LINEAR ALGEBRA

Attila Aşkar
Koç University, Istanbul

11.02.2002

Table of Contents

Chapter 1

LINEAR SPACES

1.1 Motivation

A large section of applied mathematics is linear analysis where the entities are elements of either finite dimensional (vector) or infinite dimensional (function) spaces. Similarly the relations between these elements are defined in terms of linear operators either in finite dimensional spaces (matricies) or in infinite dimensional spaces (differential or integral operators). All of the concepts in Mathematics can be built on very simple definitions and are special applications of the same ideas. This section attempts to provide the foundation of the techniques used in algebraic methods, in particular of matrices and systems of algebraic equations. Along with the practical concerns, the presentation here has the additional aims to show that the whole of the material in algebra can be derived from a simple set of axioms and definitions; that there is connection between seemengly different concepts; that the many techniques and ideas of mathematics are special aspects of the same notions as well as to bring economy for introducing the various concepts and techniques by avoiding repetitions in their presentation.

The examples in this section are chosen to be relevant to the material treated in the following sections. The presentation avoids overall general and abstract examples of spaces and operators. For developing intuition into the meaning of the ideas effort will be made to relate the definitions and the results to the familar ones from elementary algebra.

Another motivation for the linear algebraic approach is to familiarize the beginner with the modern language and formalism for further studies as this is the approach of most more advanced books in many areas and not only in mathematics. The development of the ideas follows a constructional approach where mathematics is seen as a language to express our observations. The approach is that this symbolic language is an alternative to the verbal languages. Along with being an economical means of expression, mathematics is the most convenient medium for logical deductions and, thereby its power.

1.2 Sets and linear spaces

In the contact with the external world the first observation is the existence of "objects" and "collection of objects". In more abstract terms, this is the concept of a set.

Definition
Any collection of "objects" (or "things") is a **set**.

It should be noticed that this definition of set has no structure. Once the existence of a set of objects is observed, the next attempt is to establish qualitative relations between the elements. This idea of **establishing relations** is the basic concept of **operation** and the **relations** are the basic concept of "Equations". While there is a significant amount of results that can be derived within the qualitative framework of the **set theory**, soon one is led to the concept of quantities and numbers. The fist simplest relations are to quantitatively say wath happens when one puts together any two elements denoted as $\mathbf{x}$ and $\mathbf{y}$ of the set or what happens if one takes the same element several times. These needs lead to introducing the operations of **addition** and **multiplication by a scalar**. Each of these operations is defined by four axioms obtained by abstraction from the observations in daily life. Given the elements $\mathbf{x}, y, z$,... of a set S, both of these operations are each a way to generate new elements from two or more elements of **S**. The resultant elements may or may not belong to **S**. The result of the addition of $\mathbf{x}$ and $\mathbf{y}$ is called the **sum** of $\mathbf{x}$ and $\mathbf{y}$, and is denoted as $\mathbf{x} + y$. The sum of $\mathbf{x}+y$ with $\mathbf{z}$ is denoted as $(x+y)+z$. Similarly, the result of the **multiplication by a scalar** is denoted as αx. Likewise, the multiplication of $\alpha\mathbf{x}$ by β is denoted as $\beta(\alpha x)$.

Definition (Addition). An operation is called **addition** if it satisfies the following four axioms:
Axiom **A1.** Commutativity:

$$\mathbf{x}+\mathbf{y}=\mathbf{y}+\mathbf{x}$$

Axiom **A2.** Associativity:

$$(\mathbf{x}+\mathbf{y})+\mathbf{z}=\mathbf{x}+(\mathbf{y}+\mathbf{z})$$

Axiom **A3.** Existence of the neutral (or null) element θ :

$$\mathbf{x}+\theta=\mathbf{x}$$

Axiom **A4.** Existence of an inverse (or negative) element $-\mathbf{x}$:

$$\mathbf{x}+(-\mathbf{x})=\theta$$

Definition(Multiplication by a Scalar). For α,β,γ being real numbers, an operation is called **multiplication by a scalar** if it satisfies the following axioms:
Axiom **M1.** Associativity:

$$\alpha(\beta\mathbf{x})=(\alpha\beta)\mathbf{x}$$

Axiom **M2.** Distributivity with respect to scalars:

$$(\alpha+\beta)\mathbf{x}=\alpha\mathbf{x}+\beta\mathbf{x}$$

Axiom **M3.** Distributivity with respect to the elements:

$$\alpha(\mathbf{x}+\mathbf{y})=\alpha\mathbf{x}+\alpha\mathbf{y}$$

Axiom **M4.** Existence of a neutral scalar (or the unity) such that

$$1\mathbf{x}=\mathbf{x}$$

It should be emphasised that the axioms above are abstractions made from the daily experience of counting objects. The axioms bf A1,A2, mean that the order in which the elements are added is immaterial and **A3, A4,** are ways to generate negative numbers from the natural numbers or to define subtraction. Similarly the axioms **M1, M2** and **M3** mean that the respective orders of addition and multiplication, i.e. whether we first add and then multiple or whether we first multiply and then add, is irrelevent. Likewise, **M4** implies that the number one is a neutral element in this multiplication. It must also be realised that the multiplication by a scalar is different from the product of two elements of S. This latter operation is not defined in linear algebra.

Examples

1. Let $(x_1, x_2) + (y_1, y_2) = (x_1, x_2 + y_2)$ be the definition of an operator between the elements of the vectors $x = (x_1, x_2)$ and $y = (y_1, y_2)$ in the plane. This is not a valid definition for addition since the axiom **A1** is violated (Show also that other axioms are also violeted):

$$(x_1, x_2) + (y_1, y_2) = (x_1, x_2 + y_2) \neq (y_1, y_2) + (x_1, x_2) = (y_1, y_2 + x_2).$$

Conversely to the above example, it is easely seen that the operation

$$(x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2)$$

satisfies all the axioms **A1 -A4**.
2. Let

$$\alpha(x_1, x_2) = (\alpha x_1, 0)$$

be the definition of an operation for vectors in a plane. This is not a valid definition for multiplication by a scalar since for $\alpha = 1$

$$1(x_1, x_2) = (x_1, 0) \neq (x_1, x_2)$$

i.e.,**M4** is violated. (Show however that the axioms **M1, M2** and **M3** are satisfied).
Conversely to the above example, it is easily seen that the definition

$$\alpha(x_1, x_2) = (\alpha x_1, \alpha x_2)$$

satisfies all of the axioms **M1-M4**.

Important Consequences of the Axioms

1. The zero element is unique. (Prove this as a consequence of the axioms in stating by assuming that there exist two zero elements).

2. The negative of an element is unique (Prove this as a consequence of the axioms by assuming there exist two negative elements).

3. It can be proved easily that

$$0\mathbf{x} = \theta$$

$$\alpha\theta = \theta$$

$$(-\mathbf{x}) = (-1)\mathbf{x}$$

4. $((...\alpha\mathbf{x} + \beta\mathbf{y}) + \gamma\mathbf{z}) + \delta\mathbf{w}) + ...)$ can, without ambiguity, be written as $\alpha x + \beta y + \gamma z + \delta w + ...$ since by axioms **A1, A2, M1, M2, M3** the order of these operations is immaterial.

5. $\mathbf{x} + (-\mathbf{y})$ can without ambiguity be written as $x - y$ since $\mathbf{x} + (-\mathbf{y}) = \mathbf{x} + (-\mathbf{1x}) = \mathbf{x} - \mathbf{y}$ which is the essence of subtraction.

The axioms above do not guarantee that the results of the two operations on the elements of S do not necessarily generate elements contained in S. However in some cases, the results of the two types of operations are also elements of the original set. In such a case the set is said to be **closed under the considered operation** and these are the most important cases in applications. The idea of closure leads to the concept of a **linear space**. Thus:

Definition. (Linear Space) A set which is closed under the operations of addition and multiplication by a scalar is called a **linear space**. Elements of a linear space are called **vectors**.

Examples:

1. The set of all positive real numbers is not a linear space since the negatives defined by **A4** are not elements of the set. But the set of all real numbers is a linear space.

2. The set of all discontinuous functions is not a linear space since the sum of two discontinuous functions may give a continuous function. But the set of all continuous functions is a linear space.

3. The set of numbers between -1 and 1 is not a linear space since the sum of two elements of this set (e.g.,0.5+0.6) may not be an element of this set (Find also the other violations).

4. The set of polynomials of degree n is not a linear space since the sum of two polynomials of degree n may give a polynomial of degree lees than n (e.g.,$(x^2 + ax + b) + (-x^2 + cx + d)$). But the set of polynomials of degree $\leq n$ is a linear space.

5. The set of all integers is not a linear space since the result of the multiplication by a scalar which is a fraction is not an integer.

6. The set of vectors in the plane whose second component is 1 is not a linear space since the sum of two typical vectors as $\mathbf{x} = (x_1, 1)$ and $\mathbf{y} = (y_1, 1)$ does not give a vector whose second component is 1. But the set of vectors in the plane whose second component is 0 is a linear space.

Definition. (Subspace): A linear space whose elements constitute a subset of a given linear space is called a **subspace** of the latter.

It should be emphasised that all subsets of a given linear space are not necessarily subspaces since they themselves have to satisfy the axioms defining a linear space. As an example consider the set of vectors in the plane that are of length equal to 1. This set is clearly a subset of the plane. However since e.g. the zero element is not contained in this set or since the set is not closed under addition and multiplication by a scalar, it cannot be a linear space. Thus the set of vectors of unit length in the plane is not a subspace of the whole space.

1.3 Basis and representations of vectors

In the previous paragraph the operations of addition of two vectors and the multiplication of a vector by a scalar have been defined and generalised to n vectors. Thus

$$c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_n\mathbf{v_n}$$

is a meaningful expression and is called the **linear combination** of $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n}$. Similarly

$$c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n = \theta$$

is called a **linear relation** or a **linear equation**. Given $v_1, v_2, ...v_n$ one may attempt to determine the values of the $c_1, c_2, ...c_n$ which satisfy this equation. In all cases $c_1 = c_2 = ... = c_n = 0$ is a solution and this solution is called the **trivial solution**. The question therefore is whether (1) has other solutions than the trivial one, i.e. nontrivial solutions. This brings the concept of **linear dependence** or **linear independence** of a set of vectors. Thus

Definition. (Linear Independence) A set of vectors $\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_n}$ is said to be **linearly independent** if the linear relation as in (1) has only the trivial solution $c_1 = c_2 = ... = c_n = 0$ and is said to be dependent otherwise.

Examples.

