation $\sigma$ for which $\sigma^* \sigma = \sigma \sigma^*$ is called a normal linear transformation. Clearly, a linear transformation is normal if and only if the matrix representing it (with respect to an orthonormal basis) is normal.

Theorem 2.2.5. For a normal linear transformation, eigenvectors corresponding to different eigenvalues are orthogonal.

Proof: Suppose $\sigma(\xi_1) = \lambda_1 \xi_1$ and $\sigma(\xi_2) = \lambda_2 \xi_2$ where $\lambda_1 \neq \lambda_2$. Then $\lambda_2(\xi_1, \xi_2) = (\xi_1, \lambda_2 \xi_2) = (\xi_1, \sigma(\xi_2)) = (\sigma^*(\xi_1), \xi_2) = (\lambda_1 \xi_1, \xi_2) = \lambda_1(\xi_1, \xi_2)$. Thus $(\lambda_1 - \lambda_2)$ times $(\xi_1, \xi_2) = 0$. Since $\lambda_1 - \lambda_2 \neq 0$ we see that $(\xi_1, \xi_2) = 0$, that is, $\xi_1$ and $\xi_2$ are orthogonal.

Theorem 2.2.6. A normal linear transformation $\sigma$ is an isometry if and only if all of its eigenvalues are of absolute value 1.

Proof: Suppose $\sigma$ is an isometry. Let $\lambda$ be an eigenvalue of $\sigma$ and let $\xi$ be an eigenvector corresponding to $\lambda$. Then $\|\xi\|^2 = \|\sigma(\xi)\|^2 = (\sigma(\xi); \sigma(\xi)) = (\lambda \xi, \lambda \xi) = |\lambda|^2(\xi, \xi)$. Hence $|\lambda|^2 = 1$.

On the other hand suppose $\sigma$ is a normal linear transformation and that all its eigenvalues are of absolute value 1. Since $\sigma$ is normal there exists a basis $X =$
Let $\xi_1, \ldots, \xi_n$ be the eigenvectors of $C$. Let $\lambda_i$ be the eigenvalue corresponding to $\xi_i$. Then $(C(\xi_1), C(\xi_j)) = (\lambda_i \xi_i, \lambda_j \xi_j) = \lambda_i \lambda_j (\xi_i, \xi_j) = \delta_{ij}$. Hence $C$ maps an orthonormal basis onto an orthonormal basis and is an isometry.

Although the results we state now have been already obtained, they are sufficiently useful to deserve being summarized separately. We are now considering matrices whose entries are complex numbers.

Theorem 2.2.7. If $H$ is Hermitian, then

1. $H$ is unitary similar to a diagonal matrix $D$.
2. The elements along the main diagonal of $D$ are the eigenvalues of $H$.
3. The eigenvalues of $H$ are real.

Conversely, if $H$ is normal and all its eigenvalues real, then $H$ is Hermitian.

Proof: We have already observed that a Hermitian matrix is normal so that (1) and (2) follow immediately. Since $D$ is diagonal and Hermitian $D^* = D$ and the eigenvalues are real.

Conversely, if $H$ is a normal matrix with real eigenvalues, then the diagonal form to which it is unitary similar must be real and hence Hermitian. Thus $H$ itself must be Hermitian.

Theorem 2.2.8. If $A$ is unitary, then

1. $A$ is unitary similar to a diagonal matrix $D$.
2. The elements along the main diagonal of $D$ are the eigenvalues of $A$. 

(3) The eigenvalues of $A$ are of absolute value 1. Conversely, if $A$ is normal and all its eigenvalues are of absolute value 1, then $A$ is unitary.

Proof: We have already observed that a unitary matrix is normal so that (1) and (2) follow immediately. Since $D$ is also unitary, $\overline{DD} = D^*D = I$ so that $|\lambda_i|^2 = \lambda_i^*\lambda_i = 1$, for each eigenvalue $\lambda_i$.

Conversely, if $A$ is a normal matrix with eigenvalues of absolute value 1, then from the diagonal form $D$ we have $D^*D = DD = I$ so that $D$ and $A$ are unitary.
CHAPTER III

DETERMINING A UNITARY (OR ORTHOGONAL) MATRIX OF TRANSITION

We now summarize a complete set of computational steps which will effectively determine a unitary (or orthogonal) matrix of transition for diagonalizing a given normal matrix. Let $A$ be a normal matrix.

1. Determine the characteristic matrix $C(x) = A - xI$.

2. Compute the characteristic polynomial $f(x) = |A - xI|$.

3. Determine all eigenvalues of $A$ by finding all solutions of the characteristic equation $f(x) = 0$. In any but very special or contrived examples this step is tedious and lengthy. In an arbitrarily given example we can find at best only approximate solutions. In that case all of the steps are also approximate. In some applications special information derivable from the peculiarities of the application will give information about the eigenvalues or the eigenvectors without our having to solve the characteristic equations.

4. For each eigenvalue $\lambda_i$ find the corresponding
eigenvector by solving the homogenous linear equations

\[ C(\lambda_i)X = 0. \]

(5) Find an orthonormal basis consisting of eigenvectors of A. If the eigenvalues are distinct, Theorem 2.2.5 assures us that they are mutually orthogonal. Thus all that must be done is to normalize each vector and the required orthonormal basis is obtained immediately.

