MAS277 Vector Spaces and Fourier Theory

Sarah Whitehouse

January 17, 2011

Introduction

This course involves many of the same themes as MAS201 Linear Mathematics for Applications, but takes a more abstract point of view. A central aim of the course is to help you become familiar and comfortable with mathematical abstraction and generalisation, which plays an important role in pure mathematics. This has many benefits.

- Efficiency.

  We will be able to prove a single theorem that simultaneously tells us useful things about vectors, matrices, polynomials, differential equations, and sequences satisfying a recurrence relation. Without the axiomatic approach, we would have to give five different (but very similar) proofs.

- Connections and analogies.

  We will see connections between apparently different areas of mathematics and learn to see common patterns. We will make some non-obvious but very useful analogies between different situations. For example, we will be able to define the distance or angle between two functions (by analogy with the distance or angle between two vectors in $\mathbb{R}^3$), and this will help us to understand the theory of Fourier series.

- Proofs.

  We will prove a number of things that were merely stated in MAS201. Similarly, we will give abstract proofs of some things that were previously proved using matrix manipulation. These new proofs will require a better understanding of the underlying concepts, but once you have that understanding, they will often be considerably simpler.

These Lecture Notes

These notes contain the definitions and statements of results, as well as some discussion. They do not contain proofs or details of worked examples and you will need to take your own notes of these in lectures.

1 Review of MAS201

1.1 Vector spaces

We begin with the notion of *vector space*, which was not explicit in MAS201, since that course dealt just with the most important family of examples, namely the set of vectors $\mathbb{R}^n$. For many of the ideas of this course, it may be helpful for you to try to visualise them in the setting of $\mathbb{R}^n$, and even $\mathbb{R}^3$, where your geometrical intuition is well established.

Let's think about the important properties of $\mathbb{R}^n$, the set of n-dimensional vectors with real numbers as entries. There are two crucial properties that $\mathbb{R}^n$ has.

1. Two elements of the set can be added to produce another element of the set.
2. We can scale any element of the set by a real number to produce another element of the set.

These two properties are fundamental to spaces of vectors, and we use them to give the more general definition of a *vector space*. Essentially, a vector space is going to be a set V where two elements can be added to give a third, and any element can be scaled by a real number to give another element. We have to be a little careful – we'd like the usual rules to be satisfied. For example, multiplying any element v by the real number 2, say, should be the same as taking the sum $v + v$ of the element v with itself.

Definition 1.1 A *(real) vector space* V is a set of elements (called *vectors*) such that we can add two vectors, and scale a vector by a real number, so that:

1. Given vectors u and v in V, then $u + v$ is also in V.
2. Given a vector $v \in V$, and a real number $\alpha \in \mathbb{R}$, then αv is also in V.
3. V is not empty; in particular, there is a zero vector $0_V \in V$.
4. These are subject to the usual rules: for any elements u, v and w in V and real numbers α and β, we have
   (a) $0_V + v = v$;
   (b) $u + v = v + u$;
   (c) $u + (v + w) = (u + v) + w$;
   (d) $0v = 0_V$;
   (e) $1v = v$;
   (f) $(\alpha\beta)v = \alpha(\beta v)$;
   (g) $(\alpha + \beta)v = \alpha v + \beta v$;
   (h) $\alpha(u + v) = \alpha u + \alpha v$.

There are various other things you might expect to see in this list of "usual rules", but it turns out that this list is sufficient to derive all the other "usual rules".

Example 1.2 *If $v \in V$, there is an element $-v$ such that $v + (-v) = 0_V$.*

However complicated the defintion may look, the key moral is the following:

> A vector space is a non-empty set V, where we can *add* two vectors in V to produce another element of V, and we can *scale* any vector in V by a real number to give another element of V, in such a way that the usual rules satisfied by $\mathbb{R}^n$ are satisfied.

At no point in the course will it really be necessary to remember exactly what "the usual rules" are, and it will suffice to use the informal version above.

Examples 1.3

1. In particular, we notice that $\mathbb{R}^n$ is a vector space! Here, $\mathbb{R}^n$ is, as usual, the set of vectors with n real entries. We can add two such vectors; e.g., in $\mathbb{R}^5$ we have $(1,2,4,8,16)^T + (1,-2,4,-8,16)^T = (2,0,8,0,32)^T$. (Notice that we will follow MAS201, and think of the vectors as *column* vectors, which is why we needed the transpose in our notation.)

2. The set of real polynomials of degree less than n is also a vector space; we can add two polynomials of degree less than n to get another, and we can scale a polynomial of this sort by any real number and the result will again be another polynomial of degree less than n. Indeed, we will see that this example is *isomorphic* to the previous one – this simply means that (essentially) they are two different ways of writing down the same thing; we can view the polynomial $a_0 + a_1 x + \cdots + a_{n-1}x^{n-1}$ as the real vector $(a_0 \; a_1 \; \ldots \; a_{n-1})^T$.

3. Another way to write the same set is to view it as the space of sequences of n real numbers $(a_0, a_1, \ldots, a_{n-1})$.

4. We can also consider the space of infinite sequences $\{(a_0, a_1, a_2, \ldots) \mid a_i \in \mathbb{R}\}$.

   But this differs from the space of all polynomials, which we might regard as an infinite sequence $(a_0, a_1, a_2, \ldots)$ where eventually (i.e., after some finite point) all the terms are 0.

   We therefore have two candidates for $\mathbb{R}^\infty$; this confusion will be partly responsible for us concentrating on "finite-dimensional" vector spaces.

5. In the other direction, $\{0\}$ is a trivial vector space, with just one element. We will see later that it is sensible to regard this as $\mathbb{R}^0$. We will sometimes write 0 for this trivial vector space.

In a sense, all these examples look similar; they can be viewed as sequences of real numbers. But the definition above also works in other contexts. Especially when we discuss Fourier theory, it will be useful to note that spaces of functions very often form vector spaces.

Example 1.4 The set $F(\mathbb{R})$ of all functions $\mathbb{R} \to \mathbb{R}$ is a vector space, because we can add any two functions to get a new function, and we can multiply a function by a real number to get a new function. For example, we can define

$$\begin{aligned} f(x) &= e^x, \\ g(x) &= e^{-x}, \\ h(x) &= \cosh(x) = (e^x + e^{-x})/2, \end{aligned}$$

so f, g and h are elements of $F(\mathbb{R})$. Then $f+g$ and $2h$ are again functions, and we actually have $f+g=2h$ in this example.

For this to work properly, we must insist that $f(x)$ is defined for all x, and is a real number for all x; it cannot be infinite or imaginary. Thus the rules $p(x)=1/x$ and $q(x)=\sqrt{x}$ do not define elements of $F(\mathbb{R})$.

In order to understand the above example, you need to think of a function $f:\mathbb{R}\to\mathbb{R}$ as a single object in its own right, and then think about the set $F(\mathbb{R})$ of all possible functions as a single object; later you will need to think about various different subsets of $F(\mathbb{R})$ (for example, just the continuous functions). All this may seem quite difficult to deal with. However, it is a central aim of this course for you to get to grips with this level of abstraction. So you should persevere, ask questions, study the notes and work through examples until it becomes clear to you.

Example 1.5 In practice, we do not generally want to consider the set $F(\mathbb{R})$ of *all* functions. Instead we consider the set $C(\mathbb{R})$ of continuous functions, or the set $C^\infty(\mathbb{R})$ of "smooth" functions (those which can be differentiated arbitrarily often), or the set $\mathbb{R}[x]$ of polynomial functions (e.g., $p(x)=1+x+x^2+x^3$ defines an element $p\in\mathbb{R}[x]$). If f and g are continuous then $f+g$ and αf are continuous, so $C(\mathbb{R})$ is a vector space. If f and g are smooth then $f+g$ and αf are smooth, so $C^\infty(\mathbb{R})$ is a vector space. If f and g are polynomials then $f+g$ and αf are polynomials, so $\mathbb{R}[x]$ is a vector space.

Example 1.6 We let $[a,b]$ denote the interval $\{x\in\mathbb{R}\,|\,a\leq x\leq b\}$, and we write $C[a,b]$ for the set of continuous functions $f:[a,b]\to\mathbb{R}$. For example, the rule $f(x)=1/x$ defines a continuous function on the interval $[1,2]$. (The only potential problem is at the point $x=0$, but $0\not\in[1,2]$, so we do not need to worry about it.) We thus have an element $f\in C[1,2]$.

Example 1.7 The set $M_2(\mathbb{R})$ of 2×2 matrices (with real entries) is a vector space. Indeed, if we add two such matrices, we get another 2×2 matrix; for example,

$$\begin{pmatrix}1&0\\0&4\end{pmatrix}+\begin{pmatrix}0&2\\3&0\end{pmatrix}=\begin{pmatrix}1&2\\3&4\end{pmatrix}.$$

Similarly, if we multiply a 2×2 matrix by a real number, we get another 2×2 matrix; for example,

$$7\begin{pmatrix}1&2\\3&4\end{pmatrix}=\begin{pmatrix}7&14\\21&28\end{pmatrix}.$$

We can identify $M_2(\mathbb{R})$ with $\mathbb{R}^4$, by the rule

$$\begin{pmatrix}a&b\\c&d\end{pmatrix}\to\begin{pmatrix}a\\b\\c\\d\end{pmatrix}.$$

More generally, for any m and n, the set $M_{m,n}(\mathbb{R})$ of $m\times n$ matrices is a vector space, which can be identified with $\mathbb{R}^{mn}$.

We will usually use the symbol 0, rather than 0_V, for the zero element of whatever vector space we are considering. Thus 0 could mean the vector $(0\ 0\ 0)^T$ (if we are working with $\mathbb{R}^3$) or the zero matrix $\begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ (if we are working with $M_{2,3}(\mathbb{R})$) or whatever. Occasionally it will be important to distinguish between the zero elements in different vector spaces. In that case, we will write 0_V for the zero element of V. For example, we have $0_{\mathbb{R}^2} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$ and $0_{M_2(\mathbb{R})} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$.

Example 1.8 We say that a *trigonometric polynomial of degree* n is a sum of the form

$$\frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kt + b_k \sin kt)$$

with $a_i, b_i \in \mathbb{R}$. (So these are just "finite" Fourier series.) Again, we can add any two of these, or scale by a real number, to get another trigonometric polynomial of degree n. They form a vector space. Since each trigonometric polynomial of degree n is specified by $2n+1$ coefficients $(a_0, a_1, \ldots, a_n, b_1, \ldots, b_n)$, this vector space "looks like" $\mathbb{R}^{2n+1}$.

We will consider various aspects of this space during the course.

One also has the notion of a vector space over $\mathbb{C}$ or over $\mathbb{Q}$, where we require that we can scale any vector by a *complex* number or by a *rational* number respectively (rather than a real number). In both cases, all the remaining "usual rules" must be satisfied.

Indeed, the abstract theory which we shall give in the next few lectures would work for vector spaces over any *field* (an algebraic structure where one can add, subtract, multiply and divide), but since real vector spaces are easier to visualize and to picture concretely, we'll work with them throughout the course. (In the final sections of the course, there will need to be minor modifications for complex vector spaces, but we will mention more then.)

Example 1.9 Almost all our examples of real vector spaces work over any field K. For example, the set $M_4(\mathbb{Q})$ (of 4×4 matrices whose entries are rational numbers) is a vector space over $\mathbb{Q}$. The set $\mathbb{C}[x]$ (of polynomials with complex coefficients) is a vector space over $\mathbb{C}$. Notice that since we can scale by a complex number, and $\mathbb{R} \subset \mathbb{C}$, we can certainly scale by real numbers too, so that $\mathbb{C}[x]$ is also a vector space over $\mathbb{R}$ (and, by the same argument, also over $\mathbb{Q}$). Indeed, note that $\mathbb{C}$ may also be regarded as a vector space over $\mathbb{R}$ – we can add two complex numbers, or can scale them by a real number.

