1
An Overview

This chapter gives an overview of the entire book. Since our focus turns directly to wavelets only in Chapter 5, about halfway through, beginning with an overview is useful because it enables us early on to convey an idea of what wavelets are and of the mathematical setting within which we will study them.

The idea of the *orthonormality* of a collection of vectors, together with some closely related ideas, is central to the chapter. These ideas should be familiar at least in the finite-dimensional Euclidean spaces $\mathbb{R}^n$ (and certainly in $\mathbb{R}^2$ and $\mathbb{R}^3$ in particular), but after revising the Euclidean case we will go on to examine the same ideas in certain spaces of infinite dimension, especially spaces of functions. In many ways, the central chapters of the book constitute a systematic and general investigation of these ideas, and the theory of wavelets is, from one point of view, an application of this theory.

Because this chapter is only an overview, the discussion will be rather informal, and important details will be skated over quickly or suppressed altogether, but by the end of the chapter the reader should have a broad idea of the shape and content of the book.

1.1 Orthonormality in $\mathbb{R}^n$

Recall that the **inner product** $\langle a, b\rangle$ of two vectors

$$a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix} \quad \text{and} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

in $\mathbb{R}^n$ is defined by

$$\langle a, b \rangle = a_1 b_1 + a_2 b_2 + \cdots + a_n b_n = \sum_{i=1}^{n} a_i b_i.$$

Also, the **norm** $\|a\|$ of a is defined by

$$\|a\| = \sqrt{a_1^2 + a_2^2 + \cdots + a_n^2} = \left(\sum_{i=1}^{n} a_i^2 \right)^{\frac{1}{2}} = \langle a, a \rangle^{\frac{1}{2}}.$$

➤ A few remarks on notation and terminology are appropriate before we go further. First, in elementary linear algebra, special notation is often used for vectors, especially vectors in Euclidean space: a vector might, for example, be signalled by the use of bold face, as in $\boldsymbol{a}$, or by the use of a special symbol above or below the main symbol, as in $\vec{a}$ or $\underset{\sim}{a}$. This notational complication, however, is logically unnecessary, and is normally avoided in more advanced work.

Second, you may be used to the terminology *dot product* and the corresponding notation $a \cdot b$, but we will always use the phrase *inner product* and the 'angle-bracket' notation $\langle a, b \rangle$. Similarly, $\|a\|$ is sometimes referred to as the *length* or *magnitude* of a, but we will always refer to it as the *norm*.

The inner product can be used to define the component and the projection of one vector on another. Geometrically, the **component** of a on b is formed by projecting a perpendicularly onto b and measuring the length of the projection; informally, the component measures how far a 'sticks out' in the direction of b, or 'how long a looks' from the perspective of b. Since this quantity should depend only on the direction of b and not on its length, it is convenient to define it first when b is **normalised**, or a **unit vector**, that is, satisfies $\|b\| = 1$. In this case, the component of a on b is defined simply to be

$$\langle a, b \rangle.$$

If b is not normalised, then the component of a on b is obtained by applying the definition to its **normalisation** $(1/\|b\|)\, b$, which *is* a unit vector, giving the expression

$$\left\langle a, \frac{1}{\|b\|} b \right\rangle = \frac{1}{\|b\|} \langle a, b \rangle$$

for the component (note that we must assume that b is not the zero vector here, to ensure that $\|b\| \neq 0$).

Since the component is defined as an inner product, its value can be any real number – positive, negative or zero. This implies that our initial description of the component above was not quite accurate: the component represents not just the *length* of the first vector from the perspective of the second, but also, according to its sign, how the first vector is *oriented* with respect to the second. (For the special case when the component is 0, see below.)

It is now simple to define the **projection** of a on b; this is the vector whose direction is given by b and whose length and orientation are given by the component of a on b. Thus if $\|b\| = 1$, then the projection of a on b is given by

$$\langle a, b\rangle\, b,$$

and for general non-zero b by

$$\left\langle a, \frac{1}{\|b\|} b \right\rangle \frac{1}{\|b\|} b = \frac{1}{\|b\|^2} \langle a, b\rangle\, b.$$

Vectors a and b are said to be **orthogonal** if $\langle a, b\rangle = 0$, and a collection of vectors is **orthonormal** if each vector in the set has norm 1 and every two distinct elements from the set are orthogonal. If a set of vectors is indexed by the values of a subscript, as in $v_1, v_2, \ldots$, say, then the orthonormality of the set can be expressed conveniently using the Kronecker delta: the set is orthonormal if and only if

$$\langle v_i, v_j\rangle = \delta_{i,j} \quad \text{for all } i, j,$$

where the **Kronecker delta** $\delta_{i,j}$ is defined to be 1 when $i = j$ and 0 otherwise (over whatever is the relevant range of i and j).

