0. Prelude - What is Linear Algebra?

The key idea that underlies the course is the notion of LINEARITY. Roughly speaking, a property is linear if, whenever objects $X$ and $Y$ have the property, so do $X + Y$ and $\lambda X$, where $\lambda$ is a real or complex scalar. For this to make sense, we need to be able to add our objects $X$ and $Y$ and multiply them by scalars.

The first key idea we shall meet is that of sets with a linear structure, that is sets in which there are natural definitions of addition and scalar multiplication. Simple examples are

(a) $\mathbb{R}^3$ (3-dimensional space) with

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

and

$$\lambda(x_1, x_2, x_3) = (\lambda x_1, \lambda x_2, \lambda x_3)$$

for $(x_1, x_2, x_3), (y_1, y_2, y_3) \in \mathbb{R}^3$ and $\lambda \in \mathbb{R}$.

(b) The set $C_\mathbb{R}[a, b]$ of all continuous functions $f : [a, b] \to \mathbb{R}$, with addition $f + g$ and scalar multiplication $\lambda f$ ($\lambda \in \mathbb{R}$) given by

$$(f + g)(t) = f(t) + g(t) \quad \text{and} \quad (\lambda f)(t) = \lambda f(t)$$

for $a \leq t \leq b$.

These two examples, though different in concrete terms, have similar ‘linear’ structures; they are in fact both examples of vector spaces (the precise definition will be given in Chapter 1).

Linear algebra is the study of linear structures (vector spaces) and certain mappings between them (linear mappings). To illustrate, consider the following simple examples.

(a) A $2 \times 2$ real matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ defines a mapping $\mathbb{R}^2 \to \mathbb{R}^2$ given by

$$T : \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \to \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix}.$$ 

This has the property that $T(x + y) = Tx + Ty$ and $T(\lambda x) = \lambda(Tx)$; in other words, $T$ respects the linear structure in $\mathbb{R}^2$.

(b) For $f \in C_\mathbb{R}[a, b]$, put

$$Tf = \int_a^b f(t) \, dt.$$ 

Here $T$ maps $C_\mathbb{R}[a, b]$ to $\mathbb{R}$ and satisfies

$$T(f + g) = Tf + Tg \quad \text{and} \quad T(\lambda f) = \lambda Tf$$

so that, as in (a), $T$ respects the linear structures in $C_\mathbb{R}[a, b]$ and $\mathbb{R}$. 
Such mappings $T$ that respect or preserve the linear structure are called *linear mappings*.

1. Vector spaces

Throughout, $\mathbb{R}$ and $\mathbb{C}$ denote, respectively, the real and complex numbers, referred to as *scalars*. $\mathbb{F}$ will denote either $\mathbb{R}$ or $\mathbb{C}$ when it is unimportant which set (field) of scalars we are using. $\mathbb{F}^n$ is the set of all $n$-tuples of elements of $\mathbb{F}$ and we write a typical element $x$ of $\mathbb{F}^n$ either as a column (vector)

$$x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

or as a row (vector) $x = (x_1, x_2, \ldots, x_n)$. In $\mathbb{F}^n$, addition and scalar multiplication are defined coordinatewise by

$$(x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n)$$

and

$$\lambda(x_1, x_2, \ldots, x_n) = (\lambda x_1, \lambda x_2, \ldots, \lambda x_n).$$

This is the basic example of a vector space over $\mathbb{F}$.

The formal definition of a vector space $V$ over $\mathbb{F}$ is as follows. $V$ is a non-empty set with the following structure.

- Given elements $x, y \in V$, we can add them to obtain an element $x + y$ of $V$.
- Given $x \in V$ and $\lambda \in \mathbb{F}$, we can form the product $\lambda x$, again an element of $V$.
- We need these (usually natural) operations to obey the familiar algebraic laws of arithmetic:
  
  (a) $x + y = y + x$ for all $x, y \in V$ (addition is commutative);
  
  (b) $x + (y + z) = (x + y) + z$ for all $x, y, z \in V$ (addition is associative);
  
  (c) there exists an element $0 \in V$ such that $0 + x = x (= x + 0)$ for all $x \in V$;
  
  (d) given $x \in V$, there is an element $-x \in V$ such that $x + (-x) = 0 (= (-x) + 0)$

  and

  (e) $\lambda(x + y) = \lambda x + \lambda y$;
  
  (f) $(\lambda + \mu)x = \lambda x + \mu x$;
  
  (g) $(\lambda \mu)x = \lambda(\mu x)$

  for all $x, y \in V$ and all $\lambda, \mu \in \mathbb{F}$; and finally

  (h) $1.x = x$ for all $x \in V$.

[(a)-(d) say that $V$ is an abelian group under $+$; (e)-(g) are the natural distributive laws linking addition and scalar multiplication; (h) is a natural normalization for scalar multiplication.]

We call $V$ a real (resp. complex) vector space when $\mathbb{F} = \mathbb{R}$ (resp. $\mathbb{C}$).

The examples we shall encounter in the course will comprise either sets of points in $\mathbb{F}^n$, with
algebraic operations defined coordinatewise, or sets of scalar valued functions defined on a set, with algebraic operations defined pointwise on the set. In both cases, properties (a), (b), (e), (f), (g) and (h) hold automatically. Also, $0 = 0 \cdot x$ (where the 0 on the l.h.s. is the zero element of $V$ and the 0 on the r.h.s. in the scalar 0) and $-x = (-1) \cdot x$ for every $x \in V$. The crucial properties to check are

- If $x, y \in V$, is $x + y$ in $V$?
- If $x \in V$ and $\lambda \in \mathbb{F}$, is $\lambda x$ in $V$?

When these two properties hold, we say that $V$ is closed under addition and scalar multiplication.

**Examples**

1. $\mathbb{F}^n$ with addition and scalar multiplication defined coordinatewise as above. Here (writing elements as row vectors), $0 = (0, 0, \ldots, 0)$ and $-(x_1, x_2, \ldots, x_n) = (-x_1, -x_2, \ldots, -x_n)$.

2. The set $M_{m,n}(\mathbb{F})$ of all $m \times n$ matrices with entries in $\mathbb{F}$ and with addition and scalar multiplication defined entrywise. Here 0 in $M_{m,n}(\mathbb{F})$ is just the zero matrix (all entries 0) and $-\left(a_{i,j}\right) = (-a_{i,j})$.

3. The set $P_n$ of all polynomials in a variable $t$ with real coefficients and degree less than or equal to $n$ and with addition and scalar multiplication defined pointwise in the variable $t$. This is a real vector space. (Note that the set of such polynomials of exact degree $n$ is not a vector space.)

**Lecture 2**

1. **Vector spaces contd.**

**Subspaces**

Let $V$ be a vector space over $\mathbb{F}$ and let $U \subseteq V$.

*Definition* We say that $U$ is a subspace of $V$ if, with addition and scalar multiplication inherited from $V$, $U$ is a vector space in its own right.

**(1.1) Proposition** Let $U$ be a subset of a vector space $V$ over $\mathbb{F}$. Then $U$ is a subspace of $V$ if and only if

(a) $0 \in U$;

(b) $x + y \in U$ for all $x, y \in U$;

(c) $\lambda x \in U$ for all $x \in U$ and all $\lambda \in \mathbb{F}$.

*Comments*

1. The zero element of a subspace of $V$ is the same as the zero element of $V$.

2. We could replace (a) by just saying that $U \neq \emptyset$.

3. $V$ is a subspace of itself; $\{0\}$ is also a subspace of $V$.

*Examples*

1. $\{p \in P_n : p(0) = 0\}$ is a subspace of $P_n$. 

2. \( \{ x \in \mathbb{R}^n : x_1 \geq 0 \} \) is not a subspace of \( \mathbb{R}^n \).

3. The subspaces of \( \mathbb{R}^2 \) are just \( \mathbb{R}^2 \), \{0\}, and straight lines through 0. Likewise, the subspaces of \( \mathbb{R}^3 \) are just \( \mathbb{R}^3 \), \{0\}, and straight lines or planes through 0.

### Spans

**Definitions** Let \( V \) be a vector space over \( F \).

(i) Given vectors \( v_1, v_2, \ldots, v_k \) in \( V \), we call a vector \( u \) of the form

\[
u = \lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k
\]

for some \( \lambda_1, \lambda_2, \ldots, \lambda_k \in F \) a \textit{linear combination} of \( v_1, \ldots, v_k \).

(ii) Let \( S \) be a non-empty subset of \( V \). Then the \textit{span} (or \textit{linear span}) of \( S \), denoted by \( \text{span} \, S \), is the set of all linear combinations of elements of \( S \). Thus

\[
\text{span} \, S = \{ \lambda_1 v_1 + \cdots + \lambda_k v_k : \lambda_j \in F, v_j \in S \}.
\]

### Notes

1. An alternative notation to \( \text{span} \, S \) is \( \text{lin} \, S \).

2. \( k \) here may be different for different elements of \( \text{span} \, S \).

3. \( S \) may be infinite but we only consider \textit{finite} linear combinations.

4. If \( S \) is finite, say \( S = \{ v_1, \ldots, v_m \} \), then

\[
\text{span} \, S = \{ \lambda_1 x_1 + \cdots + \lambda_m x_m : \lambda_j \in F \}.
\]

**Theorem** Let \( S \) be a non-empty subset of a vector space \( V \). Then

(i) \( \text{span} \, S \) is a subspace of \( V \);

(ii) if \( U \) is a subspace of \( V \) and \( S \subseteq U \), then \( \text{span} \, S \subseteq U \).

[Thus \( \text{span} \, S \) is the smallest subspace of \( V \) containing \( S \).]

**Examples**

1. In \( \mathbb{R}^2 \), the span of a single non-zero vector \( v = (a, b) \) is just the straight line through the point \( (a, b) \) and the origin. Similarly, in \( \mathbb{R}^3 \), the span of a single non-zero vector \( v = (a, b, c) \) is just the straight line through the point \( (a, b, c) \) and the origin.

### Lecture 3

1. **Vector spaces contd.**

Examples of spans contd.
2. Let \( v_1 = (1, 1, 2) \) and \( v_2 = (-1, 3, 1) \) in \( \mathbb{R}^3 \). Does \((1, 0, 1)\) belong to \( \text{span} \{v_1, v_2\} \)? Try to write \((1, 0, 1)\) as \( \lambda v_1 + \mu v_2 \) for some \( \lambda, \mu \in \mathbb{R} \); that is, write

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

Equating coordinates, this gives

\[
\begin{align*}
\lambda - \mu &= 1, \\
\lambda + 3\mu &= 0, \\
2\lambda + \mu &= 1.
\end{align*}
\]

The first two equations give \( \lambda = 3/4, \mu = -1/4 \), but then \( 2\lambda + \mu \neq 1 \) so the third equation is not satisfied. Thus no such \( \lambda, \mu \) exist and \((1, 0, 1)\) does not belong to \( \text{span} \{v_1, v_2\} \).

3. Let \( P \) denote the vector space of all real polynomials of arbitrary degree in a variable \( t \), with the natural definitions of addition and scalar multiplication. Then

\[
\text{span} \{1, t, t^2, \ldots, t^n\} = P_n,
\]

the vector space of all such polynomials with degree less than or equal to \( n \).

**Spanning sets**

**Definition** A non-empty subset \( S \) of a vector space \( V \) is a spanning set for \( V \) (or spans \( V \)) if \( \text{span} \ S = V \).

**Examples**

1. Let \( v_1 = (1, 0, 0), v_2 = (0, 1, 0), v_3 = (1, 1, 1) \) in \( \mathbb{R}^3 \). Then

\[
(x_1, x_2, x_3) = (x_1 - x_3)v_1 + (x_2 - x_3)v_2 + x_3v_3
\]

and so \( \{v_1, v_2, v_3\} \) spans \( \mathbb{R}^3 \).

2. Let \( v_1 = (1, 0), v_2 = (0, 2), v_3 = (3, 1) \) in \( \mathbb{R}^2 \). Then

\[
(x_1, x_2) = x_1v_1 + (x_2/2)v_2 + 0v_3
\]

or

\[
(x_1, x_2) = \frac{x_1}{2}v_1 + \left(\frac{x_2}{2} - \frac{x_1}{12}\right)v_2 + \frac{x_1}{6}v_3.
\]

So \( \text{span} \{v_1, v_2, v_3\} = \mathbb{R}^2 \) here.

**Note**

In Example 1 above, the expression for \( (x_1, x_2, x_3) \) as a linear combination of \( \{v_1, v_2, v_3\} \) is unique. However, in Example 2, the expression for \( (x_1, x_2) \) as a linear combination of \( \{v_1, v_2\} \) is not unique. Indeed, we have

\[
\text{span} \{v_1, v_2\} = \mathbb{R}^2
\]

and so \( v_3 \) is not needed to span \( \mathbb{R}^2 \). This redundancy in a spanning set will lead to the notion of a basis of a vector space.