1.2 Subspaces

A subspace of a vector space V is simply a subset $W \subseteq V$ that happens also to be a vector space. In particular, the sum of two vectors in W should again lie in W, and we should be able to scale any vector in W by any real number.

Definition 1.10 Let V be a vector space. A *subspace* of V is a subset $W \subseteq V$ such that the following conditions are satisfied.

(a) $0_V \in W$.

(b) Whenever u and v lie in W, the element $u + v$ also lies in W. (In other words, W is closed under addition.)

(c) Whenever u lies in W and α lies in $\mathbb{R}$, the element αu also lies in W. (In other words, W is closed under scalar multiplication.)

Remark 1.11 The definition can be reformulated slightly as follows: a set $W \subseteq V$ is a subspace if and only if the following conditions are satisfied.

(a) $0_V \in W$.

(d) Whenever $u, v \in W$ and $\alpha, \beta \in \mathbb{R}$ we have $\alpha u + \beta v \in W$.

Example 1.12 Given a vector space V, there are always two trivial subspaces: $\{0\}$ is always a subspace of V, and V itself is always a subspace of V.

Example 1.13 Any straight line through the origin is a subspace of $\mathbb{R}^2$. These are the only subspaces of $\mathbb{R}^2$ (except for the two trivial examples).

Example 1.14 In $\mathbb{R}^3$, any straight line through the origin is a subspace, and any plane through the origin is also a subspace. These are the only subspaces of $\mathbb{R}^3$ (except for the two trivial examples).

Example 1.15 The set $W = \{A \in M_2(\mathbb{R}) \,|\, \mathrm{trace}(A) = 0\}$ is a subspace of $M_2(\mathbb{R})$.

Example 1.16 Recall that $\mathbb{R}[x]$ denotes the set of all polynomial functions in one variable. This is a subspace of the vector space $F(\mathbb{R})$ of all functions on $\mathbb{R}$.

We write $\mathbb{R}[x]_{\leq d}$ for the set of polynomials of degree at most d, so a general element $f \in \mathbb{R}[x]_{\leq d}$ has the form

$$f(x) = a_0 + a_1 x + \cdots + a_d x^d = \sum_{i=0}^{d} a_i x^i$$

for some $a_0, \ldots, a_d \in \mathbb{R}$. It is easy to see that this is a subspace of $\mathbb{R}[x]$.

Definition 1.17 A function $f : \mathbb{R} \to \mathbb{R}$ is said to be *even* if $f(-x) = f(x)$ for all x, and *odd* if $f(-x) = -f(x)$ for all x. We write EF for the set of even functions and OF for the set of odd functions.

For example, $\cos(-x) = \cos(x)$ and $\sin(-x) = -\sin(x)$, so cos is even and sin is odd. Of course, most functions are neither even nor odd.

Example 1.18 EF is a subspace of $F(\mathbb{R})$. Similarly, OF is a subspace of $F(\mathbb{R})$.

Example 1.19 Recall that a *trigonometric polynomial of degree n* is a sum of the form

$$\frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kt + b_k \sin kt)$$

with $a_i, b_i \in \mathbb{R}$. The *even* trigonometric polynomials of degree n are those of the form

$$\frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kt,$$

and the *odd* trigonometric polynomials of degree n are those of the form

$$\sum_{k=1}^{n} b_k \sin kt.$$

The collection of all even trigonometric polynomials forms a subspace of the space of all trigonometric polynomials, as does the collection of all odd trigonometric polynomials.

Example 1.20 Consider the following sets of 3×3 matrices.

$$\begin{aligned}
U_1 &= \{\text{symmetric matrices}\} = \{A \in M_3(\mathbb{R}) \,|\, A^T = A\} \\
U_2 &= \{\text{antisymmetric matrices}\} = \{A \in M_3(\mathbb{R}) \,|\, A^T = -A\} \\
U_3 &= \{\text{trace-zero matrices}\} = \{A \in M_3(\mathbb{R}) \,|\, \operatorname{trace}(A) = 0\} \\
U_4 &= \{\text{diagonal matrices}\} = \{A \in M_3(\mathbb{R}) \,|\, A_{ij} = 0 \text{ whenever } i \neq j\} \\
U_5 &= \{\text{strictly upper-triangular matrices}\} = \{A \in M_3(\mathbb{R}) \,|\, A_{ij} = 0 \text{ whenever } i \geq j\} \\
U_6 &= \{\text{invertible matrices}\} = \{A \in M_3(\mathbb{R}) \,|\, \det(A) \neq 0\} \\
U_7 &= \{\text{noninvertible matrices}\} = \{A \in M_3(\mathbb{R}) \,|\, \det(A) = 0\}.
\end{aligned}$$

Then $U_1, \ldots, U_5$ are all subspaces of $M_3(\mathbb{R})$. We will prove this for U_1; the other cases are similar.

On the other hand, U_6 and U_7 are not subspaces.

Lemma 1.21 Let V be a vector space, and let U and W be subspaces of V. Then the following are also subspaces.

1. $U \cap W = \{v \in V \,|\, v \in U \text{ and } v \in W\}$.
2. $U + W = \{v \in V \,|\, v = u + w \text{ for some } u \in U \text{ and } w \in W\}$.

Example 1.22 If $V = \mathbb{R}^3$, $U = \{(x, 0, 0)^T \,|\, x \in \mathbb{R}\}$, and $W = \{(0, 0, z)^T \,|\, z \in \mathbb{R}\}$ then

$$U + W = \{(x, 0, z)^T \,|\, x, z \in \mathbb{R}\}.$$

Example 1.23 If $V = M_2(\mathbb{R})$ and

$$U = \left\{ \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} \;\middle|\; a, b \in \mathbb{R} \right\}, \qquad W = \left\{ \begin{pmatrix} 0 & b \\ 0 & d \end{pmatrix} \;\middle|\; b, d \in \mathbb{R} \right\},$$

then

$$U + W = \left\{ \begin{pmatrix} a & b \\ 0 & d \end{pmatrix} \;\middle|\; a, b, d \in \mathbb{R} \right\}.$$

Example 1.24 Take $V = \mathbb{R}^3$ and

$$\begin{aligned}
U &= \{(x, y, z)^T \,|\, x + 2y + 3z = 0\} \\
W &= \{(x, y, z)^T \,|\, 3x + 2y + z = 0\}.
\end{aligned}$$

We claim that $U \cap W = \{(x, -2x, x)^T \,|\, x \in \mathbb{R}\}$ and $U + W = \mathbb{R}^3$.

Definition 1.25 If V is a vector space, with subspaces U and W which have the property that $U + W = V$ and $U \cap W = 0$, then we say that V is the *direct sum* of U and W, and we write $V = U \oplus W$.

Example 1.26 Consider the space $F(\mathbb{R})$ of all functions from $\mathbb{R}$ to $\mathbb{R}$, and the subspaces EF and OF of even functions and odd functions. We claim that $F(\mathbb{R}) = EF \oplus OF$.

The same sort of statement applies for trigonometric polynomials.

We have a similar sort of thing for matrices.

Example 1.27 Put

$$\begin{aligned} V &= M_2(\mathbb{R}) \\ U &= \{A \in M_2(\mathbb{R}) \,|\, \mathrm{trace}(A) = 0\} = \left\{ \begin{pmatrix} a & b \\ c & -a \end{pmatrix} \;\middle|\; a, b, c \in \mathbb{R} \right\} \\ W &= \{\alpha . I \,|\, \alpha \in \mathbb{R}\} = \left\{ \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \;\middle|\; \alpha \in \mathbb{R} \right\}. \end{aligned}$$

We claim that $V = U \oplus W$.

The importance of direct sums is that when we write a vector $v \in V$ as the sum of $u + w$ (with $u \in U$ and $w \in W$), this expression is *unique*: if $u + w = u' + w'$, then $u - u' = w' - w$; the left-hand side is in U, the right-hand side in W – as they are equal, we see that $u - u' = w' - w \in U \cap W$, so $u - u' = w' - w = 0_V$, and so $u = u'$, $w = w'$, as claimed.

1.3 Linear independence

Two randomly-chosen vectors in $\mathbb{R}^2$ will generally not be parallel; it is an important special case if they happen to point in the same direction.

Similarly, given three vectors u, v and w in $\mathbb{R}^3$, there will usually not be any plane that contains all three vectors. This means that we can get from the origin to any point by travelling a certain (possibly negative) distance in the direction of u, then a certain distance in the direction of v, then a certain distance in the direction of w. The case where u, v and w all lie in a common plane will have special geometric significance in any purely mathematical problem, and will often have special physical significance in applied problems.

Our task in this section is to generalise these ideas, and study the corresponding special cases in an arbitrary vector space V. The abstract picture will be illuminating even in the case of $\mathbb{R}^2$ and $\mathbb{R}^3$.

Definition 1.28 Let V be a vector space, and let $\mathcal{V} = \{v_1, \dots, v_n\}$ be a set of elements of V. A *linear combination* of $\mathcal{V}$ is an expression of the form $\lambda_1 v_1 + \cdots + \lambda_n v_n$, where $\lambda_1, \dots, \lambda_n \in \mathbb{R}$. A *linear relation* occurs when a linear combination is zero, i.e., when $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$. We clearly get one relation, when $\lambda_1 = \cdots = \lambda_n = 0$, called the *trivial relation.* If there is a non-trivial linear relation, we say that the list $\mathcal{V}$ is *linearly dependent.* Otherwise, if the only relation is the trivial one, we say that the list $\mathcal{V}$ is *linearly independent.*

Lemma 1.29 *Suppose that $\mathcal{V} = \{v_1, \ldots, v_n\}$ is a linearly independent set of vectors in V. Let $v \in V$. Then there is at most one way to write v as a linear combination of the vectors in $\mathcal{V}$.*

Note that this does not say that any given vector can be expressed as a linear combination, merely that if such an expression exists (and it may not), then there is only one way to do it.

Example 1.30 Consider the following vectors in $\mathbb{R}^3$:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 4 \\ 5 \\ 6 \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} 7 \\ 8 \\ 9 \end{pmatrix}.$$

Then one finds that $\mathbf{v}_1 - 2\mathbf{v}_2 + \mathbf{v}_3 = 0$, a non-trivial linear relation, so the list $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3$ is linearly dependent.

Example 1.31 Consider the following vectors:

$$\mathbf{v}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{v}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

We claim these are linearly independent.

Example 1.32 Consider the polynomials $p_n(x) = (x+n)^2$, so

$$\begin{aligned} p_0(x) &= x^2 \\ p_1(x) &= x^2 + 2x + 1 \\ p_2(x) &= x^2 + 4x + 4 \\ p_3(x) &= x^2 + 6x + 9. \end{aligned}$$

We claim that the list p_0, p_1, p_2 is linearly independent, but that the list p_0, p_1, p_2, p_3 is linearly dependent.

We can think of linear relations as vectors $(\lambda_1, \ldots, \lambda_n)^T$ where $\lambda_1 v_1 + \cdots + \lambda_n v_n = 0$.

Example 1.33 Consider the functions

$$\begin{aligned} f_1(x) &= e^x \\ f_2(x) &= e^{-x} \\ f_3(x) &= \sinh(x) \\ f_4(x) &= \cosh(x). \end{aligned}$$

These are linearly dependent.

Example 1.34 We claim that the matrices

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad E_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad E_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

are linearly independent.

