Chapter 1

Vector spaces

These notes are 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

Let's start by recalling some definitions from *Introduction to Algebra* and *Linear Algebra I*.

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

(A) $(\mathbb{K}, +)$ is an abelian group with identity element 0;

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

(D) the *distributive law*

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

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

If you have forgotten what an abelian group is, you should refer to *Introduction to Algebra.* In fact, the only fields we'll encounter in these notes 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.

We will not stop to prove that these structures really are fields. You may have seen $\mathbb{F}_p$ referred to as $\mathbb{Z}_p$.

The above laws or axioms are the ones we should have in mind when performing manipulations involving elements of $\mathbb{K}$. However there are a lot of axioms, and a good survival technique is to have in mind a concrete field, say $\mathbb{R}$ that we are familiar with. However, it is worthwhile operating at this level of abstraction, as vector spaces over fields other than $\mathbb{R}$ and $\mathbb{C}$ have important applications. For example, much of *Coding Theory* relates to vector spaces over $\mathbb{F}_2$

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

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

such that the following rules hold:

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

(VM) Rules for scalar multiplication:

(VM1) For any $a \in \mathbb{K}$, $u, v \in V$, we have $a(u + v) = au + av$.

(VM2) For any $a, b \in \mathbb{K}$, $v \in V$, we have $(a + b)v = av + bv$.

(VM3) For any $a, b \in \mathbb{K}$, $v \in V$, we have $(ab)v = a(bv)$.

(VM4) For any $v \in V$, we have $1v = v$ (where 1 is the identity element of $\mathbb{K}$).

Since we have two kinds of elements, namely elements of $\mathbb{K}$ and elements of V, we distinguish them by calling the elements of $\mathbb{K}$ *scalars* and the elements of V *vectors*. Typically we'll use use letters around u, v, w in the alphabet to stand for vectors, and letters around a, b and c for scalars.

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*. In some sections of the course, we'll be thinking specifically of real or complex vector spaces; in others, of vector spaces over general fields. As we noted, vector spaces over other fields are very useful in some applications, for example in coding theory, combinatorics and computer science.

Example 1.3. 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 1.1). The scalar multiple av is the vector whose

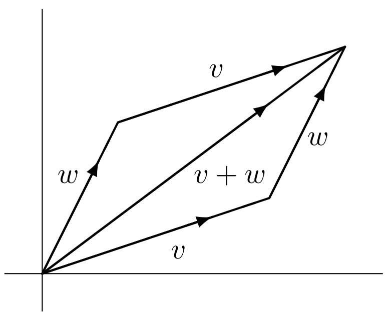

Figure 1.1: The parallelogram law

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. Thus, a vector v is just a pair (a_1, a_2) of real numbers. Now we define addition and scalar multiplication by

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

  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 may check the law $c(v + w) = cv + cw$. Let $v = (a_1, a_2)$ and $w = (b_1, b_2)$. Then we have

$$\begin{aligned}c(v + w) &= c((a_1, a_2) + (b_1, b_2))\\ &= c(a_1 + b_1, a_2 + b_2)\\ &= (ca_1 + cb_1, ca_2 + cb_2)\\ &= (ca_1, ca_2) + (cb_1, cb_2)\\ &= cv + cw.\end{aligned}$$

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

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

Example 1.5. The set $\mathbb{R}^S$ of all real functions on a set S is a vector space over $\mathbb{R}$. Vector addition is just addition of functions. Scalar multiplication is just scaling of a function by a real number.

Example 1.6. The set of all polynomials of degree $n - 1$ with coefficients in a field $\mathbb{K}$ is a vector space over $\mathbb{K}$. Vector addition is just usual addition of polynomials; scalar multiplication is just scaling of a polynomial by an element of $\mathbb{K}$. Equivalently, one can say that vector addition is coefficientwise addition, and scalar multiplication is multiplication of all coefficients by a field element. Note that from this perspective, this example is a disguised version of Example 1.4. This example was a favourite in *Linear Algebra I*!