1. Given $\mathbf{v_1} = (1, 0), \mathbf{v_2} = (0, 1), \mathbf{v_3} = (1, 2)$ the linear relation is:

$$c_1(1, 0) + c_2(0, 1) + c_3(1, 2) = (0, 0)$$

By rearrangement

$$(c_1 + c_3, c_2 + 2c_3) = (0, 0)$$

This equation has nontrivial solutions as

$$c_1 = -c_3$$

$$c_2 = -2c_3$$

for arbitrary c_3. For example

$$c_1 = 1, c_2 = 2, c_3 = -1$$

is one such nontrivial solution so that the set $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is linearly dependent.

2. Given $\mathbf{v}_1 = (1,0), \mathbf{v}_2 = (0,1)$ as the set however, there is only a trivial solution and the set $\mathbf{v}_1, \mathbf{v}_2$ is linearly independent.

3. Given $\mathbf{v}_1 = \cos^2 x, \mathbf{v}_2 = \sin^2 x, \mathbf{v}_3 = \cos 2x$, construct he linear relation

$$c_1 \cos^2 x + c_2 \sin^2 x + c_3 \cos 2x = 0$$

for studying the linear dependence of the set. Since

$$\cos 2x = \cos^2 x - \sin^2 x$$

it is seen that above there exists a nontrivial solution as

$$c_1 = -c_3, c_2 = c_3$$

for arbitrary c_3. For example $c_1 = 1, c_2 = -1, c_3 = -1$ is such a nontrivial solution. Thus the set $\cos^2 x, \sin^2 x, \cos^2 x$ is linearly dependent.

4.Given the set of functions $\{1, x, x^2\}$, the linear relation

$$c_1 + c_2 x + c_3 x^2 = 0$$

cannot have nontrivial solutions for c_1, c_2, c_3 for all values of x. Thus the set $\{1, x, x^2\}$ is linearly independent. Let us now consider a set of vectors $\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_p$ (not necessarily linearly independent) and all possible vectors y obtained as

$$\mathbf{y} = c_1 \mathbf{v_1} + c_2 \mathbf{v_2} + ... + c_p \mathbf{v_p}$$

The set of vectors $\mathbf{y}$ obtained by varying the c_i is called the **linear span** of $\{\mathbf{v_1}, \mathbf{v_2}, ..., \mathbf{v_p}\}$ For example, let

$$\mathbf{v}_1 = (1,0), \mathbf{v}_2 = (0,1), \mathbf{v}_3 = (1,2)$$

The linear span of $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\}$ is the set of vectors

$$\mathbf{y} = c_1(1,0) + c_2(0,1) + c_3(1,2)$$

or

$$\mathbf{y} = (c_1 + c_3, c_2 + 2c_3)$$

This set covers the whole plane.

At this stage, a question arises about the smallest set of vectors that are sufficient to generate to elements of the same linear space. In the example above, any two elements of the set e.g., $\mathbf{v_1}$ and $\mathbf{v_2}$ are sufficient to generate all of the vectors in the plane

$$\mathbf{y} = c_1(1,0) + c_2(1,2) = (c_1 + c_2, 2c_2)$$

Therefore, it would be convenient to chose the smallest set of vectors that is sufficient to represent the elements of a given linear space.

Definition (Basis). Any set of vectors with the least number of elements that spans a given linear space is called a **basis** for this space.

Definition (Dimension). The number of elements of a basis is called the **dimension** of the space. When this number is finite, the space is said to be **finite dimensional** and **infinite dimensional** otherwise.

It should be remarked that the basis of a linear space is not unique whereas the dimensionality d is. Moreover, any set of d linearly independent vectors can be taken as a basis. In the example above $\{\mathbf{v_1}, \mathbf{v_2}\}$ or $\{\mathbf{v_1}, \mathbf{v_3}\}$ or $\{\mathbf{v_2}, \mathbf{v_3}\}$ can be taken as a basis since any of these is sufficient to generate all of the vectors in the plane. Similarly, a single vector does not suffice to generate all the vectors in the plane. Thus the dimensionality of the plane is 2.
Similarly the set of vectors $\{1, x, x^2\}$ or $\{1 + x, 1 - x, x^2\}$ are each sufficient to generate the set of polynomials of degree ≤ 2 i.e. P_2 and that no set of two elements can generate all of the elements of P_2.
When the space is infinite dimensional, it is not always to see whether a given set of infinitely many independent vectors is a basis. For example the set of vectors $\{\cos nx : n = 0, 1, 2, ...\}$ has an infinity of elements. However this set is not a basis for the space of continuous functions since any linear combination of the $\{\cos nx\}$ is an even function where as the space of continuous functions in includes also the odd functions. The investigation of the completeness of an infinite set to be a basis for an infinite dimensional function spaces requires advanced analysis connected to the spectral theory of differential operators. From the foregoing discussion it follows that any vector of a linear space can be expressed as a linear combination

of the elements of a basis for a basis for that space. Furthermore, given a basis $\{\mathbf{v_1}, \mathbf{v_2}, ...\mathbf{v_n}\}$ and a vector $\mathbf{x}$, the representation

$$\mathbf{x} = c_1\mathbf{v_1} + c_2\mathbf{v_2} + ... + c_n\mathbf{v_n}$$

is unique. In fact, let there be another representation with coefficients, as

$$\mathbf{x} = c_1'\mathbf{v_1} + c_2'\mathbf{v_2} + ... + c_n'\mathbf{v_n}$$

Subtracting both sides of above equations, one gets.

$$\theta = (c_1 - c_1')\mathbf{v_1} + (c_2 - c_2')\mathbf{v_2} + ... + (c_n - c_n')\mathbf{v_n}$$

On the other hand the basis $\{\mathbf{v_1}, \mathbf{v_2}, ...\mathbf{v_n}\}$ is by definition an independent set and the linear relation can have only a trivial solution:

$$(c_1 - c_1') = 0, (c_2 - c_2') = 0, ..., (c_n - c_n') = 0$$

Which proves that $c_i' = c_i$. Thus, we can state the theorem below.

Teorem. (Uniqueness of Representation): The representation of a vector $\mathbf{x}$ in terms of a basis $\{\mathbf{v_1}, \mathbf{v_2}, ...\mathbf{v_n}\}$ as

$$\mathbf{x} = x_1\mathbf{v_1} + x_2\mathbf{v_2} + ... + x_n\mathbf{v_n}$$

is unique. Because of the uniqueness of the representation of a vector it is worthwhile to give a name to the set of scalars in the linear combination. Thus

Definition. (Components). The unique scalars $x_i, i = 1, ..., n$ in the representation of the vector x as a linear combination of the elements of the basis $\{\mathbf{v_1}, ...\mathbf{v_n}\}$ as

$$\mathbf{x} = x_1\mathbf{v_1} + x_2\mathbf{v_2} + ... + x_n\mathbf{v_n}$$

are called the **components** of $\mathbf{x}$ with respect to the basis $\{\mathbf{v_1}, ...\mathbf{v_n}\}$. It must however be emphasised that the components of a vector are unique with respect to a given basis. For example, the vector $\mathbf{x} = (2, 3)$ has the components $x_1 = 2$ and $x_2 = 3$ with respect to the basis

$$\mathbf{v_1} = (1, 0), \mathbf{v_2} = (0, 1)$$

The same vector has however the components $x_1' = 1$ and $x_2' = 1$ with respect to the basis

$$\mathbf{v_1'} = (1, 1), \mathbf{v_2'} = (1, 2)$$

Similarly the function

$$f(x) = 3 + x^2$$

in the space P_2 has the components $x_1 = 3, x_2 = 0$ and $x_3 = 1$ with respect to the basis

$$v_1 = 1, v_2 = x, v_3 = x^2$$

The same function has however the components $x'_1 = 1, x'_2 = 1$ and $x'_3 = 1$ with respect to the basis

$$x'_1 = 1 + x, x'_2 = 1 - x, x'_3 = 1 + x^2$$

Chapter 2

LINEAR OPERATORS AND EQUATIONS

In the preceeding chapter relations between various elements of a space have been discussed and the operators were defined between elements of the same space. However there are instances where it becomes necessary (and useful too) to relate elements belonging each to a different space. This is the idea of establishing a correspondence between an element of a space and element of another space.

2.1 Linear operators

Definition. (Mapping or Transformation or Operator). Any function f such as $B = f(A)$ where A and B are elements of some space is called a trasformation, mapping or operator.

Definition.(Domain) The set of elements on which the operator acts is called the **domain of the operator**.

Definition. (Target Space) The linear space where the mappings $T(x)$ for all x in the domain belong is called the **tagret space of the operator**.

Definition. (Range). The subset of the target space covered by the mappings $T(x)$ for all x in the domain is called the **range of the operator**.

Figure 1. Domain, target space, null space and range associated with an operator.

It should be noticed that the distinction between the latter two definitions stresses that a mapping doesn't necessarily cover all of the target space. The subset of the target space that is covered by the mapping of the domain by the transform T is denoted by T:D (or R(T)). If the target space consists of T:D alone, the operator T said to map D **onto** . Otherwise i.e., if there are elements of Y which do not belong to T:D, T is said to map D **into** Y. The study of operators can be carried further without any restrictions on the type of operators. However the simplest and perhaps the most useful class of operators, called **linear** occur in all branches of mathematics. The properties of more general ones are often obtained by approximation than by linear transformations.

Definition. (Linear Operator): An operator is said to be **linear** if it has the following properties for all elements of the domain:

$$T(x + y) = T(x) + T(y)$$

$$T(\alpha x) = \alpha T(x)$$

The meaning of linearity conveyed by the definition is that whether the sum or the multiplication is performed in the domain or the range, the result comes out to be same. Said in different words, linearity means that the order of doing the calculations is unimportant. This idea is illustrated in Fig.2.

Figure 2. Schematic representation of the linearity of operators. Equivalent to the pair of equations above, the linearity of an operator may also be defined by the combined relation:

$$T(\alpha x + \beta y) = \alpha T(x) + \beta T(y)$$

Examples.