All the above definitions are purely *algebraic*: they use nothing more than the algebraic operations permitted in a vector space and in the field of scalars $\mathbb{R}$. However, the definitions also have clear *geometric* interpretations, at least in $\mathbb{R}^2$ and $\mathbb{R}^3$, where we can visualise the geometry (see Exercise 1.2); indeed, we used this fact explicitly at a number of points in the discussion (speaking of 'how long' one vector looks from another's perspective and of the 'length' and 'orientation' of a vector, for example). We can attempt a somewhat more comprehensive list of such geometric interpretations as follows.

- A vector $a = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}$ is just a point in $\mathbb{R}^n$ (whose coordinates are of course the numbers $a_1, a_2, \ldots, a_n$).
- The norm $\|a\|$ is the Euclidean distance of the point a from the origin, and more generally $\|a - b\|$ is the distance of the point a from the point b. (Given the formula defining the norm, this is effectively just a reformulation of Pythagoras' theorem; see Exercise 1.2.)
- For a unit vector b, the inner product $\langle a, b\rangle$ gives the magnitude and orientation of a when seen from b.

- The inner product and norm are related by the formula

$$\cos\theta = \frac{\langle a, b\rangle}{\|a\|\,\|b\|},$$

  where θ is the angle between the (non-zero) vectors a and b.
- Non-zero vectors a and b are orthogonal precisely when they are perpendicular, corresponding to the case $\cos\theta = 0$ above.
- A collection of vectors is orthonormal if its members have length 1 and are mutually perpendicular.

➤ It is reasonable to ask what we can make of these geometric interpretations in $\mathbb{R}^n$ when $n > 3$ and we can no longer visualise the situation. Can we sensibly talk about vectors being 'perpendicular', or about the 'lengths' of vectors, in spaces that we cannot visualise? The algebra works similarly in all dimensions, but does the geometry?

This question is perhaps as much philosophical as mathematical, but the experience and consensus of mathematicians is that the geometric terminology and intuition which are so central to our understanding in low dimensions are simply too valuable to discard in higher dimensions, and it is therefore used uniformly, whatever the dimension of the space. One of the remarkable and beautiful aspects of linear algebra is that the geometric ideas which are so obviously meaningful and useful in $\mathbb{R}^2$ and $\mathbb{R}^3$ play just as significant a role in higher-dimensional Euclidean spaces.

Further, as hinted earlier, an underlying theme of this book is to study how the same circle of ideas – of inner products, norms, orthogonality and orthonormality, and so on – plays a vital role in the study of certain spaces of infinite dimension, and in particular in the theory of wavelets.

Let us now look at two examples in $\mathbb{R}^3$.

The first, though extremely simple, is nevertheless fundamental. Consider the vectors

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix} \quad \text{and} \quad e_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

in $\mathbb{R}^3$. These vectors form what is usually called the **standard basis** for $\mathbb{R}^3$; they form a basis, by the definition of that term, because they are linearly independent and every vector in $\mathbb{R}^3$ can be expressed uniquely as a linear combination of them. But what is especially of interest here, though it is mathematically trivial to check, is that e_1, e_2, e_3 form an *orthonormal* basis: $\langle e_i, e_j\rangle = \delta_{i,j}$ for $i, j = 1, 2, 3$. It follows from the orthonormality that in the expression for a given vector as a linear combination of e_1, e_2, e_3, the coefficients in the combination are the respective components of the vector on e_1, e_2, e_3. Specifically, if

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix},$$

then the component of a on e_i is $\langle a, e_i \rangle = a_i$ for $i = 1, 2, 3$, and we have

$$a = a_1 e_1 + a_2 e_2 + a_3 e_3 = \sum_{i=1}^{3} \langle a, e_i \rangle \, e_i.$$

For our second example, consider the three vectors

$$b_1 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad b_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \quad \text{and} \quad b_3 = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{pmatrix}.$$

It is simple to verify that these vectors are orthonormal. From orthonormality, it follows that the three vectors are linearly independent, and then, since $\mathbb{R}^3$ has dimension 3, that they form a basis for $\mathbb{R}^3$ (see Exercise 1.3). However, our main point here is to observe that we can express an arbitrary vector c as a linear combination of b_1, b_2 and b_3 nearly as easily as in the first example, by computing components.