1.2 Bases

Example 1.4 is much more general than it appears: *Every finite-dimensional vector space looks like Example 1.4.* (The meaning of "finite-dimensional" will become apparent shortly.) In *Linear Algebra I* we already verified that $\mathbb{K}^n$ is an example of a vector space over $\mathbb{K}$; in this section we go on to prove that that there are essentially no further examples.

Definition 1.7. Let V be a vector space over the field $\mathbb{K}$, and let $v_1, \dots, v_n$ be vectors in V.

(a) The vectors $v_1, v_2, \dots, v_n$ are *linearly dependent* if there are scalars $c_1, c_2, \dots, c_n$, not all zero, satisfying

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = \mathbf{0}.$$

The vectors $v_1, v_2, \dots, v_n$ are *linearly independent* if they are not linearly dependent. Equivalently, they are linearly independent if, whenever we have scalars $c_1, c_2, \dots, c_n$ satisfying

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = \mathbf{0},$$

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

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

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

(c) The list of vectors $v_1, v_2, \dots, v_n$ is a *basis* for V if it is linearly independent and spanning.

Remark 1.8. 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.

Definition 1.9. The *span* $\langle v_1, \dots, v_n \rangle$ of vectors $v_1, \dots, v_n$ is the set of all vectors that can be written as linear combinations of vectors from $v_1, \dots, v_n$:

$$\langle v_1, \dots, v_n \rangle = \big\{ c_1v_1 + c_2v_2 + \cdots + c_nv_n : (c_1, \dots, c_n) \in \mathbb{K}^n \big\}.$$

So vectors $v_1, v_2, \dots, v_n$ are spanning if $V = \langle v_1, v_2, \dots, v_n \rangle$. We will see later that the span of vectors is a vector space (or you can verify it now from the definitions).

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

Definition 1.10. Let V be a vector space over the field $\mathbb{K}$. We say that V is *finite-dimensional* if we can find vectors $v_1, v_2, \dots, v_n \in V$ that form a basis for V.

Remark 1.11. In these notes (apart from in this chapter) we are only concerned with finite-dimensional vector spaces. However, it should be noted that in various contexts, in mathematics and physics, we encounter vector spaces which are not finite dimensional.

A linearly dependent list of vectors has redundancy. It is possible to remove at least one vector from the list while keeping the span of the list the same. Here is a systematic way to do so.

Lemma 1.12. *Suppose $v_1, \dots, v_m$ is a linearly dependent list of vectors in V. There exists an index $i \in \{1, \dots, m\}$ such that*

(a) $v_i \in \langle v_1, \dots, v_{i-1} \rangle$, and

(b) $\langle v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_m \rangle = \langle v_1, \dots, v_m \rangle$.

Proof. Since $v_1, \dots, v_m$ are linearly dependent, there exist scalars $c_1, \dots, c_m$, not all zero, such that $c_1v_1 + \cdots + c_mv_m = \mathbf{0}$. Choose i to be the largest index such that $c_i \neq 0$. Then

$$v_i = -\Big(\frac{c_1}{c_i}\Big)v_1 - \cdots - \Big(\frac{c_{i-1}}{c_i}\Big)v_{i-1} \tag{1.1}$$

is an explicit expresion for v_i in terms of $v_1, \dots, v_{i-1}$, demonstrating that $v_i \in \langle v_1, \dots, v_{i-1}\rangle$. This deals with (a).

