

Chapter 2

Signal processing meets vector space

In the previous chapter we already started to classify signals according to their mathematical properties by discriminating, for instance, between finite- and infinite-length entities. As we progress in the course, we will encounter several more types of signal, both in discrete and in continuous time but, rather than trying to create a complex taxonomy, in this chapter we will show how the concept of *vector space* provides an ideal framework to describe and analyze signals irrespective of the details inherent to their type or category.

2.1 Euclidean Geometry and Vector Space

Euclidean geometry is a straightforward idealization of the physical space that we experience with our senses and therefore the concepts of length, angle, translation or scaling require no formal explanation. In this section we will show how Euclidean space can be mapped to a vector space of real-valued tuples and, consequently, how the properties of these vectors can be described in terms of standard geometrical constructs. The idea is to associate the key ingredients of vector space to intuitive geometrical concepts before generalizing them to the more abstract vector spaces that we will use later.

2.1.1 The Euclidean Plane

The vectors. Consider a vector \mathbf{x} defined as a pair of real numbers:

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix}, \quad x_0, x_1 \in \mathbb{R}; \quad (2.1)$$

we call the set of all such vectors \mathbb{R}^2 .

Consider now the Euclidean plane, equipped with a Cartesian reference system as in Figure. Each point on the plane is uniquely identified by a pair of real-valued coordinates on the two reference axes, and therefore we can associate any point on the plane to a

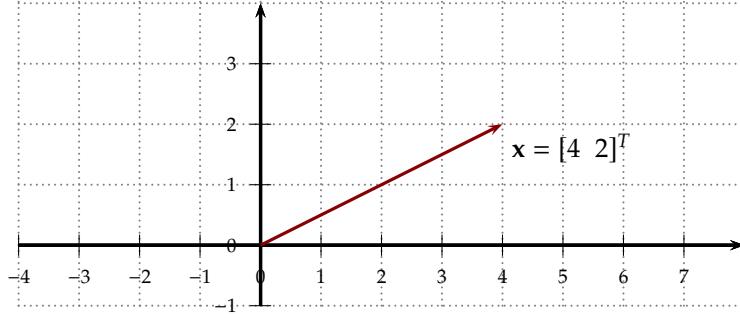


Figure 2.1: Graphical representation of a vector in \mathbb{R}^2 .

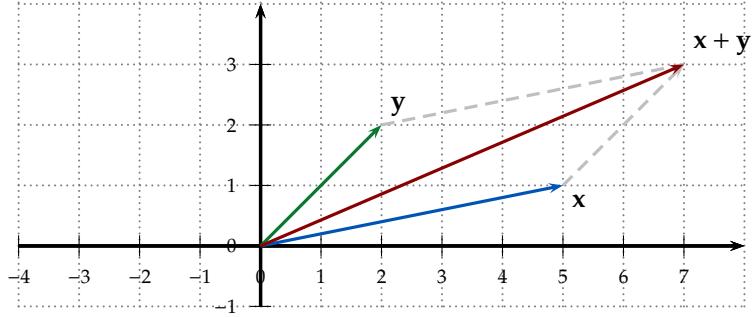


Figure 2.2: Vector addition in \mathbb{R}^2

vector in \mathbb{R}^2 . The standard graphical representation of such a vector is an arrow that from the origin (i.e. the point with coordinates¹ $[0 \ 0]^T$) reaches the point with coordinates $[x_0 \ x_1]^T$, as in Figure 2.1.

Define now the *sum* of two vectors in \mathbb{R}^2 like so:

$$\mathbf{x} + \mathbf{y} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} + \begin{bmatrix} y_0 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 + y_0 \\ x_1 + y_1 \end{bmatrix}; \quad (2.2)$$

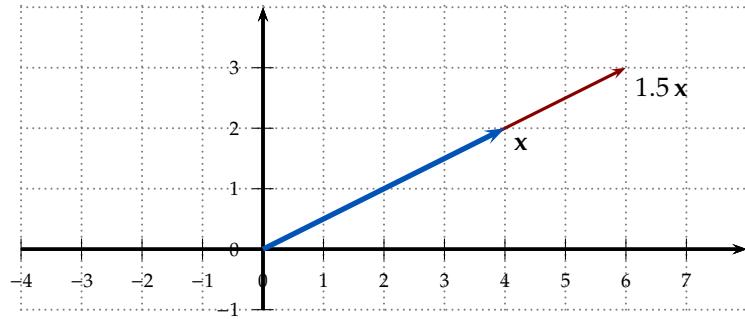
although we could have invented a different “recipe” for addition, this particular one makes a lot of sense since it corresponds to a *translation* of \mathbf{y} by \mathbf{x} : geometrically, we place \mathbf{y} at the end of \mathbf{x} and we connect the origin to the endpoint of the joined vectors as in Figure 2.2. This is the classic parallelogram rule for vector addition.

Similarly, define the scalar multiplication by a scalar $\alpha \in \mathbb{R}$ as

$$\alpha \mathbf{x} = \begin{bmatrix} \alpha x_0 \\ \alpha x_1 \end{bmatrix}; \quad (2.3)$$

again, this makes geometrical sense since it corresponds to changing the “size” of the vector, as in Figure 2.3. Given these definitions for sum and scalar multiplication, it is easy to verify that they possess all the nice properties we expect them to have, including

¹The superscript T indicates transposition so that $[0 \ 0]^T$ is a column vector.

Figure 2.3: Scalar multiplication in \mathbb{R}^2

distributivity, associativity and so on. We will formalize all this later. Multiple vectors can be joined together in *linear combinations* of the form:

$$\mathbf{y} = \sum_{n=0}^N \alpha_n \mathbf{x}_n \quad (2.4)$$

where the subscript, when applied to a boldface symbol, indicates the ordinal number of an item in a given set of vectors.

Norm, similarity and distance. We said that scalar multiplication changes the “size” of a vector but we haven’t defined a way to quantify said size. Once again, by looking at the geometric interpretation of \mathbb{R}^2 , the natural choice is using the length of the arrow that represents the vector; that is commonly called the *norm* and is represented by the symbol $\|\cdot\|$. From Figure 2.1, using Pythagoras Theorem, it’s easy to see that

$$\|\mathbf{x}\| = \sqrt{x_0^2 + x_1^2} \quad (2.5)$$

Consider now the problem of estimating the degree of *similarity* between two vectors; if we start with vectors of equal norm, the only thing left to compare is their directions or, in other words, the angle that they form. If the angle is small, the vectors will be basically oriented equally and almost interchangeable; at the other extreme, if they form a right angle, they will be maximally dissimilar in the sense that the second vector shares no directionality component with the first. The *inner product* of two vectors in \mathbb{R}^2 is a function from $\mathbb{R}^2 \times \mathbb{R}^2$ to \mathbb{R} and it is defined as:

$$\langle \mathbf{x}, \mathbf{y} \rangle = x_0 y_0 + x_1 y_1. \quad (2.6)$$

Although the definition is not instantly intuitive, the following properties will help: first of all, again by invoking Pythagoras, we can see that

$$\langle \mathbf{x}, \mathbf{x} \rangle = \|\mathbf{x}\|^2 \quad (2.7)$$

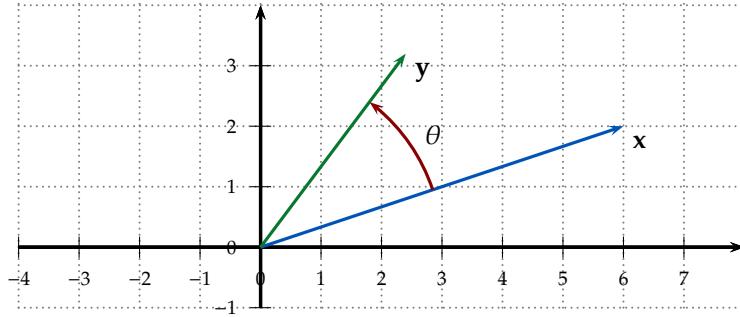


Figure 2.4: Angle between vectors in \mathbb{R}^2 .

namely, a vector's self inner product is equal to the square of the norm as defined in (2.5). Secondly, using basic trigonometry, we can rewrite the inner product as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \|\mathbf{x}\| \|\mathbf{y}\| \cos \theta \quad (2.8)$$

where θ is the angle between \mathbf{x} and \mathbf{y} as in Figure 2.4. The inner product as defined above, therefore, is a qualitative measure of similarity between vectors. Vectors forming a right angle, that is, vectors that are maximally dissimilar, will have an inner product of zero. Such vectors are called *orthogonal*.

