**MTH6140** **Linear Algebra II**

**Notes 1** **September 2010**

---

# 1 Vector spaces

This course is about linear maps and bilinear forms on vector spaces, how we represent them by matrices, how we manipulate them, and what we use this for.

## 1.1 Definitions

**Definition 1.1** A *field* is an algebraic system consisting of a non-empty set $K$ equipped with two binary operations $+$ (addition) and $\cdot$ (multiplication) satisfying the conditions:

(A) $(K,+)$ is an abelian group with identity element 0 (called *zero*);

(M) $(K\setminus\{0\},\cdot)$ is an abelian group with identity element 1;

(D) the *distributive law*

$$a(b+c)=ab+ac$$

holds for all $a,b,c\in K$.

If you don't know what an abelian group is, then

- you can find it spelled out in detail on the course information sheet *Fields and vector spaces* on the course web page;
- the only fields that we will use in this course are
  - $\mathbb{Q}$, the field of rational numbers;
  - $\mathbb{R}$, the field of real numbers;
  - $\mathbb{C}$, the field of complex numbers;
  - $\mathbb{F}_p$, the field of integers mod $p$, where $p$ is a prime number.

You will not be expected to prove that these structures are fields. You may have seen $\mathbb{F}_p$ referred to as $\mathbb{Z}_p$ in some courses.

**Definition 1.2** A *vector space* $V$ over a field $K$ is an algebraic system consisting of a non-empty set $V$ equipped with a binary operation $+$ (vector addition), and an operation of scalar multiplication

$$(a,v) \in K \times V \mapsto av \in V$$

such that the following rules hold:

(VA) $(V,+)$ is an abelian group, with identity element $0$ (the *zero vector*).

(VM) Rules for scalar multiplication:

- (VM0) For any $a \in K$, $v \in V$, there is a unique element $av \in V$.
- (VM1) For any $a \in K$, $u,v \in V$, we have $a(u+v) = au + av$.
- (VM2) For any $a,b \in K$, $v \in V$, we have $(a+b)v = av + bv$.
- (VM3) For any $a,b \in K$, $v \in V$, we have $(ab)v = a(bv)$.
- (VM4) For any $v \in V$, we have $1v = v$ (where 1 is the identity element of $K$).

Since we have two kinds of elements, namely elements of $K$ and elements of $V$, we distinguish them by calling the elements of $K$ *scalars* and the elements of $V$ *vectors*.

A vector space over the field $\mathbb{R}$ is often called a *real vector space*, and one over $\mathbb{C}$ is a *complex vector space*.

**Example 1.1** The first example of a vector space that we meet is the *Euclidean plane* $\mathbb{R}^2$. This is a real vector space. This means that we can add two vectors, and multiply a vector by a scalar (a real number). There are two ways we can make these definitions.

- The *geometric* definition. Think of a vector as an arrow starting at the origin and ending at a point of the plane. Then addition of two vectors is done by the *parallelogram law* (see Figure 1). The scalar multiple $av$ is the vector whose length is $|a|$ times the length of $v$, in the same direction if $a > 0$ and in the opposite direction if $a < 0$.

- The *algebraic* definition. We represent the points of the plane by Cartesian coordinates $(x,y)$. Thus, a vector $v$ is just a pair $(x,y)$ of real numbers. Now we define addition and scalar multiplication by

$$\begin{aligned} (x_1,y_1) + (x_2,y_2) &= (x_1 + x_2, y_1 + y_2), \\ a(x,y) &= (ax, ay). \end{aligned}$$

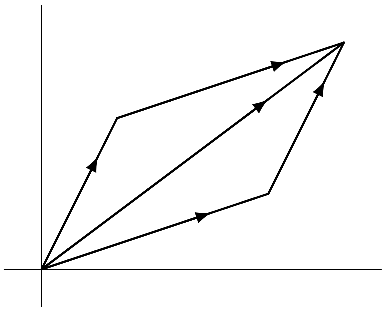

Figure 1: The parallelogram law

Not only is this definition much simpler, but it is much easier to check that the rules for a vector space are really satisfied! For example, we check the law $a(v+w) = av+aw$. Let $v = (x_1, y_1)$ and $w = (x_2, y_2)$. Then we have

$$\begin{aligned} a(v+w) &= a((x_1,y_1)+(x_2,y_2)) \\ &= a(x_1+x_2, y_1+y_2) \\ &= (ax_1+ax_2, ay_1+ay_2) \\ &= (ax_1,ay_1)+(ax_2,ay_2) \\ &= av+aw. \end{aligned}$$

In the algebraic definition, we say that the operations of addition and scalar multiplication are *coordinatewise*: this means that we add two vectors coordinate by coordinate, and similarly for scalar multiplication.