For example, take

$$c = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix}.$$

Projecting c onto each of b_1, b_2 and b_3, we obtain the components

$$\langle c, b_1 \rangle = -\frac{1}{2} - \frac{3}{\sqrt{2}}, \quad \langle c, b_2 \rangle = \frac{3}{\sqrt{2}} \quad \text{and} \quad \langle c, b_3 \rangle = -\frac{1}{2} + \frac{3}{\sqrt{2}},$$

respectively, and these quantities are the coefficients required to express c as a linear combination of b_1, b_2 and b_3. That is,

$$c = \langle c, b_1 \rangle \, b_1 + \langle c, b_2 \rangle \, b_2 + \langle c, b_3 \rangle \, b_3 = \sum_{i=1}^{3} \langle c, b_i \rangle \, b_i,$$

or, numerically,

$$\begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \left(-\tfrac{1}{2} - \tfrac{3}{\sqrt{2}}\right) \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ -\frac{1}{\sqrt{2}} \end{pmatrix} + \tfrac{3}{\sqrt{2}} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} + \left(-\tfrac{1}{2} + \tfrac{3}{\sqrt{2}}\right) \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ \frac{1}{\sqrt{2}} \end{pmatrix},$$

and of course we can check this easily by direct expansion.

➤ Although the algorithmic aspects of these ideas are not of primary interest here, it is worth making one observation about them in passing. If a basis is orthonormal, then the coefficient of a vector with respect to any given basis vector *depends only on that* *basis vector*. This is in marked contrast to the case when the basis is not orthonormal; the value of the coefficient then potentially involves *all* of the basis vectors, and finding the value may involve much more computation.

Thus orthonormality makes the *algebra* in the second example almost as simple as in the first example. Furthermore, the *geometry* is almost identical to that of the first example too, even though it is somewhat harder to visualise. The three vectors b_1, b_2 and b_3 are orthonormal, and hence define a system of mutually perpendicular axes in $\mathbb{R}^3$, just like the standard basis vectors e_1, e_2 and e_3, and the coefficients that we found for c by computing components are nothing other than the coordinates of the point c on these axes.

➤ The axes defined by b_1, b_2 and b_3, or any orthonormal basis in $\mathbb{R}^3$, can be obtained by rotation of the usual axes through some angle around some axis through the origin. However, while the usual coordinate system is a right-handed system, the coordinate system defined by an arbitrarily chosen orthogonal basis may be right- or left-handed. A linear algebra course typically explains how to compute the axis and the angle of rotation, as well as to determine the handedness of a system.

1.2 Some Infinite-Dimensional Spaces

1.2.1 Spaces of Sequences

A natural way of trying to translate our ideas so far into an infinite-dimensional space is to consider a vector space of 'infinity-tuples', instead of n-tuples for some $n \in \mathbb{N}$. Our vectors are thus infinitely long columns of the form

$$a = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \\ \vdots \end{pmatrix},$$

and we will denote the resulting vector space by $\mathbb{R}^\infty$.

We make two simple points to begin with. First, column notation for vectors becomes increasingly inconvenient as the columns become longer, so we will switch henceforth to row notation instead. Second, we already have a standard name for infinitely long row vectors

$$a = (a_1, a_2, a_3, \dots) \in \mathbb{R}^\infty :$$

they are simply **sequences**. Thus, $\mathbb{R}^\infty$ is *the vector space of all sequences of real numbers*.

There is no difficulty in checking that $\mathbb{R}^\infty$ satisfies all the required conditions to be a vector space. We will not do this here, but we will glance briefly at a couple of illustrative cases and refer to Chapter 2 for details.

The axiom of closure under vector addition requires that when we add two vectors in the space, the result is again a vector in the space. This is clear for $\mathbb{R}^\infty$, provided that we define the sum of two vectors in the obvious way, following the definition in $\mathbb{R}^n$. Thus if

$$a = (a_1, a_2, a_3, \dots) \quad \text{and} \quad b = (b_1, b_2, b_3, \dots)$$

are in $\mathbb{R}^\infty$, then we define

$$a + b = (a_1 + b_1, a_2 + b_2, a_3 + b_3, \dots);$$

that is, addition of vectors in $\mathbb{R}^\infty$ is defined *entry by entry* or *entry-wise*. Closure is now obvious: each entry $a_n + b_n$ is a real number, by the axiom of closure for addition in $\mathbb{R}$, so $a + b$ is again an element of $\mathbb{R}^\infty$. For the axiom of commutativity of vector addition, which requires that $a + b = b + a$, we have