Finally, for a quantitative measure of how far apart two vectors are, we can use the norm of their difference to compute the *distance* between vectors as

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|; \quad (2.9)$$

by using the definition of the norm in terms of the inner product, we have

$$\begin{aligned} \|\mathbf{x} - \mathbf{y}\|^2 &= \langle \mathbf{x} - \mathbf{y}, \mathbf{x} - \mathbf{y} \rangle \\ &= \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle - 2\langle \mathbf{x}, \mathbf{y} \rangle \\ &= \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 - 2\langle \mathbf{x}, \mathbf{y} \rangle \end{aligned}$$

which, when \mathbf{x} and \mathbf{y} are orthogonal, simplifies to Pythagoras theorem:

$$\|\mathbf{x} - \mathbf{y}\|^2 = \|\mathbf{x}\|^2 + \|\mathbf{y}\|^2 \quad \text{for } \mathbf{x} \perp \mathbf{y} \quad (2.10)$$

Bases and projections. Consider the two vectors

$$\delta_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \delta_1 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}; \quad (2.11)$$

obviously, for *any* vector in \mathbb{R}^2 we can write:

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} = x_0 \delta_0 + x_1 \delta_1. \quad (2.12)$$

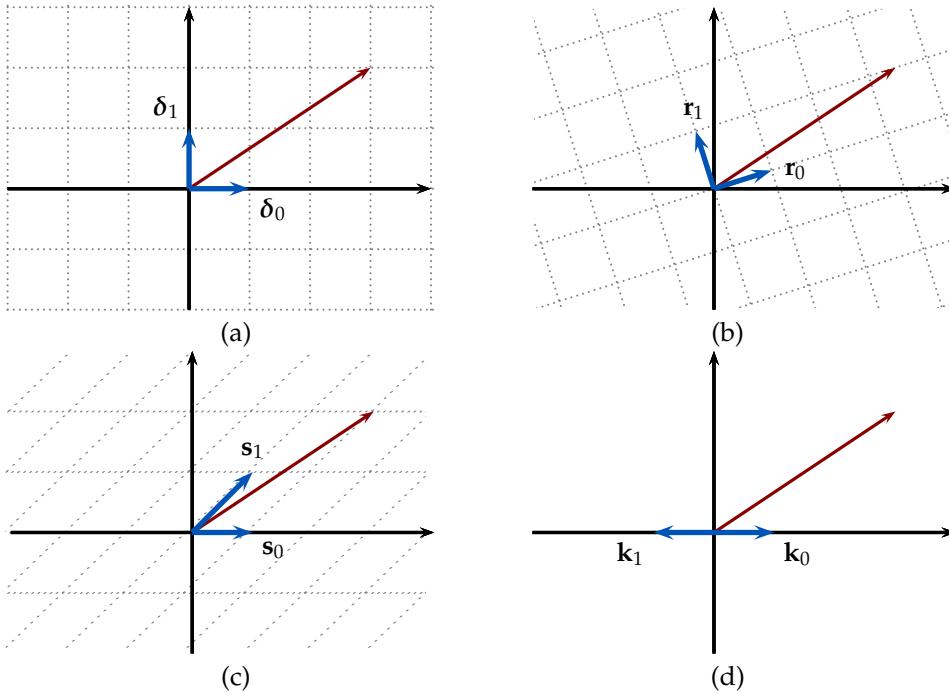


Figure 2.5: (a) canonical basis in \mathbb{R}^2 ; (b) rotated orthonormal basis; (c) non-orthogonal basis; (d) not a basis.

The fact that any vector in \mathbb{R}^2 can be expressed as a linear combination of δ_0 and δ_1 makes the set $\{\delta_0, \delta_1\}$ a *basis* for \mathbb{R}^2 . This basis is also particularly good since its elements have unit norm and are orthogonal, that is, they form an *orthonormal* basis; if we look at the way we have been drawing vectors on the plane, as exemplified in Figure 2.5-(a), we can see that the Cartesian reference grid uses δ_0 and δ_1 as the units on the horizontal and vertical axis, respectively. For this particular basis, the coefficients in the expansion (2.12) coincide with the elements of the vector; this unique fact gives $\{\delta_0, \delta_1\}$ the prestigious title of *canonical basis*.

There is however an infinity of other bases that can be used for \mathbb{R}^2 . For instance, any rotation by an angle θ of the canonical basis is also an orthonormal basis whose elements are

$$\mathbf{r}_0 = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix}, \quad \mathbf{r}_1 = \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} \quad (2.13)$$

and the grid they induce on \mathbb{R}^2 is shown in Figure 2.5-(b). Note that a basis need not be orthonormal; for instance the vectors

$$\mathbf{s}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{s}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad (2.14)$$

from a valid basis for \mathbb{R}^2 and they induce the non-rectangular grid shown in Figure 2.5-(c).

On the other hand, the vectors

$$\mathbf{k}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \quad \mathbf{k}_1 = \begin{bmatrix} -1 \\ 0 \end{bmatrix} \quad (2.15)$$

(Figure 2.5-(d)) do not form a basis for \mathbb{R}^2 since any linear combination will necessarily yield a point on the horizontal axis (the second element of the resulting vector will always be zero). Stated differently, the two vectors are not *linearly independent*.

Consider a vector \mathbf{x} in the canonical basis and pick new basis $\{\mathbf{w}_0, \mathbf{w}_1\}$; how can we find the coefficients that express \mathbf{x} in the new basis? In the Euclidean plane (and in Euclidean space in general), we can always expand the equation

$$\alpha_0 \mathbf{w}_0 + \alpha_1 \mathbf{w}_1 = \mathbf{x} \quad (2.16)$$

by writing out the vector elements explicitly and solving for α_1 and α_2 :

$$\begin{bmatrix} w_{0,0} & w_{0,1} \\ w_{1,0} & w_{1,1} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} x_0 \\ x_1 \end{bmatrix} \quad (2.17)$$

(we use the notation $w_{n,m}$ to indicate the m -th element of vector \mathbf{w}_n). For instance, looking at Figure 2.5-(c) and given $\mathbf{x} = [3 \ 2]^T$ we have

$$\begin{bmatrix} s_{0,0} & s_{0,1} \\ s_{1,0} & s_{1,1} \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \alpha_0 \\ \alpha_1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad (2.18)$$

so that

$$\mathbf{x} = 1 \mathbf{s}_0 + 2 \mathbf{s}_1. \quad (2.19)$$

If the new basis is orthonormal, however, we know that

$$\langle \mathbf{w}_h, \mathbf{w}_k \rangle = \delta[h - k]; \quad (2.20)$$

therefore, if we take the left² inner product of both sides in (2.16) with \mathbf{w}_0 and \mathbf{w}_1 in turn, we can obtain directly:

$$\alpha_0 = \langle \mathbf{w}_0, \mathbf{x} \rangle \quad (2.21)$$

$$\alpha_1 = \langle \mathbf{w}_1, \mathbf{x} \rangle \quad (2.22)$$

For instance, looking at Figure 2.5-(b) where the rotation angle is $\theta = \arctan(2/3)$, we have

$$\alpha_0 = \langle \mathbf{r}_0, \mathbf{x} \rangle = 3 \cos \theta + 2 \sin \theta \approx 3.46 \quad (2.23)$$

$$\alpha_1 = \langle \mathbf{r}_1, \mathbf{x} \rangle = -3 \sin \theta + 2 \cos \theta = 1 \quad (2.24)$$

²We use the left inner product because it mirrors the order of the operands when the change of basis is expressed as a matrix-vector multiplications as in (2.17). The order is of course irrelevant when we operate over the field of reals, since in that case the inner product is commutative; over the field of complex numbers, however, the inner product is commutative with conjugation (see definition (2.47) later) and so the order is crucial.