For part (b), suppose v is any vector in $\langle v_1, \dots, v_m\rangle$; by definition of span, $v = a_1v_1 + \cdots + a_mv_m$, for some $a_1, \dots, a_m \in \mathbb{K}$. Now substitute for v_i, using (1.1), to obtain an expression for v as a linear combination of vectors in $v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_m$. This expression demonstrates that $v \in \langle v_1, \dots, v_{i-1}, v_{i+1}, \dots, v_m\rangle$. Since v was arbitrary, part (b) follows. □

Lemma 1.13. *The length of any linearly independent list of vectors in V is less than or equal to the length of any spanning list of vectors.*

Proof. Suppose $v_1, \dots, v_n$ are linearly independent and $w_1, \dots, w_m$ are spanning. Start with the list $w_1, \dots, w_m$ and repeat the following step, which adds some vector v_i to the list and removes some w_j. For the first step, add vector v_1 to the front of the list to obtain $v_1, w_1, \dots, w_m$. Since the original list was spanning, the new one is linearly dependent as well as spanning. By Lemma 1.12, we may remove some w_j so that the remaining list is still spanning. By reindexing some of the w_j's we may write the resulting list as $v_1, w_2, w_3, \dots, w_m$.

In general, suppose, after some number of steps, the procedure has reached the spanning list $v_1, \dots, v_{k-1}, w_k, \dots, w_m$ (where some reindexing of vectors in $w_1, \dots, w_m$ has taken place). Add the vector v_k between v_{k-1} and w_k in the list. As before, the new list is linearly dependent, and we may apply Lemma 1.12 to remove one of the vectors in the list while retaining the property that the list is spanning. The important observation is the following: because $v_1, \dots, v_k$ are linearly independent, the removed vector cannot be one of the v_i's and so must be one of the w_j's. (See part (a) of Lemma 1.12.)

At each step we add one vector and remove one vector keeping the length of the list unchanged. We end up with a list of the form $v_1, \dots, v_n, w_{n+1}, \dots, v_m$. It follows that $m \geq n$. □

Remark 1.14. The proof establishes a little more than we needed. In fact we have essentially proved the *Steinitz Exchange Lemma*. (See, e.g., Wikipedia.)

Theorem 1.15. *Let V be a finite-dimensional vector space over a field $\mathbb{K}$. Then*

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

(b) any spanning list of vectors can be shortened (by removing some vectors) to a basis;

(c) any linearly independent list of vectors can be extended (by adding some vectors) to a basis.

Proof. (a) Suppose $\mathcal{B}_1$ and $\mathcal{B}_2$ are any two bases for V, of lengths n_1 and n_2 respectively. By Lemma 1.13, since $\mathcal{B}_1$ is linearly independent and $\mathcal{B}_2$ is spanning, $n_1 \leq n_2$. Also, since $\mathcal{B}_2$ is linearly independent and $\mathcal{B}_1$ is spanning, $n_2 \leq n_1$.

(b) Suppose $v_1, \dots, v_m$ is any spanning list for V. By Lemma 1.12, if this list is linearly dependent, we can remove some vector v_i from it, leaving a smaller spanning list. By repeating this step we must eventually reach a basis.

(c) Suppose $v_1, \ldots, v_m$ is a linearly independent list of vectors. If this list is not spanning then there must exist a vector $v_{m+1} \in V$ such that $v_{m+1} \notin \langle v_1, \ldots, v_m \rangle$. The extended list $v_1, \ldots, v_m, v_{m+1}$ remains linearly independent. (To see this, assume to the contrary that there exist scalars $a_1, \ldots, a_{m+1}$, not all zero, such that $a_1 v_1 + \cdots + a_{m+1} v_{m+1} = \mathbf{0}$. Since $v_1, \ldots, v_m$ are linearly independent, a_{m+1} cannot be 0. Then $v_{m+1} = -(a_0/a_{m+1})v_1 - \cdots - (a_m/a_{m+1})v_n$, and $v_{m+1} \in \langle v_1, \ldots v_m \rangle$ contrary to assumption.) By repeating this step we must eventually reach a basis. (Note that the process must terminate, since the vector space V is finite dimensional.) $\square$

Definition 1.16. The number of elements in a basis of a vector space V is called the *dimension* of V. Theorem 1.15 assures us that this parameter is well defined.

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 $\dim(V)$.