$$\begin{aligned} a + b &= (a_1, a_2, a_3, \dots) + (b_1, b_2, b_3, \dots) \\ &= (a_1 + b_1, a_2 + b_2, a_3 + b_3, \dots) \\ &= (b_1 + a_1, b_2 + a_2, b_3 + a_3, \dots) \\ &= (b_1, b_2, b_3, \dots) + (a_1, a_2, a_3, \dots) \\ &= b + a. \end{aligned}$$

Notice that the central step in this argument is an application of the law of commutativity of addition in $\mathbb{R}$, paralleling the way in which the argument for the closure axiom worked.

Now let us investigate how we might define an inner product and a norm in $\mathbb{R}^\infty$. Given vectors (that is, sequences) a and b as above, we would presumably wish, following our procedure in $\mathbb{R}^n$, to define their inner product by

$$\langle a, b \rangle = a_1 b_1 + a_2 b_2 + a_3 b_3 + \cdots = \sum_{n=1}^{\infty} a_n b_n,$$

and then to go on to say that a and b are orthogonal if $\langle a, b \rangle = 0$. But there is a problem: *the infinite series* $\sum_{n=1}^{\infty} a_n b_n$ *does not converge for all pairs of sequences* a *and* b*, so the proposed inner product is not defined for all pairs of vectors*.

It is reasonable to try to solve this problem pragmatically by simply removing all the troublesome vectors from the space, that is, by working in the largest subspace of $\mathbb{R}^\infty$ in which all of the required sums $\sum_{n=1}^{\infty} a_n b_n$ converge.

Now this description does not quite constitute a definition of the desired subspace, since it involves a condition on pairs of vectors and does not directly give us a criterion for the membership of any single vector. However, there is in fact such a criterion: the space can be defined directly as the set of all sequences $a = (a_1, a_2, a_3, \dots)$ for which the sum

$$\sum_{n=1}^{\infty} a_n^2$$

converges. This new space is denoted by ℓ^2, so we have

$$\ell^2 = \left\{ (a_1, a_2, a_3, \dots) \in \mathbb{R}^\infty : \sum_{n=1}^{\infty} a_n^2 \text{ converges} \right\}.$$

➤ The name of the space is usually read as 'little-ℓ-2', the word 'little' being needed because as we will see soon there are also 'capital-ℓ-2' or 'big-ℓ-2' spaces, which use an 'L' rather than an 'ℓ'. (Note that in some sources the '2' is written as a subscript rather than a superscript.)

It requires proof that this membership criterion for ℓ^2 solves our original problem – that ℓ^2 is a vector space and that all the desired inner products now lead to convergent sums – but this is left for Chapter 2. Notice, at least, that the norm can certainly now be defined unproblematically by the formula

$$\|a\| = \sqrt{a_1^2 + a_2^2 + \cdots} = \left(\sum_{n=1}^{\infty} a_n^2 \right)^{\frac{1}{2}} = \langle a, a \rangle^{\frac{1}{2}}.$$

Thus we have found what, after checking, turn out to be satisfactory definitions of an inner product and a norm in ℓ^2, and therefore also of the notions of orthogonality and orthonormality. What can we do with all this?

Let $a = (a_1, a_2, a_3, \dots, a_n)$ be in $\mathbb{R}^n$ for any fixed n (for notational convenience we now adopt row notation for vectors in $\mathbb{R}^n$). Then a can be expressed as a linear combination of the n standard orthonormal basis vectors

$$\begin{aligned} e_1 &= (1, 0, 0, \dots, 0), \\ e_2 &= (0, 1, 0, \dots, 0), \\ e_3 &= (0, 0, 1, \dots, 0), \\ &\vdots \\ e_n &= (0, 0, 0, \dots, 1), \end{aligned}$$

in the form

$$a = a_1 e_1 + a_2 e_2 + a_3 e_3 + \cdots + a_n e_n = \sum_{i=1}^{n} a_i e_i,$$

as we noted in detail in the case of $\mathbb{R}^3$, in the first example of the previous section.

