

# Chapter 1

## Linear Algebra

In this part of the course we will review some basic linear algebra. The topics covered include: real and complex vector spaces and linear maps, bases, matrices, inner products, eigenvalues and eigenvectors. We start from the familiar setting in two dimensions and introduce the necessary formalism to be able to work with vectors in an arbitrary number of dimensions. We end the chapter with a physical application: the study of normal modes of an oscillatory system.

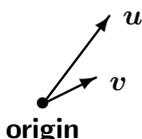
### 1.1 Vector spaces

Physics requires both *scalar* quantities like mass, temperature, charge which are uniquely specified by its magnitude in some units, e.g.,  $300^\circ K$ ,  $7 \text{ Kg}$ ,... and also *vectorial* quantities like velocity, force, angular momentum, which are specified both by a magnitude and a direction.

In the first part of the course we will study the general features shared by these vectorial quantities. As this is a course in mathematical techniques, we must abstract what these quantities have in common (the ‘mathematical’ part) while at the same time keeping a pragmatic perspective throughout (the ‘techniques’ part). This is not a mathematics course, but nevertheless a certain amount of formalism is needed. Some of you may not have seen formal definitions before, so we will start by motivating the notion of a vector space. For definiteness we will consider displacements in two dimensions; that is, in the plane.

### 1.1.1 Displacements in the plane

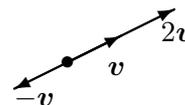
Every displacement in the plane has an initial or starting point and a final point. We will only consider displacements which have a common starting point: the **origin**.



Any point in the plane is then understood as the final point of a displacement from the origin. We will depict such displacements by an arrow starting at the origin and ending at the final point. We will denote such displacements by boldfaced letters, like  $\mathbf{u}$ ,  $\mathbf{v}$ . In lecture it is hard to write in boldface, so we use the notation  $\vec{u}$ ,  $\vec{v}$  which is not just easier to write but has the added benefit of being mnemonic, since the arrow reminds us that it is a displacement. We will say that displacements like  $\mathbf{u}$ ,  $\mathbf{v}$  are **vectors**.

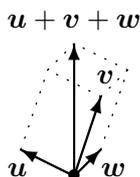
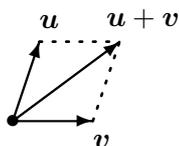
What can one do with vectors?

For example, vectors can be multiplied by real numbers (the **scalars**). If  $\lambda > 0$  is a positive real number and  $\mathbf{v}$  is a vector, then  $\lambda\mathbf{v}$  is a vector pointing in the same direction as  $\mathbf{v}$  but  $\lambda$  times as long as  $\mathbf{v}$ , e.g.,  $2\mathbf{v}$  is twice as long as  $\mathbf{v}$  but points in the same direction. In the same manner,  $-\lambda\mathbf{v}$  is a vector pointing in the direction opposite to  $\mathbf{v}$  but  $\lambda$  times as long as  $\mathbf{v}$ . We call this operation **scalar multiplication**.



This operation satisfies two properties which are plain to see from the pictures. The first says that if  $\mathbf{v}$  is any vector and  $\lambda$  and  $\mu$  are real numbers, then  $\lambda(\mu\mathbf{v}) = (\lambda\mu)\mathbf{v}$ . The second property is totally obvious from the picture:  $1\mathbf{v} = \mathbf{v}$ .

You should also be familiar from the study of, say, forces, with the fact that vectors can be added.



Indeed, if  $\mathbf{u}$  and  $\mathbf{v}$  are vectors, then their sum  $\mathbf{u} + \mathbf{v}$  is the diagonal from the origin to the opposite vertex in the parallelogram defined by  $\mathbf{u}$  and  $\mathbf{v}$ , as in the picture. This operation is called **vector addition** or simply **addition**. It follows from the picture that  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$ , so that we get the same result regardless of the order in which we add the vectors. One says that *vector addition is commutative*.

Vector addition is also *associative*. This means that, as can be seen in the picture, when adding three vectors  $\mathbf{u}$ ,  $\mathbf{v}$ , and  $\mathbf{w}$  it does not matter whether we first add  $\mathbf{u}$  and  $\mathbf{v}$  and add  $\mathbf{w}$  to the result:  $(\mathbf{u} + \mathbf{v}) + \mathbf{w}$  or whether we first add  $\mathbf{v}$  and  $\mathbf{w}$  and add the result to  $\mathbf{u}$ :  $\mathbf{u} + (\mathbf{v} + \mathbf{w})$ .

Another easy property of vector addition is the existence of a vector  $\mathbf{0}$  such that when added to any vector  $\mathbf{v}$  gives back  $\mathbf{v}$  again; that is,

$$\mathbf{0} + \mathbf{v} = \mathbf{v} \text{ for all vectors } \mathbf{v}.$$

Clearly the zero vector  $\mathbf{0}$  corresponds to the trivial displacement which starts and ends at the origin, or in other words, to no displacement at all.

Similarly, given any vector  $\mathbf{v}$  there is a vector  $-\mathbf{v}$  which obeys  $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ . We will often employ the notation  $\mathbf{u} - \mathbf{v}$  to denote  $\mathbf{u} + (-\mathbf{v})$ .

Finally, notice that scalar multiplication and addition are compatible: scalar multiplication and addition can be performed in any order:

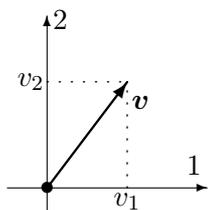
$$\lambda(\mathbf{u} + \mathbf{v}) = \lambda\mathbf{u} + \lambda\mathbf{v} \quad \text{and} \quad (\lambda + \mu)\mathbf{v} = \lambda\mathbf{v} + \mu\mathbf{v}.$$

The former identity says that *scalar multiplication is distributive over vector addition*. Notice that, in particular, it follows that  $0\mathbf{v} = \mathbf{0}$  for all  $\mathbf{v}$ .

### 1.1.2 Displacements in the plane (revisited)

There is no conceptual reason why one should not consider displacements in space, i.e., in three dimensions, as opposed to the plane. The pictures get a little harder to draw, but in principle it can still be done with better draughtsmanship than mine. In physics, though, one needs to work with vectors in more than three dimensions—in fact, as in Quantum Mechanics, one often needs to work with vectors in an infinite number of dimensions. Pictures like the ones above then become of no use, and one needs to develop a notation we can calculate with.

Let us consider again the displacements in the plane, but this time with a more algebraic notation.



The first thing we do is to draw two cartesian axes centred at the origin: axis 1 and axis 2. Then every displacement  $\mathbf{v}$  from the origin can be written as an ordered pair  $(v_1, v_2)$  of real numbers, corresponding to the components of the displacement  $\mathbf{v}$  along the cartesian axes, as in the figure.

Let us define the set

$$\mathbb{R}^2 = \{(v_1, v_2) \mid v_i \in \mathbb{R} \text{ for } i = 1, 2\}$$

of ordered pairs of real numbers.

The above notation may need some explaining. The notation ' $v_i \in \mathbb{R}$ ' is simply shorthand for the phrase ' $v_i$  is a real number;' whereas the notation

' $\{(v_1, v_2) \mid v_i \in \mathbb{R} \text{ for } i = 1, 2\}$ ' is shorthand for the phrase 'the set consisting of pairs  $(v_1, v_2)$  such that both  $v_1$  and  $v_2$  are real numbers.'

The set  $\mathbb{R}^2$  is in one-to-one correspondence with the set of displacements, for clearly every displacement gives rise to one such pair and every such pair gives rise to a displacement. We can therefore try to guess how to define the operations of vector addition and scalar multiplication in  $\mathbb{R}^2$  in such a way that they correspond to the way they are defined for displacements.

From the pictures defining addition and scalar multiplication, one sees that if  $\lambda \in \mathbb{R}$  is a real number, then

$$\lambda(v_1, v_2) = (\lambda v_1, \lambda v_2) , \quad (\text{scalar multiplication})$$

and also

$$(u_1, u_2) + (v_1, v_2) = (u_1 + v_1, u_2 + v_2) . \quad (\text{addition})$$

The zero vector corresponds with no displacement at all, hence it is given by the pair corresponding to the origin  $(0, 0)$ . It follows from the addition rule that

$$(0, 0) + (v_1, v_2) = (v_1, v_2) .$$

Similarly,  $-(v_1, v_2) = (-v_1, -v_2)$ . In fact it is not hard to show (do it!) that addition and scalar multiplication obey the same properties as they did for displacements.

The good thing about this notation is that there is no reason why we should restrict ourselves to *pairs*. Indeed, why not consider the set

$$\mathbb{R}^N = \{(v_1, v_2, \dots, v_N) \mid v_i \in \mathbb{R} \text{ for } i = 1, 2, \dots, N\} ,$$

of ordered  $N$ -*tuples* of real numbers? We can define addition and scalar multiplication in the same way as above:

(addition)

$$\begin{aligned} (u_1, u_2, \dots, u_N) + (v_1, v_2, \dots, v_N) \\ = (u_1 + v_1, u_2 + v_2, \dots, u_N + v_N) , \end{aligned}$$

(multiplication by scalars)

$$\lambda(v_1, v_2, \dots, v_N) = (\lambda v_1, \lambda v_2, \dots, \lambda v_N) \quad \text{for } \lambda \in \mathbb{R} .$$

In the homework you are asked to prove that these operations on  $\mathbb{R}^N$  obey the same properties that displacements do: commutativity, associativity, distributivity,... These properties can be formalised in the concept of an abstract vector space.

### 1.1.3 Abstract vector spaces

We are finally ready to formalise the observations made above into the definition of an abstract vector space. We say that this is an abstract vector space, because it does not refer to any concrete example.

A **real vector space** consists of the following data:

- Two sets:
  - the set of vectors, which we shall denote  $\mathbb{V}$ , and whose elements we will write as  $\mathbf{u}$ ,  $\mathbf{v}$ ,  $\mathbf{w}$ ,  $\dots$ , and
  - the set of scalars, which for a real vector space is simply the set  $\mathbb{R}$  of real numbers. We will use lowercase Greek letters from the middle of the alphabet:  $\lambda$ ,  $\mu$ ,  $\dots$  to represent real numbers.
- Two operations:
  - Scalar multiplication, which takes a scalar  $\lambda$  and a vector  $\mathbf{v}$  and produces another vector  $\lambda \mathbf{v}$ . One often abbreviates this as

$$\begin{aligned} \text{scalar multiplication} : \mathbb{R} \times \mathbb{V} &\rightarrow \mathbb{V} \\ (\lambda, \mathbf{v}) &\mapsto \lambda \mathbf{v} . \end{aligned}$$

- Vector addition, which takes two vectors  $\mathbf{u}$  and  $\mathbf{v}$  and produces a third vector denoted  $\mathbf{u} + \mathbf{v}$ . Again one can abbreviate this as

$$\begin{aligned} \text{vector addition} : \mathbb{V} \times \mathbb{V} &\rightarrow \mathbb{V} \\ (\mathbf{u}, \mathbf{v}) &\mapsto \mathbf{u} + \mathbf{v} . \end{aligned}$$

- Eight properties (or *axioms*):
  - V1 (*associativity*)  $(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$  for all  $\mathbf{u}$ ,  $\mathbf{v}$  and  $\mathbf{w}$ ;
  - V2 (*commutativity*)  $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$  for all  $\mathbf{u}$  and  $\mathbf{v}$ ;
  - V3 There exists a zero vector  $\mathbf{0}$  which obeys  $\mathbf{0} + \mathbf{v} = \mathbf{v}$  for all  $\mathbf{v}$ ;
  - V4 For any given  $\mathbf{v}$ , there exists a vector  $-\mathbf{v}$  such that  $\mathbf{v} + (-\mathbf{v}) = \mathbf{0}$ ;
  - V5  $\lambda(\mu \mathbf{v}) = (\lambda\mu) \mathbf{v}$  for all  $\mathbf{v}$ ,  $\lambda$  and  $\mu$ ;
  - V6  $1 \mathbf{v} = \mathbf{v}$  for all  $\mathbf{v}$ ;
  - V7  $(\lambda + \mu) \mathbf{v} = \lambda \mathbf{v} + \mu \mathbf{v}$  for all  $\lambda$  and  $\mu$  and  $\mathbf{v}$ ;
  - V8 (*distributivity*)  $\lambda(\mathbf{u} + \mathbf{v}) = \lambda \mathbf{u} + \lambda \mathbf{v}$  for all  $\lambda$ ,  $\mathbf{u}$  and  $\mathbf{v}$ .

This formidable looking definition might at first seem to be something you had rather forget about. Actually you will see that after using it in practice it will become if not intuitive at least more sensible. Formal definitions like this one above are meant to capture the essence of what is being defined. Every vector space is an instance of an abstract vector space, and it will inherit all the properties of an abstract vector space. In other words, we can be sure that any result that we obtain for an abstract vector space will also hold for any concrete example.

A typical use of the definition is recognising vector spaces. To go about this one has to identify the sets of vectors and scalars, and the operations of scalar multiplication and vector addition and then check that all eight axioms are satisfied. In the homework I ask you to do this for two very different looking spaces:  $\mathbb{R}^N$  which we have already met, and the set consisting of real-valued functions on the interval  $[-1, 1]$ . In the course of these lectures we will see many others.



You may wonder whether all eight axioms are necessary. For example, you may question the necessity of V4, given V3. Consider the following subset of  $\mathbb{R}^2$ :

$$\{(v_1, v_2) \mid v_i \in \mathbb{R} \text{ and } v_2 \geq 0\} \subset \mathbb{R}^2$$

consisting of pairs of real numbers where the second real number in the pair is non-negative. In terms of displacements, it corresponds to the upper half-plane. You can check that the first two axioms V1 and V2 are satisfied, and that the zero vector  $(0, 0)$  belongs to this subset. However  $-(v_1, v_2) = (-v_1, -v_2)$  whence if  $v_2$  is non-negative,  $-v_2$  cannot be non-negative unless  $v_2 = 0$ . Therefore V4 is not satisfied. In fact, neither are V5, V7 and V8 unless we restrict the scalars to be non-negative real numbers. A more challenging exercise is to determine whether V6 is really necessary.



The zero vector  $\mathbf{0}$  of axiom V3 is unique. To see this notice that if there were another  $\mathbf{0}'$  which also satisfies V3, then

$$\begin{aligned} \mathbf{0}' &= \mathbf{0} + \mathbf{0}' && \text{(by V3 for } \mathbf{0}\text{)} \\ &= \mathbf{0}' + \mathbf{0} && \text{(by V2)} \\ &= \mathbf{0} . && \text{(by V3 for } \mathbf{0}'\text{)} \end{aligned}$$

Similarly the vector  $-\mathbf{v}$  in V4 is also unique. In fact, suppose that there are two vectors  $\mathbf{u}_1$  and  $\mathbf{u}_2$  which satisfy:  $\mathbf{v} + \mathbf{u}_1 = \mathbf{0}$  and  $\mathbf{v} + \mathbf{u}_2 = \mathbf{0}$ . Then they are equal:

$$\begin{aligned} \mathbf{u}_1 &= \mathbf{0} + \mathbf{u}_1 && \text{(by V3)} \\ &= (\mathbf{v} + \mathbf{u}_2) + \mathbf{u}_1 && \text{(by hypothesis)} \\ &= \mathbf{v} + (\mathbf{u}_2 + \mathbf{u}_1) && \text{(by V1)} \\ &= \mathbf{v} + (\mathbf{u}_1 + \mathbf{u}_2) && \text{(by V2)} \\ &= (\mathbf{v} + \mathbf{u}_1) + \mathbf{u}_2 && \text{(by V1)} \\ &= \mathbf{0} + \mathbf{u}_2 && \text{(by hypothesis)} \\ &= \mathbf{u}_2 . && \text{(by V3)} \end{aligned}$$

A final word on notation: although we have defined a real vector space as two sets, vectors  $\mathbb{V}$  and real scalars  $\mathbb{R}$ , and two operations satisfying some axioms, one often simply says that ‘ $\mathbb{V}$  is a real vector space’ leaving the other bits in the definition implicit. Similarly in what follows, and unless otherwise stated, we will implicitly assume that the scalars are real, so that whenever we say ‘ $\mathbb{V}$  is a vector space’ we shall mean that  $\mathbb{V}$  is a real vector space.

### 1.1.4 Vector subspaces

A related notion to a vector space is that of a vector subspace. Suppose that  $\mathbb{V}$  is a vector space and let  $\mathbb{W} \subset \mathbb{V}$  be a subset. This means that  $\mathbb{W}$  consists of some (but not necessarily all) of the vectors in  $\mathbb{V}$ . Since  $\mathbb{V}$  is a vector space, we know that we can add vectors in  $\mathbb{W}$  and multiply them by scalars, but does that make  $\mathbb{W}$  into a vector space in its own right? As we saw above with the example of the upper half-plane, not every subset  $\mathbb{W}$  will itself be a vector space. For this to be the case we have to make sure that the following two axioms are satisfied:

S1 If  $\mathbf{v}$  and  $\mathbf{w}$  are vectors in  $\mathbb{W}$ , then so is  $\mathbf{v} + \mathbf{w}$ ; and

S2 For any scalar  $\lambda \in \mathbb{R}$ , if  $\mathbf{w}$  is any vector in  $\mathbb{W}$ , then so is  $\lambda \mathbf{w}$ .

If these two properties are satisfied we say that  $\mathbb{W}$  is a **vector subspace** of  $\mathbb{V}$ . One also often sees the phrases ‘ $\mathbb{W}$  is a subspace of  $\mathbb{V}$ ’ and ‘ $\mathbb{W}$  is a linear subspace of  $\mathbb{V}$ .’

Let us make sure we understand what these two properties mean. For  $\mathbf{v}$  and  $\mathbf{w}$  in  $\mathbb{W}$ ,  $\mathbf{v} + \mathbf{w}$  belongs to  $\mathbb{V}$  because  $\mathbb{V}$  is a vector space. The question is whether  $\mathbf{v} + \mathbf{w}$  belongs to  $\mathbb{W}$ , and S1 says that it does. Similarly, if  $\mathbf{w} \in \mathbb{W}$  is a vector in  $\mathbb{W}$  and  $\lambda \in \mathbb{R}$  is any scalar, then  $\lambda \mathbf{w}$  belongs to  $\mathbb{V}$  because  $\mathbb{V}$  is a vector space. The question is whether  $\lambda \mathbf{w}$  also belongs to  $\mathbb{W}$ , and S2 says that it does.

You may ask whether we should not also require that the zero vector  $\mathbf{0}$  also belongs to  $\mathbb{W}$ . In fact this is guaranteed by S2, because for any  $\mathbf{w} \in \mathbb{W}$ ,  $\mathbf{0} = 0 \mathbf{w}$  (why?) which belongs to  $\mathbb{W}$  by S2. From this point of view, it is S2 that fails in the example of the upper half-plane, since scalar multiplication by a negative scalar  $\lambda < 0$  takes vectors in the upper half-plane to vectors in the lower half-plane.

Let us see a couple of examples. Consider the set  $\mathbb{R}^3$  of ordered triples of real numbers:

$$\mathbb{R}^3 = \{(v_1, v_2, v_3) \mid v_i \in \mathbb{R} \text{ for } i = 1, 2, 3\} ,$$

and consider the following subsets

- $\mathbb{W}_1 = \{(v_1, v_2, 0) \mid v_i \in \mathbb{R} \text{ for } i = 1, 2\} \subset \mathbb{R}^3$ ,
- $\mathbb{W}_2 = \{(v_1, v_2, v_3) \mid v_i \in \mathbb{R} \text{ for } i = 1, 2, 3 \text{ and } v_3 \geq 0\} \subset \mathbb{R}^3$ , and
- $\mathbb{W}_3 = \{(v_1, v_2, 1) \mid v_i \in \mathbb{R} \text{ for } i = 1, 2\} \subset \mathbb{R}^3$ .

I will leave it to you as an exercise to show that  $\mathbb{W}_1$  obeys both S1 and S2 whence it is a vector subspace of  $\mathbb{R}^3$ , whereas  $\mathbb{W}_2$  does not obey S2, and  $\mathbb{W}_3$  does not obey either one. Can you think of a subset of  $\mathbb{R}^3$  which obeys S2 but not S1?

### 1.1.5 Linear independence

In this section we will introduce the concepts of linear independence and basis for a vector space; but before doing so we must introduce some preliminary notation.

Let  $\mathbb{V}$  be a vector space,  $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N$  nonzero vectors in  $\mathbb{V}$ , and  $\lambda_1, \lambda_2, \dots, \lambda_N$  scalars, i.e., real numbers. Then the vector in  $\mathbb{V}$  given by

$$\sum_{i=1}^N \lambda_i \mathbf{v}_i := \lambda_1 \mathbf{v}_1 + \lambda_2 \mathbf{v}_2 + \dots + \lambda_N \mathbf{v}_N ,$$

is called a **linear combination** of the  $\{\mathbf{v}_i\}$ . The set  $\mathbb{W}$  of *all* possible linear combinations of the  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$  is actually a vector subspace of  $\mathbb{V}$ , called the **linear span** of the  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$  or the vector subspace **spanned** by the  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$ .



Recall that in order to show that a subset of a vector space is a vector subspace it is necessary and sufficient to show that it is closed under vector addition and under scalar multiplication. Let us check this for the subset  $\mathbb{W}$  of all linear combinations of the  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$ . Let  $\mathbf{w}_1 = \sum_{i=1}^N \alpha_i \mathbf{v}_i$  and  $\mathbf{w}_2 = \sum_{i=1}^N \beta_i \mathbf{v}_i$  be any two elements of  $\mathbb{W}$ . Then

$$\begin{aligned} \mathbf{w}_1 + \mathbf{w}_2 &= \sum_{i=1}^N \alpha_i \mathbf{v}_i + \sum_{i=1}^N \beta_i \mathbf{v}_i \\ &= \sum_{i=1}^N (\alpha_i \mathbf{v}_i + \beta_i \mathbf{v}_i) && \text{(by V2)} \\ &= \sum_{i=1}^N (\alpha_i + \beta_i) \mathbf{v}_i , && \text{(by V7)} \end{aligned}$$

which is clearly in  $\mathbb{W}$ , being again a linear combination of the  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$ . Also, if  $\lambda$

is any real number and  $\mathbf{w} = \sum_{i=1}^N \alpha_i \mathbf{v}_i$  is any vector in  $\mathbb{W}$ ,

$$\begin{aligned} \lambda \mathbf{w} &= \lambda \sum_{i=1}^N \alpha_i \mathbf{v}_i \\ &= \sum_{i=1}^N \lambda (\alpha_i \mathbf{v}_i) && \text{(by V8)} \\ &= \sum_{i=1}^N (\lambda \alpha_i) \mathbf{v}_i, && \text{(by V5)} \end{aligned}$$

which is again in  $\mathbb{W}$ .

A set  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$  of nonzero vectors is said to be **linearly independent** if the equation

$$\sum_{i=1}^N \lambda_i \mathbf{v}_i = \mathbf{0}$$

has only the trivial solution  $\lambda_i = 0$  for all  $i = 1, 2, \dots, N$ . Otherwise the  $\{\mathbf{v}_i\}$  are said to be **linearly dependent**.