Remark 1.17. We allow the possibility that a vector space has dimension zero. Such a vector space contains just one vector, the zero vector $\mathbf{0}$; a basis for this vector space consists of the empty set.

Since the notion of basis of a vector space is so fundamental, it is useful in what follows to note some equivalent characterisations. These alternatives are not too difficult to verify, given Theorem 1.15.

Proposition 1.18. *The following five conditions are equivalent for a list $\mathcal{B}$ of vectors from vector space V of dimension n over $\mathbb{K}$*

(a) $\mathcal{B}$ is a basis;

(b) $\mathcal{B}$ is a maximal linearly independent list (that is, if we add any vector to the list, then the resulting list is linearly dependent);

(c) $\mathcal{B}$ is a minimal spanning list (that is, if we remove any vector from the list, then the result is no longer spanning);

(d) $\mathcal{B}$ is linearly independent and has length n;

(e) $\mathcal{B}$ is spanning and has length n.

Now let V be an n-dimensional vector space over $\mathbb{K}$. This means that there is a basis $v_1, v_2, \ldots, v_n$ 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 $c_1, c_2, \ldots, c_n \in \mathbb{K}$. The scalars $c_1, \ldots, c_n$ are the *coordinates* of v (with respect to the given basis), and the *coordinate representation* of v is the n-tuple

$$(c_1, c_2, \ldots, c_n) \in \mathbb{K}^n.$$

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 $c'_1, c'_2 \ldots, c'_n$. Subtracting these two expressions, we obtain

$$\mathbf{0} = (c_1 - c'_1)v_1 + (c_2 - c'_2)v_2 + \cdots + (c_n - c'_n)v_n.$$

Now the vectors $v_1, v_2 \ldots, v_n$ 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,

$$c_1 = c'_1, \quad c_2 = c'_2, \quad \ldots \quad c_n = c'_n.$$

Remark 1.19. In this course, the notation v_i, w_i, etc., stands for the ith vector in a sequence of vectors. It will not be used to denote the ith coordinate of the vector v (which would be a scalar). We'll use different letters for the vector and for its coordinates.

Now it is easy to check that, when we add two vectors in V, we add their coordinate representations in $\mathbb{K}^n$ (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 $\mathbb{K}^n$, as in Example 1.4.

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

Theorem 1.20. *Any n-dimensional vector space over a field $\mathbb{K}$ is isomorphic to the vector space $\mathbb{K}^n$.*

1.3 Row and column vectors

The elements of the vector space $\mathbb{K}^n$ are all the n-tuples of scalars from the field $\mathbb{K}$. 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*

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

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 in other books!)

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.

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.

We will use column vectors in these notes. So we make a formal definition:

Definition 1.21. Let V be a vector space with a basis $\mathcal{B} = (v_1, v_2, \ldots, v_n)$. If $v = c_1v_1 + c_2v_2 + \cdots + c_nv_n$, then the *coordinate representation* of v relative to the basis $\mathcal{B}$ is

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

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

$$[v]_{\mathcal{B}} = \begin{bmatrix} c_1 & c_2 & \ldots & c_n \end{bmatrix}^\top,$$

where the symbol $\top$ is read "transpose".

Note that the coordinate representation of a vector is always relative to a basis.

Let $\mathcal{B} = (v_1, \ldots, v_n)$ and $\mathcal{B}' = (v_1', \ldots, v_n')$ be different bases of an n-dimensional vector space V over the field $\mathbb{K}$. Recall from *Linear Algebra I* that there is an $n \times n$ *transitition matrix* $P_{\mathcal{B},\mathcal{B}'}$ that translates coordinate representations relative to $\mathcal{B}'$ to coordinate representations relative to $\mathcal{B}$. Specifically, $[v]_{\mathcal{B}} = P_{\mathcal{B},\mathcal{B}'}\,[v]_{\mathcal{B}'}$ for all vectors $v \in V$.