**Linear dependence and independence**

**Definition** A set of vectors \( \{v_1, v_2, \ldots, v_k\} \) in a vector space \( V \) over \( \mathbb{F} \) is said to be linearly dependent if there exist \( \lambda_1, \lambda_2, \ldots, \lambda_k \in \mathbb{F} \), not all zero, such that

\[
\lambda_1v_1 + \lambda_2v_2 + \cdots + \lambda_kv_k = 0
\]
(i.e. there is a non-trivial linear combination of the $v_j$'s equal to zero).

The set $\{v_1, v_2, \ldots, v_k\}$ is *linearly independent* if it is not linearly dependent.

(We tacitly assume here that the vectors $v_1, v_2, \ldots, v_k$ are distinct.)

**Comments**

1. If some $v_j = 0$, then $\{v_1, \ldots, v_k\}$ is linearly dependent since in that case we have

$$0.v_1 + \cdots + 0.v_{j-1} + 1.v_j + 0.v_{j+1} + \cdots + 0.v_k = 0.$$ 

2. The definition of linear independence can be reformulated as: $\{v_1, \ldots, v_k\}$ is linearly independent if

$$\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_k v_k = 0 \Rightarrow \lambda_1 = \lambda_2 = \cdots = \lambda_k = 0.$$ 

3. If $v \neq 0$ then the set $\{v\}$ is linearly independent.

**Proposition (1.3)** The vectors $v_1, \ldots, v_k$ are linearly dependent if and only if one of them is a linear combination of the others.

**Examples**

1. Let $v_1 = (1, 2, -1)$, $v_2 = (0, 1, 2)$, $v_1 = (-1, 1, 7)$ in $\mathbb{R}^3$. Are the vectors $v_1, v_2, v_3$ linearly dependent or linearly independent?

Consider $\lambda_1 v_1 + \lambda_2 v_2 + \lambda_3 v_3 = 0$. Equating coordinates leads to

$$\lambda_1 - \lambda_3 = 0; \quad 2\lambda_1 + \lambda_2 + \lambda_3 = 0; \quad -\lambda_1 + 2\lambda_2 + 7\lambda_3 = 0;$$

and then $\lambda_1 = \lambda_3$ (from the first equation), $3\lambda_1 + \lambda_2 = 0$ (from the second) and $6\lambda_1 + 2\lambda_2 = 0$ (from the third). So we get (taking $\lambda_1 = 1$ for instance), the non-trivial solution $\lambda_1 = \lambda_3 = 1$ and $\lambda_2 = -3$. Hence

$$v_1 - 3v_2 + v_3 = 0$$

and the vectors $v_1, v_2, v_3$ are linearly dependent.

2. The vectors $v_1 = (1, 1, 1, -1)$ and $v_2 = (2, 1, 1, 1)$ in $\mathbb{R}^4$ are linearly independent since neither is a scalar multiple of the other.

**Comment**

Although in this course we shall primarily be concerned with the notions of linear dependence and independence for finite sets of vectors, these notions extend to sets that may be infinite. Specifically, given an arbitrary (possibly infinite) set $S$ in a vector space $V$ over $F$, $S$ is said to be *linearly dependent* if, for some $k \in \mathbb{N}$, there exist distinct $v_1, \ldots, v_k$ in $S$ and $\lambda_1, \ldots, \lambda_k$ in $F$, with some $\lambda_j \neq 0$, such that

$$\lambda_1 v_1 + \cdots + \lambda_k v_k = 0;$$

and $S$ is *linearly independent* if it is not linearly dependent.

**Bases and dimension**

**Definition** A set of vectors $B$ in a vector space $V$ is called a basis of $V$ if

(i) span $B = V$, and (ii) $B$ is linearly independent.

**Examples**
1. Let \( v_1 = (1, 0) \) and \( v_2 = (0, 1) \) in \( \mathbb{R}^2 \). Then \( B = \{ v_1, v_2 \} \) is a basis of \( \mathbb{R}^2 \).

Lecture 4

1. Vector spaces contd.

Examples of bases contd.

2. Let \( B = \{ v_1, v_2 \} \) where \( v_1 = (3, 1) \) and \( v_2 = (1, -2) \) in \( \mathbb{R}^2 \). Here, if we try to find \( \lambda_1 \) and \( \lambda_2 \) such that

\[
(x_1, x_2) = \lambda_1 (3, 1) + \lambda_2 (1, -2)
\]

we find (solving the two coordinate equations) that we get

\[
(x_1, x_2) = \frac{2x_1 + x_2}{7} (3, 1) + \frac{x_1 - 3x_2}{7} (1, -2),
\]

giving \( \text{span} B = \mathbb{R}^2 \). Also, \( \{ v_1, v_2 \} \) is linearly independent (neither vector is a scalar multiple of the other). So \( B \) is a basis of \( \mathbb{R}^2 \).

3. \( B = \{ 1, t, t^2, \ldots, t^n \} \) is a basis of \( P_n \). It is clear that \( \text{span} B = P_n \). For linear independence, suppose that \( \lambda_0.1 + \lambda_1.t + \cdots + \lambda_n.t^n = 0 \). Differentiate \( k \) times (\( 0 \leq k \leq n \)) and put \( t = 0 \) to get \( \lambda_k k! = 0 \). Hence \( \lambda_k = 0 \) for \( k = 0, \ldots, n \) and \( B \) is linearly independent. Thus \( B \) is a basis of \( P_n \).

(1.4) Theorem Let \( B = \{ v_1, \ldots, v_n \} \) in a vector space \( V \) over \( \mathbb{F} \). Then \( B \) is a basis of \( V \) if and only if each element \( x \in V \) can be written uniquely as

\[
x = \lambda_1 v_1 + \cdots + \lambda_n v_n,
\]

with \( \lambda_1, \ldots, \lambda_n \in \mathbb{F} \).

Informally, the way to think of a basis \( B \) of \( V \) is that it has two properties:

(i) \( \text{span} B = V \), so \( B \) is big enough to generate \( V \);

(ii) \( B \) is linearly independent, so we cannot omit any element of \( B \) and still have the span equal to \( V \) (using Proposition (1.3)).

The aim now is to show that, if a particular basis of \( V \) has \( n \) elements, then every basis will have exactly \( n \) elements and then we are able to define the dimension of \( V \) to be \( n \). To do this, we need the following result.

(1.5) Theorem (Exchange Theorem) Let \( \{ u_1, \ldots, u_m \} \) span \( V \) and let \( \{ v_1, \ldots, v_k \} \) be a linearly independent subset of \( V \). Then

(i) \( k \leq m \);

(ii) \( V \) is spanned by \( v_1, \ldots, v_k \) and \( m - k \) of the \( u \)'s.

To prove this result, we use the following simple lemma.

(1.6) Lemma Let \( S \subseteq V \) and let \( u \in S \). Suppose that \( u \) is a linear combination of the other vectors in \( S \), i.e. \( u \in \text{span} (S \setminus \{ u \}) \). Then

\[
\text{span} S = \text{span} (S \setminus \{ u \}).
\]

Corollary of (1.5) Let \( \{ v_1, \ldots, v_n \} \) be a basis of \( V \). Then every basis of \( V \) has exactly \( n \) elements.
1. Vector spaces contd.

Definition  Let $V$ be a vector space over $\mathbb{F}$. If $V$ has a basis with $n$ elements, then we define the dimension of $V$, denoted by $\dim V$, to be $n$. [This is a well-defined concept by the Corollary to Theorem (1.5).] Otherwise, we say that $V$ is infinite dimensional.

Examples of bases and dimension

1. In $\mathbb{R}^n$, let $e_j = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$, where the 1 is in the $j^{\text{th}}$ coordinate. Then \{e_1, \ldots, e_n\} is a basis of $\mathbb{R}^n$ and $\dim \mathbb{R}^n = n$. This basis is often referred to as the standard basis of $\mathbb{R}^n$.

2. From an earlier example, put $v_1 = (3, 1)$ and $v_2 = (1, -2)$ in $\mathbb{R}^2$. Then \{v_1, v_2\} spans $\mathbb{R}^2$ and \{v_1, v_2\} is clearly linearly independent since neither vector is a multiple of the other. Thus \{v_1, v_2\} is also a basis of $\mathbb{R}^2$. [Note the illustration of the Corollary of Theorem (1.5) here.]

3. The set \{1, $t$, $t^2$, \ldots, $t^n$\} is a basis of $P_n$; so $\dim P_n = n + 1$.

4. The vector space, $P$, of real polynomials of arbitrary degree is infinite dimensional. [If it had a basis \{p_1, \ldots, p_n\}, then every polynomial in $P$ would have degree no more than the maximum of the degrees of $p_1, \ldots, p_n$, contradicting the fact that there are polynomials of arbitrarily large degree.]

5. Find a basis of $U = \{x = (x_1, x_2, x_3, x_4) \in \mathbb{R}^4 : x_1 + x_2 + x_3 + x_4 = 0\}$. Putting $x_4 = -x_1 - x_2 - x_3$ leads to the vectors \{(1, 0, 0, -1), (0, 1, 0, -1), (0, 0, 1, -1)\} spanning $U$ and these vectors are easily checked to be linearly independent, so form a basis of $U$; $\dim U = 3$.

Convention
If $V = \{0\}$, we say that the empty set $\emptyset$ is a basis of $V$ and that $\dim V = 0$.

The following results are simple consequences of the Exchange Theorem (1.5). This simple lemma (see Tut.Qn. 3.1) is useful in their proof.

(1.7) Lemma Let \{v_1, \ldots, v_n\} be a linearly independent subset of a vector space $V$ and let $v \in V$. If $v \notin \{v_1, \ldots, v_n\}$, then \{v, v_1, \ldots, v_n\} is also linearly independent.

(1.8) Theorem Let $\dim V = n$.

(i) A linearly independent subset of $V$ has most $n$ elements; if it has $n$ elements then it is a basis of $V$.

(ii) A spanning subset of $V$ has least $n$ elements; if it has $n$ elements then it is a basis of $V$.

(iii) A subset of $V$ containing at least $n + 1$ elements is linearly dependent; a subset of $V$ with at most $n - 1$ elements cannot span $V$. 

8
(iv) Let $U$ be a subspace of $V$. Then $U$ is finite dimensional and $\dim U \leq \dim V$. If $\dim U = \dim V$, then $U = V$.

(1.9) Theorem (Extension to a basis) Let $\{v_1, \ldots, v_k\}$ be a linearly independent in a finite dimensional vector space $V$. Then there exists a basis of $V$ containing $\{v_1, \ldots, v_k\}$ (i.e. vectors can be added to $\{v_1, \ldots, v_k\}$ to obtain a basis).

Lecture 6

1. Vector spaces contd.

Example

Let $v_1 = (1, 0, 1, 0)$ and $v_2 = (1, 1, 0, 1)$. Extend $\{v_1, v_2\}$ to obtain a basis of $\mathbb{R}^4$. (The set $\{v_1, v_2\}$ is clearly linearly independent since both vectors in it are non-zero and $v_2$ is not a multiple of $v_1$.) To do this, apply the procedure detailed in the proof of the Exchange Theorem (1.5), starting with the linearly independent set $\{v_1, v_2\}$ and the spanning set comprising the standard basis $\{e_1, e_2, e_3, e_4\}$ of $\mathbb{R}^4$. This will yield the basis $\{v_1, v_2, e_3, e_4\}$ of $\mathbb{R}^4$.

In a similar way, a spanning set can be reduced to a basis.

(1.10) Theorem Let $S$ be a finite set that spans a vector space $V$. Then some subset of $S$ forms a basis of $V$ and (hence) $V$ is finite dimensional.

Comment

The proof of this result actually shows that, given a finite set $S$ in a vector space, there is a linearly independent subset $S'$ of $S$ with span $S' = \text{span } S$ (so that $S'$ is a basis of $\text{span } S$).

Example

Let $v_1 = (1, 0, 1)$, $v_2 = (0, 1, 1)$, $v_3 = (1, -1, 0)$ and $v_4 = (1, 0, 0)$. Find a linearly independent subset of $\{v_1, v_2, v_3, v_4\}$ with the same span. Note that $\{v_1, v_2\}$ is linearly independent but $v_3 = v_1 - v_2$. So $\text{span } \{v_1, v_2, v_3, v_4\} = \text{span } \{v_1, v_2, v_4\}$. Now check whether or not $v_4$ is a linear combination of $v_1$ and $v_2$. Setting $v_4 = \lambda_1 v_1 + \lambda_2 v_2$ leads (equating coordinates) to $\lambda_1 = 1$, $\lambda_2 = 0$ and $\lambda_1 + \lambda_2 = 0$, so no such $\lambda_1, \lambda_2$ exist and $v_4$ is not a linear combination of $v_1$ and $v_2$. Thus $\{v_1, v_2, v_4\}$ is linearly independent and has the same span as $\{v_1, v_2, v_3, v_4\}$. (This span will be all of $\mathbb{R}^3$ since dim $\mathbb{R}^3 = 3$.)

Sums of subspaces

Definition The sum $U_1 + U_2$ of two subspaces $U_1, U_2$ of a vector space $V$ is defined as

$$U_1 + U_2 = \{u_1 + u_2 : u_1 \in U_1 \text{ and } u_2 \in U_2\}.$$

Compare the following result with that of Theorem (1.2).