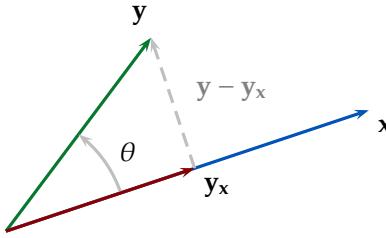


Figure 2.6: Orthogonal projection in \mathbb{R}^2 .

Bases play a fundamental role in signal processing since, as we will see in the next Chapter, the Fourier Transform can be interpreted as a change of basis in an appropriate vector space. Intuitively, a change of basis is a change of perspective: we're looking at the same data but from a different reference system and this change of viewpoint can highlight properties of a signal that are difficult to see in the original dataset.

Let's now look once again at the definition of inner product in 2.6. Consider Figure 2.6 and assume $\|x\| = 1$; the vector

$$y_x = \langle x, y \rangle x = (\|y\| \cos \theta) x \quad (2.25)$$

is called the *orthogonal projection* of y over x and it represents the best way to approximate y using only a scaled version of x (that is, using a vector that points in the direction of x). The approximation is optimal in the sense that it achieves the minimum distance (also known as the minimum *Mean Square Error*) between the original vector and any approximation based only on x :

$$\arg \min_{\alpha} \|y - \alpha x\| = \langle x, y \rangle; \quad (2.26)$$

this is illustrated graphically in Figure 2.7 and can be easily derived analytically by solving $\partial \|y - \alpha x\|^2 / \partial \alpha = 0$. Additionally, the approximation error $y - y_x$ is orthogonal to the approximation itself

$$\langle y - y_x, y_x \rangle = 0 \quad (2.27)$$

meaning that all the information of y that was expressible by x has been captured by the approximation. The expression in (2.25) assumed that x has unit norm; in the general case we need to use the normalized vector $x/\|x\|$ to obtain the general orthogonal projection formula

$$y_x = \frac{\langle x, y \rangle}{\langle x, x \rangle} x. \quad (2.28)$$

Although these results are nearly self-evident in \mathbb{R}^2 , they will be particularly interesting when generalized to spaces of higher dimensionality.

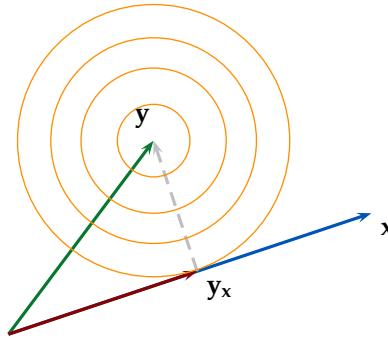


Figure 2.7: The orthogonal projection minimizes the distance between original and approximation.

2.1.2 Euclidean Space

One dimension up from the plane is the three-dimensional Euclidean space, which can be described in terms of vectors in \mathbb{R}^3 of the form

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ x_2 \end{bmatrix}, \quad x_0, x_1, x_2 \in \mathbb{R}. \quad (2.29)$$

Addition, scalar multiplication and inner product all retain their definitions (by adjusting for one more component) and their geometrical meaning. Euclidean space, however, allows us to illustrate the concept of subspace approximation better than the plane.

Subspaces and Least Squares approximation. A vector subspace is a subset of the original vector space closed under addition and scalar multiplication; consider for example a plane in 3D space: if we sum (or rescale) vectors laying on the plane we remain on the plane and therefore we have a proper subspace. Subspaces have their own bases, which can be orthonormal of course. For instance, Figure 2.8 illustrates the planar subspace spanned by the first two 3D canonical vectors

$$\delta_0 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} \quad \text{and} \quad \delta_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}.$$

The existence of orthonormal bases for subspaces provides a very elegant answer to the problem of approximating an arbitrary vector with elements of a subspace. While the problem may appear quite abstract at first, it is actually at the core of many signal processing applications: modern image or video compression algorithms, for example, proceed by approximating a signal living in a high-dimensional space with a set of basis vectors from a subspace of much lower dimensionality.

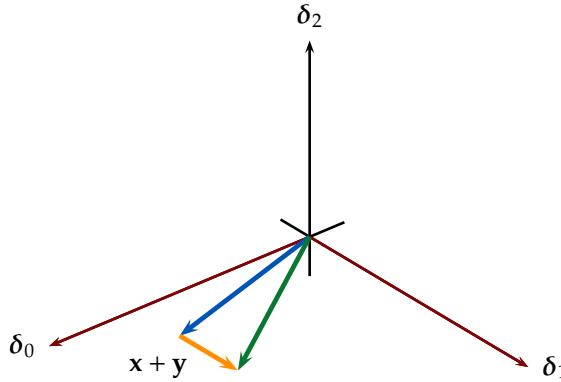


Figure 2.8: Three-dimensional Euclidean space and the planar subspace spanned by the first two canonical basis vectors; the subspace is closed under addition and scalar multiplication.

Consider a vector space V and a subspace $W \subseteq V$; assume $\{\mathbf{w}_k\}_{k=0,1,\dots,K-1}$ is an orthonormal basis for W . The *projection theorem* states that the vector

$$\hat{\mathbf{x}} = \sum_{k=0}^{K-1} \langle \mathbf{w}_k, \mathbf{x} \rangle \mathbf{w}_k \quad (2.30)$$

is the “best” approximation of $\mathbf{x} \in V$ over W . To see why, let’s first unpack the above Equation; the sum is a linear combination of orthogonal projections of \mathbf{x} onto all the basis vectors for the subspace. We already know from Equation (2.27) that each orthogonal projection is the best approximation of \mathbf{x} using a single unit norm basis vector. To show that (2.30) is the best global approximation in W we need to show that the error $\mathbf{x} - \hat{\mathbf{x}}$ is orthogonal to the approximation $\hat{\mathbf{x}}$:

$$\langle \mathbf{x} - \hat{\mathbf{x}}, \hat{\mathbf{x}} \rangle = 0; \quad (2.31)$$

this means that no further portion of the error could be accounted for by elements of the subspace and we’ve done the best we could. To prove (2.31) just substitute for (2.30), use the distributivity of the inner product and remember that the basis $\{\mathbf{w}_k\}$ is orthonormal. If the error is orthogonal to the subspace, the approximation minimizes the (square) norm of the error:

$$\arg \min_{\mathbf{y} \in W} \|\mathbf{x} - \mathbf{y}\| = \hat{\mathbf{x}}; \quad (2.32)$$

the approximation is therefore optimal in the least squares sense. An example of subspace approximation in Euclidean Space is given in Figure 2.9.