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.**

Let I denote 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, that is to say, 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.

We recall from *Linear Algebra I* some facts about transition matrices, which come directly from the definition, using uniqueness of the coordinate representation. Let $\mathcal{B}, \mathcal{B}', \mathcal{B}''$ be bases of the vector space V. Then

(a) $P_{\mathcal{B},\mathcal{B}} = I$,

(b) $P_{\mathcal{B}',\mathcal{B}} = (P_{\mathcal{B},\mathcal{B}'})^{-1}$; in particular, the transition matrix is invertible, and

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

To see that (b) holds, let's transform the coordinate representation of u relative to basis $\mathcal{B}$ by multiplication by $P_{\mathcal{B}',\mathcal{B}}$:

$$P_{\mathcal{B}',\mathcal{B}}[u]_{\mathcal{B}} = P_{\mathcal{B}',\mathcal{B}}\big(P_{\mathcal{B},\mathcal{B}'}[u]_{\mathcal{B}'}\big) = \big(P_{\mathcal{B},\mathcal{B}'}^{-1}P_{\mathcal{B},\mathcal{B}'}\big)[u]_{\mathcal{B}'} = [u]_{\mathcal{B}'}.$$

We obtain the coordinate representation of u relative to basis $\mathcal{B}'$, as desired.

To see that (c) holds, transform the coordinate representation of u relative to basis $\mathcal{B}''$ by multiplication by $P_{\mathcal{B},\mathcal{B}''}$:

$$P_{\mathcal{B},\mathcal{B}''}[u]_{\mathcal{B}''} = \big(P_{\mathcal{B},\mathcal{B}'}P_{\mathcal{B}',\mathcal{B}''}\big)[u]_{\mathcal{B}''} = P_{\mathcal{B},\mathcal{B}'}\big(P_{\mathcal{B}',\mathcal{B}''}[u]_{\mathcal{B}''}\big) = P_{\mathcal{B},\mathcal{B}'}[u]_{\mathcal{B}'} = [u]_{\mathcal{B}}.$$

We obtain the coordinate representation of u relative to basis $\mathcal{B}$, as desired.

Example 1.22. Suppose that $\mathcal{B} = (v_1, v_2)$ and $\mathcal{B}' = (v_1', v_2')$ are different bases of a 2-dimensional vector space V over $\mathbb{R}$. Since $\mathcal{B}$ is a basis of V we can express the basis vectors of $\mathcal{B}'$ in terms of $\mathcal{B}$. Suppose, in fact, that

$$v_1' = v_1 + v_2 \qquad \text{and} \qquad v_2' = 2v_1 + 3v_2.$$

Then the transition matrix from $\mathcal{B}'$ to $\mathcal{B}$ is

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

Note that the first column of $P_{\mathcal{B},\mathcal{B}'}$ is just $[v_1']_{\mathcal{B}}$, i.e., the coordinate representation of the vector v_1' relative to the basis $\mathcal{B}$, and the second column is just $[v_2']_{\mathcal{B}}$. This gives an easy way to write down $P_{\mathcal{B},\mathcal{B}'}$.

Suppose that the coordinate representation of some vector u relative to the basis $\mathcal{B}'$ is $[u]_{\mathcal{B}'} = \begin{bmatrix} a & b \end{bmatrix}^\mathsf{T}$. Then, from the definition of transition matrix, we should have

$$[u]_{\mathcal{B}} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a+2b \\ a+3b \end{bmatrix}.$$

We can check the result as follows:

$$u = av_1' + bv_2' = a(v_1 + v_2) + b(2v_1 + 3v_2) = (a+2b)v_1 + (a+3b)v_2.$$

So indeed $[u]_{\mathcal{B}} = \begin{bmatrix} a+2b & a+3b \end{bmatrix}^\mathsf{T}$ as expected.