It is easy to see that if a set  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$  of nonzero vectors is linearly dependent, then one of the vectors, say,  $\mathbf{v}_i$ , can be written as a linear combination of the remaining  $N-1$  vectors. Indeed, suppose that  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_N\}$  is linearly dependent. This means that the equation

$$\sum_{i=1}^N \lambda_i \mathbf{v}_i = \mathbf{0} \tag{1.1}$$

must have a nontrivial solution where at least one of the  $\{\lambda_i\}$  is different from zero. Suppose, for definiteness, that it is  $\lambda_1$ . Because  $\lambda_1 \neq 0$ , we can divide equation (1.1) by  $\lambda_1$  to obtain:

$$\mathbf{v}_1 + \sum_{i=2}^N \frac{\lambda_i}{\lambda_1} \mathbf{v}_i = \mathbf{0},$$

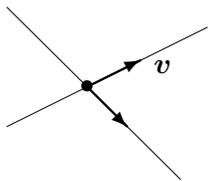
whence

$$\mathbf{v}_1 = -\frac{\lambda_2}{\lambda_1} \mathbf{v}_2 - \frac{\lambda_3}{\lambda_1} \mathbf{v}_3 - \dots - \frac{\lambda_N}{\lambda_1} \mathbf{v}_N.$$

In other words,  $\mathbf{v}_1$  is a linear combination of the  $\{\mathbf{v}_2, \dots, \mathbf{v}_N\}$ . In general and in the same way, if  $\lambda_i \neq 0$  then  $\mathbf{v}_i$  is a linear combination of  $\{\mathbf{v}_1, \dots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \dots, \mathbf{v}_N\}$ .

Let us try to understand these definitions by working through some examples.

We start, as usual, with displacements in the plane. Every nonzero displacement defines a line through the origin. We say that two displacements are **collinear** if they define the same line. In other words,  $\mathbf{u}$  and  $\mathbf{v}$  are collinear if and only if  $\mathbf{u} = \lambda \mathbf{v}$  for some  $\lambda \in \mathbb{R}$ . Clearly, any two displacements in the plane are linearly independent provided they are not collinear, as in the figure.



Now consider  $\mathbb{R}^2$  and let  $(u_1, u_2)$  and  $(v_1, v_2)$  be two nonzero vectors. When will they be linearly independent? From the definition, this will happen provided that the equation

$$\lambda_1 (u_1, u_2) + \lambda_2 (v_1, v_2) = (0, 0)$$

has no other solutions but  $\lambda_1 = \lambda_2 = 0$ . This is a system of linear homogeneous equations for the  $\{\lambda_i\}$ :

$$\begin{aligned} u_1 \lambda_1 + v_1 \lambda_2 &= 0 \\ u_2 \lambda_1 + v_2 \lambda_2 &= 0 . \end{aligned}$$

What must happen for this system to have a nontrivial solution? It will turn out that the answer is that  $u_1 v_2 = u_2 v_1$ . We can see this as follows. Multiply the top equation by  $u_2$  and the bottom equation by  $u_1$  and subtract to get

$$(u_1 v_2 - u_2 v_1) \lambda_2 = 0 ,$$

whence either  $u_1 v_2 = u_2 v_1$  or  $\lambda_2 = 0$ . Now multiply the top equation by  $v_2$  and the bottom equation by  $v_1$  and subtract to get

$$(u_1 v_2 - u_2 v_1) \lambda_1 = 0 ,$$

whence either  $u_1 v_2 = u_2 v_1$  or  $\lambda_1 = 0$ . Since a nontrivial solution must have at least one of  $\lambda_1$  or  $\lambda_2$  nonzero, we are forced to have  $u_1 v_2 = u_2 v_1$ .

### 1.1.6 Bases

Let  $\mathbb{V}$  be a vector space. A set  $\{\mathbf{e}_1, \mathbf{e}_2, \dots\}$  of nonzero vectors is said to be a **basis** for  $\mathbb{V}$  if the following two axioms are satisfied:

- B1 The vectors  $\{\mathbf{e}_1, \mathbf{e}_2, \dots\}$  are linearly independent; and
- B2 The linear span of the  $\{b\mathbf{e}_1, \mathbf{e}_2, \dots\}$  is all of  $\mathbb{V}$ ; in other words, any  $\mathbf{v}$  in  $\mathbb{V}$  can be written as a linear combination of the  $\{\mathbf{e}_1, \mathbf{e}_2, \dots\}$ .

The vectors  $\mathbf{e}_i$  in a basis are known as the **basis elements**.

There are two basic facts about bases which we mention without proof. First of all, *every vector space has a basis*, and in fact, unless it is the trivial vector space consisting only of  $\mathbf{0}$ , it has infinitely many bases. However not every vector space has a *finite* basis; that is, a basis with a finite number of elements. If a vector space *does* possess a finite basis  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$  then it is said to be **finite-dimensional**. Otherwise it is said to be **infinite-dimensional**. We will deal mostly with finite-dimensional vector spaces in this part of the course, although we will have the chance of meeting some infinite-dimensional vector spaces later on.

The second basic fact is that if  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$  and  $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_M\}$  are two bases for a vector space  $\mathbb{V}$ , then  $M = N$ . In other words, every basis has the same number of elements, which is therefore an intrinsic property of the vector space in question. This number is called the **dimension** of the vector space. One says that  $\mathbb{V}$  has dimension  $N$  or that it is  $N$ -dimensional. In symbols, one writes this as  $\dim \mathbb{V} = N$ .

From what we have said before, any two displacements which are non-collinear provide a basis for the displacements on the plane. Therefore this vector space is two-dimensional.

Similarly, any  $(v_1, v_2)$  in  $\mathbb{R}^2$  can be written as a linear combination of  $\{(1, 0), (0, 1)\}$ :

$$(v_1, v_2) = v_1(1, 0) + v_2(0, 1) .$$

Therefore since  $\{(1, 0), (0, 1)\}$  are linearly independent, they form a basis for  $\mathbb{R}^2$ . This shows that  $\mathbb{R}^2$  is also two-dimensional.

More generally for  $\mathbb{R}^N$ , the set given by the  $N$  vectors

$$\{(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)\}$$

is a basis for  $\mathbb{R}^N$ , called the **canonical basis**. This shows that  $\mathbb{R}^N$  has dimension  $N$ .

Let  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_p\}$  be a set of  $p$  linearly independent vectors in a vector space  $\mathbb{V}$  of dimension  $N \geq p$ . Then they are a basis for the vector subspace  $\mathbb{W}$  of  $\mathbb{V}$  which they span. If  $p = N$  they span the full space  $\mathbb{V}$ , whence they are a basis for  $\mathbb{V}$ . It is another basic fact that any set of linearly independent vectors can be completed to a basis.

One final remark: the property B2 satisfied by a basis guarantees that any vector  $\mathbf{v}$  can be written as a linear combination of the basis elements, but does not say whether this can be done in more than one way. In fact, the linear combination turns out to be unique.



Let us prove this. For simplicity, let us work with a finite-dimensional vector space  $\mathbb{V}$  with a basis  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$ . Suppose that a vector  $\mathbf{v} \in \mathbb{V}$  can be written as a linear

combination of the  $\{\mathbf{e}_i\}$  in two ways:

$$\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i \quad \text{and} \quad \mathbf{v} = \sum_{i=1}^N v'_i \mathbf{e}_i .$$

We will show that  $v_i = v'_i$  for all  $i$ . To see this consider

$$\begin{aligned} \mathbf{0} &= \mathbf{v} - \mathbf{v} \\ &= \sum_{i=1}^N v_i \mathbf{e}_i - \sum_{i=1}^N v'_i \mathbf{e}_i \\ &= \sum_{i=1}^N (v_i - v'_i) \mathbf{e}_i . \end{aligned}$$

But because of **B1**, the  $\{\mathbf{e}_i\}$  are linearly independent, and by definition this means that the last of the above equations admits only the trivial solution  $v_i - v'_i = 0$  for all  $i$ . The numbers  $\{v_i\}$  are called the **components** of  $\mathbf{v}$  relative to the basis  $\{\mathbf{e}_i\}$ .

Bases can be extremely useful in calculations with vector spaces. A clever choice of basis can help tremendously towards the solution of a problem, just like a bad choice of basis can make the problem seem very complicated. We will see more of them later, but first we need to introduce the second main concept of linear algebra, that of a linear map.

## 1.2 Linear maps

In the previous section we have learned about vector spaces by studying objects (subspaces, bases,...) living in a fixed vector space. In this section we will look at objects which relate different vector spaces. These objects are called linear maps.

### 1.2.1 Linear maps

Let  $\mathbb{V}$  and  $\mathbb{W}$  be two vector spaces, and consider a map  $A : \mathbb{V} \rightarrow \mathbb{W}$  assigning to each vector  $\mathbf{v}$  in  $\mathbb{V}$  a unique vector  $A(\mathbf{v})$  in  $\mathbb{W}$ . We say that  $A$  is a **linear map** (or a **homomorphism**) if it satisfies the following two properties:

- L1 For all  $\mathbf{v}_1$  and  $\mathbf{v}_2$  in  $\mathbb{V}$ ,  $A(\mathbf{v}_1 + \mathbf{v}_2) = A(\mathbf{v}_1) + A(\mathbf{v}_2)$ ; and
- L2 For all  $\mathbf{v}$  in  $\mathbb{V}$  and  $\lambda \in \mathbb{R}$ ,  $A(\lambda \mathbf{v}) = \lambda A(\mathbf{v})$ .

In other words, a linear map is compatible with the operations of vector addition and scalar multiplication which define the vector space; that is, it does not matter whether we apply the map  $A$  before or after performing these operations: we will get the same result. One says that ‘linear maps respect addition and scalar multiplication.’

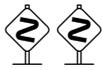
Any linear map  $A : \mathbb{V} \rightarrow \mathbb{W}$  sends the zero vector in  $\mathbb{V}$  to the zero vector in  $\mathbb{W}$ . Let us see this. (We will use the notation  $\mathbf{0}$  *both* for the zero vector in  $\mathbb{V}$  and for the zero vector in  $\mathbb{W}$  as it should be clear from the context which one we mean.) Let  $\mathbf{v}$  be any vector in  $\mathbb{V}$  and let us apply  $A$  to  $\mathbf{0} + \mathbf{v}$ :

$$A(\mathbf{0} + \mathbf{v}) = A(\mathbf{0}) + A(\mathbf{v}) ; \quad (\text{by L1})$$

but because  $\mathbf{0} + \mathbf{v} = \mathbf{v}$ ,

$$A(\mathbf{v}) = A(\mathbf{0}) + A(\mathbf{v}) ,$$

which says that  $A(\mathbf{0}) = \mathbf{0}$ , since the zero vector is unique.



Any linear map  $A : \mathbb{V} \rightarrow \mathbb{W}$  gives rise to a vector subspace of  $\mathbb{V}$ , known as the **kernel** of  $A$ , and written  $\ker A$ . It is defined as the subspace of  $\mathbb{V}$  consisting of those vectors in  $\mathbb{V}$  which get mapped to the zero vector of  $\mathbb{W}$ . In other words,

$$\ker A := \{ \mathbf{v} \in \mathbb{V} \mid A(\mathbf{v}) = \mathbf{0} \in \mathbb{W} \} .$$

To check that  $\ker A \subset \mathbb{W}$  is really a vector subspace, we have to make sure that axioms S1 and S2 are satisfied. Suppose that  $\mathbf{v}_1$  and  $\mathbf{v}_2$  belong to  $\ker A$ . Let us show that so does their sum  $\mathbf{v}_1 + \mathbf{v}_2$ :

$$\begin{aligned} A(\mathbf{v}_1 + \mathbf{v}_2) &= A(\mathbf{v}_1) + A(\mathbf{v}_2) && (\text{by L1}) \\ &= \mathbf{0} + \mathbf{0} && (\text{because } A(\mathbf{v}_i) = \mathbf{0}) \\ &= \mathbf{0} , && (\text{by V3 for } \mathbb{W}) \\ \therefore \mathbf{v}_1 + \mathbf{v}_2 &\in \ker A . \end{aligned}$$

This shows that S1 is satisfied. Similarly, if  $\mathbf{v} \in \ker A$  and  $\lambda \in \mathbb{R}$  is any scalar, then

$$\begin{aligned} A(\lambda \mathbf{v}) &= \lambda A(\mathbf{v}) && (\text{by L2}) \\ &= \lambda \mathbf{0} && (\text{because } A(\mathbf{v}) = \mathbf{0}) \\ &= \mathbf{0} , && (\text{follows from V7 for } \mathbb{W}) \\ \therefore \lambda \mathbf{v} &\in \ker A ; \end{aligned}$$

whence S2 is also satisfied. Notice that we used both properties L1 and L2 of a linear map.

There is also a vector subspace, this time of  $\mathbb{W}$ , associated with  $A : \mathbb{V} \rightarrow \mathbb{W}$ . It is called the **image** of  $A$ , and written  $\text{im } A$ . It consists of those vectors in  $\mathbb{W}$  which can be written as  $A(\mathbf{v})$  for some  $\mathbf{v} \in \mathbb{V}$ . In other words,

$$\text{im } A := \{ \mathbf{w} \in \mathbb{W} \mid \mathbf{w} = A(\mathbf{v}) \text{ for some } \mathbf{v} \in \mathbb{V} \} .$$

To check that  $\text{im } A \subset \mathbb{W}$  is a vector subspace we must check that S1 and S2 are satisfied. Let us do this. Suppose that  $\mathbf{w}_1$  and  $\mathbf{w}_2$  belong to the image of  $A$ . This means that there are vectors  $\mathbf{v}_1$  and  $\mathbf{v}_2$  in  $\mathbb{V}$  which obey  $A(\mathbf{v}_i) = \mathbf{w}_i$  for  $i = 1, 2$ . Therefore,

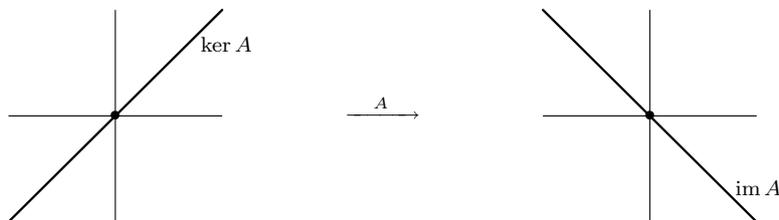
$$\begin{aligned} A(\mathbf{v}_1 + \mathbf{v}_2) &= A(\mathbf{v}_1) + A(\mathbf{v}_2) && (\text{by L1}) \\ &= \mathbf{w}_1 + \mathbf{w}_2 , \end{aligned}$$

whence  $\mathbf{w}_1 + \mathbf{w}_2$  belong to the image of  $A$ . Similarly, if  $\mathbf{w} = A(\mathbf{v})$  belongs to the image of  $A$  and  $\lambda \in \mathbb{R}$  is any scalar,

$$\begin{aligned} A(\lambda \mathbf{v}) &= \lambda A(\mathbf{v}) && (\text{by L2}) \\ &= \lambda \mathbf{w} , \end{aligned}$$

whence  $\lambda \mathbf{w}$  also belongs to the image of  $A$ .

As an example, consider the linear transformation  $A : \mathbb{R}^2 \rightarrow \mathbb{R}^2$  defined by  $(x, y) \mapsto (x - y, y - x)$ . Its kernel and image are pictured below:



A linear map  $A : \mathbb{V} \rightarrow \mathbb{W}$  is said to be **one-to-one** (or **injective** or a **monomorphism**) if  $\ker A = \mathbf{0}$ . The reason for the name is the following. Suppose that  $A(\mathbf{v}_1) = A(\mathbf{v}_2)$ . Then because of linearity,  $A(\mathbf{v}_1 - \mathbf{v}_2) = \mathbf{0}$ , whence  $\mathbf{v}_1 - \mathbf{v}_2$  belongs to the kernel. Since the kernel is zero, we have that  $\mathbf{v}_1 = \mathbf{v}_2$ .

Similarly a linear map  $A : \mathbb{V} \rightarrow \mathbb{W}$  is said to be **onto** (or **surjective** or an **epimorphism**) if  $\text{im } A = \mathbb{W}$ , so that every vector of  $\mathbb{W}$  is the image under  $A$  of some vector in  $\mathbb{V}$ . If this vector is unique, so that  $A$  is also one-to-one, we say that  $A$  is an **isomorphism**. If  $A : \mathbb{V} \rightarrow \mathbb{W}$  is an isomorphism, one says that  $\mathbb{V}$  is **isomorphic** to  $\mathbb{W}$ , and we write this as  $\mathbb{V} \cong \mathbb{W}$ . As we will see below, ‘being isomorphic to’ is an equivalence relation.

Notice that if  $\mathbb{V}$  is an  $N$ -dimensional real vector space, any choice of basis  $\{\mathbf{e}_i\}$  induces an isomorphism  $A : \mathbb{V} \rightarrow \mathbb{R}^N$ , defined by sending the vector  $\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i$  to the ordered  $N$ -tuple made out from its components  $(v_1, v_2, \dots, v_N)$  relative to the basis. Therefore we see that all  $N$ -dimensional vector spaces are isomorphic to  $\mathbb{R}^N$ , and hence to each other.

An important property of linear maps is that once we know how they act on a basis, we know how they act on any vector in the vector space. Indeed, suppose that  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$  is a basis for an  $N$ -dimensional vector space  $\mathbb{V}$ . Any vector  $\mathbf{v} \in \mathbb{V}$  can be written uniquely as a linear combination of the basis elements:

$$\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i .$$

Let  $A : \mathbb{V} \rightarrow \mathbb{W}$  be a linear map. Then

$$\begin{aligned} A(\mathbf{v}) &= A\left(\sum_{i=1}^N v_i \mathbf{e}_i\right) \\ &= \sum_{i=1}^N A(v_i \mathbf{e}_i) && \text{(by L1)} \\ &= \sum_{i=1}^N v_i A(\mathbf{e}_i) . && \text{(by L2)} \end{aligned}$$

Therefore if we know  $A(\mathbf{e}_i)$  for  $i = 1, 2, \dots, N$  we know  $A$  on *any* vector.



The dual space.

## 1.2.2 Composition of linear maps

Linear maps can be composed to produce new linear maps. Let  $A : \mathbb{V} \rightarrow \mathbb{W}$  and  $B : \mathbb{U} \rightarrow \mathbb{V}$  be linear maps connecting three vector spaces  $\mathbb{U}$ ,  $\mathbb{V}$  and  $\mathbb{W}$ . We can define a third map  $C : \mathbb{U} \rightarrow \mathbb{W}$  by composing the two maps:

$$\mathbb{U} \xrightarrow{B} \mathbb{V} \xrightarrow{A} \mathbb{W} .$$

In other words, if  $\mathbf{u} \in \mathbb{U}$  is any vector, then the action of  $C$  on it is defined by first applying  $B$  to get  $B(\mathbf{u})$  and then applying  $A$  to the result to obtain  $A(B(\mathbf{u}))$ . The resulting map is written  $A \circ B$ , so that one has the composition rule:

$$(A \circ B)(\mathbf{u}) := A(B(\mathbf{u})) . \quad (1.2)$$

This new map is linear because  $B$  and  $A$  are, as we now show. It respects addition:

$$\begin{aligned} (A \circ B)(\mathbf{u}_1 + \mathbf{u}_2) &= A(B(\mathbf{u}_1 + \mathbf{u}_2)) \\ &= A(B(\mathbf{u}_1) + B(\mathbf{u}_2)) && \text{(by L1 for } B\text{)} \\ &= A(B(\mathbf{u}_1)) + A(B(\mathbf{u}_2)) && \text{(by L1 for } A\text{)} \\ &= (A \circ B)(\mathbf{u}_1) + (A \circ B)(\mathbf{u}_2) ; \end{aligned}$$

and it also respects scalar multiplication:

$$\begin{aligned} (A \circ B)(\lambda \mathbf{u}) &= A(B(\lambda \mathbf{u})) \\ &= A(\lambda B(\mathbf{u})) && \text{(by L2 for } B\text{)} \\ &= \lambda A(B(\mathbf{u})) && \text{(by L2 for } A\text{)} \\ &= \lambda (A \circ B)(\mathbf{u}) . \end{aligned}$$

Thus  $A \circ B$  is a linear map, known as the **composition** of  $A$  and  $B$ . One usually reads  $A \circ B$  as ‘ $B$  composed with  $A$ ’ (notice the order!) or ‘ $A$  pre-composed with  $B$ .’

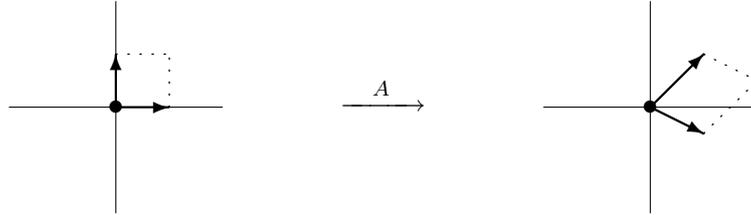


Notice that if  $A$  and  $B$  are isomorphisms, then so is  $A \circ B$ . In other words, composition of isomorphisms is an isomorphism. This means that if  $\mathbb{U} \cong \mathbb{V}$  and  $\mathbb{V} \cong \mathbb{W}$ , then  $\mathbb{U} \cong \mathbb{W}$ , so that the property of being isomorphic is transitive. This property is also symmetric: if  $A : \mathbb{V} \rightarrow \mathbb{W}$  is an isomorphism,  $A^{-1} : \mathbb{W} \rightarrow \mathbb{V}$  is too, so that  $\mathbb{V} \cong \mathbb{W}$  implies  $\mathbb{W} \cong \mathbb{V}$ . Moreover it is also reflexive, the identity map  $\mathbb{1} : \mathbb{V} \rightarrow \mathbb{V}$  provides an isomorphism  $\mathbb{V} \cong \mathbb{V}$ . Hence the property of being isomorphic is an equivalence relation.

## 1.2.3 Linear transformations

An important special case of linear maps are those which map a vector space to itself:  $A : \mathbb{V} \rightarrow \mathbb{V}$ . These linear maps are called **linear transformations**

(or **endomorphisms**). Linear transformations are very easy to visualise in two dimensions:



A linear transformation sends the origin to the origin, straight lines to straight lines, and parallelograms to parallelograms.

Composition of two linear transformation is another linear transformation. In other words, we can think of composition of linear transformations as some sort of multiplication. This multiplication obeys a property reminiscent of the associativity **V1** of vector addition. Namely, given three linear transformations  $A$ ,  $B$  and  $C$ , then

$$(A \circ B) \circ C = A \circ (B \circ C) . \quad (1.3)$$

To see this simply apply both sides of the equation to  $\mathbf{v} \in \mathbb{V}$  and use equation (1.2) to obtain in both cases simply  $A(B(C(\mathbf{v})))$ . By analogy, we say that composition of linear transformations is **associative**. Unlike vector addition, composition is *not* commutative; that is, in general,  $A \circ B \neq B \circ A$ .

Let  $\mathbb{1} : \mathbb{V} \rightarrow \mathbb{V}$  denote the **identity transformation**, defined by  $\mathbb{1}(\mathbf{v}) = \mathbf{v}$  for all  $\mathbf{v} \in \mathbb{V}$ . Clearly,

$$\mathbb{1} \circ A = A \circ \mathbb{1} = A , \quad (1.4)$$

for any linear transformations  $A$ . In other words,  $\mathbb{1}$  is an **identity** for the composition of linear transformations. Given a linear transformation  $A : \mathbb{V} \rightarrow \mathbb{V}$ , it may happen that there is a linear transformation  $B : \mathbb{V} \rightarrow \mathbb{V}$  such that

$$B \circ A = A \circ B = \mathbb{1} . \quad (1.5)$$

If this is the case, we say that  $A$  is **invertible**, and we call  $B$  its **inverse**. We then write  $B = A^{-1}$ .