(1)The Trasformation

$$T(x) = T(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3)$$

is a linear trasformation, since T

$$T(x+y) = T(x_1+y_1, x_2+y_2, x_3+y_3) = (x_1+y_1+x_2+y_2, x_2+y_2+x_3+y_3) =$$

$$(x_1 + x_2, x_2 + x_3) + (y_1 + y_2, y_2 + y_3) = T(x) + T(y)$$

$$T(\alpha x) = T(\alpha x_1, \alpha x_2, \alpha x_3) =$$

$$(\alpha x_1 + \alpha x_2, \alpha x_2 + \alpha x_3) = \alpha T(x)$$

(2) The Transformation $T(x) = x^2$ is not a linear trasformation, since,

$$T(x + y) = (x + y)^2 = x^2 + y^2 + 2xy =$$

$$T(x) + T(y) + 2xy \neq T(x) + T(y)$$

$$T(\alpha x) = (\alpha x)^2 = \alpha^2 x^2 = \alpha^2 T(x) \neq \alpha T(x)$$

In the study of linear spaces, it was seen that the zero element is a special one as being a neutral element. Its mapping in the range and the set of points in the domain that are mapped onto the zero of range are also special. Thus,

Definition. (Null Space) The set of elements of the domain that are mapped onto the zero of the range constitute a linear subspace of the domain which is called the **null space** of the operator.

Examples.

(1) For the linear transformation

$$T(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3)$$

the null space consists of the elements with

$$x_1 + x_2 = 0, x_2 + x_3 = 0$$

This set forms a linear space of the vectors with components

$$x = (x_1, -x_1, x_1).$$

(2) For the linear operator $T(f(x)) = \frac{d^n f}{dx^n}$ the null space consists of the function $f(x)$ such that $\frac{d^n f}{dx^n} = 0$ the null space is P_{n-1}

(3) For the operator T such that

$$T(f(x)) = \int_{-1}^{1} \mid f(x) \mid dx$$

the null space contains only the zero element.

(4) The null space of the linear operator T such that

$$T(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3, x_3 + x_1)$$

contains only the zero element, since the system

$$x_1 + x_2 = 0,$$

$$x_2 + x_3 = 0,$$

$$x_3 + x_1 = 0$$

has only one solution:$x_1 = x_2 = x_3 = 0$, i.e. $x = \theta$
Some operators establish a unique correspondence between pairs of elements with each belonging respectively to the domain and the range. In other words the mapping may be such that it associates only one element of the domain to an element of the range. Thus,

Definition (one-to-one mapping):
A mapping T is said to be one-to-one if for each element of the range there exists at most one element of the domain,i.e. $T(x) = T(y)$ implies $x = y$
Connected with this concept, there is a quete powerful and simple theorem that allows to determine whether a given linear operator is in fact a one-to-one mapping. Thus,

Teorem. (One-to-one mapping):
If the null space of a linear operator consists of the zero element alone, then the mapping is one-to-one and vice versa.

Proof: Let there be two elements x and y of the domain that are mapped onto the same element z of the range: $T(x) = z$ and $T(y) = z$. Subtracting the two equations side by side while keeping the linearity of T, one obtains:

$$T(x - y) = \theta$$

This result means that the vector $x - y$ is in the null space of T, which by hypothesis contains no element other than the θ. Thus $x - y = \theta$ or $x = y$ which contradicts the initial assumption.

Conversely, if the mapping is one-to-one, for $T(x) = z, T(y) = z$ implies that $x = y$ subtracting side by side gives $T(\theta) = \theta$ for all possible x and y which proves that the null space can contain no other element than θ.
Before going into the study of certain general properties of operators it is worthwhile to motivate this study by remarking the relationship between mapping and equations. An important part of the mathematical sciences deals with the solutions of equations of the form $A(x) = b$ where A may be an algebraic operator (i.e.,a matrix), a differential operator or an integral operator; x is an unknown vector in the former case and an unknown function in the latter two cases and b is a is a known vector or a known function accordingly. Consequently, solving an equation amonts to determining the vector or function x whose mapping b is given. It should also be remarked that this process is in the reserve order of the idea of mapping. A special but important class of equations, called homogeneous equations, corresponds to the case when $b = \theta$ i.e., $A(x) = \theta$. Thus, the solution of the problem amounts to determining the null space of the mapping A.

2.2 Algebra of linear operators

In the study of linear spaces, certain operations, namely addition and multiplication by a scalar have been defined. Similar operations can also be defined for linear operators. However in these definitions, it should always be kept in mind that the definition of an operator becomes complete only when its domain and range are also specified and that an operator is defined by its action on an element. Thus, the sum of two product of two operators are defined by their action on test vectors in their domain and these operations require certain compatibility conditions for their definition. The sum and the product of two operators are defined below and illustrated in Figures 3 and 4.

Definition (Sum of two operators).
The sum R of two linear operators T and S is defined as

$$R(x) = (T + S)(x) = T(x) + S(x)$$

Figure 3. Schematic representation of the sum of two operators. A necessary condition for this definition to be meaningful is that T and S have the same domain and range.

Definition (Zero or Neutral Operator).

The operator O such that $O(x) = \theta$ for all x in its domain is called the **zero** or **neutral operator.**
The following consequences of the definitions above are deduced readily. The proofs are left to the reader.

Theorem: (Commutativity in Addition)

$$T + S = S + T$$

Theorem: (Linearity of the Sum of Two Operators)

The sum of two linear operators is a linear operator.

Theorem: (Neutral Operator):
For O being the zero operator

$$T + O = O + T = T$$

for all T compatible with it.

The neutral operator allows to define the negative operator S for a given T as: $T + S = 0$.

Example.
For the linear operators T and S as:

$$T(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3)$$

$$S(x_1, x_2, x_3) = (3x_1 - 2x_2, x_1 + x_3)$$

the sum operator $R = T + S$ is defined as:

$$R(x_1, x_2, x_3) = T(x_1, x_2, x_3) + S(x_1, x_2, x_3) = (4x_1 - x_2, x_1 + x_2 + 2x_3)$$

Similar to addition of two operators, the product is also defined by the action of the operators on a test element. This operation also requires certain compatibility conditions. Thus,

Definition.(Product of Two Operators).
The product R of two linear operators R and S is defined as

$$R(x) = (TS)(x) = T(S(x)).$$

This definition is illustrated in Figure 4. A necessary condition for this definition to be meaningful is that T admits the range of S as its domain. Furthermore, from the definition there is no reason for the product to be commutative. Thus, $TS \neq ST$, in general.

Figure 4.

Schematic representation of the product of operators.

The following consequences of the definitions above are deduced realite. The proofs are left to the reader.

Theorem. (Linearity of the Product Operator). The product TS of two linear operators T and S is also a linear operator.

Example.
For the linear operators

$$T(x_1, x_2, x_3) = (x_1 + x_2, x_2 + x_3)$$

$$S(y_1, y_2) = (2y_1 - y_2, y_2 + 3y_2)$$

the product $R = ST$ is obtained as

$$R(x_1, x_2, x_3) = S(T(x_1, x_2, x_3)) =$$

$$S(x_1 + x_2, x_2 + x_3) =$$

$$2((x_1 + x_2) - (x_2 + x_3), (x_2 + x_3) + 3(x_2 + x_3))$$

$$(2x_1 + x_2 - x_3, x_1 + 4x_2 + 3x_3)$$

On the other hand, the product TS is incompatible since the range of S is E_2 which is not the domain of T and thus TS is undefined.

As for the operation of addition, it is essential to define a neutral operator under product, i.e. an identity operator, in order to introduce the concept of an operator. Thus,

Definition. (Identity Operator).

The operator I such that $I(x) = x$ for all x in its domain is called the **identity operator**.

An immediate consequence of this definition is that the domain and the range of I must be of the same structure since both admit the same x as an element.
Based on the definition of the identity operator, the inverse of a linear operator T can be defined as an operator such that its product with T gives the identity operator. However, unlike for addition, the product being non-commutative in general, the inverse depends on whether it multiplies the original operator on the left or the right.
Definition. (Right Inverse).
The operator R such that $TR = I$ is called the right inverse of T.

Definition. (Left Inverse). The operator L such that $LT = I$ is called the left inverse of T.

Although the existence of the left and their uniqueness are somewhat confusing concepts, the following theorem, given without proof, is a powerful criterion to this respect. Thus,

Theorem. (Uniqueness of the Inverse).
If a linear operator T is a one-to-one mapping, it has a unique left inverse which is at the same time a unique inverse.

In this case, the inverse is indicated by T^{-1} and $T^{-1}T = TT^{-1} = I$.
A concept closely related to product is that of power. This involves the product of an operator by itself several times. By the definition of the product, for the compatibility of the product of an operator T with itself, it follows that the domain and the range must have the same structure. Thus,

Definition. (Power of an Operator). For n a positive integer and where the product is defined:

$$T^{n+1} = T^n T$$

$$T^0 = I$$

The case of n being a positive integer follows by simple induction. However, the case of n being negative also brings no additional complication. A negative integer power $T^{-n}(n > 0)$ is to be interpreted as the n^{th} power of T^{-1}, if this latter exists.

Examples.
1. For T such as

$$T(x_1, x_2) = (x_1 + x_2, x_1 - x_2)$$

the operator T^2 is

$$T^2(x_1, x_2) = T(T(x_1, x_2)) = T(x_1 + x_2, x_1 - x_2) = (2x_1, 2x_2)$$

2. For T such as

$$T(f) = \frac{df}{dx} + f$$

the operator T^2 is

$$T^2(f) = T(T(f)) = T(\frac{df}{dx} + f) = \frac{d}{dx}(\frac{df}{dx} + f) + \frac{df}{dx} + f =$$

$$\frac{d^2 f}{dx^2} + 2\frac{df}{dx} + f$$

Notice that T is a first order linear differential operator. Its square is also a linear operator, but is a second order linear differential operator.