The transition matrix from $\mathcal{B}$ to $\mathcal{B}'$ is the inverse of $P_{\mathcal{B},\mathcal{B}'}$:

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

Finally, suppose $\mathcal{B}'' = (v_1'', v_2'')$ is a third basis of V, related to $\mathcal{B}'$ by $v_1'' = 3v_1' - 2v_2'$ and $v_2'' = -2v_1' + v_2'$. Then

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

and

$$P_{\mathcal{B},\mathcal{B}''} = P_{\mathcal{B},\mathcal{B}'} P_{\mathcal{B}',\mathcal{B}''} = \begin{bmatrix} 1 & 2 \\ 1 & 3 \end{bmatrix} \begin{bmatrix} 3 & -2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ -3 & 1 \end{bmatrix}.$$

Note that this example provides additional insight into why matrix multiplication is defined the way it is: in this instance, it provides the correct rule for composing transition matrices.

1.4 Subspaces and direct sums

Definition 1.23. Suppose V is a vector space over $\mathbb{K}$. We say that U is a *subspace* of V if U is a subset of V, and U is itself a vector space (with respect the same operations of vector addition and scalar multiplication).

We write $U \leq V$ to mean "U is a subspace of V".

Lemma 1.24. *Suppose U is a non-empty subset of a vector space V. The following conditions are equivalent:*

1. *U is a subspace of V;*

2. *U is closed under vector addition and scalar multiplication. (That is to say, $u+u' \in U$ and $cu \in U$ for any vectors $u, u' \in U$ and scalar $c \in \mathbb{K}$.)*

Proof. Since any vector space is closed under vector addition and scalar multiplication, it is clear that (1) implies (2).

Suppose now that (2) holds. For any vector $u \in U$, we know that $-u = (-1)u$ is in U (by closure under scalar multiplication). Also, since U is non-empty, the additive identity $\mathbf{0} = u - u$ is in U. So (2) assures us that the operations of vector addition, taking the inverse of a vector, and scalar multiplication all make sense in U; moreover, U contains an additive identity. The vector space axioms (VA) and (VM) for U are inherited from V: since they hold in the larger set, they certainly hold in the smaller. (Go through all five axioms and convince yourself of this fact.) $\square$

Subspaces can be constructed in various ways:

(a) Recall that the span of vectors $v_1, \ldots, v_k \in V$ is the set

$$\{c_1v_1 + c_2v_2 + \cdots + c_kv_k : c_1, \ldots, c_k \in \mathbb{K}\}.$$

This is a subspace of V. Moreover, vectors $v_1, \ldots, v_k$ are spanning in this subspace.

(b) Let U and W be subspaces of V. Then

- the *intersection* $U \cap W$ is the set of all vectors belonging to both U and W;
- the *sum* $U + W$ is the set $\{u + w : u \in U, w \in W\}$ of all sums of vectors from the two subspaces.

Both $U \cap W$ and $U + W$ are subspaces of V.

We will just check (a) here, leaving (b) as an exercise. By Lemma 1.24, we just need to check closure under vector addition and scalar multiplication. So suppose $v = c_1v_1 + \cdots + c_kv_k$ and $v' = c_1'v_1 + \cdots + c_k'v_k$ are vectors in the span $\langle v_1, \ldots, v_k\rangle$ of $v_1, \ldots, v_k \in V$. Then $v + v' = (c_1v_1 + \cdots + c_kv_k) + (c_1'v_1 + \cdots + c_k'v_k) = (c_1 + c_1')v_1 + \cdots + (c_k + c_k')v_k$, which is clearly also in the span $\langle v_1, \ldots, v_k\rangle$. Also for any $a \in \mathbb{K}$, we have $av = a(c_1v_1) + \cdots + a(c_kv_k) = (ac_1)v_1 + \cdots + (ac_k)v_k$, which is again clearly in $\langle v_1, \ldots, v_k\rangle$.