The composition of two invertible linear transformations is again invertible. Indeed one has

$$(A \circ B)^{-1} = B^{-1} \circ A^{-1} .$$

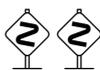


To show this we compute

$$\begin{aligned} B^{-1} \circ A^{-1} \circ (A \circ B) &= B^{-1} \circ A^{-1} \circ (A \circ B) && \text{(by equation (1.3))} \\ &= B^{-1} \circ A^{-1} \circ A \circ B && \text{(by equation (1.3))} \\ &= B^{-1} \circ (\mathbb{1} \circ B) && \text{(by equation (1.5))} \\ &= B^{-1} \circ B && \text{(by equation (1.4))} \\ &= \mathbb{1} , && \text{(by equation (1.5))} \end{aligned}$$

and similarly

$$\begin{aligned}
 (A \circ B) \circ (B^{-1} \circ A^{-1}) &= A \circ B \circ B^{-1} \circ A^{-1} && \text{(by equation (1.3))} \\
 &= A \circ \mathbf{1} \circ A^{-1} && \text{(by equation (1.3))} \\
 &= A \circ \mathbf{1} \circ A^{-1} && \text{(by equation (1.5))} \\
 &= A \circ A^{-1} && \text{(by equation (1.4))} \\
 &= \mathbf{1} . && \text{(by equation (1.5))}
 \end{aligned}$$



This shows that the invertible transformations of a vector space  $\mathbb{V}$  form a **group**, called the **general linear group** of  $\mathbb{V}$  and written  $GL(\mathbb{V})$ .

A group is a set  $G$  whose elements are called **group elements**, together with an operation called **group multiplication** and written simply as

$$\begin{aligned}
 \text{group multiplication} : G \times G &\rightarrow G \\
 (x, y) &\mapsto xy
 \end{aligned}$$

satisfying the following three axioms:

G1 group multiplication is **associative**:

$$(xy)z = x(yz) \quad \text{for all group elements } x, y \text{ and } z.$$

G2 there exists an **identity element**  $e \in G$  such that

$$ex = xe = x \quad \text{for all group elements } x.$$

G3 every group element  $x$  has an **inverse**, denoted  $x^{-1}$  and obeying

$$x^{-1}x = xx^{-1} = e .$$

If in addition the group obeys a fourth axiom

G4 group multiplication is **commutative**:

$$xy = yx \quad \text{for all group elements } x \text{ and } y,$$

then we say that the group is **commutative** or **abelian**, in honour of the Norwegian mathematician Niels Henrik Abel (1802-1829).

When the group is abelian, the group multiplication is usually written as a group addition:  $x + y$  instead of  $xy$ . Notice that axioms V1—V4 for a vector space say that, under vector addition, a vector space is an abelian group.

Groups are extremely important objects in both mathematics and physics. It is an ‘algebraic’ concept, yet its uses transcend algebra; for example, it was using the theory of groups that quarks were originally postulated in particle physics. The fact that we now think of quarks as elementary particles and not simply as mathematical construct is proof of how far group theory has become a part of our description of nature at its most fundamental.

## 1.2.4 The vector space of linear maps

Now we point out that linear maps themselves also form a vector space! In order to do this, we have to produce the two operations: vector addition and scalar multiplication, and show that they satisfy the eight axioms **V1—V8**. Let  $A$  and  $B$  be linear maps  $\mathbb{V} \rightarrow \mathbb{W}$ , let  $\lambda \in \mathbb{R}$  be a scalar, and let  $\mathbf{v} \in \mathbb{V}$  be any vector. Then we define the two operations by

$$\begin{aligned} \text{(addition)} \\ (A + B)(\mathbf{v}) = A(\mathbf{v}) + B(\mathbf{v}) , \end{aligned} \tag{1.6}$$

$$\begin{aligned} \text{(scalar multiplication)} \\ (\lambda A)(\mathbf{v}) = \lambda A(\mathbf{v}) . \end{aligned} \tag{1.7}$$

Having defined these two operations we must check that the axioms are satisfied. We leave this as an exercise, except to note that the zero vector is the transformation which sends every  $\mathbf{v} \in \mathbb{V}$  to  $\mathbf{0} \in \mathbb{W}$ . The rest of the axioms follow from the fact that  $\mathbb{W}$  is a vector space.



This is a general mathematical fact: the space of functions  $f : X \rightarrow Y$  always inherits whatever algebraic structures  $Y$  possesses simply by defining the operations pointwise in  $X$ .

Let  $\mathcal{L}(\mathbb{V}, \mathbb{W})$  denote the vector space of linear maps  $\mathbb{V} \rightarrow \mathbb{W}$ . What is its dimension? We will see in the next section when we talk about matrices that its dimension is given by the product of the dimensions of  $\mathbb{V}$  and  $\mathbb{W}$ :

$$\dim \mathcal{L}(\mathbb{V}, \mathbb{W}) = \dim \mathbb{V} \dim \mathbb{W} . \tag{1.8}$$

In particular the space  $\mathcal{L}(\mathbb{V}, \mathbb{V})$  of linear transformations of  $\mathbb{V}$  has dimension  $(\dim \mathbb{V})^2$ . We will call this space  $\mathcal{L}(\mathbb{V})$  from now on.

Because  $\mathcal{L}(\mathbb{V})$  is a vector space, its elements can be added and as we saw above, composition allows us to multiply them too. It turns out that these two operations are compatible:

$$A \circ (B + C) = (A \circ B) + (A \circ C) \tag{1.9}$$

$$(A + B) \circ C = (A \circ C) + (B \circ C) . \tag{1.10}$$



Let us prove the left and right distributivity properties. Let  $A$ ,  $B$ , and  $C$  be linear transformations of a vector space  $\mathbb{V}$  and let  $\mathbf{v} \in \mathbb{V}$  be an arbitrary vector. Then

$$\begin{aligned} (A \circ (B + C))(\mathbf{v}) &= A((B + C)(\mathbf{v})) && \text{(by equation (1.2))} \\ &= A(B(\mathbf{v}) + C(\mathbf{v})) && \text{(by equation (1.6))} \\ &= A(B(\mathbf{v})) + A(C(\mathbf{v})) && \text{(because } A \text{ is linear)} \\ &= (A \circ B)(\mathbf{v}) + (A \circ C)(\mathbf{v}) , && \text{(by equation (1.2))} \end{aligned}$$

which proves (1.9). Similarly

$$\begin{aligned} ((A + B) \circ C)(\mathbf{v}) &= (A + B)(C(\mathbf{v})) && \text{(by equation (1.2))} \\ &= A(C(\mathbf{v})) + B(C(\mathbf{v})) && \text{(by equation (1.6))} \\ &= (A \circ C)(\mathbf{v}) + (B \circ C)(\mathbf{v}), && \text{(by equation (1.2))} \end{aligned}$$

which proves (1.10).

Composition of linear transformations is also compatible with scalar multiplication:

$$(\lambda A) \circ B = A \circ (\lambda B) = \lambda (A \circ B). \quad (1.11)$$



In fact, we can summarise the properties (1.9), (1.10) and (1.11) in a very simple way using concepts we have already introduced. Given a linear transformation  $A$  of  $\mathbb{V}$  we will define two operations on  $\mathcal{L}(\mathbb{V})$ , left and right multiplication by  $A$ , as follows:

$$\begin{aligned} L_A : \mathcal{L}(\mathbb{V}) &\rightarrow \mathcal{L}(\mathbb{V}) && \text{and} && R_A : \mathcal{L}(\mathbb{V}) &\rightarrow \mathcal{L}(\mathbb{V}) \\ B &\mapsto A \circ B && && B &\mapsto B \circ A. \end{aligned}$$

Then equations (1.9), (1.10) and (1.11) simply say that  $L_A$  and  $R_A$  are linear transformations of  $\mathcal{L}(\mathbb{V})$ !



The vector space  $\mathcal{L}(\mathbb{V})$  of linear transformations of  $\mathbb{V}$  together with the operation of composition, the identity  $\mathbb{1}$ , the distributive properties (1.9) and (1.10), and the condition (1.11) is an **associative algebra with identity**.

An **algebra** is a vector space  $\mathbb{A}$  together with a multiplication

$$\begin{aligned} \text{multiplication} : \mathbb{A} \times \mathbb{A} &\rightarrow \mathbb{A} \\ (A, B) &\mapsto AB, \end{aligned}$$

obeying the following axioms, where  $A, B, C \in \mathbb{A}$  and  $\lambda \in \mathbb{R}$ :

- A1 (left distributivity)  $A(B + C) = AB + AC$ ;
- A2 (right distributivity)  $(A + B)C = AC + BC$ ;
- A3  $A(\lambda B) = (\lambda A)B = \lambda(AB)$ .

If in addition  $\mathbb{A}$  obeys the axiom

- A4 (identity) There exists  $\mathbb{1} \in \mathbb{A}$  such that  $\mathbb{1}A = A\mathbb{1} = A$ ;

then it is an **algebra with identity**. If instead  $\mathbb{A}$  obeys the axiom

- A5 (associativity)  $A(BC) = (AB)C$ ;

it is an **associative algebra**. Finally if it obeys all five axioms, it is an **associative algebra with identity**.

It is a general fact that the invertible elements of an associative algebra with identity form a group.

## 1.2.5 Matrices

Matrices are intimately linked to linear maps. Let  $A : \mathbb{V} \rightarrow \mathbb{W}$  be a linear map between two *finite-dimensional* vector spaces. Let  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$  be a basis for  $\mathbb{V}$  and let  $\{\mathbf{f}_1, \mathbf{f}_2, \dots, \mathbf{f}_M\}$  be a basis for  $\mathbb{W}$ . Let us write each  $A(\mathbf{e}_i)$  as a linear combination of the basis elements  $\{\mathbf{f}_j\}$ :

$$A(\mathbf{e}_i) = \sum_{j=1}^M A_{ji} \mathbf{f}_j, \quad (1.12)$$

where we have introduced a real number  $A_{ji}$  for each  $i = 1, 2, \dots, N$  and  $j = 1, 2, \dots, M$ , a total of  $NM$  real numbers. Now let  $\mathbf{v}$  be a vector in  $\mathbb{V}$  and consider its image  $\mathbf{w} = A(\mathbf{v})$  under  $A$ . We can expand both  $\mathbf{v}$  and  $\mathbf{w}$  as linear combinations of the respective bases:

$$\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i \quad \text{and} \quad \mathbf{w} = \sum_{j=1}^M w_j \mathbf{f}_j. \quad (1.13)$$

Let us now express the  $w_j$  in terms of the  $v_i$ :

$$\begin{aligned} \mathbf{w} &= A(\mathbf{v}) \\ &= A\left(\sum_{i=1}^N v_i \mathbf{e}_i\right) && \text{(by the first equation in (1.13))} \\ &= \sum_{i=1}^N A(v_i \mathbf{e}_i) && \text{(by L1)} \\ &= \sum_{i=1}^N v_i A(\mathbf{e}_i) && \text{(by L2)} \\ &= \sum_{i=1}^N v_i \sum_{j=1}^M A_{ji} \mathbf{f}_j && \text{(by equation (1.12))} \\ &= \sum_{j=1}^M \left(\sum_{i=1}^N A_{ji} v_i\right) \mathbf{f}_j, && \text{(rearranging the sums)} \end{aligned}$$

whence comparing with the second equation in (1.13) we obtain the desired result:

$$w_j = \sum_{i=1}^N A_{ji} v_i. \quad (1.14)$$

To visualise this equation, let us arrange the components  $\{v_i\}$  and  $\{w_j\}$  of  $\mathbf{v}$  and  $\mathbf{w}$  as ‘column vectors’  $\mathbf{v}$  and  $\mathbf{w}$ , and the real numbers  $A_{ji}$  as an  $M \times N$  matrix  $\mathbf{A}$ . Then equation (1.14) can be written as

$$\mathbf{w} = \mathbf{A} \mathbf{v} ,$$

or explicitly as

$$\begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_M \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} .$$

Therefore the matrix

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{pmatrix}$$

represents the linear map  $A : \mathbb{V} \rightarrow \mathbb{W}$  relative to the bases  $\{\mathbf{e}_i\}$  and  $\{\mathbf{f}_j\}$  of  $\mathbb{V}$  and  $\mathbb{W}$ . It is important to stress that the linear map  $A$  is more fundamental than the matrix  $\mathbf{A}$ . If we choose different basis, the matrix for the linear map will change (we will see this in detail below), but the map itself does not. However if we fix bases for  $\mathbb{V}$  and  $\mathbb{W}$ , then there is a one-to-one correspondence between linear maps  $\mathbb{V} \rightarrow \mathbb{W}$  and  $M \times N$  matrices.



The commuting square: linear maps to matrices...

We saw in Section 1.2.4 that the space  $\mathcal{L}(\mathbb{V}, \mathbb{W})$  of linear maps  $\mathbb{V} \rightarrow \mathbb{W}$  is a vector space in its own right. How are the operations of vector addition and scalar multiplication defined for the matrices? It turns out that they are defined entry-wise as for real numbers. Let us see this. The matrix corresponding to the sum of two linear maps  $A$  and  $A'$  is given by

$$(A + A')(\mathbf{e}_i) = \sum_{j=1}^M (A + A')_{ji} \mathbf{f}_j .$$

On the other hand, from equation (1.6) we have that

$$\begin{aligned} (A + A')(\mathbf{e}_i) &= A(\mathbf{e}_i) + A'(\mathbf{e}_i) \\ &= \sum_{j=1}^M A_{ji} \mathbf{f}_j + \sum_{j=1}^M A'_{ji} \mathbf{f}_j \\ &= \sum_{j=1}^M (A_{ji} + A'_{ji}) \mathbf{f}_j . \end{aligned}$$

Therefore we see that the matrix of the sum is the sum of the matrices:

$$(A + A')_{ji} = A_{ji} + A'_{ji} ;$$

or in other words, the sum of two matrices is performed entry-by-entry:

$$\begin{aligned} & \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{pmatrix} + \begin{pmatrix} A'_{11} & A'_{12} & \cdots & A'_{1N} \\ A'_{21} & A'_{22} & \cdots & A'_{2N} \\ \vdots & \vdots & & \vdots \\ A'_{M1} & A'_{M2} & \cdots & A'_{MN} \end{pmatrix} \\ &= \begin{pmatrix} A_{11} + A'_{11} & A_{12} + A'_{12} & \cdots & A_{1N} + A'_{1N} \\ A_{21} + A'_{21} & A_{22} + A'_{22} & \cdots & A_{2N} + A'_{2N} \\ \vdots & \vdots & & \vdots \\ A_{M1} + A'_{M1} & A_{M2} + A'_{M2} & \cdots & A_{MN} + A'_{MN} \end{pmatrix} . \end{aligned}$$

Similarly, scalar multiplication is also performed entry-by-entry. If  $\lambda \in \mathbb{R}$  is a scalar and  $A$  is a linear map, then on the one hand we have

$$(\lambda A)(\mathbf{e}_i) = \sum_{j=1}^M (\lambda A)_{ji} \mathbf{f}_j ,$$

but from equation (1.7) we have that

$$\begin{aligned} (\lambda A)(\mathbf{e}_i) &= \lambda A(\mathbf{e}_i) \\ &= \lambda \sum_{j=1}^M A_{ji} \mathbf{f}_j \\ &= \sum_{j=1}^M \lambda A_{ji} \mathbf{f}_j , \end{aligned}$$

so that the matrix of  $\lambda A$  is obtained from the matrix of  $A$  by multiplying each entry by  $\lambda$ :

$$(\lambda A)_{ji} = \lambda A_{ji} ;$$

explicitly,

$$\lambda \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{pmatrix} = \begin{pmatrix} \lambda A_{11} & \lambda A_{12} & \cdots & \lambda A_{1N} \\ \lambda A_{21} & \lambda A_{22} & \cdots & \lambda A_{2N} \\ \vdots & \vdots & & \vdots \\ \lambda A_{M1} & \lambda A_{M2} & \cdots & \lambda A_{MN} \end{pmatrix} .$$

The vector space of  $M \times N$  matrices has a ‘canonical’ basis given by the matrices  $\mathbf{E}_{ji}$  all of whose entries are zero except for the entry sitting in the intersection of the  $j$ th column and the  $i$ th row, which is 1. They are clearly linearly independent and if  $\mathbf{A}$  is any matrix with entries  $A_{ji}$  then

$$\mathbf{A} = \sum_{i=1}^N \sum_{j=1}^M A_{ji} \mathbf{E}_{ji} ,$$

so that their span is the space of all  $M \times N$  matrices. Therefore they form a basis for this space. The matrices  $\mathbf{E}_{ji}$  are known as **elementary matrices**. Clearly there are  $MN$  such matrices, whence the dimension of the space of  $M \times N$  matrices, and hence of  $\mathcal{L}(\mathbb{V}, \mathbb{W})$ , is  $MN$  as claimed in equation (1.8).

Now consider a third vector space  $\mathbb{U}$  of dimension  $P$  and with basis  $\{\mathbf{g}_1, \mathbf{g}_2, \dots, \mathbf{g}_P\}$ . Then a linear map  $B : \mathbb{U} \rightarrow \mathbb{V}$  will be represented by an  $N \times P$  matrix

$$\mathbf{B} = \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1P} \\ B_{21} & B_{22} & \cdots & B_{2P} \\ \vdots & \vdots & & \vdots \\ B_{N1} & B_{N2} & \cdots & B_{NP} \end{pmatrix}$$

relative to the chosen bases for  $\mathbb{U}$  and  $\mathbb{V}$ ; that is,

$$B(\mathbf{g}_k) = \sum_{i=1}^N B_{ik} \mathbf{e}_i . \tag{1.15}$$

The composition  $A \circ B : \mathbb{U} \rightarrow \mathbb{W}$  will now be represented by an  $M \times P$  matrix whose entries  $C_{jk}$  are given by

$$(A \circ B)(\mathbf{g}_k) = \sum_{j=1}^M C_{jk} \mathbf{f}_j . \tag{1.16}$$

The matrix of  $A \circ B$  can be expressed in terms of the matrices  $\mathbf{A}$  and  $\mathbf{B}$ . To

see this, let us compute

$$\begin{aligned}
(A \circ B)(\mathbf{g}_k) &= A(B(\mathbf{g}_k)) && \text{(by equation (1.2))} \\
&= A\left(\sum_{i=1}^N B_{ik} \mathbf{e}_i\right) && \text{(by equation (1.15))} \\
&= \sum_{i=1}^N B_{ik} A(\mathbf{e}_i) && \text{(since } A \text{ is linear)} \\
&= \sum_{i=1}^N B_{ik} \sum_{j=1}^M A_{ji} \mathbf{f}_j && \text{(by equation (1.12))} \\
&= \sum_{j=1}^M \left(\sum_{i=1}^N A_{ji} B_{ik}\right) \mathbf{f}_j . && \text{(rearranging sums)}
\end{aligned}$$

Therefore comparing with equation (1.16) we see that

$$C_{jk} = \sum_{i=1}^N A_{ji} B_{ik} , \quad (1.17)$$

which is nothing else but matrix multiplication:

$$\begin{aligned}
&\begin{pmatrix} C_{11} & C_{12} & \cdots & C_{1P} \\ C_{21} & C_{22} & \cdots & C_{2P} \\ \vdots & \vdots & & \vdots \\ C_{M1} & C_{M2} & \cdots & C_{MP} \end{pmatrix} \\
&= \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & & \vdots \\ A_{M1} & A_{M2} & \cdots & A_{MN} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} & \cdots & B_{1P} \\ B_{21} & B_{22} & \cdots & B_{2P} \\ \vdots & \vdots & & \vdots \\ B_{N1} & B_{N2} & \cdots & B_{NP} \end{pmatrix} .
\end{aligned}$$

In other words,

the matrix of  $A \circ B$  is the matrix product  $AB$ .

Let us consider now linear transformations  $\mathcal{L}(\mathbb{V})$  of an  $N$ -dimensional vector space  $\mathbb{V}$  with basis  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_N\}$ . Matrices representing linear transformations  $\mathbb{V} \rightarrow \mathbb{V}$  are now a square  $N \times N$  matrices. We can add them and multiply them as we do real numbers, except that multiplication is not

commutative: for two matrices  $A$  and  $B$  one has that, in general,  $AB \neq BA$ . Let  $A$  be an  $N \times N$  matrix. If there exists another matrix  $B$  which obeys

$$AB = BA = I$$

where  $I$  is the identity matrix, then we say that  $A$  is **invertible**. Its **inverse**  $B$  is written  $A^{-1}$ . A matrix which is not invertible is called **singular**. Clearly a matrix is singular if and only if its determinant is zero.

A useful fact is that a matrix is invertible if and only if its determinant is different from zero. This allows us to show that the product of invertible elements is again invertible. To see this notice that the determinant of a product is the product of the determinants:

$$\boxed{\det(AB) = \det A \det B,} \quad (1.18)$$

and that this is not zero because neither are  $\det A$  nor  $\det B$ . In fact, the inverse of a product  $AB$  is given by

$$\boxed{(AB)^{-1} = B^{-1}A^{-1}.} \quad (1.19)$$

(Notice the order!)



Matrices, just like  $\mathcal{L}(V)$ , form an associative algebra with identity. The algebra of  $N \times N$  real matrices is denoted  $\text{Mat}_N(\mathbb{R})$ . The invertible elements form a group, which is denoted  $\text{GL}_N(\mathbb{R})$ , the general linear group of  $\mathbb{R}^N$ .

### 1.2.6 Change of basis

We mentioned above that a linear map is more fundamental than the matrix representing it relative to a chosen basis, for the matrix changes when we change the basis but the linear map remains unchanged. In this Section we will explore how the matrix of a linear map changes as we change the basis. We will restrict ourselves to linear transformations, but the results here extend straightforwardly to linear maps between different vector spaces.

Let  $V$  be an  $N$ -dimensional vector space with basis  $\{e_i\}$ , and let  $A : V \rightarrow V$  be a linear transformation with matrix  $A$  relative to this basis. Let  $\{e'_i\}$  be another basis. We want to know what the matrix  $A'$  representing  $A$  relative to this new basis is. By definition, the matrix  $A'$  has entries  $A'_{ji}$  given by

$$A(e'_i) = \sum_{j=1}^N A'_{ji} e'_j. \quad (1.20)$$

Because  $\{\mathbf{e}_i\}$  is a basis, we can express each element  $\mathbf{e}'_i$  of the primed basis in terms of them:

$$\mathbf{e}'_i = \sum_{j=1}^N S_{ji} \mathbf{e}_j , \quad (1.21)$$

for some  $N^2$  numbers  $S_{ji}$ . We have written this equation in such a way that it looks as if  $S_{ji}$  are the entries of a matrix. This is with good reason. Let  $S : \mathbb{V} \rightarrow \mathbb{V}$  be the linear transformation defined by  $S(\mathbf{e}_i) = \mathbf{e}'_i$  for  $i = 1, 2, \dots, N$ . Then using the explicit expression for  $\mathbf{e}'_i$  we see that

$$S(\mathbf{e}_i) = \sum_{j=1}^N S_{ji} \mathbf{e}_j ,$$

so that  $S_{ji}$  are indeed the entries of a matrix  $\mathbf{S}$  relative to the basis  $\{\mathbf{e}_i\}$ . We can compute both sides of equation (1.20) separately and compare. The right-hand side gives

$$\begin{aligned} A(\mathbf{e}'_i) &= A\left(\sum_{j=1}^N S_{ji} \mathbf{e}_j\right) && \text{(by equation (1.21))} \\ &= \sum_{j=1}^N S_{ji} A(\mathbf{e}_j) && \text{(since } A \text{ is linear)} \\ &= \sum_{j=1}^N S_{ji} \sum_{k=1}^N A_{kj} \mathbf{e}_k && \text{(by equation (1.12))} \\ &= \sum_{k=1}^N \sum_{j=1}^N A_{kj} S_{ji} \mathbf{e}_k . && \text{(rearranging sums)} \end{aligned}$$

On the other hand, the left-hand side gives

$$\begin{aligned} \sum_{j=1}^N A'_{ji} \mathbf{e}'_j &= \sum_{j=1}^N A'_{ji} \sum_{k=1}^N S_{kj} \mathbf{e}_k && \text{(by equation (1.21))} \\ &= \sum_{k=1}^N \sum_{j=1}^N S_{kj} A'_{ji} \mathbf{e}_k . && \text{(rearranging sums)} \end{aligned}$$

Comparing the two sides, we see that

$$\sum_{j=1}^N A_{kj} S_{ji} = \sum_{j=1}^N S_{kj} A'_{ji} ,$$

or in terms of matrices,

$$\mathbf{A} \mathbf{S} = \mathbf{S} \mathbf{A}' . \quad (1.22)$$

Now,  $S$  is invertible. To see this use the fact that because  $\{\mathbf{e}'_i\}$  is also a basis, we can write each  $\mathbf{e}_i$  in terms of the  $\{\mathbf{e}'_i\}$ :

$$\mathbf{e}_i = \sum_{j=1}^N T_{ji} \mathbf{e}'_j . \quad (1.23)$$

By the same argument as above, the  $N^2$  numbers  $T_{ji}$  are the entries of a matrix which, relative to the primed basis, represents the linear transformation  $T : \mathbb{V} \rightarrow \mathbb{V}$  defined by  $T(\mathbf{e}'_i) = \mathbf{e}_i$ . The linear transformations  $S$  and  $T$  are mutual inverses:

$$S(T(\mathbf{e}'_i)) = S(\mathbf{e}_i) = \mathbf{e}'_i \quad \text{and} \quad T(S(\mathbf{e}_i)) = T(\mathbf{e}'_i) = \mathbf{e}_i ,$$

so that  $T \circ S = S \circ T = \mathbb{1}$ ; or in other words,  $T = S^{-1}$ .