Using coordinates, this example can be generalised.

**Example 1.2** Let $n$ be any positive integer and $K$ any field. Let $V = K^n$, the set of all $n$-tuples of elements of $K$. Then $V$ is a vector space over $K$, where the operations are defined coordinatewise:

$$\begin{aligned} (a_1,a_2,\ldots,a_n)+(b_1,b_2,\ldots,b_n) &= (a_1+b_1,a_2+b_2,\ldots,a_n+b_n), \\ c(a_1,a_2,\ldots,a_n) &= (ca_1,ca_2,\ldots,ca_n). \end{aligned}$$

## 1.2 Bases

This example is much more general than it appears: *Every finite-dimensional vector space looks like Example 1.2*. Here's why.

**Definition 1.3** Let $V$ be a vector space over the field $K$, and let $v_1, \dots, v_n$ be vectors in $V$.

(a) The vectors $v_1, v_2, \dots, v_n$ are *linearly independent* if, whenever we have scalars $c_1, c_2, \dots, c_n$ satisfying

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = 0,$$

then necessarily $c_1 = c_2 = \cdots = 0$.

(b) The vectors $v_1, v_2, \dots, v_n$ are *spanning* if, for every vector $v \in V$, we can find scalars $c_1, c_2, \dots, c_n \in K$ such that

$$v = c_1v_1 + c_2v_2 + \cdots + c_nv_n.$$

In this case, we write $V = \langle v_1, v_2, \dots, v_n \rangle$.

(c) The vectors $v_1, v_2, \dots, v_n$ form a *basis* for $V$ if they are linearly independent and spanning.

**Remark** Linear independence is a property of a *list* of vectors. A list containing the zero vector is never linearly independent. Also, a list in which the same vector occurs more than once is never linearly independent.

I will say "Let $B = (v_1, \dots, v_n)$ be a basis for $V$" to mean that the list of vectors $v_1, \dots, v_n$ is a basis, and to refer to this list as $B$.

**Definition 1.4** Let $V$ be a vector space over the field $K$. We say that $V$ is *finite-dimensional* if we can find vectors $v_1, v_2, \dots, v_n \in V$ which form a basis for $V$.

**Remark** In this course we are only concerned with finite-dimensional vector spaces. If you study Functional Analysis, you will meet vector spaces which are not finite dimensional.

**Proposition 1.1** The following three conditions are equivalent for the vectors $v_1, \dots, v_n$ of the vector space $V$ over $K$:

(a) $v_1, \dots, v_n$ is a basis;

(b) $v_1, \dots, v_n$ is a maximal linearly independent set (that is, if we add any vector to the list, then the result is no longer linearly independent);

(c) $v_1, \dots, v_n$ is a minimal spanning set (that is, if we remove any vector from the list, then the result is no longer spanning).

The next theorem helps us to understand the properties of linear independence.

**Theorem 1.2 (The Exchange Lemma)** Let $V$ be a vector space over $K$. Suppose that the vectors $v_1, \dots, v_n$ are linearly independent, and that the vectors $w_1, \dots, w_m$ are linearly independent, where $m > n$. Then we can find a number $i$ with $1 \le i \le m$ such that the vectors $v_1, \dots, v_n, w_i$ are linearly independent.

**Proof** See the course information sheet *The Exchange Lemma* for a proof.

**Corollary 1.3** Let $V$ be a finite-dimensional vector space over a field $K$. Then

(a) any two bases of $V$ have the same number of elements;

(b) any linearly independent set can be extended to a basis.

The number of elements in a basis is called the *dimension* of the vector space $V$. We will say "an $n$-dimensional vector space" instead of "a finite-dimensional vector space whose dimension is $n$". We denote the dimension of $V$ by $\dim(V)$.