Even when a multiple eigenvalue \( \lambda_i \) occurs, theorem 2.2.2 assures us that an orthonormal basis of eigenvectors exists. Hence there is no difficulty in obtaining a basis of eigenvectors. The problem is that the different eigenvectors corresponding to the multiple eigenvalue \( \lambda_i \) are not automatically orthogonal; however, that is easily remedied. All we need to do is to take a basis of eigenvectors and use the Gram-Schmidt orthonormalization process. The vectors obtained in this way will still be eigenvectors since they are linear combinations of eigenvectors corresponding to the same eigenvalue. Vectors from different eigenspaces will be orthogonal because of Theorem 2.2.5. Since eigenspaces are seldom of very high dimensions, the amount of work in applying the Gram-Schmidt process is usually quite nominal.

We now give several examples to illustrate the computational procedures and the various diagonalization theorems. Remember that these examples are contrived so that the characteristic equation can easily be solved.
Randomly given examples of high order are likely to result in vexingly difficult characteristic equations.

Example 1, a real symmetric matrix with distinct eigenvalues. Let

\[
A = \begin{bmatrix}
1 & -2 & 0 \\
-2 & 2 & -2 \\
0 & -2 & 3
\end{bmatrix}
\]

We first determine the characteristic matrix,

\[
C(x) = \begin{bmatrix}
1-x & -2 & 0 \\
-2 & 2-x & -2 \\
0 & -2 & 3-x
\end{bmatrix}
\]

and then the characteristic polynomial,

\[
f(x) = \det(C(x)) = -x^3 + 6x^2 - 3x - 10 = -(x+1)(x-2)(x-5).
\]

The eigenvalues are \( \lambda_1 = -1, \lambda_2 = 2, \lambda_3 = 5 \).

Solving the equations \( C(\lambda_i)X = 0 \) we obtain the eigenvectors \( \varphi_1 = (2,2,1), \varphi_2 = (-2,1,2), \varphi_3 = (1,-2,2) \).

Theorem 2.2.5 assures us that these eigenvectors are orthogonal. Normalizing them we obtain the orthonormal basis

\[
X = \left\{ \varphi_1 = \frac{1}{3}(2,2,1), \varphi_2 = \frac{1}{3}(-2,1,2), \varphi_3 = \frac{1}{3}(1,-2,2) \right\}.
\]

The orthogonal matrix of transition is

\[
P = \frac{1}{3} \begin{bmatrix}
2 & -2 & 1 \\
2 & 1 & -2 \\
1 & 2 & 2
\end{bmatrix}.
\]
Example 2, a Hermitian matrix. Let

\[ A = \begin{bmatrix} 2 & 1-i \\ 1+i & 3 \end{bmatrix}. \]

Then

\[ C(x) = \begin{bmatrix} 2-x & 1-i \\ 1+i & 3-x \end{bmatrix}, \]

and \( f(x) = x^2 - 5x + 4 = (x-1)(x-4) = 0 \) is the characteristic equation. The eigenvalues are \( \lambda_1 = 1 \) and \( \lambda_2 = 4 \).

(The example is contrived so that the eigenvalues are rational, but the fact that they are real is assured by Theorem 2.2.7). Corresponding to \( \lambda_1 = 1 \) we obtain the normalized eigenvector \( \xi_1 = 1/\sqrt{3}(-1, 1+i) \). The unitary matrix of transition is

\[ U = 1/\sqrt{3} \begin{bmatrix} -1+i & 1 \\ 1 & 1+i \end{bmatrix}. \]

Example 3, an orthogonal matrix. Let

\[ A = \frac{1}{3} \begin{bmatrix} 1 & -2 & 2 \\ -2 & 1 & 2 \\ -2 & -2 & -1 \end{bmatrix}. \]

This orthogonal matrix is real but not symmetric. Therefore, it is unitary similar to a diagonal matrix but it is not orthogonal similar to a diagonal matrix. We have

\[ C(x) = \begin{bmatrix} 1/3 - x & -2/3 & 2/3 \\ -2/3 & 1/3 - x & 2/3 \\ -2/3 & -2/3 & -1/3 - x \end{bmatrix}, \]
and hence \(-x^3 + \frac{1}{3}x^2 - \frac{1}{3}x + 1 = -(x-1)(x^2 + \frac{2}{3}x + 1) = 0\) is the characteristic equation. Notice that the real eigenvalues of an orthogonal matrix are particularly easy to find since they must be of absolute value 1. The eigenvalues are \(\lambda_1 = 1, \lambda_2 = \frac{-1 + 2\sqrt{2}i}{3},\) and \(\lambda_3 = \frac{-1 - 2\sqrt{2}i}{3}.\) The corresponding normalized eigenvectors are \(\xi_1 = \frac{1}{\sqrt{2}}(1,-1,0), \xi_2 = \frac{1}{\sqrt{2}}(1,1,\sqrt{2}i),\) and \(\xi_3 = \frac{1}{\sqrt{2}}(1,1,-\sqrt{2}i).\) Thus the unitary matrix of transition is

\[
U = \begin{bmatrix}
\frac{1}{\sqrt{2}} & 1/2 & 1/2 \\
-\frac{1}{\sqrt{2}} & 1/2 & 1/2 \\
0 & i/\sqrt{2} & -i/\sqrt{2}
\end{bmatrix}.
\]
BIBLIOGRAPHY