Note that the approximation provided by a set of orthonormal vectors can be *successively refined*. Assume you have an approximation using the first m basis vectors:

$$\hat{\mathbf{x}}_m = \sum_{k=0}^{m-1} \langle \mathbf{w}_k, \mathbf{x} \rangle \mathbf{w}_k \quad (2.33)$$

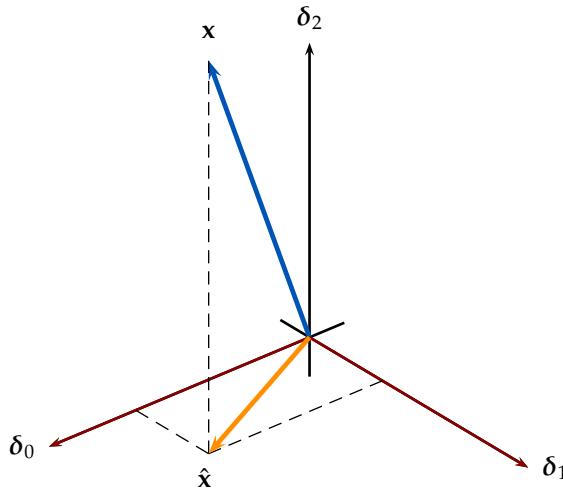


Figure 2.9: Least-squares approximation of a 3D vector onto the subspace spanned by the first two canonical basis vectors.

the optimal approximation using another vector is simply

$$\hat{\mathbf{x}}_{m+1} = \hat{\mathbf{x}}_m + \langle \mathbf{w}_m, \mathbf{x} \rangle \mathbf{w}_m \quad (2.34)$$

While this may seem obvious, it is actually a small miracle, since it does not hold for more general, non-orthonormal bases; a classic application of this property is used in algorithms for variable-bit-rate transmission of audio and video.

2.1.3 Higher Dimensions

Formally, it's quite easy to increase the number of dimensions beyond three: just consider the set of all tuples of N real numbers

$$\mathbf{x} = \begin{bmatrix} x_0 \\ x_1 \\ \vdots \\ x_{N-1} \end{bmatrix}. \quad (2.35)$$

By extending the Euclidean definitions of sum, scalar multiplication and inner product in a straightforward manner we obtain a vector space called \mathbb{R}^N . While formally equivalent to Euclidean space, \mathbb{R}^N does not correspond to any physical entity that we can visualize; yet, the intuition gained in 2 and 3 dimensions can help us understand the properties of higher-dimensional spaces that defy graphical representation.

Similarly, we can define vectors as tuples of complex numbers; the resulting vector space, \mathbb{C}^N , cannot be represented graphically for N greater than one, but properties like norm,

orthogonality and distance remain intuitive by analogy to Euclidean space. In general, no matter how abstract the vector space, we can always try to resort to the geometrical interpretation of \mathbb{R}^2 or \mathbb{R}^3 to get an intuitive grasp on the space's properties.

2.2 Vector Space: a Formal Definition

2.2.1 The Axioms

Following are the axiomatic definitions for a vector space equipped with an inner product. The stuff is dry, no question, but it simply encodes at an abstract level the very reasonable properties we have seen for Euclidean space³.

Vector Space. A vector space $H(V, S)$ is defined in terms of V , a set of objects called *vectors*, and S , a set of numbers called *scalars*; note that we will always assume $S = \mathbb{C}$. For the vector space to have any use, we need to be able to resize and combine vectors together and this is achieved via two operations called, respectively, *addition* and *scalar multiplication*. For $H(V, S)$ to be a vector space, these two operations, while arbitrarily definable, must fulfill the following set of axioms:

For any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and any $\alpha, \beta \in S$:

- operations don't leave the vector space:

$$\mathbf{x} + \mathbf{y} \in V \tag{2.36}$$

$$\alpha\mathbf{x} \in V \tag{2.37}$$

- addition is commutative:

$$\mathbf{x} + \mathbf{y} = \mathbf{y} + \mathbf{x} \tag{2.38}$$

- addition is associative:

$$(\mathbf{x} + \mathbf{y}) + \mathbf{z} = \mathbf{x} + (\mathbf{y} + \mathbf{z}) \tag{2.39}$$

- scalar multiplication is distributive:

$$\alpha(\mathbf{x} + \mathbf{y}) = \alpha\mathbf{x} + \alpha\mathbf{y} \tag{2.40}$$

$$(\alpha + \beta)\mathbf{x} = \alpha\mathbf{x} + \beta\mathbf{x} \tag{2.41}$$

$$\alpha(\beta\mathbf{x}) = (\alpha\beta)\mathbf{x} \tag{2.42}$$

³Although the analogy is a bit tenuous, readers with more than a passing acquaintance with object-oriented programming could be excused for drawing a parallel between the concept of vector space and that of an interface class: internal variables and implementations in derived classes can be arbitrary as long as they conform with the interface.

- there exists a null vector $\mathbf{0} \in V$ that is the additive identity:

$$\mathbf{x} + \mathbf{0} = \mathbf{0} + \mathbf{x} = \mathbf{x} \quad (2.43)$$

- for every $\mathbf{x} \in V$ there exists an additive inverse $-\mathbf{x} \in V$:

$$\mathbf{x} + (-\mathbf{x}) = (-\mathbf{x}) + \mathbf{x} = \mathbf{0} \quad (2.44)$$

- there exists a scalar $1 \in S$ that is the scalar multiplication identity:

$$1 \cdot \mathbf{x} = \mathbf{x} \cdot 1 = \mathbf{x} \quad (2.45)$$

Note that no restrictions are set on what's "inside" a vector or on the internal mechanics of the vector operations; this leads to the great power of generalization of the vector space paradigm.

Inner Product Space. In order to compare vectors in a space, we need some function that takes two vectors and returns some measure of their similarity. The usual choice is to equip the space with an *inner product* (or *dot product*), namely, a function defined on $V \times V$ that returns a scalar; remember that our set of scalars is \mathbb{C} . Again, the inner product can be defined arbitrarily, as long as it satisfies the following properties:

For any $\mathbf{x}, \mathbf{y}, \mathbf{z} \in V$ and $\alpha \in \mathbb{C}$:

- the inner product is distributive:

$$\langle \mathbf{x} + \mathbf{y}, \mathbf{z} \rangle = \langle \mathbf{x}, \mathbf{z} \rangle + \langle \mathbf{y}, \mathbf{z} \rangle \quad (2.46)$$

- the inner product is commutative, with complex conjugation:

$$\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle^* \quad (2.47)$$

- the inner product scales with respect to scalar multiplication⁴:

$$\langle \alpha \mathbf{x}, \mathbf{y} \rangle = \alpha^* \langle \mathbf{x}, \mathbf{y} \rangle \quad (2.48)$$

$$\langle \mathbf{x}, \alpha \mathbf{y} \rangle = \alpha \langle \mathbf{x}, \mathbf{y} \rangle \quad (2.49)$$

- the self-product is real-valued and nonnegative:

$$\langle \mathbf{x}, \mathbf{x} \rangle \geq 0 \quad (2.50)$$

- the self-product is zero only for the null vector:

$$\langle \mathbf{x}, \mathbf{x} \rangle = 0 \Leftrightarrow \mathbf{x} = \mathbf{0}. \quad (2.51)$$

⁴See footnote on page 6 as for the reason why we conjugate the scalar associated to the left operand.

Hilbert Space. The axioms for addition and scalar multiplication require that the result of these operations remain in the original vector space; as a consequence, all linear combinations of a *finite* number of vectors from V will live in V as well. However, that is not sufficient to guarantee that the limit of a convergent sequence of vectors in V is also in V . Closure of a vector space under convergent limiting operations is an independent property called *completeness* and a inner-product space that is complete is called a *Hilbert space*. Completeness is a rather technical property and one that is often difficult to prove rigorously⁵. While there are no elementary examples of a non-complete vector space, to gain some intuition think of the set of rational numbers \mathbb{Q} ; this set is not complete because there are many sequences in \mathbb{Q} that converge to irrational numbers. Consider for instance, the sequence

$$s_k = \sum_{n=1}^k \frac{1}{n^2};$$

while it's easy to see that $s_k \in \mathbb{Q}$ for all values of k , it can be shown that

$$\lim_{k \rightarrow \infty} s_k = \frac{\pi^2}{6} \notin \mathbb{Q}.$$

The set of rational numbers \mathbb{R} , on the other hand, is complete.