(1.11) Theorem $U_1 + U_2$ is a subspace of $V$ containing $U_1$ and $U_2$. Further, if $W$ is a subspace of $V$ with $U_1, U_2 \subseteq W$, then $U_1 + U_2 \subseteq W$. Hence $U_1 + U_2$ is the smallest subspace of $V$ that contains both $U_1$ and $U_2$.

Examples

1. In $\mathbb{R}^2$, let $U_1 = \text{span } \{v_1\}$ and $U_2 = \text{span } \{v_2\}$, where $v_1, v_2 \neq 0$ and neither is a multiple of the other. Then $U_1 + U_2 = \mathbb{R}^2$. Note that each $v \in \mathbb{R}^2$ can be written uniquely as $v = u_1 + u_2$ with $u_1 \in U_1$ and $u_2 \in U_2$.  


2. In $P_4$, let $U_1$ be the subspace of all polynomials of degree less than or equal to 3 and let $U_2$ be the subspace of all even polynomials of degree less than or equal to 4. Then $U_1 + U_2 = P_4$. Note that, here, the expression of a given $p \in P_4$ as $p = p_1 + p_2$, with $p_1 \in U_1$ and $p_2 \in U_2$, is not unique.

Direct sums

Suppose that $U_1 \cap U_2 = \{0\}$. Then each $u \in U_1 + U_2$ can be expressed uniquely as $u = u_1 + u_2$, with $u_1 \in U_1$ and $u_2 \in U_2$. In this situation, we call write $U_1 \oplus U_2$ for $U_1 + U_2$ and call this the direct sum of $U_1$ and $U_2$.

Comment

In Example 1 above, $\mathbb{R}^2 = U_1 \oplus U_2$ whilst, in Example 2, $P_4$ is not the direct sum of $U_1$ and $U_2$.

(1.12) Theorem (The dimension theorem) Let $U_1$ and $U_2$ be subspaces of a finite dimensional vector space. Then

(i) $\dim(U_1 + U_2) = \dim U_1 + \dim U_2 - \dim(U_1 \cap U_2)$;

(ii) if $U_1 \cap U_2 = \{0\}$, then $\dim(U_1 \oplus U_2) = \dim U_1 + \dim U_2$.

Lecture 7

1. Vector spaces contd.

Comment

Compare the formula in Theorem (1.12)(i) with the formula

$$\#(A \cup B) = \#(A) + \#(B) - \#(A \cap B)$$

for finite sets $A, B$, where $\#$ denotes ‘number of elements in’.

Exercise

Verify the results of Theorem (1.12) in the cases of the Examples 1 and 2 of sums of subspaces in Lecture 6.

Coordinates

Let $B = \{v_1, \ldots, v_n\}$ be a basis of a vector space $V$. Then each $v \in V$ can be written uniquely as

$$v = \lambda_1 v_1 + \cdots + \lambda_n v_n$$

with the coefficients $\lambda_j \in \mathbb{F}$. We call $\lambda_1, \ldots, \lambda_n$ the coordinates of $v$ with respect to the basis $B$; we also call

(a) the row vector $(\lambda_1, \ldots, \lambda_n)$ the coordinate row vector of $v$ w.r.t. $B$; and

(b) the column vector

$$\begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}$$

the coordinate column vector of $v$ w.r.t. $B$.

Note: The order in which the vectors in $B$ are listed is important when we consider row or column coordinate vectors.

Examples
1. Let $B$ be the standard basis $\{e_1, \ldots, e_n\}$ in $\mathbb{F}^n$ ($e_j = (0, 0, \ldots, 0, 1, 0, \ldots, 0)$ with the $1$ in the $j^{th}$ place). If $x = (x_1, \ldots, x_n) \in \mathbb{F}^n$, then the column coordinate vector of $x$ w.r.t. $B$ is just $\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$.

2. In $\mathbb{R}^2$, consider the basis $B = \{(1, 0), (2, 3)\}$. Then the coordinate column vector of $(4, 9) \in \mathbb{R}^2$ w.r.t. $B$ is $\left( \begin{array}{c} -2 \\ 3 \end{array} \right)$ since $(4, 9) = -2(1, 0) + 3(2, 3)$.

**Change of basis**

Let $B_1 = \{v_1, \ldots, v_n\}$ and $B_2 = \{w_1, \ldots, w_n\}$ be two bases of a vector space $V$. Express each vector in $B_2$ in terms of $B_1$ thus:

$$w_1 = p_{11}v_1 + p_{21}v_2 + \cdots + p_{n1}v_n$$
$$w_2 = p_{12}v_1 + p_{22}v_2 + \cdots + p_{n2}v_n$$

$$\begin{array}{c}
\vdots \\
\end{array}$$

$$(*)$$
$$w_n = p_{1n}v_1 + p_{2n}v_2 + \cdots + p_{nn}v_n.$$  

Note that $(*)$ can be written as

$$w_j = \sum_{i=1}^{n} p_{ij}v_i$$

for $j = 1, \ldots, n$. We call the matrix

$$P = \begin{pmatrix}
    p_{11} & p_{12} & \cdots & p_{1n} \\
p_{21} & p_{22} & \cdots & p_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
p_{n1} & p_{n2} & \cdots & p_{nn}
\end{pmatrix}$$

the *change of basis matrix* from $B_1$ to $B_2$. It is important to note that the columns of the matrix $P$ are the rows of the coefficients in $(*)$ (or, equivalently, the columns of $P$ are the column coordinate vectors of $w_1, \ldots, w_n$ w.r.t. $B_1$).

Note that, writing $B_1$ and $B_2$ as rows $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_n)$ respectively, $(*)$ can be written as

$$(w_1, \ldots, w_n) = (v_1, \ldots, v_n)P.$$  

**Example**

Let $B_1 = \{e_1, e_2, e_3\}$ be the standard basis of $\mathbb{R}^3$ and let $B_2 = \{w_1, w_2, w_3\}$ be the basis given by

$$w_1 = (1, 0, 1) ; \ w_2 = (1, 1, 0) ; \ w_3 = (1, 1, 1).$$

Then

$$w_1 = 1.e_1 + 0.e_2 + 1.e_3$$
$$w_2 = 1.e_1 + 1.e_2 + 0.e_3$$
$$w_3 = 1.e_1 + 1.e_2 + 1.e_3.$$
and so the change of basis matrix from $B_1$ to $B_2$ is

$$P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}.$$ 

We also have

$$e_1 = 1.w_1 + 1.w_2 - 1.w_3$$
$$e_2 = -1.w_1 + 0.w_2 + 1.w_3$$
$$e_3 = 0.w_1 - 1.w_2 + 1.w_3,$$

so the change of basis matrix from $B_2$ to $B_1$ is

$$Q = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ -1 & 1 & 1 \end{pmatrix}.$$ 

Lecture 8

1. Vector spaces contd.

Comments

1. Notice that, in the example at the end of Lecture 7, the columns of $P$ are just the elements of $B_2$ written as column vectors. This true is true more generally.

Important fact

Let $B_1 = \{e_1, \ldots, e_n\}$ be the standard basis of $\mathbb{F}^n$ and let $B_2 = \{v_1, \ldots, v_n\}$ be another basis. Suppose that

$$v_1 = (\alpha_{11}, \alpha_{21}, \ldots, \alpha_{n1}), \quad v_2 = (\alpha_{12}, \alpha_{22}, \ldots, \alpha_{n2}), \ldots, \quad v_n = (\alpha_{1n}, \alpha_{2n}, \ldots, \alpha_{nn})$$

(that is, $v_j = (\alpha_{1j}, \alpha_{2j}, \ldots, \alpha_{nj})$ for $j = 1, \ldots, n$). Then the change of basis matrix from $B_1$ to $B_2$ is the matrix

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1n} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nn} \end{pmatrix},$$

the columns of which are just the vectors $v_j$ written as column vectors.

2. Note also that, in this same example, $PQ = QP = I_3$, the $3 \times 3$ identity matrix. Hence the change of basis matrix $P$ from $B_1$ to $B_2$ is invertible and $P^{-1} = Q$, the change of basis matrix from $B_2$ to $B_1$. This is true in general.

(1.13) Theorem Let $B_1 = \{v_1, \ldots, v_n\}$ and $B_2 = \{w_1, \ldots, w_n\}$ be two bases of a vector space $V$. Let $P$ be the change of basis matrix from $B_1$ to $B_2$ and $Q$ the change of basis matrix from $B_2$ to $B_1$. Then $P$ is invertible and $Q = P^{-1}$. 

12
The importance of a change of basis matrix is that it gives a way of relating the column coordinates of a vector relative to two bases. Note that we can express the fact that \( \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} \) is the coordinate column vector of a vector \( v \) w.r.t. a basis \( \{v_1, \ldots, v_n\} \) by writing

\[
v = (v_1, \ldots, v_n) \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix}.
\]

Formally, (*) denotes \( v = v_1 \lambda_1 + \cdots + v_n \lambda_n \), but we interpret this as the usual expression \( v = \lambda_1 v_1 + \cdots + \lambda_n v_n \) for \( v \) in terms of \( \{v_1, \ldots, v_n\} \).

**Theorem (1.14)** Let \( B_1 = \{v_1, \ldots, v_n\} \) and \( B_2 = \{w_1, \ldots, w_n\} \) be two bases of a vector space \( V \). Let \( P \) be the change of basis matrix from \( B_1 \) to \( B_2 \) and \( Q \) the change of basis matrix from \( B_2 \) to \( B_1 \) (so that \( Q = P^{-1} \)). Let \( v \in V \) have coordinate column vector \( \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} \) w.r.t. \( B_1 \) and coordinate column vector \( \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} \) w.r.t. \( B_2 \). Then

\[
\begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix} = Q \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{bmatrix} = P \begin{bmatrix} \mu_1 \\ \vdots \\ \mu_n \end{bmatrix}.
\]

**Important Comment**

Note that, although \( P \) expresses \( B_2 \) in terms of \( B_1 \), it is the inverse \( P^{-1} \) that is used to express \( B_2 \)-coordinates in terms of \( B_1 \)-coordinates.

**Examples**

1. Let \( B_1 \) be the standard basis of \( \mathbb{R}^2 \) and let \( B_2 \) be the basis \( \{(3, -1), (1, 1)\} \). The change of basis matrix from \( B_1 \) to \( B_2 \) is

\[
P = \begin{pmatrix} 3 & 1 \\ -1 & 1 \end{pmatrix}
\]

and that from \( B_2 \) to \( B_1 \) is

\[
P^{-1} = \frac{1}{4} \begin{pmatrix} 1 & -1 \\ 1 & 3 \end{pmatrix}.
\]

[General fact: The \( 2 \times 2 \) matrix \( \begin{pmatrix} a & b \\ c & d \end{pmatrix} \) is non-singular (invertible) if and only if \( ad - bc \neq 0 \) and in this case

\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.
\]

Thus the coordinates of \((x_1, x_2) \in \mathbb{R}^2\) w.r.t. \( B_2 \) are given by

\[
\frac{1}{4} \begin{pmatrix} 1 & -1 \\ 1 & 3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \frac{1}{4} \begin{pmatrix} x_1 - x_2 \\ x_1 + 3x_2 \end{pmatrix}.
\]

Thus \((x_1, x_2) = \frac{1}{4}(x_1 - x_2)(3, -1) + \frac{1}{4}(x_1 + 3x_2)(1, 1)\).
2. Let $B_1$ be the standard basis of $\mathbb{R}^3$ and let $B_2$ be the basis $\{w_1, w_2, w_3\}$ where 

$w_1 = (1, 0, 1), \ w_2 = (1, 1, 0), \ w_3 = (1, 1, 1)$. Then the change of basis from $B_2$ to $B_1$ is

$$
\begin{pmatrix}
1 & -1 & 0 \\
1 & 0 & -1 \\
-1 & 1 & 1
\end{pmatrix}
$$

(see the example at the end of Lecture 7). Hence the coordinates of the vector $(2, 3, 1) \in \mathbb{R}^3$ w.r.t. $B_2$ are given by

$$
\begin{pmatrix}
1 & -1 & 0 \\
1 & 0 & -1 \\
-1 & 1 & 1
\end{pmatrix}
\begin{pmatrix}
2 \\
3 \\
1
\end{pmatrix} = \begin{pmatrix}
-1 \\
1 \\
2
\end{pmatrix}.
$$

Thus $(2, 3, 1) = -w_1 + w_2 + 2w_3$.

Lecture 9

2. Linear Mappings

Let $U$ and $V$ be two vector spaces over $\mathbb{F}$.

Definition A linear mapping (or linear transformation) from $U$ to $V$ is a map $T : U \rightarrow V$ such that

$$(*) \quad T(x + y) = Tx + Ty \text{ and } T(\lambda x) = \lambda Tx$$

for all $x, y \in U$ and all $\lambda \in \mathbb{F}$.

Notes

1. $T$ is a function defined on $U$ with values in $V$. We usually write $Tx$ rather than $T(x)$ for its value at $x \in U$.

2. Note that $(*)$ can be replaced by the single condition

$$T(\lambda x + \mu y) = \lambda Tx + \mu Ty$$

for all $x, y \in U$ and all $\lambda, \mu \in \mathbb{F}$.