2.3 Linear transformation represented by matrices

Consider a linear operator A such as: $A(x) = y$, where x and y are elements of finite dimensional spaces with dimensions m and n respectively Let $e_1, e_2, ..., e_m$ be a basis for the domain and $f_1, f_2, ..., f_n$

for the range of A. Thus:

$$x = x_1 e_1 + x_2 e_2 + ... + x_m e_m = \sum_{j=1}^{m} x_j e_j$$

$$y = y_1 f_1 + y_2 f_2 + ... + y_n f_n = \sum_{i=1}^{n} y_i f_i$$

Since the mapping $A(x) = y$ above is defined for all x in the domain, particular the mapping of each e_i is expressible in terms of the f_j. Thus

$$A(e_j) = \sum_{i=1}^{m} a_{ij} f_j, \quad j = 1, 2...m$$

Due to the linearity of A,

$$A(x) = \sum_{j=1}^{m} x_j A(e_j)$$

The substitution of $A(e_j)$ from the representation above yields:

$$\sum_{j=1}^{m} x_j \sum_{i=1}^{m} a_{ij} f_j = \sum_{i=1}^{n} y_i f_i$$

Since m and n are finite, the order of summatiom may always be interchanged and the above equation yields:

$$\sum_{j=1}^{m} \{\sum_{i=1}^{m} (a_{ij} x_j - y_i)\} f_i = 0$$

Since f_i is a basis, the linear relation above implies each of coefficient to be zero. Thus:

$$\sum_{i=1}^{m} a_{ij} x_j = y_i, i = 1, 2, ...n$$

The importance of the representation of the mapping $A(x) = y$ in terms of scalars as above lies in the fact that the linear operator A may be represented by the two-dimensional scalar array $\{a_{ij}\}$. This array is called as the matrix representation of the linear operator A or simply the matrix A. The scalars $\{a_{ij}\}$ are called the components of the matrix denoted as A. Their interpretation is clear: $\{a_{ij}\}$ is the i^{th} component of the transformation of e_i in the basis $\{f_i\}$. The possibility of representing the operator A by a set of scalars $\{a_{ij}\}$ brings great simplifications in applications. The scalar are arranged as a two-dimensional array with the first index indicating the row and the second the column where it is to be placed. Pictorially, the matrix is represented with $A(e_j)$ being its j^{th} column:

$$\begin{bmatrix} A(e_1) & A(e_2) & \cdots & A(e_m) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m}) \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

The importance of matrices in applied mathematics lies in the fact that they arise in
(a) algebraic equations with a finite number of unknowns;
(b) difference equations;
(c) approximations to differential and integral operators;
(d) systems of differential equations;
(e) coordinate transformations;
(f) probability theory;
(g) input-output analysis;
(h) linear programming;
(i) Markov chains, to name a few areas.

2.4 Determinants

For square matrices, an additional quantity is defined: the determinant. Eventually, practical rules will be deduced for calculating the determinant of a square matrix. It seems worthwhile however to deduce those practical rules from a basic definition of determinant rather than giving them as a set of empirical rules.

Definition. (determinant of a Square Matrix)
The scalar $d, d = (A_1, A_2, ...A_n)$ defined by the following axioms is called the **determinant** of the square matrix A with the n - dimensional row vectors $A_1, A_2, ...A_n$

Axiom D1: Homogeneity in each row:

$$d(A_1, ..., tA_k, ..., A_n) = td(A_1, ..., A_k, ..., A_n)$$

Axiom D2: Additivity in each row:

$$d(A_1, ..., A_k + C, ..., A_n) = d(A_1, ..., A_k, ..., A_n) + d(A_1, ..., C, ..., A_n)$$

Axiom D3: Zero determinant for equal rows:

$$d(A_1, ..., A_k, A_p, ..., A_n) = 0$$

if $A_k = A_1$

Axiom D4: Determinant for the unit matrix:

$$d(I_1, I_2, ..., I_n) = 1,$$

where $I_k = (0, ..., 1, ...0)$.

Some Immediate Consequences of the Axioms.

(a) The determinant of a square matrix is unique.
(b) The determinant is zero if any row is the zero vector;
i.e. $d(A_1, ..., 0, ..., A_n) = 0$
(c) The determinant changes sign if any two rows are interchanged;
i.e.

$$d(A_1, ..., A_k, A_p, ..., A_n) = -d(A_1, ..., A_p, A_k, ..., A_n)$$

(d) The determinant doesn't change by adding a multiple of a row to any other; i.e.

$$d(A_1, ..., A_k + \alpha A_p, ..., A_p, A_n) =$$

$$d(A_1, ..., A_k, A_p, ..., A_n)$$

(e) The determinant of a diagonal matrix is the product of its diagonal elements.

(f) The determinant of a triangular matrix is the product of its diagonal elements.

(g) The determinant of the product of two matrixes is equal to the product of their determinants; i.e. $\det(\mathbf{AB}) = (\det\mathbf{A})(\det\mathbf{B})$.

(h) If A is nonsingular, then $\det A \neq 0$ and $\det\mathbf{A}^{-1} = (\det\mathbf{A})^{-1}$ (proof: If $\mathbf{A}^{-1}$ exists, $\mathbf{AA}^{-1} = \mathbf{I}$ The theorem follows by taking the determinants and using(**g**)).

Axiom D_1 and results **(c)** and **(d)** above are called the **elementary operations**. These operational properties of determinants suggest a practical method called the Gauss-Jordan process. This process is a systematic way of reducing a matrix to its triangular form through the use of the elementary operations.

2.5 Rank of a matrix and systems of equations

In the representation $y = A(x)$, we saw by the linearity of A that y is of the form

$$y = x_1 A(e_1) + x_2 A(e_2) + ... + x_m A(e_m)$$

Hence typical vector y in the range of A is expressed as a linear combination of $\{A(e_j)\}$. This means that the range of A is the linear span of the set of vectors $\{A(e_j)\}$. Therefore the rank of A, i.e. the dimension for the range of A is equal to the number of linearly independent vectors in this set. On the other hand, it is seen that the vectors constitute the columns in the matrix representation of A. Cast in the matrix language, the rank of a matrix is therefore equal to the number of its linearly independent columns.
A common application of matrices is in systems of linear equations. Indeed, the linear mapping $A(x) = y$ is called an equation when $y = b$ is given. Written in the matrix representation, the linear mapping $A(x) = b$ becomes a system of algebraic equations for the unknowns $x_1, x_2, ..., x_m$ of the form:

$$a_{11}x_1 + a_{12}x_2 + ... + a_{1m}x_m = b_1$$

$$a_{21}x_1 + a_{22}x_2 + ... + a_{2m}x_m = b_2$$

$$\cdots\cdots$$

$$a_{n1}x_1 + a_{n2}x_2 + ... + a_{nm}x_m = b_n$$

The above system of equation may be written in compact form as $A.x = b$ where the matrix A has the elements a_{ij}, the vectors x and b respectively have the elements $x_1, x_2, ..., x_n$ and $b_1, b_2, ...b_n$.The equation above may be interpreted as a linear mapping of the vector x in

E_m, onto the vector b in E_n and the solution of system of equations amounts to determining the vector x when its mapping b is known. A practical method of solution, called the Gaussian elimination is based on the use of elementary operations in order to bring 1 on the diagonal of A and have zeros below the diagonal is the end of these operations.
Interpreted in operator formalism for $Ax = b$ to have no solution means that b is not an element in the range of A. Conversely if the system has solutions, it means that b is an element of the range of A. However the case with a unique solution implies that A is a one-to-one mapping while the case with an infinity of solutions implies that A is not a one-to-one operator.
Moreover it chould be recalled that the columns of the matrix representation of A span its range. Hence b will be in the range of A if it is linearly dependent on the columns of A. If b is linearly dependent on the columns of A, the rank of the matrix obtained from A by adding b as a last column will be the same as that of A. This latter matrix represented as $[A|b]$ is called the **augmented matrix**. Contrary to above, if b is not in the range of A, it will not be linearly dependent on the columns of the matrix representation of A. This means that the columns of the augmented matrix will have one more linearly independent vector, i.e. rank $[A|b]$ will be larger than rank $[A]$.
Clearly, if b is in the range of A, i.e. rank$A$ = rank $[A|b]$, one needs further information to decide whether the solution is unique or not. This can be readily answered by consideration of the null space of A. Indeed if the null space of A is not trivial, any solution x gives further solutions of the form $x + k$ where k is an element of the null space, i.e. $A(k) = 0$. Conversely, if the mill space of A contains only the zero element, an x such as $A(x) = b$ is going to be the only solution. Therefore the criterion for deciding about the uniqueness of the solution is the nullity of a to be zero. On the other hand it is known as a theorem that dimension of the domain of A = nullity of A + rank of A. Hence, if the nullity is zero, the dimension of the domain of A equals the rank of A. On the other hand, x being an element of the domain of A, the number of its components equals the dimension of the domain. Thus the following theorem:

Theorem. (Existence of Solution of Algebraic Equations). The system of algebraic equations $Ax = b$ has a solution if Rank$A$ = Rank$[A|b]$ and conversely no solution exists if Rank$A$ = Rank$[A|b]$. The use of determinants allows to make the above ideals more operational by calculating ranks as from non-vanishing determinants.

Chapter 3

EIGENVALUE PROBLEM

3.1 Transformations of matrices.

The basis concept of a matrix was obtained in the previous section as the image of a linear operator with respect to two bases, one for the domain and one for the range. Consequently, the same linear operator may be represented by different matrices depending on the choice of the bases for the domain and target space. Certain forms, i.e., diagonal or Jordan, may be preferred by their simplicity to a full matrix. Thus a change in the bases or a transformation of the bases, amounts in a change or transformation of the matrix. Rather than naming certain transformations, effort will be made to motivate how various transformations arise.

Consider first the mapping $\mathbf{Ax} = \mathbf{b}$. Let $\mathbf{Q}$ be a matrix such that $\mathbf{x} = \mathbf{Qy}$ and $\mathbf{b} = \mathbf{Pc}$. The substitution of the transformation yields: $\mathbf{AQy} = \mathbf{Pc}$ or for a nonsingular P, $(\mathbf{P}^{-1}\mathbf{AQ})\mathbf{y} = \mathbf{c}$

In the case where $P = Q$, the above transformation reads: $(\mathbf{Q}^{-1}\mathbf{AQ})\mathbf{y} = \mathbf{c}$. The matrix $\mathbf{B} = \mathbf{Q}^{-1}\mathbf{AQ}$ is said to be obtained from A by a **similarity transformation.** We shall later study methods of constructing the matrix $\mathbf{Q}$ such that $\mathbf{Q}^{-1}\mathbf{AQ}$ yields a particularly simple form: diagonal or Jordan.

As a second example, consider the expression $A = \mathbf{x}^T\mathbf{Ax}$ called a **quadratic form**. Again the transformation $x = Qy$ in view of $\mathbf{x}^T = (\mathbf{Qy})^T = \mathbf{y}^T\mathbf{Q}^T$ yields: $\mathbf{A} = \mathbf{y}^T(\mathbf{Q}^T\mathbf{AQ})\mathbf{y}$. Thus new matrix $\mathbf{B} = \mathbf{Q^TAQ}$ is obtained by a transformation of A that is called a **congruence transformation**.