2.2.2 Norms and Distances

The properties of the inner-product induce a topology on the vector space, namely, once we have an inner product, we can start to compare vectors and define qualitative and quantitative measures of similarity. The most important concept is orthogonality

$$\langle \mathbf{x}, \mathbf{y} \rangle = 0,$$

which encodes the notion of maximal difference between vectors. The *norm* of a vector is defined as the square root of the self inner product

$$\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} \tag{2.52}$$

and it describes quantitatively the size of a vector. The norm also extends the concept of Euclidean distance to arbitrary vector spaces as

$$d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|; \tag{2.53}$$

this distance measure is also known as the Mean Square Error between vectors.

Triangle Inequality To quickly show the power of the abstract definitions above, let's formally prove the *triangle inequality* in an generic inner product space $H(V, \mathbb{C})$. The statement is simple: given any two vectors $\mathbf{x}, \mathbf{y} \in V$, it is always

$$\|\mathbf{x} + \mathbf{y}\| \leq \|\mathbf{x}\| + \|\mathbf{y}\| \tag{2.54}$$

⁵In this course, we will really need completeness only once when we prove the sampling theorem, and we will tacitly assume the property without proof.

where the norm is the natural norm induced by the inner product as in (2.52). On the Euclidean plane, the above relation is obvious since it states that, for any triangle, the sum of the length of two sides can't be smaller than the length of the third side. The result is however valid for all inner-product spaces and, to prove the inequality in the general case, we first need an intermediate result, called the **Cauchy-Schwarz inequality**:

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|. \quad (2.55)$$

To prove this, assume $\mathbf{x}, \mathbf{y} \neq \mathbf{0}$ (otherwise the result is obvious) and consider the orthogonal projection (2.28) of \mathbf{y} onto \mathbf{x}

$$\mathbf{y}_x = \frac{\langle \mathbf{x}, \mathbf{y} \rangle}{\langle \mathbf{x}, \mathbf{x} \rangle} \mathbf{x};$$

call $\mathbf{d} = \mathbf{y} - \mathbf{y}_x$ the approximation error and remember that the error is orthogonal to \mathbf{y}_x ; you can look at a graphical representation of the vectors involved in the proof in Figure 2.10. Since $\mathbf{y} = \mathbf{y}_x + \mathbf{d}$ we can write

$$\|\mathbf{y}\|^2 = \langle \mathbf{y}_x + \mathbf{d}, \mathbf{y}_x + \mathbf{d} \rangle \quad (2.56)$$

$$= \|\mathbf{y}_x\|^2 + \|\mathbf{d}\|^2 + \langle \mathbf{y}_x, \mathbf{d} \rangle + \langle \mathbf{d}, \mathbf{y}_x \rangle \quad (2.57)$$

$$= \|\mathbf{y}_x\|^2 + \|\mathbf{d}\|^2 \quad (2.58)$$

$$\geq \|\mathbf{y}_x\|^2 \quad (2.59)$$

$$= \frac{|\langle \mathbf{x}, \mathbf{y} \rangle|^2}{\|\mathbf{x}\|^2} \quad (2.60)$$

where the cross inner products have vanished because of the orthogonality of the error. Getting back to the triangle inequality, we have

$$\|\mathbf{x} + \mathbf{y}\|^2 = \langle \mathbf{x} + \mathbf{y}, \mathbf{x} + \mathbf{y} \rangle \quad (2.61)$$

$$= \langle \mathbf{x}, \mathbf{x} \rangle + \langle \mathbf{x}, \mathbf{y} \rangle + \langle \mathbf{y}, \mathbf{x} \rangle + \langle \mathbf{y}, \mathbf{y} \rangle \quad (2.62)$$

$$= \|\mathbf{x}\|^2 + 2 \operatorname{Re}\{\langle \mathbf{x}, \mathbf{y} \rangle\} + \|\mathbf{y}\|^2 \quad (2.63)$$

$$\leq \|\mathbf{x}\|^2 + 2|\langle \mathbf{x}, \mathbf{y} \rangle| + \|\mathbf{y}\|^2 \quad (2.64)$$

$$\leq \|\mathbf{x}\|^2 + 2\|\mathbf{x}\|\|\mathbf{y}\| + \|\mathbf{y}\|^2 \quad (2.65)$$

$$= (\|\mathbf{x}\| + \|\mathbf{y}\|)^2 \quad (2.66)$$

where we have used (2.47), the fact that $|z| \geq \operatorname{Re}\{z\}$ for all $z \in \mathbb{C}$ and, of course, the Cauchy-Schwarz inequality. Please note that nowhere in the proof we made use of the internal structure of the vectors; we only used the axioms.

Other Norms We should mention in passing that, although the norm induced by the inner product is the natural norm for a Hilbert space, we are free to define alternative quantitative measures of a vector's size. In Euclidean space, for instance, the so-called

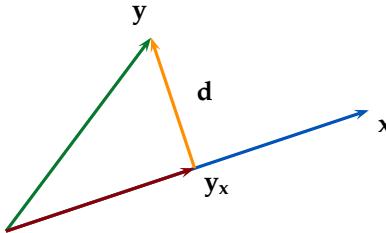


Figure 2.10: Graphical representation of the vectors used in the proof of the Cauchy-Schwarz inequality.

p -norm is defined as

$$\|\mathbf{x}\|_p = \left(\sum_n |x_n|^p \right)^{1/p}. \quad (2.67)$$

Notable values include

- the standard norm for $p = 2$
- the so-called “Manhattan” norm for $p = 1$
- the supremum norm for $p = \infty$:

$$\|\mathbf{x}\|_\infty = \max_n |x_n| \quad (2.68)$$

The latter, in particular, will reappear in a major role when we study FIR filter design. Norms not derived from the inner product, in order to be valid norms, must be *designed* to satisfy the triangle inequality.

2.2.3 Bases and Subspaces

Given a vector space $H(V, \mathbb{C})$, a subset of N vectors $S = \{\mathbf{s}_k\}_k$, is said to be linearly independent if the equation

$$\sum_{k=0}^{N-1} \beta_k \mathbf{s}_k = \mathbf{0}$$

only admits the solution $\beta_k = 0$ for $k = 0, 1, \dots, N - 1$. Linear independence states that the subset S contains no redundancies, that is, none of its members can be expressed as a linear combination of the others. Now, suppose we can determine a subset of N vectors $W = \{\mathbf{w}_k\}_k \subseteq H$ so that any element $\mathbf{x} \in H$ can be written as a linear combination of elements of W :

$$\mathbf{x} = \sum_{k=0}^{N-1} \alpha_k \mathbf{w}_k;$$

in this case we say that W is a *spanning set* for H . If the set is also linearly independent (i.e., it contains no redundancy) then we call W a *basis* for H . If W is a basis, the coefficients α_k are unique, as a consequence of linear independence. The laborious process of finding the analysis coefficients α_k for an arbitrary vector \mathbf{x} is greatly simplified if the basis is orthonormal:

$$\langle \mathbf{w}_k, \mathbf{w}_h \rangle = \begin{cases} 1 & \text{for } h = k \\ 0 & \text{for } h \neq k \end{cases};$$

in this case, it is simply

$$\alpha_k = \langle \mathbf{w}_k, \mathbf{x} \rangle.$$

If a basis for H contains N elements, it is relatively straightforward (at least for finite N) to show that any other basis will also contain exactly N vectors; N is therefore called the *dimensionality* of H . As a consequence, for finite-dimensional vector spaces, any set $S = \{\mathbf{s}_k\}_k$ of N linearly-independent vectors will constitute a basis; indeed, if S did not span H then there exists at least one nonzero element $\mathbf{x} \in H$ so that

$$\mathbf{x} \neq \sum_{k=0}^{N-1} \alpha_k \mathbf{s}_k \quad \text{for all } \alpha_k,$$

that is, $S' = \{\mathbf{s}_0, \dots, \mathbf{s}_{N-1}, \mathbf{x}\}$ is a set of $N + 1$ linearly independent vectors. Now, if S' spans H , it will be $\dim(H) = N + 1$, which is a contradiction; otherwise we can repeat the procedure and augment S' until we have a basis, but, a fortiori, the dimensionality of H will be larger than N .