**Proof** Let us see how the corollary follows from the Exchange Lemma.

(a) Let $(v_1, \dots, v_n)$ and $(w_1, \dots, w_m)$ be two bases for $V$. Suppose, for a contradiction, that they have different numbers of elements; say that $n < m$, without loss of generality. Both lists of vectors are linearly independent; so, according to the Exchange Lemma, we can add some vector $w_i$ to the first list to get a larger linearly independent list. This means that $v_1, \dots, v_n$ was not a maximal linearly independent set, and so (by Proposition 1.1) not a basis, contradicting our assumption. We conclude that $m = n$, as required.

(b) Let $(v_1, \dots, v_n)$ be linearly independent and let $(w_1, \dots, w_m)$ be a basis. Necessarily $n \le m$, since otherwise we could add one of the $v$s to $(w_1, \dots, w_m)$ to get a larger linearly independent set, contradicting maximality. But now we can add some $w$s to $(v_1, \dots, v_n)$ until we obtain a basis.

**Remark** We allow the possibility that a vector space has dimension zero. Such a vector space contains just one vector, the zero vector 0; a basis for this vector space consists of the empty set.

Now let $V$ be an $n$-dimensional vector space over $K$. This means that there is a basis $v_1, v_2, \dots, v_n$ for $V$. Since this list of vectors is spanning, every vector $v \in V$ can be expressed as

$$v = c_1 v_1 + c_2 v_2 + \cdots + c_n v_n$$

for some scalars $c_1, c_2, \dots, c_n \in K$. The scalars $c_1, \dots, c_n$ are the *coordinates* of $v$ (with respect to the given basis), and the *coordinate representation* of $v$ is the $n$-tuple

$$(c_1, c_2, \dots, c_n) \in K^n.$$

Now *the coordinate representation is unique*. For suppose that we also had

$$v = c_1' v_1 + c_2' v_2 + \cdots + c_n' v_n$$

for scalars $c_1', c_2' \dots, c_n'$. Subtracting these two expressions, we obtain

$$0 = (c_1 - c_1')v_1 + (c_2 - c_2')v_2 + \cdots + (c_n - c_n')v_n.$$

Now the vectors $v_1, v_2 \dots, v_n$ are linearly independent; so this equation implies that $c_1 - c_1' = 0$, $c_2 - c_2' = 0$, ..., $c_n - c_n' = 0$; that is,

$$c_1 = c_1', \quad c_2 = c_2', \quad \dots \quad c_n = c_n'.$$

Now it is easy to check that, when we add two vectors in $V$, we add their coordinate representations in $K^n$ (using coordinatewise addition); and when we multiply a vector $v \in V$ by a scalar $c$, we multiply its coordinate representation by $c$. In other words, addition and scalar multiplication in $V$ translate to the same operations on their coordinate representations. This is why we only need to consider vector spaces of the form $K^n$, as in Example 1.2.

Here is how the result would be stated in the language of abstract algebra:

**Theorem 1.4** Any $n$-dimensional vector space over a field $K$ is isomorphic to the vector space $K^n$.

## 1.3 Row and column vectors

The elements of the vector space $K^n$ are all the $n$-tuples of scalars from the field $K$. There are two different ways that we can represent an $n$-tuple: as a row, or as a column. Thus, the vector with components 1, 2 and $-3$ can be represented as a *row vector*

$$[1, 2, -3]$$

or as a column vector

$$\begin{bmatrix} 1 \\ 2 \\ -3 \end{bmatrix}.$$

(Note that we use square brackets, rather than round brackets or parentheses. But you will see the notation $(1, 2, -3)$ and the equivalent for columns.)

Both systems are in common use, and you should be familiar with both. The choice of row or column vectors makes some technical differences in the statements of the theorems, so care is needed.

There are arguments for and against both systems. Those who prefer row vectors would argue that we already use $(x,y)$ or $(x,y,z)$ for the coordinates of a point in 2- or 3-dimensional Euclidean space, so we should use the same for vectors. The most powerful argument will appear when we consider representing linear maps by matrices.

Those who prefer column vectors point to the convenience of representing, say, the linear equations

$$\begin{aligned} 2x+3y &= 5, \\ 4x+5y &= 9 \end{aligned}$$

in matrix form

$$\begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 5 \\ 9 \end{bmatrix}.$$

Statisticians also prefer column vectors: to a statistician, a vector often represents data from an experiment, and data are usually recorded in columns on a datasheet.