If the matrix Q is such that $\mathbf{Q^T} = \mathbf{Q^{-1}}$, it is called an **orthonor-**

malal matrix and the similarity transformation involving $\mathbf{Q}$ is called an **orthogonal transformation**. It should be noted that an orthogonal transformation is simultaneously a congruence transformation. The following theorem expresses an important property or orthogonal transformation.
Theorem. An orthogonal transformation preserves norm and scalar product.
Proof. For the orthogonal transformation $\mathbf{Q}$ such that $\mathbf{x} = \mathbf{Qu}$ and $\mathbf{y} = \mathbf{Qv}$

$$(\mathbf{x}, \mathbf{y}) = \mathbf{x^T} \cdot \mathbf{y} = \mathbf{u^T Q^T} \cdot \mathbf{Qv} = \mathbf{u^T} \cdot \mathbf{v}$$

With $\mathbf{x} = \mathbf{y}$, the above results yields $|\mathbf{x}| = |\mathbf{u}|$. The geometric interpretation of orthogonal transforms that it preserves the lengths and angles, i.e. the geometry of shapes.
Example: The matrix

$$\mathbf{Q} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

represents a rotation by an angle α in two-dimensional Euclidean space (i.e. the plane). It is easily seen that $Q^{-1} = Q^T$.
A further property of orthogonal matrices is expressed by the following theorem.
Theorem. For an orthogonal matrix $\mathbf{Q}, \det\mathbf{Q} = 1 \text{or} - 1$.
Proof. Since $\mathbf{Q^{-1}} = \mathbf{Q^T}$ for an orthogonal matrix, then $Q^T Q = I$. Thus $\det(\mathbf{Q^T}) det(\mathbf{Q}) = 1$. As $\det(Q^T) = \det(Q)$, there follows $(\det(\mathbf{Q})^2 = 1$ and $\det(\mathbf{Q}) = \pm 1$.
As a geometric interpretation in E_3, orthogonal matrices with $\det\mathbf{Q} = 1$ correspond to a pure rotation and orthogonal matrices with
$\det\mathbf{Q} = -1$ correspond to a rotation and a reflection. Equivalently, the trasformation
with $\det\mathbf{Q} = 1$ represents a mapping from a right-handed (or left-handed) system to a right-handed (or left-handed)) system. Conversely the
transformation with $\det\mathbf{Q} = -1$ represents a mapping from a right-handed (or left-handed) system to a left handed (or left-handed)system.
Examples: For

$$\mathbf{Q} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}$$

$$\mathbf{P} = \begin{bmatrix} \cos\alpha & \sin\alpha \\ \sin\alpha & -\cos\alpha \end{bmatrix}$$

correspond respectively to transformations from a right-handed system to a right-handed system and to a left-handed system.

3.2 Eigenvalue problem for matrices.

For every square matrix there exist certain special vectors called characteristic or eigen vectors that are mapped onto themselves:

$$\mathbf{Ae} = \lambda \mathbf{e}$$

The scalar λ in this mapping is called the Eigenvalue.
The eigenvalues and eigenvectors are important for they usually carry an important information as well as they are auxiliary tools for solving mathematical problems related to matrices such as the Diagonalization of matrices.
Introducing the identity matrix $\mathbf{I}$, the eigenvalue equation above may be rewritten as a homogeneous equation:

$$(\mathbf{A} - \lambda \mathbf{I})\mathbf{e} = 0$$

The non-trivial solutions of this equation exist for

$$\det[\mathbf{A} - \lambda \mathbf{I}] = 0$$

This equation is a polynomial of degree n in λ, with n being the order of the matrix $\mathbf{A}$, since the determinant contains the n -tuple products of its elements. The polynomial in λ defined above is called the **characteristic polynomial**; is of degree n and hence has n (real or complex, dictinct or multiple) roots. The roots $\lambda_1, \lambda_2, ..., \lambda_n$ of the characteristic polynomial are called the **characteristic values** or the **eigenvalues** and the set of eigenvalues is called the **spectrum** of the matrix.
Example. We seek the eigenvalues and eigenvectors for the matrix

$$\mathbf{A} = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$$

The characteristic polynomial is:

$$P(\lambda) = det[\mathbf{A} - \lambda \mathbf{I}] = \lambda^2 - 6\lambda + 8$$

The roots of $P(\lambda) = 0$ are the eigenvalues: $\lambda_1 = 2$ and $\lambda_2 = 4$. To find the eigenvectors, we use the equations

$$(\mathbf{A} - \lambda_1 \mathbf{I})\mathbf{e}^{(1)} = 0$$

$$(\mathbf{A} - \lambda_2 \mathbf{I})\mathbf{e}^{(2)} = 0$$

For $\lambda_1 = 2$, there exists only one independent equation for $\mathbf{e}^{(1)} = (v_1^{(1)}, v_2^{(1)})$:

$$v_1^{(1)} + v_2^{(1)} = 0$$

For an arbitrary choice of $v_1^{(1)} = 1$, the eigenvector is determined as: $\mathbf{e}^{(1)} = (1, -1)$. By the similar calculations for λ_2, one obtains again one independent equation for $\mathbf{e}^{(1)} = (v_1^{(2)}, v_2^{(2)})$:

$$-v_1^{(2)} + v_2^{(2)} = 0$$

For the arbitrary choice of $v_1^{(2)} = 1$ the eigenvector is obtained to be: $\mathbf{e}^{(2)} = (1, 1)$.
Viewed in operator formalism, the eigenvalue problem amounts to determining the special values λ_k, for $k = 1, 2, ..., n$ for which the operator $A - \lambda_k I$ has a nontrivial null space. The corresponding eigenvector (or if λ_k is a multiple root, set of the corresponding eigenvectors) is a basis for the null space.
The following theorem illustrates the importance of the set of eigenvectors of a matrix.

Theorem. (Eigenvectors corresponding to distinct eigenvalues). The eigenvectors corresponding to distinct eigenvalues are linearly independent.

bf Proof. We seek to show that the linear relation $\Sigma c_k \mathbf{e}^{(k)} = \theta$ has only the trivial solution $c_1 = c_2 = ... = c_n = 0$. The proof is by induction by proving, that the eigenvector $\mathbf{e}^{(2)}$ is linearly independent from $\mathbf{e}^{(1)}$, that is linearly independent from $\mathbf{e}^{(2)}$ and $\mathbf{e}^{(1)}$ and so on.
First consider the relationship

$$c_1 \mathbf{e}^{(1)} + c_2 \mathbf{e}^{(2)} = 0$$

Applying the matrix $\mathbf{A}$ and making use of linearity and that $\mathbf{e}^{(1)}, \mathbf{e}^{(2)}$ are eigenvectors one has:

$$\mathbf{A}(c_1 \mathbf{e}^{(1)} + c_2 \mathbf{e}^{(2)}) = c_1 \mathbf{A}\mathbf{e}^{(1)} + c_2 \mathbf{A}\mathbf{e}^{(2)} = c_1 \lambda_1 \mathbf{e}^{(1)} + c_2 \lambda_2 \mathbf{e}^{(2)}$$

The replacement of $c_2\mathbf{e}^{(2)}$ by $-c_1\mathbf{e}^{(1)}$ from the starting linear relationship

$$c_1\mathbf{e}^{(1)} + c_2\mathbf{e}^{(2)} = 0$$

in the above equation yields:

$$c_1(\lambda_1 - \lambda_2)\mathbf{e}^{(1)} = 0$$

Since λ_1 and λ_2 are different, the only possibility is $c_1 = 0$. The use of $c_1 = 0$ in the linear relatioship
$c_1\mathbf{e}^{(1)} + c_2\mathbf{e}^{(2)} = 0$ yields $c_2 = 0$, which proves that the linear relationship has only the trivial solution for the coefficients c_1 and c_2 so that $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$ are linearly independent.

Discussion on multiple eigenvalues. The linear independence of the eigenvectors suggests the possibility of using them as a basis for the domain and range of the matrix. In the case where all of the eigenvectors are distinct, there are n linearly independent vectors for an n dimensional space. Hence, this set forms a basis. However an obstacle arises if all of the n eigenvalues of the matrix are not distinct. In the case , the question of existence of m_i linearly independent eigenvectors corresponding to the eigenvalue λ_i of multiplicity m_i is answered by the study of the rank of $\mathbf{A} - \lambda_i\mathbf{I}$. Indeed, the dimension N_i of the null space of $\mathbf{A} - \lambda_i\mathbf{I}$ is equal to $N_i = n - rank(\mathbf{A} - \lambda_i\mathbf{I})$. Hence, the number of linearly independent eigenvectors of $\mathbf{A}$ is N_i. Consequently, if N_i is equal to the multiplicity m_i of the eigenvalue λ_i, there corresponds m_i linearly independent eigenvectors of $\mathbf{e}^{(1)}$.

Examples. 1. For the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 2 & 1 \\ 1 & 3 & 1 \\ 1 & 2 & 2 \end{bmatrix}$$

the characteristic polynomial is

$$P(\lambda) = \det[\mathbf{A} - \lambda\mathbf{I}] = -(\lambda^3 - 7\lambda^2 + 11\lambda - 5)$$

The eigenvalues are found to be:

$$\lambda_1 = 5\lambda_2 = \lambda_3 = 1; m_2 = 2$$

Thus, Rank($\mathbf{A} - 5\mathbf{I}$) = 2, Rank($\mathbf{A} - \mathbf{I}$) = 1. For the distinct root $\lambda_1 = 5$ the eigenvector is found from

$$(\mathbf{A} - \boldsymbol{\lambda_1}\mathbf{I})\mathbf{e} = \begin{bmatrix} -3 & 2 & 1 \\ 1 & -2 & 1 \\ 1 & 2 & -3 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

It is seen by inspection that the third equation is a linear combination of first two (-1 times the first equation +2 times the second). Thus there exists two linearly independent equations. The arbitrary choice of $\mathbf{v}_1 = 1$ yields the eigenvector $\mathbf{e}^{(1)} = (1, 1, 1)$. For $\lambda_2 = \lambda_3$

$$(\mathbf{A} - \boldsymbol{\lambda_2}\mathbf{I})\mathbf{e} = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 2 & 1 \\ 1 & 2 & 1 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

It is seen that only one linearly independent equation exists and the solution admits two arbitrary coefficients. Thus the two arbitrary choices $v_2 = 0$ and $v_3 = -1$ yields the eigenvector $\mathbf{e}^{(2)} = (1, 0, -1)$. Similarly, the second arbitrary choice as $v_2 = -1$ and $v_3 = 0$ yields the eigenvector $e^{(3)} = (2, -1, 0)$. These are easily checked to be linearly independent as predicted by the theorem.

2. For the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 1 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

the eigenvalues are found to be: $\lambda_1 = \lambda_2 = \lambda_3 = 1; m_3 = 3$. On the other hand, it is easily checked that Rank$(\mathbf{A} - \mathbf{I}) = 1$. Consequently, there corresponds only $3 - 1 = 2$ linearly independent vectors to $\lambda_1 = 1$. In fact $(\mathbf{A} - \mathbf{I})\mathbf{e} = \theta$ yields:

$$(\mathbf{A} - \mathbf{I})\mathbf{e} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

*

$$(\mathbf{A} - \lambda_1\mathbf{I})\mathbf{e} = \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix}$$

=

$$\mathbf{A} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

Thus only v_1 and one v_2 of v_3 may be chosen. Thus two linearly independent solutions exist. An example is:

$$\mathbf{e}^{(1)} = (0, -2, 1), \mathbf{e}^{(2)} = (2, -1, 0)$$

The eigenvalue $\lambda_1 = 1$ of multiplicity $m_3 = 3$ can generate only two linearly independent eigenvectors and is said to be of degeneracy 1.