It seems natural to try to do the same thing in ℓ^2, since one of our aims was to try to extend finite-dimensional ideas to the infinite-dimensional case. Thus we would hope to say that any $a = (a_1, a_2, a_3, \dots) \in \ell^2$ can be written as a linear combination of *the infinite orthonormal sequence of vectors*

$$e_1 = (1, 0, 0, 0, \dots), \quad e_2 = (0, 1, 0, 0, \dots), \quad e_3 = (0, 0, 1, 0, \dots), \dots$$

in the form

$$a = a_1e_1 + a_2e_2 + a_3e_3 + \cdots = \sum_{n=1}^{\infty} a_ne_n,$$

where the coefficient of e_n is the component

$$a_n = \langle a, e_n \rangle$$

of a on e_n, as in the finite-dimensional case.

But another problem arises: the axioms of a vector space only allow a *finite* number of vectors to added together, and hence only allow *finite linear combinations* of vectors to be formed. Therefore, if we want to justify the above very natural expression for a, we will have to find a way of giving meaning (at least in some circumstances) to 'infinite linear combinations' of vectors. This will be an important topic for detailed discussion in Chapter 4.

➤ The specific difficulty here is almost exactly the same as in the case of infinite series of real numbers. The ordinary laws of arithmetic only allow us to add together finitely many real numbers, and when we wish to add infinitely many, as in an infinite series, we have to develop appropriate definitions and results to justify the process. Specifically, we define partial sums and then take limits, and this will be exactly the route we follow in the present case as well when we return to the issue in detail in Chapter 4.

A related issue is raised by the phrase 'orthonormal basis', which we have used a number of times. For an integer n, it is correct to say (as we have) that the collection $e_1, e_2, \dots, e_n$ is an orthonormal basis for $\mathbb{R}^n$, simply because the collection is both orthonormal and a basis. However, the collection $e_1, e_2, e_3, \dots$ is *not* a basis for ℓ^2, and we consequently cannot correctly refer to this collection as an orthonormal basis for ℓ^2. (See Exercise 1.6 and Subsection 2.3.2 for the claim that $e_1, e_2, e_3, \dots$ do not form a basis for ℓ^2.)

This is the case, at any rate, as long as we continue to use the term 'basis' in the ordinary sense of linear algebra, which only allows finite linear combinations. Once we have found a satisfactory definition of an infinite linear combination, however, we will be able to expand the scope of application of the term 'basis' in a way that will make it correct after all to call the collection $e_1, e_2, e_3, \dots$ an orthonormal basis for ℓ^2.

Although the use of the phrase 'orthonormal basis' in this extended sense will not be formally justified until Chapter 4, we will nevertheless make informal use of it a few times in the remainder of this chapter.

1.2.2 Spaces of Functions

We now take a further step away from the familiar Euclidean spaces $\mathbb{R}^n$ and consider vector spaces of functions defined on an interval I of the real line. For most of the discussion, the interval I will either be $[-\pi, \pi]$ or the whole real line $\mathbb{R}$, but there is no need initially to restrict our choice. Our first attempt to specify a useful space of functions on I might be to consider $F(I)$, the collection of all functions $f\colon I \to \mathbb{R}$. This is indeed a vector space, if we define operations pointwise. Thus for $f, g \in F(I)$, we define $f + g$ by setting $(f + g)(x) = f(x) + g(x)$ for all $x \in I$, and it is clear that $f + g$ is again a function from I to $\mathbb{R}$, giving closure of $F(I)$ under addition. Scalar multiplication is handled similarly.

Further, we can easily write down what might appear to be a reasonable definition of an inner product on $F(I)$ by working in analogy to the inner product definitions that we introduced earlier in $\mathbb{R}^n$ and ℓ^2. In $\mathbb{R}^n$, for example, the definition

$$\langle a, b\rangle = \sum_{i=1}^{n} a_i b_i$$

multiplies corresponding entries of the n-tuples a and b and sums the resulting terms and so, given two functions f and g in $F(I)$, we might define

$$\langle f, g\rangle = \int_I fg,$$

since this definition multiplies corresponding values of the functions f and g and 'sums' the resulting terms, provided that we are prepared to think of integration as generalised summation.

➤ Note the two notational simplifications we have used here for integration. First, we can if we wish represent the region of integration in a definite integral as a subscript to the integral sign; thus

$$\int_{[-\pi,\pi]} \ldots \quad \text{and} \quad \int_{-\pi}^{\pi} \ldots$$

mean the same thing, as do

$$\int_{\mathbb{R}} \ldots \quad \text{and} \quad \int_{-\infty}^{\infty} \ldots\,.$$