1.4 Spanning sets

Definition 1.35 Given a set $\mathcal{V} = \{v_1, \dots, v_n\}$ of elements of a vector space V, we write $\mathrm{sp}(\mathcal{V})$ for the set of all vectors $w \in V$ that can be written in the form $w = \lambda_1 v_1 + \cdots + \lambda_n v_n$ for some $\lambda_1, \dots, \lambda_n \in \mathbb{R}$.

Notice that $\mathrm{sp}(\mathcal{V})$ is a subspace of V.

Definition 1.36 We say that $\mathcal{V}$ is a *spanning set* for V if every vector in V can be written as a linear combination of elements of $\mathcal{V}$, i.e., if $\mathrm{sp}(\mathcal{V}) = V$.

In a sense, a spanning set is opposite to a linear independent set: if $\mathcal{V}$ is linearly independent, then every vector has at most one expression as a linear combination of elements of $\mathcal{V}$, whereas if $\mathcal{V}$ is a spanning set, every vector has at least one expression as a linear combination of elements of $\mathcal{V}$.

There is an obvious spanning set (the *standard basis*) for $\mathbb{R}^n$.

Definition 1.37 Let $\mathbf{e}_i$ be the vector in $\mathbb{R}^n$ whose ith entry is 1, with all other entries being zero. For example, in $\mathbb{R}^3$ we have

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \qquad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{e}_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}.$$

The vectors $\mathbf{e}_1, \dots, \mathbf{e}_n$ span $\mathbb{R}^n$. Indeed, any vector $\mathbf{x} \in \mathbb{R}^n$ can be written as $x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$, which is a linear combination of $\mathbf{e}_1, \dots, \mathbf{e}_n$, as required. For example, in $\mathbb{R}^3$ we have

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x_1 \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} + x_3 \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + x_3\mathbf{e}_3.$$

Example 1.38 The set $1, x, \dots, x^n$ spans $\mathbb{R}[x]_{\leq n}$. Indeed, any element of $\mathbb{R}[x]_{\leq n}$ is a polynomial of the form $f(x) = a_0 + a_1 x + \cdots + a_n x^n$, and so is visibly a linear combination of $1, x, \dots, x^n$.

Example 1.39 Consider the vectors

$$\mathbf{u}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{u}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \qquad \mathbf{u}_3 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \mathbf{u}_4 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

We claim that these span $\mathbb{R}^4$.

Example 1.40 Consider the polynomials $p_i(x) = (x + i)^2$. We claim that the list $p_{-2}, p_{-1}, p_0, p_1, p_2$ spans $\mathbb{R}[x]_{\leq 2}$.

Example 1.41 Put $V = \{f \in C^\infty(\mathbb{R}) \mid f'' + f = 0\}$. We claim that the functions sin and cos span V.

Definition 1.42 A vector space V is *finite-dimensional* if there is a finite set $\mathcal{V} = \{v_1, \dots, v_n\}$ of elements of V that spans V.

Example 1.43 Using our earlier examples of spanning sets, we see that the spaces $\mathbb{R}^n$, $M_{n,m}(\mathbb{R})$ and $\mathbb{R}[x]_{\leq n}$ are finite-dimensional.

Example 1.44 The space $\mathbb{R}[x]$ is not finite-dimensional.

1.5 Basis and dimension

Definition 1.45 A *basis* for a vector space V is a set $\mathcal{V}$ of elements of V that is linearly independent and also spans V.

Recall that to be a spanning set, any vector $v \in V$ must be expressible as a linear combination of elements of $\mathcal{V}$. However, Lemma 1.29 tells us that there is now exactly one way to do this. That is, if $\mathcal{V}$ is a basis, every vector in V has a *unique* expression as a linear combination of vectors in $\mathcal{V}$.

Example 1.46 We will find a basis for the space V of antisymmetric 3×3 matrices.

Example 1.47 Put $V = \{A \in M_3(\mathbb{R}) \mid A^T = A \text{ and } \operatorname{trace}(A) = 0\}$. We will find a basis for V.

Definition 1.48 Suppose that V is a finite-dimensional vector space. Then the *dimension* of V is the number of elements in a basis of V.

Hang on! This definition makes a number of assumptions which we ought to be rather careful about! We haven't even proved that every finite-dimensional vector space has a basis at all. Even if that were true, could there be some funny space V, with two bases $\mathcal{V}_1$ and $\mathcal{V}_2$ with different numbers of elements?

Luckily, we'll see a bit later that every finite-dimensional vector space does have a basis. Firstly, though, we'll show that two different bases do have the same number of elements. But it's a bit fiddly to compare two bases, so the argument is slightly tricky.

We'll first prove that the number of elements in a linearly independent set can never exceed the number of elements in a spanning set. (Hopefully your intuition suggests that this is to be expected.)

Lemma 1.49 (Steinitz Exchange Lemma) *Suppose that V is a finite-dimensional vector space. Suppose that $\mathcal{V} = \{v_1, \dots, v_r\}$ is a linearly independent set and that $\mathcal{W} = \{w_1, \dots, w_s\}$ is a spanning set for V. Then $r \leq s$.*

Corollary 1.50 *Let V be a finite-dimensional vector space, and suppose $\mathcal{V} = \{v_1, \dots, v_r\}$ and $\mathcal{V}' = \{v'_1, \dots, v'_s\}$ are two bases for V. Then $r = s$.*

Similarly, any linearly independent set in an n-dimensional vector space can have at most n elements (compare with a basis, which spans), and any spanning set in an n-dimensional vector space must have at least n elements (compare with a basis, which is linearly independent).

The existence of the standard basis tells us that the dimension of $\mathbb{R}^n$ is n. Further, in Example 1.46, the dimension of V is 3, and in Example 1.47, the dimension of V is 5. You should hopefully see that, roughly, the dimension can be regarded as the number of "parameters" one needs to specify in order to define the vector. For example, in $\mathbb{R}^n$, a vector is defined by specifying n co-ordinates.

Example 1.51 There are several interesting bases for the space $Q = \mathbb{R}[x]_{\leq 2}$ of polynomials of degree at most two. A typical element $f \in Q$ has $f(x) = ax^2 + bx + c$ for some $a, b, c \in \mathbb{R}$. So the dimension of Q is 3. We will show that each of the following is a basis.

- The set p_0, p_1, p_2, where $p_i(x) = x^i$.
- The set

$$\begin{aligned} q_0(x) &= (x^2 - 3x + 2)/2 \\ q_1(x) &= -x^2 + 2x \\ q_2(x) &= (x^2 - x)/2. \end{aligned}$$

- The set

$$\begin{aligned} r_0(x) &= 1 \\ r_1(x) &= \sqrt{3}(2x - 1) \\ r_2(x) &= \sqrt{5}(6x^2 - 6x + 1). \end{aligned}$$

Example 1.52 Put $V = \{f \in \mathbb{R}[x]_{\leq 4} \mid f(1) = f(-1) = 0 \text{ and } f'(1) = f'(-1)\}$. Let $p(x) = x^3 - x$ and $q(x) = x^4 - 2x^2 + 1 = (x^2 - 1)^2$. Then we claim that p, q is a basis for V.

Example 1.53 A *magic square* is a 3×3 matrix in which every row, every column and both diagonals add up to the same number. Let V be the set of magic squares, which is easily seen to be a subspace of $M_3(\mathbb{R})$. We will find a basis for V.

Now we'll prove a few slightly technical results about bases and dimension. Hopefully none will seem too surprising, although the proofs can get a little intricate.

For example, we haven't yet shown that finite-dimensional vector spaces have any basis at all – but of course they will.

Proposition 1.54 *Let V be a vector space, and let $\mathcal{V} = \{v_1, \ldots, v_n\}$ be a finite set of elements of V that spans V. Then some subset $\mathcal{V}' \subseteq \mathcal{V}$ is a basis for V.*

Thus every finite-dimensional vector space has a basis: just find a spanning set (which exists by definition of "finite-dimensional") and choose an appropriate subset.

Now that we've proved that any spanning set can by shrunk to a basis, let's go the other way, and explain that any linearly independent set can be expanded to a basis.

Proposition 1.55 *Let V be an n-dimensional vector space, and let $\mathcal{V} = \{v_1, \ldots, v_p\}$ be a linearly independent set of elements of V. Then $p \leq n$, and $\mathcal{V}$ can be extended to a set $\mathcal{V}' = \{v_1, \ldots, v_n\}$ such that $\mathcal{V}'$ is a basis of V.*

Proposition 1.56 *Let V be a finite-dimensional vector space, and let W be a subspace of V. If W has a basis $\mathcal{W} = \{v_1, \dots, v_p\}$, then we can extend this set to a basis $\mathcal{V} = \{v_1, \dots, v_n\}$ for V. In particular, $\dim(W) \leq \dim(V)$.*

Further, if $\dim(W) = \dim(V)$, then $W = V$.

Proposition 1.57 *Let V be an n-dimensional vector space.*

(a) Any spanning set for V with exactly n elements is linearly independent, and so is a basis.

(b) Any linearly independent set in V with exactly n elements is a spanning set, and so is a basis.

Proposition 1.58 *Let V be a finite-dimensional vector space, and let U and W be subspaces of V. Then*

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

Proposition 1.59 *Suppose that V is a vector space, with subspaces U and W satisfying $U \cap W = 0$ and $U + W = V$ (recall that this means that $V = U \oplus W$, the direct sum of U and W). If $\mathcal{U}$ is a basis for U and $\mathcal{W}$ is a basis for W, then $\mathcal{U} \cup \mathcal{W}$ is a basis for V.*

1.6 Bases and matrices

Suppose that V is an n-dimensional vector space, and that we have two bases $\mathcal{V} = \{v_1, \dots, v_n\}$ and $\mathcal{V}' = \{v_1', \dots, v_n'\}$. Given any vector $v \in V$, we can express it as a linear combination of the v_i:

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

As $\mathcal{V}'$ is also a basis, we can also express v as a linear combination of the v_i':

$$v = \mu_1 v_1' + \cdots + \mu_n v_n'.$$

How are these related?

Proposition 1.60 *1. In the above situation,*

$$\begin{pmatrix} \mu_1 \\ \vdots \\ \mu_n \end{pmatrix} = B \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \end{pmatrix},$$

where

$$B = \begin{pmatrix} \beta_{11} & \dots & \beta_{1n} \\ \vdots & & \vdots \\ \beta_{n1} & \dots & \beta_{nn} \end{pmatrix}$$

is given by $v_i = \beta_{1i} v_1' + \cdots + \beta_{ni} v_n'$.

2. In the case where $\mathcal{V}'$ is the standard basis, the matrix B has columns the vectors of $\mathcal{V}$.

3. *In the general case, $B = (A')^{-1}A$, where A is the matrix with columns the vectors of $\mathcal{V}$ and A' is is the matrix with columns the vectors of $\mathcal{V}'$.*

The matrix B is called a change-of-basis matrix.

Example 1.61 Suppose $V = \mathbb{R}^3$, with the two bases

$$\mathcal{V} = \{v_1 = (1,1,1)^T, v_2 = (0,1,1)^T, v_3 = (0,0,1)^T\} \text{ and}$$
$$\mathcal{V}' = \{v_1' = (1,0,0)^T, v_2' = (1,0,1)^T, v_3' = (0,1,1)^T\}.$$

Suppose that the vector v can be written as $av_1 + bv_2 + cv_3$ and $\alpha v_1' + \beta v_2' + \gamma v_3'$. Express α, β and γ in terms of a, b and c.