Following the terminology of Linear Algebra I, *I will use column vectors in these notes.* So we make a formal definition:

**Definition 1.5** Let $V$ be a vector space with a basis $B = (v_1, v_2, \dots, v_n)$. If $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$, then the *coordinate representation* of $v$ relative to the basis $B$ is

$$[v]_B = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}.$$

In order to save space on the paper, we often write this as

$$[v]_B = [c_1, c_2, \dots, c_n]^\top.$$

The symbol $\top$ is read “transpose”.

## 1.4 Change of basis

The coordinate representation of a vector is always relative to a basis. We now have to look at how the representation changes when we use a different basis.

**Definition 1.6** Let $B = (v_1, \dots, v_n)$ and $B' = (v'_1, \dots, v'_n)$ be bases for the $n$-dimensional vector space $V$ over the field $K$. The *transition matrix* $P$ from $B$ to $B'$ is the $n \times n$ matrix whose $j$th column is the coordinate representation $[v'_j]_B$ of the $j$th vector of $B'$ relative to $B$. If we need to specify the bases, we write $P_{B,B'}$.

**Proposition 1.5** Let $B$ and $B'$ be bases for the $n$-dimensional vector space $V$ over the field $K$. Then, for any vector $v \in V$, the coordinate representations of $v$ with respect to $B$ and $B'$ are related by

$$[v]_B = P\,[v]_{B'}.$$

**Proof** Let $p_{ij}$ be the $i,j$ entry of the matrix $P$. By definition, we have

$$v'_j = \sum_{i=1}^{n} p_{ij} v_i.$$

Take an arbitrary vector $v \in V$, and let

$$[v]_B = [c_1, \dots, c_n]^\top, \qquad [v]_{B'} = [d_1, \dots, d_n]^\top.$$

This means, by definition, that

$$v = \sum_{i=1}^{n} c_i v_i = \sum_{j=1}^{n} d_j v'_j.$$

Substituting the formula for $v'_j$ into the second equation, we have

$$v = \sum_{j=1}^{n} d_j \left( \sum_{i=1}^{n} p_{ij} v_i \right).$$

Reversing the order of summation, we get

$$v = \sum_{i=1}^{n} \left( \sum_{j=1}^{n} p_{ij} d_j \right) v_i.$$

Now we have two expressions for $v$ as a linear combination of the vectors $v_i$. By the uniqueness of the coordinate representation, they are the same: that is,

$$c_i = \sum_{j=1}^{n} p_{ij} d_j.$$

In matrix form, this says

$$\begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = P \begin{bmatrix} d_1 \\ \vdots \\ d_n \end{bmatrix},$$

or in other words

$$[v]_B = P\,[v]_{B'},$$

as required.

In this course, we will see four ways in which matrices arise in linear algebra. Here is the first occurrence: **matrices arise as transition matrices between bases of a vector space.**

The next corollary summarises how transition matrices behave. Here $I$ denotes the *identity matrix*, the matrix having 1s on the main diagonal and 0s everywhere else. Given a matrix $P$, we denote by $P^{-1}$ the *inverse* of $P$, the matrix $Q$ satisfying $PQ = QP = I$. Not every matrix has an inverse: we say that $P$ is *invertible* or *non-singular* if it has an inverse.

**Corollary 1.6** Let $B, B', B''$ be bases of the vector space $V$.

(a) $P_{B,B} = I$.

(b) $P_{B',B} = (P_{B,B'})^{-1}$.

(c) $P_{B,B''} = P_{B,B'}P_{B',B''}$.

This follows from the preceding Proposition. For example, for (b) we have

$$[v]_B = P_{B,B'}\,[v]_{B'}, \qquad [v]_{B'} = P_{B',B}\,[v]_B,$$

so

$$[v]_B = P_{B,B'}P_{B',B}\,[v]_B.$$

By the uniqueness of the coordinate representation, we have $P_{B,B'}P_{B',B} = I$.

**Corollary 1.7** The transition matrix between any two bases of a vector space is invertible.

This follows immediately from (b) of the preceding Corollary.