Examples

1. Let $T : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ be defined by

$$T : \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} \rightarrow \begin{pmatrix} 1 & -1 & 2 \\ 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix},$$

so that $T(x_1, x_2, x_3) = (x_1 - x_2 + 2x_3, 3x_1 + x_2)$. Then $T$ is linear.

2. $T : P_n \rightarrow P_{n+1}$ given by $(Tp)(t) = tp(t)$ is linear.

3. $T : P_n \rightarrow P_n$ given by $(Tp)(t) = p'(t) + 2p(t)$ is linear.
4. The linear mappings \( F^n \to F^m \) are precisely the mappings of the form

\[
T : \begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n \
\end{pmatrix} \to A \begin{pmatrix}
x_1 \\
x_2 \\
\vdots \\
x_n 
\end{pmatrix},
\]

where \( A \) is an \( m \times n \) matrix (fixed for a given \( T \)).

**Technical terms**

Let \( T : U \to V \) be a linear mapping.

(a) If \( Tu = v \) (\( u \in U, v \in V \)), then \( v \) is the image of \( u \) under \( v \).

(b) \( T \) is one-one (or injective) if, for \( u_1, u_2 \in U, u_1 \neq u_2 \Rightarrow Tu_1 \neq Tu_2 \) (equivalently, \( Tu_1 = Tu_2 \Rightarrow u_1 = u_2 \)).

(c) \( T \) is onto \( V \) (or surjective) if, given \( v \in V, v = Tu \) for some \( u \in U \).

(d) The kernel of \( T \) is the set

\[
\ker T = \{ u \in U : Tu = 0 \}
\]

(e) The image of \( T \) is the set

\[
\text{im } T = \{ Tu : u \in U \};
\]

\( \text{im } T \) is also called the range of \( T \), sometimes denoted by \( \text{ran } T \).

**Theorem**

Let \( T : U \to V \) be linear. Then

(i) \( T0 = 0 \);

(ii) \( \ker T \) and \( \text{im } T \) are subspaces of \( U \) and \( V \) respectively;

(iii) \( T \) is one-one (or injective) if and only if \( \ker T = \{0\} \).

**Example**

Let \( T : \mathbb{R}^2 \to \mathbb{R}^3 \) be given by

\[
T \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ -1 & 0 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix},
\]

so that \( T(x_1, x_2) = (x_1 + 2x_2, -x_1, x_1 + x_2) \). Then \( \ker T = \{0\} \), so that \( T \) is one-one, whilst \( \text{im } T = \text{span}\{(1, -1, 1), (2, 0, 1)\} \).

**Note**

This example illustrates the important fact that, if \( T : F^n \to F^m \) is a linear mapping defined by

\[
T \begin{pmatrix} x_1 \\ x_2 \\
\vdots \\
x_n \end{pmatrix} = A \begin{pmatrix} x_1 \\ x_2 \\
\vdots \\
x_n \end{pmatrix},
\]

where \( A \) is an \( m \times n \) matrix, then \( \text{im } T \) equals the span (in \( F^m \)) of the columns of \( A \).
2. Linear Mappings contd.

The matrix of a linear mapping

Let \( T : U \to V \) be linear, and suppose also that \( B_1 = \{ u_1, \ldots, u_n \} \) is a basis of \( U \) and that \( B_2 = \{ v_1, \ldots, v_m \} \) is a basis of \( V \). The linearity of \( T \) means that \( T \) is completely determined by the values of \( Tu_1, Tu_2, \ldots, Tu_n \). Suppose then that

\[
Tu_1 = a_{11}v_1 + a_{21}v_2 + \cdots + a_{m1}v_m \\
Tu_2 = a_{12}v_1 + a_{22}v_2 + \cdots + a_{m2}v_m \\
\vdots \\
Tu_n = a_{1n}v_1 + a_{2n}v_2 + \cdots + a_{mn}v_m.
\]

Note that (*) can be written as

\[
Tu_j = \sum_{i=1}^{m} a_{ij}v_i 
\]

for \( j = 1, \ldots, n \). We call the matrix

\[
A = \begin{pmatrix}
    a_{11} & a_{12} & \cdots & a_{1n} \\
    a_{21} & a_{22} & \cdots & a_{2n} \\
    \vdots & \vdots & \ddots & \vdots \\
    a_{m1} & a_{m2} & \cdots & a_{mn}
\end{pmatrix}
\]

the matrix of \( T \) with respect to \( B_1 \) and \( B_2 \).

Notes

1. The columns of \( A \) are coefficients in the rows of (*).
2. (*) can be written as

\[
(Tu_1, \ldots, Tu_n) = (v_1, \ldots, v_m) A.
\]
3. \( A \) is an \( m \times n \) matrix, where \( n = \dim U \) and \( m = \dim V \).
4. If \( U = V \) and \( B_1 = B_2 = B \), we just refer to \( A \) as the matrix of \( T \) w.r.t. \( B \).
5. Let \( B_1 \) and \( B_2 \) be two bases of \( U \), let \( T \) be the identity mapping \( Tx = x \) \((x \in U)\) and consider \( T \) as a linear mapping from \( U \) equipped with basis \( B_1 \) to \( U \) equipped with basis \( B_2 \). Then the matrix of \( T \) is the change of basis matrix from \( B_2 \) to \( B_1 \).

Examples

1. Let \( T \) be linear mapping \( P_n \to P_{n+1} \) defined by \( (Tp)(t) = tp(t) \). Then the matrix of \( T \) w.r.t. the bases \( \{1, t, t^2, \ldots, t^n\} \) and \( \{1, t, t^2, \ldots, t^{n+1}\} \) of \( P_n \) and \( P_{n+1} \) respectively is the \((n+1) \times n\) matrix

\[
\begin{pmatrix}
    0 & 0 & 0 & \cdots & 0 \\
    1 & 0 & 0 & \cdots & 0 \\
    0 & 1 & 0 & \cdots & 0 \\
    0 & 0 & 1 & \cdots & 0 \\
    \vdots & \vdots & \vdots & \ddots & \vdots \\
    0 & 0 & \cdots & \cdots & 0 \end{pmatrix}
\]
For instance, the first two and final columns of this matrix come from noting that, since 
\[ T(1) = t, \quad T(t) = t^2 \quad \text{and} \quad T(t^n) = t^{n+1}, \]
we have
\[
T(1) = 0.1 + 1.t + 0.t^2 + \cdots + 0.t^n + 0.t^{n+1} \\
T(t) = 0.1 + 0.t + 1.t^2 + \cdots + 0.t^n + 0.t^{n+1} \\
T(t^n) = 0.1 + 0.t + 0.t^2 + \cdots + 0.t^n + 1.t^{n+1}.
\]

2. Let \( T : \mathbb{F}^n \to \mathbb{F}^m \) be given by
\[
T : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \to C \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix},
\]
where \( C \) is an \( m \times n \) matrix. Then \( C \) is the matrix of \( T \) w.r.t. the standard bases of \( \mathbb{F}^n \) and \( \mathbb{F}^m \).

3. Let \( T : \mathbb{R}^3 \to \mathbb{R}^3 \) be defined by \( T(x_1, x_2, x_3) = (x_2, x_3, x_1) \). Consider the basis \( B = \{ u_1, u_2, u_3 \} \) of \( \mathbb{R}^3 \), where
\[
 u_1 = (1, 0, 0) ; \quad u_2 = (1, 2, 3) ; \quad u_3 = (-1, 1, 1).
\]
Then \( Tu_1 = (0, 0, 1) = \lambda_1 u_1 + \lambda_2 u_2 + \lambda_3 u_3 \), where (after equating the three coordinates and solving the resulting equation for \( \lambda_1, \lambda_2, \lambda_3 \) \( \lambda_1 = -3, \lambda_2 = 1 \) and \( \lambda_3 = -3 \). Hence
\[
Tu_1 = -3u_1 + 1.u_2 - 2.u_3.
\]
Similarly,
\[
Tu_2 = (2, 3, 1) = 11u_1 - 2u_2 + 7u_3 \\
Tu_3 = (1, 1, -1) = 8u_1 - 2u_2 + 5u_3.
\]
So the matrix of \( T \) w.r.t. \( B \) is
\[
\begin{pmatrix} -3 & 11 & 8 \\ 1 & -2 & -2 \\ -2 & 7 & 5 \end{pmatrix}.
\]

Coordinates

The matrix of a linear mapping is related to the corresponding mapping of coordinates as follows.

(2.2) Theorem Let \( T : U \to V \) be linear and have matrix \( A = (a_{ij}) \) w.r.t. bases \( B_1 = \{ u_1, \ldots, u_n \} \) and \( B_2 = \{ v_1, \ldots, v_m \} \) of \( U \) and \( V \) respectively. Let \( u \in U \) have coordinate column vector
\[
\lambda = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}
\]
*w.r.t. \( B_1 \). Then \( Tu \) has coordinate column vector
\[
A\lambda = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix}
\]
Example

Let $T : U \to V$ have matrix

\[
\begin{pmatrix}
1 & 2 & -1 \\
-1 & 3 & 6
\end{pmatrix}
\]

w.r.t. bases $\{u_1, u_2, u_3\}$ of $U$ and $\{v_1, v_2\}$ of $V$. Then $T(3u_1 - u_2 + 2u_3)$ has coordinate column vector

\[
\begin{pmatrix}
1 & 2 & -1 \\
-1 & 3 & 6
\end{pmatrix}
\begin{pmatrix}
3 \\
-1
\end{pmatrix} = \begin{pmatrix}
-1 \\
6
\end{pmatrix}
\]

w.r.t. $\{v_1, v_2\}$, so that $T(3u_1 - u_2 + 2u_3) = -v_1 + 6v_2$.

Change of bases

(2.3) Theorem

Let $T : U \to V$ be linear, with matrix $A$ w.r.t. bases $B_1$ of $U$ and $B_2$ of $V$. Let $B'_1$ be another basis of $U$ and $B'_2$ another basis of $V$. Then the matrix $C$ of $T$ w.r.t. $B'_1$ and $B'_2$ is given by

\[
C = Q^{-1}AP,
\]

where $P$ is the change of basis matrix from $B_1$ to $B'_1$ and $Q$ is the change of basis matrix from $B_2$ to $B'_2$.

Lecture 11

2. Linear Mappings contd.

The rank theorem

Let $T : U \to V$ be linear and recall that $\ker T = \{u \in U : Tu = 0\}$ is a subspace of $U$, whilst $\im T = \{Tu : u \in U\}$ is a subspace of $V$. Suppose that $U$ and $V$ are finite dimensional.

Definitions

(i) The nullity of $T$, $\text{null}(T)$, is defined by

\[\text{null}(T) = \dim \ker T.\]

(ii) The rank of $T$, $\text{rk}(T)$, is defined by

\[\text{rk}(T) = \dim \im T.\]

[Alternative notation: $\text{n}(T)$ and $\text{r}(T)$]

(2.4) Theorem (The rank theorem) $\text{rk}(T) + \text{null}(T) = \dim U$.

Example

Let $T : \mathbb{R}^3 \to \mathbb{R}^4$ be defined by

\[T(x_1, x_2, x_3) = (x_1 - x_2, x_2 - x_3, x_3 - x_1, x_1 + x_2 - 2x_3).\]

Then

\[x = (x_1, x_2, x_3) \in \ker T \iff x_1 - x_2 = x_2 - x_3 = x_3 - x_1 = x_1 + x_2 - 2x_3 = 0 \]

\[\iff x_1 = x_2 = x_3 \iff x = x_1(1, 1, 1);\]
thus \( \ker T = \text{span}\{(1, 1, 1)\} \) and \( \text{null}(T) = 1 \). Writing \((x_1 - x_2, x_2 - x_3, x_3 - x_1, x_1 + x_2 - 2x_3)\) as

\[
x_1(1, 0, -1, 1) + x_2(-1, 1, 0, 1) + x_3(0, -1, 1, -2),
\]
we see that

\[
\text{im } T = \text{span}\{(1, 0, -1, 1), (-1, 1, 0, 1), (0, -1, 1, -2)\}.
\]
The first two of these vectors are clearly linearly independent but

\[
(0, -1, 1, -2) = -(1, 0, -1, 1) - (-1, 1, 0, 1).
\]
Hence

\[
\text{im } T = \text{span}\{(1, 0, -1, 1), (-1, 1, 0, 1)\},
\]
\{(1, 0, -1, 1), (-1, 1, 0, 1)\} is a basis of \( \text{im } T \) and \( \text{rk}(T) = 2 \). Thus

\[
\text{rk}(T) + \text{null}(T) = 2 + 1 = 3 = \dim \mathbb{R}^3
\]
as expected.

Note: \( \dim V \) does not appear in the statement of the rank theorem; this is not surprising as the space into which \( T \) is mapping could be adjusted (say, enlarged) without changing \( \text{im } T \).

Corollary to the rank theorem Let \( T : U \rightarrow V \) be linear, where \( U \) and \( V \) are finite dimensional. We have the following.

(i) \( \text{rk}(T) \leq \dim U \).