3.3 Eigenvalue problem for symmetric matrices

If the matrices are specialised to be symmetric, their eigenvectors enjoy two important properties expressed in the following theorems.
Theorem. (Real values). The eigenvalues of real symmetric matrices are real.
Proof: The characterstic polynomial cooresponding to a real matrix is real.
Therefore if it has a complex root λ_i, its complex conjugate λ_i^* is also a root. Consequently, for the eigenvector $\mathbf{e}_i$ corresponding to $\lambda_i, \mathbf{e}_i^*$ is the eigenvalue corresponding to λ_i^*

$$\mathbf{A}\mathbf{e}_i = \lambda_i \mathbf{e}_i$$

$$\mathbf{A}\mathbf{e}_i^* = \lambda_i^* \mathbf{e}_i^*$$

Multiplying the first equation by $\mathbf{e}_i^{*T}$ on the left; second equation by $\mathbf{e}_i$ on the right after taking its transpose yields:

$$\mathbf{e}_i^{*T}.T.e_i = \lambda_i \mathbf{e}_i^{*T}\mathbf{e}_i$$

$$\mathbf{e}_i^{T^*} T e_i = \lambda_i^* \mathbf{e}_i^{*T}\mathbf{e}_i$$

Subtracting the second equation above from the first, one has:

$$0 = (\lambda_i - \lambda_i^*)\mathbf{e}_i^{*T}\mathbf{e}_i$$

Since $\mathbf{e}_i^{*T^*}\mathbf{e}_i$ is the norm of the vector $\mathbf{e}_i$ is positive, i.e. non-zero. Consequently, $(\lambda_i - \lambda_i^*) = 0$, which implies that λ_i is real valued. It should be noticed also that the real valuedness of the eigenvalues implies the eigenvectors are to be real valued as well, since these are calculated from the set of linear equations with real coefficients.

Theorem. (orthogonality of eigenvectors).
The eigenvectors corresponding to different eigenvalues of a real symmetric matrix are orthogonal.
Proof. For the eigenvalues λ_1 and λ_2 with λ_1 and λ_2 being different, consider the eigenvalue problems:

$$\mathbf{A}\mathbf{e}_1 = \lambda_1 \mathbf{e}_1,$$

$$\mathbf{A}\mathbf{e}_2 = \lambda_2 \mathbf{e}_2$$

Multiplying the first equation by $\mathbf{e}_2^T$ on the left; the second equation by $\mathbf{e}_1$ on the right after taking its transpose yields:

$$\mathbf{e}_2^T \cdot \mathbf{A}\mathbf{e}_1 = \lambda_1 \mathbf{e}_2^T \cdot \mathbf{e}_1$$

$$\mathbf{e}_2^T \cdot \mathbf{A}\mathbf{e}_1 = \lambda_2 \mathbf{e}_2^T \cdot \mathbf{e}_1$$

Subtracting the second eqution above from the first, one has:

$$0 = (\lambda_1 - \lambda_2)\mathbf{e}_2^T \cdot \mathbf{e}_1$$

Since $(\lambda_1 - \lambda_2)$ is non-zero, $\mathbf{e}_2^T \cdot \mathbf{e}_1 = 0$ which is the orthogonality of $\mathbf{e}_1$ and $\mathbf{e}_2$.

Theorem. (Multiple eigenvalues).
For a symmetric matrix with the eigenvalue λ_i of multiplicity m_i the dimension of the null space is m_i. (For a proof, the reader is

referred to more specialised books in the literature).

The significance of this theorem lies in that for a symmetric matrix with the eigenvalue λ_i of multiplicity m_i, it is always possible to find m_i linearly independent eigenvectors.
Furthermore, since it is always possible to orthogonalise a given basis, it is possible to construct the m_i eigenvectors to be mutually orthogonal. It chould be noticed that while the eigenvectors corresponding to distinct eigenvalues are automatically orthogonal, those corresponding to the same eigenvalue have to be "constructed" as orthogonal. From the preceding two theorems we conclude that a symmetric matrix of order n, always has n orthogonal eigenvectors.

Example: Given

$$\mathbf{A} = \begin{bmatrix} 7 & -2 & 1 \\ -2 & 10 & -2 \\ 1 & -2 & 7 \end{bmatrix}$$

We seek to solve the eigenvalue problem for the matrix. The characteristic polynomial is

$$P(\lambda) = \det[\mathbf{A} - \lambda \mathbf{I}] =$$

$$-(\lambda^3 - 24\lambda^2 + 180\lambda - 432)$$

The eigenvalues are found to be:

$$\lambda_1 = \lambda_2 = 6; m_1 = 2, \lambda_3 = 12$$

For the multiple root $\lambda_1 = \lambda_2 = 6$, the eigenvector is found from

$$(\mathbf{A} - 6\mathbf{I})\mathbf{e} = 1 \begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

It is seen by inspection that the second and third rows are respectively -2 and 1 times the first. Thus there exists only one linearly independent equation. This equation has two arbitrary variables, say v_2 and v_3. Thus:

$$\mathbf{e}^{(1)} = (1, 0, -1),$$

$$\mathbf{e}^{(2)} = (1, 1, 1)$$

are two linearly independent solutions, which are constructed to be orthogonal.
For $\lambda_3 = 12$ the eigenvector is easily found from $(\mathbf{A} - 12\mathbf{I})e = 0$:

$$(A - 12I)e = 1 \begin{bmatrix} -5 & -2 & 1 \\ -2 & -2 & -2 \\ 1 & -2 & -5 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

It is easily noticed by inspection that the last row is equal (-1) times the first row , plus 2 times the second row so that it is not linearly independent. For the arbitrary choice of $v_3 = 1$, the eigenvector is found to be

$$\mathbf{e}^{(3)} = (1, -2, 1)$$

It chould be noticed that $\mathbf{e}^{(3)}$ is automatically orthogonal to $\mathbf{e}^{(1)}$ and $\mathbf{e}^{(2)}$.
This above example shows that for a symmetric matrix even a multiple eigenvalue always has a sufficient number of independent eigenvectors.

3.4 Diagonalization of matrices.

The matrix was introduced as a representation (or an image) of a linear transformation for a given of bases for the domain and the range of the operator. Consequently, different matrices depending on the choice of the bases may represent the same linear transformation. For a linear mapping with the domain and target space of dimension n, suppose there exist n linearly independent eigenvectors. These eigenvectors can be used as a basis for both the domain and the target space. Consequently, by the matrix represetation of the linear transformation as shown in Section 2.,we immediately have:

$$\mathbf{A}(\mathbf{e}_j) = \lambda_j \mathbf{e}_j = \Sigma a_{ij} \mathbf{e}_j$$

It thus follows that $a_{ij} = \lambda_j$ while all other a_{ij} are zero; i.e. the matrix with the elements defined above by using the eigenvectors of the operator is diagonal. It should be noticed that for the above construction, the eigenvalues need not be distinct while there suffices to have n linearly independent eigenvectors.
The above reasoning can be made operationally more convenient in the matrix notation. To appreciate its importance, let us recall

that the transformations $\mathbf{x} = \mathbf{Qy}, \mathbf{b} = \mathbf{Qc}$ in the matrix equation $Ax = b$ yields: $(\mathbf{Q}^{-1}\mathbf{AQ})\mathbf{y} = \mathbf{c}$. The transformation above would appear simplest if the matrix $(\mathbf{Q}^{-1}\mathbf{AQ})$ was in diagonal form. Towards obtaining the similarity transformation that diagonalises $\mathbf{A}$, consider collecting all of the n eigenvalue equations for the matrix together. We thus have:

$$\mathbf{A}\mathbf{e}_1 = \lambda_1 \mathbf{e}_1$$

$$\mathbf{A}\mathbf{e}_2 = \lambda_2 \mathbf{e}_2$$

$$\cdots\cdots\cdots$$

$$\mathbf{A}\mathbf{e}_n = \lambda_n \mathbf{e}_n$$

Since all of the eigenvectors multiply the same $\mathbf{A}$ on the right, the set of n vector equations above may be collected together into a single matrix equation by writing the eigenvectors one after the other to form the columns of a matrix:

$$A \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_3 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} \lambda_1\mathbf{e}_1 & \lambda_2\mathbf{e}_2 & \cdots & \lambda_n\mathbf{e}_n \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} =$$

Identifying the matrix with the eigenvectors as its columns by $\mathbf{Q}$, the right hand side of the equation above may be written as:

$$\begin{bmatrix} \lambda_1\mathbf{e}_1 & \lambda_2\mathbf{e}_2 & \cdots & \lambda_n\mathbf{e}_n \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_3 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

or defining Λ as the diagonal matrix with the eigenvalues on the diagonal, the above equation can be expressed in the compact form:

$$\mathbf{AQ} = \mathbf{Q}\boldsymbol{\Lambda}$$

The matrix Q being made of linearly independent vectors as its columns,is nonsingular,i.e. $\mathbf{Q}^{-1}$ exists. Thus multiplying the above equation by $\mathbf{Q}^{-1}$ on the left results in the similarity transformation that reduces the matrix to a diagonal form as:

$$\mathbf{Q^{-1}AQ} = \mathbf{\Lambda}$$

Thus we have proven the following theorem.

Theorem. (Diagonalization of Matrices).
If the square matrix $\mathbf{A}$ with n columns and rows possesses n linearly independent eigenvectors, then it can always be put in its diagonal form by the similarity transformation $\mathbf{Q^{-1}AQ}$ where the matrix Q has the n eigenvectors as its columns.

A special situation exists when the matrix A is symmetric as a result of the orthogonality of its eigenvectors. In constructing $\mathbf{Q}$ for a symmetric matrix, if the eigenvecotrs are-normalised, Q becomes an orthonormal matrix. The following theorem specifies this property.