**Remark** We see that, to express the coordinate representation w.r.t. the new basis in terms of that w.r.t. the old one, we need the inverse of the transition matrix:

$$[v]_{B'} = P_{B,B'}^{-1}[v]_B.$$

**Example** Consider the vector space $\mathbb{R}^2$, with the two bases

$$B = \left( \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \end{bmatrix} \right), \qquad B' = \left( \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 2 \\ 3 \end{bmatrix} \right).$$

The transition matrix is

$$P_{B,B'} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix},$$

whose inverse is calculated to be

$$P_{B',B} = \begin{bmatrix} 3 & -2 \\ -1 & 1 \end{bmatrix}.$$

So the theorem tells us that, for any $x, y \in \mathbb{R}$, we have

$$\begin{bmatrix} x \\ y \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \end{bmatrix} = (3x - 2y) \begin{bmatrix} 1 \\ 1 \end{bmatrix} + (-x + y) \begin{bmatrix} 2 \\ 3 \end{bmatrix},$$

as is easily checked.

## 1.5 Subspaces and direct sums

**Definition 1.7** A non-empty subset of a vector space is called a *subspace* if it contains the sum of any two of its elements and any scalar multiple of any of its elements. We write $U \leq V$ to mean "$U$ is a subspace of $V$".

A subspace of a vector space is a vector space in its own right.
Subspaces can be constructed in various ways:

(a) Let $v_1, \ldots, v_n \in V$. The *span* of $(v_1, \ldots, v_n)$ is the set

$$\{c_1 v_1 + c_2 v_2 + \cdots + c_n v_n : c_1, \ldots, c_n \in K\}.$$

This is a subspace of $V$. Moreover, $(v_1, \ldots, v_n)$ is a spanning set in this subspace. We denote the span of $v_1, \ldots, v_n$ by $\langle v_1, \ldots, v_n \rangle$.

(b) Let $U_1$ and $U_2$ be subspaces of $V$. Then

- the *intersection* $U_1 \cap U_2$ is the set of all vectors belonging to both $U_1$ and $U_2$;
- the *sum* $U_1 + U_2$ is the set $\{u_1 + u_2 : u_1 \in U_1, u_2 \in U_2\}$ of all sums of vectors from the two subspaces.

Both $U_1 \cap U_2$ and $U_1 + U_2$ are subspaces of $V$.

The next result summarises some properties of these subspaces.

**Proposition 1.8** Let $V$ be a vector space over $K$.

(a) For any $v_1, \dots, v_n \in V$, the dimension of $\langle v_1, \dots, v_n \rangle$ is at most $n$, with equality if and only if $v_1, \dots, v_n$ are linearly independent.

(b) For any two subspaces $U_1$ and $U_2$ of $V$, we have

$$\dim(U_1 \cap U_2) + \dim(U_1 + U_2) = \dim(U_1) + \dim(U_2).$$

An important special case occurs when $U_1 \cap U_2$ is the zero subspace $\{0\}$. In this case, the sum $U_1 + U_2$ has the property that each of its elements has a *unique* expression in the form $u_1 + u_2$, for $u_1 \in U_1$ and $u_2 \in U_2$. For suppose that we had two different expressions for a vector $v$, say

$$v = u_1 + u_2 = u_1' + u_2', \qquad u_1, u_1' \in U_1, u_2, u_2' \in U_2.$$

Then

$$u_1 - u_1' = u_2' - u_2.$$

But $u_1 - u_1' \in U_1$, and $u_2' - u_2 \in U_2$; so this vector is in $U_1 \cap U_2$, and by hypothesis it is equal to 0, so that $u_1 = u_1'$ and $u_2 = u_2'$; that is, the two expressions are not different after all! In this case we say that $U_1 + U_2$ is the *direct sum* of the subspaces $U_1$ and $U_2$, and write it as $U_1 \oplus U_2$. Note that

$$\dim(U_1 \oplus U_2) = \dim(U_1) + \dim(U_2).$$

The notion of direct sum extends to more than two summands, but is a little complicated to describe. We state a form which is sufficient for our purposes.

**Definition 1.8** Let $U_1, \dots, U_r$ be subspaces of the vector space $V$. We say that $V$ is the *direct sum* of $U_1, \dots, U_r$, and write

$$V = U_1 \oplus \dots \oplus U_r,$$

if every vector $v \in V$ can be written uniquely in the form $v = u_1 + \cdots + u_r$ with $u_i \in U_i$ for $i = 1, \dots, r$.

**Proposition 1.9** If $V = U_1 \oplus \cdots \oplus U_r$, then

(a) $\dim(V) = \dim(U_1) + \cdots + \dim(U_r)$;

(b) if $B_i$ is a basis for $U_i$ for $i = 1, \dots, r$, then $B_1 \cup \cdots \cup B_r$ is a basis for $V$.