(ii) Suppose that \( \dim U = \dim V \). Then \( T \) is one-one (i.e. is injective) if and only if it maps \( U \) onto \( V \) (i.e. is surjective). In this case, \( T \) is bijective and the inverse mapping \( T^{-1} : V \rightarrow U \) is also linear.

Comments

1. It is clear that \( \text{rk}(T) \leq \dim V \). Result (i) is saying something less obvious.

2. When \( T \) is bijective as in (ii), we say that \( T \) is invertible or is a linear isomorphism.

Rank and nullity of a matrix

Let \( A \) be an \( m \times n \) matrix. Then there is an associated linear mapping \( T_A : \mathbb{F}^n \rightarrow \mathbb{F}^m \) given by

\[
T_A : \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \rightarrow A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.
\]

The rank of \( A \), \( \text{rk}(A) \), is defined to be the rank of \( T_A \) and the nullity of \( A \), \( \text{null}(A) \), is defined to be the nullity of \( T_A \).

Since \( \text{im } T_A \) is the span in \( \mathbb{F}^m \) of the columns of \( A \), \( \text{rk}(A) \) is sometimes referred to as the column rank of \( A \). It equals the maximum number of linearly independent columns of \( A \). One can also consider the maximum number of linearly independent rows of \( A \) (equivalently, the dimension of the span in \( \mathbb{F}^n \) of the rows of \( A \)). This is called the row rank of \( A \). It can be shown, though this is not at all obvious, that the row and column ranks of a matrix are always equal (Tut. Qn. 7.1 will indicate how this is proved).
Notice also that, if \( A = (a_{ij}) \), then \( \text{null}(A) \) is the maximum number of linearly independent solutions of the system of linear equations

\[
\begin{align*}
  a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= 0 \\
  a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= 0 \\
  \vdots \\
  a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= 0.
\end{align*}
\]

Sums and products of linear mappings

Let \( S, T : U \rightarrow V \) be two linear mappings. Then we define \( S + T : U \rightarrow V \) by

\[(S + T)u = Su + Tu \quad (u \in U)\]

and, for \( \lambda \in \mathbb{F} \), we define \( \lambda S : U \rightarrow V \) by

\[(\lambda S)u = \lambda Su \quad (u \in U).\]

It is easy to check that \( S + T \) and \( \lambda S \) are linear mappings and that, if \( S \) has matrix \( A \) and \( T \) has matrix \( B \) w.r.t. some bases of \( U \) and \( V \), then \( S + T \) has matrix \( A + B \) and \( \lambda S \) has matrix \( \lambda A \) w.r.t. the same bases.

**Comment:** Denote by \( L(U, V) \) the set of all linear mappings from \( U \) into \( V \). Then, with the above definitions of addition and scalar multiplication, \( L(U, V) \) is itself a vector space.

Suppose now that \( T : U \rightarrow V \) and \( S : V \rightarrow W \) are linear. We get the composition \( R = S \circ T : U \rightarrow W \), defined by

\[Ru = S(Tu) \quad (u \in U).\]

Again it is easy to check that \( R \) is linear. For linear mappings, we normally write such a composition as \( R = ST \) and call \( R \) the product of \( S \) and \( T \).

**Example**

Let \( T : \mathbb{R}^2 \rightarrow \mathbb{R}^3 \) and \( S : \mathbb{R}^3 \rightarrow \mathbb{R}^2 \) be defined by

\[
T(x_1, x_2) = (x_1 + x_2, 0, x_2) \quad \text{and} \quad S(x_1, x_2, x_3) = (x_2, x_3 - x_1).
\]

Then \( ST : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) is given by

\[
(ST)(x_1, x_2) = S(x_1 + x_2, 0, x_2) = (0, -x_1).
\]

We also have \( TS : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) given by

\[
(TS)(x_1, x_2, x_3) = T(x_2, x_3 - x_1) = (x_2 + x_3 - x_1, 0, x_3 - x_1).
\]

**Note:** \( ST \) and \( TS \) are completely different mappings.
2. Linear Mappings contd.

(2.5) Theorem Let $T : U \rightarrow V$ and $S : V \rightarrow W$ be linear mappings. Suppose that $T$ has matrix $A$ w.r.t. bases $B_1$ of $U$ and $B_2$ of $V$, and $S$ has matrix $C$ w.r.t. the bases $B_2$ of $V$ and $B_3$ of $W$. Then $ST$ has matrix $CA$ w.r.t. $B_1$ and $B_3$.

Notes

1. Given $S, T \in L(U, U)$, both products $ST$ and $TS$ are defined. However, when $\dim U > 1$, it may not be the case that $ST$ equals $TS$. This is because, when $n > 1$, there always exist $n \times n$ matrices $A, B$ such that $AB \neq BA$.

2. When $T \in L(U, U)$, we can form positive powers $T^n$ of $T$ by taking $n$ compositions of $T$ with itself. It is customary to define $T^0$ to be the identity mapping $I_U = u$ on $U$. Further, when $T$ is invertible, negative powers of $T$ are defined by setting $T^{-n} = (T^{-1})^n$ for $n = 1, 2, \ldots$. With these definitions, the expected index laws hold for the range of indices for which they make sense. For instance,

$$ (T^m)(T^n) = T^{(m+n)} $$

holds for general $T$ when $m, n \geq 0$ and holds for invertible $T$ when $m, n \in \mathbb{Z}$.

Singular and non-singular matrices

Let $A$ be an $n \times n$ matrix with entries in $\mathbb{F}^n$ and let $I_n$ be the $n \times n$ identity matrix (i.e. it has 1’s on the diagonal and 0’s elsewhere). The following is a consequence of the Corollary to the rank theorem.

(2.6) Theorem The following statements are equivalent.

(i) There exists an $n \times n$ matrix $B$ such that $BA = I_n$.

(ii) There exists an $n \times n$ matrix $B$ such that $AB = I_n$.

(iii) The columns of $A$ are linearly independent.

(iv) $Ax = 0 \Rightarrow x = 0$ [where $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$].

Definition When (i)-(iv) hold, we say that $A$ is non-singular; otherwise, $A$ is said to be singular. Thus $A$ is singular precisely when its columns are linearly dependent. It can be shown that $A$ is singular if and only if $\det A = 0$ (equivalently, $A$ is non-singular if and only if $\det A \neq 0$).

Linear Equations

Let $T : U \rightarrow V$, let $v \in V$ be given, and suppose that we are interested in solving the equation

$$ Tx = v \quad (\dagger) $$

i.e. we wish to find all $x \in U$ such that $Tx = v$. The basic simple result is the following.

(2.7) Theorem The equation $(\dagger)$ has a solution if and only if $v \in \text{im} T$. Assume that this is the case and let $x = x_0$ be some solution of $(\dagger)$. Then the general solution of $(\dagger)$ is

$$ x = x_0 + u, $$
where $u \in \ker T$.

Comment

This result is illustrated, for instance, in the following.

1. Let $A$ be an $m \times n$ matrix and consider the system of $m$ linear equations $Ax = b$ in the variables $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$, where $b = \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix}$. This has a solution if and only if $b$ is in the span of the columns of $A$ and, in that case, the general solution has the form $x = x_0 + u$, where $x_0$ is a particular solution and $u$ is the general solution of the associated homogeneous equation $Ax = 0$.

2. Let $\mathcal{L}$ be a linear differential operator (e.g., a linear combination of derivatives $d^kx/dt^k$) and consider the equation $\mathcal{L}x = f$, where $f(t)$ is some given function. Then the general solution of this equation has the form $x(t) = x_0(t) + u(t)$, where $x_0(t)$ is a solution of the equation (called, in this context, a particular integral), and $u(t)$ is the general solution of the associated homogeneous equation $\mathcal{L}x = 0$ (called the complementary function).

Eigenvalues, eigenvectors and diagonalization

Let $S : U \to U$ be a linear mapping, where $U$ is a vector space over $\mathbb{F}$.

Definition. A scalar $\lambda \in \mathbb{F}$ is an eigenvalue of $T$ if there exists $u \neq 0$ such that $Tu = \lambda u$. Such a (non-zero) vector $u$ is called an eigenvector associated with $\lambda$. The set 

$$\{ u : Tu = \lambda u \},$$

which equals $\ker(T - \lambda I)$ where $I$ is the identity operator $Iu = u$ on $U$, is called the eigenspace corresponding to $\lambda$. It is clearly a subspace of $U$ (being the kernel of a linear mapping).

We also have the following related definition.

Definition. We say that $T$ is diagonalizable if $U$ has a basis consisting of eigenvectors of $T$.

We then have the following simple result that explains the terminology.

(2.8) Theorem. $T$ is diagonalizable if and only if there exists a basis $B = \{ u_1, \ldots, u_n \}$ of $U$ w.r.t. which the matrix of $T$ is diagonal, i.e., has the form

$$\begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
& \cdots & \cdots & \cdots \\
0 & \cdots & 0 & \lambda_n
\end{pmatrix}.$$

Suppose now that $A$ is an $n \times n$ matrix, with an associated linear mapping $T_A : x \to Ax$ from $\mathbb{F}^n$ to itself. Then $T_A$ is diagonalizable if and only if there exists a non-singular $n \times n$ matrix $P$ such that $P^{-1}AP$ is diagonal. [$P$ will be the change of basis matrix from the standard basis of $\mathbb{F}^n$ to the basis consisting of eigenvectors of $T_A$. In these circumstances, we say that $A$ is diagonalizable.

(2.9) Theorem. Let $\lambda_1, \ldots, \lambda_n$ be distinct eigenvalues of $T$, with corresponding eigenvectors $u_1, \ldots, u_n$. Then $\{ u_1, \ldots, u_n \}$ is linearly independent.

Corollary. If $T : U \to U$ is linear, $\dim U = n$ and $T$ has $n$ distinct eigenvalues, then $T$ is diagonalizable.
Calculation of eigenvalues

Let $T : U \to U$ be linear, suppose that $\dim U = n$ and let $T$ have matrix $A$ w.r.t. some basis of $U$. Then

$$\lambda \text{ is an eigenvalue of } T \iff \ker(T - \lambda I) \neq \{0\}$$
$$\iff T - \lambda I \text{ is not invertible}$$
$$\iff A - \lambda I_n \text{ is singular}$$
$$\iff \det(A - \lambda I_n) = 0.$$ 

Here $I_n$ denotes the $n \times n$ identity matrix.

Notice that $\det(A - \lambda I_n)$ is a polynomial in $\lambda$ of degree $n$. It is called the characteristic polynomial of $T$; its roots (in $F$) are precisely the eigenvalues of $T$. It is easy to check, using the change of bases formula for the matrix of a linear mapping (Theorem (2.3)), that the characteristic polynomial depends on $T$ but not the particular matrix $A$ representing $T$.

The underlying scalar field is important here. If $F = \mathbb{C}$, then $T$ will always have an eigenvalue, since every complex polynomial has a root in $\mathbb{C}$ (the so-called Fundamental Theorem of Algebra). However, when $F = \mathbb{R}$, the characteristic polynomial may have no real roots if $n$ is even, though it will always have at least one real root if $n$ is odd. Thus, a linear mapping on an even dimensional real vector space might not have any eigenvalues, though such a mapping on an odd dimensional real vector space will always have an eigenvalue.

Comment

Starting with an $n \times n$ matrix $A$ with entries in $F$, the polynomial $\det(A - \lambda I_n)$ is also referred to as the characteristic polynomial of $A$. Its roots are the eigenvalues of the associated linear mapping $T_A : x \to Ax$ on $F^n$.

Lecture 13

3. Inner product spaces

We now extend the idea of an inner (or scalar) product in $\mathbb{R}^2$ and $\mathbb{R}^3$ to more general vector spaces. In this context, it is necessary to distinguish between real and complex vector spaces.

**Definition** Let $V$ be a vector space over $F = \mathbb{R}$ or $\mathbb{C}$. An inner product on $V$ is a map that associates with each pair of vectors $x, y \in V$ a scalar $\langle x, y \rangle \in F$ (more formally, $x, y \to \langle x, y \rangle$ is a function from $V \times V$ to $F$) such that, for all $x, y$ and $z \in V$ and all $\lambda, \mu \in F$,

(i) $\langle y, x \rangle = \begin{cases} 
\langle x, y \rangle & \text{when } F = \mathbb{R} \\
\overline{\langle x, y \rangle} & \text{when } F = \mathbb{C};
\end{cases}$

(ii) $\langle \lambda x + \mu y, z \rangle = \lambda \langle x, z \rangle + \mu \langle y, z \rangle$;

(iii) $\langle x, x \rangle \geq 0$, with equality if and only if $x = 0$.

A real (resp. complex) inner product space is a vector space over $\mathbb{R}$ (resp. over $\mathbb{C}$) endowed with an inner product.

Comments
1. It follows from (i) that, even when $F = C$, $\langle x, x \rangle \in R$ for all $x \in V$; thus (iii) both makes sense and strengthens this. Further, it is easy to see that (ii) implies that $\langle 0,0 \rangle = 0$, so that, for (iii), we need $\langle x, x \rangle \geq 0$ and $\langle x, x \rangle = 0 \Rightarrow x = 0$.