Theorem 1.25. *Let V be a vector space over $\mathbb{K}$. For any two subspaces U and W of V, we have*

$$\dim(U \cap W) + \dim(U + W) = \dim(U) + \dim(W).$$

Proof. Let $v_1, \ldots, v_i$ be a basis for $U \cap W$. By Theorem 1.15(c) we can extend this basis to a basis $v_1, \ldots, v_i, u_1, \ldots, u_j$ of U and a basis $v_1, \ldots, v_i, w_1, \ldots, w_k$ of W. If we can show that $v_1, \ldots, v_i, u_1, \ldots, u_j, w_1, \ldots, w_k$ is a basis of $U + W$ then we are done, since then

$$\dim(U \cap W) = i, \quad \dim(U) = i + j, \quad \dim(W) = i + k, \quad \text{and} \quad \dim(U + V) = i + j + k,$$

and both sides of the identity we are aiming to prove are equal to $2i + j + k$.

Since every vector in U (respectively W) can be expressed as a linear combination of $v_1, \ldots, v_i, u_1, \ldots, u_j$ (respectively $v_1, \ldots, v_1, w_1, \ldots, w_k$), it is clear that the list of vectors $v_1, \dots, v_1, u_1, \dots, u_j, w_1, \dots, w_k$ spans $U + W$. So we just need to show that the list $v_1, \dots, v_1, u_1, \dots, u_j, w_1, \dots, w_k$ is linearly independent.

Consider any linear relationship

$$a_1 v_1 + \cdots + a_i v_i + b_1 u_1 + \cdots + b_j u_j + c_1 w_1 + \cdots + c_k w_k = \mathbf{0};$$

we need to show that $a_1, \dots, a_i, b_1, \dots, b_j, c_1, \dots, c_k$ are all zero. Writing

$$c_1 w_1 + \cdots + c_k w_k = -a_1 v_1 - \cdots - a_i v_i - b_1 u_1 - \cdots - b_j u_j,$$

we see that $c_1 w_1 + \cdots + c_k w_k \in U$. But, by construction, $c_1 w_1 + \cdots + c_k w_k \in W$, so in fact $c_1 w_1 + \cdots + c_k w_k \in U \cap W$. Since $v_1, \dots, v_i$ is a basis for $U \cap W$ we have

$$c_1 w_1 + \cdots + c_k w_k = d_1 v_1 + \cdots d_i v_i,$$

for some scalars $d_1, \dots d_i$. But this implies that $c_1 = \cdots = c_k = 0$ (and, incidentally, $d_1 = \cdots = d_i = 0$), since $v_1, \dots, v_i, w_1, \dots, w_k$ is a basis for W and hence linearly independent. A similar argument establishes $b_1 = \cdots = b_k = 0$. But now $a_1 = \cdots = a_i = 0$, since the list $v_1, \dots, v_i$ is linearly independent. $\square$

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

$$v = u + w = u' + w', \qquad \text{for some } u, u' \in U \text{ and } w, w' \in W.$$

Then

$$u - u' = w' - w.$$

But $u - u' \in U$, and $w' - w \in W$, and hence

$$u - u' = w' - w \in U \cap W = \{\mathbf{0}\}.$$

It follows that $u = u'$ and $w = w'$; that is, the two expressions for v are not different after all! In this case we say that $U + W$ is the *direct sum* of the subspaces U and W, and write it as $U \oplus W$. Note that

$$\dim(U \oplus W) = \dim(U) + \dim(W).$$

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 1.26. 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

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

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 $i = 1, \dots, r$.

There is an equivalent characterisation of direct sum that will be useful later.

Lemma 1.27. *Suppose $U_1, \ldots, U_r$ are subspaces of V, and $V = U_1 + \cdots + U_r$. Then the following are equivalent:*

(a) V is the direct sum *of $U_1, \ldots, U_r$.*

(b) For all vectors $u_1 \in U_1, \ldots, u_r \in U_r$, it is the case that $u_1 + \cdots + u_r = \mathbf{0}$ implies $u_1 = \cdots = u_r = \mathbf{0}$.