Theorem. (Orthogonality of the Matrix of Eigenvectors). The matrix Q having as its columns the normalised eigenvectors of a symmetric matrix is an orthonormal matrix.
Proof: For

$$\mathbf{Q} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix}$$

implies

$$\mathbf{Q^TQ} = \begin{bmatrix} \mathbf{e}^{T_1} & \cdots & \cdots & \rightarrow \\ \mathbf{e}^{T_2} & \cdots & \cdots & \rightarrow \\ \cdots & \cdots & \cdots & \cdots \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{e}^{T_n} & \cdots & \cdots & \rightarrow \end{bmatrix} \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ \downarrow & \downarrow & \downarrow & \downarrow \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Thus $\mathbf{Q^TQ} = \mathbf{I}$ or $\mathbf{Q^T} = \mathbf{Q^{-1}}$. In view of this property of $\mathbf{Q}$, the similarity transformation in becomes also a congruence transformation or an orthogonal transformation.

Example:
We seek to find the orthogonal transformation that diagonalises the matrix of the example in Section 3.3:

$$\mathbf{A} = \begin{bmatrix} 7 & -2 & 1 \\ -2 & 10 & -2 \\ 1 & -2 & 7 \end{bmatrix}$$

The eigenvalues are found to be:

$$\lambda_1 = 12,$$

$$\lambda_2 = \lambda_3 = 6; m_2 = 2$$

and the corresponding eigenvectors were

$$e^{(1)} = (1, 0, -1); e^{(2)} = (1, 1, 1); e^{(3)} = (1, -2, 1)$$

With the normalisation of the eigenvectors, the matrix $\mathbf{Q}$ then is obtained as:

$$\mathbf{Q} = \begin{bmatrix} 1/2 & 1/3 & 1/6 \\ 0 & 1/3 & -2/6 \\ -1/2 & 1/3 & 1/6 \end{bmatrix}$$

It can directly be checked that $Q^T Q = I, Q^T = Q^{-1}$ and

$$\mathbf{Q^T A Q} = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 12 \end{bmatrix}$$

This example illustrates that even when a symmetric matrix has multiple eigenvalues, it is possible to diagonals it.

Example: We seek to find the orthogonal transformation that diagonalises the matrix

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$$

The eigenvalues are obtained to be: $\lambda_1 = -1, \lambda_2 = 5$. The corresponding eigenvectors are:

$$\mathbf{e}_1 = (1, -1), \mathbf{e}_2 = (1, 2)$$

The matrix $\mathbf{Q}$, its inverse and the similarity transformations are calculated to be:

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}$$

$$\mathbf{Q^{-1}} = \begin{bmatrix} 2/3 & -1/3 \\ 1/3 & 1/3 \end{bmatrix}$$

$$\mathbf{Q^{-1}AQ} = \begin{bmatrix} -1 & 0 \\ 0 & 5 \end{bmatrix}$$

It is also readily checked that the eigenvectors are not orthogonal.

Example: We seek to find the orthogonal transformation that diagonalises the matrix

$$\mathbf{A} = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix}$$

The eigenvalues and the corresponding eigenvectors are obtained to be: $\lambda_1 = \lambda_2 = 2$. For the eigenvalue with multiplicity two, in this case only one eigenvector can be obtained and hence the matrix Q cannot be constructed. Consequently, this matrix cannot be diagonalised. The matrix in this example is said to be in the Jordan form. When a matrix cannot be diagonalised, it can always be reduced to a Jordan form.

3.5 Quadratic forms and definite matrices

.

The quadratic form

$$A = \Sigma_{i=1}^{N} \Sigma_{j=1}^{N} a_{ij} x_i x_j = a_{11} x_1^2 + 2a_{12} x_1 x_2 + a_{22} x_2^2 + \cdots$$

appears in many applications representing for instance kinetic energy, potential energy, a target function to optimize, a surface, etc. A as a function of the n variables $\{x_j, j = 1, 2, ...N\}$, may be written by introducing the symmetric matrix A with elements $a_{ij} = a_{ji}$ as

$$A = x^T \cdot \mathbf{A} x$$

The quadratic form would take on its simplest form in a coordinate system where only the squares of the cootdinates appeared. To this aim, consider a coordinate transformation as $x = Qy$. Substitution of x above into the quadratic form results in

$$A = y^T \cdot (\mathbf{Q}^T \mathbf{A} \mathbf{Q}) y$$

Sect 3.3 about the diagonalization of symmetric matrices shows that with Q having the normalized eigenvectors of $\mathbf{A}$ as its columns, $\mathbf{Q}^T \mathbf{A} \mathbf{Q}$ results in a diagonal matrix. Furthermore, the diagonal elements of ??? are the eigenvalues of $\mathbf{A}$. Thus, in the transformed coordinates $y_1, y_2, ..., y_n, A$ reads:

$$A = \Sigma \lambda_i y_i^2 = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \lambda_3 y_3^2 + ...$$

The above representation of in not having mixed terms is said to be the **canonical form**. This form of the quadratic expression allows one to understand the shape of the function that is not transparent in the $x_1, x_2, ...x_n$ coordinates.
In many applications, the physical requirements are such that A be always positive and be zero only for $x = 0$.

Definition. (Positive definiteness). The quadratic form $A = x^T \cdot \mathbf{A} x$ and the associated matrix $\mathbf{A}$ are called **positive definite**, if $A > 0$ for all values of x_i and $A = 0$ only for $x = 0$.
The canonical representation of the quadratic form provides the necessary and sufficient conditions of positive definiteness: if all λ_i are positive, A is too and conversely if A is positive definite, the λ_i are

necessarily positive also. Hence we state the following theorem.

Theorem. (Positive definiteness). The necessary and sufficient conditions for a quadratic form and the associated matrix to be positive definite are that all of its eigenvalues be positive.

Examples. Finding the canonical representations of quadratic forms.

1. For

$$A = 5x_1^2 + 4x_1x_2 + 2x_2^2 = x^T \cdot Ax$$

we have

$$\mathbf{A} = \begin{bmatrix} 5 & 2 \\ 2 & 2 \end{bmatrix}$$

The corresponding eigenvalues and eigenvectors are found to be:

$$\lambda_1 = 6, \lambda_2 = 1$$

and

$$u_1 = (2, 1)/5^{1/2}, u_2 = (-1, 2)/5^{1/2}$$

Thus, for $x = Q.y$, with Q having u_1 and u_2 as its columns, we find the canonical form as:

$$A = 6y_1^2 + y_2^2$$

Geometrically, A takes as constants represents a family of concentric ellipses. In the (x_1, x_2) plane, the principal axes are along u_1 and u_2 and are of length $(A/6)^{1/2}$ and $A^{1/2}$ respectively. Similarly the function $A = A(x_1, x_2)$ represents an elliptic paraboloid surface in the three dimensional (x_1, x_2, A) space.

2. For

$$A = -x_1^2 + 4x_1x_2 + 2x_2^2 = x^T \cdot Ax$$

we have

$$\mathbf{A} = \begin{bmatrix} -1 & 2 \\ 2 & 2 \end{bmatrix}$$

The corresponding eigenvalues and eigenvectors are found to be:

$$\lambda_1 = 3, \lambda_2 = -2$$

and

$$u_1 = (1,2)/5^{1/2}, u_2 = (-2,1)/5^{1/2}$$

Thus, for $x = \mathbf{Q}y$, with $\mathbf{Q}$ having u_1 and u_2 as its columns, we find the canonical form as:

$$A = 3y_1^2 - 2y_2^2$$

Geometrically, A takes as constants represents a family of hyperbolas. In the (x_1, x_2) plane, the principal axes are along u_1 and u_2 and are of length $(A/3)^{1/2}$ and $(A/2)^{1/2}$ respectively. Similarly the function $A = A(x_1, x_2)$ represents a hyperbolic paraboloid surface, i.e. a saddle point geometry in the three dimensional (x_1, x_2, A) space.
3. For

$$A = x_1^2 + 4x_1x_2 + 4x_2^2$$

we have

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$$

The corresponding eigenvalues and eigenvectors are found to be:

$$\lambda_1 = 5, \quad \lambda_2 = 0$$

and

$$u_1 = (1,2)/5^{1/2}, \quad u_2 = (-2,1)/5^{1/2}$$

Thus, for $x = Qy$, with Q having u_1 and u_2 as its columns, we find the canonical form as:

$$A = 5y_1^2$$

Geometrically, $A =$ constant represents a family of parabolic cylinders with the axis in the u_2 direction in the (x_1, x_2) plane.

3.6 Functions of matrices

.

The powers of the matrix $\mathbf{A}$ are defined as:

$$\mathbf{A}^0 = \mathbf{I}, \mathbf{A}^k = \mathbf{A}^{k-1}\mathbf{A}$$

From this definition it is clear that the matrix $\mathbf{A}^k$ commutes with $\mathbf{A}^l$ and that

$$\mathbf{A}^k\mathbf{A}^l = \mathbf{A}^l\mathbf{A}^k = \mathbf{A}^{k+1}$$

The product $c_k\mathbf{A}^k$ of any power of $\mathbf{A}$ by a scalar c_k is called a monomial in the matrix A. The sum of a finite number of monomials with $k = 1, 2, ..., p$ is called a polynomial of degree p in the matrix A:

$$P(\mathbf{A}) = \Sigma c_k \mathbf{A}^k$$

Theorem. (eigenvalues of a power of a matrix). For λ_i being an eigenvalue of a matrix $\mathbf{A}$ and v_i being the corresponding eigenvector, λ_i^k is an eigenvalue and u_i is the corresponding eigenvector of $\mathbf{A}^k$.

Proof: The eigenvalue equation $\mathbf{A}u_i = \lambda_i u_i$, after multiplicition by $\mathbf{A}$ yields:

$$\mathbf{A}^2 u_i = \lambda_i(\mathbf{A}u_i) = \lambda_i^2 u_i$$

Multiplying consecutively the above equation by A and using the eigenvalue equation each time, we generate $\mathbf{A}^k u_i = \lambda_i^k u_i$ which is the proof of the theorem.

Corollary (eigenvalues of matrix polynomials). For λ_i being an eigenvalue of a matrix A and u_i being the corresponding eigenvector, $P(\lambda_i)$ is an eigenvalue and u_i is the corresponding eigenvector of $P(\mathbf{A})$.
The proof follows readily by using the result above for each power in each of the terms of a polynomial:

$$P(\mathbf{A})u_i = \Sigma c_k \mathbf{A}^k u_i = (\Sigma c_k \lambda_i^k) u_i = P(\lambda_i) u_i$$

The following theorem highlights the functions of matrices and is of practical and theoretical importance.

Theorem. (Cayley-Hamilton). Every matrix satisfies its own characteristic polynomial $P(\lambda) = \det[\mathbf{A} - \lambda\mathbf{I}]$.