2. For a real inner product space $V$, properties (i) and (ii) together imply that

$$\langle x, \lambda y + \mu z \rangle = \lambda \langle x, y \rangle + \mu \langle x, z \rangle$$

for all $x, y, z \in V$ and all $\lambda, \mu \in R$, and so $\langle x, y \rangle$ is linear in the second variable $y$ as well as in the first variable $x$. However, for a complex inner product space $V$,

$$\langle x, \lambda y + \mu z \rangle = \lambda \bar{x} \langle y, x \rangle + \mu \langle x, z \rangle$$

for all $x, y, z \in V$ and all $\lambda, \mu \in C$. In this case, we say that $\langle x, y \rangle$ is conjugate linear in the second variable $y$ (it is linear in the first variable $x$ by (ii)).

Examples

1. In $\mathbb{R}^n$, define

$$\langle x, y \rangle = x_1 y_1 + x_2 y_2 + \cdots + x_n y_n$$

for $x = (x_1, \ldots, x_n)$ and $y = (y_1, \ldots, y_n)$. Then $\langle \cdot, \cdot \rangle$ is an inner product, referred to as the standard inner product on $\mathbb{R}^n$.

2. On $\mathbb{R}^2$, define $\langle x, y \rangle = x_1 y_1 + 4x_2 y_2$. Then $\langle \cdot, \cdot \rangle$ is an inner product on $\mathbb{R}^2$.

3. In $\mathbb{C}^n$, define

$$\langle z, w \rangle = z_1 \bar{w}_1 + z_2 \bar{w}_2 + \cdots + z_n \bar{w}_n$$

for $z = (z_1, \ldots, z_n)$ and $w = (w_1, \ldots, w_n)$. Then $\langle \cdot, \cdot \rangle$ is an inner product, referred to as the standard inner product on $\mathbb{C}^n$.

4. $\langle z, w \rangle = z_1 \bar{w}_1 + 4z_2 \bar{w}_2$ defines an inner product on $\mathbb{C}^2$.

5. On $\mathbb{R}^2$, define $\langle x, y \rangle = 2x_1 y_1 + x_1 y_2 + x_2 y_1 + x_2 y_2$. Properties (i) and (ii) for an inner product are easily seen to be satisfied. Also

$$\langle x, x \rangle = 2x_1^2 + 2x_1 x_2 + x_2^2$$

$$= 2(x_1 + x_2/2)^2 + x_2^2/2$$

$$\geq 0,$$

with equality in the last line if and only if $x_1 = x_2 = 0$. Thus $\langle \cdot, \cdot \rangle$ is inner product on $\mathbb{R}^2$.

6. Now define $\langle x, y \rangle = x_1 y_1 + 2x_1 y_2 + 2x_2 y_1 + x_2 y_2$ on $\mathbb{R}^2$. Again it is easy to see that properties (i) and (ii) of an inner product are satisfied. However,

$$\langle x, x \rangle = x_1^2 + 4x_1 x_2 + x_2^2$$

$$= (x_1 + 2x_2)^2 - 3x_2^2$$

$$< 0,$$

provided $x_2 \neq 0$ and $x_1 = -2x_2$. Thus $\langle \cdot, \cdot \rangle$ is not an inner product on $\mathbb{R}^2$ in this case.
7. Fix $a < b$ and, for $p, q \in P_n$, define

$$\langle p, q \rangle = \int_a^b p(t)q(t) \, dt.$$ 

Then $\langle \cdot, \cdot \rangle$ is an inner product on $P_n$. Properties (i) and (ii) follow from the linearity of the integral, whilst (iii) follows from the fact that, if $\varphi : [a, b] \to \mathbb{R}$ is continuous, $\varphi(t) \geq 0$ on $[a, b]$ and $\int_a^b \varphi(t) \, dt = 0$, then $\varphi \equiv 0$ on $[a, b]$. Thus, with this inner product, $P_n$ is a real inner product space.

8. Let $V = C_C[a, b]$ be the vector space of all continuous complex valued functions defined on $[a, b]$, where $a < b$. (Continuity for complex valued functions is defined as the natural extension of the definition for real valued functions, using the modulus in $\mathbb{C}$ to measure ‘nearness’). For $f, g \in V$, define

$$\langle f, g \rangle = \int_a^b f(t)\overline{g(t)} \, dt.$$ 

(For a complex valued function $\varphi$ on $[a, b]$, the integral $\int_a^b \varphi(t) \, dt$ is defined to be $\int_a^b \text{Re}\varphi(t) \, dt + i \int_a^b \text{Im}\varphi(t) \, dt$.)

It is straightforward to check that $\langle \cdot, \cdot \rangle$ satisfies (i) and (ii). Also, $\langle f, f \rangle = \int_a^b |f(t)|^2 \, dt \geq 0$, with equality if and only if $f(t) \equiv 0$ on $[a, b]$ (using the continuity of the non-negative function $|f(t)|^2$). Hence $\langle \cdot, \cdot \rangle$ also satisfies (iii) and so defines an inner product on the complex vector space $C_C[a, b]$

Lecture 14

3. Inner product spaces contd.

Norms

Suppose that $V, \langle \cdot, \cdot \rangle$ is a real or complex inner product space. Then $\langle x, x \rangle \geq 0$ for all $x \in V$.

**Definition** The norm of $x \in V$, $\|x\|$, is defined as

$$\|x\| = \langle x, x \rangle^{1/2},$$

where $^{1/2}$ denotes the non-negative square root.

**(3.1) Theorem (The Cauchy-Schwarz inequality)** We have

$$|\langle x, y \rangle| \leq \|x\| \|y\|$$

for all $x, y \in V$.

Applications of the Cauchy-Schwarz inequality

1. Let $a_1, \ldots, a_n, b_1, \ldots, b_n \in \mathbb{R}$. Then

$$|a_1b_1 + \cdots + a_nb_n| \leq (a_1^2 + \cdots + a_n^2)^{1/2}(b_1^2 + \cdots + b_n^2)^{1/2}.$$
2. Let \( p, q \) be two real polynomials (thought of as in \( P_n \) for some \( n \)) and let \( a < b \). Then, applying the Cauchy-Schwarz inequality with the inner product \( \langle p, q \rangle = \int_a^b p(t)q(t) \, dt \), we get
\[
\left| \int_a^b p(t)q(t) \, dt \right| \leq \left\{ \int_a^b |p(t)|^2 \, dt \right\}^{1/2} \left\{ \int_a^b |q(t)|^2 \, dt \right\}^{1/2}.
\]

(3.2) Theorem The norm \( \| \cdot \| \) in an inner product space \( V \) satisfies

(i) \( \| x \| \geq 0 \) with equality if and only if \( x = 0 \);
(ii) \( \| \lambda x \| = |\lambda| \| x \| \);
(iii) \( \| x + y \| \leq \| x \| + \| y \| \) [the triangle inequality]

for all \( x, y, z \in V \) and all \( \lambda \in \mathbb{F} \).

Orthogonal and orthonormal sets

Definitions

(i) Two vectors \( x, y \) in an inner product space are **orthogonal** if \( \langle x, y \rangle = 0 \). We write this as \( x \perp y \).

(ii) A set of vectors \( \{ u_1, \ldots, u_k \} \) in an inner product space is

(a) **orthogonal** if \( \langle u_i, u_j \rangle = 0 \) when \( i \neq j \);

(b) **orthonormal** if \( \langle u_i, u_j \rangle = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \).

Comments

1. If \( \{ u_1, \ldots, u_k \} \) is orthonormal, then \( \| u_j \| = 1 \) and so \( u_j \neq 0 \) for all \( j \). If \( \{ u_1, \ldots, u_k \} \) is orthogonal and \( u_j \neq 0 \) for all \( j \), then the set \( \{ \frac{u_j}{\| u_j \|} : j = 1, \ldots, k \} \) is orthonormal. We say that \( \{ u_1, \ldots, u_k \} \) has been **normalized**.

2. The term ‘orthonormal’ is often abbreviated to ‘o.n.’.

(3.3) Theorem An orthogonal set \( \{ u_1, \ldots, u_k \} \) of non-zero vectors is linearly independent. In particular, an o.n. set \( \{ u_1, \ldots, u_k \} \) is linearly independent.

Suppose now that \( V \) is an inner product space and \( \dim V = n \). Let \( \{ u_1, \ldots, u_n \} \) be an orthogonal set of non-zero vectors in \( V \). Then \( \{ u_1, \ldots, u_n \} \) is a basis of \( V \) and so each vector \( x \in V \) can be expressed uniquely as \( x = \lambda_1 u_1 + \cdots + \lambda_n u_n \). In this situation, it easy to compute the coordinates \( \lambda_j \) of \( x \in V \).

(3.4) Theorem Let \( V \) be an \( n \)-dimensional inner product space.

(i) Let \( \{ u_1, \ldots, u_n \} \) be an orthogonal set of non-zero vectors in \( V \), and hence a basis of \( V \). Then
\[
x = \sum_{j=1}^{n} \frac{\langle x, u_j \rangle}{\| u_j \|^2} u_j
\]

for all \( x \in V \).
Let \( \{e_1, \ldots, e_n\} \) be an orthonormal set of vectors in \( V \), and hence an o.n. basis of \( V \). Then
\[
x = \sum_{j=1}^{n} \langle x, e_j \rangle e_j
\]
and
\[
\|x\|^2 = \sum_{j=1}^{n} |\langle x, e_j \rangle|^2 \quad \text{(Parseval’s equality)}
\]
for all \( x \in V \). Thus, if \( x = \sum_{j=1}^{n} \lambda_j e_j \), then
\[
\|x\|^2 = \sum_{j=1}^{n} |\lambda_j|^2.
\]

Lecture 15

3. Inner product spaces contd.

The Gram-Schmidt process

(3.5) Theorem Let \( V \) be a finite dimensional inner product space. Then \( V \) has an orthogonal, and hence an o.n basis.

The proof of this result gives a procedure, called the Gram-Schmidt process, for constructing from a given basis of \( V \) an orthogonal basis. Normalization will then yield an o.n basis.

Example

Let \( V \) be the subspace of \( \mathbb{R}^4 \) with basis \( \{v_1, v_2, v_3\} \), where
\[
v_1 = (1, 1, 0, 0), \quad v_2 = (1, 0, 1, 1), \quad v_3 = (1, 0, 0, 1).
\]
(It is easy to check that these three vectors are linearly independent; for instance, \( v_2 \) and \( v_3 \) are linearly independent since they are not multiples of each other and, by considering the second coordinate, \( v_1 \notin \text{span}\{v_2, v_3\} \).) An orthogonal basis \( \{u_1, u_2, u_3\} \) of \( V \) can be constructed from \( \{v_1, v_2, v_3\} \) as follows.

Let \( u_1 = v_1 = (1, 1, 0, 0) \). Now put
\[
u_2 = v_2 - \lambda u_1 = (1, 0, 1, 1) - \lambda(1, 1, 0, 0) = (1 - \lambda, -\lambda, 1, 1)
\]
and choose \( \lambda \) so that \( u_2 \perp u_1 \). For this, we need
\[
\langle u_2, u_1 \rangle = 1 - \lambda - \lambda = 0 \quad \text{or} \quad \lambda = 1/2,
\]
giving \( u_2 = (\frac{1}{2}, -\frac{1}{2}, 1, 1) \). Now put
\[
u_3 = v_3 - \lambda_1 u_1 - \lambda_2 u_2 = (1, 0, 0, 1) - \lambda_1(1, 1, 0, 0) - \lambda_2(\frac{1}{2}, -\frac{1}{2}, 1, 1)
\]
and choose \( \lambda_1, \lambda_2 \) so that \( u_3 \perp u_1, u_2 \). For \( u_3 \perp u_1 \), we need
\[
\langle u_3, u_1 \rangle = \langle (1, 0, 0, 1), (1, 1, 0, 0) \rangle - \lambda_1 \langle (1, 1, 0, 0), (1, 1, 0, 0) \rangle = 0
\]
i.e. $1 - 2\lambda_1 = 0$ or $\lambda_1 = 1/2$. (Note that we have used the fact that $u_1 \perp u_2$ when calculating $\langle u_3, u_1 \rangle$ here; the effect is that we get an equation in $\lambda_1$ alone.)

For $u_3 \perp u_2$, we need

$$\langle u_3, u_2 \rangle = \langle (1, 0, 0, 1), (1/2, -1/2, 1, 1) \rangle - \lambda_2\langle (1/2, -1/2, 1, 1), (1/2, -1/2, 1, 1) \rangle = 0,$$

giving

$$\frac{1}{2} + 1 - \lambda_2\left(\frac{1}{4} + \frac{1}{4} + 1 + 1\right) = 0$$

i.e. $\lambda_2 = \frac{3/2}{5/2} = 3/5$. Hence

$$u_3 = (1, 0, 0, 1) - \frac{1}{2}(1, 1, 0, 0) - \frac{3}{5}\left(\frac{1}{2}, -\frac{1}{2}, 1, 1\right)$$

$$= \left(\frac{1}{5}, -\frac{1}{5}, -\frac{3}{5}, \frac{2}{5}\right)$$

$$= \frac{1}{5}(1, -1, -3, 2).$$

We have thus constructed an orthogonal basis $\{u_1, u_2, u_3\}$ of $V$ given by

$$u_1 = (1, 1, 0, 0), \quad u_2 = u_2 = \left(\frac{1}{2}, -\frac{1}{2}, 1, 1\right) = \frac{1}{2}(1, -1, 2, 2), \quad u_3 = \frac{1}{5}(1, -1, -3, 2).$$

Stripping out the fractions, we get

$$(1, 1, 0, 0), \quad (1, -1, 2, 2), \quad (1, -1, -3, 2).$$

It useful to check that these three vectors are indeed orthogonal in pairs. Note also that, to avoid quite so many fractions, we could have replaced $u_2 = \left(\frac{1}{2}, -\frac{1}{2}, 1, 1\right)$ by $(1, -1, 2, 2)$ earlier when calculating $u_3$.