Finally, let us observe that the change-of-basis matrix is always invertible. Indeed, given two bases, $\mathcal{V}$ and $\mathcal{V}'$, and an expression for a vector $v \in V$ in terms of $\mathcal{V}$, we can get the expression in terms of $\mathcal{V}'$ by multiplying by the change-of-basis matrix. But we can go backwards too, and now express it in terms of $\mathcal{V}$ by multiplying by the change-of-basis matrix from $\mathcal{V}'$ to $\mathcal{V}$. These two matrices must be inverse to each other, since we get back the vector we start with.

Warning! In this section, if we take the basis vectors in a different order, we will get a different matrix. Indeed, changing the order of the elements in either $\mathcal{V}$ or $\mathcal{V}'$ will change the order of the rows or columns of the change-of-basis matrix. Unlike the earlier sections (linear independence, spanning, etc.), then, the order matters: we should therefore not consider the bases as *sets*, where there is no ordering, but rather as *lists*, when we do have an order. We'll pass over this distinction rather blithely, supposing when we need to that there is an order on our sets.

2 Linear maps

2.1 Definition and basic properties

Matrices can be viewed as functions between vector spaces. Indeed, if $A \in M_{m,n}(\mathbb{R})$, then we get a map

$$\begin{array}{rcl} \mathbb{R}^n & \to & \mathbb{R}^m \\ \mathbf{v} & \mapsto & A\mathbf{v} \end{array}$$

We are going to make this a little more abstract in this chapter. We'll study more general maps between vector spaces which have similar properties to these matrix maps. Indeed, we will study the maps between vector spaces which have all the same properties that matrices do when the vector spaces involved are $\mathbb{R}^n$.

Definition 2.1 Let V and W be vector spaces, and let $\phi : V \to W$ be a function (so for each element $v \in V$ we have an element $\phi(v) \in W$). We say that ϕ is *linear* if it satisfies the following two conditions.

(a) For any v and v' in V, we have $\phi(v + v') = \phi(v) + \phi(v')$ in W.

(b) For any $\alpha \in \mathbb{R}$ and $v \in V$ we have $\phi(\alpha v) = \alpha\phi(v)$ in W.

By taking $\alpha = v = 0$ in (b), we see that a linear map must satisfy $\phi(0) = 0$. Further simple arguments also show that $\phi(v - v') = \phi(v) - \phi(v')$.

The definition can be reformulated slightly as follows. A map $\phi : V \to W$ is linear if and only if

(c) for any $\alpha, \alpha' \in \mathbb{R}$ and any $v, v' \in V$ we have $\phi(\alpha v + \alpha' v') = \alpha\phi(v) + \alpha'\phi(v')$.

Example 2.2 Let A be a fixed $m \times n$ matrix. Given a vector $\mathbf{v}$ of length n (so $\mathbf{v} \in \mathbb{R}^n$), we can multiply A by $\mathbf{v}$ in the usual way to get a vector $A\mathbf{v}$ of length m. We can thus define $\phi_A : \mathbb{R}^n \to \mathbb{R}^m$ by $\phi_A(\mathbf{v}) = A\mathbf{v}$. It is clear that $A(\mathbf{v} + \mathbf{v}') = A\mathbf{v} + A\mathbf{v}'$ and $A(\alpha\mathbf{v}) = \alpha.A\mathbf{v}$, so ϕ_A is a linear map. We will see later that every linear map from $\mathbb{R}^n$ to $\mathbb{R}^m$ has this form.

Example 2.3 Consider the functions $f, g, h, j : \mathbb{R} \to \mathbb{R}$ given by $f(x) = 2x, g(x) = x^2, h(x) = \sin(x), j(x) = \exp(x)$. The function f is linear, but g, h and j are not.

Similarly, all the maps $\mu_m : \mathbb{R} \to \mathbb{R}$ given by $\mu_m(x) = mx$ for some constant m are all linear ($\mu_m = \phi_{(m)}$); indeed, all the linear maps $\mathbb{R} \to \mathbb{R}$ are of this form. Note also that the graph of μ_m is a straight line of slope m through the origin; this is essentially the reason for the word "linear". Also similarly, the maps $x \mapsto x^\alpha$ are never linear for any $\alpha \neq 1$.

Example 2.4 For any $\mathbf{v} \in \mathbb{R}^2$, we let $\rho(\mathbf{v})$ be the vector obtained by rotating $\mathbf{v}$ through 90 degrees anticlockwise around the origin. Then $\rho : \mathbb{R}^2 \to \mathbb{R}^2$ is linear.

Example 2.5 For any $\mathbf{v} \in \mathbb{R}^2$, we let $\tau(\mathbf{v})$ be the vector obtained by reflecting $\mathbf{v}$ across the line $y = 0$. Then $\tau : \mathbb{R}^2 \to \mathbb{R}^2$ is linear.

Example 2.6 Define $\theta : \mathbb{R}^2 \to \mathbb{R}$ by $\theta(\mathbf{v}) = \|\mathbf{v}\|$, so $\theta\begin{pmatrix} x \\ y \end{pmatrix} = \sqrt{x^2 + y^2}$. This is not linear.

Example 2.7 Define $\sigma : \mathbb{R}^2 \to \mathbb{R}^2$ by $\sigma\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x+1 \\ y-1 \end{pmatrix}$. Then σ is not linear.

Example 2.8 Given vectors $\mathbf{u} = \begin{pmatrix} u_1 \\ u_2 \\ u_3 \end{pmatrix}$ and $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \\ v_3 \end{pmatrix}$ in $\mathbb{R}^3$, recall that the inner product and cross product are defined by

$$\begin{aligned} \mathbf{u}.\mathbf{v} &= u_1v_1 + u_2v_2 + u_3v_3 \\ \mathbf{u} \times \mathbf{v} &= \begin{pmatrix} u_2v_3 - u_3v_2 \\ u_3v_1 - u_1v_3 \\ u_1v_2 - u_2v_1 \end{pmatrix}. \end{aligned}$$

Fix a vector $\mathbf{a} \in \mathbb{R}^3$. Define $\kappa : \mathbb{R}^3 \to \mathbb{R}$ by $\kappa(\mathbf{v}) = \mathbf{a}.\mathbf{v}$ and $\lambda : \mathbb{R}^3 \to \mathbb{R}^3$ by $\lambda(\mathbf{v}) = \mathbf{a} \times \mathbf{v}$. Then both κ and λ are linear.

Example 2.9 For any continuous function $f : \mathbb{R} \to \mathbb{R}$, we write

$$I(f) = \int_0^1 f(x)\,dx \in \mathbb{R}.$$

This defines a map $I : C(\mathbb{R}) \to \mathbb{R}$. Then I is a linear map.

Now let's see that every linear map $\mathbb{R}^n \to \mathbb{R}^m$ is given by a matrix, and then we'll deduce the same result for arbitrary finite dimensional vector spaces.

Proposition 2.10 *Every linear map $\theta : \mathbb{R}^n \to \mathbb{R}^m$ has the form ϕ_A (as in Example 2.2) for some $m \times n$ matrix A (which is uniquely determined). Thus, a linear map $\theta : \mathbb{R}^n \to \mathbb{R}^m$ is essentially the same thing as an $m \times n$ matrix.*

Notice also that the same kind of argument will apply to any linear map between two finite-dimensional vector spaces, once bases have been chosen. Indeed, let $\theta : V \to W$ be a linear map between two vector spaces, with (ordered) bases $\mathcal{V} = v_1, \ldots, v_n$ and $\mathcal{W} = w_1, \ldots, w_m$. Then the same argument works: given $\mathbf{x} = x_1 v_1 + \cdots + x_n v_n \in V$, we see that $\theta(\mathbf{x}) = x_1\theta(v_1) + \cdots + x_n\theta(v_n) \in W$. Each vector $\theta(v_i) \in W$, so can be written $a_{1i}w_1 + \cdots + a_{mi}w_m$. Then form $A = \begin{pmatrix} a_{11} & \ldots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \ldots & a_{mn} \end{pmatrix}$, and notice that the same argument shows that $\theta(\mathbf{x}) = y_1 w_1 + \cdots + y_m w_m$, where

$$\begin{pmatrix} y_1 \\ \vdots \\ y_m \end{pmatrix} = A \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}.$$

It's important to notice that this view of linear maps as matrices is only valid once bases have been chosen – only then can we identify vectors as linear combinations of basis vectors, and write the map as a matrix in terms of these bases.

Example 2.11 For any smooth function $f : \mathbb{R} \to \mathbb{R}$ we write $D(f) = f'$ and $L(f) = f'' + f$. These are again smooth functions, so we have maps $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ and $L : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$. These are linear.

Also consider the subspace V spanned by $\mathcal{V} : p = \sin, q = \cos, r = \exp$, a subspace of $F(\mathbb{R})$. Then $D : V \to V$ and $L : V \to V$ are linear maps. Write them as matrices with respect to the basis p, q, r.

Example 2.12 For any 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$, the trace and determinant are defined by $\operatorname{trace}(A) = a + d \in \mathbb{R}$ and $\det(A) = ad - bc \in \mathbb{R}$. We thus have two functions $\operatorname{trace}, \det : M_2(\mathbb{R}) \to \mathbb{R}$. Show that trace is linear and write it as a matrix with respect to the basis E_1, E_2, E_3, E_4 of $M_2(\mathbb{R})$, where

$$E_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \qquad E_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \qquad E_3 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \qquad E_4 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix},$$

and the basis 1 of $\mathbb{R}$. Also show that det is not linear.

None of this is really restricted to 2×2 matrices. For any n we have a map $\mathrm{trace} : M_n(\mathbb{R}) \to \mathbb{R}$ given by $\mathrm{trace}(A) = \sum_{i=1}^n A_{ii}$, which is again linear. We also have a determinant map $\det : M_n(\mathbb{R}) \to \mathbb{R}$ which satisfies $\det(\alpha I) = \alpha^n$; this shows that det is not linear, except in the case where $n = 1$.

Example 2.13 We can define a map $\mathrm{trans} : M_n(\mathbb{R}) \to M_n(\mathbb{R})$ by $\mathrm{trans}(A) = A^T$. Here as usual, A^T is the transpose of A, which is obtained by flipping A across the main diagonal. For example:

$$\begin{pmatrix} 1 & 2 & 3 \\ 0 & 4 & 5 \\ 0 & 0 & 6 \end{pmatrix}^T = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 4 & 0 \\ 3 & 5 & 6 \end{pmatrix}.$$

This is linear.

We've already stressed that if $\phi : V \to W$ is a linear map, then once we choose bases $\mathcal{V}$ and $\mathcal{W}$ for V and W respectively, the map ϕ can be regarded as a matrix. But what if we were to choose different bases $\mathcal{V}'$ and $\mathcal{W}'$?

Proposition 2.14 *Suppose the linear map $\phi : V \to W$ is represented by the matrix A when we use bases $\mathcal{V}$ and $\mathcal{W}$, and by A' when we use $\mathcal{V}'$ and $\mathcal{W}'$. Then $A' = C^{-1}AB$, where B is the change-of-basis matrix from $\mathcal{V}'$ to $\mathcal{V}$ and C is the change-of-basis matrix from $\mathcal{W}'$ to $\mathcal{W}$.*

You should regard this multiplication as: first write your vector with respect to the basis $\mathcal{V}'$. Then multiply by B to get the vector in terms of the basis $\mathcal{V}$. With respect to the basis $\mathcal{V}$, the map ϕ is given by the matrix A, which gives $\phi(v)$ in terms of the basis $\mathcal{W}$. Finally, multiply by C^{-1} to write $\phi(v)$ in terms of $\mathcal{W}'$.