Proof. The above result $P(\mathbf{A})u_i = P(\lambda_i)u_i$ holds for any polynomial. In particular for $P(\lambda)$ being the characteristic polynomial, by definition $P(\lambda_i) = 0$. Thus,

$$P(\mathbf{A})u_i = 0$$

For u_i being a nontrivial vector and the result above being true for the whole set of eigenvectors which form a basis, we have

$$P(\mathbf{A}) = O$$

where **O** is the the null matrix.
Written explicitly, the characteristic polynomial $P(\lambda)$ is of degree n of the form:

$$P(\lambda) = (-1)^{-1}(\lambda_i^n - I_1\lambda^{n-1} + I_2\lambda^{n-2} + ... + (-1)^{-n}I_n)$$

where $I_1, I_2, I_3, ..., I_n$ are some specific numbers, called as the invariants of the matrix A. Thus, $P(\mathbf{A}) = 0$ yields:

$$\mathbf{A}^n = I_1\mathbf{A}^{n-1} - I_2\mathbf{A}^{n-2} + I_3\mathbf{A}^{n-3} + ... + (-1)^n I_n\mathbf{I}$$

Viewed in this form, the Cayley-Hamilton Theorem states that the n^{th} power of an $(n \times n)$ matrix is expressible in terms of powers of A up to $n-1$. Taking the product of the above representation by A,we find:

$$\mathbf{A}^{n+1} = I_1\mathbf{A}^n - I_2\mathbf{A}^{n-1} + I_3\mathbf{A}^{n-2} + ... + +(-1)^n I_n\mathbf{A}$$

Above, $\mathbf{A}^n$ may be substituted by the expression above to obtain $\mathbf{A}^{n+1}$ in terms of powers of A up to $n-1$ again. Proceeding in this direction, we prove the following corollary.

Corollary. (powers of a matrix) Any power $k > n-1 of \mathbf{A}^k$ with $\mathbf{A}$ being an $n \times n$ matrix is expressible as a linear combination of $\mathbf{I}, \mathbf{A}, ..., \mathbf{A}^{n-1}$:

$$\mathbf{A}^k = b_0 I + b_1\mathbf{A} + b_2\mathbf{A}^2 + ... + b_{n-1}\mathbf{A}^{n-1}$$

The use of the above corollary permits to state even a more general result about polynomials of matrices.

Corollary. (Polynomials of a matrix) Any polynomial $P(\mathbf{A})$ of order $p \geq n$ of the matrix $\mathbf{A}$ of order n, is expressible as a linear combination of $\mathbf{I}, \mathbf{A}, ..., \mathbf{A}^{n-1}$

$$P(\mathbf{A}) = b_0\mathbf{I} + b_1\mathbf{A} + b_2\mathbf{A}^2 + ... + b_{n-1}\mathbf{A}^{n-1}$$

The proof follows immediately from the previous corollary for each term of power $k \geq n$ in the polynomial.

The previous corollary is also extended to cover convergent infinite series of. Consequently, any matrix function $f(A)$ such that

$f(x)$ has a convergent Taylor series (e.g. $e^x, \sin x$, etc.) is expressible as a linear combination of $\mathbf{I}, \mathbf{A}, ..., \mathbf{A}^{n-1}$:

$$f(\mathbf{A}) = b_0\mathbf{I} + b_1\mathbf{A} + b_2\mathbf{A}^2 + ... + b_{n-1}\mathbf{A}^{n-1}$$

Similarly, the result on the eigenvalues and eigenvectors for $P(\mathbf{A})$ may also be generalized to any function $f(\mathbf{A})$ as:

$$f(\mathbf{A})u_i = f(\lambda_i)u_i$$

This means that the eigenvectors u_i of A are also the eigenvectors of any function $f(\mathbf{A})$ with the eigenvalues $f(\lambda_i)$

A method for calculating functions of matrices.

The method for calculating powers of a matrix by the use of the Cayley-Hamilton theorem is not practical for high powers and is impossible for infinite series of the matrix. Instead, an indirect method based on the theorems above proves more convenient. We proved that it is possible to express any $f(A)$ as

$$f(\mathbf{A}) = b_0\mathbf{I} + b_1\mathbf{A} + b_2\mathbf{A}^2 + ... + b_{n-1}\mathbf{A}^{n-1}$$

where $b_0, b_1, b_2, ..., b_{n-1}$ are yet unknown coefficients. Thus, for any eigenvalue λ_i, $f(\mathbf{A})u_i = f(\lambda_i)u_i$ or in view of the definition of $f(\mathbf{A})$:

$$(b_0 + b_1\lambda_i + b_2\lambda_i^2 + ... + b_{n-1}\lambda_i^{n-1})u_i = f(\lambda_i)u_i$$

Consequently for each of $i = 1, 2, ..., n$ we equate the coefficients by removing the eigenvector u_i to have n equations of the form:

$$b_0 + b_1\lambda_i + b_2\lambda_i^2 + ... + b_{n-1}\lambda_i^{n-1} = f(\lambda_i), i = 0, 1, ...n - 1.$$

We can solve for the unknowns $b_i, i = 0, 1, ..., n - 1$

A method for calculating functions of matrices using diagonalization.

An alternative for calculating functions of matrices utilizes the diagonal form. The collective use of the eigenvectors of a matrix A to build the modal matrix Q having the eigenvectors as its columns yields the diagonal form:

$$Q^{-1}f(A)Q = \begin{bmatrix} f(\lambda_1) & 0 & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & 0 & \cdots & 0 \\ 0 & 0 & f(\lambda_3) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & \cdots & f(\lambda_{n-1}) & 0 \\ 0 & 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix}$$

The use of the products of the above expression by respectively $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ on the right and left yields:

$$f(\mathbf{A}) = \mathbf{Q} \begin{bmatrix} f(\lambda_1) & 0 & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & 0 & \cdots & 0 \\ 0 & 0 & f(\lambda_3) & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \\ 0 & 0 & \cdots & f(\lambda_{n-1}) & 0 \\ 0 & 0 & 0 & \cdots & f(\lambda_n) \end{bmatrix} \mathbf{Q}^{-1}$$

Examples:

1. We wish to calculate A^3 by using the Cayley-Hamilton Theorem for

$$\mathbf{A} = \begin{bmatrix} 1 & 2 \\ 4 & 3 \end{bmatrix}$$

The characteristic polynomial

$$P(\lambda) = \det|\mathbf{A} - \lambda \mathbf{I}|$$

is:

$$P(\lambda) = \lambda^2 - 4\lambda - 5.$$

By Cayley-Hamilton theorem

$$P(\mathbf{A}) = 0,$$

i.e.

$$\mathbf{A}^2 - 4\mathbf{A} - 5\mathbf{I} = 0.$$

Thus, $\mathbf{A}^2 = 4\mathbf{A} + 5\mathbf{I}$. Forming the product of the above equation with $\mathbf{A}$, we have:

$$\mathbf{A}^3 = 4\mathbf{A}^2 + 5\mathbf{A}$$

$\mathbf{A}^2$ on the right hand side may be substituted by the use of $P(\mathbf{A}) = 0$ to obtain:

$$\mathbf{A}^3 = 4(\mathbf{A} + 5\mathbf{I}) + 5\mathbf{A}$$

Thus

$$\mathbf{A}^3 = 20\mathbf{I} + 21\mathbf{A} =$$

$$\mathbf{A} = \begin{bmatrix} 41 & 42 \\ 84 & 83 \end{bmatrix}$$

2. Calculate $\mathbf{A}^3$ by the method using eigenvalues. The eigenvalues of A in the last example are: $\lambda_1 = -1$ and $\lambda_2 = 5$. By the general representation, it is always possible to express $\mathbf{A}^3$ as a combination of $\mathbf{A}$ and $\mathbf{I} = \mathbf{A}^0$:

$$\mathbf{A}^3 = b_0\mathbf{I} + b_1\mathbf{A}$$

In order to determine the coefficients b_0 and b_1, we take the of by u_1 and u_2 and make use of the eigenvalues to obtain:

$$\mathbf{A}^3 u_1 = (b_0\mathbf{I} + b_1\mathbf{A})u_1$$

$$\mathbf{A}^3 u_2 = (b_0\mathbf{I} + b_1\mathbf{A})u_2$$

In the above equating the coefficients of u_1 and u_2 and substituting $\lambda_1 = -1$ and $\lambda_2 = 5$ from above, we have

$$b_0 - b_1 = -1$$

and

$$b_0 + 5b_1 = 125$$

Solving for b_0 and b_1, we find $b_0 = 20$ and $b_1 = 21$. The use of b_0 and b_1 in A^3 yields the same result as that obtained by the direct use of the Cayley-Hamilton Theorem.

3. We wish to calculate $\mathbf{A}^3$ by using the diagonalization of $\mathbf{A}$ in the preceding example. For $\mathbf{A}$ the eigenvalues were found to be $\lambda_1 = -1$ and $\lambda_2 = 5$. The corresponding eigenvectors are: $u_1 = (1, -1)$ and $u_2 = (1, 2)$. The modal matrix $\mathbf{Q}$ and its inverse respectively become:

$$\mathbf{Q} = \begin{bmatrix} 1 & 1 \\ -1 & 2 \end{bmatrix}$$

$$\mathbf{Q^{-1}} = \begin{bmatrix} 2/3 & -1/3 \\ 1/3 & 1/3 \end{bmatrix}$$

The result it obtained as:

$$A^3 = Q \begin{bmatrix} (-1)^3 & 0 \\ 0 & 5^3 \end{bmatrix} Q^{-1}$$

This product yields the same result as those obtained by the two previous medhods.

4. For A as in the preceding example, we wish to evaluate e^A. In this case, the direct method proves impossible, since e^A results, when Taylor expanded, in an infinite series. We proceed therefore, by the other two methods.
For $e^{\mathbf{A}} = b_0\mathbf{I} + b_1\mathbf{A}$ yields by the same eigenvalue considerations as above and with $\lambda_1 = -1$ and $\lambda_2 = 5$ as given above,

$b_0 - b_1 = e^{-1}$ and $b_0 + 5b_1 = e^5$ b_0 and b_1 are solved to be: $b_0 = (5e^{-1} + e^5)/6$ and $b_1 = (-e^{-1} + e^5)/6$. Substituting b_0, b_1, I and A yields the result for e^A.
Similarly, as a second method, the use of diagonalization with $\mathbf{Q}$ and $\mathbf{Q}^{-1}$ as in example 3,

$$e^{\mathbf{A}} = \mathbf{Q} \begin{bmatrix} e^{-1} & 0 \\ 0 & e^5 \end{bmatrix} \mathbf{Q}^{-1}$$

The products above yields the same result as by the previous method in a simpler way.