Therefore, we can multiply both sides of equation (1.22) by  $S^{-1}$  on the left to obtain

$$\boxed{\mathbf{A}' = \mathbf{S}^{-1} \mathbf{A} \mathbf{S} .} \quad (1.24)$$

The operation above taking  $\mathbf{A}$  to  $\mathbf{A}'$  is called **conjugation** by  $\mathbf{S}$ . One says that the matrices  $\mathbf{A}$  and  $\mathbf{A}'$  are **conjugate**. (This is not to be confused with the notion of *complex conjugation*.)



Change of basis for linear maps.

## 1.2.7 Matrix invariants

Certain properties of square matrices do not change when we change the basis; one says that they are **invariants** of the matrix or, more precisely, of the linear map that the matrix represents.

For example, the determinant is one such invariant. This can be seen by computing the determinant to both sides of equation (1.24) and using equation (1.18), to obtain that  $\det \mathbf{A}' = \det \mathbf{A}$ . This implies that also the property of being invertible is invariant.

Another invariant is the **trace** of a matrix, defined as the sum of the diagonal elements, and written  $\text{tr} \mathbf{A}$ . Explicitly, if  $\mathbf{A}$  is given by

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{pmatrix}$$

then its trace  $\text{tr } A$  is given by

$$\text{tr } A = \sum_{i=1}^N A_{ii} = A_{11} + A_{22} + \cdots + A_{NN} . \quad (1.25)$$

A matrix whose trace vanishes is said to be **traceless**.

The fact that the trace is indeed an invariant, will follow from some fundamental properties of the trace, which we discuss now. The trace satisfies the following property:

$$\boxed{\text{tr } (A B) = \text{tr } (B A) .} \quad (1.26)$$



Let us prove this. Let  $A, B : \mathbb{V} \rightarrow \mathbb{V}$  be linear maps with matrices  $A$  and  $B$  relative to some fixed basis. The matrix product  $AB$  is the matrix of the composition  $A \circ B$ . Computing the trace of the product, using equations (1.17) and (1.25), we find

$$\begin{aligned} \text{tr } (A B) &= \sum_{i=1}^N \sum_{j=1}^N A_{ij} B_{ji} \\ &= \sum_{j=1}^N \sum_{i=1}^N B_{ji} A_{ij} && \text{(rearranging the sums)} \\ &= \sum_{i=1}^N \sum_{j=1}^N B_{ij} A_{ji} && \text{(relabelling the sums)} \\ &= \text{tr } (B A) . \end{aligned}$$

The fact which allows us to relabel the summation indices is known as the **Shakespeare Theorem**: “a dummy index by any other name...” The modern version of this theorem is due to Gertrude Stein: “a dummy index is a dummy index is a dummy index.”

It follows from equation (1.26) that

$$\boxed{\text{tr } (A B C) = \text{tr } (C A B) = \text{tr } (B C A) ,} \quad (1.27)$$

which is often called the **cyclic property** of the trace. Using this property and computing the trace to both sides of equation (1.24), we see that  $\text{tr } A' = \text{tr } A$ , as claimed. Notice that the trace of the identity  $N \times N$  matrix  $I$  is  $\text{tr } I = N$ .



Because the trace is an invariant, it actually defines a function on the vector space of linear maps  $\mathcal{L}(\mathbb{V})$ . The trace of a linear map  $A : \mathbb{V} \rightarrow \mathbb{V}$  is defined as the trace of any matrix of  $A$  relative to some basis. Invariance says that it does not depend on which basis we choose to compute it with respect to. As a function  $\text{tr} : \mathcal{L}(\mathbb{V}) \rightarrow \mathbb{R}$ , the trace is actually linear. It is an easy exercise to prove that

$$\text{tr}(A + B) = \text{tr } A + \text{tr } B \quad \text{and} \quad \text{tr}(\lambda A) = \lambda \text{tr } A .$$

There are other properties of a matrix which are *not* invariant under arbitrary change of basis; but are nevertheless important. For example, given a matrix  $\mathbf{A}$  let its **transpose**, denoted  $\mathbf{A}^t$ , be the matrix whose  $(i, j)$  entry equals the  $(j, i)$  entry of  $\mathbf{A}$ . Explicitly,

$$\mathbf{A} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1N} \\ A_{21} & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ A_{N1} & A_{N2} & \cdots & A_{NN} \end{pmatrix} \Rightarrow \mathbf{A}^t = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{N1} \\ A_{12} & A_{22} & \cdots & A_{N2} \\ \vdots & \vdots & \ddots & \vdots \\ A_{1N} & A_{2N} & \cdots & A_{NN} \end{pmatrix} .$$

In other words,  $\mathbf{A}^t$  is obtained from  $\mathbf{A}$  by reflecting the matrix on the main diagonal, and because reflection is an involution, it follows that

$$\boxed{(\mathbf{A}^t)^t = \mathbf{A} .} \quad (1.28)$$

It follows from the expression for  $\mathbf{A}^t$  that the diagonal entries are not changed, and hence that

$$\boxed{\text{tr } \mathbf{A}^t = \text{tr } \mathbf{A} .} \quad (1.29)$$

It is also easy to see that

$$\boxed{(\mathbf{A}\mathbf{B})^t = \mathbf{B}^t \mathbf{A}^t} \quad (1.30)$$

and also that

$$\boxed{(\mathbf{A} + \mathbf{B})^t = \mathbf{A}^t + \mathbf{B}^t} \quad \text{and} \quad \boxed{(\lambda\mathbf{A})^t = \lambda\mathbf{A}^t .} \quad (1.31)$$

From the former equation it follows that

$$\boxed{(\mathbf{A}^{-1})^t = (\mathbf{A}^t)^{-1} .} \quad (1.32)$$

A less obvious identity is

$$\boxed{\det \mathbf{A}^t = \det \mathbf{A} ,} \quad (1.33)$$

which follows from the fact that the row expansion of the determinant of  $\mathbf{A}^t$  is precisely the column expansion of the determinant of  $\mathbf{A}$ . A matrix is said to be **symmetric** if  $\mathbf{A}^t = \mathbf{A}$ . It is said to be **antisymmetric** or **skew-symmetric** if  $\mathbf{A}^t = -\mathbf{A}$ . Notice that an antisymmetric matrix is traceless, since

$$\text{tr } \mathbf{A} = \text{tr } \mathbf{A}^t = \text{tr}(-\mathbf{A}) = -\text{tr } \mathbf{A} .$$

The converse is of course false: a traceless matrix need not be antisymmetric.

Generic matrices are neither symmetric nor antisymmetric, yet any matrix is the sum of a symmetric matrix and an antisymmetric matrix. Indeed, adding and subtracting  $\frac{1}{2}\mathbf{A}^t$  in a clever way, we see that

$$\mathbf{A} = \frac{1}{2}(\mathbf{A} + \mathbf{A}^t) + \frac{1}{2}(\mathbf{A} - \mathbf{A}^t) .$$

But now, using equations (1.28) and (1.31), we see that  $\frac{1}{2}(\mathbf{A} + \mathbf{A}^t)$  is symmetric and  $\frac{1}{2}(\mathbf{A} - \mathbf{A}^t)$  antisymmetric.

A matrix  $\mathbf{O}$  is said to be **orthogonal** if its transpose is its inverse:

$$\mathbf{O}^t \mathbf{O} = \mathbf{O} \mathbf{O}^t = \mathbf{I} .$$



The property of being symmetric or antisymmetric is not invariant under arbitrary changes of basis, but it will be preserved under certain types of changes of basis, e.g., under orthogonal changes of basis.

## 1.3 Inner products

Vectors in physics are usually defined as objects which have both a magnitude and a direction. In that sense, they do not quite correspond to the mathematical notion of a vector as we have been discussing above. In our definition of an abstract vector space as in the discussion which followed, there is no mention of how to compute the magnitude of a vector. In this section we will remedy this situation. Geometrically the magnitude of a vector is simply its length. If we think of vectors as displacement, the magnitude is the distance away from the origin. In order to define distance we will need to introduce an *inner product* or *scalar product*, as it is often known.

### 1.3.1 Norms and inner products

Let us start by considering displacements in the plane. The length  $\|\mathbf{v}\|$  of the displacement  $\mathbf{v} = (v_1, v_2)$  is given by the Pythagorean theorem:  $\|\mathbf{v}\|^2 = v_1^2 + v_2^2$ . This length obeys the following properties which are easily verified. First of all it is a non-negative quantity  $\|\mathbf{v}\|^2 \geq 0$ , vanishing precisely for the zero displacement  $\mathbf{0} = (0, 0)$ . If we rescale  $\mathbf{v}$  by a real number  $\lambda$ :  $\lambda \mathbf{v} = (\lambda v_1, \lambda v_2)$ , its length rescales by the absolute value of  $\lambda$ :  $\|\lambda \mathbf{v}\| = |\lambda| \|\mathbf{v}\|$ . Finally, the length obeys the so-called triangle inequality:  $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ . This is obvious pictorially, since the shortest distance between two points in the plane is the straight line which joins them. In any case we will prove it later in much more generality.

Now consider  $\mathbb{R}^N$ . We can define a notion of length by generalising slightly what was done above: if  $(v_1, v_2, \dots, v_N) \in \mathbb{R}^N$ , then define its length by

$$\|(v_1, v_2, \dots, v_N)\| = \sqrt{v_1^2 + v_2^2 + \dots + v_N^2} .$$

It again satisfies the same three properties described above.

We can formalise this into the notion of a norm in a vector space. By a **norm** in a real vector space  $\mathbb{V}$  we mean a function  $\|\cdot\| : \mathbb{V} \rightarrow \mathbb{R}$  assigning a real number to every vector in  $\mathbb{V}$  in such a way that the following three properties are satisfied for every vector  $\mathbf{v}$  and  $\mathbf{w}$  and every scalar  $\lambda$ :

N1  $\|\mathbf{v}\| \geq 0$ , and  $\|\mathbf{v}\| = 0$  if and only if  $\mathbf{v} = \mathbf{0}$ ;

N2  $\|\lambda \mathbf{v}\| = |\lambda| \|\mathbf{v}\|$ ; and

N3 (triangle inequality)  $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$ .

The study of normed vector spaces is an important branch of modern mathematics (cf., one of the 1998 Fields Medals). In physics, however, it is fair to say that the more important notion is that of an inner product. If a norm allows us to calculate lengths, an inner product will allow us to also calculate angles.

Consider again the case of displacements in two dimensions, or equivalent  $\mathbb{R}^2$ . Let us define now a function which assigns a real number to two displacements  $\mathbf{v} = (v_1, v_2)$  and  $\mathbf{w} = (w_1, w_2)$ :

$$\langle \mathbf{v}, \mathbf{w} \rangle := v_1 w_1 + v_2 w_2 .$$

This is usually called the **dot product** and is written  $\mathbf{v} \cdot \mathbf{w}$ . We will not use this notation.

Clearly,  $\langle \mathbf{v}, \mathbf{v} \rangle = \|\mathbf{v}\|^2$ , so that this construction also incorporates a norm. If we write the displacements using polar coordinates:  $\mathbf{v} = \|\mathbf{v}\| (\cos \theta_1, \sin \theta_1)$  and similarly for  $\mathbf{w} = \|\mathbf{w}\| (\cos \theta_2, \sin \theta_2)$ , then we can compute:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \|\mathbf{v}\| \|\mathbf{w}\| \cos (\theta_1 - \theta_2) . \quad (1.34)$$

In other words,  $\langle \cdot, \cdot \rangle$  is essentially the angle between the two displacements. More generally we can consider  $\mathbb{R}^N$  and define its dot product as follows. If  $\mathbf{v} = (v_1, v_2, \dots, v_N)$  and  $\mathbf{w} = (w_1, w_2, \dots, w_N)$ , then

$$\langle \mathbf{v}, \mathbf{w} \rangle := \sum_{i=1}^N v_i w_i = v_1 w_1 + v_2 w_2 + \dots + v_N w_N .$$

The dot product satisfies the following properties. First of all it is symmetric:  $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ . It is also linear in the right-hand slot:  $\langle \mathbf{v}, \mathbf{w} + \mathbf{w}' \rangle = \langle \mathbf{v}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w}' \rangle$  and  $\langle \mathbf{v}, \lambda \mathbf{w} \rangle = \lambda \langle \mathbf{v}, \mathbf{w} \rangle$ ; and using the symmetry also in the left-hand slot. It is also important that the function  $\|\mathbf{v}\| := \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$  is a norm. The only non-obvious thing is to prove the triangle inequality for the norm, but we will do this below in all generality. The vector space  $\mathbb{R}^N$  with the dot product defined above is called  **$N$ -dimensional Euclidean space**, and is denoted  $\mathbb{E}^N$ . As a vector space, of course,  $\mathbb{E}^N = \mathbb{R}^N$ , but  $\mathbb{E}^N$  serves to remind us that we are talking about the vector space with the dot product. Notice that in terms of column vectors:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_N \end{pmatrix} \quad \text{and} \quad \mathbf{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix},$$

the dot product is given by

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^t \mathbf{w} = (v_1 \quad v_2 \quad \cdots \quad v_N) \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_N \end{pmatrix} = \sum_{i=1}^N v_i w_i.$$

More generally, we define an **inner product** (or **scalar product**) on a real vector space  $\mathbb{V}$  to be a function  $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{R}$  taking pairs of vectors to real numbers and obeying the following axioms:

IP1  $\langle \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{v} \rangle$ ;

IP2  $\langle \mathbf{u}, \lambda \mathbf{v} + \mu \mathbf{w} \rangle = \lambda \langle \mathbf{u}, \mathbf{v} \rangle + \mu \langle \mathbf{u}, \mathbf{w} \rangle$ ; and

IP3  $\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle > 0$  for all  $\mathbf{v} \neq \mathbf{0}$ .

Notice that IP1 and IP2 together imply that  $\langle \lambda \mathbf{u} + \mu \mathbf{v}, \mathbf{w} \rangle = \lambda \langle \mathbf{u}, \mathbf{w} \rangle + \mu \langle \mathbf{v}, \mathbf{w} \rangle$ .

Let  $\{\mathbf{e}_i\}$  be a basis for  $\mathbb{V}$ . Because of IP1 and IP2, it is enough to know what the inner product of any two basis elements is to be know what it is on any two vectors. Indeed, let  $\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i$  and  $\mathbf{w} = \sum_{i=1}^N w_i \mathbf{e}_i$  be any two vectors. Then their inner product is given by

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} \rangle &= \left\langle \sum_{i=1}^N v_i \mathbf{e}_i, \sum_{j=1}^N w_j \mathbf{e}_j \right\rangle \\ &= \sum_{i,j=1}^N v_i w_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle. \end{aligned} \quad (\text{using IP1,2})$$

In other words, all we need to know in order to compute this are the real numbers  $G_{ij} := \langle \mathbf{e}_i, \mathbf{e}_j \rangle$ . These can be thought of as the entries of a matrix  $\mathbf{G}$ . If we think of  $\mathbf{v}$  as a column vector  $\mathbf{v}$  in  $\mathbb{R}^N$  whose entries are the components of  $\mathbf{v}$  relative to the basis  $\{\mathbf{e}_i\}$ , and the same for  $\mathbf{w}$ , we can compute their inner product using matrix multiplication:

$$\langle \mathbf{v}, \mathbf{w} \rangle = \mathbf{v}^t \mathbf{G} \mathbf{w} .$$

The matrix  $\mathbf{G}$  is not arbitrary. First of all from IP1 it follows that it is symmetric:

$$G_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \langle \mathbf{e}_j, \mathbf{e}_i \rangle = G_{ji} .$$

Furthermore IP3 imposes a strong condition known as **positive-definiteness**. We will see at the end of this section what this means explicitly. Let us simply mention that IP3 implies that the only vector which is orthogonal to all vectors is the zero vector. This condition is weaker than IP3. It is often desirable to relax IP3 in terms of this condition. Such inner products are called **non-degenerate**. Non-degeneracy means that the matrix  $\mathbf{G}$  is invertible, so that its determinant is non-zero.

Here comes a point which confuses many people, so *pay attention!* Both inner products and linear transformations are represented by matrices relative to a basis, but they are very different objects. In particular, they transform differently under a change of basis and this means that even if the matrices for a linear transformation and an inner product agree numerically in a given basis, they will generically *not* agree with respect to a different basis. Let us see this in detail. Let  $\{\mathbf{e}'_i\}$  be a new basis, with  $\mathbf{e}'_i = S(\mathbf{e}_i)$  for some linear transformation  $S$ . Relative to  $\{\mathbf{e}_i\}$  the linear transformation  $S$  is represented by a matrix  $\mathbf{S}$  with entries  $S_{ji}$  given by equation (1.21). Let  $\mathbf{G}'$  denote the matrix describing the inner product in the new basis: its entries  $G'_{ij}$  are given by

$$\begin{aligned} G'_{ij} &= \langle \mathbf{e}'_i, \mathbf{e}'_j \rangle && \text{(by definition)} \\ &= \left\langle \sum_{k=1}^N S_{ki} \mathbf{e}_k, \sum_{l=1}^N S_{lj} \mathbf{e}_l \right\rangle && \text{(by equation (1.21))} \\ &= \sum_{k,l=1}^N S_{ki} S_{lj} \langle \mathbf{e}_k, \mathbf{e}_l \rangle && \text{(using IP1,2)} \\ &= \sum_{k,l=1}^N S_{ki} G_{kl} S_{lj} . \end{aligned}$$

In other words,

$$\boxed{G' = S^t G S} , \quad (1.35)$$

to be contrasted with the analogous formula (1.24) for the behaviour of the matrix of a linear transformation under a change of basis.



Notice, however, that under an orthogonal change of basis, so that  $S^{-1} = S^t$ , then both inner products and linear maps transform the same way.

### 1.3.2 The Cauchy–Schwartz and triangle inequalities

In this section we prove that  $\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v} \rangle}$  is indeed a norm. Because axioms N1 and N2 are obvious from the axioms of the inner product, all we really need to prove is the triangle inequality. This inequality will follow trivially from another inequality called the Cauchy–Schwartz inequality, and which is itself quite useful. Consider equation (1.34). Because the cosine function obeys  $|\cos \theta| \leq 1$ , we can deduce an inequality from equation (1.34). Namely that for any two displacements  $\mathbf{v}$  and  $\mathbf{w}$  in the plane,

$$|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\| ,$$

with equality if and only if the angle between the two displacements is zero; in other words, if the displacements are collinear. The above inequality is called the two-dimensional Cauchy–Schwartz inequality. This inequality actually holds in any vector space with an inner product (even if it is infinite-dimensional).

Let  $\mathbf{v}$  and  $\mathbf{w}$  be any two vectors in a vector space  $\mathbb{V}$  with an inner product  $\langle \cdot, \cdot \rangle$ . Let  $\lambda$  be a real number and let us consider the following inequality:

$$\begin{aligned} 0 &\leq \|\mathbf{v} - \lambda \mathbf{w}\|^2 \\ &= \langle \mathbf{v} - \lambda \mathbf{w}, \mathbf{v} - \lambda \mathbf{w} \rangle && \text{(by definition)} \\ &= \|\mathbf{v}\|^2 + \|\lambda \mathbf{w}\|^2 - 2\langle \mathbf{v}, \lambda \mathbf{w} \rangle && \text{(expanding and using IP1,2)} \\ &= \|\mathbf{v}\|^2 + \lambda^2 \|\mathbf{w}\|^2 - 2\lambda \langle \mathbf{v}, \mathbf{w} \rangle . && \text{(using IP2)} \end{aligned}$$

Now we want to make a clever choice of  $\lambda$  which allows us to partially cancel the last two terms against each other. This way we can hope to get an inequality involving only two terms. The clever choice of  $\lambda$  turns out to be  $\lambda = \langle \mathbf{v}, \mathbf{w} \rangle / \|\mathbf{w}\|^2$ . Inserting this into the above equation and rearranging the terms a little, we obtain the following inequality

$$\|\mathbf{v}\|^2 \geq \frac{\langle \mathbf{v}, \mathbf{w} \rangle^2}{\|\mathbf{w}\|^2} .$$

Taking the (positive) square root and rearranging we arrive at the **Cauchy–Schwartz inequality**:

$$\boxed{|\langle \mathbf{v}, \mathbf{w} \rangle| \leq \|\mathbf{v}\| \|\mathbf{w}\| .} \quad (1.36)$$

The triangle inequality now follows easily. Let us expand  $\|\mathbf{v} + \mathbf{w}\|^2$  as follows:

$$\begin{aligned} \|\mathbf{v} + \mathbf{w}\|^2 &= \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle \\ &= \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2\langle \mathbf{v}, \mathbf{w} \rangle && \text{(using IP1,2)} \\ &\leq \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2|\langle \mathbf{v}, \mathbf{w} \rangle| && \text{(since } x \leq |x|) \\ &\leq \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2\|\mathbf{v}\| \|\mathbf{w}\| && \text{(using Cauchy–Schwartz)} \\ &= (\|\mathbf{v}\| + \|\mathbf{w}\|)^2 . \end{aligned}$$

Taking the (positive) square root we arrive at the **triangle inequality**:

$$\boxed{\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\| .} \quad (1.37)$$

### 1.3.3 Orthonormal bases and Gram–Schmidt

Throughout this section we will let  $\mathbb{V}$  be an  $N$ -dimensional real vector space with an inner product  $\langle \cdot, \cdot \rangle$ .

We say that two vectors  $\mathbf{v}$  and  $\mathbf{w}$  are **orthogonal** (written  $\mathbf{v} \perp \mathbf{w}$ ) if their inner product vanishes:  $\langle \mathbf{v}, \mathbf{w} \rangle = 0$ . Any nonzero vector can be normalised to have unit norm simply dividing by its norm:  $\mathbf{v}/\|\mathbf{v}\|$  has unit norm. A basis  $\{\mathbf{e}_i\}$  is said to be **orthonormal** if

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (1.38)$$

In other words, the basis elements in an orthonormal basis are mutually orthogonal and are normalised to unit norm. Notice that the matrix representing the inner product relative to an orthonormal basis is the identity matrix.