In practice, it is usually easier not to use this theory, but to use a more *ad hoc* method. Given a vector $v \in V$ in terms of $\mathcal{V}'$, rewrite it in terms of $\mathcal{V}$. Then apply the original matrix to compute $\phi(v) \in W$ in terms of $\mathcal{W}$. And then work out how to rewrite $\phi(v)$ in terms of $\mathcal{W}'$. Compare the original vector in terms of $\mathcal{V}'$ and $\phi(v)$ in terms of $\mathcal{W}'$, and work out the corresponding matrix.

Let's consider the special case where ϕ is a map from a (finite dimensional) vector space V to itself.

Proposition 2.15 *Let $\phi : V \to V$ be a linear map. The determinant of any matrix representing ϕ is the same, i.e. this determinant does not depend on the choice of basis. The same is true for the trace and the characteristic polynomial.*

Definition 2.16 *We define the determinant of a linear map $\phi : V \to V$ to be the determinant of the matrix of ϕ with respect to any basis. Similarly for the trace and characteristic polynomial.*

Since the characteristic polynomial is unaffected by a change of basis, so are its roots, and thus we have a sensible notion of an *eigenvalue* of the linear map, namely a root of the characteristic polynomial with respect to any basis. (Since we are dealing with *real* vector spaces in this section, our eigenvalues only have corresponding eigenvectors if they are real, but for more general complex vector spaces, every eigenvalue comes with corresponding eigenvectors.)

Definition 2.17 *Let V be a finite-dimensional vector space over $\mathbb{C}$, and let $\phi : V \to V$ be a linear map (note that both the source and target vector spaces here are the same). Let λ be a complex number. An eigenvector for ϕ with eigenvalue λ is a nonzero vector $v \in V$ such that $\phi(v) = \lambda v$.*

Suppose we choose a basis $\mathcal{V}$ for V, and let A be the matrix of ϕ with respect to $\mathcal{V}$ and $\mathcal{V}$. Then the eigenvalues of ϕ are the same as the eigenvalues of the matrix A, which are the roots of the characteristic polynomial $\det(tI - A)$, as we've already noted.

Example 2.18 Put $V = \mathbb{R}[x]_{\leq 4}$, and define $\phi : V \to V$ by $\phi(f)(x) = f(-x)$, so $\phi(x^k) = (-1)^k x^k$. Find the corresponding matrix P (with respect to $1, x, x^2, x^3, x^4$) and its characteristic polynomial, eigenvalues and eigenvectors.

In general, if ϕ is a linear map from one (finite dimensional) real vector space to itself, and if λ is a real root of the characteristic polynomial of ϕ (i.e., of the matrix of ϕ with respect to any basis), then there will be an eigenvector v such that $\phi(v) = \lambda v$.

More generally, if $\phi : V \to V$ is a linear map from a complex vector space to itself, we know that there will be eigenvalues (roots of the characteristic polynomial), and that each will have a corresponding eigenvector. However, you should recall from MAS201 that if λ is a root of the characteristic polynomial with some multiplicity k, then there need not be k linearly independent eigenvectors – there will be at least one, but there could be anywhere between 1 and k eigenvectors.

Example 2.19 Put $V = \mathbb{R}[x]_{\leq 4}$, and define $\phi : V \to V$ by $\phi(f)(x) = f(x+1)$. We claim that 1 is the only eigenvalue.

2.2 Kernels and images

Definition 2.20 Let V and W be vector spaces, and let $\phi : V \to W$ be a linear map. Then we write

$$\begin{aligned} \ker(\phi) &= \{v \in V \mid \phi(v) = 0\} \\ \operatorname{im}(\phi) &= \{w \in W \mid w = \phi(v) \text{ for some } v \in V\}. \end{aligned}$$

Example 2.21 Define $\pi : \mathbb{R}^3 \to \mathbb{R}^3$ by $\pi \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ y \\ z \end{pmatrix}$. Write down $\ker(\pi)$ and $\operatorname{im}(\pi)$.

Remark 2.22 We can write simultaneous linear equations as a matrix equation $A\mathbf{x} = \mathbf{b}$ (or in other words, $\phi_A(\mathbf{x}) = \mathbf{b}$). Then the system of linear equations has a solution if $\mathbf{b}$ is in the image of the linear map ϕ_A. Further, if $\mathbf{x}$ and $\mathbf{x}'$ are two solutions, then $A(\mathbf{x} - \mathbf{x}') = 0$, so if $\ker(\phi_A) = \{0\}$, the solution must be unique (as then $\mathbf{x} - \mathbf{x}' = 0$), if it exists.

Proposition 2.23 *Let V and W be vector spaces, and let $\phi : V \to W$ be a linear map. Then $\ker(\phi)$ is a subspace of V and $\operatorname{im}(\phi)$ is a subspace of W.*

Example 2.24 Define $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ by $\phi((x, y, z)^T) = (x - y, y - z, z - x)^T$. Then

$$\begin{aligned} \ker(\phi) &= \{(x, y, z)^T \in \mathbb{R}^3 \mid x = y = z\} = \{(t, t, t)^T \mid t \in \mathbb{R}\}, \\ \operatorname{im}(\phi) &= \{(x, y, z)^T \in \mathbb{R}^3 \mid x + y + z = 0\} = \{(x, y, -x - y)^T \mid x, y \in \mathbb{R}^2\}. \end{aligned}$$

Example 2.25 Define $\phi : M_n(\mathbb{R}) \to M_n(\mathbb{R})$ by $\phi(A) = A - A^T$ (which is linear). Then clearly $\phi(A) = 0$ precisely when $A = A^T$, i.e., A is a symmetric matrix. Thus

$$\ker(\phi) = \{n \times n \text{ symmetric matrices}\}.$$

We claim that also

$$\operatorname{im}(\phi) = \{n \times n \text{ antisymmetric matrices}\}.$$

Recall that a map $\phi : V \to W$ is *surjective* if every element $w \in W$ has the form $\phi(v)$ for some $v \in V$. Moreover, ϕ is said to be *injective* if whenever $\phi(v) = \phi(v')$ we have $v = v'$.

Proposition 2.26 *Let V and W be vector spaces, and let $\phi : V \to W$ be a linear map. Then ϕ is injective if and only if $\ker(\phi) = \{0\}$ and ϕ is surjective if and only if $\operatorname{im}(\phi) = W$.*

Definition 2.27 We say that a linear map $\phi : V \to W$ is an *isomorphism* if it is a bijection, so there is an inverse map $\phi^{-1} : W \to V$ with $\phi(\phi^{-1}(w)) = w$ for all $w \in W$, and $\phi^{-1}(\phi(v)) = v$ for all $v \in V$. (It turns out that ϕ^{-1} is automatically a *linear* map – we leave this as an exercise.) We say that V and W are *isomorphic* if there exists an isomorphism from V to W.

Example 2.28 We can now rephrase part of Example 1.7 as follows: there is an isomorphism $\phi : M_2(\mathbb{R}) \to \mathbb{R}^4$ given by $\phi\begin{pmatrix} a & b \\ c & d \end{pmatrix} = (a, b, c, d)^T$, so $M_2(\mathbb{R})$ is isomorphic to $\mathbb{R}^4$. Similarly, the space $M_{m,n}(\mathbb{R})$ is isomorphic to $\mathbb{R}^{mn}$.

Corollary 2.29 *$\phi : V \to W$ is an isomorphism if and only if $\ker(\phi) = 0$ and $\operatorname{im}(\phi) = W$.*

Example 2.30 Define $\phi : \mathbb{R}[x]_{\leq 2} \to \mathbb{R}^2$ by $\phi(f) = (f(1), f'(1))^T$. More explicitly, we have

$$\phi(ax^2 + bx + c) = (a + b + c, 2a + b)^T.$$

Then ϕ is surjective but not injective.

Example 2.31 Consider the map $\phi : \mathbb{R}^3 \to \mathbb{R}^3$ given by

$$\phi((x, y, z)^T) = (x - y, y - z, z - x)^T.$$

This is neither injective nor surjective.

Suppose that $\phi : V \to W$ is a linear map between two vector spaces. We've already explained that $\ker(\phi)$ is a subspace of V and $\operatorname{im}(\phi)$ is a subspace of W. The dimension of $\ker(\phi)$ is sometimes called the *nullity* of ϕ, and the dimension of $\operatorname{im}(\phi)$ is the *rank* of ϕ. Let's now briefly explain this latter terminology. After choosing a basis $\mathcal{V}$ for V and $\mathcal{W}$ for W, the linear map ϕ is represented by multiplication by a matrix, A say. The map is given on $v \in V$ by taking the vector of coefficients with respect to $\mathcal{V}$, and multiplying by A, to get the vector of coefficients with respect to $\mathcal{W}$. Then the image is spanned by the columns of the matrix, so the dimension of the image is the same as the dimension of the column space of the matrix: you will recall from MAS201 that this is the definition of the (column) rank of the matrix.

There is a relation between the dimensions of the kernel and image of a linear map.

Theorem 2.32 (Rank-Nullity) *Let $\phi : V \to W$ be a linear map between two vector spaces and suppose that V is finite-dimensional. Then*

$$\dim\ker(\phi) + \dim\operatorname{im}(\phi) = \dim(V).$$

Let's now view this result in terms of matrices, and we'll see that we recover some of the row-reduction results from MAS201. Using the notation from the proof of Theorem 2.32, put $w_i = \phi(v_{n+i})$ for $1 \le i \le m$, and extend $\{w_1, \dots, w_m\}$ to a basis $\mathcal{W} = \{w_1, \dots, w_m, w_{m+1}, \dots, w_{m+r}\}$ for W. Then with respect to the bases $\mathcal{V} = \{v_1, \dots, v_n, v_{n+1}, \dots, v_{n+m}\}$ and $\mathcal{W}$, the map ϕ has the following properties:

(a) $\phi(v_i) = 0$ for $1 \le i \le n$;

(b) $\phi(v_{n+i}) = w_i$ for $1 \le i \le m$;

(c) $\{v_1, \dots, v_n\}$ is a basis for $\ker(\phi)$;

(d) $\{w_1, \dots, w_m\}$ is a basis for $\operatorname{im}(\phi)$.

The matrix of ϕ with respect to the bases $\mathcal{V}$ and $\mathcal{W}$ is therefore:

$$A = \left(\begin{array}{c|c} 0_{n,m} & I_m \\ 0_{n,r} & 0_{m,r} \end{array} \right),$$

which should remind you of some results from MAS201 on row reductions of matrices.

Example 2.33 Consider the map $\phi : M_2(\mathbb{R}) \to M_2(\mathbb{R})$ given by $\phi(A) = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} A \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, or equivalently

$$\begin{aligned} \phi\begin{pmatrix} a & b \\ c & d \end{pmatrix} &= \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} b+d & a+c \\ b+d & a+c \end{pmatrix} \\ &= (b+d)\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} + (a+c)\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}. \end{aligned}$$

Find bases for $M_2(\mathbb{R})$ such that $\phi(v_1) = \phi(v_2) = 0$, $\phi(v_3) = w_1$ and $\phi(v_4) = w_2$, so the matrix of ϕ with respect to these bases is $\begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$.

2.3 Maps out of $\mathbb{R}^n$

Let V be a vector space, and let $\mathcal{V} = v_1, \dots, v_n$ be a list of elements of V. We have a linear map $\mu_{\mathcal{V}} : \mathbb{R}^n \to V$, given by

$$\mu_{\mathcal{V}}((\lambda_1, \dots, \lambda_n)^T) = \lambda_1 v_1 + \cdots + \lambda_n v_n.$$

By definition, a linear relation between the v_i is just a vector $\ell = (\lambda_1, \dots, \lambda_n)^T \in \mathbb{R}^n$ such that $\mu_{\mathcal{V}}(\ell) = 0$.