To obtain an o.n. basis $\{e_1, e_2, e_3\}$, we normalize to get

$$e_1 = \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{2}}(1, 1, 0, 0), \quad e_2 = \frac{u_2}{\|u_2\|} = \frac{1}{\sqrt{10}}(1, -1, 2, 2), \quad e_3 = \frac{u_3}{\|u_3\|} = \frac{1}{\sqrt{15}}(1, -1, -3, 2).$$

Consider $v = v_1 + v_2 - v_3 = (1, 1, 1, 0) \in V$. We can write $v$ as $\lambda_1 e_1 + \lambda_2 e_2 + \lambda_3 e_3$, the coefficients here being given by

$$\lambda_1 = \langle v, e_1 \rangle = \frac{2}{\sqrt{2}} = \sqrt{2}, \quad \lambda_2 = \langle v, e_2 \rangle = \frac{2}{\sqrt{10}} = \frac{\sqrt{10}}{5}, \quad \lambda_3 = \langle v, e_3 \rangle = -\frac{3}{\sqrt{15}} = -\frac{\sqrt{15}}{5}.$$ 

Notice that

$$\lambda_1^2 + \lambda_2^2 + \lambda_3^2 = 2 + \frac{10}{25} + \frac{15}{25} = 3 = \|v\|^2,$$

as expected by Parseval’s equality.

Another example

Find a basis of $P_2$ that is orthogonal w.r.t. the inner product given by $\langle p, q \rangle = \int_0^1 p(t)q(t) \, dt$. To do this, start with the basis $\{v_1, v_2, v_3\}$, where $v_1(t) = 1$, $v_2(t) = t$, $v_3(t) = t^2$, and apply the Gram-Schmidt process.
Put \( u_1 = v_1 \); then \( u_2(t) = v_2(t) - \lambda u_1(t) = t - \lambda \). For \( u_2 \perp u_1 \), we need
\[
\int_0^1 (t - \lambda) \, dt = 0, \quad \text{i.e.} \quad \lambda = 1/2.
\]
This gives \( u_2(t) = t - 1/2 \). Now put \( u_3 = v_3 - \lambda_1 u_1 - \lambda_2 u_2 \). For \( u_3 \perp u_1 \), we need
\[
\int_0^1 (v_3(t) - \lambda_1 u_1(t)) u_1(t) \, dt = 0, \quad \text{i.e.} \quad \int_0^1 t^2 \, dt - \lambda_1 \int_0^1 1 \, dt = 0.
\]
So \( \lambda_1 = 1/3 \).

For \( u_3 \perp u_2 \), we need
\[
\int_0^1 (v_3(t) - \lambda_2 u_2(t)) u_2(t) \, dt = 0, \quad \text{i.e.} \quad \int_0^1 t^2(t - 1/2) \, dt - \lambda_2 \int_0^1 (t - 1/2)^2 \, dt = 0.
\]
This leads to
\[
\lambda_2 = \frac{\int_0^1 t^2(t - 1/2) \, dt}{\int_0^1 (t - 1/2)^2 \, dt} = \frac{1/12}{1/12} = 1.
\]
Hence \( u_3(t) = t^2 - 1/3 - (t - 1/2) = t^2 - t + 1/6 \) and we obtain the orthogonal basis
\[
1, \ t - 1/2, \ t^2 - t + 1/6.
\]

Lecture 16

3. Inner product spaces contd.

Orthogonal complements and orthogonal projections

Let \( U \) be a subspace of an inner product space \( V \).

Definition The orthogonal complement of \( U \) is the set \( U^\perp \) defined by
\[
U^\perp = \{ x \in V : \langle x, u \rangle = 0 \text{ for all } u \in U \}.
\]

(3.6) Theorem Let \( U \) be a subspace of a finite dimensional inner product space \( V \). Then each \( x \in V \) can be written uniquely as
\[
x = y + z,
\]
where \( y \in U \) and \( z \in U^\perp \). Thus \( V = U \oplus U^\perp \).

Terminology Let \( V \) and \( U \) be as in Theorem (3.6). Given \( x \in V \), define the orthogonal projection of \( x \) onto \( U \) to be the unique \( y \in U \) with \( x = y + z \), where \( z \in U^\perp \), and write \( y = P_U x \). We can then think of \( P_U \) as a mapping \( V \rightarrow V \). This mapping is called the orthogonal projection of \( V \) onto \( U \). The following are its main properties.

(3.7) Theorem With \( V, U \) and \( P_U \) as above, the following hold.

(i) \( P_U \) is a linear mapping with \( \text{im} \, P_U = U \) and \( \ker P_U = U^\perp \).
(ii) If \( \{e_1, \ldots, e_k\} \) is an o.n. basis of \( U \), then \( P_U \) is given by the formula

\[
P_U x = \sum_{j=1}^{k} \langle x, e_j \rangle e_j \quad (x \in V).
\]

If \( \{u_1, \ldots, u_k\} \) is an orthogonal basis of \( U \), then \( P_U \) is given by the formula

\[
P_U x = \sum_{j=1}^{k} \frac{1}{\|u_j\|^2} \langle x, u_j \rangle u_j \quad (x \in V).
\]

(iii) \( P_U^2 = P_U \).

(iv) \( x \in U \) if and only if \( x = P_U x \).

(v) For \( x \in V \),

\[
\|x\|^2 = \|P_U x\|^2 + \|x - P_U x\|^2.
\]

[Strictly, we should consider the case when \( U = \{0\} \) separately. Then \( U^\perp = V \) and \( P_U = 0 \).]

Nearest point to a subspace.

In an inner product space, the norm is used to measure distance, with \( \|x - y\| \) being thought of as the distance between \( x \) and \( y \).

(3.8) Theorem Let \( U \) be a subspace of a finite dimensional inner product space \( V \) and let \( x \in V \). Then

\[
\|x - u\| \geq \|x - P_U x\|
\]

for all \( u \in U \), with equality if and only if \( u = P_U x \). Thus, with distance measured using the norm, \( P_U x \) is the unique nearest point of \( U \) to \( x \).

Connection with Fourier series

Let \( V \) denote the space of \( 2\pi \)-periodic continuous functions \( f : \mathbb{R} \to \mathbb{C} \) and give \( V \) the inner product defined by

\[
\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \overline{g(t)} \, dt.
\]

For \( n \in \mathbb{Z} \), let \( e_n \in V \) be the function \( e_n(t) = e^{int} \). Then \( \{e_n : n \in \mathbb{Z}\} \) is an orthonormal set and, for \( f \in V \), \( \langle f, e_n \rangle \) is the \( n^{th} \) (complex) Fourier coefficient

\[
c_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t)e^{-int} \, dt
\]

of \( f \). Thus the Fourier series of \( f \) can be written as

\[
\sum_{n=-\infty}^{\infty} \langle f, e_n \rangle e_n.
\]

This suggests that, in some sense, the Fourier series of \( f \) is its expansion in terms of the o.n. ‘basis’ \( \{e_n : n \in \mathbb{Z}\} \). This is, indeed, a reasonable interpretation but, to make it more rigorous, one has to decide how to make sense of the convergence of the Fourier series of \( f \in V \) and also
whether \( V \) is the right space to consider these matters. What is true is that, for \( f \in V \) with Fourier series
\[
\sum_{n=-\infty}^{\infty} c_n e_n,
\]
\[
\|f - \sum_{n=-N}^{N} \langle f, e_n \rangle e_n \| = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t) - \sum_{n=-N}^{N} c_n e_n(t)|^2 dt \right\}^{1/2} \to 0
\]
as \( N \to \infty \). Further, fixing \( N \in \mathbb{N} \), amongst all the functions \( g \in \text{span}\{e_n : n = -N, \ldots, N\} \), the \( N \)th partial sum \( \sum_{n=-N}^{N} c_n e_n \) of the Fourier series of \( f \) is the unique function \( g \) that minimizes
\[
\|f - g\| = \left\{ \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(t) - g(t)|^2 dt \right\}^{1/2}.
\]
In fact, \( V \) is not quite the correct space to discuss these issues — this will be discussed in more detail in later years.

Lecture 17

4. Adjoints and self-adjointness

Assume throughout this chapter that \( V \) is a finite dimensional inner product space. The aim is to show that, given a linear mapping \( T : V \to V \), there is a (unique) linear mapping \( T^* : V \to V \) such that \( \langle T^* x, y \rangle = \langle x, Ty \rangle \) for all \( x, y \in V \) and then to study those mappings for which \( T = T^* \).

(4.1) Theorem  Given a linear mapping \( T : V \to V \), there is a unique linear mapping \( T^* : V \to V \) such that
\[
\langle T^* x, y \rangle = \langle x, Ty \rangle
\]
for all \( x, y \in V \). Furthermore, if \( \{e_1, \ldots, e_n\} \) is an o.n. basis of \( V \) and \( T \) has matrix \( A = (a_{ij}) \) with respect to \( \{e_1, \ldots, e_n\} \), then \( T^* \) has matrix \( C = (c_{ij}) \) with respect to \( \{e_1, \ldots, e_n\} \), where
\[
c_{ij} = \begin{cases} a_{ji} & \text{when } \mathbb{F} = \mathbb{R} \\ \overline{a_{ji}} & \text{when } \mathbb{F} = \mathbb{C} \end{cases}
\]

Notation and terminology

1. The mapping \( T^* \) is called the adjoint of \( T \).

2. Given a matrix \( A = (a_{ij}) \), its transpose is the matrix \( A^t = (a_{ji}) \) and, when the entries are complex, its conjugate transpose is \( A^* = (\overline{a_{ji}}) = (\overline{A})^t \). Thus, for a given linear mapping \( T : V \to V \) with matrix \( A \) with respect to some o.n. basis of \( V \),

(a) if \( V \) is a real inner product space, \( T^* \) has matrix \( A^t \),

(b) if \( V \) is a complex inner product space, \( T^* \) has matrix \( A^* \)

with respect to the same o.n. basis.

Definition A linear mapping \( T : V \to V \) is said to be self-adjoint if \( T^* = T \), i.e. if \( \langle Tx, y \rangle = \langle x, Ty \rangle \) for all \( x, y \in V \).

Thus \( T \) is self-adjoint if and only if its matrix \( A \) w.r.t. any (and hence every) o.n. basis satisfies
(a) \( A^t = A \) if \( V \) is a real inner product space;
(b) \( A^* = A \) if \( V \) is a complex inner product space.

Definitions A matrix \( A \) for which \( A^t = A \) is called symmetric whilst (in the case of complex entries) it is called hermitian when \( A^* = A \).

Although a linear mapping on a complex (finite dimensional) vector space always has an eigenvalue, it is possible for a linear mapping on a real vector space not to have any eigenvalues. However, this phenomenon cannot occur for a self-adjoint mapping. Indeed, more can be said about eigenvalues and eigenvectors of self-adjoint linear mappings.

(4.2) Theorem Let \( T : V \to V \) be self-adjoint.

(i) Every eigenvalue of \( T \) is real (even when \( V \) is a complex inner product space).
(ii) Let \( \lambda, \mu \) be distinct eigenvalues of \( T \) with corresponding eigenvectors \( u, v \). Then \( \langle u, v \rangle = 0 \).
(iii) \( T \) has an eigenvalue (even when \( V \) is a real inner product space).

Lecture 18


Our aim now is to show that, given a self-adjoint \( T \) on \( V \), \( V \) has an o.n. basis consisting of eigenvectors of \( T \) (and so, in particular, \( T \) is diagonalizable). The following result will be useful in doing this, as well as being of interest in its own right.

(4.3) Theorem Let \( T : V \to V \) be self-adjoint and let \( U \) be a subspace of \( V \) such that \( T(U) \subseteq U \). Then \( T(U^\perp) \subseteq U^\perp \) and the restrictions of \( T \) to \( U \) and \( U^\perp \) are self-adjoint.

We now have the main result of this chapter.

(4.4) Theorem (The finite dimensional spectral theorem) Let \( V \) be a finite dimensional inner product space and let \( T : V \to V \) be a self-adjoint linear mapping. Then \( V \) has an o.n. basis consisting of eigenvectors of \( T \).