The components of a vector  $\mathbf{v}$  relative to an orthonormal basis  $\{\mathbf{e}_i\}$  are very easy to compute. Let  $\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i$ , and take its inner product with

$\mathbf{e}_j$ :

$$\begin{aligned} \langle \mathbf{e}_j, \mathbf{v} \rangle &= \langle \mathbf{e}_j, \sum_{i=1}^N v_i \mathbf{e}_i \rangle \\ &= \sum_{i=1}^N v_i \langle \mathbf{e}_j, \mathbf{e}_i \rangle && \text{(using IP2)} \\ &= v_j . && \text{(using equation (1.38))} \end{aligned}$$

This shows that orthonormal vectors are automatically linearly independent. Indeed, suppose that  $\{\mathbf{e}_i\}$  are orthonormal vectors. Then suppose that a linear combination is the zero vector:

$$\sum_i \lambda_i \mathbf{e}_i = \mathbf{0} .$$

Taking the inner product of both sides of this equality with  $\mathbf{e}_j$  we find, on the left-hand side  $\lambda_j$  and on the right-hand side 0, hence  $\lambda_j = 0$  and thus the  $\{\mathbf{e}_i\}$  are linearly independent.

We now discuss an algorithmic procedure by which *any* basis can be modified to yield an orthonormal basis. Let  $\{\mathbf{f}_i\}$  be any basis whatsoever for  $\mathbb{V}$ . We will define iteratively a new basis  $\{\mathbf{e}_i\}$  which will be orthonormal. The procedure starts as follows. We define

$$\mathbf{e}_1 = \frac{\mathbf{f}_1}{\|\mathbf{f}_1\|} ,$$

which has unit norm by construction. We now define  $\mathbf{e}_2$  starting from  $\mathbf{f}_2$  but making it orthogonal to  $\mathbf{e}_1$  and normalising it to unit norm. A moment's thought reveals that the correct definition is

$$\mathbf{e}_2 = \frac{\mathbf{f}_2 - \langle \mathbf{f}_2, \mathbf{e}_1 \rangle \mathbf{e}_1}{\|\mathbf{f}_2 - \langle \mathbf{f}_2, \mathbf{e}_1 \rangle \mathbf{e}_1\|} .$$

It has unit norm by construction, and it is clearly orthogonal to  $\mathbf{e}_1$  because

$$\langle \mathbf{f}_2 - \langle \mathbf{f}_2, \mathbf{e}_1 \rangle \mathbf{e}_1, \mathbf{e}_1 \rangle = \langle \mathbf{f}_2, \mathbf{e}_1 \rangle - \langle \mathbf{f}_2, \mathbf{e}_1 \rangle \|\mathbf{e}_1\|^2 = 0 .$$

We can continue in this fashion and at each step define  $\mathbf{e}_i$  as  $\mathbf{f}_i + \dots$  divided by its norm, where the omitted terms are a linear combination of the  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{i-1}\}$  defined in such a way that the  $\mathbf{e}_i$  is orthogonal to them. For a finite-dimensional vector space, this procedure stops in a finite time

and we are left with an orthonormal basis  $\{\mathbf{e}_i\}$ . The general formulae for the  $\mathbf{e}_i$  is

$$\mathbf{e}_i = \frac{\mathbf{f}_i - \sum_{j=1}^{i-1} \langle \mathbf{f}_i, \mathbf{e}_j \rangle \mathbf{e}_j}{\|\mathbf{f}_i - \sum_{j=1}^{i-1} \langle \mathbf{f}_i, \mathbf{e}_j \rangle \mathbf{e}_j\|} . \quad (1.39)$$

Notice that this formula is *recursive*: it defines  $\mathbf{e}_i$  in terms of  $\mathbf{f}_i$  and the  $\{\mathbf{e}_{j < i}\}$ .



Studying this formula we see that each  $\mathbf{e}_i$  is a linear combination

$$\mathbf{e}_i = \sum_{j=1}^i S_{ji} \mathbf{f}_j , \quad (1.40)$$

where  $S_{ii}$  is positive, since it is given by  $S_{ii} = 1/\|\mathbf{f}_i + \dots\|$ . Now let  $S$  be the linear transformation defined by  $S(\mathbf{f}_i) = \mathbf{e}_i$ . Relative to the original basis  $\{\mathbf{f}_i\}$ ,  $S$  has a matrix  $\mathbf{S}$  with entries  $S_{ji}$  defined by

$$\mathbf{e}_i = \sum_{j=1}^N S_{ji} \mathbf{f}_j .$$

Comparing with equation (1.40) we see that  $S_{ji} = 0$  for  $j > i$ , so that all the entries of  $\mathbf{S}$  below the main diagonal are zero. We say that  $\mathbf{S}$  is **upper triangular**. The condition  $S_{ii} > 0$  says that the diagonal entries are positive.

We can turn equation (1.39) around and notice that  $\mathbf{f}_i$  is in turn given as a linear combination of  $\{\mathbf{e}_{j \leq i}\}$ . The linear transformation  $T$  defined by  $\mathbf{f}_i = T(\mathbf{e}_i)$ , which is the inverse of  $S$ , has a matrix  $\mathbf{T}$  relative to the  $\{\mathbf{e}_i\}$  basis which is also upper triangular with positive entries on the main diagonal. Now the matrix  $\mathbf{G}$  with entries  $G_{ij} = \langle \mathbf{f}_i, \mathbf{f}_j \rangle$  representing the inner product on the  $\{\mathbf{f}_i\}$  basis, is now given by

$$\mathbf{G} = \mathbf{T}^t \mathbf{T} .$$

In other words, since the  $\{\mathbf{f}_i\}$  were an arbitrary basis,  $\mathbf{G}$  is an arbitrary matrix representing an inner product. We have learned then that this matrix can always be written as a "square"  $\mathbf{T}^t \mathbf{T}$ , where  $\mathbf{T}$  is an upper triangular matrix with positive entries in the main diagonal.

### 1.3.4 The adjoint of a linear transformation

Throughout this section we will let  $\mathbb{V}$  be an  $N$ -dimensional real vector space with an inner product  $\langle \cdot, \cdot \rangle$ .

Let  $A : \mathbb{V} \rightarrow \mathbb{V}$  be a linear transformation. A linear transformation is uniquely defined by its *matrix elements*  $\langle A(\mathbf{v}), \mathbf{w} \rangle$ . Indeed, if  $A'$  is another linear transformation with  $\langle A'(\mathbf{v}), \mathbf{w} \rangle = \langle A(\mathbf{v}), \mathbf{w} \rangle$  for all  $\mathbf{v}$  and  $\mathbf{w}$ , then we claim that  $A = A'$ . To see this notice that

$$\begin{aligned} 0 &= \langle A'(\mathbf{v}), \mathbf{w} \rangle - \langle A(\mathbf{v}), \mathbf{w} \rangle \\ &= \langle A'(\mathbf{v}) - A(\mathbf{v}), \mathbf{w} \rangle . \end{aligned} \quad (\text{using IP1,2})$$

Since this is true for all  $\mathbf{w}$ , it says that the vector  $A'(\mathbf{v}) - A(\mathbf{v})$  is orthogonal to all vectors, and in particular to itself. Therefore it has zero norm and by IP3 it is the zero vector. In other words,  $A'(\mathbf{v}) = A(\mathbf{v})$  for all  $\mathbf{v}$ , which means that  $A = A'$ .

Given a linear transformation  $A : \mathbb{V} \rightarrow \mathbb{V}$  we define its **adjoint** relative to the inner product, as the linear transformation  $A^\dagger : \mathbb{V} \rightarrow \mathbb{V}$  with matrix elements

$$\boxed{\langle A^\dagger(\mathbf{v}), \mathbf{w} \rangle = \langle \mathbf{v}, A(\mathbf{w}) \rangle .} \quad (1.41)$$

The adjoint operation obeys several properties. First of all, taking adjoint is an involution:

$$\boxed{A^{\dagger\dagger} = A .} \quad (1.42)$$

Moreover it is a linear operation

$$\boxed{(\lambda A + \mu B)^\dagger = \lambda A^\dagger + \mu B^\dagger ,} \quad (1.43)$$

which reverses the order of a composition:

$$\boxed{(A \circ B)^\dagger = B^\dagger \circ A^\dagger .} \quad (1.44)$$



These properties are easily proven. The method of proof consists in showing that both sides of each equation have the same matrix elements. For example, the matrix elements of the double adjoint  $A^{\dagger\dagger}$  are given by

$$\begin{aligned} \langle A^{\dagger\dagger}(\mathbf{v}), \mathbf{w} \rangle &= \langle \mathbf{v}, A^\dagger(\mathbf{w}) \rangle && \text{(by equation (1.41))} \\ &= \langle A^\dagger(\mathbf{w}), \mathbf{v} \rangle && \text{(by IP1)} \\ &= \langle \mathbf{w}, A(\mathbf{v}) \rangle && \text{(by equation (1.41))} \\ &= \langle A(\mathbf{v}), \mathbf{w} \rangle ; && \text{(by IP1)} \end{aligned}$$

whence they agree with the matrix elements of  $A$ .

Similarly, the matrix elements of  $(\lambda A + \mu B)^\dagger$  are given by

$$\begin{aligned} \langle (\lambda A + \mu B)^\dagger(\mathbf{v}), \mathbf{w} \rangle &= \langle \mathbf{v}, (\lambda A + \mu B)(\mathbf{w}) \rangle && \text{(by equation (1.41))} \\ &= \lambda \langle \mathbf{v}, A(\mathbf{w}) \rangle + \mu \langle \mathbf{v}, B(\mathbf{w}) \rangle && \text{(using IP2)} \\ &= \lambda \langle A^\dagger(\mathbf{v}), \mathbf{w} \rangle + \mu \langle B^\dagger(\mathbf{v}), \mathbf{w} \rangle && \text{(by equation (1.41))} \\ &= \langle (\lambda A^\dagger + \mu B^\dagger)(\mathbf{v}), \mathbf{w} \rangle , && \text{(using IP1,2)} \end{aligned}$$

which agree with the matrix elements of  $\lambda A^\dagger + \mu B^\dagger$ .

Finally, the matrix elements of  $(A \circ B)^\dagger$  are given by

$$\begin{aligned} \langle (A \circ B)^\dagger(\mathbf{v}), \mathbf{w} \rangle &= \langle \mathbf{v}, (A \circ B)(\mathbf{w}) \rangle && \text{(by equation (1.41))} \\ &= \langle \mathbf{v}, A(B(\mathbf{w})) \rangle && \text{(by equation (1.2))} \\ &= \langle A^\dagger(\mathbf{v}), B(\mathbf{w}) \rangle && \text{(by equation (1.41))} \\ &= \langle B^\dagger(A^\dagger(\mathbf{v})), \mathbf{w} \rangle && \text{(by equation (1.41))} \\ &= \langle (B^\dagger \circ A^\dagger)(\mathbf{v}), \mathbf{w} \rangle , && \text{(by equation (1.2))} \end{aligned}$$

which agree with the matrix elements of  $B^\dagger \circ A^\dagger$ .

A linear transformation is said to be **symmetric** if  $A^\dagger = A$ . It is said to be **orthogonal** if  $A^\dagger \circ A = A \circ A^\dagger = \mathbb{1}$ . In particular, orthogonal transformations preserve inner products:

$$\begin{aligned} \langle A(\mathbf{v}), A(\mathbf{w}) \rangle &= \langle \mathbf{v}, A^\dagger(A(\mathbf{w})) \rangle && \text{(by equation (1.41))} \\ &= \langle \mathbf{v}, (A^\dagger \circ A)(\mathbf{w}) \rangle && \text{(by equation (1.2))} \\ &= \langle \mathbf{v}, \mathbf{w} \rangle . && \text{(since } A \text{ is orthogonal)} \end{aligned}$$



Notice that in the above we only used the condition  $A^\dagger \circ A = \mathbb{1}$  but not  $A \circ A^\dagger = \mathbb{1}$ . In a finite-dimensional vector space one implies the other, but in infinite dimensional vector spaces it may happen that a linear transformation which preserves the inner product obeys  $A^\dagger \circ A = \mathbb{1}$  but does *not* obey  $A \circ A^\dagger = \mathbb{1}$ . (Maybe an example?)

To justify these names, notice that *relative to an orthonormal basis* the matrix of a symmetric transformation is symmetric and the matrix of an orthogonal transformation is orthogonal, as defined in Section 1.2.7. This follows because the matrix of the adjoint of a linear transformation is the transpose of the matrix of the linear transformation.

Let us prove this. Let  $\{\mathbf{e}_i\}$  be an orthonormal basis and let  $A : \mathbb{V} \rightarrow \mathbb{V}$  be a linear transformation. The matrix  $\mathbf{A}$  of  $A$  relative to this basis has entries  $A_{ij}$  defined by

$$A(\mathbf{e}_i) = \sum_{j=1}^N A_{ji} \mathbf{e}_j .$$

The entries  $A_{ij}$  are also given by matrix elements:

$$\begin{aligned} \langle A(\mathbf{e}_i), \mathbf{e}_j \rangle &= \left\langle \sum_{k=1}^N A_{ki} \mathbf{e}_k, \mathbf{e}_j \right\rangle \\ &= \sum_{k=1}^N A_{ki} \langle \mathbf{e}_k, \mathbf{e}_j \rangle && \text{(using IP1,2)} \\ &= A_{ji} . && \text{(using equation (1.38))} \end{aligned}$$

In other words, relative to an orthonormal basis, we have the following useful formula:

$$\boxed{A_{ij} = \langle \mathbf{e}_i, A(\mathbf{e}_j) \rangle} . \quad (1.45)$$

From this it follows that the matrix of the adjoint  $A^\dagger$  relative to this basis is given by  $\mathbf{A}^t$ . Indeed,

$$\begin{aligned} A_{ij}^\dagger &= \langle A^\dagger(\mathbf{e}_j), \mathbf{e}_i \rangle \\ &= \langle \mathbf{e}_j, A(\mathbf{e}_i) \rangle && \text{(using equation (1.41))} \\ &= \langle A(\mathbf{e}_i), \mathbf{e}_j \rangle && \text{(using IP1)} \\ &= A_{ji} . \end{aligned}$$

Therefore if  $A^\dagger = A$ , then  $\mathbf{A} = \mathbf{A}^t$ , and the matrix is symmetric. Similarly, if  $A \circ A^\dagger = A^\dagger \circ A = \mathbb{1}$ , then  $\mathbf{A}^t \mathbf{A} = \mathbf{A} \mathbf{A}^t = \mathbf{I}$ , and the matrix is orthogonal.

Notice that equations (1.42), (1.43) and (1.44) for the linear transformations are now seen to be consequences of equations (1.28), (1.31) and (1.30) applied to their matrices relative to an orthonormal basis.

### 1.3.5 Complex vector spaces

Much of what we have been saying about vector spaces remains true if we substitute the scalars and instead of real numbers consider complex numbers. Only the notion of an inner product will have to be changed in order for it to become useful. Inner products on complex vector spaces will be the subject of the next section; in this one, we want to emphasise those aspects of vectors spaces which remain unchanged when we extend the scalars from the real to the complex numbers.

As you know, complex numbers themselves can be understood as a real vector space of dimension two; that is, as  $\mathbb{R}^2$ . If  $z = x + iy$  is a complex number with  $x, y$  real and  $i = \sqrt{-1}$ , then we can think of  $z$  as the pair  $(x, y) \in \mathbb{R}^2$ . Addition of complex numbers corresponds to vector addition in  $\mathbb{R}^2$ . Indeed, if  $z = x + iy$  and  $w = u + iv$  then  $z + w = (x + u) + i(y + v)$ , which is precisely what we expect from the vector addition  $(x, y) + (u, v) = (x + u, y + v)$ . Similarly, multiplication by a *real* number  $\lambda$  corresponds to scalar multiplication in  $\mathbb{R}^2$ . Indeed,  $\lambda z = (\lambda x) + i(\lambda y)$ , which is in agreement with  $\lambda(x, y) = (\lambda x, \lambda y)$ . However the complex numbers have more structure than that of a mere vector space. Unlike vectors in a general vector space, complex numbers can be multiplied: if  $z = x + iy$  and  $w = u + iv$ , then  $zw = (xu - yv) + i(xv + yu)$ . Multiplication is commutative:  $wz = zw$ .



In a sense, complex numbers are more like matrices than vectors. Indeed, consider the  $2 \times 2$  matrices of the form

$$\begin{pmatrix} a & -b \\ b & a \end{pmatrix}.$$

If we take the matrix product

$$\begin{pmatrix} x & -y \\ y & x \end{pmatrix} \begin{pmatrix} u & -v \\ v & u \end{pmatrix} = \begin{pmatrix} xu - yv & -(xv + yu) \\ xv + yu & xu - yv \end{pmatrix},$$

we see that we recover the multiplication of complex numbers. Notice that the complex number  $i$  is represented by the matrix

$$\mathbf{J} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

which obeys  $\mathbf{J}^2 = -\mathbf{I}$ .

A real matrix  $\mathbf{J}$  obeying  $\mathbf{J}^2 = -\mathbf{I}$  is called a **complex structure**.

We now briefly review some basic facts about complex numbers. Although you should be familiar with the following concepts, I will briefly review them here just to set the notation. As we have seen complex number can be added and multiplied. So far that is as with the real numbers, but in addition there is a notion of **complex conjugation**:  $z = x + iy \mapsto z^* = x - iy$ . Clearly conjugation is an involution:  $(z^*)^* = z$ . It also obeys  $(zw)^* = z^*w^*$ . A complex number  $z$  is said to be **real** if it is invariant under conjugation:  $z^* = z$ . Similarly a complex number is said to be **imaginary** if  $z^* = -z$ . Given  $z = x + iy$ ,  $z$  is real if and only if  $y = 0$ , whereas  $z$  is imaginary if and only if  $x = 0$ . If  $z = x + iy$ ,  $x$  is said to be the **real part** of  $z$ , written  $x = \operatorname{Re} z$ , and  $y$  is said to be the **imaginary part** of  $z$ , written  $y = \operatorname{Im} z$ . Notice that the imaginary part of a complex number is a real number, not an imaginary number! Given a complex number  $z$ , the product  $zz^*$  is real:  $(zz^*)^* = zz^*$ . It is written  $|z|^2$  and it is called the **modulus** of  $z$ . If  $z = x + iy$ , then  $|z|^2 = x^2 + y^2$ , which coincides with the squared norm  $\|(x, y)\|^2$  of the corresponding vector in the plane. Notice that the modulus is multiplicative:  $|zw| = |z||w|$  and invariant under conjugation:  $|z^*| = |z|$ .

After this flash review of complex numbers, it is possible to define the notion of a complex vector space. There is really very little to do. Everything that was said in Sections 1.1 and 1.2 still holds provided we replace *real* with *complex* everywhere. An abstract complex vector space satisfies the same axioms, except that the scalars are now complex numbers as opposed to real numbers. Vector subspaces work the same way. Bases and linear independence also work in the same way, linear combinations being now complex linear combinations. The canonical example of a complex vector space is  $\mathbb{C}^N$ , the set of ordered  $N$ -tuples of complex numbers:  $(z_1, z_2, \dots, z_N)$ , with the operations defined slot-wise as for  $\mathbb{R}^N$ . The canonical basis  $\{(1, 0, \dots, 0), (0, 1, \dots, 0), \dots, (0, 0, \dots, 1)\}$  still spans  $\mathbb{C}^N$ , but where we now take complex linear combinations. As a result  $\mathbb{C}^N$  has (complex) dimension  $N$ . If we only allowed ourselves to take real linear combinations, then in order to span  $\mathbb{C}^N$  we would need in addition the  $N$  vectors  $\{(i, 0, \dots, 0), (0, i, \dots, 0), \dots, (0, 0, \dots, i)\}$ , showing that as a *real* vector space,  $\mathbb{C}^N$  is  $2N$ -dimensional.

Linear maps and linear transformations are now complex linear and matrices and column vectors now have complex entries instead of real entries. Matrix invariants like the trace and the determinant are now complex numbers instead of real numbers. There is one more operation we can do with complex matrices, and that is to take complex conjugation. If  $\mathbf{A}$  is a complex  $N \times M$  matrix, then  $\mathbf{A}^*$  is the  $N \times M$  matrix whose entries are simply the

complex conjugates of the entries in  $\mathbf{A}$ . Clearly, for square matrices,

$$\det(\mathbf{A}^*) = (\det \mathbf{A})^* \quad \text{and} \quad \text{tr}(\mathbf{A}^*) = (\text{tr} \mathbf{A})^* .$$

The only significant difference between real and complex vector spaces is when we introduce inner products, which we do now.

### 1.3.6 Hermitian inner products

We motivated the introduction of inner products as a way to measure, in particular, lengths of vectors. The need to compute lengths was motivated in turn by the fact that the vectorial quantities used in physics have a magnitude as well as a direction. Magnitudes, like anything else that one ever measures experimentally, are positive (or at least non-negative) real numbers. However if we were to simply extend the dot product from  $\mathbb{R}^N$  to  $\mathbb{C}^N$ , we would immediately notice that for  $\mathbf{z} = (z_1, z_2, \dots, z_N) \in \mathbb{C}^N$ , the dot product with itself

$$\mathbf{z} \cdot \mathbf{z} = \sum_{i=1}^N z_i z_i ,$$

gives a complex number, not a real number. Hence we cannot understand this as a length. One way to generate a positive real number is to define the following inner product on  $\mathbb{C}^N$ :

$$\langle \mathbf{z}, \mathbf{w} \rangle = \sum_{i=1}^N z_i^* w_i ,$$

where  $\mathbf{z} = (z_1, z_2, \dots, z_N)$  and  $\mathbf{w} = (w_1, w_2, \dots, w_N)$ . It is then easy to see that now

$$\langle \mathbf{z}, \mathbf{z} \rangle = \sum_{i=1}^N z_i^* z_i = \sum_{i=1}^N |z_i|^2 ,$$

so that this is a non-negative real number, so that it can be interpreted as a norm. The above inner product obeys the following property, in contrast with the dot product in  $\mathbb{R}^N$ : it is not symmetric, so rather than IP1 it obeys  $\langle \mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{z} \rangle^*$ .

This suggests the following definition. A complex valued function  $\langle \cdot, \cdot \rangle : \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{C}$  taking pairs of vectors to complex numbers is called a **hermitian inner product** if the following axioms are satisfied:

HIP1  $\langle \mathbf{z}, \mathbf{w} \rangle = \langle \mathbf{w}, \mathbf{z} \rangle^*$ ;

HIP2  $\langle \mathbf{x}, \lambda \mathbf{z} + \mu \mathbf{w} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle + \mu \langle \mathbf{x}, \mathbf{w} \rangle$ ; and

HIP3  $\|\mathbf{z}\|^2 = \langle \mathbf{z}, \mathbf{z} \rangle > 0$  for all  $\mathbf{z} \neq \mathbf{0}$ ,

where here  $\lambda$  and  $\mu$  are complex scalars.

Except for the fact that  $\langle \cdot, \cdot \rangle$  is a complex function, the only obvious difference is HIP1. Using HIP1 and HIP2 we see that

$$\begin{aligned} \langle \lambda \mathbf{z} + \mu \mathbf{w}, \mathbf{x} \rangle &= \langle \mathbf{x}, \lambda \mathbf{z} + \mu \mathbf{w} \rangle^* && \text{(by HIP1)} \\ &= (\lambda \langle \mathbf{x}, \mathbf{z} \rangle + \mu \langle \mathbf{x}, \mathbf{w} \rangle)^* && \text{(by HIP2)} \\ &= \lambda^* \langle \mathbf{x}, \mathbf{z} \rangle^* + \mu^* \langle \mathbf{x}, \mathbf{w} \rangle^* \\ &= \lambda^* \langle \mathbf{z}, \mathbf{x} \rangle + \mu \langle \mathbf{w}, \mathbf{x} \rangle, && \text{(using HIP1)} \end{aligned}$$

so that  $\langle \cdot, \cdot \rangle$  is complex linear in the second slot but only conjugate linear in the first. One says that hermitian inner products are **sesquilinear**, which means ‘one and a half’ linear.