Finally, mutually orthogonal vectors are inherently linearly independent; pick a set of N nonzero orthogonal vectors $\{\mathbf{w}_k\}_k$ and assume we can write

$$\sum_{k=0}^{N-1} c_k \mathbf{w}_k = \mathbf{0}$$

with at least one nonzero coefficient c_i . By the properties of the inner product, since $\langle \mathbf{0}, \mathbf{x} \rangle = 0$, we have

$$\langle \mathbf{0}, \mathbf{w}_i \rangle = \langle \sum_{k=0}^{N-1} c_k \mathbf{w}_k, \mathbf{w}_i \rangle = c_i \|\mathbf{w}_i\|^2 = 0.$$

As a consequence, if we can find a set of N orthogonal vectors in a space of dimension N , we automatically have found a basis.

2.3 Vector Spaces in Signal Processing

After an informal introduction using Euclidean space and a precise axiomatic definition, we are now ready to see how vector spaces are going to be used for signal processing in the rest of the course.

2.3.1 Discrete-time Signals

As we have seen in the previous chapter, in signal processing it is common practice to use the expression $x[n]$ to indicate a discrete-time signal in its entirety; this is an abuse of notation, since the term $x[n]$ in fact pinpoints the value of the signal $x[\cdot]$ at a specific index value n . Vector notation, on the other hand, provides us with a much more compact and precise symbolism: we can use the symbol \mathbf{x} to indicate an entire signal and the expression $x[n]$ to indicate its n -th value.

Finite-length signals Length- N signals are simply ordered tuples of N complex numbers and therefore they belong to \mathbb{C}^N :

$$\mathbf{x} = \begin{bmatrix} x[0] \\ x[1] \\ \vdots \\ x[N-1] \end{bmatrix} \quad (2.69)$$

Addition and scalar multiplication are defined in the intuitive, straightforward way. The inner product is defined as

$$\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{n=0}^{N-1} x^*[n] y[n]. \quad (2.70)$$

We will soon see that, in signal processing, a *filter* works by implementing an inner product; indeed, by filtering a data set, we are evaluating the local degree of similarity of the input with a prototype “shape”, known as the filter’s impulse response. Similarly, many communication systems work by comparing the incoming waveform to a set of predefined “pulses” that encode different symbols; the measure of similarity that is used for the selection of the closest pulse is also an inner product.

The canonical basis for \mathbb{C}^N is also intuitive:

$$\delta_0 = \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad \delta_1 = \begin{bmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{bmatrix}, \quad \dots \quad \delta_{N-1} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{bmatrix}. \quad (2.71)$$

Another very important basis for \mathbb{C}^N is the Fourier basis $\{\mathbf{w}_k\}$ with

$$w_k[n] = e^{j\frac{2\pi}{N}nk}; \quad (2.72)$$

we will study the Fourier basis in detail in the next chapter.

Given a basis $\{\mathbf{s}_k\}$, we know that we can express any vector in \mathbb{C}^N as a linear combination of basis elements:

$$\mathbf{x} = \sum_{k=0}^{N-1} \alpha_k \mathbf{s}_k; \quad (2.73)$$

if the basis is orthonormal, the coefficients α_k can be easily computed using the *analysis formula*

$$\alpha_k = \langle \mathbf{s}_k, \mathbf{x} \rangle. \quad (2.74)$$

In \mathbb{C}^N a particularly nice thing happens: the set of N coefficients α_k is itself a member of \mathbb{C}^N and this opens up the full linear algebra toolbox for us to use. For example, remembering the definition of inner product in (2.70), we can build an $N \times N$ matrix \mathbf{M} by stacking the conjugates of the basis vectors as rows⁶

$$\mathbf{M} = \begin{bmatrix} \mathbf{s}_0^H \\ \hline \mathbf{s}_1^H \\ \hline \vdots \\ \hline \mathbf{s}_{N-1}^H \end{bmatrix} \quad (2.75)$$

and, with this, obtain all the coefficient α_k in one go using the matrix-vector product:

$$\boldsymbol{\alpha} = \mathbf{M}\mathbf{x}. \quad (2.76)$$

Note that \mathbf{M}^H is a matrix whose columns are simply the basis vectors side by side; we can therefore write

$$\mathbf{x} = \mathbf{M}\boldsymbol{\alpha} \quad (2.77)$$

since

$$\mathbf{M}\mathbf{M}^H = \mathbf{M}^H\mathbf{M} = \mathbf{I}. \quad (2.78)$$

The possibility of applying the constructs of linear algebra to signal manipulation further allows us to cast all linear operations on signals as matrix-vector multiplications.

Periodic Signals Periodic signals of period N are completely described by N consecutive samples and therefore they are equivalent to finite-length signals of length N ; they also live in \mathbb{C}^N and, when the periodicity is to be made explicit, we will sometimes use the notation $\tilde{\mathbb{C}}^N$.

Infinite-length signals From a purely formal point of view, we can let N in \mathbb{C}^N grow to infinity and, by keeping the same definitions for sum, scalar multiplication and inner product, we obtain the vector space \mathbb{C}^∞ (sometimes also notated as $\mathbb{C}^\mathbb{Z}$ since, in signal processing, we are usually using both positive and negative indexes). Because the resulting space is now infinite-dimensional, we need to exercise much caution; for example, it

⁶The notation \mathbf{x}^H denotes the *Hermitian transpose* i.e. a vector or a matrix where the original elements have been transposed *and* conjugated.

is immediate to see that the inner product can diverge even for elementary signals such as a constant signal whose samples are all equal to one; indeed

$$\langle \mathbf{x}, \mathbf{x} \rangle = \sum_{n=-\infty}^{\infty} x^*[n]x[n] = \sum_{n=-\infty}^{\infty} 1 = \infty. \quad (2.79)$$

To avoid these type of problems we can consider instead the set of infinite-length vectors whose Euclidean norm is finite, that is, the space of square-summable sequences; the space is called $\ell_2(\mathbb{Z})$, with

$$\mathbf{x} \in \ell_2(\mathbb{Z}) \iff \|\mathbf{x}\|^2 = \sum_{n \in \mathbb{Z}} |x[n]|^2 < \infty, \quad (2.80)$$

and is also known as the space of *finite-energy* signals, given the definition of energy in the previous chapter. In this space the inner product is always well defined, as can be shown by using the Cauchy-Schwarz inequality. Be forewarned, however, that there are a lot of interesting signals that are not in $\ell_2(\mathbb{Z})$ and we will see how to deal with them while trying to preserve the vector space formalism.

In the case of infinite dimensions, the concept of basis also becomes tricky since it is not necessarily safe to allow for linear combinations of vectors with an infinite number of terms in the sum. In $\ell_2(\mathbb{Z})$, however, all is well: the canonical basis for the space is the extension to infinite dimensions of the canonical basis for \mathbb{C}^N , that is, an infinite set of vectors in which only a single element at a time is equal to one.

2.3.2 Function Spaces

Function vector spaces, as the name implies, are spaces in which the vectors are functions; since we know from calculus how to combine, rescale and measure the difference between functions, we can re-cast those operations in such a way that they comply with the axioms of a vector space. Here we are really performing a task of profound abstraction because our vectors are no longer tuples of numbers but sophisticated mathematical entities.