An illustration of the use of the spectral theorem

Suppose that \( T : V \to V \) is self-adjoint and that every eigenvalue \( \lambda \) of \( T \) satisfies \( \lambda \geq 1 \). Then

\[
\langle Tx, x \rangle \geq \|x\|^2 \quad \text{for all} \ x \in V.
\]

Proof Let \( \{e_1, \ldots, e_n\} \) be an o.n. basis of \( V \) with \( Te_j = \lambda_j e_j \) for \( j = 1, \ldots, n \), where \( \lambda_j \geq 1 \) for all \( j \). Let \( x \in V \). We have

\[
x = x_1 e_1 + \cdots + x_n e_n,
\]

where \( x_j = \langle x, e_j \rangle \), and

\[
Tx = x_1 \lambda_1 e_1 + \cdots + x_n \lambda_n e_n.
\]
Then
\[ \langle Tx, x \rangle = \left( \sum_{j=1}^{n} x_j \lambda_j e_j, \sum_{k=1}^{n} x_k e_k \right) \]
\[ = \sum_{j=1}^{n} \lambda_j |x_j|^2 \quad \text{since } \langle e_j, e_k \rangle = \delta_{ij} \]
\[ \geq \sum_{j=1}^{n} |x_j|^2 \quad \text{since each } \lambda_j \geq 1 \]
\[ = \|x\|^2 \quad \text{by Parseval’s equality.} \]

Here $\delta_{ij}$ is the Kronecker delta symbol, defined as
\[ \delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j. \end{cases} \]

Lecture 19

4 Adjoints and self-adjointness contd.

Real symmetric matrices

Let $A$ be a real $n \times n$ matrix. This has an associated mapping $T_A : \mathbb{R}^n \to \mathbb{R}^n$ given by
\[ T_A : x \to Ax, \]
where we are thinking of $x$ as a column vector. Then $A$ is the matrix of $T_A$ and $A^t$ is the matrix of $(T_A)^*$ w.r.t. the standard basis of $\mathbb{R}^n$.

Suppose now that $A$ is symmetric (i.e. $A^t = A$). Then $T_A$ is self-adjoint since in this case $(T_A)^* = T_{A^t} = T_A$.

Aside: This can also be seen as follows. Note that, writing vectors in $\mathbb{R}^n$ as columns and thinking of these as $n \times 1$ matrices, we can express the standard inner product of $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ and
\[ y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \] as $\langle x, y \rangle = x^t y = \left( x_1, \ldots, x_n \right) \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$. We then have
\[ \langle Ax, y \rangle = (Ax)^t y = (x^t A^t)y = (x^t A)y = x^t (Ay) = \langle x, Ay \rangle, \]
which shows that $T_A$ is self-adjoint.

Applying the spectral theorem to $T_A$, we conclude that there is an o.n. basis $\{u_1, \ldots, u_n\}$ of $\mathbb{R}^n$
consisting of eigenvectors of $T_A$, say $T_A u_j = \lambda_j u_j$ for $j = 1, \ldots, n$. Consider the $n \times n$ matrix $R = (u_1 \ldots u_n)$ (i.e. $u_j$ is the $j^{th}$ column of $R$). Then $Au_j$ is the $j^{th}$ column of $AR$ and so

$$AR = (Au_1 \ldots Au_n) = (\lambda_1 u_1 \ldots \lambda_n u_n)$$

$$= (u_1 \ldots u_n) \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}$$

$$= RD,$$

where $D$ is the diagonal matrix

$$\begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix}.$$

Now $R^t$ is the $n \times n$ matrix with $j^{th}$ row $u^t_j$ (where $u_j$ is a column and so $u^t_j$ is the vector $u_j$ written as a row). Thus

$$R^t R = \begin{pmatrix} u^t_1 \\ \vdots \\ u^t_n \end{pmatrix} (u_1 \ldots u_n),$$

so the $ij^{th}$ entry of $R^t R$ is $u^t_i u_j = \langle u_i, u_j \rangle = \delta_{ij}$. Thus

$$R^t R = I_n,$$

the $n \times n$ identity matrix.

Thus $R$ is invertible and $R^{-1} = R^t$. (It could have been noted earlier that $R$ is invertible since its columns are linearly independent. However, the above calculation shows that its inverse can be written down without any further work.)

**Definition** An orthogonal matrix is an $n \times n$ real matrix $R$ (for some $n$) such that $R^t R = I_n$ [equivalently, $R$ is invertible and $R^{-1} = R^t$ or the columns (or rows) of $R$ form an o.n. basis of $\mathbb{R}^n$].

Returning to the above equation $AR = RD$, premultiplication by $R^t$ gives $R^t AR = D$. It is easy to show (by considering the characteristic polynomial of $A$, which equals the characteristic polynomial of $R^{-1} AR = R^t AR = D$) that the diagonal elements of $D$ are precisely the eigenvalues of $A$, repeated according to their multiplicity as roots of the characteristic polynomial.

### (4.5) Theorem (The spectral theorem for real symmetric matrices)

**Let $A$ be a real $n \times n$ symmetric matrix. Then there is an $n \times n$ orthogonal matrix $R$ such that**

$$R^t AR = R^{-1} AR = \begin{pmatrix} \lambda_1 & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \lambda_n \end{pmatrix},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the (necessarily real) eigenvalues of $A$, repeated according to their multiplicity as roots of the characteristic polynomial of $A$.

**Example**
Let
\[ A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}. \]

Find an orthogonal matrix \( R \) such that \( R^tAR \) is diagonal.

There are three steps to do this.

1. **Find the eigenvalues of** \( A \). We have (after some computation)
   \[
   \det(A - \lambda I_3) = (1 - \lambda)^2(4 - \lambda).
   \]
   Thus the eigenvalues, which are the roots of \( \det(A - \lambda I_3) = 0 \), are 1 and 4. (As a check, their sum counting multiplicities is \( 1 + 1 + 4 = 6 \) and this equals the trace of \( A \), the sum of the diagonal elements of \( A \). This will always be the case.)

2. **Find o.n. bases of the corresponding eigenspaces.**

   Eigenspace corresponding to \( \lambda = 1 \):
   We need to solve \( (A - 1.I_3)x = 0 \), i.e. \( x_1 + x_2 + x_3 = 0 \), leading to
   \[
   x = (x_1, x_2, -x_1 - x_2) = x_1(1, 0, -1) + x_2(0, 1, -1).
   \]
   Thus \( \{v_1, v_2\} \) is a basis of this eigenspace, where \( v_1 = (1, 0, -1) \) and \( v_2 = (0, 1, -1) \). An application of the Gram-Schmidt process to \( \{v_1, v_2\} \) gives the orthogonal basis
   \( \{(1, 0, -1), (-1/2, 1, -1/2)\} \) or (stripping out the fractions) \( \{(1, 0, -1), (-1/2, 1, -1/2)\} \). Normalizing, we get the o.n. basis \( \{e_1, e_2\} \) where
   \[
   e_1 = \frac{1}{\sqrt{2}}(1, 0, -1); \quad e_2 = \frac{1}{\sqrt{6}}(-1, 2, -1).
   \]

   Eigenspace corresponding to \( \lambda = 4 \):
   For this we need to solve \( (A - 4I_3)x = 0 \), leading to the three equations
   \[
   \begin{align*}
   -2x_1 + x_2 + x_3 &= 0 \\
   x_1 - 2x_2 + x_3 &= 0 \\
   x_1 + x_2 - 2x_3 &= 0.
   \end{align*}
   \]
   The third of these equations is just minus the sum of the first two and so is redundant. Eliminating \( x_3 \) from the first two gives \( x_1 = x_2 \) and then \( x_3 = x_1 \) from the first, leading to the general solution \( x = x_1(1, 1, 1) \). Thus an o.n. basis for this eigenspace consists of the single vector
   \[
   e_3 = \frac{1}{\sqrt{3}}(1, 1, 1).
   \]

3. **Obtain the required orthogonal matrix.** Now form the matrix \( R = (e_1 \, e_2 \, e_3) \), with the \( e_j \)'s written as column vectors; i.e. let
   \[
   R = \begin{pmatrix}
   \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
   0 & \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\
   -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}}
   \end{pmatrix}.
   \]
   This matrix is orthogonal and
   \[
   R^tAR = A = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 4 \end{pmatrix}.
   \]
4 Adjoint and self-adjointness contd.

Hermitian matrices

Suppose that $A = (a_{ij}) \in M_n(\mathbb{C})$ is hermitian, i.e. $A^* = A$ or $a_{ji} = \overline{a_{ij}}$ for all $i, j$. In this case, the associated mapping $T_A : \mathbb{C}^n \to \mathbb{C}^n$ given by

$$T_A : z \mapsto Az,$$

where, as in the discussion of real symmetric matrices, we are thinking of $z$ as a column vector, is self-adjoint. We have already seen this, but a direct proof is as follows.

Write a typical vector of $\mathbb{C}^n$ as a column vector $z = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$, thought of as an $n \times 1$ matrix.

Then $z^* = (\overline{z}_1 \ldots \overline{z}_n)$ and hence

$$\langle w, z \rangle = w_1 \overline{z}_1 + \cdots + w_n \overline{z}_n = z^* w.$$  

We then have

$$\langle Aw, z \rangle = z^*(Aw) = (z^*A)w = (z^*A^*)w = (Az)^*w = \langle w, Az \rangle,$$

showing that $T_A$ is self-adjoint.

Again, we can apply the spectral theorem to $T_A$ and conclude that there is an o.n. basis $\{u_1, \ldots, u_n\}$ of $\mathbb{C}^n$ consisting of eigenvectors of $T_A$, say $T_Au_j = \lambda_j u_j$ for $j = 1, \ldots, n$, where each $\lambda_j \in \mathbb{R}$ since the eigenvalues of self-adjoint mappings are always real. Guided by the discussion of real symmetric matrices, consider the $n \times n$ matrix $U = (u_1 \ldots u_n)$ (i.e. $u_j$ is the $j^{th}$ column of $U$). Then $Au_j$ is the $j^{th}$ column of $AU$ and so

$$AU = (Au_1 \ldots Au_n) = (\lambda_1 u_1 \ldots \lambda_n u_n)$$

$$= (u_1 \ldots u_n) \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

$$= UD,$$

where $D$ is the diagonal matrix

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$

Here we have

$$U^*U = \begin{pmatrix} u_1^* \\ \vdots \\ u_n^* \end{pmatrix}(u_1 \ldots u_n) = I_n,$$

as the $ij^{th}$ entry of $U^*U$ is $u_i^* u_j = \langle u_j, u_i \rangle = \delta_{ij}$. Thus

$$U^*U = I_n,$$

the $n \times n$ identity matrix.
and $U^{-1} = U^*$.

**Definition** A unitary matrix is a (complex) $n \times n$ real matrix $U$ (for some $n$) such that $U^* U = I_n$ [equivalently, $U$ is invertible and $U^{-1} = U^*$ or the columns (or rows) of $U$ form an o.n. basis of $\mathbb{C}^n$].

The spectral theorem then takes on the following form.

**(4.6) Theorem (The spectral theorem for hermitian matrices)** Let $A$ be an $n \times n$ (complex) hermitian matrix. Then there is an $n \times n$ unitary matrix $U$ such that

$$U^* A U = U^{-1} A U = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the (necessarily real) eigenvalues of $A$, repeated according to their multiplicity as roots of the characteristic polynomial of $A$.

Comments about orthogonal and unitary matrices

1. Let $R$ be an $n \times n$ orthogonal matrix, so that $R^T R = I_n$. Then

$$\langle Rx, Ry \rangle = (Rx)^T (Ry) = x^T R^T Ry = x^T y = \langle x, y \rangle$$

for all $x, y \in \mathbb{R}^n$. Thus the mapping $x \to Rx$ from $\mathbb{R}^n$ to $\mathbb{R}^n$ preserves the inner product. In particular, $\|Rx\| = \|x\|$ for all $x \in \mathbb{R}^n$ since

$$\|Rx\|^2 = \langle Rx, Rx \rangle = \langle x, x \rangle = \|x\|^2.$$

Conversely, suppose that $A$ is an $n \times n$ real matrix and $\langle Ax, Ay \rangle = \langle x, y \rangle$ for all $x, y \in \mathbb{R}^n$. Then $\langle A^T Ax, y \rangle = \langle Ax, Ay \rangle = \langle x, y \rangle$ for all $x, y$, from which it follows that $A^T A = I_n$, i.e. $A$ is orthogonal. In fact, it suffices to assume that $\|Ax\| = \|x\|$ for all $x \in \mathbb{R}^n$ to conclude that $A$ is orthogonal. To prove this, note the expression for an inner product in terms of the associated norm given in Tutorial Qn. 7.7.

2. In much the same way, if $U$ is a unitary $n \times n$ matrix, then

$$\langle Uw, Uz \rangle = \langle w, z \rangle$$

and, in particular,

$$\|Uw\| = \|w\|$$

for all $w, z \in \mathbb{C}^n$. Conversely, if $A$ is an $n \times n$ complex matrix such that $\langle Aw, Az \rangle = \langle w, z \rangle$ for all $w, z \in \mathbb{C}^n$ (or just $\|Aw\| = \|w\|$ for all $w \in \mathbb{C}^n$), then $A$ is unitary.