Just as in the real case, the inner product of any two vectors is determined by the matrix of inner products relative to any basis. Let  $\{\mathbf{e}_i\}$  be a basis for  $\mathbb{V}$ . Let  $\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i$  and  $\mathbf{w} = \sum_{i=1}^N w_i \mathbf{e}_i$  be any two vectors. Then their inner product is given by

$$\begin{aligned} \langle \mathbf{v}, \mathbf{w} \rangle &= \left\langle \sum_{i=1}^N v_i \mathbf{e}_i, \sum_{j=1}^N w_j \mathbf{e}_j \right\rangle \\ &= \sum_{i,j=1}^N v_i^* w_j \langle \mathbf{e}_i, \mathbf{e}_j \rangle. \end{aligned} \quad \text{(using HIP1,2)}$$

In other words, all we need to know in order to compute this are the complex numbers  $H_{ij} := \langle \mathbf{e}_i, \mathbf{e}_j \rangle$ , which can be thought of as the entries of a matrix  $\mathbf{H}$ . If we think of  $\mathbf{v}$  as a column vector  $\mathbf{v}$  in  $\mathbb{C}^N$  whose entries are the components of  $\mathbf{v}$  relative to the basis  $\{\mathbf{e}_i\}$ , and the same for  $\mathbf{w}$ , we can compute their inner product using matrix multiplication:

$$\langle \mathbf{v}, \mathbf{w} \rangle = (\mathbf{v}^*)^t \mathbf{H} \mathbf{w} .$$

We saw in the real case that the analogous matrix there was symmetric and positive-definite, reflecting the similar properties of the inner product. In the complex case, we expect that  $\mathbf{H}$  should still be positive-definite but that instead of symmetry it should obey a property based on HIP1. Indeed, it follows from HIP1 that

$$H_{ij} = \langle \mathbf{e}_i, \mathbf{e}_j \rangle = \langle \mathbf{e}_j, \mathbf{e}_i \rangle^* = H_{ji}^* .$$

This means that the matrix  $\mathbf{H}$  is equal to its conjugate transpose:

$$\mathbf{H} = (\mathbf{H}^*)^t . \quad (1.46)$$

Such matrices are called **hermitian**. Property **HIP3** means that  $\mathbf{H}$  is positive-definite, so that in particular it is non-degenerate.

Let us see how  $\mathbf{H}$  transforms under a change of basis. Let  $\{\mathbf{e}'_i\}$  be a new basis, with  $\mathbf{e}'_i = S(\mathbf{e}_i)$  for some complex linear transformation  $S$ . Relative to  $\{\mathbf{e}_i\}$  the linear transformation  $S$  is represented by a matrix  $\mathbf{S}$  with entries  $S_{ji}$  given by equation (1.21). Let  $\mathbf{H}'$  denote the matrix describing the inner product in the new basis: its entries  $H'_{ij}$  are given by

$$\begin{aligned}
H'_{ij} &= \langle \mathbf{e}'_i, \mathbf{e}'_j \rangle && \text{(by definition)} \\
&= \left\langle \sum_{k=1}^N S_{ki} \mathbf{e}_k, \sum_{l=1}^N S_{lj} \mathbf{e}_l \right\rangle && \text{(by equation (1.21))} \\
&= \sum_{k,l=1}^N S_{ki}^* S_{lj} \langle \mathbf{e}_k, \mathbf{e}_l \rangle && \text{(using HIP1,2)} \\
&= \sum_{k,l=1}^N S_{ki}^* H_{kl} S_{lj} .
\end{aligned}$$

In other words,

$$\boxed{H' = (\mathbf{S}^*)^t \mathbf{H} \mathbf{S}} , \tag{1.47}$$

to be contrasted with the analogous formula (1.35).

The Cauchy–Schwarz and triangle inequalities are still valid for hermitian inner products. The proofs are essentially the same as for the real case. We will therefore be brief.

In order to prove the Cauchy–Schwarz inequality, we start the following inequality, which follows from **HIP3**,

$$\|\mathbf{v} - \lambda \mathbf{w}\|^2 \geq 0 ,$$

and choose  $\lambda \in \mathbb{C}$  appropriately. Expanding this out using **HIP1** and **HIP2** we can rewrite it as

$$\|\mathbf{v}\|^2 + |\lambda|^2 \|\mathbf{w}\|^2 - \lambda \langle \mathbf{v}, \mathbf{w} \rangle - \lambda^* \langle \mathbf{w}, \mathbf{v} \rangle \geq 0 .$$

Hence if we choose  $\lambda = \langle \mathbf{w}, \mathbf{v} \rangle / \|\mathbf{w}\|^2$ , we turn the inequality into

$$\|\mathbf{v}\|^2 - \frac{|\langle \mathbf{v}, \mathbf{w} \rangle|^2}{\|\mathbf{w}\|^2} \geq 0 ,$$

which can be rewritten as

$$|\langle \mathbf{v}, \mathbf{w} \rangle|^2 \leq \|\mathbf{v}\|^2 \|\mathbf{w}\|^2 .$$

Taking square roots (all quantities are positive) we obtain the Cauchy–Schwarz inequality (1.36).

In order to prove the triangle inequality, we start with

$$\begin{aligned}
\|\mathbf{v} + \mathbf{w}\|^2 &= \langle \mathbf{v} + \mathbf{w}, \mathbf{v} + \mathbf{w} \rangle \\
&= \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2 \operatorname{Re} \langle \mathbf{v}, \mathbf{w} \rangle \\
&\leq \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2|\langle \mathbf{v}, \mathbf{w} \rangle| && \text{(since } \operatorname{Re} z \leq |z| \ \forall z \in \mathbb{C} \text{)} \\
&\leq \|\mathbf{v}\|^2 + \|\mathbf{w}\|^2 + 2\|\mathbf{v}\| \|\mathbf{w}\| && \text{(by Cauchy–Schwarz)} \\
&= (\|\mathbf{v}\| + \|\mathbf{w}\|)^2 ;
\end{aligned}$$

whence taking square roots we obtain the triangle inequality (1.37).

The complex analogue of an orthonormal basis is a unitary basis. Explicitly, a basis  $\{\mathbf{e}_i\}$  is said to be **unitary** if

$$\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \delta_{ij} := \begin{cases} 1 & \text{if } i = j \\ 0 & \text{otherwise.} \end{cases} \quad (1.48)$$

The components of a vector  $\mathbf{v}$  relative to a unitary basis  $\{\mathbf{e}_i\}$  can be computed by taking inner products, just as in the real case. Let  $\mathbf{v} = \sum_{i=1}^N v_i \mathbf{e}_i$ , and take its inner product with  $\mathbf{e}_j$ :

$$\begin{aligned}
\langle \mathbf{e}_j, \mathbf{v} \rangle &= \langle \mathbf{e}_j, \sum_{i=1}^N v_i \mathbf{e}_i \rangle \\
&= \sum_{i=1}^N v_i \langle \mathbf{e}_j, \mathbf{e}_i \rangle && \text{(using HIP2)} \\
&= v_j . && \text{(using equation (1.48))}
\end{aligned}$$

This shows that unitary vectors are automatically linearly independent.

One still has the Gram–Schmidt procedure for hermitian inner products. It works essentially in the same way as in the real case, so we will not spend much time on this. Consider a basis  $\{\mathbf{f}_i\}$  for  $\mathbb{V}$ . Define the following vectors:

$$\mathbf{e}_i = \frac{\mathbf{f}_i - \sum_{j=1}^{i-1} \langle \mathbf{e}_j, \mathbf{f}_i \rangle \mathbf{e}_j}{\|\mathbf{f}_i - \sum_{j=1}^{i-1} \langle \mathbf{e}_j, \mathbf{f}_i \rangle \mathbf{e}_j\|} .$$

It is easily checked that they are a unitary basis. First of all each  $\mathbf{e}_i$  is clearly normalised, because it is defined as a vector divided by its norm; and moreover if  $i > j$ , then  $\mathbf{e}_i$  is clearly orthogonal to  $\mathbf{e}_j$ .

Finally, we discuss the adjoint of a complex linear map relative to a hermitian inner product. Let  $A : \mathbb{V} \rightarrow \mathbb{V}$  be a complex linear map. We

define its adjoint  $A^\dagger$  by equation (1.41), where now  $\langle \cdot, \cdot \rangle$  is a hermitian inner product. The properties (1.42) and (1.44) still hold, and are proven in exactly the same way.

Only property (1.43) changes, reflecting the sesquilinear nature of the inner product. Indeed notice that

$$\begin{aligned}
\langle (\lambda A + \mu B)^\dagger \mathbf{v}, \mathbf{w} \rangle &= \langle \mathbf{v}, (\lambda A + \mu B) \mathbf{w} \rangle && \text{(by (1.41))} \\
&= \lambda \langle \mathbf{v}, A \mathbf{w} \rangle + \mu \langle \mathbf{v}, B \mathbf{w} \rangle && \text{(by HIP2)} \\
&= \lambda \langle A^\dagger \mathbf{v}, \mathbf{w} \rangle + \mu \langle B^\dagger \mathbf{v}, \mathbf{w} \rangle && \text{(by (1.41))} \\
&= (\lambda^* \langle \mathbf{w}, A^\dagger \mathbf{v} \rangle + \mu^* \langle \mathbf{w}, B^\dagger \mathbf{v} \rangle)^* && \text{(by HIP1)} \\
&= \langle \mathbf{w}, (\lambda^* A^\dagger + \mu^* B^\dagger) \mathbf{v} \rangle^* && \text{(by HIP2)} \\
&= \langle (\lambda^* A^\dagger + \mu^* B^\dagger) \mathbf{v}, \mathbf{w} \rangle ; && \text{(by HIP1)}
\end{aligned}$$

whence

$$\boxed{(\lambda A + \mu B)^\dagger = \lambda^* A^\dagger + \mu^* B^\dagger .} \quad (1.49)$$

A complex linear transformation  $A$  is said to be **hermitian** if  $A^\dagger = A$ , and it is said to be **anti-hermitian** (also **skew-hermitian**) if  $A^\dagger = -A$ . As in the real case, the nomenclature can be justified by noticing that the matrix of a hermitian transformation relative to a unitary basis is hermitian, as defined in equation (1.46). The proof is similar to the proof of the analogous statement in the real case. Indeed,

$$\begin{aligned}
A_{ij}^\dagger &= \langle A^\dagger(\mathbf{e}_j), \mathbf{e}_i \rangle && \text{(by equation (1.45))} \\
&= \langle \mathbf{e}_j, A(\mathbf{e}_i) \rangle && \text{(using equation (1.41))} \\
&= \langle A(\mathbf{e}_i), \mathbf{e}_j \rangle^* && \text{(using HIP1)} \\
&= A_{ji}^* . && \text{(by equation (1.45))}
\end{aligned}$$

Therefore if  $A^\dagger = A$ , then  $\mathbf{A} = (\mathbf{A}^*)^t$ , and the matrix is hermitian. Notice that if  $\mathbf{A}$  is a hermitian matrix, then  $i\mathbf{A}$  is antihermitian, hence unlike the real case, the distinction between hermitian and anti-hermitian is trivial.

Let us say that a linear transformation  $U$  is **unitary** if  $U^\dagger \circ U = U \circ U^\dagger = \mathbb{1}$ . In this case, its matrix  $\mathbf{U}$  relative to a unitary basis obeys  $(\mathbf{U}^*)^t \mathbf{U} = \mathbf{U} (\mathbf{U}^*)^t = \mathbf{I}$ . This means that the conjugate transpose is the inverse,

$$\mathbf{U}^{-1} = (\mathbf{U}^*)^t . \quad (1.50)$$

Not surprisingly, such matrices are called **unitary**. Finally let us notice that

just as in the real case, a unitary transformation preserves the inner product:

$$\begin{aligned}\langle U(\mathbf{v}), U(\mathbf{w}) \rangle &= \langle \mathbf{v}, U^\dagger(U(\mathbf{w})) \rangle && \text{(by equation (1.41))} \\ &= \langle \mathbf{v}, (U^\dagger \circ U)(\mathbf{w}) \rangle && \text{(by equation (1.2))} \\ &= \langle \mathbf{v}, \mathbf{w} \rangle . && \text{(since } U \text{ is unitary)}\end{aligned}$$

## 1.4 The eigenvalue problem and applications

In this section we study perhaps the most important aspect of linear algebra from a physical perspective: the so-called eigenvalue problem. We mentioned when we introduced the notion of a basis that a good choice of basis can often simplify the solution of a problem involving linear transformations. Given a linear transformation, it is hard to imagine a better choice of basis than one in which the matrix is diagonal. However not all linear transformations admit such a basis. Understanding which transformations admit such basis is an important part of linear algebra; but one whose full solution requires more machinery than the one we will have available in this course. We will content ourselves with showing that certain types of linear transformation of use in physics do admit a diagonal basis. We will finish this section with two applications of these results: one to mathematics (quadratic forms) and one to physics (normal modes).

### 1.4.1 Eigenvectors and eigenvalues

Throughout this section  $\mathbb{V}$  shall be an  $N$ -dimensional complex vector space.

Let  $A : \mathbb{V} \rightarrow \mathbb{V}$  be a complex linear transformation. Let  $\mathbf{v} \in \mathbb{V}$  be a *nonzero* vector which obeys

$$A\mathbf{v} = \lambda\mathbf{v} \quad \text{for some } \lambda \in \mathbb{C}. \quad (1.51)$$

We say that  $\mathbf{v}$  is an **eigenvector** of  $A$  with **eigenvalue**  $\lambda$ . Let  $\{\mathbf{e}_i\}$  be a basis for  $\mathbb{V}$ . Let  $\mathbf{v}$  be the column vector whose entries are the components  $v_i$  of  $\mathbf{v}$  relative to this basis:  $\mathbf{v} = \sum_i v_i \mathbf{e}_i$ ; and let  $\mathbf{A}$  be the matrix representing  $A$  relative to this basis. Then equation (1.51) becomes

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} . \quad (1.52)$$

Rewriting this as

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

we see that the matrix  $\mathbf{A} - \lambda\mathbf{I}$  annihilates a nonzero vector, whence it must have zero determinant:

$$\det(\mathbf{A} - \lambda\mathbf{I}) = 0 . \quad (1.53)$$

Let  $\lambda$  be an eigenvalue of  $A$ . The set of eigenvectors of  $A$  with eigenvalue  $\lambda$ , together with the zero vector, form a vector subspace  $\mathbb{V}_\lambda$  of  $\mathbb{V}$ , known as the **eigenspace** of  $A$  with eigenvalue  $\lambda$ .



It is easy to prove this: all one needs to show is that  $\mathbb{V}_\lambda$  is closed under vector addition and scalar multiplication. Indeed, let  $\mathbf{v}$  and  $\mathbf{w}$  be eigenvectors of  $A$  with eigenvalue  $\lambda$  and let  $\alpha, \beta$  be scalars. Then

$$\begin{aligned} A(\alpha \mathbf{v} + \beta \mathbf{w}) &= \alpha A(\mathbf{v}) + \beta A(\mathbf{w}) && \text{(by L1,2)} \\ &= \alpha \lambda \mathbf{v} + \beta \lambda \mathbf{w} && \text{(by equation (1.51))} \\ &= \lambda (\alpha \mathbf{v} + \beta \mathbf{w}) , \end{aligned}$$

whence  $\alpha \mathbf{v} + \beta \mathbf{w}$  is also an eigenvector of  $A$  with eigenvalue  $\lambda$ .



That  $\mathbb{V}_\lambda$  is a subspace also follows trivially from the fact that it is the kernel of the linear transformation  $A - \lambda \mathbb{1}$ .

The dimension of the eigenspace  $\mathbb{V}_\lambda$  is called the **multiplicity** of the eigenvalue  $\lambda$ . One says that an eigenvalue  $\lambda$  is **non-degenerate** if  $\mathbb{V}_\lambda$  is one-dimensional and **degenerate** otherwise.

A linear transformation  $A : \mathbb{V} \rightarrow \mathbb{V}$  is **diagonalisable** if there exists a basis  $\{\mathbf{e}_i\}$  for  $\mathbb{V}$  made up of eigenvectors of  $A$ . In this basis, the matrix  $A$  representing  $A$  is a diagonal matrix:

$$A = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix} ,$$

where not all of the  $\lambda_i$  need be distinct. In this basis we can compute the trace and the determinant very easily. We see that

$$\begin{aligned} \text{tr}(A) &= \lambda_1 + \lambda_2 + \cdots + \lambda_N = \sum_{i=1}^N \lambda_i \\ \det(A) &= \lambda_1 \lambda_2 \cdots \lambda_N = \prod_{i=1}^N \lambda_i . \end{aligned}$$

Therefore the trace is the sum of the eigenvalues and the determinant is their product. This is independent of the basis, since both the trace and the determinant are invariants.



This has a very interesting consequence. Consider the identity:

$$\prod_{i=1}^N \exp(\lambda_i) = \exp \left( \sum_{i=1}^N \lambda_i \right) .$$

We can interpret this identity as an identity involving the diagonal matrix  $A$ :

$$\det(\exp(A)) = \exp(\operatorname{tr}(A)) ,$$

where the exponential of a matrix is defined via its Taylor series expansion:

$$\exp(A) = I + A + \frac{1}{2}A^2 + \frac{1}{3!}A^3 + \dots = \sum_{n=1}^{\infty} \frac{1}{n!} A^n ,$$

so that for a diagonal matrix, it is simply the exponential of its diagonal entries. Now notice that under a change of basis given by  $A \mapsto A'$ , where  $A'$  is given by equation (1.24),

$$\begin{aligned} \exp(A') &= \sum_{n=1}^{\infty} \frac{1}{n!} (A')^n \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} (S^{-1} A S)^n && \text{(by equation (1.24))} \\ &= \sum_{n=1}^{\infty} \frac{1}{n!} S^{-1} A^n S \\ &= S^{-1} \exp(A) S ; \end{aligned}$$

whence because the trace and determinant are invariants

$$\det \exp(A') = \exp \operatorname{tr}(A') .$$

Hence this equation is still true for diagonalisable matrices. In fact, it follows from the fact (see next section) that diagonalisable matrices are dense in the space of matrices, that this identity is true for arbitrary matrices:

$$\boxed{\det(\exp(A)) = \exp(\operatorname{tr}(A))} . \tag{1.54}$$

This is an extremely useful formula, particularly in quantum field theory and statistical mechanics, where it is usually applied to define the determinant of infinite-dimensional matrices.

## 1.4.2 Diagonalisability

Throughout this section  $\mathbb{V}$  is an  $N$ -dimensional complex vector space.

It turns out that not every linear transformation is diagonalisable, but many of the interesting ones in physics will be. In this section, which lies somewhat outside the main scope of this course, we will state the condition for a linear transformation to be diagonalisable.

Fix a basis for  $\mathbb{V}$  and let  $A$  be the matrix representing  $A$  relative to this basis. Let us define the following polynomial

$$\boxed{\chi_A(t) = \det(A - tI)} , \tag{1.55}$$

known as the **characteristic polynomial** of the matrix  $A$ . Under a change of basis, the matrix  $A$  changes to the matrix  $A'$  given by equation (1.24). The characteristic polynomial of the transformed matrix  $A'$  is given by

$$\begin{aligned}
 \chi_{A'}(t) &= \det(A' - tI) \\
 &= \det(S^{-1}AS - tI) && \text{(by equation (1.24))} \\
 &= \det(S^{-1}(A - tI)S) && \text{(since } S^{-1}IS = I) \\
 &= \det(S^{-1}) \det(A - tI) \det(S) && \text{(by equation (1.18))} \\
 &= \frac{1}{\det(S)} \chi_A(t) \det(S) \\
 &= \chi_A(t) .
 \end{aligned}$$

In other words, the characteristic polynomial is a matrix invariant and hence is a property of the linear transformation  $A$ . We will therefore define the **characteristic polynomial**  $\chi_A(t)$  of a linear transformation  $A : \mathbb{V} \rightarrow \mathbb{V}$  as the polynomial  $\chi_A(t)$  of the matrix which represents it relative to any basis. By the above calculation it does not depend on the basis.

The characteristic polynomial is a polynomial of order  $N$  where  $N$  is the complex dimension of  $\mathbb{V}$ . Its highest order term is of the form  $(-1)^N t^N$  and its zeroth order term is the determinant of  $A$ , as can be seen by evaluating  $\chi_A(t)$  at  $t = 0$ . In other words,

$$\chi_A(t) = \det(A) + \dots + (-1)^N t^N .$$

Equation (1.53) implies that every eigenvalue  $\lambda$  of  $A$  is a root of its characteristic polynomial:  $\chi_A(\lambda) = 0$ . Conversely it is possible to prove that every root of the characteristic polynomial is an eigenvalue of  $A$ ; although the multiplicities need not correspond: the multiplicity of the eigenvalue is never larger than that of the root.

This gives a method to compute the eigenvalues and eigenvectors of a linear transformation  $A$ . We simply choose a basis and find the matrix  $A$  representing  $A$ . We compute its characteristic polynomial and find its roots. For each root  $\lambda$  we solve the system of linear homogeneous equations:

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

This approach rests on the following general fact, known as the *Fundamental Theorem of Algebra*: every complex polynomial has a root. In fact, any complex polynomial of order  $N$  has  $N$  roots counted with multiplicity. In particular, the characteristic polynomial factorises into a product of monomials:

$$P(t) = (\lambda_1 - t)^{m_1} (\lambda_2 - t)^{m_2} \dots (\lambda_k - t)^{m_k} ,$$

where all the  $\lambda_i$  are distinct and where  $m_i \geq 1$  are positive integers. Clearly each  $\lambda_i$  is a root and  $m_i$  is its multiplicity. Each  $\lambda_i$  is also an eigenvalue of  $A$ , but  $m_i$  is *not* necessarily the multiplicity of the eigenvalue  $\lambda_i$ . Consider the matrix

$$A = \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix},$$

where  $a \neq 0$  is any complex number. Its characteristic polynomial is given by

$$\chi_A(t) = \det(A - tI) = \begin{vmatrix} 1-t & a \\ 0 & 1-t \end{vmatrix} = (1-t)^2 = 1 - 2t + t^2.$$

Hence the only root of this polynomial is 1 with multiplicity 2. The number 1 is also an eigenvalue of  $A$ . For example, an eigenvector  $\mathbf{v}$  is given by

$$\mathbf{v} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

However, the multiplicity of the eigenvalue 1 is only 1. Indeed, if it were 2, this would mean that there are two linearly independent eigenvectors with eigenvalue 1. These eigenvectors would then form a basis, relative to which  $A$  would be the identity matrix. But if  $A = I$  relative to some basis,  $A' = I$  relative to any other basis, since the identity matrix is invariant under change of basis. This violates the explicit expression for  $A$  above.