Proposition 2.34 *Any linear map* $\phi : \mathbb{R}^n \to V$ *has the form* $\phi = \mu_{\mathcal{V}}$ *for some list* $\mathcal{V} = v_1, \ldots, v_n$ *of elements of* V*. Thus a linear map* $\mathbb{R}^n \to V$ *is essentially the same thing as a set of* n *elements of* V*.*

Example 2.35 Consider the map $\phi : \mathbb{R}^3 \to M_3(\mathbb{R})$ given by

$$\phi \begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} a & a+b & a \\ a+b & a+b+c & a+b \\ a & a+b & a \end{pmatrix}.$$

Find a list $\mathcal{A}$ of elements of $M_3(\mathbb{R})$ such that $\phi = \mu_{\mathcal{A}}$.

Proposition 2.36 *Suppose* $\mathcal{V}$ *is a list of elements of* V*. Then*

(a) $\mathcal{V}$ *is linearly independent if and only if* $\mu_{\mathcal{V}}$ *is injective;*

(b) $\mathcal{V}$ *spans if and only if* $\mu_{\mathcal{V}}$ *is surjective;*

(c) $\mathcal{V}$ *is a basis if and only if* $\mu_{\mathcal{V}}$ *is bijective.*

Corollary 2.37 *If* V *is an* n*-dimensional vector space, then there is an isomorphism* $\mathbb{R}^n \to V$*.*

In other words, every n-dimensional vector space "looks like" $\mathbb{R}^n$ in some sense. Let's reinterpret some of our earlier results on bases in terms of these maps $\mu_{\mathcal{V}}$. Recall that we have explained that if V and W are finite dimensional vector spaces, then a linear map $V \to W$ is "essentially" given by matrix multiplication.

Proposition 2.38 *If* $\theta : V \to W$ *is a linear map between finite-dimensional vector spaces, with bases* $\mathcal{V} = v_1, \ldots, v_n$ *and* $\mathcal{W} = w_1, \ldots, w_m$*, then* $\mu_{\mathcal{W}}(\phi_A(\mathbf{x})) = \theta(\mu_{\mathcal{V}}(\mathbf{x}))$*, where* $A = (a_{ij})$ *is the matrix computed from* $\theta(v_j) = a_{1j}w_1 + \cdots + a_{mj}w_m$*.*

And we can reinterpret change-of-basis results in this language.

Lemma 2.39 *If* $\mathcal{V}$ *and* $\mathcal{V}'$ *are bases for the* n*-dimensional vector space* V*, and* A *is the change-of-basis matrix, then* $\mu_{\mathcal{V}}(\phi_A(\mathbf{x})) = \mu_{\mathcal{V}'}(\mathbf{x})$ *for* $\mathbf{x} \in \mathbb{R}^n$*.*

These two combine to give a slick proof of the result we saw earlier.

Proposition 2.40 *Let* $\theta : V \to W$ *be a linear map between finite-dimensional vector spaces. Suppose that we have two bases* $\mathcal{V}$ *and* $\mathcal{V}'$ *for* V*, with change-of-basis matrix* B *and two bases* $\mathcal{W}$ *and* $\mathcal{W}'$ *for* W *with change-of-basis matrix* C*. Let* A *be the matrix of* θ *with respect to* $\mathcal{V}$ *and* $\mathcal{W}$*, and let* A' *be the matrix with respect to* $\mathcal{V}'$ *and* $\mathcal{W}'$*. Then* $A' = C^{-1}AB$*.*

3 Inner product spaces and Fourier theory

The axioms for vector spaces which we gave in the first chapter mark our attempt to abstract all the algebraic structure of $\mathbb{R}^n$ into a framework which allows us to prove similar results simultaneously in other situations.

However, $\mathbb{R}^3$ has additional *geometrical* structure, namely some notion of *angle*, which one sees most straightforwardly in vector dot products; the dot product of two vectors in $\mathbb{R}^3$ is the product of their lengths, multiplied by the cosine of the angle between them. Indeed, this allows us to measure the angle in terms of the dot product. If we can abstract the idea of dot product to other vector spaces, we will similarly obtain a notion of angle in other situations, where the geometrical notion of angle seems not to apply.

3.1 Definition of inner products

Definition 3.1 Let V be a vector space over $\mathbb{R}$. An *inner product* on V is a rule that gives a number $\langle u, v\rangle \in \mathbb{R}$ for each $u, v \in V$, with the following properties.

(a) $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ for all $u, v, w \in V$.

(b) $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$ for all $u, v \in V$ and $\alpha \in \mathbb{R}$.

(c) $\langle u, v\rangle = \langle v, u\rangle$ for all $u, v \in V$.

(d) We have $\langle v, v\rangle \geq 0$ for all $v \in V$, and $\langle v, v\rangle = 0$ if and only if $v = 0$.

Given an inner product, we will write $\|v\| = \sqrt{\langle v, v\rangle}$, and call this the *norm* of v.

Unlike most of the other things we have done, this does not immediately generalise to fields K other than $\mathbb{R}$. The reason is that axiom (d) involves the condition $\langle v, v\rangle \geq 0$, and in an arbitrary field K (such as $\mathbb{C}$, for example) we do not have a good notion of positivity (would you say that $1 - i$ was "positive"?). We will indicate very briefly how one can fix things up in the complex case later.

Example 3.2 We can define an inner product on $\mathbb{R}^n$ by

$$\langle (x_1, \ldots, x_n)^T, (y_1, \ldots, y_n)^T\rangle = \sum_{i=1}^{n} x_i y_i = x_1y_1 + x_2y_2 + \cdots + x_ny_n.$$

Notice that if $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ then we can regard $\mathbf{x}$ and $\mathbf{y}$ as $n \times 1$ matrices, so $\mathbf{x}^T$ is a $1 \times n$ matrix, and then $\mathbf{x}^T\mathbf{y}$ is a 1×1 matrix, or in other words a number. This number is just $\langle \mathbf{x}, \mathbf{y}\rangle$. This is most easily explained by example: in the case $n = 4$ we have

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}^T \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix} = x_1y_1 + x_2y_2 + x_3y_3 + x_4y_4 = \langle \mathbf{x}, \mathbf{y}\rangle.$$

Notice that if we use the standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$, and $\mathbf{x}$ is the vector $(x_1 \ldots x_n)^T$, then $\langle \mathbf{x}, \mathbf{e}_i\rangle = x_i$, so taking inner products with the standard basis vectors recovers the coordinates.

Although we will always use the standard inner product on $\mathbb{R}^n$, there are many other inner products. Given any basis $\mathcal{V}$, we can take two vectors $\mathbf{u}$ and $\mathbf{v}$, write them in terms of the basis $\mathcal{V}$, and take the same definition as the product above, but using the co-ordinates with respect to the basis $\mathcal{V}$.

Example 3.3 We can define an inner product on $C[0,1]$ by

$$\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx.$$

Example 3.4 The *Fourier inner product space* is $C[-\pi,\pi]$, the space of continuous functions $[-\pi,\pi]\to\mathbb{R}$, with the inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(t)g(t)\,dt.$$

The verification that this is an inner product is very similar to Example 3.3. Notice that when we do the usual computation of Fourier series, and compute

$$\int_{-\pi}^{\pi} f(t)\cos nt\,dt, \qquad \int_{-\pi}^{\pi} f(t)\sin nt\,dt,$$

we are actually computing the inner products $\langle f(t), \cos nt\rangle$ and $\langle f(t), \sin nt\rangle$; as in Example 3.2, you should view this as recovering the co-ordinates with respect to some basis. We'll make this a little more precise later. For future reference, we recall that $\langle \cos mt, \sin nt\rangle = 0$ for all m, n, and that $\langle \cos mt, \cos nt\rangle = 0$ for all $m \neq n$, and is π if $m = n > 0$ or 2π if $m = n = 0$; similarly $\langle \sin mt, \sin nt\rangle = 0$ for all $m \neq n$, and is π if $m = n > 0$ or 0 if $m = n = 0$. Thus $1, \cos t, \sin t, \cos 2t, \sin 2t, \cos 3t, \sin 3t, \ldots$ is an *orthogonal sequence.*

Example 3.5 We can define an inner product on the space $M_n(\mathbb{R})$ by

$$\langle A, B\rangle = \operatorname{trace}(AB^T).$$

Then

$$\langle A, B\rangle = \operatorname{trace}(AB^T) = \sum_{i=1}^{n} (AB^T)_{ii} = \sum_{i=1}^{n}\sum_{j=1}^{n} A_{ij}B_{ij}.$$

In other words $\langle A, B\rangle$ is the sum of the entries of A multiplied by the corresponding entries in B. Thus, if we identify $M_n(\mathbb{R})$ with $\mathbb{R}^{n^2}$ in the usual way, then our inner product on $M_n(\mathbb{R})$ corresponds to the standard inner product on $\mathbb{R}^{n^2}$.

We now briefly discuss the analogue of inner products for complex vector spaces. Most of the proofs above that certain spaces are inner product spaces use the fact that $x^2 \geq 0$ for all $x \in \mathbb{R}$, and of course this ceases to be true if we work over $\mathbb{C}$ (indeed, x^2 won't even be real in general). But, given a complex number $z = x + iy$, we write $\overline{z}$ for the complex conjugate, which is $x - iy$, and we know that the quantity $|z|^2 = z\overline{z} = x^2 + y^2$ is real and non-negative, and is only zero when $z = 0$. In order that we can use this in our proofs, we make the following definition which replaces the definition of inner product in this setting.

Definition 3.6 *Let V be a vector space over $\mathbb{C}$. A Hermitian form on V is a rule that gives a number $\langle u, v\rangle \in \mathbb{C}$ for each $u, v \in V$, with the following properties:*

(a) $\langle u+v, w\rangle = \langle u, w\rangle + \langle v, w\rangle$ *for all* $u, v, w \in V$.

(b) $\langle \alpha u, v\rangle = \alpha\langle u, v\rangle$ *for all* $u, v \in V$ *and* $\alpha \in \mathbb{C}$.

(c) $\langle u, v\rangle = \overline{\langle v, u\rangle}$ *for all* $u, v \in V$. *In particular,* $\langle u, u\rangle = \overline{\langle u, u\rangle}$, *so* $\langle u, u\rangle$ *is real.*

(d) For all $u \in V$ *we have* $\langle u, u\rangle \geq 0$ *(which is meaningful because* $\langle u, u\rangle \in \mathbb{R}$*), and* $\langle u, u\rangle = 0$ *if and only if* $u = 0$.

Note that (b) and (c) together imply that $\langle u, \alpha v\rangle = \overline{\alpha}\langle u, v\rangle$ *(we say that this is "semilinear in the second variable").*

Given an inner product, we will write $\|u\| = \sqrt{\langle u, u\rangle}$, and call this the *norm* of u.

Let's give some examples of Hermitian forms (proofs are omitted, but are pretty similar to the earlier proofs).

- We can define a Hermitian form on $\mathbb{C}^n$ by

  $$\langle \mathbf{u}, \mathbf{v}\rangle = u_1\overline{v_1} + \cdots + u_n\overline{v_n}.$$

  This gives

  $$\|\mathbf{u}\|^2 = \langle \mathbf{u}, \mathbf{u}\rangle = |u_1|^2 + \cdots + |u_n|^2.$$

- For any $n \times m$ matrix A over $\mathbb{C}$, we let $A^\dagger$ be the complex conjugate of the transpose of A, so for example

  $$\begin{pmatrix} 1+i & 2+i & 3+i \\ 4+i & 5+i & 6+i \end{pmatrix}^\dagger = \begin{pmatrix} 1-i & 4-i \\ 2-i & 5-i \\ 3-i & 6-i \end{pmatrix}.$$

  The above Hermitian form on $\mathbb{C}^n$ can then be rewritten as

  $$\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{v}^\dagger\mathbf{u} = \overline{\mathbf{u}^\dagger\mathbf{v}}.$$

  We can define a Hermitian form on $M_n(\mathbb{C})$ by $\langle A, B\rangle = \mathrm{trace}(B^\dagger A)$. If we identify $M_n(\mathbb{C})$ with $\mathbb{C}^{n^2}$ in the usual way, then this is just the same as the Hermitian form in the example above.