In this course we will only be interested in two specific function spaces, the space of square-integrable functions over an interval and the space of bandlimited functions on the real line; we will wait until later to introduce the latter, but the former is easily described: consider an interval I and the set of complex-valued functions of a real variable $\mathbf{x} = x(t)$ such that

$$\int_I |x(t)|^2 dt < \infty. \quad (2.81)$$

The space of all such functions is denoted by $L_2(I)$; addition and scalar multiplication are defined in the intuitive manner:

$$\mathbf{x} + \mathbf{y} = x(t) + y(t) \quad (2.82)$$

$$\alpha \mathbf{x} = \alpha x(t) \quad (2.83)$$

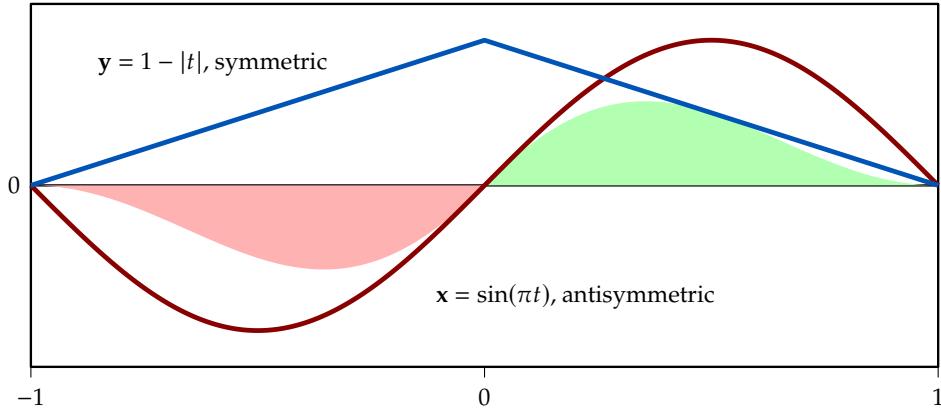


Figure 2.11: Orthogonality of symmetric and antisymmetric functions. The shaded areas show the integration area in the inner product.

while the inner product is

$$\langle \mathbf{x}, \mathbf{y} \rangle = \int_I x^*(t) y(t) dt. \quad (2.84)$$

Note that the norm induced by the inner product is

$$\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle = \int_I |x(t)|^2 dt \quad (2.85)$$

so that all elements of $L_2(I)$ have a well-defined norm.

The concept of orthogonality carries on, with the same meaning, to function spaces. Consider for instance the two vectors

$$\mathbf{x} = \sin(\pi t) \quad (2.86)$$

$$\mathbf{y} = 1 - |t| \quad (2.87)$$

over $I = [-1, 1]$, as shown in Figure 2.11. Their inner product is the integral

$$\int_{-1}^1 (1 - |t|) \sin(\pi t) dt$$

which, graphically, is the sum of the positive green area and the negative red areas in the picture; since this sum is zero, the functions are orthogonal. This makes sense since the first function is odd and the second is even: clearly no symmetric shape can be captured by an antisymmetric curve. Similarly, two sinusoids at different multiples of a fundamental frequency are also orthogonal, as shown in Figure 2.12; the orthogonality of harmonically-related sinusoids is one of the key concepts in signal processing.

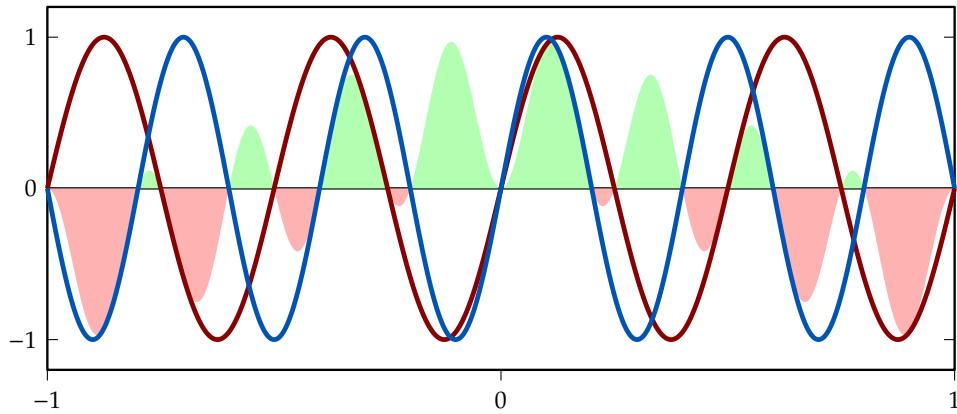


Figure 2.12: Orthogonality of sinusoids $x = \sin(4\pi t)$ and $y = \sin(5\pi t)$ over $[-1, 1]$; the sinusoids' frequencies are multiples of the fundamental frequency for the interval. The sum of the shaded areas yields the value of the integral in the inner product, which is zero.

Function vector spaces possess bases as well, obviously. For instance, an orthonormal basis for $L_2([-\pi, \pi])$ is the well-known Fourier basis

$$\{(1/\sqrt{2\pi}) e^{jnt}\}_{n \in \mathbb{Z}}. \quad (2.88)$$

The number of vectors is understandably infinite but, perhaps surprisingly, they forms a *countable* set. While it's easy to verify that the vectors are orthogonal, proving that they indeed span $L_2([-\pi, \pi])$ is a much harder task — unfortunately, this is very much the rule for all non-trivial, infinite-dimensional basis sets, as we will have to admit again for the space of bandlimited functions later on in the course.

2.4 Appendix: Approximation via Legendre Polynomials

In this section we will illustrate how the concept of subspace projection can lead to optimal polynomial approximation (in the min-square-error sense) for functions over an interval. This example will give us the chance to talk about polynomial vector spaces, orthonormalization procedures and optimal approximation via subspace projections with a concrete case study. While the section is not a direct signal processing application, approximation by projection is routinely used in compression algorithm and the concepts illustrated here will show the flexibility of the vector space approach.

Let's start by considering $P_N(I)$, the vector space of all polynomials of degree up to $N - 1$ over the interval I ; a generic element of $P_N(I)$ is of the form

$$\mathbf{a} = a_0 + a_1 t + \dots + a_{N-1} t^{N-1}, \quad a_k \in \mathbb{R}, t \in I. \quad (2.89)$$

Addition and scalar multiplication are defined according to standard polynomial algebra and, since $P_N(I) \in L_2(I)$, we will use the standard definition for the inner product as

in (2.84). A natural basis for $P_N(I)$ is the family

$$\{1, t, t^2, \dots, t^{N-1}\} \quad (2.90)$$

since every vector is a linear combination of powers of t . The natural basis, however, is not orthonormal so before we can address the issue of subspace approximation we need to rectify that.

Gram-Schmidt Orthonormalization A basis for a vector (sub)space can be transformed into an orthonormal basis via the Gram-Schmidt iteration. Given a basis $\{\mathbf{s}_k\}$, the procedure progressively “whittles” away from each basis vector the portions that are linearly dependent to the previously processed elements. In detail, the orthonormal basis $\{\mathbf{u}_k\}$ is built one vector at a time by first converting \mathbf{s}_m into \mathbf{p}_m , a vector orthogonal to the set of the previously-computed $\{\mathbf{u}_0, \dots, \mathbf{u}_{m-1}\}$, and then by normalizing \mathbf{p}_m and adding it to the pool:

$$\mathbf{p}_m = \mathbf{s}_m - \sum_{k=0}^{m-1} \langle \mathbf{s}_m, \mathbf{u}_k \rangle \mathbf{u}_k \quad (2.91)$$

$$\mathbf{u}_m = \mathbf{p}_m / \|\mathbf{p}_m\|. \quad (2.92)$$

A graphical example of the orthonormalization procedure for two vectors in \mathbb{R}^2 is shown in Figure 2.13.