A result known as the **Cayley–Hamilton Theorem** states that any matrix  $A$  satisfies the following polynomial equation:

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

where  $\mathbf{0}$  means the matrix all of whose entries are zero, and where a scalar  $a$  is replaced by the scalar matrix  $aI$ . For example, consider the matrix  $A$  above:

$$\begin{aligned} \chi_A(A) &= I - 2A + A^2 \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - 2 \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix} + \begin{pmatrix} 1 & a \\ 0 & 1 \end{pmatrix}^2 \\ &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} - \begin{pmatrix} 2 & 2a \\ 0 & 2 \end{pmatrix} + \begin{pmatrix} 1 & 2a \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \end{aligned}$$

The Cayley–Hamilton theorem shows that any  $N \times N$  matrix obeys an  $N$ -th order polynomial equation. However in some cases an  $N \times N$  matrix

$A$  obeys a polynomial equation of smaller order. The polynomial  $\mu_A(t)$  of smallest order such that

$$\mu_A(A) = 0 ,$$

is called the **minimal polynomial** of the matrix  $A$ . One can show that the minimal polynomial divides the characteristic polynomial. In fact, if the characteristic polynomial has the factorisation

$$\chi_A(t) = (\lambda_1 - t)^{m_1} (\lambda_2 - t)^{m_2} \cdots (\lambda_k - t)^{m_k} ,$$

the minimal polynomial has the factorisation

$$\mu_A(t) = (\lambda_1 - t)^{n_1} (\lambda_2 - t)^{n_2} \cdots (\lambda_k - t)^{n_k} ,$$

where  $1 \leq n_i \leq m_i$ . The main result in this topic is that a matrix  $A$  is diagonalisable if and only if all  $n_i = 1$ . For the non-diagonalisable matrix above, we see that its characteristic polynomial equals its minimal polynomial, since  $A \neq I$ .

In particular this shows that if all eigenvalues of a linear transformation are non-degenerate, then the linear transformation is diagonalisable. Given any matrix, one need only perturb it infinitesimally to lift any degeneracy its eigenvalues might have. This then implies that the diagonalisable matrices are dense in the space of matrices; that is, infinitesimally close to any non-diagonalisable matrix there is one which is diagonalisable. This is key to proving many identities involving matrices. If an identity of the form  $f(A) = 0$  holds for diagonalisable matrices then it holds for any matrix *provided* that  $f$  is a *continuous* function.

Computing the minimal polynomial of a linear transformation is not an easy task, hence it is in practice not very easy to decide whether or not a given linear transformation is diagonalisable. Luckily large classes of linear transformations can be shown to be diagonalisable, as we will now discuss.

### 1.4.3 Spectral theorem for hermitian transformations

Throughout this section  $\mathbb{V}$  is an  $N$ -dimensional complex vector space with a hermitian inner product  $\langle \cdot, \cdot \rangle$ .

Let  $A : \mathbb{V} \rightarrow \mathbb{V}$  be a hermitian linear transformation:  $A^\dagger = A$ . We will show that it is diagonalisable. As a corollary we will see that unitary transformations  $U : \mathbb{V} \rightarrow \mathbb{V}$  such that  $U^\dagger \circ U = U \circ U^\dagger = \mathbb{1}$  are also diagonalisable. These results are known as the spectral theorems for hermitian and unitary transformations.

We will first need to show two key results about the eigenvalues and eigenvectors of a hermitian transformation. First we will show that the eigenvalues

of a hermitian transformation are real. Let  $\mathbf{v}$  be an eigenvector of  $A$  with eigenvalue  $\lambda$ . Then on the one hand,

$$\begin{aligned}\langle A(\mathbf{v}), \mathbf{v} \rangle &= \langle \lambda \mathbf{v}, \mathbf{v} \rangle \\ &= \lambda^* \langle \mathbf{v}, \mathbf{v} \rangle ;\end{aligned}\quad \text{(by sesquilinearity)}$$

whereas on the other hand,

$$\begin{aligned}\langle A(\mathbf{v}), \mathbf{v} \rangle &= \langle \mathbf{v}, A^\dagger(\mathbf{v}) \rangle && \text{(by equation (1.41))} \\ &= \langle \mathbf{v}, A(\mathbf{v}) \rangle && \text{(since } A \text{ is hermitian)} \\ &= \langle \mathbf{v}, \lambda \mathbf{v} \rangle \\ &= \lambda \langle \mathbf{v}, \mathbf{v} \rangle . && \text{(by HIP2)}\end{aligned}$$

Hence,

$$(\lambda - \lambda^*) \|\mathbf{v}\|^2 = 0 .$$

Since  $\mathbf{v} \neq \mathbf{0}$ , HIP3 implies that  $\|\mathbf{v}\|^2 \neq 0$ , whence  $\lambda = \lambda^*$ .

The second result is that eigenvectors corresponding to different eigenvalues are orthogonal. Let  $\mathbf{v}$  and  $\mathbf{w}$  be eigenvectors with distinct eigenvalues  $\lambda$  and  $\mu$ , respectively. Then on the one hand,

$$\begin{aligned}\langle A(\mathbf{v}), \mathbf{w} \rangle &= \langle \lambda \mathbf{v}, \mathbf{w} \rangle \\ &= \lambda \langle \mathbf{v}, \mathbf{w} \rangle .\end{aligned}\quad \text{(since } \lambda \text{ is real)}$$

On the other hand,

$$\begin{aligned}\langle A(\mathbf{v}), \mathbf{w} \rangle &= \langle \mathbf{v}, A^\dagger(\mathbf{w}) \rangle && \text{(by equation (1.41))} \\ &= \langle \mathbf{v}, A(\mathbf{w}) \rangle && \text{(since } A \text{ is hermitian)} \\ &= \langle \mathbf{v}, \mu \mathbf{w} \rangle \\ &= \mu \langle \mathbf{v}, \mathbf{w} \rangle . && \text{(by HIP2)}\end{aligned}$$

Hence,

$$(\lambda - \mu) \langle \mathbf{v}, \mathbf{w} \rangle ,$$

whence if  $\lambda \neq \mu$ ,  $\mathbf{v} \perp \mathbf{w}$ .

Now we need a basic fact: every hermitian transformation has at least one eigenvalue.



This can be shown using variational calculus. Consider the expression

$$f(\mathbf{v}) \equiv \langle \mathbf{v}, A(\mathbf{v}) \rangle .$$

We claim that  $f(\mathbf{v})$  is a real number:

$$\begin{aligned}
f(\mathbf{v})^* &= \langle \mathbf{v}, A(\mathbf{v}) \rangle^* \\
&= \langle A(\mathbf{v}), \mathbf{v} \rangle && \text{(by HIP1)} \\
&= \langle \mathbf{v}, A^\dagger(\mathbf{v}) \rangle && \text{(by equation (1.41))} \\
&= \langle \mathbf{v}, A(\mathbf{v}) \rangle && \text{(since } A \text{ is hermitian)} \\
&= f(\mathbf{v}) .
\end{aligned}$$

Therefore  $f$  defines a continuous quadratic function from  $\mathbb{V}$  to the real numbers. We would like to extremise this function. Clearly,

$$f(\alpha \mathbf{v}) = |\alpha|^2 f(\mathbf{v}) ,$$

and this means that by rescaling  $\mathbf{v}$  we can make  $f(\mathbf{v})$  be as large or as small as we want. This is not the type of extremisation that we are interested: we want to see in which direction is  $f(\mathbf{v})$  extremal. One way to do this is to restrict ourselves to vectors such that  $\|\mathbf{v}\|^2 = 1$ . This can be imposed using a Lagrange multiplier  $\lambda$ . Extremising  $f(\mathbf{v})$  subject to the constraint  $\|\mathbf{v}\|^2 = 1$ , can be done by extremising the expression

$$I(\mathbf{v}, \lambda) = f(\mathbf{v}) - \lambda (\|\mathbf{v}\|^2 - 1) .$$

The variation of  $I$  yields the following expression:

$$\delta I = 2 \langle \delta \mathbf{v}, (A - \lambda I) \mathbf{v} \rangle - \delta \lambda (\|\mathbf{v}\|^2 - 1) .$$

Therefore the variational equations are  $\|\mathbf{v}\|^2 = 1$  and

$$A \mathbf{v} = \lambda \mathbf{v} ,$$

where we have used the non-degeneracy of the inner product and the fact that we want  $\delta I = 0$  for all  $\delta \lambda$  and  $\delta \mathbf{v}$ . Therefore this says that the extrema of  $I$  are the pairs  $(\mathbf{v}, \lambda)$  where  $\mathbf{v}$  is a normalised eigenvalue of  $A$  with eigenvalue  $\lambda$ . The function  $I(\mathbf{v}, \lambda)$  takes the value  $I(\mathbf{v}, \lambda) = \lambda$  at such a pair; whence the maxima and minima correspond to the largest and smallest eigenvalues. It remains to argue that the variational problem has solution. This follows from the compactness of the space of normalised vectors, which is the unit sphere in  $\mathbb{V}$ . The function  $f(\mathbf{v})$  is continuous on the unit sphere and hence attains its maxima and minima in it.

We are now ready to prove the spectral theorem. We will first assume that the eigenvalues are non-degenerate, for ease of exposition and then we will relax this hypothesis and prove the general result.

Let  $\mathbf{v}_1$  be a normalised eigenvector of  $A$  with eigenvalue  $\lambda_1$ . It exists from the above discussion and it is the only such eigenvector, up to scalar multiplication, by the non-degeneracy hypothesis. The eigenvalue is real as we saw above. Choose vectors  $\{\mathbf{e}_2, \mathbf{e}_3, \dots\}$  such that  $\{\mathbf{v}_1, \mathbf{e}_2, \dots\}$  is a basis for  $\mathbb{V}$  and apply the Gram–Schmidt procedure if necessary so that it is a unitary basis. Let us look at the matrix  $\mathbf{A}$  of  $A$  in such a basis. Because  $\mathbf{e}_1$  is an eigenvector, one has

$$\langle A(\mathbf{v}_1), \mathbf{e}_j \rangle = \langle \lambda_1 \mathbf{v}_1, \mathbf{e}_j \rangle = \lambda_1 \langle \mathbf{v}_1, \mathbf{e}_j \rangle = 0 ,$$

and similarly

$$\langle A(\mathbf{e}_j), \mathbf{v}_1 \rangle = \langle \mathbf{e}_j, A(\mathbf{v}_1) \rangle = \langle \mathbf{e}_j, \lambda_1 \mathbf{v}_1 \rangle = \lambda_1 \langle \mathbf{e}_j, \mathbf{v}_1 \rangle = 0 .$$

Moreover

$$\langle \mathbf{v}_1, A(\mathbf{v}_1) \rangle = \langle \mathbf{v}_1, \lambda_1 \mathbf{v}_1 \rangle = \lambda_1 \|\mathbf{e}_1\|^2 = \lambda_1 .$$

This means that the matrix takes the form

$$\begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & A_{22} & \cdots & A_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & A_{N2} & \cdots & A_{NN} \end{pmatrix} . \quad (1.56)$$

The submatrix

$$\begin{pmatrix} A_{22} & \cdots & A_{2N} \\ \vdots & \ddots & \vdots \\ A_{N2} & \cdots & A_{NN} \end{pmatrix} ,$$

is still hermitian, since for  $i, j = 2, \dots, N$ ,

$$A_{ij} = \langle \mathbf{e}_i, A(\mathbf{e}_j) \rangle = \langle A(\mathbf{e}_j), \mathbf{e}_i \rangle^* = \langle \mathbf{e}_j, A(\mathbf{e}_i) \rangle^* = A_{ji}^* .$$

Now we can apply the procedure again to this  $(N-1) \times (N-1)$  matrix: we find a normalised eigenvector  $\mathbf{v}_2$ , which by assumption corresponds to a non-degenerate eigenvalue  $\lambda_2$ . Starting with this eigenvector we build a unitary basis  $\{\mathbf{v}_2, \mathbf{e}'_3, \dots\}$  for the  $(N-1)$ -dimensional subspace spanned by the  $\{\mathbf{e}_2, \mathbf{e}_3, \dots\}$ . The submatrix  $\mathbf{A}^{(N-1)}$  then takes the form analogous to the one in equation (1.56), leaving an  $(N-2) \times (N-2)$  submatrix which is again still hermitian. We can apply the same procedure to this smaller matrix, and so on until we are left with a  $1 \times 1$  hermitian matrix, i.e., a real number:  $\lambda_N$ . The basis  $\{\mathbf{v}_i\}$  formed by the eigenvectors is clearly unitary, since each  $\mathbf{v}_i$  is normalised by definition and is orthogonal to the preceding  $\{\mathbf{v}_{j < i}\}$  by the way they were constructed. The matrix of  $A$  relative to this basis is then

$$\mathbf{A} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix} ,$$

with real eigenvalues  $\lambda_i$ .

The case with degenerate eigenvalues works along similar lines. We start with an eigenvalue  $\lambda_1$  and consider the eigenspace  $\mathbb{V}_{\lambda_1}$ . It may be that the

dimension  $m_1$  of  $\mathbb{V}_{\lambda_1}$  is larger than 1. In any case, every vector in  $\mathbb{V}_{\lambda_1}$  is an eigenvector of  $A$ . Use Gram–Schmidt to find a unitary basis  $\{\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_{m_1}\}$  for  $\mathbb{V}_{\lambda_1}$ . Complete this basis to a unitary basis  $\{\mathbf{v}_1, \dots, \mathbf{v}_{m_1}, \mathbf{e}_{m_1+1}, \dots, \mathbf{e}_N\}$  for  $\mathbb{V}$ , which can be done using Gram–Schmidt if necessary again. The matrix  $A$  representing  $A$  in this basis is given by

$$A = \begin{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_1 \end{pmatrix} & \\ & \begin{pmatrix} A_{m_1+1,m_1+1} & \cdots & A_{m_1+1,N} \\ \vdots & \ddots & \vdots \\ A_{N,m_1+1} & \cdots & A_{NN} \end{pmatrix} \end{pmatrix},$$

where the off-diagonal blocks have vanishing entries because

$$\langle \mathbf{e}_i, A(\mathbf{v}_j) \rangle = \lambda_1 \langle \mathbf{e}_i, \mathbf{v}_j \rangle = 0.$$

The submatrix

$$\begin{pmatrix} A_{m_1+1,m_1+1} & \cdots & A_{m_1+1,N} \\ \vdots & \ddots & \vdots \\ A_{N,m_1+1} & \cdots & A_{NN} \end{pmatrix}$$

is again hermitian, so we can apply the procedure again to it, until we are left with a basis  $\{\mathbf{v}_i\}$  of eigenvectors of  $A$ , so that the matrix is diagonal.

In summary, suppose that we start with a hermitian matrix  $A$ , thought of as the matrix of a hermitian linear transformation  $A$  relative to a unitary basis. Then the above iterative procedure produces a unitary basis relative to which the matrix for  $A$  is diagonal. Because the initial and final basis are unitary, the change of basis transformation  $U$  is unitary. In other words, given a hermitian matrix  $A$  there is a unitary matrix  $U$  such that

$$A' = U^{-1} A U = (U^*)^t A U$$

is diagonal. In other words,

Every hermitian matrix can be diagonalised by a unitary transformation.

In fact the unitary matrix  $U$  above can be written down explicitly in terms of the normalised eigenvectors of  $A$ . Let  $\{\mathbf{v}_i\}$  be a set of normalised eigenvectors which are mutually orthogonal. This is guaranteed if they have

different eigenvalues, and in the case of degenerate eigenvalues by Gram–Schmidt. Consider the matrix

$$\mathbf{U} = \begin{pmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N \\ \downarrow & \downarrow & \cdots & \downarrow \end{pmatrix} .$$

We claim first of all that  $\mathbf{U}$  is unitary. Indeed,

$$(\mathbf{U}^*)^t = \begin{pmatrix} \leftarrow & (\mathbf{v}_1^*)^t & \rightarrow \\ \leftarrow & (\mathbf{v}_2^*)^t & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & (\mathbf{v}_N^*)^t & \rightarrow \end{pmatrix} .$$

Hence

$$\begin{aligned} (\mathbf{U}^*)^t \mathbf{U} &= \begin{pmatrix} \leftarrow & (\mathbf{v}_1^*)^t & \rightarrow \\ \leftarrow & (\mathbf{v}_2^*)^t & \rightarrow \\ \vdots & \vdots & \vdots \\ \leftarrow & (\mathbf{v}_N^*)^t & \rightarrow \end{pmatrix} \begin{pmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N \\ \downarrow & \downarrow & \cdots & \downarrow \end{pmatrix} \\ &= \begin{pmatrix} (\mathbf{v}_i^*)^t \mathbf{v}_j \end{pmatrix} \\ &= \mathbf{I} , \end{aligned}$$

since the  $\{\mathbf{v}_i\}$  form a unitary basis. Moreover, paying attention to the way matrix multiplication is defined and using that the  $\{\mathbf{v}_i\}$  are eigenvectors of  $\mathbf{A}$ , we find

$$\begin{aligned} \mathbf{A} \mathbf{U} &= \begin{pmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \lambda_1 \mathbf{v}_1 & \lambda_2 \mathbf{v}_2 & \cdots & \lambda_N \mathbf{v}_N \\ \downarrow & \downarrow & \cdots & \downarrow \end{pmatrix} \\ &= \begin{pmatrix} \uparrow & \uparrow & \cdots & \uparrow \\ \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_N \\ \downarrow & \downarrow & \cdots & \downarrow \end{pmatrix} \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix} \\ &= \mathbf{U} \mathbf{A}' . \end{aligned}$$

In other words,  $\mathbf{A}' = \mathbf{U}^{-1} \mathbf{A} \mathbf{U}$  just as above.

There is a real version of this result: if  $\mathbf{A}$  is real and symmetric, then it is hermitian. We can diagonalise it with a unitary transformation, which is

also real, whence it is orthogonal. This then yields the spectral theorem for symmetric real matrices, which says that any real symmetric matrix can be diagonalised by an orthogonal transformation.

We can now understand the positive-definiteness condition on the matrix representing an inner product on a vector space. We saw in Section 1.3.1 that the matrix  $\mathbf{G}$  of an inner product in a real vector space is symmetric. Hence it can be diagonalised by an orthogonal transformation. From equation (1.35), it follows that there is a basis relative to which the matrix of the inner product is diagonal. Let  $\{\mathbf{e}_i\}$  be such a basis and let  $\langle \mathbf{e}_i, \mathbf{e}_j \rangle = \lambda_i \delta_{ij}$ . If  $\mathbf{v} = \sum_i v_i \mathbf{e}_i$  is any vector, then

$$\|\mathbf{v}\|^2 = \langle \mathbf{v}, \mathbf{v} \rangle = \sum_i \lambda_i v_i^2 .$$

Axiom IP3 says that this quantity has to be positive for all nonzero vectors  $\mathbf{v}$ , which clearly implies that  $\lambda_i > 0$  for all  $i$ . Therefore a symmetric matrix is positive definite if and only if all its eigenvalues are positive. A similar statement also holds for hermitian inner products, whose proof is left as an exercise.



It is not just hermitian matrices that can be diagonalised by unitary transformations. Let us say that a linear transformation  $N$  is **normal** if it commutes with its adjoint

$$N^\dagger \circ N = N \circ N^\dagger . \tag{1.57}$$

Then it can be proven that a  $N$  is diagonalisable by a unitary transformation. As an example consider the  $3 \times 3$  matrix

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

considered in the Exercises. We saw that it was diagonalisable by a unitary transformation, yet it is clearly not hermitian. Nevertheless it is easy to check that it is normal. Indeed,

$$(\mathbf{P}^*)^t = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} ,$$

so that

$$\mathbf{P} (\mathbf{P}^*)^t = (\mathbf{P}^*)^t \mathbf{P} = \mathbf{I} ;$$

in other words, it is unitary.

It follows from the spectral theorem for hermitian transformations that unitary transformations can also be diagonalised by unitary transformations. This is known as the **Cayley transformation**, which is discussed in detail in the Problems. It follows from the Cayley transformation that the eigenvalues of a unitary matrix take values in the unit circle in the complex plane. This can also be seen directly as follows. Let  $U$  be a unitary transformation and

let  $\mathbf{v}$  be an eigenvector with eigenvalue  $\lambda$ . Then consider  $\|U(\mathbf{v})\|^2$ . Because  $U$  is unitary,

$$\|U(\mathbf{v})\|^2 = \|\mathbf{v}\|^2 ,$$

but because  $\mathbf{v}$  is an eigenvector,

$$\|U(\mathbf{v})\|^2 = \|\lambda \mathbf{v}\|^2 = |\lambda|^2 \|\mathbf{v}\|^2 ,$$

whence  $|\lambda|^2 = 1$ .

Spectral theorems are extremely powerful in many areas of physics and mathematics, and in the next sections we will discuss two such applications. However the real power of the spectral theorem manifests itself in quantum mechanics, although the version of the theorem used there is the one for self-adjoint operators in an infinite-dimensional Hilbert space, which we will not have the opportunity to discuss in this course.

#### 1.4.4 Application: quadratic forms

In this section we discuss a mathematical application of the spectral theorem for real symmetric transformations.

Let us start with the simplest case of a two-dimensional quadratic form. By a **quadratic form** on two variables  $(x_1, x_2)$  we mean a quadratic polynomial of the form

$$Q(x_1, x_2) = ax_1^2 + 2bx_1x_2 + cx_2^2 , \quad (1.58)$$

for some real constants  $a, b, c$ . By a **quadric** we mean the solutions  $(x_1, x_2)$  of an equation of the form

$$Q(x_1, x_2) = d ,$$

where  $d$  is some real number and  $Q$  is a quadratic form. For example, we can take

$$Q_1(x_1, x_2) = x_1^2 + x_2^2 ,$$

in which case the quadrics  $Q_1(x_1, x_2) = d$  for  $d > 0$  describe a circle of radius  $\sqrt{d}$  in the plane coordinatised by  $(x_1, x_2)$ . To investigate the type of quadric that a quadratic form gives rise to, it is convenient to diagonalise it: that is, change to coordinates  $(y_1, y_2)$  for which the mixed term  $y_1 y_2$  in the quadratic form is not present. To tie this to the spectral theorem, it is convenient to rewrite this in terms of matrices. In terms of the column vector  $\mathbf{x} = (x_1, x_2)^t$ , the general two-dimensional quadratic form in equation (1.58) can be written as

$$Q(x_1, x_2) = \mathbf{x}^t \mathbf{Q} \mathbf{x} ,$$

where  $\mathbf{Q}$  is the matrix

$$\mathbf{Q} = \begin{pmatrix} a & b \\ b & c \end{pmatrix} .$$

Because  $\mathbf{Q}$  is symmetric, it can be diagonalised by an orthogonal transformation which is built out of the normalised eigenvectors as was explained in the previous section. Hence there is an orthogonal matrix  $\mathbf{O}$  such that  $\mathbf{Q} = \mathbf{O} \mathbf{D} \mathbf{O}^t$ , where  $\mathbf{D}$  is a diagonal matrix with entries  $\lambda_i$ , for  $i = 1, 2$ . That means that in terms of the new coordinates

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{O}^t \mathbf{x} ,$$

the quadratic form is diagonal

$$Q(y_1, y_2) = \mathbf{y}^t \mathbf{D} \mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 .$$

We can further rescale the coordinates  $\{y_i\}$ :  $z_i = \mu_i y_i$ , where  $\mu_i$  is real. This means that relative to the new coordinates  $z_i$ , the quadratic form takes the form

$$Q(z_1, z_2) = \varepsilon_1 z_1^2 + \varepsilon_2 z_2^2 ,$$

where  $\varepsilon_i$  are  $0, \pm 1$ .

We can distinguish three types of quadrics, depending on the relative signs of the eigenvalues:

1. ( $\varepsilon_1 \varepsilon_2 = 1$ ) In this case the eigenvalues have the same sign and the quadric is an ellipse.
2. ( $\varepsilon_1 \varepsilon_2 = -1$ ) In this case the eigenvalues have different sign and the quadric is a hyperbola.
3. ( $\varepsilon_1 \varepsilon_2 = 0$ ) In this case one of the eigenvalues is zero, and the quadric consists of a pair of lines.