Proof. (1) $\implies$ (2). Suppose $u_1 + \cdots + u_r = \mathbf{0}$, where $u_1 \in U_1, \ldots, u_r \in U_r$. Certainly $u_1 = \cdots = u_r = \mathbf{0}$ is one way this situation may occur. But the definition of direct sum tells us that such an expression is unique. So, indeed, $u_1 = \cdots = u_r = \mathbf{0}$ as required.

(2) $\implies$ (1). Suppose $v \in V$ and that $v = u_1 + \cdots + u_r$ and $v = u_1' + \cdots + u_r'$ are two ways of expressing v, with $u_1, u_1' \in U_1, \ldots, u_r, u_r' \in U_r$. Then

$$(u_1 - u_1') + \cdots + (u_r - u_r') = (u_1 + \cdots + u_r) - (u_1' + \cdots + u_r') = v - v = \mathbf{0}.$$

From condition (2), we deduce that $u_1 - u_1' = \cdots = u_r - u_r' = \mathbf{0}$. Thus, $u_1 = u_1', \ldots, u_r = u_r'$ as required. □

Note the similarity between the condition described in Lemma 1.27(b) and the definition of linear independence. In fact, $v_1, \ldots, v_n$ is a basis for a vector space V if and only if $V = \langle v_1 \rangle \oplus \cdots \oplus \langle v_n \rangle$. In a sense, a direct sum generalises the concept of basis.

Lemma 1.28. *If $V = U_1 \oplus \cdots \oplus U_r$, then*

(a) if $\mathcal{B}_i$ is a basis for U_i for $i = 1, \ldots, r$, then $\mathcal{B} = (\mathcal{B}_1, \cdots, \mathcal{B}_r)$, i.e., the concatenation of the lists $\mathcal{B}_1, \ldots, \mathcal{B}_r$, is a basis for V;

(b) $\dim(V) = \dim(U_1) + \cdots + \dim(U_r)$.

Proof. Since every vector $v \in V$ may be expressed as $v = u_1 + \cdots + u_r$ with $u_i \in U_i$, and every $u_i \in U_i$ may be expressed as a linear combination of basis vectors in $\mathcal{B}_i$, we see that V is contained in the span of $\mathcal{B}$. So we just need to verify that the list $\mathcal{B}$ is linearly independent.

Let $d_i = \dim(U_i)$ and $\mathcal{B}_i = (u_{i,1}, \ldots, u_{i,d_i})$, for $1 \le i \le r$, be an explicit enumeration of the basis vectors $\mathcal{B}_i$. Suppose that some linear combination of the basis vectors $\mathcal{B}$ sums to $\mathbf{0}$. We can express this linear combination as $u_1 + \cdots + u_r = \mathbf{0}$, where $u_i = a_{i,1}u_{i,1} + \cdots + a_{i,d_i}u_{i,d_i}$ for some scalars $a_{i,1}, \ldots, a_{i,d_i} \in \mathbb{K}$.

By Lemma 1.27, $u_i = \mathbf{0}$ for all $1 \le i \le r$. Then, since $\mathcal{B}_i$ is a basis and hence linearly independent, $a_{i,1} = \cdots = a_{i,d_i} = 0$. Since the linear combination of basis vectors $\mathcal{B}$ was arbitrary, we deduce that $\mathcal{B}$ is linearly independent.

This deals with part (a). Part (b) follows immediately, since

$$\dim(V) = |\mathcal{B}| = |\mathcal{B}_1| + \cdots + |\mathcal{B}_r| = \dim(U_1) + \cdots + \dim(U_r).$$

□

Remark 1.29. The results in this chapter apply to all finite dimensional vector spaces over $\mathbb{K}$, regardless of the field $\mathbb{K}$. In our proofs, we used nothing beyond the general axioms of a field. In some later chapters we need restrict our attention to particular fields, typically $\mathbb{R}$ or $\mathbb{C}$.