Legendre Polynomials Let’s apply the Gram-Schmidt algorithm to the naive basis for $P_N([-1, 1])$:

- $m = 0$: $\mathbf{s}_0 = 1$
 - $\mathbf{p}_0 = \mathbf{s}_0 = 1$
 - $\|\mathbf{p}_0\|^2 = 2$
 - $\mathbf{u}_0 = \mathbf{p}_0 / \|\mathbf{p}_0\| = \sqrt{1/2}$
- $m = 1$: $\mathbf{s}_1 = t$
 - $\langle \mathbf{s}_1, \mathbf{u}_0 \rangle = \int_{-1}^1 t / \sqrt{2} = 0$
 - $\mathbf{p}_1 = \mathbf{s}_1 = t$
 - $\|\mathbf{p}_1\|^2 = 2/3$
 - $\mathbf{u}_1 = \sqrt{3/2} t$
- $m = 2$: $\mathbf{s}_2 = t^2$
 - $\langle \mathbf{s}_2, \mathbf{u}_0 \rangle = \int_{-1}^1 t^2 / \sqrt{2} = 2/3\sqrt{2}$
 - $\langle \mathbf{s}_2, \mathbf{u}_1 \rangle = \int_{-1}^1 t^3 / \sqrt{2} = 0$

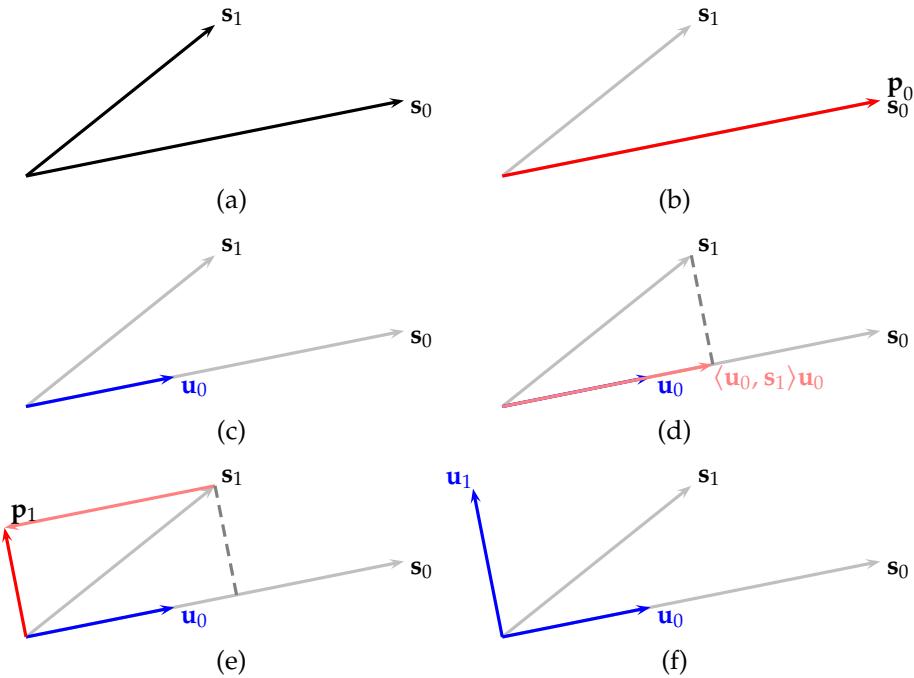


Figure 2.13: The Gram-Schmidt algorithm applied to a basis for \mathbb{R}^2 .

- $\mathbf{p}_2 = \mathbf{s}_2 - (2/3\sqrt{2})\mathbf{u}_0 = t^2 - 1/3$
- $\|\mathbf{p}_2\|^2 = 8/45$
- $\mathbf{u}_2 = \sqrt{5/8}(3t^2 - 1)$
- $m = 3$: $\mathbf{s}_2 = t^3$
- ...

The resulting family of orthonormal basis vectors is known under the name of Legendre polynomials; the basis for $P_6([-1, 1])$ is plotted in Figure 2.14.

Subspace approximation Now that we have an orthonormal polynomial basis for the interval, let's try to approximate a function via subspace projection. We will pick an element from $L_2([-1, 1])$

$$x = \sin t$$

and project it over the basis for $P_3([-1, 1])$ that we computed before. We will then compare the result to another well known linear approximation for the sine, namely its Taylor series expansion.

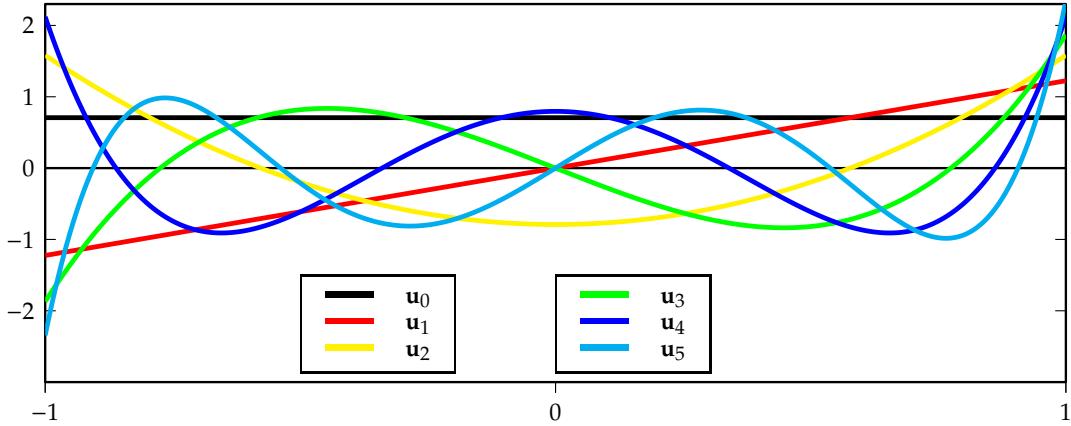


Figure 2.14: The Legendre basis for $P_6([-1, 1])$.

The expansion coefficients over the Legendre basis are readily computed:

$$\alpha_0 = \langle \mathbf{x}, \mathbf{u}_0 \rangle = \langle \sin t, \sqrt{1/2} \rangle = 0 \quad (2.93)$$

$$\alpha_1 = \langle \mathbf{x}, \mathbf{u}_1 \rangle = \langle \sin t, \sqrt{3/2} t \rangle \approx 0.7377 \quad (2.94)$$

$$\alpha_2 = \langle \mathbf{x}, \mathbf{u}_2 \rangle = \langle \sin t, \sqrt{5/8}(3t^2 - 1) \rangle = 0; \quad (2.95)$$

note that we need to explicitly compute just the value for α_1 since, in the other two cases, the inner products are between an even and an odd function, which we know to be orthogonal. In the end, the least-squares approximation to $\sin t$ over $P_3([-1, 1])$ is:

$$\hat{\mathbf{x}} = \alpha_1 \mathbf{u}_1 \approx 0.9035 t. \quad (2.96)$$

By comparison, the second-degree Taylor approximation of the sine around zero is the function $y = t$. Although y and $\hat{\mathbf{x}}$ seem very similar, let's look at the approximation error over the interval as shown in Figure 2.15. Although the Taylor approximation is more precise around the origin, we can see that the Legendre approximation minimizes the *global* error, i.e. the mean square error over the entire interval.

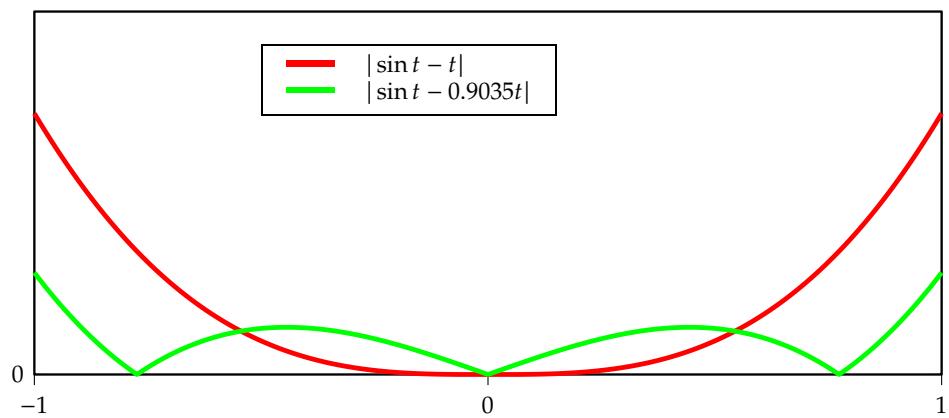


Figure 2.15: Approximation error using the optimal Legendre projection vs Taylor series.