The general case is not much more complicated. Let  $\mathbb{V}$  be a real vector space of dimension  $N$  with an inner product. By a **quadratic form** we mean a symmetric bilinear form  $Q : \mathbb{V} \times \mathbb{V} \rightarrow \mathbb{R}$ . In other words,  $Q$  satisfies axioms IP1 and IP2 of an inner product, but IP3 need *not* be satisfied. Associated to every quadratic form there is a linear transformation in  $\mathbb{V}$ , which we also denote  $Q$ , defined as follows

$$\langle \mathbf{v}, Q(\mathbf{w}) \rangle = Q(\mathbf{v}, \mathbf{w}) .$$

Symmetry of the bilinear form implies that the linear transformation  $Q$  is also symmetric:

$$\langle \mathbf{v}, Q(\mathbf{w}) \rangle = Q(\mathbf{v}, \mathbf{w}) = Q(\mathbf{w}, \mathbf{v}) = \langle \mathbf{w}, Q(\mathbf{v}) \rangle = \langle Q(\mathbf{v}), \mathbf{w} \rangle .$$

Hence it can be diagonalised by an orthogonal transformation. Relative to an orthonormal basis  $\{\mathbf{e}_i\}$  for  $\mathbb{V}$ ,  $Q$  is represented by a symmetric matrix  $\mathbf{Q}$ . Let  $\mathbf{O}$  be an orthogonal matrix which diagonalises  $\mathbf{Q}$ ; that is,  $\mathbf{Q} = \mathbf{O} \mathbf{D} \mathbf{O}^t$ , with  $\mathbf{D}$  diagonal.

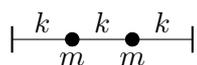
We can further change basis to an orthogonal basis whose elements are however no longer normalised, in such a way that the resulting matrix  $\mathbf{D}'$  is still diagonal with all its entries either  $\pm 1$  or  $0$ . Let  $(n_+, n_-, n_0)$  denote, respectively, the number of positive, negative and zero diagonal entries of  $\mathbf{D}'$ . There is a result, known as **Sylvester's Law of Inertia**, which says that the numbers  $(n_+, n_-, n_0)$  are an invariant of the quadratic form, so that they can be computed from the matrix of the quadratic form relative to any basis. A quadratic form is said to be **non-degenerate** if  $n_0 = 0$ . It is said to be **positive-definite** if  $n_- = n_0 = 0$ , and **negative-definite** if  $n_+ = n_0 = 0$ . Clearly a quadratic form is an inner product when it is positive-definite. A non-degenerate quadratic form, which is not necessarily positive- or negative-definite, defines a **generalised inner product** on  $\mathbb{V}$ . There are two integers which characterise a non-degenerate quadratic form: the dimension  $N$  of the vector space, and the **signature**  $n_+ - n_-$ . Notice that if the signature is bounded above by the dimension: the bound being saturated when the quadratic form is positive-definite. There are plenty of interesting non-degenerate quadratic forms which are not positive-definite. For example, Minkowski spacetime in the theory of special relativity possesses a quadratic form with dimension 4 and signature 2.

### 1.4.5 Application: normal modes

This section discusses the powerful method of normal modes to decouple interacting mechanical systems near equilibrium. It is perhaps not too exaggerated to suggest that theoretical physicists spend a large part of their lives studying the problem of normal modes in one way or another.

We start with a simple example.

Consider an idealised one-dimensional mechanical system consisting of two point masses each of mass  $m$  connected by springs to each other and to two fixed ends. We will neglect gravity, friction and the mass of the springs. The springs obey Hooke's law with spring constant  $k$ . We assume that the system is at equilibrium when the springs are relaxed, and we want to study the system around



equilibrium; that is, we wish to study small displacements of the masses. We let  $x_i$  for  $i = 1, 2$  denote the displacements from equilibrium for each of the two point masses, as shown below.



Then the potential energy due to the springs is the sum of the potential energies of each of the springs:

$$\begin{aligned} V &= \frac{1}{2}k x_1^2 + \frac{1}{2}k (x_2 - x_1)^2 + \frac{1}{2}k x_2^2 \\ &= k (x_1^2 + x_2^2 - x_1 x_2) . \end{aligned}$$

The kinetic energy is given by

$$T = \frac{1}{2}m\dot{x}_1^2 + \frac{1}{2}m\dot{x}_2^2 .$$

The equations of motion are then, for  $i = 1, 2$ ,

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{x}_i} = - \frac{\partial V}{\partial x_i} .$$

Explicitly, we have the following coupled system of second order ordinary differential equations:

$$\begin{aligned} m\ddot{x}_1 &= -2kx_1 + kx_2 \\ m\ddot{x}_2 &= -2kx_2 + kx_1 . \end{aligned}$$

Let us write this in matrix form. We introduce a column vector  $\mathbf{x}^t = (x_1, x_2)$ . Then the above system of equations becomes

$$\ddot{\mathbf{x}} = -\omega^2 \mathbf{K} \mathbf{x} , \tag{1.59}$$

where  $\mathbf{K}$  is the matrix

$$\mathbf{K} = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix} ,$$

and where we have introduced the notation

$$\omega \equiv \sqrt{\frac{k}{m}} .$$

Notice that  $\mathbf{K}$  is symmetric, hence it can be diagonalised by an orthogonal transformation. Let us find its eigenvalues and its eigenvectors. The characteristic polynomial of  $\mathbf{K}$  is given by

$$\chi_{\mathbf{K}}(\lambda) = \begin{vmatrix} 2 - \lambda & -1 \\ -1 & 2 - \lambda \end{vmatrix} = (2 - \lambda)^2 - 1 = (\lambda - 1)(\lambda - 3) ,$$

from which it follows that it has as roots  $\lambda = 1, 3$ . The normalised eigenvectors corresponding to these eigenvalues are

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \quad \text{and} \quad \mathbf{v}_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ -1 \end{pmatrix} ,$$

respectively. We build the following matrix  $\mathbf{O}$  out of the normalised eigenvectors

$$\mathbf{O} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} .$$

One can check that  $\mathbf{O}$  is orthogonal:  $\mathbf{O}^t = \mathbf{O}^{-1}$ . One can also check that

$$\mathbf{K} = \mathbf{O} \mathbf{D} \mathbf{O}^t ,$$

where  $\mathbf{D}$  is the diagonal matrix

$$\mathbf{D} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix} .$$

Inserting this expression into equation (1.59), we see that

$$\ddot{\mathbf{x}} = -\omega^2 \mathbf{O} \mathbf{D} \mathbf{O}^t \mathbf{x} .$$

In terms of the new variables

$$\mathbf{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \mathbf{O}^t \mathbf{x} ,$$

the equation of motion (1.59) becomes

$$\ddot{\mathbf{y}} = -\omega^2 \mathbf{D} \mathbf{y} . \tag{1.60}$$

Because the matrix  $\mathbf{D}$  is diagonal, the equations of motion for the new variables  $y_i$  are now decoupled:

$$\ddot{y}_1 = -\omega^2 y_1 \quad \text{and} \quad \ddot{y}_2 = -3\omega^2 y_2 .$$

One can now easily solve these equations,

$$\begin{aligned}y_1(t) &= A_1 \cos(\omega_1 t + \varphi_1) \\y_2(t) &= A_2 \cos(\omega_2 t + \varphi_2) ,\end{aligned}$$

where  $\omega_1 = \omega$ ,  $\omega_2 = \sqrt{3}\omega$  and  $A_i$  and  $\varphi_i$  are constants to be determined from the initial conditions. The physical variables in the original problem are the displacements  $x_i$  of each of the point masses. They can be found in terms of the new decoupled variables  $y_i$  simply by inverting the change of variables (1.60). Explicitly,

$$\begin{aligned}x_1(t) &= \frac{A_1}{\sqrt{2}} \cos(\omega_1 t + \varphi_1) + \frac{A_2}{\sqrt{2}} \cos(\omega_2 t + \varphi_2) \\x_2(t) &= \frac{A_1}{\sqrt{2}} \cos(\omega_1 t + \varphi_1) - \frac{A_2}{\sqrt{2}} \cos(\omega_2 t + \varphi_2) .\end{aligned}$$

Variables like the  $y_i$  which decouple the equations of motion are called the **normal modes** of the mechanical system. Their virtue is that they reduce an interacting (i.e., coupled) mechanical system around equilibrium to a set of independent free oscillators. Each of these free oscillators are mathematical constructs: the normal modes do not generally correspond to the motion of any of the masses in the original system, but they nevertheless possess a certain “physicality” and it is fruitful to work with them as if they were physical. The original physical variables can then be understood as linear combinations of the normal modes as we saw above. The frequencies  $\omega_i$  of the normal modes are known as the **characteristic frequencies** of the mechanical system. In particle physics, for example, the elementary particles are the normal modes and their masses are the characteristic frequencies.

To illustrate the simplification in the dynamics which results from considering the normal modes, in Figure 1.1 we have sketched the motion of the two masses in the problem and of the two normal modes, with time running horizontally to the right.

Notice also that although the motion of each of the normal modes is periodic, the system as a whole is not. This is due to the fact that the characteristic frequencies are not rational multiples of each other.



Let us see this. Suppose that we have two oscillators with frequencies  $\omega_1$  and  $\omega_2$ . That means that the oscillators are periodic with periods  $T_1 = 2\pi/\omega_1$  and  $T_2 = 2\pi/\omega_2$ . The combined system will be periodic provided that  $N_1 T_1 = N_2 T_2$  for some integers  $N_i$ . But this means that

$$\frac{\omega_1}{\omega_2} = \frac{N_1}{N_2} ,$$

which is a rational number. In the problem treated above, the ratio

$$\frac{\omega_1}{\omega_2} = \frac{1}{\sqrt{3}} = \frac{\sqrt{3}}{3}$$

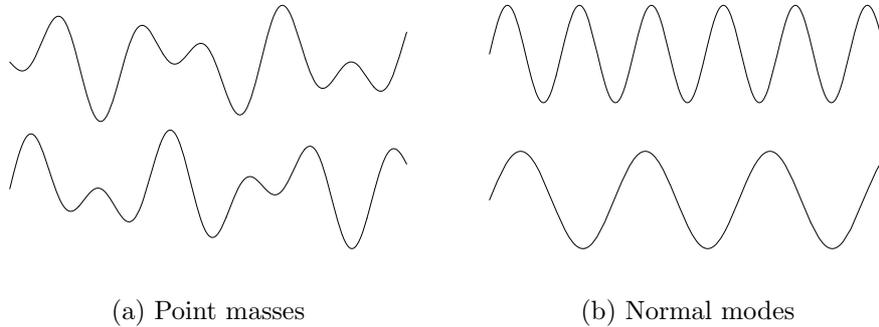


Figure 1.1: Dynamics of point masses and normal modes.

is irrational. Therefore the motion is aperiodic.

If we were to plot the trajectory of the system in the plane, with the trajectory of one of the point masses along the  $x$ -axis and the trajectory of the other point mass along the  $y$ -axis, we see that the orbit never repeats, and that we end up filling up the available configuration space. In Figure 1.2 we have plotted the cumulative trajectory of the system after letting it run for  $T$  units of time, for different values of  $T$ . As you can see, as  $T$  grows the system has visited more and more points in the available configuration space. Asymptotically, as  $T \rightarrow \infty$ , the system will have visited the whole available space.

### 1.4.6 Application: near equilibrium dynamics

In this section we will consider a more general mechanical system near equilibrium. Throughout the section  $\mathbb{V}$  will be a real finite-dimensional vector space with an inner product.

Consider a mechanical system whose configuration space is  $\mathbb{V}$ . For example, it could be a system of  $n$  point particles in  $d$  dimensions, and then  $\mathbb{V}$  would be an  $(nd)$ -dimensional vector space. In the previous section we discussed the case of a one-dimensional system consisting of two point particles, so that  $\mathbb{V}$  was two-dimensional. In the Problems we looked at systems with three-dimensional  $\mathbb{V}$ . In this section we are letting  $\mathbb{V}$  be arbitrary but finite-dimensional.

The potential energy is given by a function  $V : \mathbb{V} \rightarrow \mathbb{R}$ . The configurations of mechanical equilibrium are those for which the gradient of the potential vanishes. Hence let us consider one such equilibrium configuration  $\mathbf{q}_0 \in \mathbb{V}$ :

$$\nabla V|_{\mathbf{q}_0} = \mathbf{0} .$$

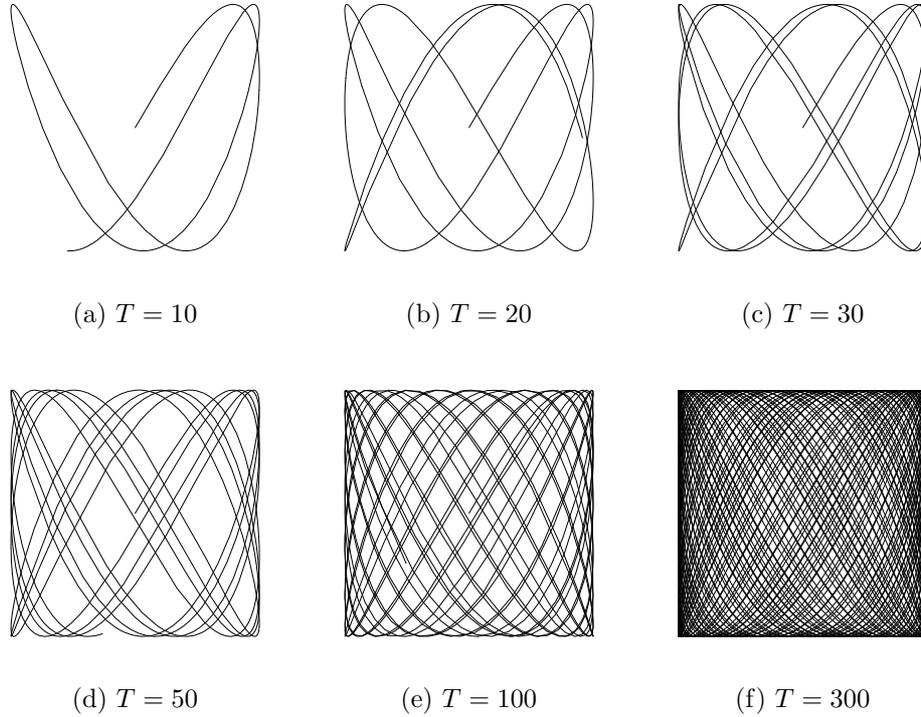


Figure 1.2: Trajectory of the mechanical system at different times.

Because the potential energy is only defined up to an additive constant, we are free to choose it such that  $V(\mathbf{q}_0) = 0$ . We can therefore expand the potential function  $V$  about  $\mathbf{q}_0$  and the first contribution will be quadratic:

$$\begin{aligned} V(\mathbf{q}) &= V(\mathbf{q}_0) + \langle \nabla V|_{\mathbf{q}_0}, \mathbf{q} - \mathbf{q}_0 \rangle + \frac{1}{2} \langle \mathbf{q} - \mathbf{q}_0, H(\mathbf{q} - \mathbf{q}_0) \rangle + \dots \\ &= \frac{1}{2} \langle \mathbf{q} - \mathbf{q}_0, H(\mathbf{q} - \mathbf{q}_0) \rangle, \end{aligned}$$

where  $H : \mathbb{V} \rightarrow \mathbb{V}$  is a symmetric linear transformation known as the **Hessian** of  $V$  at  $\mathbf{q}_0$ . Explicitly, if we choose an orthonormal basis  $\{\mathbf{e}_i\}$  for  $\mathbb{V}$ , then let  $\mathbf{q} = \sum_i q_i \mathbf{e}_i$  define some coordinates  $q_i$  for the configuration space. Then relative to this basis the Hessian of  $V$  has matrix elements

$$H_{ij} = \langle \mathbf{e}_i, H(\mathbf{e}_j) \rangle = \left. \frac{\partial^2 V}{\partial q_i \partial q_j} \right|_{\mathbf{q}_0},$$

which shows manifestly that it is symmetric:  $H_{ij} = H_{ji}$ . Let us define  $\mathbf{x} = \mathbf{q} - \mathbf{q}_0$  to be the displacements about equilibrium. These will be our dynamical variables. The potential energy in the quadratic approximation is given by

$$V = \frac{1}{2} \langle \mathbf{x}, H(\mathbf{x}) \rangle.$$

We will make the assumption that the kinetic energy is quadratic in the velocities  $\dot{\mathbf{x}}$ :

$$T = \frac{1}{2} \langle \dot{\mathbf{x}}, M(\dot{\mathbf{x}}) \rangle ,$$

where the **mass matrix**  $M$  is assumed to be symmetric and positive-definite; that is, all its eigenvalues are positive.

We will now analyse the dynamics of small displacements from equilibrium following the following prescription:

1. we will standardise the kinetic energy by diagonalising and normalising the mass matrix; and
2. we will then diagonalise the potential energy and solve for the normal modes and characteristic frequencies of this system.

Both steps make use of the spectral theorem for symmetric transformations. To do the first step notice that relative to an orthonormal basis  $\{\mathbf{e}_i\}$  for  $\mathbb{V}$ ,  $\mathbf{x} = \sum_i x_i \mathbf{e}_i$  and we can form a column vector

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

out of the components of  $\mathbf{x}$ . Relative to this basis, the mass matrix  $M$  and the Hessian  $H$  have matrices  $\mathbf{M}$  and  $\mathbf{H}$ , respectively. By assumption both are symmetric, and  $\mathbf{M}$  is in addition positive-definite. The kinetic and potential energies become

$$T = \frac{1}{2} \dot{\mathbf{x}}^t \mathbf{M} \dot{\mathbf{x}} \quad \text{and} \quad V = \frac{1}{2} \mathbf{x}^t \mathbf{H} \mathbf{x} .$$

Because  $\mathbf{M}$  is symmetric, there is an orthogonal matrix  $\mathbf{O}_1$  such that  $\mathbf{M}' = \mathbf{O}_1^t \mathbf{M} \mathbf{O}_1$  is diagonal with positive entries. Let  $\mathbf{D}_1$  be the diagonal matrix whose entries are the (positive) square roots of the diagonal entries of  $\mathbf{M}'$ . In other words,  $\mathbf{M}' = \mathbf{D}_1^2$ . We can therefore write

$$\mathbf{M} = \mathbf{O}_1 \mathbf{D}_1^2 \mathbf{O}_1^t = (\mathbf{O}_1 \mathbf{D}_1) (\mathbf{O}_1 \mathbf{D}_1)^t ,$$

where we have used that  $\mathbf{D}_1^t = \mathbf{D}_1$  since it is diagonal. Introduce then the following variables

$$\mathbf{y} = (\mathbf{O}_1 \mathbf{D}_1)^t \mathbf{x} = \mathbf{D}_1 \mathbf{O}_1^t \mathbf{x} .$$

We can invert this change of variables as follows:

$$\mathbf{x} = \mathbf{O}_1 \mathbf{D}_1^{-1} \mathbf{y} ,$$

where we have used that  $\mathbf{O}_1$  is orthogonal, so that  $\mathbf{O}_1^t = \mathbf{O}_1^{-1}$ . This change of variables accomplishes the first step outlined above, since in terms of  $\mathbf{y}$ , the kinetic energy becomes simply

$$T = \frac{1}{2} \dot{\mathbf{y}}^t \dot{\mathbf{y}} = \frac{1}{2} \|\dot{\mathbf{y}}\|^2 .$$

Similarly, the potential energy has become

$$V = \frac{1}{2} \mathbf{y}^t \mathbf{K} \mathbf{y} ,$$

where the matrix  $\mathbf{K}$  is defined by

$$\mathbf{K} = \mathbf{D}_1^{-1} \mathbf{O}_1^t \mathbf{H} \mathbf{O}_1 \mathbf{D}_1^{-1} ,$$

which is clearly symmetric since  $\mathbf{H}$  and  $\mathbf{D}_1$  are. Therefore we can find a second orthogonal matrix  $\mathbf{O}_2$  such that  $\mathbf{O}_2^t \mathbf{K} \mathbf{O}_2$  is diagonal; call this matrix  $\mathbf{D}$ . Let us define a new set of variables

$$\mathbf{z} = \mathbf{O}_2^t \mathbf{y} ,$$

relative to which the kinetic energy remains simple

$$T = \frac{1}{2} \|\mathbf{O}_2 \dot{\mathbf{z}}\|^2 = \frac{1}{2} \|\dot{\mathbf{z}}\|^2 ,$$

since orthogonal matrices preserve norms, and the potential energy diagonalises

$$V = \frac{1}{2} \mathbf{z}^t \mathbf{D} \mathbf{z} .$$

Because  $\mathbf{D}$  is diagonal, the equations of motion of the  $\mathbf{z}$  are decoupled:

$$\ddot{\mathbf{z}} = -\mathbf{D} \mathbf{z} ,$$

whence the  $\mathbf{z}$  are the normal modes of the system. Let  $\mathbf{D}$  have entries

$$\mathbf{D} = \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_N \end{pmatrix} ,$$

Then the equations of motion for the normal modes are

$$\ddot{z}_i = -\lambda_i z_i .$$

We can distinguish three types of solutions:

1. ( $\lambda_i > 0$ ) The solution is oscillatory with characteristic frequency  $\omega_i = \sqrt{\lambda_i}$ :

$$z_i(t) = A_i \cos(\omega_i t + \varphi_i) .$$

2. ( $\lambda_i = 0$ ) The solution is linear

$$z_i(t) = a_i + b_i t .$$

Such a normal mode is said to be a **zero mode**, since it has zero characteristic frequency.

3. ( $\lambda_i < 0$ ) The solution is exponential

$$z_i(t) = A_i \exp\left(\sqrt{|\lambda_i|}t\right) + B_i \exp\left(-\sqrt{|\lambda_i|}t\right) .$$

If all eigenvalues  $\lambda_i$  are positive the equilibrium point is said to be **stable**, if they are all non-negative then it is **semi-stable**, whereas if there is a negative eigenvalue, then the equilibrium is **unstable**. The signs of the eigenvalues of the matrix  $D$  agree with the sign of the eigenvalues of the Hessian matrix of the potential at the equilibrium point. The different types of equilibria are illustrated in Figure 1.3, which shows the behaviour of the potential function around an equilibrium point in the simple case of a two-dimensional configuration space  $\mathbb{V}$ . The existence of zero modes is symptomatic of *flat directions* in the potential along which the system can evolve without spending any energy. This usually signals the existence of some continuous symmetry in the system. In the Figure we see that the semi-stable equilibrium point indeed has a flat direction along which the potential is constant. In other words, translation along the flat direction is a symmetry of the potential function.

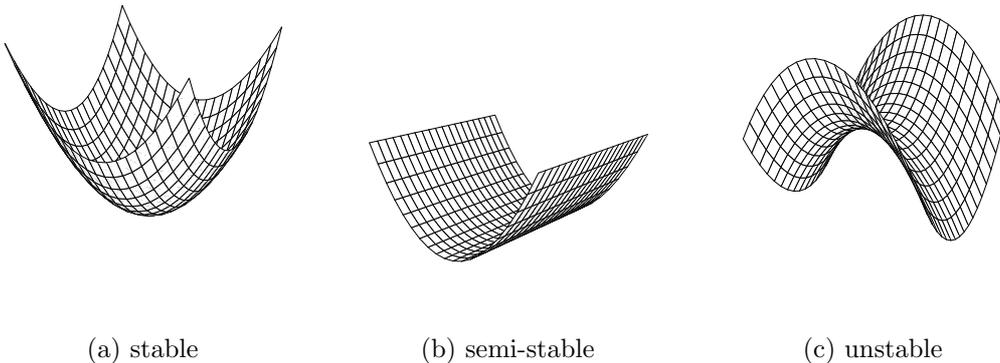


Figure 1.3: Different types of equilibrium points.