- We can define a Hermitian form on $\mathbb{C}[t]$ by

  $$\langle f, g\rangle = \int_0^1 f(t)\overline{g(t)}\, dt.$$

  This gives

  $$\|f\|^2 = \langle f, f\rangle = \int_0^1 |f(t)|^2\, dt.$$

Our results about inner products are mostly also true for Hermitian forms, but they need to be adjusted slightly by putting complex conjugates or absolute value signs in appropriate places. We will not go through the statements or proofs, but they are mostly rather similar to those over $\mathbb{R}$.

3.2 The Cauchy-Schwarz inequality

If $\mathbf{v}$ and $\mathbf{w}$ are vectors in $\mathbb{R}^2$ or $\mathbb{R}^3$, you should be familiar with the fact that

$$\langle \mathbf{v}, \mathbf{w} \rangle = \|\mathbf{v}\| \, \|\mathbf{w}\| \, \cos(\theta),$$

where θ is the angle between $\mathbf{v}$ and $\mathbf{w}$. In particular, as the cosine lies between -1 and 1, we see that $|\langle \mathbf{v}, \mathbf{w} \rangle| \le \|\mathbf{v}\| \, \|\mathbf{w}\|$. We would like to extend all this to arbitrary inner-product spaces. We'll only consider inner product spaces over $\mathbb{R}$ for the remainder of the course, although there are analogues for Hermitian forms over complex vector spaces.

Theorem 3.7 (The Cauchy-Schwarz inequality) *Let V be an inner product space over $\mathbb{R}$, and let v and w be elements of V. Then*

$$|\langle v, w \rangle| \le \|v\| \, \|w\|,$$

with equality if and only if v and w are linearly dependent.

Example 3.8 We claim that for any vector $\mathbf{x} \in \mathbb{R}^n$, we have

$$|x_1 + \cdots + x_n| \le \sqrt{n}\sqrt{x_1^2 + \cdots + x_n^2}.$$

Example 3.9 We claim that for any continuous function $f : [0,1] \to \mathbb{R}$ we have

$$\left| \int_0^1 (1 - x^2) f(x) \, dx \right| \le \sqrt{\frac{8}{15}} \sqrt{\int_0^1 f(x)^2 \, dx}.$$

Example 3.10 Suppose that A is a nonzero $n \times n$ matrix over $\mathbb{R}$. We claim that

(a) $\operatorname{trace}(A)^2 \le n \operatorname{trace}(AA^T)$, with equality if and only if A is a multiple of the identity.

(b) $\operatorname{trace}(A^2) \le \operatorname{trace}(AA^T)$, with equality if and only if A is either symmetric or antisymmetric.

It is natural to define the "angle" between two elements in an inner product space so that $\langle v, w \rangle = \|v\| \, \|w\| \, \cos(\theta)$, just as is true for $\mathbb{R}^3$.

Definition 3.11 Let V be an inner product space over $\mathbb{R}$, and let v and w be nonzero elements of V, so $\|v\| \, \|w\| > 0$. Put $c = \langle v, w \rangle / (\|v\| \, \|w\|)$. The Cauchy-Schwarz inequality tells us that $-1 \le c \le 1$, so there is a unique angle $\theta \in [0, \pi]$ such that $\cos(\theta) = c$. We call this the *angle between v and w*. In particular, we'll say that v and w are *orthogonal* when the angle between them is $\theta = \pi/2$, or, equivalently, when $\langle v, w \rangle = 0$.

Example 3.12 Take $V = C[0,1]$ (with the usual inner product), and $v(t) = 1$, and $w(t) = t$. Calculate the angle between v and w.

Example 3.13 Take $V = M_3(\mathbb{R})$ (with the usual inner product) and

$$A = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 2 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

Calculate the angle between A and B.

3.3 Orthogonality and projections

Definition 3.14 Let V be a vector space with inner product. We say that a set $\mathcal{V} = \{v_1, \ldots, v_n\}$ of elements of V is *orthogonal* if we have $\langle v_i, v_j \rangle = 0$ for all $i \neq j$. We say that the set is *strictly orthogonal* if it is orthogonal, and all the elements v_i are nonzero. We say that the set is *orthonormal* if it is orthogonal, and also $\langle v_i, v_i \rangle = 1$ for all i (we say that v_i is a *unit vector* in this case).

Remark 3.15 If $\mathcal{V}$ is a strictly orthogonal set then we can define an orthonormal set $\hat{v}_1, \ldots, \hat{v}_n$ by $\hat{v}_i = v_i/\|v_i\|$.

Example 3.16 The standard basis $\mathbf{e}_1, \ldots, \mathbf{e}_n$ for $\mathbb{R}^n$ is an orthonormal set.

As a consequence of the material on the Fourier theory handout, we also deduce the following example.

Example 3.17 The sequence $1, \cos t, \sin t, \cos 2t, \sin 2t, \cos 3t, \sin 3t, \ldots$ is a strictly orthogonal set in the Fourier inner product space $C[-\pi, \pi]$. Indeed, the sequence

$$\frac{1}{\sqrt{2\pi}}, \frac{1}{\sqrt{\pi}}\cos t, \frac{1}{\sqrt{\pi}}\sin t, \frac{1}{\sqrt{\pi}}\cos 2t, \frac{1}{\sqrt{\pi}}\sin 2t, \ldots, \frac{1}{\sqrt{\pi}}\cos nt, \frac{1}{\sqrt{\pi}}\sin nt, \ldots$$

is orthonormal.

Combining the handout with the ideas from the previous section, we can work out angles between certain trigonometric functions easily.

Example 3.18 Consider the Fourier inner product space $C[-\pi, \pi]$, with its usual inner product

$$\langle f, g \rangle = \int_{-\pi}^{\pi} f(t)g(t)\,dt.$$

If $f(t) = \cos 5t$, and $g(t) = \cos 2t \cos 3t$, calculate the angle between f and g.

Let's now prove some useful short technical results. We begin with a version of Pythagoras's Theorem.

Lemma 3.19 *Let $v_1, \ldots, v_n$ be an orthogonal set, and put $v = v_1 + \cdots + v_n$. Then*

$$\|v\| = \sqrt{\|v_1\|^2 + \cdots + \|v_n\|^2}.$$

We are going to construct bases for inner product spaces and their subspaces consisting of orthogonal sequences. It turns out that such sequences are automatically linearly independent.

Lemma 3.20 *Any strictly orthogonal sequence is linearly independent.*

Definition 3.21 Let V be a vector space with inner product, and let W be a subspace. We then put

$$W^{\perp} = \{v \in V \mid \langle v, w \rangle = 0 \text{ for all } w \in W\}.$$

This is called the *orthogonal complement* of W.

We'll show in the next two results that V is the direct sum of W and its complement $W^{\perp}$. Recall that this means that $W \cap W^{\perp} = \{0\}$ and $V = W + W^{\perp}$.

Lemma 3.22 *We have* $W \cap W^{\perp} = \{0\}$.

Proposition 3.23 *Let V be a vector space with inner product, and let W be a subspace. Suppose that we have a strictly orthogonal set $\mathcal{W} = \{w_1, \ldots, w_p\}$ that spans W, and we define*

$$\pi(v) = \frac{\langle v, w_1 \rangle}{\langle w_1, w_1 \rangle} w_1 + \cdots + \frac{\langle v, w_p \rangle}{\langle w_p, w_p \rangle} w_p$$

(for all $v \in V$). Then $\pi(v) \in W$ and $v - \pi(v) \in W^{\perp}$, so $v = \pi(v) + (v - \pi(v)) \in W + W^{\perp}$. In particular, we have $V = W \oplus W^{\perp}$.

The map $\pi : V \to W$ should be regarded as taking a vector, and projecting it onto W, in the same way that we can take a vector $(x, y)^T \in \mathbb{R}^2$ and map it to $(0, y)^T$ on the y-axis. Let's show that $\pi(v)$ is the closest vector in W to v.

Proposition 3.24 *Let $\mathcal{W}$ and π be as in Proposition 3.23. Then $\pi(v)$ is the point in W that is closest to v.*

You may recall from MAS202 that Parseval's Theorem is an identity for Fourier coefficients. We will discuss this a little further right at the end of the course. Here's a version for more general inner product spaces.

Proposition 3.25 *Let V be a vector space with inner product, and let $\mathcal{W} = \{w_1, \ldots, w_p\}$ be an orthonormal set in V. Then for any $v \in V$ we have*

$$\|v\|^2 \geq \sum_{i=1}^{p} \langle v, w_i \rangle^2.$$

Moreover, this inequality is actually an equality if and only if $v \in \mathrm{sp}(\mathcal{W})$.

3.4 The Gram-Schmidt procedure

We'll now develop a very important tool for calculations, the *Gram-Schmidt procedure.* As a consequence, we will see that any finite-dimensional inner product space has an orthogonal (and therefore orthonormal) basis. The idea is to take any basis, and to adjust it so that it becomes orthogonal – we run through the basis elements in turn; each time we subtract off some linear combination of the previous basis elements in such a way that it becomes orthogonal to them, and continue this until the end.

Theorem 3.26 *Let V be a vector space with inner product, and let $\mathcal{U} = u_1, \ldots, u_n$ be a linearly independent list of elements of V. Then there is a strictly orthogonal sequence $\mathcal{V} = v_1, \ldots, v_n$ such that $\mathrm{sp}(v_1, \ldots, v_i) = \mathrm{sp}(u_1, \ldots, u_i)$ for all i.*

Of course, we then see that there is an *orthonormal* sequence $\hat{v}_1, \ldots, \hat{v}_n$ such that $\mathrm{sp}(\hat{v}_1, \ldots, \hat{v}_i) = \mathrm{sp}(u_1, \ldots, u_i)$ for all i, simply by finding a strictly orthogonal set $v_1, \ldots, v_n$ as in the theorem, and putting $\hat{v}_i = v_i/\|v_i\|$ as above.

Essentially, the idea of the proof is that we take $v_k = u_k - \lambda_1 v_1 - \cdots - \lambda_{k-1} v_{k-1}$, and choose $\lambda_1, \ldots, \lambda_{k-1}$ to make this orthogonal to $v_1, \ldots, v_{k-1}$. To make it orthogonal to v_j, we note that

$$\langle v_k, v_j \rangle = \langle u_k, v_j \rangle - \lambda_1 \langle v_1, v_j \rangle - \cdots - \lambda_{k-1} \langle v_{k-1}, v_j \rangle,$$

and as $v_1, \ldots, v_{k-1}$ are orthogonal, $\langle v_i, v_j \rangle = 0$ except when $i = j$, so that

$$\langle v_k, v_j \rangle = \langle u_k, v_j \rangle - \lambda_j \langle v_j, v_j \rangle,$$

so v_k and v_j are orthogonal if and only if we choose $\lambda_j = \frac{\langle u_k, v_j \rangle}{\langle v_j, v_j \rangle}$.

Thus

$$v_k = u_k - \frac{\langle u_k, v_1 \rangle}{\langle v_1, v_1 \rangle} v_1 - \cdots - \frac{\langle u_k, v_{k-1} \rangle}{\langle v_{k-1}, v_{k-1} \rangle} v_{k-1}.$$

Example 3.27 Consider the following elements of $\mathbb{R}^5$:

$$u_1 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0 \\ 1 \\ 1 \\ 0 \\ 0 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 1 \\ 0 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}.$$

Apply the Gram-Schmidt procedure to get an orthogonal basis for $U = \mathrm{sp}(u_1, u_2, u_3, u_4)$.

Example 3.28 Consider the space $V = \mathbb{R}[x]_{\leq 2}$ with inner product $\langle p, q \rangle = \int_{-1}^{1} p(x)q(x)\,dx$. Apply the Gram-Schmidt procedure to the usual basis $1, x, x^2$ to get an orthonormal basis for V.

Example 3.29 Consider the matrix $P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, and let V be the space of 3×3 symmetric matrices of trace zero. Find the matrix $Q \in V$ that is closest to P.

3.5 Adjoints of linear maps

Remark 3.30 Recall that linear maps are essentially just matrices; we've seen that every linear map (between finite-dimensional vector spaces) is represented by a matrix after choosing bases. One simple operation that we can do on an $m \times n$-matrix is to transpose it. This gives an $n \times m$-matrix. Thus the transpose of a matrix of a linear map from $\mathbb{R}^n \to \mathbb{R}^m$ is the matrix of a linear map $\mathbb{R}^m \to \mathbb{R}^n$. This map has another property too. Let A be an $n \times m$ matrix over $\mathbb{R}$, giving a linear map $\phi_A : \mathbb{R}^m \to \mathbb{R}^n$ by $\phi_A(\mathbf{v}) = A\mathbf{v}$. The transpose of A is then an $m \times n$ matrix, giving a linear map $\phi_{A^T} : \mathbb{R}^n \to \mathbb{R}^m$. If $\mathbf{u} \in \mathbb{R}^n$, and $\mathbf{v} \in \mathbb{R}^m$, we have

$$\langle \phi_A(\mathbf{u}), \mathbf{v} \rangle = \langle A\mathbf{u}, \mathbf{v} \rangle = (A\mathbf{u})^T \mathbf{v} = \mathbf{u}^T A^T \mathbf{v} = \langle \mathbf{u}, A^T \mathbf{v} \rangle = \langle \mathbf{u}, \phi_{A^T}(\mathbf{v}) \rangle.$$

Definition 3.31 Let V and W be real vector spaces with inner products. Let $\phi : V \to W$ and $\psi : W \to V$ be linear maps (over $\mathbb{R}$). We say that ϕ is *adjoint* to ψ if we have $\langle \phi(v), w \rangle = \langle v, \psi(w) \rangle$ for all $v \in V$ and $w \in W$. We write $\psi = \hat{\phi}$ to indicate that ψ is the adjoint of ϕ.

In particular, ϕ_A and ϕ_{A^T} in Remark 3.30 are adjoint.

Example 3.32 Consider the vector spaces $\mathbb{R}^2$ (with the usual inner product) and $\mathbb{R}[x]_{\leq 2}$ (with inner product $\langle f, g\rangle = \int_0^1 f(x)g(x)\,dx$). Define maps $\phi : \mathbb{R}[x]_{\leq 2} \to \mathbb{R}^2$ and $\psi : \mathbb{R}^2 \to \mathbb{R}[x]_{\leq 2}$ by

$$\phi(f) = \begin{pmatrix} f(0) \\ f(1) \end{pmatrix}; \qquad \psi \begin{pmatrix} p \\ q \end{pmatrix} = (30p + 30q)x^2 - (36p + 24q)x + (9p + 3q).$$

We claim that ϕ is adjoint to ψ.

Proposition 3.33 *Let V and W be finite-dimensional inner product spaces. Let $\phi : V \to W$ be a linear map. Then there is a unique map $\psi : W \to V$ that is adjoint to ϕ.*

Definition 3.34 Let V be a finite-dimensional vector space over $\mathbb{C}$ with Hermitian form. A *self-adjoint operator* on V is a linear map $\phi : V \to V$ such that ϕ is its own adjoint, i.e., $\langle \phi(v_1), v_2\rangle = \langle v_1, \phi(v_2)\rangle$.

Theorem 3.35 *If $\phi : V \to V$ is a self-adjoint operator, then every eigenvalue of ϕ is real.*

Theorem 3.36 *If $\phi : V \to V$ is a self-adjoint operator, then one can choose an orthonormal basis $\mathcal{V} = v_1, \ldots, v_n$ for V such that each v_i is an eigenvector of ϕ.*

The following lemma will be useful in the proof.

Lemma 3.37 *Let $\phi : V \to V$ be a self-adjoint operator, and let W be a subspace of V such that $\phi(W) \subseteq W$ (i.e., $\phi(w) \in W$ for all $w \in W$). Then $\phi(W^\perp) \subseteq W^\perp$.*

From Remark 3.30, an example of a self-adjoint operator is a real symmetric matrix; we deduce that its eigenvalues are real, and that the eigenvectors are orthogonal.

3.6 Fourier series

We now begin to re-interpret some aspects of classical Fourier Theory from the viewpoint of this course. This gives us little new – no beautiful new formulae for π, for example! – but it should reassure us that Fourier Theory fits well into this topic.

In MAS202, you defined the Fourier series of a continuous function f with period 2π. The coefficients a_n and b_n in the Fourier series $S_f(t)$ are chosen so as to ensure that $\int_{-\pi}^{\pi} (f(t) - S_f(t)) \cos nt\,dt = 0$ and $\int_{-\pi}^{\pi} (f(t) - S_f(t)) \sin nt\,dt = 0$ for all n, so that $f(t) - S_f(t)$ is orthogonal to each $\cos nt$ and $\sin nt$. This is certainly good evidence that $f(t) - S_f(t)$ might be the zero function, but it is by no means a proof: there might be nonzero functions orthogonal to each of the trigonometric functions. In this section, we will shed more light on this.

Definition 3.38 We say that a function $f : \mathbb{R} \to \mathbb{R}$ is *periodic* if $f(t + 2\pi) = f(t)$ for all $t \in \mathbb{R}$; such functions are then determined by their values on $[-\pi, \pi]$. We let $C^{2\pi}$ be the set of all continuous periodic functions from $\mathbb{R}$ to $\mathbb{R}$. This is a vector space over $\mathbb{R}$. We define an inner product on $C^{2\pi}$ by

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(t)g(t).dt$$

Recall that a *trigonometric polynomial of degree* n is a sum of the form

$$\frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kt + b_k \sin kt).$$

In some sense, these are polynomials, as the relations

$$\begin{aligned} \cos kt + \cos(k-2)t &= 2\cos(k-1)t\,\cos t \\ \sin(k+1)t - \sin(k-1)t &= 2\cos kt\,\sin t \end{aligned}$$

imply inductively that $\cos kt$ can be written as a polynomial of degree k in $\cos t$, and that $\sin(k+1)t$ is the product of $\sin t$ and a polynomial of degree k in $\cos t$. This shows that the space of functions spanned by

$$\mathcal{B}_1 = \{1, \cos t, \cos 2t, \ldots, \cos nt, \sin t, \sin 2t, \ldots, \sin nt\}$$

is the same as that spanned by

$$\mathcal{B}_2 = \{1, \cos t, \cos^2 t, \ldots, \cos^n t, \sin t, \sin t \cos t, \ldots, \sin t \cos^{n-1} t\}.$$

Indeed, the members of $\mathcal{B}_1$ are linearly independent (by Lemma 3.20). On the other hand, each member of $\mathcal{B}_1$ can be expressed as a linear combination of elements of $\mathcal{B}_2$ by the observation above, so $\operatorname{span}(\mathcal{B}_1) \subseteq \operatorname{span}(\mathcal{B}_2)$. But $\mathcal{B}_1$ consists of $2n+1$ linearly independent functions, so spans a space of dimension $2n+1$. As $|\mathcal{B}_2| = 2n+1$, $\operatorname{span}(\mathcal{B}_2)$ has dimension at most $2n+1$, and therefore it must equal $2n+1$ as it contains $\operatorname{span}(\mathcal{B}_1)$. So $\mathcal{B}_1$ and $\mathcal{B}_2$ have the same span, and either may be used as a basis for the trigonometric polynomials of degree n.

Note that the even trigonometric polynomials are those spanned by $\{\cos nt\}$, or equivalently by $\{\cos^n t\}$, and these really are therefore polynomial expressions in $\cos t$.

Let's make the following abbreviations:

$$\begin{aligned} c_n(t) &= \cos nt \qquad (\text{for } n \geq 0) \\ s_n(t) &= \sin nt \qquad (\text{for } n > 0). \end{aligned}$$

We know that the sequence $\mathcal{C}_n = c_0, c_1, s_1, c_2, s_2, \ldots, c_n, s_n$ is an orthogonal sequence. It satisfies $\|s_k\|^2 = \pi = \|c_k\|^2$ for $k > 0$, and $\|c_0\|^2 = 2\pi$. Denote the set of linear combinations of $\{c_0, c_1, s_1, \ldots, c_n, s_n\}$, the trigonometric polynomials of degree n, by $\mathcal{T}_{\leq n} = \operatorname{sp}(\mathcal{C}_n) \subset C^{2\pi}$. Given f, let's write $\pi_n(f)$ for the projection onto $\mathcal{T}_{\leq n}$; by Proposition 3.23,

$$\pi_n(f) = \frac{1}{2\pi}\langle f, c_0\rangle c_0 + \frac{1}{\pi}\sum_{k=1}^{n}\langle f, c_k\rangle c_k + \frac{1}{\pi}\sum_{k=1}^{n}\langle f, s_k\rangle s_k.$$

By Proposition 3.24, $\pi_n(f)$ is the closest point to f lying in $\mathcal{T}_{\leq n}$, so $\|f - \pi_n(f)\|$ can be regarded as the distance from f to $\mathcal{T}_{\leq n}$. Recall that the usual Fourier coefficients are just $a_k = \langle f, c_k\rangle/\pi$ and $b_k = \langle f, s_k\rangle/\pi$, so $\pi_n(f)$ can also be viewed as the truncated Fourier series.

Theorem 3.39 (L^2-convergence) *For any* $f \in C^{2\pi}$ *we have* $\|f - \pi_n(f)\| \to 0$ *as* $n \to \infty$.

The proof is not examinable and will not be covered in lectures. See the appendix on the course web page if interested in the details.

The theorem says that by taking n to be sufficiently large, we can make the distance from f to $\mathcal{T}_{\leq n}$ as small as we like. In other words, f can be very well approximated by a trigonometric polynomial of sufficiently high degree.

Corollary 3.40 (Parseval's Identity) *For any $f \in C^{2\pi}$ we have*

$$\|f\|^2 = \langle f, c_0\rangle^2/2\pi + \sum_{k=1}^{\infty}\langle f, c_k\rangle^2/\pi + \sum_{k=1}^{\infty}\langle f, s_k\rangle^2/\pi.$$

We also would like to know that the trigonometric polynomials somehow form all of $C^{2\pi}$, i.e., that there aren't any other periodic functions orthogonal to all the ones we already know about. This is also a corollary of Theorem 3.39.

Corollary 3.41 *If $f \in C^{2\pi}$, and if*

$$\int_{-\pi}^{\pi} f(t)\cos nt\,dt = \int_{-\pi}^{\pi} f(t)\sin nt\,dt = 0$$

for each $n = 0, 1, 2, \ldots$, then $f(t) = 0$ for all $t \in [-\pi, \pi]$.

Corollary 3.42 *If $f \in C^{2\pi}$, and S_f denotes its Fourier series, then $S_f(t) = f(t)$ for all $t \in [-\pi, \pi]$.*

The End